manuscripta geodaetica (1995) 20:204–230

manuscripta geodaetica

The theory of optimal linear estimation for continuous fields of measurements

Fernando Sansó and Giovanna Sona

Dipartimento I.I.A.R. – Politecnico di Milano, Piazza Leonardo Da Vinci 32, I-20133 Milano, Italy

Received 30 December 1993; Accepted 16 November 1994

Summary

The main problem of linear estimation theory in infinite dimensional spaces is presented and its typical difficulties are illustrated.
The use of Wiener measures to represent continuous observation equations is carefully analysed in relation to the physical description of measurements and to the mathematical limit when the number of observations grows to infinity.
The overdetermined problem is solved by applying the Wiener principle of minimizing the mean square estimation error; the solution is proved to exist and to be unique under very general conditions on the observation operators.
Examples coming from space geodesy, potential theory and image analysis are presented to prove the effectiveness of the method and its applicability in different contexts.

1. Introduction

In this paragraph we would like to discuss the main differences and difficulties encountered when we try to generalize the usual least squares estimation theory to infinite dimensional spaces.
A classical scheme in linear estimation theory is the following: assume the set of measurements to be collected in a vector Y belonging to some linear vector space H_Y, also called the space of observables, with finite dimensions n_Y; the theoretical value y is constrained to belong to some linear (proper) manifold in H_Y, which for instance is described in parametric form as

$$\{ y = Ax + a \; ; \; x \in H_X \}$$

where the vector x ranges in the so-called parameter space H_X, also finite dimensional, with $n_X < n_Y$, a is a constant vector in H_Y, A is a linear operator (matrix) from H_X into H_Y.
The relation

$$y = Ax + a \qquad (1.1)$$

reflects the physics and the geometry of the observational process, i.e. a description of all what is known of the physical and geometric relations between the quantities relevant to the experiment, including those which describe the internal states of the instruments; in (1.1) the vector a represents just a fixed translation of the range of A, $\mathcal{R}_A$, away from the origin of H_Y and it can be eliminated by a coordinate shift, therefore from now on we suppose a = 0 and the manifold of the admissible values will coincide with $\mathcal{R}_A$. The vector of observations Y does not belong to the manifold of the admissible values, because the model (1.1) is imperfect in describing the measurement process, i.e. we rather have an observational model

$$Y = Ax + \nu \qquad (1.2)$$

where the discrepancies, collected in the vector ν, are unknown ; the vector ν has an erratic behaviour so that it is usually described in statistical terms rather than as a deterministic quantity; classical hypotheses for ν are that it is a sample from a zero mean variable,

$$E\{\nu\} = 0 \qquad (1.3)$$

and that its covariance structure is known, in terms of the covariance matrix of the components of ν , i.e.

$$C_{\nu\nu} = E\,\{\nu\nu^{+}\} \quad . \qquad (1.4)$$

The operator A is supposed to be of full rank, i.e.

$$Ax = 0 \;\Rightarrow\; x = 0 \qquad (1.5)$$

so that there is a one to one correspondence between $y \in \mathcal{R}_A$ and $x \in H_X$.

In this case the estimation problem can be formulated as follows: find $y \in \mathcal{R}_A$ (or equivalently find x) so that it is reasonably close to Y.

The closeness concept requires that H_Y be endowed with a norm, and this can take the fairly general form

$$|y_P| = (y^{+}Py)^{1/2} = (\Sigma y_i y_k P_{ik})^{1/2} \qquad (1.6)$$

($P^{+}=P$, $P > 0$, i.e. is positive definite)

with P any symmetric positive matrix; (1.6) is in fact the general form of a norm compatible with a scalar product, i.e. with a Hilbert space structure of H_Y, which will be mostly useful in the sequel.

What is now a reasonable distance between Y and y ?

Without any further specification a sensible choice seems to be the minimum distance between Y and $y \in \mathcal{R}_A$:

$$\hat{y} : \hat{\nu}^{+}P\,\hat{\nu} = |Y - \hat{y}|_P^2 = \min \qquad (1.7)$$

$$(\hat{y} = A\hat{x})$$

this is the least squares (l.s.) criterion in the general deterministic form, leading to the solution

$$\begin{cases} \hat{x} = N^{-1}A^{+}P\,Y \\ \hat{y} = A\,\hat{x} \\ (N = A^{+}P\,A) \end{cases} \qquad (1.8)$$

This solution supplies an unbiased estimate (in fact a uniformly unbiased estimate, i.e. $E\{\hat{y}\}=y$, $\forall\, y = E\{Y\} \in \mathcal{R}_A$), however it has the disadvantage to depend on the metric, P.

The deterministic l.s. approach leaves open a norm choice problem, which can be "solved" only by a full exploitation of our knowledge of the stochastic behaviour of ν, contained in $C_{\nu\nu}$.

In fact from (1.2), (1.8), after applying the proper covariance propagation, we know that

$$C_{\hat{x}\hat{x}} = N^{-1}A^{+}P\,C_{\nu\nu}P\,A\,N^{-1} \quad . \qquad (1.9)$$

Since the covariance matrix represents in a sense the spread of the distribution of the vector $\hat{x}$ it is natural to search for a $C_{\hat{x}\hat{x}}$ as small as possible, for then the probability of finding $\hat{x}$ close to x will increase.

It is a remarkable and peculiar result of l.s. theory that the family (1.9) admits an extremal point[1], in fact a set of matrices, even if they are positive definite, need not to admit a minimum: in our case however this happens, for instance for

$$P = \lambda\, C_{\nu\nu}^{-1} \qquad (1.10)$$

<u>Remark 1.1</u>

In literature it is not usually stressed very much that this minimum is not unique, and not only for the arbitrariness of the factor λ, in fact for instance any other metric associated with a matrix

$$P = C_{\nu\nu}^{-1} + \Lambda \quad , \qquad (1.11)$$

with Λ symmetric, positive and such that

$$\Lambda\, A = 0 \; ,$$

will also cause (1.9) to achieve the same value, $C_{\hat{x}\hat{x}}$, and hence the

---

[1] A matrix A is said to be smaller than B if B - A is a positive definite matrix i.e. if $B - A \geq 0$; A is the minimum of a family of matrices $\{A_\alpha\}$ indexed by the general index α, if A belongs to $\{A_\alpha\}$ and $A_\alpha - A \geq 0$ uniformly in α.

minimum.

Remark 1.2

By minimizing $C_{\hat{x}\hat{x}}$ we also minimize the covariance matrix of any other linear function of $\hat{x}$, for instance that of $\hat{y} = A\hat{x}$, namely

$$C_{\hat{y}\hat{y}} = A\, C_{\hat{x}\hat{x}}\, A^{+} \quad . \tag{1.12}$$

The above discussion is somehow the core of least squares theory, providing the bridge between the deterministic approach (i.e. the idea of minimizing $|Y - y|_P$ whatever is Y not belonging to $\mathcal{R}_A$) and the stochastic approach which consists in searching for the best unbiased estimator (i.e. the one minimizing the covariance matrix $C_{\hat{x}\hat{x}}$) in the class of linear functions of the random variable Y. In fact we can summarize this crucial point of Gauss-Markov theory by the following

Theorem 1.1

The least squares estimators $\hat{x}, \hat{y}$ (for(1.8)) for the problem (1.2), (1.3), (1.4), coincide with the best linear unbiased estimators (BLUE) if the metric in H_Y is associated with the positive symmetric matrix

$$P = \lambda\, C^{-1}_{\nu\nu} \quad . \tag{1.13}$$

This is also the point which is not possible to transfer to the infinite dimensional case, so that, as we shall show in a moment, a l.s. estimation principle cannot provide any more an optimal linear estimator; however, fortunately enough, it is still possible to construct best linear unbiased estimators which give answers to many problems of great interest, in particular when we try to model continuous fields of measurements. Naturally this problem is not new in literature and it has been studied in conjunction with underdetermined problems by applying all the machinery of estimation theory or, more recently, concepts from theoretical informatics (cfr. Dermanis 1991, Backus 1970a, 1970b, 1970c, Tarantola 1987, Sanso' 1990); yet a clear and rigorous solution of the purely overdetermined problems in infinite dimensional spaces is unknown to the authors.

The situation we would like to analyse is as follows: assume we have a set of measurements ordered by the continuous parameter $t \in T$; t can be a time but it could also be a point of $\mathbb{R}^n$ or it could range over a manifold, e.g. on a sphere; accordingly T can be the whole real line or part of it, $\mathbb{R}^n$ or a subset of it, the whole sphere S or a part of it etc.

Hence the set of observations will be (for the moment) considered as a function $Y(t)$, $t \in T$ and we will assume it belongs to some Hilbert space H_Y.

On the other hand $Y(t)$ will always be split into a regular part, or theoretical value $y(t)$, and a disturbance $\nu(t)$ described in terms of stochastic variables, i.e. a stochastic process: it is to be stressed that, with this specification, claiming $Y \in H_Y$ means that naturally $y(t) \in H_Y$ and $\{\nu(t)\}$ has a realization in H_Y with probability one (Pr=1).

The theoretical value $y(t)$ is again supposed to belong to some subspace that we represent in parametric form as the range of an operator A

$$\begin{cases} y = Ax \\ x \in H_X \end{cases} \tag{1.14}$$

In order to avoid complications it is useful to assume that A is an operator from $H_X \to \mathcal{R}_A \subset H_Y$ which is also a bijection (i.e. it is into and onto); this means on one side that

$$A\, x = 0 \quad \Rightarrow \quad x = 0 \quad , \tag{1.15}$$

i.e. the problem is not rank deficient, on the other side we expect

$$\mathcal{R}_A = \{A\, x \ ,\ x \in H_X\} \tag{1.16}$$

to be a closed set in H_Y, i.e. a (proper) subspace of H_Y; this last requirement is essentially a condition on the topology of H_X in relation to the topology of H_Y and to the operator A, and to fulfil it we shall assume that A is a continuous operator from H_X into H_Y.

Remark 1.3

We mention that a "natural" topology for H_X is

$$\| x \|_X \equiv \| A x \|_Y \qquad (1.17)$$

which is indeed a norm as a consequence of (1.15).
With this topology in fact A is an isometry between H_X and $\mathcal{R}_A$.

As for the disturbance $\{\nu\}$, we suppose that it is a stochastic process, with zero mean, i.e. we assume the model $Y = y + \nu$ to be unbiased,

$$E\{\nu\} = 0 \quad , \qquad (1.18)$$

and with such a regular distribution as to admit a covariance operator, i.e.

$$E\{\langle w,\nu\rangle_Y^2\} = \langle w,Cw\rangle_Y \ . \qquad (1.19)$$

It is possible to see that with the definition (1.19), and the subsequent

$$E\{\langle w,\nu\rangle_Y \langle u,\nu\rangle_Y\} = \langle w,Cu\rangle_Y \ , \qquad (1.20)$$

the operator C turns out to be selfadjoint and positive definite; furthermore if we make the reasonable assumption that the disturbance ν affects all directions in H_Y (for otherwise there would be a direction $\tilde{y}$ along which we can make perfectly errorless measurements) we find that C is strictly positive definite, i.e.

$$C w = 0 \Rightarrow w = 0 \quad . \qquad (1.21)$$

Remark 1.4

Let us further assume that H_Y is chosen in such a way as to contain ν and that $\|\nu\|^2_{H_Y}$ has a finite expectation; in this case C is not only positive and selfadjoint, but compact and it admits a spectral decomposition

$$C e_n = \lambda_n e_n \qquad (1.22)$$

with $\{e_n\}$ a complete orthonormal system in H_Y and λ_n real positive eigenvalues satisfying the Hilbert-Schmidt condition

$$\sum \lambda_n < +\infty \ .$$

According to (1.20) we see that putting

$$\nu_n = \langle \nu, e_n\rangle_Y \qquad (1.23)$$

we have

$$\sigma^2(\nu_n) = \langle e_n, Ce_n\rangle_Y = \lambda_n \ , \qquad (1.24)$$

showing that $\{\lambda_n\}$ are the variances of ν in the directions of $\{e_n\}$.

At this point it might seem straightforward to follow the same way of reasoning as for $\mathbb{R}^n$ and introduce in H_Y a possibly more restrictive norm

$$H_C \ ; \ u \to \langle u, C^{-1}u\rangle_Y = \sum \frac{u_n^2}{\lambda_n} \qquad (1.25)$$

and perform the minimization of $\|\nu\|_C^2 = \|Y - A x\|_C^2$: however this is not possible because in general ν does not belong any more to H_C with Pr = 1.

