A Theorem of Insensitivity of the Collocation Solution to Variations of the Metric of the Interpolation Space

F. Sansò, G. Venuti
DIIAR, Sez. Rilevamento,
Politecnico di Milano, Piazza Leonardo da Vinci 32, I-20133 Milano, Italy

C.C. Tscherning
Department of Geophysics,
University of Copenhagen, Juliane Maries Vej 30, DK-200 Copenhagen Oe, Denmark

Abstract: The collocation approach to the estimation of a field from observed functionals, is known, by examples and simulations, to display a not very strong dependence from the choice of the specific reproducing kernel-covariance function.

In fact the situation is similar to the case of the dependence of least squares parameters on the weight of observations.

The paper, after recalling the basic theory according to its deterministic and stochastic interpretation, shows that the variation of the sought solution is infinitesimal with both, the variation of the metric of the intepolation space going to zero and the quantity of information carried by the observations going to hunderd percent on the specific functional of the field that we want to predict. The combined effect of the two gives an infinitesimal of the second order, namely a theorem of "insensitivity" of the solution to the metric of the interpolation space. Different simulations show the action of this particular effect.

Keywords: Collocation, Reproducing kernel, Covariance, Metric variation.

1 The two versions of collocation and the stability problem

1.a The deterministic version

This is a simple approximation version where the field $u(t)$, $(t \in T)$ with T any set compatible with the subsequent hypothesis, is assumed to belong to some Hilbert space $\mathcal{H}(T)$ and we have performed on $u(t)$ a finite number (N) of observations which can be represented as bounded linear (or linearized)[1] functionals on $\mathcal{H}(T)$, i.e. via Riesz theorem

$$L_i(u) = \langle h_i, u\rangle_{\mathcal{H}} = Y_i (i = 1, 2, \dots N) \quad (1.1)$$
$$h_i, u \in \mathcal{H} \ .$$

For the sake of simplicity we assume to have "exact observations", i.e. without measurement error. In case we want to include the evaluation functionals

$$\delta_t(u) = u(t) \qquad \forall t \in T \quad (1.2)$$

in the set of bounded functionals on $\mathcal{H}$ we know that then $\mathcal{H}$ must be a R.K.H.S., i.e. there is a $K(t, t')$ such that (this is useful but not strictly necessary!)

$$\langle K(t, t'), u(t')\rangle \equiv u(t) \ \forall t \in T, \forall u \in \mathcal{H} \ . \quad (1.3)$$

The kernel K is in biunivocal relation with the Hilbert space structure of $\mathcal{H}$, i.e. with the definition of scalar product or, equivalently, with the definition of norm in $\mathcal{H}$.

Here we have a sensible bifurcation of the theory, with two variants:

P1) we want to estimate a field u, with an optimal estimator $\widehat{u}$, satisfying (1.2) and we decide that optimal here means of minimum norm (maximum smoothing)

$$\|\widehat{u}\|^2 = \underset{\langle h_i, u\rangle = y_i}{\text{Min}} \|u\|^2 \ , \quad (1.4)$$

[1] Note: this is a consistency hypothesis allowing us to say that we do not know which $u \in \mathcal{H}$ is our field (or signal) but we know that in any way we can compute the observational functionals on it.

P2) we want to estimate some given bounded linear functional of k, i.e.

$$L_0(u) = \langle h_0, u\rangle \tag{1.5}$$

with an estimator $\widehat{y}_0(u)$ linear in the observations such that the relative error

$$\frac{|\widehat{y}_0(u) - L_0(u)|}{\|u\|} = \mathcal{E}_r(y_0, u)$$

satisfies the minimax principle

$$\underset{\widehat{y}_0 = \sum \lambda_i y_i}{\text{Min}} \ \underset{u}{\text{Max}} \, \mathcal{E}_r(y_0, u) \ . \tag{1.6}$$

