IDENTIFICATION — INVERSE PROBLEMS FOR PARTIAL DIFFERENTIAL EQUATIONS: A STOCHASTIC FORMULATION

A.V. Balakrishnan*

System Science Department; UCLA, USA

1. Introduction

This paper presents a stochastic formulation of a class of identification problems for partial differential equations, known as 'inverse' problems in the mathematical-physics literature. By introducing stochastic processes to model errors in observation as well as 'disturbance' we can provide a precise formulation to interpret what appear to be 'ad hoc' techniques, especially in the treatment of 'inverse' problems. More importantly, we can model unknown sources as stochastic disturbances leading to more general 'inverse' problems than considered hitherto.

Important as inverse problems are in the area of Geophysical applications, where they were studied initially, they would appear to be equally if not more important, in the newer application areas such as modelling and optimization of Environmental Systems, particularly in Water Resources. Diffusion equations arise for instance in stream pollution problems as well as in underground water flow problems, and these models would appear to be well enough founded to attempt system identification. In fact with ever increasing feasibility in high speed and low cost of digital and hybrid computation, the scope of such inverse problems in modelling and simulation is bound to widen even further.

We shall only deal with Cauchy problems for partial differential equations with continuous time observation (as opposed to 'discrete' time). A crucial point then is that the familiar 'Wiener process' modelling of observation errors is unrealistic, as pointed out in [1]. Instead we shall employ a 'white-noise' theory throughout, the relevant notion being explained in Section 2. This involves handling finitely-additive cylinder measures but for linear systems it actually provides a simpler set-up in dealing with partial differential equations. A second feature of our approach is the use of semigroup theory because within the restriction to time-invariant systems (a natural assumption for identification problems) it separates the more general and abstract structural aspects of the problems from the technicalities of the particular partial differential equation set-up involved,

*Research supported in part under AFOSR Grant No. 73-2492, Applied Math Division, USAF.

and retains the similarity to the familiar finite-dimensional formulation as much as possible.

2. The White Noise Process:

Let H denote a real separable Hilbert space and let

$$W = L_2((0,T);H)$$

so that W is a real separable Hilbert space also. Introduce the Gauss measure μ on W on the ring of cylinder sets. This measure is completely described by specifying that for any h in W, with ω denoting 'points' in W:

$$\int_H \exp i[\omega,h] \quad d\mu \quad = \quad \exp(-||h||^2/2)$$

For each ω, let

$$n(t,\omega) \quad = \quad \omega(t)$$

where of course $\omega(t)$ is only defined a.e. We call the 'process' $n(t,\omega)$ 'white noise'. More generally we shall continue to call

$$L\omega$$

white noise again, if L is a linear bounded transformation mapping W into another Hilbert space, and LL^* has a bounded inverse. Not every Borel measurable function $f(\omega)$ (mapping W into a Hilbert space H_1) is a random variable in the usual sense since μ being only finitely additive, need not be extendable as a countably additive measure on every sub-sigma-algebra of Borel sets. We shall mention only two classes of function where this is possible, (and is of significance to us); referring to the work of Gross [2], and [3] for further information on the subject. The first function is a linear transformation: $f(\omega) = L\omega$. This is a random variable in the sense that μ can be extended to be countably additive on the inverse images of Borel sets, if and only if L is Hilbert-Schmidt, and in that case

$$E(||L\omega||^2) = \text{Tr. } LL^* < \infty$$

The second functional is the homogeneous polynomial of degree two, scalar valued:

$$f(\omega) = [L\omega,\omega]$$

This is a random variable if and only if $(L+L^*)$ is nuclear. See [3]. We shall also need to use the theory of estimation involving linear random variables and stochastic semigroup equations; see [3], [7] for both of these topics.

3. A Class of Inverse Problems

Our point of departure will be a particular class of inverse problems for linear p.d.e., studied in depth by Lavrentiev et al [5]. We quote it with a slight modification in notation: Given the telegraphist's equation:

$$\left.\begin{aligned} &\frac{\partial^2 f}{\partial t^2} = \Delta f + a(\cdot)f + u(t,.) \text{ in}\Omega;\ t > 0 \\ &\quad f(0,.) = 0 = \frac{\partial f}{\partial t}(0,.) \\ &\Omega = [x,y,z] \text{ with } z > 0 \\ &\frac{\partial f}{\partial z}(t,.) = 0\Big|_{z=0} \end{aligned}\right\} \qquad (3.1)$$