It is sufficient to show a counterexample, in fact one of particular importance (in this respect cfr. also the discussion in Sacerdote and Sanso', 1985):

Example 1.1

Let ν be normal and assume $\nu \in H_Y$ with Pr = 1 ; then

$$\nu_n = \langle \nu, e_n\rangle_Y \sim N[0,\lambda_n] \qquad (1.26)$$

so that

$$\|\nu\|_C^2 = \lim_{N\to\infty} \sum_1^N \frac{\nu_n^2}{\lambda_n} = \lim_{N\to\infty} \chi_N^2 \qquad (1.27)$$

and

$$\Pr\{ \|\nu\|_C^2 < +\infty\} = \Pr\{\lim_{N\to\infty} \sum_1^N \frac{\nu_n^2}{\lambda_n} < +\infty\} =$$

$$= \Pr\{\lim \chi_N^2 < +\infty \} \qquad (1.28)$$

On the other hand

$$\Pr\{\lim \chi_N^2 < +\infty\} = 1 - \Pr\{\lim \chi_N^2 = +\infty\} \le$$

$$\le 1 - \Pr\{\chi_{\bar{N}}^2 > L\} \qquad \forall\ \bar{N},\ L \quad (1.29)$$

by fixing L and letting $\bar{N} \to \infty$, since

$$\Pr\{\ \chi_{\bar{N}}^2 > L\ \} \to 1 \quad ,$$

we derive the result that

$$\Pr\{\ \|\nu\|_C^2 < +\infty\ \} = 0 \quad .$$

Remark 1.5

The above example is a particular case of a more general statement which can be presented in the form of a negative theorem:

Theorem 1.2

It is not possible to construct an isotropic distribution on a Hilbert space H, i.e. it is not possible that at the same time

$$\Pr\ \{\nu \in H\} = 1 \qquad (1.29)$$

and the variables

$$<\nu, e>_H = \nu(e) \qquad (1.30)$$

be identically distributed whatever is e ($\|e\|_H = 1$), without being $\nu \equiv 0$.

In fact, excluding the trivial case $\nu \equiv 0$, (1.29) implies for any o.n. sequence $\{e_n\}$

$$\Pr\ \{\ \sum \nu_n^2 < +\infty\ \} = 1$$

i.e.

$$\Pr\ \{\lim_{n\to\infty} \nu_n = 0\ \} = 1 \quad ; \qquad (1.31)$$

(1.31) means that $\nu_n \to 0$ in Pr almost everywhere, what in turn implies convergence in probability and also convergence in law (see Papoulis 1965). Since the limit is 0, i.e. a variable ω with distribution P_0

$$\begin{cases} P_0\{\omega = 0\} = 1 \\ P_0\{\omega \ne 0\} = 0 \end{cases} ,$$

then the distributions P_n of ν_n must satisfy

$$\lim_{n\to\infty} P_n = P_0 \quad , \qquad (1.32)$$

the convergence being in the weak sense for measures.
On the other hand if P_n are to be all identical we must also have

$$P_n \equiv P_0 \quad \Rightarrow \quad \Pr\ \{\nu_n = 0\} = 1 \quad ,$$

i.e. $\nu \equiv 0$ with Pr = 1, what we have excluded from the beginning.

The above statements break the similarity with the finite dimensional case, so that we are left with the best linear estimation theory, as the only possible approach, which, as we will show in the next paragraphs, is indeed successful.
To this aim, in fact, the only thing we need is to give some meaning to the coupling between ν and $w \in H_Y$, i.e. to the symbol

$$\nu(w) = <\nu, w>_Y$$

in spite of the fact that $\nu \notin H_Y$, because for instance $\{\nu(e_n)\}$ are identically distributed normal variables.
This is possible indeed by using the concepts of generalized stochastic processes and of bounded stochastic functionals; these are generalizations developed from the basic concept of Wiener integral (cfr. Hida T., 1980; Ito R., 1984; Lamperti J., 1977), to which we shall refer mainly in the sequel because its construction is close to the physical modelling of continuous fields of measurements.
So we shall work mainly with Wiener measures modelling essentially a type of independent white noises ν, leaving the generalizations, which seem straightforward, to future works.
Some work in this direction is indeed already present in geodetic literature (cf.Sanso', 1988; Keller, 1989), however here we try to give a systematic and general solution to the problem.

2. Measurements and Wiener measures

In this paragraph we try to show how Wiener measures arise in a natural way as mathematical tools describing the limit of a discrete configuration; the bases of the theory of Wiener integral are also briefly reviewed.(cfr. also Sanso' 1988)

The observations of a "smooth" function y(t) (t being in general a vector variable ranging in a D-dimensional region T) at sample points $\{t_i\}$ can be modelled as

$$Y_i = y(t_i) + \nu_i \qquad (i=1,2....N) \quad (2.1)$$

and we shall assume ν_i to be independent Gaussian noises with the same variance σ^2_ν,

$$E\{\nu_i\} = 0 \quad , \quad E\{\nu_i \nu_k\} = \delta_{ik}\sigma^2_\nu \ . \quad (2.2)$$

Why do we want to represent the situation described in (2.1) and (2.2) by a continuous limit?

The answer is that it is possible in many cases nowadays to obtain a large amount of measurements, so that depending on the spectral structure of y(t), we might be able to obtain even more information than the resolution sought for our function.

Accordingly it might be more meaningful to perform approximate computations with the continuous model, rather than throwing away information or using it piecewise to prevent our computations from blowing up.

The next crucial question is: under which conditions can we make a continuous model of (2.1)?

We will come back later to this question, however we can guess a first answer: when the distance between the points t_i is so small that y(t) has no significant variation between neighbouring points, when we compare it with the amplitude of ν.

Under these circumstances the information carried by the measurements is not so much related precisely to the points where they are taken but rather to the way they accumulate in that particular region, i.e. to their density.

That is the same reason why we describe the matter as a continuum, for those phenomena which have no significant fluctuations on a microscopic scale.

The third point to be stressed is that when the unknowns to be estimated are vectors in infinite-dimensional spaces (as it is with y(t)) there is a substantial difference between estimating a functional of the unknown or the whole vector. The following example is a very typical illustration of this statement.

Example 2.1

Assume y(t) to be in $L^2(0,2\pi)$ and that it is possible to measure with a suitable instrument the Fourier coefficients $\{a_n, b_n\}$ of y(t) expressed as

$$a_n = \frac{1+\delta_{n0}}{\pi}\int_0^{2\pi} y(t)\cos nt \, dt$$

$$b_n = \frac{1}{\pi}\int_0^{2\pi} y(t)\sin nt \, dt \quad . \qquad (2.3)$$

The measured quantities are by hypothesis

$$a_{on} = a_n + \nu_{an} \quad ,$$

$$b_{on} = b_n + \nu_{bn} \qquad (2.4)$$

with ν_{an}, ν_{bn} all normal independent noises with the same variance σ^2_ν.

Hence, although we are able to estimate with a_{on}, b_{on} a complete set of functionals of y(t), we will never be able to reconstruct an unbiased estimate of y(t) because the formal series

$$\sum (a_n \cos nt + b_n \sin nt)$$

would have L^2 error norm

$$\sum (\nu^2_{an} + \nu^2_{bn})$$

to which the reasoning of Example 1.1 applies.

After these preparatory discussions we go back to (2.1) and try to see how we can construct a meaningful limit, for $N \to \infty$, of the scheme (2.1),(2.2) .

To this aim we want to rewrite (2.1) as a relation between discrete measures; first we split the region T where the measurements are taken into non-overlapping cells $\{T_k\}$ such that one and only one t_k falls in T_k; we also put

$$\begin{aligned} \mu_k &= \text{meas } \{T_k\} \quad \text{(Lebesgue meas.)} \\ \delta_k &= \text{diam } \{T_k\} \end{aligned} \tag{2.5}$$

We must note that when N changes also the partition $\{T_k\}$ has to change, so sometimes we might write T_k^N instead of T_k and μ_k^N instead of μ_k etc, in order to make explicit the dependence on N.
In order to perform a limit for $N \to \infty$ we must make a choice on the way the points $\{t_i\}$ are chosen : we assume that for $N \to \infty$ we have for any measurable set A,

$$\lim_{N\to\infty} \frac{N_A}{N} = \int_A \rho(t)\, dt \quad , \tag{2.6}$$

where N_A is the number of points $\{t_i\}$ falling in A; the function $\rho(t)$, which is positive on T, represents a relative density of measurements and clearly satisfies the relation $\int_T \rho(t)\, dt = 1$.

Thinking of Bernoulli's theorem, it is clear that (2.6) can be realized for instance in terms of independent samplings from the distribution $\rho(t)$, the limit in this case being understood in a stochastic sense.
If we rewrite (2.6), for a sufficiently small A, in the approximate form

$$\frac{N_A}{N} \sim \rho(t_A)\, \mu_A \qquad (t_A \in A)$$

we see that in the region A the measure of T_i will be about

$$\mu_i \sim \frac{\mu_A}{N_A} \sim \frac{1}{\rho(t_i)N} \tag{2.7}$$

Now we construct a discrete measure μ_{YN} supported by the points $\{t_i\}$ such that

$$\mu_{YN}(t_i) = Y_i \mu_i \quad ; \tag{2.8}$$

in this way we associate with any Lebesgue measurable set A the measure

$$\mu_{YN}(A) = \sum_{t_i \in A} Y_i \mu_i \tag{2.9}$$

This μ_{YN} is our prototype of stochastic measure, associated with a field of measurements; it is in fact a collection of stochastic variables, indexed by the measurable sets $\{A\}$, which, owing to (2.1), can be written in the form

$$\mu_{YN} = \mu_{yN} + \mu_{wN} \tag{2.10}$$

with

$$\mu_{yN}(A) = \sum_{t_i \in A} y(t_i)\mu_i \tag{2.11}$$

$$\mu_{wN}(A) = \sum_{t_i \in A} \nu_i \mu_i . \tag{2.12}$$

Due to our hypotheses (2.2) it is clear that

$$E\, \{\mu_{YN}(A)\} = \mu_{yN}(A) \quad \forall\, A \tag{2.13}$$

while, for every measurable A,

$$E\, \{\mu_{wN}(A)\} = 0 \tag{2.14}$$

$$E\, \{\mu_{wN}(A)^2\} = \sigma_\nu^2 \cdot \sum_{t_i \in A} \mu_i^2 \quad ; \tag{2.15}$$

μ_{wN} is our prototype of a discrete Wiener measure.

<u>Remark 2.1</u>

In writing the observation equations in the form (2.10) we have introduced an element of arbitrariness related to the choice of the partition $\{T_i\}$ and this might seem dangerous in the construction of a sound theory; on the other hand it is clear that what we want to achieve is a limit of our scheme such that it is representative of our discrete reality (i.e. close to it in a suitable sense), so it will be enough to show that the limit model we arrive at, by letting

$$\delta = \max \delta_i \to 0 ,$$

is indeed independent on the way the partition $\{T_i\}$ is chosen.

We will not prove this here, but, just to fix our ideas, since it will be useful to maintain the oscillation of

$y(t)$ over each T_i as small as possible, we can think of $\{T_i\}$ as a partition such that $\Sigma_i \delta_i$ = min or other similar conditions; for instance $\{T_i\}$ could very well be a triangulation of minimal perimeter.

Remark 2.2

Since we want to take the limit of μ_{YN} for $N \to \infty$, we must be prepared to define in which sense this limit has to be understood.
As for μ_{yN} there are no difficulties: for each A, $\mu_{yN}(A)$ is a sequence of constants, so its limit is the usual limit in $\mathbb{R}$ and it is simply (cfr. (2.11))

$$\lim_{N\to\infty} \mu_{yN}(A) = \int_A y(t)dt \quad ; \qquad (2.16)$$

since $y(t)$ has been assumed for the moment to be smooth, e.g. continuous, (2.16) is in fact just the definition of the integral in the Peano sense.
For $\mu_{wN}(A)$ things are slightly more complicated; this is a sequence of stochastic variables with zero mean and finite variance and one way of defining their limit is in mean square sense, i.e. we want to say that

$$\mu_w(A) = \lim_{N\to\infty} \mu_{wN}(A) \qquad (2.17)$$

in the sense that

$$E\{[\mu_w(A) - \mu_{wN}(A)]^2\} \to 0 \qquad (2.18)$$

To make this possible we must first be sure that $E\{\mu_{wN}(A)^2\}$ be bounded, or even better that it has a reasonable limit for every A.
For instance if we had taken $\sigma_\nu^2 = C$ (constant) then we would immediately have ($\varepsilon = \max \mu_i$)

$$E\{\mu_{wN}^2(A)\} \le C\varepsilon \sum \mu_i = C\varepsilon\mu(A) \to 0; \quad (2.19)$$

although surprising it is perfectly reasonable that, if we measure with higher and higher resolution and constant variance a continuous function, we tend to have an infinite number of observations even on smaller and smaller intervals so that by applying the $1/\sqrt{N}$ law to the mean we can achieve a perfect knowledge of the point value of $y(t)$ as the limit of a sequence of mean values.