The solutions are trivial and well known: let us introduce the notations

$$\begin{aligned}
&\underline{Y} = \{Y_i, i = 1 \ldots N\}\\
&\underline{h}(t) = \{h_i(t), i = 1 \ldots N\}\\
&H = \langle \underline{h}(t), \underline{h}^+(t)\rangle \equiv \langle h_i(t), h_j(t)\rangle\\
&\text{Span } (\underline{h}) = \left\{\underline{\lambda}^+ \underline{h}(t), \underline{\lambda} \in R^N\right\} \equiv S_N\\
&P = P_N = \text{orthogonal projection on } S_N.
\end{aligned}$$

As we all know the projector P can be explicitely constructed from the Grahamian H as the operator with kernel[2]

$$P(t, t') = \underline{h}^+(t) H^{-1} \underline{h}(t') \ ; \tag{1.7}$$

in fact

$$\langle P(t, t'), v(t')\rangle = 0 \quad \forall v \perp S_N$$

and $\forall w \in S_N$ (i.e. $w(t') = \underline{\lambda}^+ \underline{h}(t') = \underline{h}^+(t')\underline{\lambda}$)

$$\begin{aligned}
&\langle P(t, t'), w(t')\rangle = \\
&= \underline{h}^+(t) H^{-1} \underline{h}^+(t) H^{-1} \left\langle \underline{h}(t'), \underline{h}^+(t')\right\rangle \underline{\lambda} = \\
&= \underline{h}^+(t) H^{-1} H \underline{\lambda} = \underline{h}^+(t)\underline{\lambda} = w(t).
\end{aligned}$$

Solution of P1:

$$\widehat{u} = Pu = \underline{h}^+(t) H^{-1} \underline{y} \ ; \tag{1.8}$$

this is nothing but observing that Pu is univocally fixed by $\underline{y}$ and then invoking the famous theorem on orthogonal projections in Hilbert spaces.

Solution of P2:

We put

$$\widehat{y}_0(u) = \underline{\lambda}^+ \underline{y} = \underline{\lambda}^+ \langle \underline{h}, u\rangle = \left\langle \underline{\lambda}^+ \underline{h}, u\right\rangle$$

and note that

$$\widehat{y}_0(u) - L_0(u) = \left\langle \underline{\lambda}^+ \underline{h} - h_0, u\right\rangle . \tag{1.9}$$

Therefore, by Schwarz inequality,

$$\begin{aligned}
&\underset{u}{\text{Max}} \frac{|Y_0(u) - L_0(u)|}{\|u\|} = \underset{u}{\text{Max}} \frac{\left|\left\langle \underline{\lambda}^+ \underline{h} - h_0, u\right\rangle\right|}{\|u\|} = \\
&= \|\underline{\lambda}^+ \underline{h} - h_0\| \ .
\end{aligned} \tag{1.10}$$

Since $\underline{\lambda}^+ \underline{h} \in S_N$, it is indeed

$$\begin{aligned}
&\underset{\underline{\lambda}}{\text{Min}} \, \|\underline{\lambda}^+ \underline{h} - h_0\| \Rightarrow \underline{\lambda}^+ \underline{h} = \widehat{h}_0 = P h_0 = \\
&\underline{\lambda}^+(t) H^{-1} \langle \underline{h}, h_0\rangle
\end{aligned} \tag{1.11}$$

Therefore

$$\widehat{y}_0(u) = \left\langle \widehat{h}_0, u\right\rangle = \langle P h_0, u\rangle . \tag{1.12}$$

All that leads to a noteworthy conclusion.

Theorem: the problems P1) and P2) are dual one with respect to the other and their solutions are equivalent in the sense that

$$L_0(\widehat{u}) = \langle h_0, \widehat{u}\rangle = \left\langle \widehat{h}_0, u\right\rangle = \widehat{y}_0(u). \tag{1.13}$$

□ In fact

$$\langle h_0, Pu\rangle = \langle P h_0, u\rangle$$

because P (as orthogonal projector) is selfadjoint. □

Therefore we might conclude that the deterministic collocation problem can be formulated along two dual variants leading to one and the same solution.