The 'inverse' problem is to determine the unknown function $a(\cdot)$, given the solution of (3.1) $f(t,x_1,y_1,z_1)$, at the point (x_1,y_1,z_1), and the input $u(t,.)$. [Alternately, we may wish to determine a functional on $f(t,.)$ following Marchuk [6]. This can bring in further simplification.] Their technique (in outline, consult the paper for the details) is to first solve the Cauchy problem with the same initial and boundary conditions:

$$\frac{\partial f_o}{\partial t} = \Delta f_o + u(t,.) \qquad (3.2)$$

and <u>linearize</u> (3.1) about this solution to obtain, setting

$$h = f - f_o \qquad (3.3)$$

the linear equation (linear also in $a(\cdot)$):

$$\frac{\partial^2 h}{\partial t^2} = \Delta h + a(\cdot)\, f_o(t,.) \qquad (3.4)$$

with the same initial and boundary conditions for h. This linearization technique is also basic in Marchuk's work [6]. We have thus the problem of determining $a(\cdot)$ from (3.4) and the given observation:

$$y(t) = h(t,x_1,y_1,z_1) = f(t,x_1,y_1,z_1) - f_o(t,x_1,y_1,y_1,z_1) \qquad (3.5)$$

They proceed to show that this solution of this problem has the property of uniqueness, and obtain formulas for determining $a(\cdot)$ in special cases and otherwise introduce techniques based on integral geometry.

Our first step is to obtain an abstract formulation of this problem using the theory of semigroups of linear operators. Not only does such a formulation yield a useful measure of generality; it also enables us to see more clearly the relationship of this 'inverse' problem to the 'identification' problem as studied in the engineering literature. As is well known, we can recast (3.1) as an 'ordinary' differential equation in the Hilbert space $H = L_2(\Omega)$ as follows (see [7] for an elementary exposition):

$$\dot{x}(t) = Ax(t) + Bx(t) + u(t) \tag{3.6}$$

where A is the infinitesimal generator of a strongly continuous semigroup S(t) (corresponding to the Cauchy problem with homogeneous boundary conditions) u(t) denoting the function u(t,.) as an element of H for each t; x(0) = 0, and finally, B is a linear bounded transformation of H into H that is to be determined given u(t), $0 \leq t \leq T$ and the 'observation'

$$f(t;\ x_1,y_1,z_1) = Cx(t) \qquad 0 < t < T$$

In this example C is a linear transformation defined on the subdomain of continuous functions of H and is an unbounded, uncloseable operator, and hence our natural inclination to make C to be a bounded (continuous) operator has to be revised. On the other hand one can adduce 'physical' arguments for making C bounded; we may argue that any measuring device or instrument must correspond to a spatial smoothing and that 'pointwise' measurements are impossible. This view is shared by Marchuk [6]. Fortunately it is possible to construct a theory in which measurement at a point (or points) of Ω or its boundary can be permitted. Thus we assume that C is unbounded but that for each x in H.

$$S(t)x \in \cdot\ D(C) \qquad \text{for } t > 0 \ \text{ a.e.}$$

and that for $h(\cdot)$ in $L_2([0,T];H) = W$

$$\int_0^t S(t-\sigma)h(\sigma)d\sigma \in D(C) \qquad \text{a.e.} \quad 0 < t < T$$

and that

$$C\int_0^t S(t-\sigma)h(\sigma)d\sigma = \int_0^t CS(t-\sigma)h(\sigma)d\sigma$$

and the righthand side defines a linear bounded transformation of W into

$$W_o = L_2([0,T];H_o)$$

where H_o is the range space of C and at least in theory can be a Hilbert space <u>not</u> necessarily finite dimensional. Needless to say, this assumption is satisfied in the present example. Let us next examine the method of solution of Lavrentiev [5]. We define $x_o(t)$ by:

$$\dot{x}_o(t) = Ax_o(t) + Bx_o(t) + u(t);\ x_o(0) = 0$$

Then defining

$$z(t) = x(t) - x_o(t)$$

we note that

$$\dot{z}(t) = Az(t) + Bx_o(t) + Bz(t)$$

Following the 'linearization' technique now means that we omit the third term $Bz(t)$ or redefine $z(t)$ by:

$$\dot{z}(t) = Az(t) + Bx_o(t);\ z(0) = 0 \tag{3.7}$$

and since $x_o(t) \varepsilon D(C)$ a.e., so does $z(t)$ and we obtain:

$$y(t) = Cz(t) \quad \text{a.e.} \tag{3.8}$$

The problem is that of determining B from (3.7) and (3.8). Thanks to the linearization we have:

$$y(t) = \int_o^t CS(t-\sigma)Bx_o(\sigma)d\sigma \qquad \text{a.e.} \quad 0 < t < T \tag{3.9}$$