This tells us that the variances of $\{\nu_i^N\}$ have to be variable with N if we want to receive a limit where the variability of the noise, associated with a set A of positive measure is positive too.
It is easy to recognize that a good choice is the following: let $\{\nu_{oi}$, $i=1,2..\}$ be an infinite sequence sampled independently from the same distribution $N[0,\sigma_0^2]$ and let us stipulate that

$$\nu_i^N = \sqrt{N}\, \nu_{oi} \quad ; \qquad (2.20)$$

in this way we have

$$(\sigma_\nu^N)^2 = N\sigma_0^2 \qquad (\sigma_0 = \text{constant})$$

and from (2.15) and (2.7) we find

$$E\{\mu_{wN}^2(A)\} = \sigma_0^2 \sum_{t_i \in A} (N\mu_i)\mu_i \cong$$

$$\cong \sigma_0^2 \sum_{t_i \in A} \frac{\mu_i}{\rho(t_i)} \to \sigma_0^2 \int_A \frac{dt}{\rho(t)} \qquad (2.21)$$

Before refining the mathematics of this reasoning we must answer an important question: isn't it too artificial, from a physical point of view, to let the variance of noise to tend to infinity so as to obtain a finite positive limit?
In our opinion the answer is no. In fact assume for instance ρ to be constant on A so that $\bar{\rho}\mu_A = N_A/N$; then the points t_i are uniformly distributed in A and we can take T_i to be of constant measure $\mu_i = \mu_A/N_A$, so that

$$E\{\mu_{wN}^2(A)\} = \sigma_\nu^2 \frac{\mu_A^2}{N_A} = \frac{\sigma_\nu^2}{N} \frac{\mu_A}{\bar{\rho}} \quad . \qquad (2.22)$$

Now we can easily understand how the limit can be taken in such a way that (2.22) remains constant: it is enough that $\bar{\rho}$ remains constant (e.g. we double

N_A every time we double N) and on the same time σ_ν^2/N remains constant. This corresponds essentially to the idea that we increase fictitiously the number of points in A, N_A, but on the same time we increase the variance so that the total information contained in A (or, if you like, its random variability) is kept constant.
In order to make this approach acceptable we need therefore two results: first to prove that indeed μ_{wN} tends stochastically to a random measure, that we will call Wiener measure, second to give some condition under which the continuous limit can be considered as a good model for the discrete reality.
So we shall first sketch the proof of the following theorem

<u>Theorem 2.1</u>

The sequence $\mu_{wN}(A)$ is stochastically convergent in mean square sense for any measurable set A and we can define its limit $\mu_w(A)$ as the Wiener measure of A.

We need to prove convergency only for the whole T because the set function $E\{|\mu_{wN+p}(A) - \mu_{wN}(A)|^2\}$ is indeed increasing with A.
We have

$$\mu_{wN+p}(T) - \mu_{wN}(T) =$$

$$= \sum_{1}^{N} {}_i \nu_{oi}\sqrt{N+p}\, \mu_i^{N+p}\left(1-\frac{\sqrt{N}\,\mu_i^N}{\sqrt{N+p}\,\mu_i^{N+p}}\right) +$$

$$+ \sum_{N+1}^{N+p} {}_i \nu_{oi}\sqrt{N+p}\, \mu_i^{N+p} = M_1 + M_2 \qquad (2.23)$$

Let us consider the first term. First we note that, recalling (2.7),

$$\left|1 - \frac{\sqrt{N}\,\mu_i^N}{\sqrt{N+p}\,\mu_i^{N+p}}\right| \cong$$

$$\cong \left|1 - \sqrt{\frac{N+p}{N}}\right| \leq \frac{p}{N}\,; \qquad (2.24)$$

then we have

$$E\{M_1^2\} \leq \sigma_0^2 \sum_{1}^{N} {}_i (N+p)(\mu_i^{N+p})^2 \left(\frac{p}{N}\right)^2 \cong$$

$$\cong \sigma_0^2 \left(\sum_{1}^{N} {}_i \frac{\mu_i^{N+p}}{\rho(t_i)}\right)\left(\frac{P}{N}\right)^2 \qquad (2.25)$$

and since $\sum_{1}^{N+p} {}_i \frac{\mu_i^{N+p}}{\rho(t_i)}$ is convergent according to (2.21), we see that

$$E\{M_1^2\} \to 0 \text{ for } N \to \infty\,.$$

As for the second term we have

$$E\{M_2^2\} \cong \sigma_0^2 \sum_{N+1}^{N+p} {}_i \frac{\mu_i^{N+p}}{\rho(t_i)}\,,$$

which is indeed convergent to zero.
Therefore we have

$$\lim_{N\to\infty} E\{|\mu_{wN+p}(T) - \mu_{wN}(T)|^2\} = 0 \qquad (2.26)$$

and the same is true for any smaller measurable A.

Consequently we can define

$$\mu_w(A) = \lim_{N\to\infty} \mu_{wN}(A) \qquad (2.27)$$

From (2.21) and (2.26) we can derive the main properties of the Wiener measures, namely for any measurable A, B

$$\boxed{\begin{aligned} &E\{\mu_w(A)\} = 0 \\ &E\{\mu_w(A)\mu_w(B)\} = \sigma_0^2 \int_{A\cap B} \frac{dt}{\rho(t)} \end{aligned}} \qquad \begin{aligned}&(2.28)\\&(2.29)\end{aligned}$$

<u>Remark 2.3</u>

We have not yet explicitly mentioned a hypothesis which proves to be necessary in the above computations, i.e.

that $1/\rho(t)$ be an integrable function. Even more, we shall assume that ρ is bounded from above and below over all T, $\rho_1 \geq \rho(t) \geq \rho_0$, meaning that the density of measurements can vary significantly in T but remaining commensurable from point to point.

Now that the field of Wiener measures has been suitably defined, we can represent the observation equations (2.10) in the limit form

$$\boxed{\begin{cases} \mu_Y(A)=\mu_y(A) + \mu_w(A) \\ \mu_y(A)=\int_A y(t)\,dt \end{cases}} \qquad (2.30)$$

or in their equivalent infinitesimal formulation

$$\mu_Y(dt) = y(t)\,dt + \mu_w(dt)\ . \qquad (2.31)$$

Now we try to define integrals of functions over the measure $\mu_Y(dt)$.

To this aim we must define $\int_T f(t)d\mu_w(t)$, i.e. the so-called Wiener integral. Please note that from now on we will use as equivalent notations

$$d\mu_\omega(\bar{t}) = \mu_\omega\left[(\bar{t},\bar{t}+dt)\right] = \mu_\omega(dt)$$

First let us assume that f is a piecewise constant function, over a partition $\{T_i\}$ of T,

$$f(t) = \sum f_i \chi_i(t) \qquad (2.32)$$

$$\left(\chi_i(t) = \begin{cases} 1 & t \in T_i \\ 0 & t \notin T_i \end{cases}\right)$$

then we put by definition

$$I(f) = \int_T f(t)d\mu_w(t) = \sum_i f_i \mu_w(T_i) \qquad (2.33)$$

It is easy to recognize that

$$E\{I(f)\} = 0 \qquad (2.34)$$

$$E\{I^2(f)\} = \sum_i f_i^2 \int_{T_i} \frac{dt}{\rho(t)} =$$

$$= \int_{T_i} f^2(t)\,\frac{dt}{\rho(t)} \qquad (2.35)$$

These two relations show that I is in fact an isometry between piecewise constant functions with norm

$$\|f\|^2_{1/\rho} = \int_T f^2(t)\,\frac{dt}{\rho(t)} \qquad (2.36)$$

and L^2_w i.e. the space of random variables with finite variance which are linear combinations of $\{\mu_w(A)\}$. Let us note that the first linear space is in fact dense in $L^2(T)$ because (2.36) is equivalent to the simple L^2 norm as a consequence of the relation $\rho_1 \geq \rho(t) \geq \rho_o$ and moreover any L^2 function can be approximated by a piecewise constant function.

It follows that I(f) can be extended by continuity to any function $f \in L^2(T)$ so that

$$I(f) = \int_T f(t)d\mu_w(t) \qquad (2.37)$$

becomes an L^2_w variable satisfying the two main properties

$$E\{I(f)\} = 0$$

$$E\{I(f)I(g)\} = \int_T f(t)g(t)\frac{dt}{\rho(t)} \qquad (2.38)$$

for any f, $g \in L^2(T)$.

With this new definition we can also nicely represent in a "weak form" the observation equation (2.31), namely we can write

$$\int_T f(t)d\mu_Y(t) = \int_T f(t)y(t)dt +$$

$$+ \int_T f(t)d\mu_w(t) \qquad \forall\ f \in L^2(T) \qquad (2.39)$$

This form will be directly used in the next paragraph; in any way we can see now what is the "natural" degree of smoothness we must impose on y(t), in fact in order that (2.39) be meaningful and finite for every $f \in L^2(T)$ it is enough that $y \in L^2(T)$ too.

The last point we try to discuss shortly is what are the conditions under which it is reasonable to use the continuous limit.
Essentially we must compare the observation equations in the discrete form

$$Y_i\mu_i = y(t_i)\mu_i + \nu_i\mu_i \quad , \quad (i=1,2..N) \qquad (2.40)$$

and their continuous counterpart

$$\mu_Y(T_i) = \int_{T_i} y(t)dt + \mu_w(T_i) \qquad (2.41)$$

and verify whether they are close to each other.
In particular for the noise part we have just to verify that the two variances are equal, for naturally $\nu_i\mu_i$ and $\mu_w(T_i)$ cannot be stochastically close since $\mu_w(T_i)$ is constructed by adding to ν_i an infinite number of independent noises suitably scaled.
So we must compare $\sigma_0(\int_{T_i} \frac{dt}{\rho})^{1/2}$ with $\sqrt{N}\ \sigma_0\mu_i$ which are the r.m.s. respectively of $\mu_w(T_i)$ and $\nu_i\mu_i$: it is easy to realize that assuming $\rho = \rho_i =$ constant over T_i, due to the relation $N\ \rho_i\mu_i = 1$, we have exactly

$$\frac{\sigma_0^2}{N\ \mu_i^2\sigma_0^2} \int \frac{dt}{\rho} = \frac{\mu_i}{N\ \mu_i^2\rho_i} = 1$$

This means that if ρ is a smooth, almost constant function over the cells T_i, we expect the two variances to be very close; in this sense note should be taken that where $\rho(t)$ has sudden variations it is preferable to use a stepwise discontinuous $\rho(t)$ rather than imposing a continuous interpolation with a very steep $\rho(t)$.
As for the deterministic part the relevant quantity to be evaluated is

$$\left|y(t_i)\mu_i - \int_{T_i} y(t)dt\right| = \left|\int_{T_i} [y(t)-y(t_i)]dt\right| \quad ; \qquad (2.42)$$

the continuous limit will be acceptable if for every i this quantity is much smaller than the noise level, namely $\sigma_\nu\mu_i$.
This statement can be expressed by the global relation

$$\sum_i \frac{\left|\int_{T_i} [y(t) - y(t_i)]dt\right|}{\sigma_\nu\mu_i} << N \qquad (2.43)$$

In order to elaborate a little on (2.43) let us assume y to be twice differentiable and that the point t_i is the barycenter of T_i; then we have

$$\int_{T_i} [y(t) - y(t_i)]dt \cong$$

$$\cong \frac{1}{2} \int_{T_i} (t-t_i)^+ y''(t_i)(t-t_i)dt =$$

$$= \frac{1}{2}\ \mathrm{Tr}\ y''(t_i)C_i \qquad (2.44)$$

where C_i is a form-factor matrix of the order of δ^{D+2}, with D the dimension of the vector t. If T_i is symmetric around t_i then C_i will be simply proportional to the identity matrix; for instance if T_i is a cube of side L_i around t_i we have

$$C_i = \frac{1}{12}(L_i)^{D+2}I = \frac{\mu_i}{12}\mu_i^{2/D}I \quad . \qquad (2.45)$$

Using (2.45) in (2.44) and (2.43) we get

$$\sum_1^N {}_i \frac{|\ \mathrm{Tr}\ y''(t_i)|\mu_i}{24\ \sigma_\nu N\ \mu_i^{1-2/D}} << 1 \quad ; \qquad (2.46)$$

finally exploiting (2.7) we arrive at the relation

$$\frac{1}{24\ \sigma_\nu N^{2/D}} \sum_i |\mathrm{Tr}\ y''(t_i)|\mu_i\rho(t_i)^{1-2/D} \cong$$

$$\cong \frac{1}{24\ \sigma_\nu N^{2/D}} \int_T |\mathrm{Tr}\ y''(t)|\rho(t)^{1-2/D}dt << 1 \ . \qquad (2.47)$$

The inequality (2.47) is simple enough

to be used in practice as we show in the following simple example.