Now our purpose is to study the stability of the solution (1.8) when we give some change to the metric of the space $\mathcal{H}$; the result will be that it is difficult to say how much $\widehat{u}$ varies in norm, but one can much better study the variation of $\widehat{y}_0 = \langle h_0, \widehat{u}\rangle$, in terms of the relative error (1.6), showing that this becomes less and less sensitive to a norm change in $\mathcal{H}$, the closer h_0 gets to S_N. This in particular has to happen for N sufficiently large. A motivation for looking into this

[2] Note: this implicitely assumes that $\{h_i(t)\}$ are linearly independent functionals, what seems very reasonable since observing here without noise we never do redundant observations.

problem is first of all a logical one; in fact as far as the choice of the particular norm, among the many equivalent, defining the same Hilbert space $\mathcal{H}$ as a set, is arbitrary we would like to be sure that the result does not depend critically on it. Moreover, there are cases (for instance when $\mathcal{H}$ has a reproducing kernel K) in which a relatively small change in $K(t,t')$ can give a much simpler structure to H, simplifying the numerical burden of calculating $H^{-1}\underline{y}$, which could be very large if we have very many data and H is not a sparse matrix.

1.b The stochastic version

In this case $u = u(t,\omega)$ is interpreted as a generalized random field (GRF) (we assume zero average for simplicity on the Hilbert space $\mathcal{H}$ [Sansò, 1986].

This means that $u(t,\omega)$ has realizations which do not belong to $\mathcal{H}$ [Tscherning, 1977] nevertheless we can give a meaning to expressions like

$$L_i(u) = \langle h_i, u\rangle = Y_i \qquad (1.14)$$

as r.v. on the space $H(u)$ obtained in $\mathcal{L}^2(\omega)$[3] by closing the linear combinations of the type $\sum \lambda_i u(t_i\omega)$.

Already the idea that $u(t,\omega) \in \mathcal{L}^2(\omega)$ implicitely means that $\mathcal{H}$ must have a reproducing kernel; as a matter of fact, recalling the definition of covariance operator [Lauritzen, 1973]

$$E\{\langle \delta_t, u\rangle \langle \delta_{t'}, u\rangle\} = \langle \delta_t, C\delta_{t'}\rangle = C(t,t') \qquad (1.15)$$

which is by definition the covariance function of u. If we choose that the covariance operator C has to be the identity, we see that

$$C(t,t') = \langle \delta_t \delta_{t'}\rangle = K(t,t'), \qquad (1.16)$$

i.e. the covariance function of the GRF u coincides with the reproducing kernel of $\mathcal{H}$. Otherwise stated: we use as Hilbert space to produce the approximate solutions, the space which has $C(t,t')$ as reproducing kernel. This opens the doors to the empirical estimation of $K(t,t')$, if we assume suitable invariance properties for u.

As we know in this case the problem of approximating $\widehat{u}$ from observations (1.14) cannot be put in the form of problem P1), but only problem P2) is meaningful: namely we want to approximate the r.v.

$$Y_0 = \langle h_0, u\rangle \qquad (1.17)$$

from the observation vector (1.14) $\underline{Y}$. The approximation now is evaluated in terms of $\mathcal{L}^2(\omega)$, i.e. in terms of variance of

$$Y_0 - \underline{\lambda}^+\underline{Y} = \langle h_0 - \underline{\lambda}^+\underline{h}, u\rangle \qquad (1.18)$$