We can clearly see that this is expressible as:

$$y = LB$$

where L is a linear bounded transformation mapping the Banach space $E(H)$, of linear bounded transformations on H into H, into W_o. The uniqueness theorem of Lavrentiev is then equivalent to saying that zero is not in the point spectrum of L. One method of determining B is the familiar least squares technique:

$$\text{Minimize} \quad ||y - LB||^2 \tag{3.10}$$

where B ranges over a known subspace of $E(H)$. The important point here is that for any 'practical' algorithm the subspace has to be finite dimensional. Let M denote the closed subspace in which B is known to lie. Let L^* denote the adjoint of L with respect to M. Then denoting by B the minimizing solution (assuming that one exists in M) we can write the familiar 'pseudo-inverse'

$$B = (L^*L)^{-1}\ L^*y \tag{3.11}$$

An important point to note is the appearance of the adjoint operation, a feature emphasized (properly) by Marchuk [6].

Before we go on to the stochastic formulation let us briefly examine a second example of the inverse problem to indicate how to include non-homogeneous boundary conditions. This is also a problem considered by Lavrentiev [5, p. 39], except that we shall change the domain. Thus let Ω denote the sphere in R^3:

$$x^2 + y^2 + z^2 < a^2$$

and consider the heat conduction problem:

$$\frac{\partial f}{\partial t} = \Delta f + u(t)\ g(x,y,z) \qquad t > 0,\ \text{in } \Omega$$
$$f(t,.)\Big|_{\Gamma} = h(t); \qquad f(0;\ x,y,z) = 0 \tag{3.12}$$

where Γ is the boundary of Ω and $h(t)$ is an element of $H_\Gamma = L_2(\Gamma)$, a.e. in t, and $h(\cdot)$ is an element of $W = L_2([0,T];\ H_\Gamma)$. Let Γ_o denote the 2-sphere

$$x^2 + y^2 + z^2 = c^2 \qquad c^2 < a^2$$

Then the 'observation' or 'measurement' is the function:

$$f(t;x,y,z), \qquad (x,y,z)\varepsilon\Gamma_o$$

The inverse problem is to determine the function $g(x,y,z)$ given the scalar function $u(t)$ and the observation. This is a simpler problem in that it is actually 'linear' as we shall see. To obtain the abstract formulation we begin with the Cauchy problem corresponding to the homogeneous boundary value problem:

$$\frac{\partial f}{\partial t} = \Delta f\ ; \quad f(t,.)\Big|_{\Gamma} = 0\ ; \quad f(0;.) \text{ given}$$

Let $S(t)$ denote the semigroup and A the generator with denoting $L_2(\Omega)$ as before. Next let us note that the domain is such that the Dirichlet problem:

$$\Delta f = 0\ ; \quad f(\cdot)\Big|_{\Gamma} = h$$

where h is an element of H_Γ, has a unique solution given by

$$f = Dh$$

where D is a bounded linear operator mapping H_Γ into H. We shall call D the Dirichlet operator. The solution to (3.12) can now be expressed (see [4] for details):

$$x(t) = -\int_0^t AS(t-\sigma)D\ h(\sigma)d\sigma + \int_0^t S(t-\sigma)\ Bu(\sigma)d\sigma \qquad \text{a.e.} \tag{3.13}$$

where B denotes the element $g(\cdot)$ assumed to be in H, and we stress that $x(t)$ is defined only a.e. in general. Next let C denote the mapping corresponding to the observation. C is clearly defined on the subspace of continuous functions in H and maps this subspace into $H_o = L_2(\Gamma_o)$. Also C satisfies the conditions imposed earlier in the abstract version of the first problem. Thus we can write for the observation:

$$y(t) = -\int_0^t CAS(t-\sigma)\ Dh(\sigma)d\sigma + \int_0^t CS(t-\sigma)\ Bu(\sigma)d\sigma \qquad \text{a.e.,} \tag{3.14}$$

and each term on the right is an element of $W_o = L_2([0,T];H_o)$. We need to determine B from (3.14) knowing $h(\cdot)$ and $u(\cdot)$. No linearization is required in this problem since (3.14) is already linear in B. Moreover writing LB as before to denote the mapping given by

$$\int_0^t CS(t-\sigma)\ Bu(\sigma)d\sigma$$

we need to solve the linear equation

$$LB = r$$

where $r(\cdot)$ denotes the function:

$$r(t) = y(t) + \int_0^t CAS(t-\sigma)Dh(\sigma)d\sigma$$

and the problem is thus reduced to the one already considered except that no linearizing approximation is needed.