Example 2.2

Assume t to be a time, so that D = 1, the set T to be the interval $[0,2\pi]$ and take as measured function

$$y(t) = A \sin nt \quad ;$$

in this case y" is just a scalar and indeed

$$|\mathrm{Tr}\, y''| = A\, n^2 |\sin nt| \quad ;$$

let us further assume that the density of measurements ρ is constant, so that $\rho = (2\pi)^{-1}$.
Taking into account that

$$\int_0^{2\pi} |\sin nt| dt = 4 \ ,$$

by applying (2.47) we find

$$\frac{1}{24\sigma_\nu N^2} (2\pi)\ 4\ n^2 A \ll 1 \quad ,$$

i.e. essentially

$$\frac{n^2}{N^2} \frac{A}{\sigma_\nu} \ll 1 \tag{2.48}$$

which seems like a type of generalized Nyquist relation.
It has to be stressed that in the linear relation between frequency, n, and number of points, N, now the signal to noise ratio is essential.

3. Optimal linear estimation

We start this paragraph by proving the results, recalled in the introduction, in a non-standard way which is particularly suited to treat observation equations in weak (or dual) form, like (2.39) (cfr. also Dermanis A., 1991).
So we go back to the finite dimensional model,

$$Y = y + \nu = Ax + \nu \tag{3.1}$$

$$(E\{\nu\} = 0,\ E\{\nu\nu^+\} = C)$$

and we write it in a weak form by taking the scalar product (here we use the simple euclidean topology) with any vector η in H_Y.
We get

$$\eta^+ Y = \eta^+ Ax + \eta^+ \nu \tag{3.2}$$

by using the rules of transposition this writes

$$\begin{cases} \eta^+ Y = \xi^+ x + \eta^+ \nu \\ \xi = A^+ \eta \end{cases} \tag{3.3}$$

so that we read directly that the linear functional $\eta^+ Y$ of the observations Y is an unbiased estimator of the linear functional of x, $\xi^+ x$, because $E\{\eta^+\nu\}=0$ whatever is y.
Now we consider two questions: first, is it true that we can always find η such that $A^+\eta=\xi$ for a given ξ ? The answer is yes because the system $A^+\eta=\xi$ has more unknowns than equations and we assumed the matrix A to be of full rank. This means that if we want to estimate $\xi^+ x$ from the observations we can always find η such that $\eta^+ Y$ is an unbiased estimator of $\xi^+ x$.
Second question: are there many η such that $A^+\eta=\xi$? Indeed our hypothesis above implies that there are $(\infty)^{n_y - n_x}$ solutions of that system. This urges a rational choice among them; so we can look at the variance of those estimators, which can be computed directly from (3.3)

$$\sigma^2(\eta^+ Y) = \eta^+ C \eta \ . \tag{3.4}$$

Naturally we prefer an estimator with smaller variance so η can be chosen in such a way as to satisfy the minimum property

$$\begin{cases} \eta^+ C\eta = \min \\ A^+ \eta = \xi \end{cases} \tag{3.5}$$

with the customary Lagrange multiplier approach we find (λ being the multiplier)

$$\begin{cases} C\eta = A\lambda \\ A^+ \eta = \xi \end{cases} \tag{3.6}$$

leading to the normal equation

$$(A^+ C^{-1} A)\lambda = \xi \tag{3.7}$$

and to the solution

$$\eta = C^{-1} A (A^+ C^{-1} A)^{-1} \xi \quad . \tag{3.8}$$

So the sought estimator is

$$\eta^+ Y = \xi^+ (A^+C^{-1}A)^{-1}A^+C^{-1}Y \quad . \qquad (3.9)$$

In the finite dimensional case we may conclude that

$$\eta^+ Y = \xi^+ \hat{x}$$

so that not only we have estimated the functional $\xi^+ x$ but we have on the same time an estimate for the whole vector x, namely

$$\hat{x} = (A^+C^{-1}A)^{-1}A^+C^{-1}Y \qquad (3.10)$$

which indeed coincides with (1.8) when the choice (1.10) is made.
The above reasoning can be repeated almost identically for the infinite dimensional case, except the last conclusion, because as we have observed in the Example 2.1 even if we are able to estimate all linear functionals of the unknown x, it is not true that there is a certain $\hat{x} \in H_x$ giving rise to all these functionals.
So let us first establish precisely the <u>observational</u> <u>model</u> ; we assume that we have several domains D_i (i=1,2...n), some of which may also coincide; on each D_i a field of continuous measurements is given, $d\mu_i(t_i)$, so that if two D_i coincide we are representing two fields measured on the same domain; the observational equations in infinitesimal form are

$$d\mu_{Y_i} = y_i(t_i)dt_i + d\mu_{wi} \qquad (3.11)$$

with

$$y_i(t_i) = A_i x \quad i = 1,2...n \qquad (3.12)$$

and with noises (Wiener measures) characterized by

$$E\{d\mu_{wi}\} = 0 \qquad (3.13)$$

$$E\{d\mu_{wi}(t_i)d\mu_{wj}(t_j)\} = \delta_{ij}\frac{\sigma_{0i}^2}{\rho_i(t_i)}dt_i \; ; \qquad (3.14)$$

the equations (3.11) to (3.14) can be written in weak form, introducing n arbitrary functions $f_i \in L^2(D_i)$ (i=1,2 ..n),

$$\sum_i \int_{D_i} f_i d\mu_{Y_i} =$$

$$= \sum_i \int_{D_i} f_i y_i dt_i + \sum_i \int_{D_i} f_i d\mu_{wi} \qquad (3.15)$$

$$E\left\{\int_{D_i} f_i d\mu_{wi} \int_{D_j} f_j d\mu_{wj}\right\} =$$

$$= \delta_{ij}\int_{D_i} \frac{\sigma_{0i}^2}{\rho_i(t_i)} f_i^2(t_i)dt_i \quad ; \qquad (3.16)$$

introducing the vectors

$$d\underline{\mu}_Y = \begin{bmatrix} \vdots \\ d\mu_{Yi} \\ \vdots \end{bmatrix} , \quad d\underline{\mu}_w = \begin{bmatrix} \vdots \\ d\mu_{wi} \\ \vdots \end{bmatrix} ,$$

$$\underline{A} = \begin{bmatrix} \vdots \\ A_i \\ \vdots \end{bmatrix} , \quad \underline{y} = \begin{bmatrix} \vdots \\ y_i(t_i) \\ \vdots \end{bmatrix} ,$$

the Hilbert space $H_Y = \sum_{\oplus} L^2(D_i)$ with the scalar product

$$\langle \underline{f}, \underline{y} \rangle_Y = \sum_i \int_{D_i} f_i(t_i) y_i(t_i) dt_i \; ,$$

the matrix operator in H_Y

$$C\,\underline{f} = \begin{bmatrix} \ddots & & 0 \\ & \frac{\sigma_{0i}^2}{\rho(t_i)} & \\ 0 & & \ddots \end{bmatrix} \begin{bmatrix} \vdots \\ f_i \\ \vdots \end{bmatrix} = \begin{bmatrix} \vdots \\ \frac{\sigma_{0i}^2}{\rho_i} f_i \\ \vdots \end{bmatrix}$$

and the stochastic linear functional

$$L_{\underline{f}}(d\underline{\mu}_Y) = \sum_i \int_{D_i} f_i d\mu_{Yi}$$

$$L_{\underline{f}}(d\underline{\mu}_w) = \sum_i \int_{D_i} f_i d\mu_{wi}$$

the precise definition of which has been the object of section 2, we can finally represent our observational model through the equations

$$L_{\underline{f}}(d\underline{\mu}_Y) = \langle \underline{f}, \underline{y} \rangle_Y + L_{\underline{f}}(d\underline{\mu}_w) \quad (3.17)$$

$$\underline{y} = \underline{A}\, x \quad (3.18)$$

$$E\,\{L_{\underline{f}}(d\underline{\mu}_w)\} = 0 \quad (3.19)$$

$$E\,\{L_{\underline{f}}(d\underline{\mu}_w)^2\} = \langle \underline{f}, \underline{C}\underline{f} \rangle_Y \quad . \quad (3.20)$$

Following the suggestion of Remark 1.3, we stipulate that x ranges in a Hilbert space with norm

$$\|x\|_X^2 = \|\underline{A}\, x\|_Y^2 \quad ; \quad (3.21)$$

this implies that the manifold of the admissible values for $\underline{y}$, i.e. $\mathcal{R}_A$, is closed in H_Y and that there is an isometric relation between $\mathcal{R}_A$ and $\mathcal{D}_A$; in particular we must require that

$$\underline{A}x = 0 \Rightarrow x = 0 \quad (3.22)$$

i.e. that $\underline{A}$ is a full rank operator, for otherwise (3.21) would not be a norm. Now it is enough to recall the definition of the adjoint operator $\underline{A}^+$, namely

$$\langle \underline{A}^+ \underline{f}, x \rangle_X \equiv \langle \underline{f}, \underline{A}x \rangle_Y \quad (3.23)$$

to see that (3.17) can be written as

$$L_{\underline{f}}(d\underline{\mu}_Y) = \langle \underline{A}^+ \underline{f}, x \rangle_X + L_{\underline{f}}(d\underline{\mu}_w) \quad (3.24)$$

so that $L_{\underline{f}}(d\underline{\mu}_Y)$ is indeed an unbiased estimator of $\langle g, x \rangle_X$ with $g = \underline{A}^+\underline{f}$; the variance of this estimator is given by (3.20).

Our next step will be to search for the $\underline{f} \in H_Y$ that minimizes (3.20) under the condition

$$\underline{A}^+\underline{f} = g \quad , \quad (3.25)$$

for a given $g \in H_X$. Before undertaking this step however we must understand properly the nature of equation (3.25) and answer the two questions: is there a solution of (3.25) for any g and, in case, how many solutions are there?

To give an answer we recall that $\underline{A}$ is an isometry between H_X and $\mathcal{R}_A$ and thus, being a linear closed manifold in H_Y containing 0, it is a Hilbert space itself, with the same norm as H_Y; therefore $\underline{A}^+$, restricted to $\mathcal{R}_A$, is indeed the inverse of $\underline{A}$, in fact $\forall\, x \in H_X$

$$\langle x, x \rangle_X = \langle \underline{A}x, \underline{A}x \rangle_Y = \langle \underline{A}^+\underline{A}x, x \rangle_X$$

implying that

$$\underline{A}^+\underline{A} = I \qquad (\text{in } H_X) \ . \quad (3.26)$$

Then the restriction of $\underline{A}^+$ to $\mathcal{R}_A$ is as a matter of fact the isometry $\underline{A}^+ : \mathcal{R}_A \to H_X$ inverse of the isometry $\underline{A} : H_X \to \mathcal{R}_A$.

It follows that the equation (3.25) has always at least the solution

$$\underline{f} = \underline{A}\ g \quad . \quad (3.27)$$

Let us consider now the action of $\underline{A}^+$ on $\mathcal{R}_A^\perp$ (the orthogonal complement of $\mathcal{R}_A$).

If $\underline{\eta} \in \mathcal{R}_A^\perp$, then $\forall\, \xi \in H_X$

$$\langle \underline{A}^+ \underline{\eta}, \xi \rangle_X = \langle \underline{\eta}, \underline{A}\xi \rangle_Y = 0$$

implying that

$$\underline{A}^+\underline{\eta} = 0 \quad .$$

Therefore the equation (3.25), written in the form

$$\underline{A}^+\underline{f}^{\parallel} + \underline{A}^+\underline{f}^{\perp} = g \ ,$$

where $\underline{f}^{\parallel}$ is the projection of $\underline{f}$ on $\mathcal{R}_A$ while $\underline{f}^{\perp}$ is its projection on $\mathcal{R}_A^\perp$, has the general solution

$$\begin{cases} \underline{f}^{\parallel} = A\ g \\ \underline{f}^{\perp} \ \text{ whatever in } R_A^\perp \end{cases} \quad (3.28)$$

Since $\mathcal{R}_A^\perp$ is a closed subspace of H_Y, (3.28) represents a closed linear manifold in H_Y.