In fact the Wiener-Kolmogorov principle is formulated as: find $\underline{\lambda}$ such that

$$E\{[Y_0 - \underline{\lambda}^+\underline{Y}]^2\} = \mathrm{Min}\,.$$

But, recallig that $C \equiv I$ in $\mathcal{H}$,

$$E\{[Y_0 - \underline{\lambda}^+\underline{Y}]^2\} = \|h_0 - \underline{\lambda}^+\underline{h}\|^2 \qquad (1.19)$$

so that we find as before the solutions

$$\widehat{h}_0 = \underline{h}^+(t)H^{-1}\langle \underline{h}, h_0\rangle = Ph_0 \quad (1.20)$$

$$\widehat{Y}_0 = \langle \underline{h}_0, h\rangle = \underline{\lambda}^+\underline{Y} = \langle h_0, \underline{h}^+\rangle H^{-1}y\ . \qquad (1.21)$$

while (1.20) is exactly the same as (1.11), (1.2) is different from (1.12) in that now $\widehat{Y}_0$ and $\underline{Y}$ are r.v. This is the reason why for the deterministic version we will study the variation of $\widehat{Y}_0$ in terms of its maximum relative variation, while for the stochastic versions, one should consider the variation of $\widehat{Y}_0$ in terms of its variance.

Now if we observe that

$$\langle h_0, \underline{h}\rangle = L_{0t}\underline{L}_{t'}C(t,t')$$
$$\langle \underline{h}, \underline{h}^+\rangle = L_t\underline{L}_{t'}C(t,t'),$$

we immediately see that to study the sensitivity of (1.21) to norm changes is the same as studying its sensitivity to covariance changes. In fact this follows from our choice

$$C = I \qquad C(t,t') = K(t,t')$$

and the fact that K identifies the metric in $\mathcal{H}$.

This explains the importance of this study because typically our model for the covariance function is not very much close to empirical values and these in any way are themselves not true values.

Then it is very consolating if we can affirm that the final solution is only weakly dependent on the choice of $C(t,t')$, at least when h_0 can be predicted reasonably well from $\underline{\lambda}^+\underline{h}$.

[3]Note: this is the space of r.v. with finite variance on the same probability space $(\Omega, \mathcal{A}, P)$ on which $u(t,\omega)$ is defined.

2 The solution stability theorem

The purpose of this paragraph is to prove a theorem of stability of the solutions (1.12), (1.21) with respect to first order variations in the definition of norm in $\mathcal{H}$.

This will be done through 7 statements: the first 1), 2), 3) statements establish a general representation of the variation of the norm in $\mathcal{H}$ and the subsequent variations (to the first order) of Riesz representers and of their mutual scalar products; statements 4) and 5) study the stability of $d\widehat{Y}_0$ as a general function of the norm variation in $\mathcal{H}$ and in the particular case that this is given through a reproducing kernel K; statements 6) and 7) apply the former case to the stochastic version, giving a more precise meaning to the concept of a small variation of the covariance/reproducing kernel $dK = dC$.

STATEMENT 1: let $\mathcal{H}$ be given with scalar product $\langle , \rangle$ and assume that to the same set of functions $\mathcal{H}$ is given a slightly different metric with scalar product $\langle , \rangle'$ such a way that the new norm is equivalent to the former; then we have for a suitable bounded selfadjoint operator dQ in $\mathcal{H}$

$$\langle n, v\rangle' = \langle u, (I + dQ)v\rangle \tag{2.1}$$

with

$$0 < \alpha^2 < (I + dQ) < \beta^2 \ . \tag{2.2}$$

□ This is a simple consequence of Lax Milgram theorem [Yosida, 1978], implying the existence of a bounded symmetric operator Q such that

$$\langle u, v\rangle' = \langle u, Qv\rangle \ ;$$

it is then enough to put

$$Q = I + dQ$$

and to recall that by hypothesis

$$\alpha\|u\| \le \|u\|' \le \beta\|u\|$$

to prove the statement.