4. Stochastic Formulation

Having established a more general abstract setting using semi-group theory we proceed now to the stochastic formulation. This we shall do in two stages. In the first stage, stochastic aspects will arise primarily by modelling errors in the measurement (the observation) as an additive stochastic term. In practice of course other sources of error such as calibration errors, bias errors may be far more significant; but the point is that these 'systematic' errors are supposed to be known and even after they are corrected for there will always nevertheless remain a random error component which in an electrical instrument will be the shot or thermal noise. It is best modelled as a Gaussian process of large bandwidth; or in theory as 'white noise'.

Let H denote a separable Hilbert space and let $A(\theta)$ parametrised by θ denote a family of infinitesimal generators of semigroups (strongly continuous at the origin) $S(\theta;t)$. Our observation now takes the form:

$$v(t) = Cx(\theta_o;t) + n(t) \qquad 0 \leq t \leq T \tag{4.1}$$

where θ_o is the true parameter value (unknown to the experimenter or observer) and $n(t)$ is a white Gaussian process, and we need to estimate θ_o based on the observation of duration T. The observation $v(t)$ has its range in a Hilbert space H_o, and for each T the function $v(t)$ is such that

$$v(\cdot) \in W_o = L_2((0,T);H_o)$$

The noise process $n(\cdot)$ has its range also in W_o and to say that it is white Gaussian, it suffices to specify the characteristic function of the corresponding weak distribution by: (E denoting expectation, [,] denoting inner product in W_o):

$$E(\exp i[n,h]) = \exp -1/2(d[h,h]), \; h \in W_o \tag{4.2}$$

where d is a positive number denoting the component-by-component error variance. C is a linear, possibly unbounded operator, but subject then to the assumptions placed in the previous section. Finally $x(\theta_o,t)$ satisfies

$$\dot{x}(\theta_o;t) = A(\theta_o)\; x(\theta_o;t) + B(\theta_o)\; u(t) \tag{4.3}$$

$$x(\theta_o;0) = 0$$

where $B(\theta)$ also parametrized by θ is a linear bounded operator, $u(\cdot)$ is the known source or input process

$$u(\cdot) \in W_i = L_2((0,T);H_i)$$

where H_i is a separable Hilbert space and $B(\theta)$ maps H_i into H. We also allow the parameter θ to have its range in a Hilbert space M or in any algorithmic theory in the neighborhood of a known value θ_1. We may also take the view that θ is Gaussian with mean θ_1 and variance (operator) Λ.

The next basic question concerns the notion of a good or optimal estimate. Since the observation is now stochastic any operation on the data will also be stochastic and we resort to the classical stastical estimation theory to settle this. Thus if we assume that θ is an unknown parameter, we shall accept asymptotically unbiased and asymptotically consistent estimates as optimal. As is well known, one way to obtain such estimates is to invoke the '(a posteriori) maximal likelihood' technique. The likelihood functional, or more properly, the Radon-Nikodym derivative of the measure induced by the process $v(\cdot)$ on W_o with respect to the Gauss measure (induced by $n(\cdot)$), is given by:

$$p(T;v/\theta) = \exp - 1/2(1/d)\; \{[Cx(\theta;.),Cx(\theta;.)] - 2[Cx(\theta;.),v]\} \;\ldots \tag{4.4}$$

for each fixed θ, and T, where $x(\theta;t)$ is the unique solution of:

$$\dot{x}(\theta;t) = A(\theta)x(\theta;t) + B(\theta)\; u(t);\; x(\theta;0) = 0$$