Our target is now to find the solution of the variational problem

$$\begin{cases} \langle \underline{f}, \underline{C}\underline{f} \rangle_Y = \min \\ \underline{A}^+\underline{f} \ = \ g \end{cases} \quad . \quad (3.29)$$

Let us first observe that if we assume, as we have done in section 2, that the density functions ρ_i satisfy

$$0 < \rho_{0i} \le \rho_i \le \rho_{1i}\ , \quad (\rho_{0i},\rho_{1i}\ \text{const.})$$

then

$$a^2\langle \underline{f},\underline{f}\rangle_Y \le \langle \underline{f},C\underline{f}\rangle_Y \le b^2\langle \underline{f},\underline{f}\rangle_Y \qquad (3.30)$$

with

$$a^2 = \mathrm{Inf}\ \frac{\sigma^2_{0i}}{\rho_{1i}} \qquad b^2 = \mathrm{Sup}\ \frac{\sigma^2_{0i}}{\rho_{0i}} \quad . \qquad (3.31)$$

Consequently $\langle \underline{f},C\underline{f}\rangle_Y$ can be viewed as an equivalent norm in H_Y, so that (3.29) means that we are looking for the element of minimum C-norm in the closed manifold described by (3.28); this element is known to exist and to be unique. So our problem is just to find the formal solution of (3.29).

By applying a standard technique we introduce a Lagrange multiplier $\lambda \in H_X$ and we form the new target functional

$$\phi(\underline{f}) = \frac{1}{2}\ \langle \underline{f},C\underline{f}\rangle_Y - \langle \underline{A}^+\underline{f},\lambda\rangle_X =$$

$$= \frac{1}{2}\ \langle \underline{f},C\underline{f}\rangle_Y - \langle \underline{f},\underline{A}\lambda\rangle_Y \quad ; \qquad (3.32)$$

setting to zero the first variation of (3.32) we receive

$$C\underline{f} = \underline{A}\lambda,$$

i.e.

$$\underline{f} = C^{-1}\underline{A}\lambda$$

which substituted in the second of (3.29) gives

$$\underline{A}^+C^{-1}\underline{A}\lambda = g \quad .$$

The normal operator $\underline{A}^+C^{-1}\underline{A}$ is indeed bounded, it is selfadjoint and even strictly positive definite in view of the relation (recall (3.30))

$$\langle \lambda,\underline{A}^+C^{-1}\underline{A}\lambda\rangle_X = \langle \underline{A}\lambda,C^{-1}\underline{A}\lambda\rangle_Y \ge$$

$$\ge \frac{1}{b^2}\ \langle \underline{A}\lambda,\underline{A}\lambda\rangle_Y = \frac{1}{b^2}\ \langle \lambda,\lambda\rangle_X \quad ; \qquad (3.33)$$

so its inverse exists and is bounded too.

Therefore we obtain the solution

$$\begin{aligned}\lambda &= (\underline{A}^+C^{-1}\underline{A})^{-1}g\\ \underline{f} &= C^{-1}\underline{A}(\underline{A}^+C^{-1}\underline{A})^{-1}g \quad :\end{aligned} \qquad (3.34)$$

this means that the stochastic functional $L_{\underline{f}}(d\underline{\mu}_Y)$ with $\underline{f}$ given by (3.34) is the minimum variance unbiased estimator of $\langle \underline{g},x\rangle_X$.

To be complete let us also compute the variance of $L_{\underline{f}}(d\underline{\mu}_Y)$; from (3.20) we have

$$\sigma^2\{L_{\underline{f}}(d\underline{\mu}_Y)\} = \langle \underline{f},\ C\ \underline{f}\rangle_Y$$

$$\langle C^{-1}\underline{A}(\underline{A}^+C^{-1}\underline{A})^{-1}g, C\ C^{-1}\underline{A}(\underline{A}^+C^{-1}\underline{A})^{-1}g\rangle_Y =$$

$$= \langle g,(\underline{A}^+C^{-1}\underline{A})^{-1}g\rangle_X \qquad (3.35)$$

<u>Remark 3.1</u>

The topology we have chosen for H_X is not unique but just the simplest; any other topology with respect to which $\underline{A}$ is a continuous injection $H_X \to H_Y$, so that $\mathcal{R}_A$ is a closed subspace of H_Y, would do with the <u>same</u> solution formulae. Naturally the definition of $\underline{A}^+$ depends on the definition of the scalar product in H_X in such a way that in the end the estimator $L_{\underline{f}}(d\underline{\mu}_Y)$ is always the same. Note also that in case the functional of x we want to estimate is given directly, F(x), then its representer g in H_X is as well dependent on the definition of the scalar product in H_X.

<u>Remark 3.2</u>

We could have multiplied at the beginning the observation equation (3.11) by $\dfrac{\sqrt{\rho_i(t_i)}}{\sigma_{0i}}$, thus obtaining

$$d\mu_{zi} = z_i(t_i)dt_i + d\mu_{vi} \qquad (3.36)$$

with

$$d\mu_{zi} = \frac{\sqrt{\rho_i}}{\sigma_{0i}} d\mu_{Yi}$$

$$z_i(t_i) = \frac{\sqrt{\rho_i}}{\sigma_{0i}} y_i$$

$$d\mu_{vi} = \frac{\sqrt{\rho_i}}{\sigma_{0i}} d\mu_{wi}$$

$$E\{d\mu_{vi} d\mu_{vj}\} = \delta_{ij} dt_i \quad .$$

For this new model we have

$$C = I \qquad (\text{in } H_Y)$$

and, recalling (3.26),

$$\underline{A}^+ C^{-1} \underline{A} = \underline{A}^+ \underline{A} = I \qquad (\text{in } H_X)$$

so that the sought estimator of $<g,x>_{H_X}$ is simply

$$L_{\underline{f}}(d\underline{\mu}_z) = \sum_i \int_{D_i} f_i d\mu_{zi}$$

with

$$\underline{f} = \underline{A}\, g \tag{3.37}$$

Indeed with this modification we have changed also $\underline{A}$ as well as $\underline{A}^+$ and g through the change of scalar product in H_X, which depends on the definition of $\underline{A}$.
All this is analogous to the reduction of the observation equations to the same weight in the finite-dimensional case.

Remark 3.3

We want to prove that it is not possible to write the estimator $L_{\underline{f}}(d\underline{\mu}_Y)$ in the form $<g,\hat{x}>_X$, with $\hat{x}$ a process with realization in H_X and with bounded squared norm.

In fact if

$$L_{\underline{f}}(d\underline{\mu}_Y) = <g,\hat{x}>_X \quad , \tag{3.38}$$

since

$$E\{L_{\underline{f}}(d\underline{\mu}_Y)\} = <g,x>_X \quad ,$$

we must have

$$\hat{x} = x + \hat{\xi} \quad , \quad <g,\hat{\xi}>_X = L_{\underline{f}}(d\underline{\mu}_w) \quad ; \tag{3.39}$$

then

$$E\{<g,\hat{\xi}>_X^2\} = <g, C_{\xi\xi} g>_X \tag{3.40}$$

by definition of covariance operator.
Comparing (3.40) with (3.35) we see that it should be

$$C_{\xi\xi} = (\underline{A}^+ C^{-1} \underline{A})^{-1} \quad . \tag{3.41}$$

On the other hand, if $E\{\|\hat{x}\|_X^2\} < +\infty$ by hypothesis, the same must be true for $\hat{\xi} = \hat{x} - x$, since $x \in H_X$.
So if $\{e_n\}$ is a complete orthonormal system in H_X we should have

$$+\infty > E\{\|\hat{\xi}\|_X^2\} = E\{\textstyle\sum_n <\hat{\xi}, e_n>_X^2\} =$$

$$= \sum_{1}^{+\infty}{}_n <e_n, C_{\xi\xi} e_n>_{H_X} = \mathrm{Tr}C_{\xi\xi} \quad ; \tag{3.42}$$

therefore $\|\hat{\xi}\|_X^2$ is bounded in the average iff $C_{\xi\xi}$ is a finite trace operator.
But this hypothesis contradicts (3.41) since $\forall\, g \in H_X$,

$$<g, \underline{A}^+ C^{-1} \underline{A} g>_X = <\underline{A}g, C^{-1} \underline{A} g>_Y \le$$

$$\le \frac{1}{a^2} <\underline{A}g, \underline{A}g>_Y = \frac{1}{a^2} <g,g>_X$$

with the positive constant a defined in (3.30), so that

$$<\underline{e}_n, (\underline{A}^+ C^{-1} \underline{A})^{-1} \underline{e}_n>_X \ge$$

$$\ge a^2 <e_n, e_n>_X = a^2 > 0$$

and indeed

$$\mathrm{Tr}\ (\underline{A}^+ C^{-1} \underline{A})^{-1} = +\infty \quad .$$

This does not mean that we could not find a larger space $\tilde{H}_X \supset H_X$ such that $\hat{x} \in \tilde{H}_X$ with probability one and has even a bounded mean square norm there: in this case we might be able to write, in a generalized sense,

$$L_{\underline{f}}(d\underline{\mu}_Y) = \langle g, \hat{x} \rangle_X$$

with $\hat{x} \in \tilde{H}_X \supset H_X$ and $g \in \tilde{H}^*_X \subset H_X$.

We observe now that the estimator $L_{\underline{f}}(d\underline{\mu}_Y)$, with $\underline{f}$ given by (3.34) is invariant under the multiplication of C by an arbitrary constant. This means that we can compute the estimator even if the operator C has the form

$$C = \sigma_0^2\, Q \tag{3.43}$$

with Q known and σ_0^2 unknown. In practice this means that we know C apart from a proportionality factor.
We are left therefore with the problem of estimating σ_0^2, which is also classical of least squares theory.
On the other hand it is enough to rewrite the observation equations in the form

$$L_{\underline{f}}(d\underline{\mu}_Y) = \langle \underline{A}^+ \underline{f}, x \rangle_X + L_{\underline{f}}(d\underline{\mu}_w)$$

to realize that if $\underline{f}$ is orthogonal to $\mathcal{R}_A$, so that $\underline{A}^+ \underline{f} = 0$, we have directly

$$L_{\underline{f}}(d\underline{\mu}_Y) = L_{\underline{f}}(d\underline{\mu}_w) \ , \tag{3.44}$$

$$(\underline{A}^+ \underline{f} = 0) \quad ;$$

whence

$$L_{\underline{f}}(d\underline{\mu}_Y) \sim N\left[0\ ,\ \sigma_0^2 \langle \underline{f}, Q\underline{f} \rangle_Y\right] . \tag{3.45}$$

Assume then to be able to construct a sequence $\{\underline{f}_n\}$ of elements of H_Y such that

$$\underline{A}^+ \underline{f}_n = 0 \ .$$

By applying the Schmidt orthonormalization we can derive another sequence $\{\underline{h}_n\}$ such that

$$\underline{A}^+ \underline{h}_n = 0$$

holds as well, but at the same time

$$\langle \underline{h}_n, Q\underline{h}_m \rangle_Y = \delta_{nm} \quad . \tag{3.46}$$

It follows that not only

$$L_{\underline{h}n}(d\underline{\mu}_Y) = L_{\underline{h}n}(d\underline{\mu}_w)$$

is a sequence of N $[0, \sigma_0^2]$ variables, as implied by (3.45), but they are also independent as

$$E\left\{L_{\underline{h}n}(d\underline{\mu}_Y) L_{\underline{h}m}(d\underline{\mu}_Y)\right\} = \sigma_0^2 \langle \underline{h}_n, Q\underline{h}_m \rangle_Y = 0$$

$$\forall\ n \neq m \quad .$$

Therefore the following distributional result holds

$$\sum_1^N{}_n\, L_{\underline{h}n}(d\underline{\mu}_Y)^2 = \chi_N^2 \sigma_0^2 \tag{3.47}$$

This already proves that

$$\hat{\sigma}_0^2 = \frac{1}{N} \sum_1^N{}_n\, L_{\underline{h}n}(d\underline{\mu}_Y)^2 \tag{3.48}$$

is an unbiased estimator of σ_0^2; moreover since

$$\lim_{N\to\infty} \frac{\chi_N^2}{N} = 1$$

holds both in probability and in mean square sense, we see that if the sequence $\{\underline{h}_n\}$ is infinite we can achieve the exact (with Pr = 1) limit

$$\sigma_0^2 = \lim_{N\to\infty} \frac{1}{N} \sum_1^N{}_n\, L_{\underline{h}n}(d\underline{\mu}_Y)^2 \ . \tag{3.49}$$

<u>Remark 3.4</u>

The same problem, or even a more complex one, can be solved by another limit property; in fact we can prove that given a field of measurements $d\mu_Y(t)$ $(t \in T)$, such that

$$d\mu_Y(t) = y(t)dt + d\mu_w(t)$$

with $y \in L^2(T)$ and

$$E\{d\mu_w^2\} = \sigma_0^2 \frac{dt}{\rho(t)} \quad ,$$

we can estimate σ_0^2, with Pr = 1, so that in the more complex model (3.11) every σ_{0i}^2 can be estimated from the corresponding field $d\mu_{Yi}$.