Let us notice that

$$(I + dQ) \ge \alpha^2 I$$

implies existence and boundedness of $(I+dQ)^{-1}$. Moreover, since dQ is assumed as "small", e.g. such that $\|dQ\| < 1$, we have as well the relation

$$(I + dQ)^{-1} = I - dQ$$

exact to the first order in dQ. □

STATEMENT 2: when a linear bounded functional $L(u)$ is directly defined on $\mathcal{H}$, we know that, through Riesz theorem, there is a representer h of L such that

$$L(u) \equiv \langle h, u\rangle \ ; \tag{2.3}$$

now if we change the scalar product in $\mathcal{H}$, also the representer h will change to h' such that

$$L(u) \equiv \langle h', u\rangle' \ ; \tag{2.4}$$

if we put

$$h' = h + dh$$

then it is, to the first order,

$$dh = -dQh \tag{2.5}$$

□ Equating (2.4) and (2.5) and recalling (2.1) we get to the first order

$$\begin{aligned}\langle h, u\rangle &= \langle h', u\rangle' = \langle h + dh, (I - dQ)u\rangle = \\ &= \langle h, u\rangle + \langle dh, u\rangle + \langle h, dQu\rangle = \\ &= \langle h, u\rangle + \langle dh + dQh, u\rangle \ , \quad \forall u \in \mathcal{H} \ ,\end{aligned}$$

yelding (2.5). □

STATEMENT 3: let h_1, h_2 be two representer of L_1, L_2 with the scalar product $\langle , \rangle$ and h_1', h_2' the representers of the same functionals with the scalar product $\langle , \rangle'$; then we have

$$\begin{aligned} d\{\langle h_1, h_2\rangle\} &= \langle h_1', h_2'\rangle' - \langle h_1, h_2\rangle = \\ &= -\langle dQh_1, h_2\rangle = -\langle h_1 dQh_2\rangle \ . \end{aligned} \tag{2.6}$$

□ From (2.5) and (2.1) we can write to the first order in dQ

$$\begin{aligned}\langle h_1', h_2'\rangle' &= \langle (I - dQ)h_1, (I + dQ)(I - dQ)h_2\rangle = \\ &= \langle (I - dQ)h_1, h_2\rangle = \langle h_1, h_2\rangle - \langle dQh_1, h_2\rangle \ ,\end{aligned}$$

because $(I-dQ)(I+dQ)=I-dQ^2$. This proves the first of (2.6): the second follows because dQ is selfadjoint.

We note that from (2.6) the formulas

$$d\langle h_0,\underline{h}\rangle=-\langle h_0,dQ\underline{h}\rangle \quad (2.7)$$

$$dH=d\langle \underline{h},\underline{h}^+\rangle=-\langle \underline{h},dQ\underline{h}^+\rangle \quad (2.8)$$

follow. □

STATEMENT 4: let $\widehat{y}_0$ be the estimator of y_0

$$\widehat{y}_0=\langle h_0,\underline{h}^+\rangle H^{-1}\underline{y} \ : \quad (2.9)$$

if we give a variation to the scalar product in $\mathcal{H}$, holding fixed the observations $\underline{y}$, we get a variation $d\widehat{y}_0$ such that

$$\sup_{\|u\|=1}|d\widehat{y}_0|\equiv\|PdQ(I-P)h_0\|\leq \quad (2.10)$$
$$\leq\|dQ\|\cdot\|(I-P)h_0\|$$

□ Recalling (2.7), (2.8) and (1.7) we find

$$\begin{aligned} d\widehat{y}_0 &= -\langle h_0,dQ\underline{h}^+\rangle H^{-1}\underline{y}+\\ &\quad +\langle h_0,\underline{h}^+\rangle H^{-1}\langle\underline{h},dQ\underline{h}^+\rangle H^{-1}\underline{y}=\\ &= -\langle dQ(I-P)h_0,\underline{h}^+\rangle H^{-1}\langle\underline{h},u\rangle=\\ &= \langle dQ(I-P)h_o,Pu\rangle=\\ &= -\langle PdQ(I-P)h_0,u\rangle \ . \end{aligned}$$

Then the first of (2.10) follows by Schwarz inequality and the second from the fact that for any two bounded operators A,B we have $\|AB\|\leq\|A\|\|B\|$ and $\|P\|=1$. □