Maximizing the 'likelihood' ratio is seen to be identical with the least squares technique. In practice we seek a root of

$$q(T;v/\theta) = \text{Log}\; p(T;v/\theta)$$

in the neighborhood of the known parameter θ_1. Thus we seek a root of

$$[L_1\theta,\; v - Cx(\theta;.)] = 0 \tag{4.5}$$

where the operator L_1 is the Frechet derivative or gradient mapping:

$$L_1\theta = \frac{d}{d\lambda} Cx(\theta_1 + \lambda\theta;.)\Big|_{\lambda=0} , \quad \theta \in M \tag{4.6}$$

(where $Cx(\theta;.)$ is considered as an element of W_o). We note immediately that in the first example in section 3 the linearization technique of Lavrentiev is precisely this if we use the approximation:

$$Cx(\theta;.) \approx Cx(\theta_1;.) + L_1\theta \tag{4.7}$$

so that we are 'solving':

$$[L_1\theta, \ v - Cx(\theta_1;.) - L_1\theta] = 0$$

or

$$L_1^* L_1\theta = L_1^*(v - Cx(\theta_1;.))$$

or alternately we are minimizing:

$$[L_1\theta, \ L_1\theta] - 2 \ [L_1\theta, \ v - Cx(\theta_1;.)]$$

The uniqueness now is guaranteed by the 'non-singularity of $L_1^* L_1$ and we may view the Lavrentiev result as establishing this for the particular problem considered.

The stochastic formulation provides a lot more than merely a justification of the linearization procedures of Lavrentiev and Marchuk. In the first place as also noted by Marchuk independently [6], we can repeat the linearization or more properly, give an iteration technique for fixed T as follows:

$$\theta_{n+1} = \theta_n + (L_n^* L_n)^{-1} \ (v - Cx(\theta_n;.)) \tag{4.8}$$

where L_n is the gradient operator at $\theta = \theta_n$. This is a Newton-Raphson type algorithm which is not limited to linear systems. Moreover we can show the role played by T, the duration of the data. Thus we get first an 'identifiability' criterion:

$$\underset{T \to \infty}{\text{limit}} \ (1/T) \ (L_o^* L_o) \quad \text{is non-singular} \tag{4.9}$$

where L_o is the gradient at $\theta = \theta_o$ for fixed T. Under this condition we can show that for sufficiently large T a root exists in a suitably small neighborhood of the unknown θ_o, and that this root (approximated by the Newton-Raphson algorithm (4.8)) satisfies the conditions of asymptotic unbiassedness and consistency. These considerations are made much simpler in example two of section 3 because of the linear dependence of $x(\theta;.)$ on θ in that case. In other words the gradient L does NOT depend on θ. Moreover in both examples it is possible to obtain differential equations (updating equations) for the estimate as a function of T and actually a degenerate linear Kalman-Bucy type equation in the case of example 2. See [8].

5. Stochastic Formulation Continued: Source Noise

Not only does the stochastic formulation of inverse problems lead to a precise meaning of the algorithms used; but it also allows us to consider a situation which has so far not been considered. This is the inclusion of errors in modelling the source or forcing function or input; or, alternately, the inclusion of sources not accounted for by the experimenter. The lack of introduction of unknown sources is not surprising since as we shall see the estimation algorithms then cannot be given any physical explanation such as 'least squares fit' or other 'wave-form matching' criterion.

We shall present only the mathematical theory here exploiting the semigroup theoretic formulation already developed. Thus the observation has the same form as before:

$$v(t) = Cx(\theta_o;t) + n(t) \tag{5.1}$$

where we use the same notation as section 4 but $x(\theta;t)$ is itself stochastic now defined by the stochastic equation:

$$\dot{x}(\theta_o;t) = A(\theta_o)\, x(\theta_o;t) + B(\theta_o)\, u(t) + F(\theta_o)\, n_s(t) \tag{5.2}$$

where the difference from (4.3) is the third term, which is a Gaussian stochastic process with (for each T)

$$n_s(\cdot) \in W_s = L_2([0,T];H_s)$$

where H_s is a separable Hilbert space and $F(\theta)$ for each θ is a linear bounded operator mapping H_s into H. The process $n_s(\cdot)$ is white Gaussian with

$$E(\exp i[n_s,h]) = \exp - 1/2[h,h]; \qquad h \in W_s$$

and is independent of the observation noise process $n(\cdot)$.