In fact let us take a partition of T in non-overlapping sets $\{T_i\}$; if we form

$$\sum_i \mu_Y(T_i)^2 = \sum_i\Big(\int_{T_i} y(t)dt\Big)^2 +$$

$$+\ 2\sum_i\Big(\int_{T_i} y(t)dt\Big)\mu_w(T_i) + \sum_i \mu_w(T_i)^2$$

we find that

$$E\Big\{\sum_i \mu_Y(T_i)^2\Big\} =$$

$$= \sum_i\Big(\int_{T_i} y(t)dt\Big)^2 + \sigma_0^2 \sum_i \int_{T_i}\frac{dt}{\rho(t)} \quad .$$

Therefore, since

$$\sum_i\Big(\int_{T_i} y(t)dt\Big)^2 \le \sum_i \int_{T_i} y(t)^2 dt\cdot \operatorname{Sup}_i \mu_i$$

$$(\mu_i = \text{Lebesgue measure of } T_i)$$

we have, with $\delta = \operatorname{Sup}_i \operatorname{diam}(T_i)$,

$$\lim_{\delta\to 0} E\Big\{\sum_i \mu_Y(T_i)^2\Big\} = \sigma_0^2 \int_T \frac{dt}{\rho(t)} \quad . \qquad (3.50)$$

Furthermore it is not difficult to see that

$$\lim_{\delta\to 0} \sigma^2\Big\{\sum_i \Big(\int_{T_i} y(t)dt\Big)\mu_w(T_i)\Big\} = 0$$

$$\lim_{\delta\to 0} \sigma^2\Big\{\sum_i \mu_w(T_i)^2\Big\} = 0 \quad .$$

Consequently we find that in a mean square sense

$$\sigma_0^2 = \lim_{\delta\to 0}\Big(\int_T \frac{dt}{\rho(t)}\Big)^{-1} \sum_i \mu_Y(T_i)^2 \quad . \qquad (3.51)$$

We conclude this paragraph by mentioning that the above theory can be extended to more general cases; for instance let us assume that the disturbances $\{\nu_i\}$, instead of being white noises, admit a smooth covariance function

$$\delta_{ij} C_i(t,\tau) = E\,\{\nu_i(t)\nu_j(\tau)\}. \qquad (3.52)$$

In this case the weak observation equations become

$$\sum_i \int_{D_i} f_i(t_i) Y_i(t_i) dt_i =$$

$$= \sum_i \int_{D_i} f_i(t_i) y_i(t_i) dt_i +$$

$$+ \sum_i \int_{D_i} f_i(t_i)\nu_i(t_i) dt_i \qquad (3.53)$$

which we can write symbolically as

$$(\underline{f},\underline{Y}) = (\underline{f},\underline{y}) + (\underline{f},\underline{\nu}) \quad . \qquad (3.54)$$

The variance of the disturbing term is then

$$E\,\{(\underline{f},\underline{\nu})^2\} = (\underline{f},C\underline{f}) = \sum_i (f_i, C_i f_i)_{L^2(D_i)}$$

$$(f_i, C_i f_i)_{L^2(D_i)} =$$

$$= \int_{D_i}\int_{D_i} f_i(t) C_i(t,\tau) f_i(\tau) dt d\tau \ ,$$

showing that in general we must have $\underline{f} \in H_C$ with norm

$$\|\underline{f}\|_C^2 = (\underline{f},C\underline{f})$$

Consequently in order that the term $(\underline{f},\underline{y})$ remain bounded $\underline{y}$ has to belong to $H_C^* = H_{C^{-1}}$ with norm

$$\|\underline{y}\|_{C^{-1}}^2 = (\underline{y}, C^{-1}\underline{y}) \quad .$$

The following inclusions are then obvious

$$H_C^* = H_{C^{-1}} \subset \sum L^2(D_i) \subset H_C \ ,$$

the embeddings being completely continuous.

Now introducing the transpose operator $\underline{A}^*$ such that

$$(\underline{f},\underline{A}x) = \langle \underline{A}^*\underline{f}, x\rangle_X \ ,$$

with

$$\|x\|_X = \|\underline{A}x\|_{C^{-1}} \quad ,$$

we can proceed as above.
It is then easy to show that

$$\underline{A}^* C^{-1} \underline{A} = I \quad ,$$

in analogy to (3.26) and that the minimum variance unbiased estimator of <g,x> is (f,Y) with

$$\underline{f} = C^{-1} \underline{A}\, g \quad . \qquad (3.55)$$

This is indeed not a formal solution since

$$\|\underline{f}\|_C^2 = (C^{-1}\underline{A}g, CC^{-1}\underline{A}g) =$$

$$= (C^{-1}\underline{A}g, \underline{A}g) = \|\underline{A}g\|^2_{C^{-1}} = \|g\|_X^2$$

which is bounded since $g \in H_X$.

We will not dwell more on this generalization, however, since with the main theory already developed here we can solve many interesting problems as we will show in the next paragraph.

4. Examples

In this paragraph we try to show how the theory presented in section 3 can be applied to some examples taken from various branches of geodesy, geophysics and photogrammetry.
The first example in particular is used for the purpose of training the reader and of checking accurately all the various aspects formerly discussed.

Example 4.1

This example is taken from satellite geodesy where we have instruments (PRARE) capable of measuring at the same time range and range rate from a point on the earth's surface to a satellite; similar is the situation with GPS where we can measure phase and phase rate (by the Doppler effect) of the signal transmitted from satellite to the ground. The density of measurements in time is in these cases so high that we can represent them as fields of continuous measurements. We assume that the two densities $\rho_1(t)$, $\rho_2(t)$ are constant and are the same over a time interval T.

Moreover before applying a much more complicated data reduction involving sophisticated geodetic models one might be interested in using the inner redundancy contained in the two functions $\{x(t), \dot{x}(t)\}$, for instance to check that the instrument is regularly operating.
So let us start by writing the observation equations, in differential form,

$$\begin{cases} d\mu_{o1} = y_1(t)dt + d\mu_{w1} \\ d\mu_{o2} = y_2(t)dt + d\mu_{w2} \end{cases} , \qquad (4.1)$$

$$\begin{cases} \underline{Y} = \begin{vmatrix} y_1 \\ y_2 \end{vmatrix} = \begin{vmatrix} x(t) \\ \dot{x}(t) \end{vmatrix} = \underline{A}x(t) \\ \underline{A} = \begin{vmatrix} I \\ D \end{vmatrix} \qquad (D = \frac{d}{dt}) \end{cases} \qquad (4.2)$$

the stochastic properties of $d\mu_{w1}$, $d\mu_{w2}$ are that these are independent Wiener measures with

$$\begin{cases} E\{d\mu_{w1}^2\} = \sigma_1^2 dt \\ E\{d\mu_{w2}^2\} = \sigma_2^2 dt \end{cases} \qquad (4.3)$$

Note should be taken that we have assumed

$$\rho_1 = \rho_2 = \text{const.} = 1/T \quad ,$$

so that comparing with (3.14), we find

$$\sigma_1^2 = \sigma_{o1}^2 T \quad , \quad \sigma_2^2 = \sigma_{o2}^2 T \quad , \qquad (4.4)$$

showing that σ_1 and σ_2 have measure units different from σ_{o1}, σ_{o2}, which in turn have the same dimensions as the measurements (in our case distance and velocity).
After this remark we put for the sake of simplicity T = 1.
Let us note that here we must have:

$$\|\underline{Y}\|_Y^2 = \int_0^1 [y_1^2(t) + y_2^2(t)]dt \qquad (4.5)$$

$$\|x\|_X^2 = \|\underline{A}x\|_Y^2 = \int_0^1 [x^2(t) + \dot{x}^2(t)]dt \qquad (4.6)$$

so that H_X is nothing but the Sobolev space of square integrable functions with their first derivative.

We can observe that, from the physical point of view, (4.5), (4.6) are very odd expressions since we add quantities like y_1^2, y_2^2 or x^2, $\dot{x}^2$, which have different physical dimensions; yet in the end everything will return to compatible formulae due to the presence of the matrix C^{-1}.
The first step we need now to make is to find the operator $\underline{A}^+$; going back to the definition we have that if

$$\underline{A}^+\underline{f} = u \quad (\underline{f} = \begin{vmatrix} f_1 \\ f_2 \end{vmatrix}) \qquad (4.7)$$

then $\forall\ x \in H_X$

$$\langle u,x\rangle_X = \langle \underline{f},\underline{A}x\rangle_Y \quad , \qquad (4.8)$$

i.e. the formula

$$\int_0^1 \{ux + \dot{u}\dot{x}\}dt = \int_0^1 \{f_1 x + f_2 \dot{x}\}dt \qquad (4.9)$$

has to hold as an identity in x.
We can rearrange (4.9) in the form

$$\int_0^1 (u - f_1)x\ dt = -\int_0^1 (\dot{u} - f_2)\dot{x}\ dt \qquad (4.10)$$

so recognizing from the definition of generalized derivative that, at least in a distribution sense,

$$(u - f_1) = \frac{d}{dt}(\dot{u} - f_2) \qquad (4.11)$$

If we substitute back (4.11) in the first member of (4.10) and we integrate by parts we find that the identity is fully realized iff

$$\dot{u} - f_2\big|_{t=1} = 0 \qquad \dot{u} - f_2\big|_{t=0} = 0 \quad .(4.12)$$

We come to the conclusion that $u = \underline{A}^+\underline{f}$ if

$$\begin{cases} \frac{d}{dt}(\dot{u} - f_2) = (u - f_1) \\ \dot{u}(0) = f_2(0) \quad , \quad \dot{u}(1) = f_2(1) \end{cases} \qquad (4.13)$$

One remark is in order here. Rigorously speaking if $f_2 \in L^2(01)$ the values $f_2(0)$, $f_2(1)$ are not in general defined so that the boundary condition in (4.13) can hardly be meaningful: however one can always think of approximating f_2 with a sequence of functions such that $\dot{f}_2$ too belongs to L^2, so that the pointwise values of f_2 are also defined, and, once the solution u of the problem (4.13) is found, one realizes that it can be extended by continuity to any $f_2 \in L^2$.
We will do exactly that applying (4.13) to define our normal equation. In fact our next step is to write and solve the normal equation

$$(\underline{A}^+C^{-1}\underline{A})\lambda = g \qquad (4.14)$$

Let us recall that g is that element of H_X which represents the functional F(x) we want to estimate, i.e.

$$F(x) = \langle g,x\rangle_X \quad . \qquad (4.15)$$

We observe that, since $x \in H_X$ means $x \in L^2$, $\dot{x} \in L^2$, the evaluation functional, namely

$$F(x) = x(\bar{t}) \qquad (\bar{t}\ \text{fixed})$$

is indeed a bounded functional in H_X.
In fact it is possible to show that in this case the representer of F is

$$G(t,\bar{t}) = -\frac{1}{2}e^{|t-\bar{t}|} + \frac{Ch(1-\bar{t})e^t}{2\ Sh\ 1} +$$

$$+ \frac{e\ Ch\ \bar{t}\ e^{-t}}{2\ Sh\ 1} \quad , \qquad (4.16)$$

for this is essentially the Green's function of the operator $(1 - \frac{d^2}{dt^2})$ satisfying the boundary conditions

$$\dot{G}(1,\bar{t}) = 0 \quad , \quad \dot{G}(0,\bar{t}) = 0 \quad , \quad (4.17)$$

so that taking

$$\langle G,x\rangle_X = \int_0^1 \{G(t,\bar{t})x(t) + \dot{G}(t,\bar{t})\dot{x}(t)\}dt$$

we obtain after an integration by parts

$$\langle G,x\rangle_X = \int_0^1 \{G - \ddot{G}\}\ x\ dt = x(\bar{t}) \quad .$$

Now in order to write and solve (4.14)

we have just to put

$$\underline{f} = C^{-1}\underline{A}\lambda = \begin{vmatrix} 1/\sigma_1^2 & \lambda \\ 1/\sigma_2^2 & \dot{\lambda} \end{vmatrix} \qquad (4.18)$$

in (4.13) and write g instead of u, thus arriving at the problem

$$\begin{cases} \frac{d}{dt}(\dot{g} - 1/\sigma_2^2\,\dot{\lambda}) = (g - 1/\sigma_1^2\,\lambda) \\ \dot{g}(0) - 1/\sigma_2^2\,\dot{\lambda}(0) = 0 \\ \dot{g}(1) - 1/\sigma_2^2\,\dot{\lambda}(1) = 0 \end{cases} \qquad (4.19)$$

It is simpler to transform (4.19), posing

$$\gamma = g - 1/\sigma_2^2\,\lambda$$

so that, calling

$$\alpha^2 = \sigma_2^2 / \sigma_1^2 \quad ,$$

we have

$$\begin{cases} \ddot{\gamma} - \alpha^2\gamma = (1 - \alpha^2)g \\ \dot{\gamma}(0) = 0 \quad , \quad \dot{\gamma}(1) = 1 \end{cases} \qquad (4.20)$$

The Green's function of this problem is analogous to (4.16), and we call it

$$G(t,\bar{t};\alpha) = \frac{1}{2\alpha} e^{\alpha|\bar{t}-t|} - \frac{\mathrm{Ch}(1 - \bar{t})e^{\alpha t}}{2\alpha\,\mathrm{Sh}\,\alpha} +$$

$$- \frac{e^{\alpha}\mathrm{Ch}\,\bar{t}\,e^{-\alpha t}}{2\alpha\,\mathrm{Sh}\,\alpha} \quad ; \qquad (4.21)$$

with the help of this function we can write the solution of (4.20) as

$$\gamma(t) = (1 - \alpha^2)\int_0^1 G(t,\tau,\alpha)g(\tau)d\tau \quad , \qquad (4.22)$$

which finally allows to compute

$$\lambda(t) = \sigma_2^2\,[g(t) - \gamma(t)] \qquad (4.23)$$

From λ we can compute $\underline{f}$ through (4.18) (cf. also (3.34)) and also the estimator

$$L_{\underline{f}}(d\underline{\mu}_Y) = \int_0^1 f_1 d\mu_{o1} + f_2 d\mu_{o2} =$$

$$= \int_0^1 \left\{ \frac{1}{\sigma_1^2}\lambda\, d\mu_{o1} + \frac{1}{\sigma_2^2}\dot{\lambda}\, d\mu_{o2} \right\} \qquad (4.24)$$

which is our ultimate goal.

We can observe that if we use in (4.22) the representer of the evaluation functional (cfr.(4.16)) then by $L_{\underline{f}}(d\underline{\mu}_Y)$ we have an unbiased estimate of the pointwise value $x(\bar{t})$; this can be repeated for every $\bar{t} \in \overline{01}$ so that in this case we can indeed find a function $\hat{x}(t)$ which is an unbiased estimator of $x(t)$, however in view of Remark 3.3 we cannot expect $\hat{x}(t)$ to be so regular as to belong to H_X, almost surely.

Without going into too many details we mention also that the variance of $L_{\underline{f}}(d\underline{\mu}_{oY})$ can be computed from (3.35):

$$\sigma^2(L_{\underline{f}}(d\underline{\mu}_{oY})) = \langle g,\lambda\rangle_X \quad ;$$

if as g we take $G(t,\bar{t})$ (cfr.(4.16)), i.e. the functional of evaluation at $\bar{t}$ and if λ is the corresponding solution, say $\lambda(t,\bar{t})$, computed from (4.23), (4.22) with $g = G(t,\bar{t})$, then we have simply

$$\sigma^2(L_{\underline{f}}(d\underline{\mu}_{oY})) = \lambda(\bar{t},\bar{t}) \quad .$$

Example 2

This example is taken as an extreme idealization of a foreseen satellite mission aiming at improving our knowledge of the physics of the solid earth: the so-called Aristoteles project. (cfr. Brovelli, Migliaccio and Sanso' 1991; Bassanino and Migliaccio 1991)

In the gradiometric part of the mission it is supposed that we can reconstruct three components of the gravity gradient tensor, through gradiometric measurements.

In particular we can assume that the second radial derivative of the potential, T_{rr}, is measured with a very high density on a sphere at satellite's altitude (i.e. a sphere S_s of radius R_s).

The measurements are assumed to be of

equal variance and uniformly distributed. Similarly we assume to know on the earth's surface, taken as a sphere S with radius R, the field of gravity anomalies, i.e.

$$\Delta g = -T_r - \frac{2}{R} T \qquad (4.25)$$

which is also assumed to be measured with constant variance and uniform density.
Our purpose is to estimate functionals of the gravity anomalous potential T.
The observational model is then

$$\begin{cases} d\mu_{Y1} = (-T_r - \frac{2}{R} T)R^2 d\sigma + d\mu_{w1} \\ d\mu_{Y2} = (T_{rr})R_s^2 d\sigma + d\mu_{w2} \end{cases} \qquad (4.26)$$

with $d\mu_{w1}$, $d\mu_{w2}$ mutually independent Wiener measures and

$$\begin{cases} E\{d\mu_{w1}^2\} = \sigma_1^2 R^2 d\sigma \\ E\{d\mu_{w2}^2\} = \sigma_2^2 R_s^2 d\sigma \end{cases} . \qquad (4.27)$$

The H_Y space is a space of $\underline{y} = \begin{vmatrix} y_1 \\ y_2 \end{vmatrix}$ with $y_1 \in L^2(S)$, $y_2 \in L^2(S_s)$, i.e.

$$\|\underline{y}\|_Y^2 = \frac{1}{4\pi} \int_\sigma \{y_1^2 R^2 + y_2^2 R_s^2\} d\sigma \quad ; \qquad (4.28)$$

moreover, with

$$\underline{A} = \begin{bmatrix} -\frac{\partial}{\partial r} - \frac{2}{R} \Big|_{r=R} \\ \frac{\partial^2}{\partial r^2} \Big|_{r=R_s} \end{bmatrix}$$

H_X becomes the space of functions harmonic outside the sphere S with norm

$$\|T\|_X^2 = \|AT\|_Y^2 =$$

$$= \frac{1}{4\pi} \int \left\{ \left(-T_r - \frac{2}{R} T \right)^2 R^2 + T_{rr}^2 R_s^2 \right\} d\sigma \quad ; \qquad (4.29)$$

this formula can also be expressed in terms of harmonic components, i.e. after developing

$$T = \sum_0^{+\infty}{}_l \sum_{-l}^{l}{}_m T_{lm} \left(\frac{R}{r} \right)^{l+1} Y_{lm}$$

we have

$$\|T\|_X^2 = \sum_{lm} c_l T_{lm}^2 , \qquad (4.30)$$

$$c_l = (l-1)^2 + \frac{(l+1)^2 (l+2)^2}{R^2} q^{2l+4}$$

$$q = \frac{R}{R_s}$$

Let us observe also here that the apparent dimensional inconsistencies will be solved in the final formulae due to the presence of the C^{-1} matrix.
From the relation

$$\underline{A}^+ \underline{f} = g \iff \langle g, T\rangle_X = \langle \underline{f}, \underline{A}T\rangle_Y \quad ,$$

which can be expressed in spectral terms as

$$\sum c_l g_{lm} T_{lm} =$$

$$= \frac{1}{4\pi} \int \left\{ f_1 \left(-T_r - \frac{2}{R} T \right) R^2 + f_2 \left(T_{rr} \right) R_s^2 \right\} d\sigma =$$

$$= \sum \left\{ f_{1lm} R(l-1) T_{lm} + f_{2lm} (l+1)(l+2) q^{l+1} T_{lm} \right\}$$

we derive directly

$$g_{lm} = \frac{R(l-1)}{c_l} f_{1lm} + \frac{(l+1)(l+2)}{c_l} q^{l+1} f_{2lm} . \qquad (4.31)$$

Now to obtain the normal equation we have only to substitute the relation

$$\underline{f} = C^{-1} \underline{A} \lambda \quad , \qquad (4.32)$$

expressed in spectral terms, into (4.31).
Since by hypothesis $\lambda \in H_X$ it has also to enjoy a series development like T,

$$\lambda = \sum \lambda_{lm} \left(\frac{R}{r} \right)^{l+1} Y_{lm} \qquad (4.33)$$

so we find

$$(\underline{f})_{lm} = C^{-1} \left| \begin{matrix} \dfrac{(l-1)}{R} \\ \dfrac{(l+1)(l+2)}{R^2}\, q^{l+3} \end{matrix} \right| \lambda_{lm} =$$

$$= \left| \begin{matrix} (l-1)/\, \sigma_1^2 R \\ \dfrac{(l+1)(l+2)}{\sigma_2^2 R^2}\, q^{l+3} \end{matrix} \right| \lambda_{lm} \qquad (4.34)$$

Substituting (4.34) in (4.31) we get

$$g_{lm} = N_l \lambda_{lm} \ , \qquad (4.35)$$

$$N_l = \frac{1}{c_l} \left\{ \frac{(l-1)^2}{\sigma_1^2} + \frac{(l+1)^2(l+2)^2}{\sigma_2^2 R^2} q^{2l+4} \right\} = \frac{\tilde{C}_l}{c_l} \ ;$$

let us note that for $l \to \infty$

$$N_l \cong \frac{1}{\sigma_1^2} \ ,$$

so that g and λ have the same degree of regularity, as it was to be expected.
Back substituting we find the optimal estimator of $\langle g, T \rangle_X$ in the form

$$L_{\underline{f}}(d\underline{\mu}_Y) = \sum_{lm} \frac{g_{lm}}{N_l} \left[\frac{(l-1)}{\sigma_1^2 R} \int Y_{lm} d\mu_{o1} + \right.$$

$$\left. + \frac{(l+1)(l+2)}{\sigma_2^2 R^2} q^{l+3} \int Y_{lm} d\mu_{o2} \right] \ . \qquad (4.36)$$

In particular if we wanted to estimate T_{lm}, since indeed in this case

$$\langle g, T \rangle_X = T_{lm} \Leftrightarrow g = \frac{1}{C_l} Y_{lm} \left(\frac{R}{r} \right)^{l+1} \ ,$$

we find from (4.36)

$$\hat{T}_{lm} = \frac{1}{c_l N_l} \left[\frac{(l-1)}{\sigma_1^2 R} \int Y_{lm} d\mu_{o1} + \right.$$

$$\left. + \frac{(l+1)(l+2)}{\sigma_2^2 R^2} q^{l+3} \int Y_{lm} d\mu_{o2} \right] \ . \qquad (4.37)$$

It is interesting to compute the variance of $\hat{T}_{lm}$, in fact, recalling that $c_l N_l = C_l$, we find

$$\sigma^2(\hat{T}_{lm}) = \frac{1}{\tilde{C}_l^2} \left[\frac{(l-1)^2}{\sigma_1^4 R^2} \sigma_1^2 R^2 4\pi + \right.$$

$$\left. + \frac{(l+1)^2 (l+2)^2}{\sigma_2^4 R^4} q^{2l+6} \sigma_2^2 R_s^2 4\pi \right] =$$

$$= \frac{4\pi}{\tilde{C}_l} \underset{l \to \infty}{\cong} \frac{4\pi\sigma_1^2}{(l-1)^2} \ .$$

This shows that the estimation errors $\varepsilon_{lm} = \hat{T}_{lm} - T_{lm}$ are such that

$$E \left\{ \sum_{lm} \varepsilon_{lm}^2 \right\} \cong \sum_l \frac{(2l+1) 4\pi\sigma_1^2}{(l-1)^2} = +\infty \ .$$

For this reason the stochastic process $\sum \varepsilon_{lm} Y_{lm}$ has no realizations in $L^2(S)$ as it happens to $\sum \hat{T}_{lm} Y_{lm}$ as well.
Naturally by repeating the reasoning with a slightly less demanding norm, like $\sum \varepsilon_{lm}^4$, we would show that indeed T has realizations in a space of functions like L^4 with probability 1.
In conclusion we find once more that the function T (or better its trace on S) can be estimated as a whole, its estimate being in L^p, $(p > 2)$, but not in H_X, as we know.

<u>Example 3</u>

Let us assume that we have N images of a piece of land characterized by a horizontal projection Q;

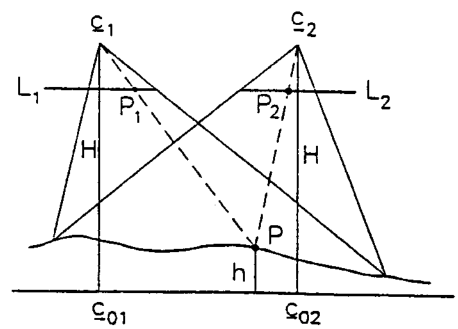

Fig. 4.1

on each image L_i we measure the grey density g_i at plate points P_i which are images of the point P ; what is really measured is the grey content over pixels (cells) of constant size; these pixels are so small and so many that we want to describe the measurement fields as continuous.

So we shall assume that the measurement density is constant as well as the noise variance and that these are the same for all images.

The basic observation equations are derived from a particular hypothesis on the radiation properties of the land surface, i.e. we shall assume it to be a Lambertian surface, so that the density $g_i(P_i)$ on the images is exactly the same as that seen by an observer looking at P from infinity in zenithal direction; we will call it the orthophoto density G(P).

The main unknowns of this problem will be the field G(P) as well as the field of geometric heights h(P): let us see how these enter into the observational model.

To make things easier we shall assume a very simple geometry, although it would be possible to develop a complete theory along these lines; so we assume that the projection centers $\underline{c}_i$ are all at the same height H and that the "plates" (or the instruments) collecting the images are horizontal and oriented parallel to each other. Centers $\underline{c}_i$ and orientations are known in a terrestrial coordinate system.