We need to make a basic assumption; that for each θ of interest the process

$$Cx(\theta;.)$$

where $x(\theta;.)$ is defined by

$$\dot{x}(\theta;t) = A(\theta)\, x(\theta;t) + B(\theta)u(t) + F(\theta)\, n_s(t); \; x(\theta;0) = 0 \tag{5.3}$$

is such that the corresponding covariance operator $R(\theta)$ defined by

$$E(\exp i[h,Cx(\theta;.)]) = \exp - 1/2\, [R(\theta)h,h] \tag{5.4}$$

is trace-class (nuclear). In that case we can again invoke the method of maximum likelihood to obtain asymptotically unbiased, consistent estimates.

First of all under the assumption of nuclearity on $R(\theta)$ the (finitely additive) measure induced by the process $v(\cdot)$ (for each assumed θ) is absolutely continuous with respect to the Gauss measure induced by $n(\cdot)$. Hence we may invoke the criterion of maximum likelihood to obtain optimal estimates. Thus the likelihood functional (R-N derivative) is given by:

$$p(T;v/\theta) = \exp - 1/2d \quad [C\hat{x}(\theta;.), C\hat{x}(\theta;.)] - 2[C\hat{x}(\theta;.), v(\cdot)] + 2 \text{ Tr.} \int_0^T C(CP(t))^* dt \qquad \ldots (5.5)$$

where $\hat{x}(\theta;.)$ is the solution of:

$$\hat{x}(\theta;t) = A(\theta)\hat{x}(\theta;t) + B(\theta)u(t) + (CP(t))^* (1/d)(v(t) - Cx(\theta;t)) \; x(\theta;0) = 0$$

$$[\dot{P}(t)x,y] = [P(t)A(\theta)^* x,y] + [P(t)x,A(\theta)^* y] + [Fx,Fy] - (1/d)[CP(t)x,CP(t)y]$$

$P(0) = 0$; x, y in the domain of $A(\theta)^*$

For each T our estimate is the root of the gradient of

$$q(T;v/\theta) = \text{Log } p(T;v/\theta) \qquad (5.6)$$

in a neighborhood of θ_1. The root will exist and be unique in a small enough neighborhood of θ_o under the identifiability condition which is now that

$$\lim_{T \to \infty} (1/T)(L_o^* L_o) \text{ be non-singular,}$$

where L_o is the gradient operator (Frechet derivative):

$$L_o\theta = \frac{d}{d\lambda} Cx(\theta_o + \lambda\theta)\Big|_{\lambda=0} ; \; \theta \in M$$

Interpreting L_n as the similar derivative at $\theta = \theta_n$, the algorithm for estimation for each T now becomes:

$$\theta_{n+1} = \theta_n + (L_n^* L_n)^{-1} (L_n^*(v - C\hat{x}(\theta_n;.) - K_n)$$

where K_n is the Frechet derivative of

$$\text{Tr} \int_0^T C(CP(t))^* \, dt$$

with respect to in M at $\theta = \theta_n$. For a treatment of the case of noise on the boundary, See [4]. Note that in (5.5) the term non-linear in the data:

$$[Cx(\theta;.), v(\cdot)]$$

has the form

$$[L\omega,\omega]$$

where L is trace-class, and is a random variable as noted in Section 2.

REFERENCES

1. A.V. Balakrishnan: "On the approximation of Ito integrals by band-limited processes" SIAM Journal on Control 1974.
2. L. Gross: Harmonic Analysis in a Hilbert Space, American Mathematical Society Memories No. 46, 1963.
3. A.V. Balakrishnan: Stochastic Optimization Theory in Hilbert Spaces International Journal on Applied Mathematics and Optimization, Vol. 1, No. 2, 1974.
4. A.B. Balakrishnan: Identification and Stochastic Control of a Class of Distributed Systems with Boundary Noise, IRIA Conference on Stochastic Control and Computer Modelling, June, 1974. To be published by Springer-Verlag.
5. Lavrentiev, Romanov, Vasiliev: Multidimensional Inverse Problems for differential Equations, Lecture Notes in Mathematics, No. 167, Springer Verlag, 1970.
6. G. Marchuk: Perturbation Theory and the Statement of Inverse Problems, Lecture Notes in Computer Science, Vol. 4, Part II, Springer Verlag, 1973.
7. A.V. Balakrishnan: Introduction to Optimization Theory in a Hilbert space Lecture Notes in OR and Mathematical Systems, Springer-Verlag, Russian version, MIR, Moscow, 1974.
8. A.V. Balakrishnan: "Stochastic System Identification Techniques", in Stochastic Optimization and Control, edited by H.F. Karreman, John Wiley & Sons, 1968.