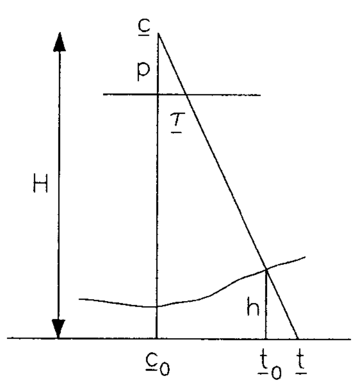

Fig.4.2 - The geometry of the projectivity: s = p/H is the scale of the image

The relation between the horizontal coordinates $\underline{t}_o = (t_{o1}, t_{o2})$ of a point $\underline{P}$ and the plate coordinates $\underline{\tau} = (\tau_1, \tau_2)$ of the image of P are given by the equation (cfr. Fig.4.2)

$$\begin{cases} \underline{t} - \underline{c}_0 = \dfrac{1}{s}\underline{\tau} \\[2ex] \underline{t} - \underline{t}_0 = \dfrac{h(\underline{t}_o)}{H}(\underline{t} - \underline{c}_0) \end{cases} ; \qquad (4.38)$$

given $\underline{t}_o$ and h we can compute $\underline{t}$ from the second and $\underline{\tau}$ from the first or given $\underline{\tau}$ and h we can derive $\underline{t}$ from the first and then $\underline{t}_o$ from the second. This process is indeed non-linear; however, if we assume that h can have only small variations as compared to H, we can directly approximate the second of (4.38) with the equation

$$\underline{t} - \underline{t}_0 = \frac{h(\underline{t})}{H}(\underline{t} - \underline{c}_0) . \qquad (4.39)$$

Now we can write the model equations in the form

$$g_i(\underline{P}_i) = G(\underline{P}) \qquad (4.40)$$

which can be parameterized as

$$g_i(\underline{\tau}_i) = G(\underline{t}_o) \qquad (i=1,2...N) \qquad (4.41)$$

and, since $\underline{\tau}_i$ is in one to one correspondence with $\underline{t}$ through the relation $\underline{\tau}_i = s(\underline{t} - \underline{c}_{oi})$, we can rewrite (4.41) in the form

$$g_i(\underline{t}) = G(\underline{t}_o) \quad . \qquad (4.42)$$

Written in this form the equation would be completely outside the theory presented here, because it implies a non-linear relation with $h(\underline{t})$, however we can readily linearize it using (4.38) and we find

$$g_i(\underline{t}) \cong G(\underline{t}) - \nabla_t G(\underline{t})\frac{h(\underline{t}_o)}{H}\cdot(\underline{t} - \underline{c}_{oi}) \cong$$

$$\cong G(\underline{t}) - \left[\frac{1}{H}\nabla_t G(\underline{t})\cdot(\underline{t} - \underline{c}_{oi})\right]h(\underline{t}) =$$

$$= G(\underline{t}) - A_i(\underline{t})h(\underline{t}) \quad . \qquad (4.43)$$

The coefficients A_i do depend from the unknown G; yet, since they don't need to be computed very accurately, they can be easily estimated by an averaging smoothed version of the observations.
These will be our basic observation equations holding for $t \in Q$ and $i = 1,2...N$.
Since the observable quantities are g_i we can then write a complete model, including the measurement noise, as

$$d\mu_{oi} = \left[G(\underline{t}) - A_i(\underline{t})h(\underline{t})\right]dt + d\mu_{wi}(\underline{\tau}); \qquad (4.44)$$

now in general the observations are on the image L_i over a small area $d\tau$ to which will correspond on the terrain an area

$$dt = \frac{1}{s^2}\, d\tau \quad . \qquad (4.45)$$

If $d\mu_{wi}$ is a Wiener measure such that

$$E\{d\mu_{wi}^2\} = \sigma^2 d\tau$$

then the same noise can be taken as being distributed on Q with variance

$$E\{d\mu_{wi}^2\} = \sigma^2 s^2 dt \quad ,$$

so the observational model has finally the form

$$d\mu_{oi} = g_i(t)dt + d\mu_{wi}(t)$$

where all the measurements are considered as taken on Q.
Here

$$\underline{y} = \begin{vmatrix} \vdots \\ g_i \\ \vdots \end{vmatrix}, \quad \underline{A} = \begin{vmatrix} 1 & \vdots \\ & -A_i(\underline{t}) \\ & \vdots \end{vmatrix}, \quad x = \begin{vmatrix} G(\underline{t}) \\ h(\underline{t}) \end{vmatrix}$$

so that

$$<\underline{y},\underline{y}>_Y = \int_Q \sum g_i^2(t)dt \quad ;$$

the vector operator $\underline{A}$ is here purely algebraic and in this case

$$<\underline{A}x,\underline{A}x>_Y = \int_Q \sum_1^N{}_i (G - A_i h)^2 dt =$$

$$= \int_Q (NG^2 - 2Gh\Sigma A_i + h^2\Sigma A_i^2)dt \qquad (4.46)$$

The quadratic form in (4.46) has to be strictly positive by hypothesis.
On the other hand if the matrix

$$K = \begin{vmatrix} N & -\Sigma A_i \\ -\Sigma A_i & \Sigma A_i^2 \end{vmatrix} \qquad (4.47)$$

were not strictly positive we would have all the $A_i(\underline{t})$ equal to each other, i.e.

$$A_i(\underline{t}) = \eta(\underline{t}) \quad , \qquad (4.48)$$

so that it could be impossible to discriminate in (4.44) between G and h, remaining estimable only their combination $G - \eta h$.
The above conclusion is pointwise, so we might have some particular area where (4.48) holds, for instance because $G(\underline{t})$ is constant so that $A_i \equiv 0$.
So, excluding this case, we have

$$\beta x^+ x \ \geq x^+ K x \geq \alpha x^+ x$$

showing that the norm (4.46) has to be equivalent to a simple L^2 norm,

$$\beta\int\{G^2 + h^2\}dt \geq <\underline{A}x,\underline{A}x>_Y \geq$$

$$\geq \alpha\int\{G^2 + h^2\}dt \quad . \qquad (4.49)$$

The fact that we change the topology in H_x does not affect our formalism, but only the definition of $\underline{A}^+$ which by the way in this case is just the algebraic transpose of $\underline{A}$.
Therefore, recalling that

$$E\{d\mu_{wi}(t)d\mu_{wj}(t)\} = \sigma^2 s^2 dt \delta_{ij}$$

so that

$$<\underline{f},C\underline{f}>_Y = \sigma^2 s^2 <\underline{f},\underline{f}>_Y \quad ,$$

we find that the sought estimator is just

$$\underline{f} = \underline{A}(\underline{A}^+\underline{A})^{-1}g \quad , \qquad (4.50)$$

with $\underline{A}^+\underline{A} = K$ (cfr.(4.47)) and with g the given functional of x

$$\langle g,x\rangle_X = \int_Q \{g_1 G + g_2 h\} dt \quad .$$

The explicit form of (4.50) is

$$f_j = \frac{1}{\Delta}\left\{\left[\sum_i A_i^2 - A_j \sum_i A_i\right] g_1 + \left[\sum_i A_i - NA_j\right] g_2\right\} \tag{4.51}$$

$$(\Delta = N \sum A_i^2 - (\sum A_i)^2) \quad ,$$

supplying the required estimate of $\langle g,x\rangle_X$ as

$$L_{\underline{f}}(d\underline{\mu}_Y) = \int_Q \sum f_j d\mu_{oj} \quad .$$

We note that in particular if we wish to estimate a functional of G, or a functional of h, we have only to put just $g_2 \equiv 0$, or respectively $g_1 \equiv 0$.

Another remark needed here is that, in contrast to the previous two examples, neither for G nor for h it is possible to obtain an unbiased estimate: in fact the evaluation functional is not bounded in L^2 and so it cannot have a representer g of bounded H_X norm.

As a final point we report that by applying the formula (3.35) it is also possible to estimate the variance of the estimate of $\langle g,x\rangle$ which turns out to be

$$\sigma^2\left\{L_f(d\underline{\mu}_Y)\right\} =$$

$$= \int_Q \frac{1}{\Delta}\left\{ \sum_i A_i^2 g_1^2 + 2 \sum_i A_i g_1 g_2 + N g_2^2 \right\} dt.$$

5. Conclusions

We have presented a theory which seems to be sufficiently general to cope with the problem of optimal estimating a field x when continuous fields of measurements are taken on fields y_i linearly dependent on x; the relation between x and $\{y_i\}$ has to be injective and its inverse has to be surjective like in classical least squares schemes. Essential to solve the problem has been the understanding that the correct way of describing a "white noise" disturbance of the observations is through the concept of Wiener measures, what has forced us to write the observation equations in terms of measures or in a weak form exploiting the concept of Wiener integral.

A particular care has been given to the process of transition from a discrete to a continuous model and the problem of their equivalence has been assessed in terms of a kind of generalized Nyquist relation (cfr. (2.47) or (2.48)).

The approach seems to be useful to perform error propagation analysis and to produce approximate estimators when large amounts of data have to be treated.

Some examples show how the method works in very different instances ; among them the case of overdetermined boundary value problem is of particular importance and has required a careful analysis, first of all in order to define what is a stochastic boundary value problem; on the other hand in this field the theory has already been fruitfully applied for both noise propagation analyses and simulations (cfr. Sacerdote F., Sanso' F., 1991; Brovelli M., Migliaccio F., Sanso' F., 1991).

This type of estimation problems can be (and in fact have been) generalized to cases in which part of the observations are discrete (this means that one type of measurements is performed at such a small number of points, as compared with the other observations, that we could not describe it as a continuum) and also the unknown x includes, beyond one or more functions, a number of discrete parameters.

The main limit of this theory is its restriction to linear problems. However the generalization of this approach to non-linear observation equations seems to be considerably more difficult requiring Ito's theory of the representation of non-linear functionals of a white noise and a deeper understanding of non-linear estimation theory.

Yet this question deserves to be studied because first of all many problems are structurally strongly non-linear and even more also those problems which we would guess to be weakly non-linear need first of all to be understood in their full non-linearity if we want to be able to say that in some sense a linearized problem supplies approximate solutions: for this reason the item of non-linear estimation with random fields will be object for us of further research.

References

Backus G., 1970a Inference from inadequate and inaccurate data: I, Proceedings of the National Academy of Sciences, 65, 1, 1-105.

Backus G., 1970b Inference from inadequate and inaccurate data: II,Proceedings of the National Academy of Sciences, 65, 2, 281-287.

Backus G., 1970c Inference from inadequate and inaccurate data: III, Proceedings of the National Academy of Sciences, 67, 1, 282-289.

Bassanino M., Migliaccio F., 1991 "A BVP approach to the reduction of spaceborne GPS and accelerometric observations" IUGG XX General Assembly - IAG - Wien, 11-24 August.

Brovelli M., Migliaccio F., Sanso' F., 1991 "A BVP approach to the reduction of spaceborne gradiometry : theory and simulations" IUGG XX General Assembly - IAG - Wien, 11-24 August.

Dermanis A., 1991 "A unified approach to linear estimation and prediction" IUGG XX General Assembly - IAG Sect.IV - Wien, 11-24 August.

Hida T., 1980 "Brownian Motion" - Springer Verlag.

Ito R., 1984 "Foundations of stochastic differential equations in infinite dimensional spaces" CBMS-NSF Regional Conference Series in Applied Mathematics N.47 - Society for Industrial and Applied Mathematics.

Keller W., 1989 "On the Treatment of an Overdetermined BVP by Pseudodifferential Operators" Proceedings of II Hotine - Marussi Symposium on Mathematical Geodesy - Pisa - 5-8-June.

Lamperti J., 1977 "Stochastic processes. A survey of the mathematical theory" Springer Verlag - Applied Mathematical Sciences Vol.23.

Papoulis A., 1965 "Probability, Random Variables and Stochastic Processes" McGraw Hill Book Company - New York.

Sacerdote F., Sanso' F., 1985 "Overdetermined boundary value problems in physical geodesy" Manuscripta Geodaetica Vol.10, N.3.

Sacerdote F., Sanso' F., 1991 "On a rigorous continuous model for digital photogrammetry" ISPRS - Intercommission WG III/VI - Proceedings of the tutorial on "Mathematical aspects of data analysis" - Milan, 7 May 1991.

Sanso' F., 1988 "The Wiener integral and the overdetermined boundary value problems of physical geodesy" - Manuscripta Geodaetica Vol.13, N.2.

Sanso' F., 1990 "On the foundation of various approaches to improperly posed problems" Fisica de la Tierra N.2 - Editorial de la Universidad Complutense.

Tarantola A., 1987 "Inverse Problem Theory" - Elsevier.