Remark: formula (2.10) is our fundamental result for the deterministic version of collocation. In fact (2.10) shows first that the maimum relative error, $\frac{|d\widehat{y}_0|}{\|u\|}=\mathcal{E}_r$, is indeed linear in $\|dQ\|$; but on the same time we have that the sensitivity of $\mathcal{E}_r$ to $\|dQ\|$ is controlled by $\|(I-P)h_0\|$. Therefore, if h_0 is close enough to S_N, or if we assume that N increases with

$$\left[\bigcup_N S_N\right]\equiv\mathcal{H} \ ,$$

then this sensitivity is itself small or it goes to 0 when $N\to\infty$ because $Ph_0\to h_0$ in $\mathcal{H}$.

STATEMENT 5: if $\mathcal{H}$ is endowed with a reproducing kernel $K(t,t')$ with the scalar product $\langle,\rangle$ and with $K'(t,t')=K(t,t')+dK(t,t')$ with the scalar product $\langle,\rangle'$, then the operator dQ corresponds to

$$dQu=-\langle dK(t,t'),u(t')\rangle \ , \quad \forall u\in\mathcal{H} \quad (2.11)$$

□ In fact, to the first order and $\forall u\in\mathcal{H}$

$$\begin{aligned} u(t) &\equiv \langle K'(t,t'),u(t')\rangle'\equiv\\ &\equiv \langle K(t,t')+dK(t,t'),(I+dQ)u\rangle\equiv\\ &\equiv \langle K(t,t'),u(t')\rangle+\langle dK(t,t'),u(t')\rangle+\\ &+ \langle K(t,t'),(dQu)(t')\rangle\equiv u(t)+(dQu)(t)+\\ &+ \langle dK(t,t'),u(t')\rangle \ , \end{aligned}$$

which proves (2.11).

Let us notice also that from

$$\langle u_1,u_2\rangle=L_{1t}L_{2t'}K(t,t') \quad (2.12)$$

we derive straightforwardly

$$\begin{aligned} d\langle u_1,u_2\rangle &= -\langle h_1,dQh_2\rangle=\\ &= L_{1t}L_{2t'}dK(t,t') \end{aligned} \quad (2.13)$$

as one can also directly verify. □

STATEMENT 6: let $\widehat{Y}_0$ be the random estimator

$$\widehat{Y}_0=\langle h_0,\underline{h}^+\rangle H^{-1}\underline{Y} \ ,$$

and dQ the operator of the variation of the metric in $\mathcal{H}$, then

$$\begin{aligned} E\{d\widehat{Y}_0^2\} &= \|PdQ(I-P)h_0\|^2\leq\\ &\leq\|dQ\|^2\|(I-P)h_0\|^2 \ . \end{aligned} \quad (2.14)$$

□ In fact, as we saw in Statement 4,

$$d\widehat{Y}_0=-\langle PdQ(I-P)h_0,u\rangle \ . \quad (2.15)$$

Now we have only to recall that u is a G.R.F. on $\mathcal{H}$ with covariance operator $C\equiv I$ (i.e. with covariance function $C(t,t')\equiv K(t,t')$) and (2.14) descends immediately. □

Remark: this is our second fundamental result for the stochastic version. It seems noteworthy that it coincides perfectly with the result (2.10) for the deterministic version.

STATEMENT 7: when $\mathcal{H}$ is endowed with a reproducing kernel $K(t,t')$, including the stochastic version when $K(t,t') = C(t,t')$, the factor controlling the stability of either solutions $\widehat{y}_0, \widehat{Y}_0$ can be explicitely computed by the formula

$$\begin{cases} \|P\ dQ(I-P)h_0\|^2 = \underline{\Lambda}^+ H^{-1} \underline{\Lambda} \\ \underline{\Lambda} = \underline{L}_t L_{0t'} dK(t,t') + \\ - \left[\underline{L}_t \underline{L}_{t'}^+ dK(t,t')\right] H^{-1} \left[\underline{L}_t L_{0t'} K(t,t')\right] \ . \end{cases}$$

□ Recall that

$$P^2 = P = \underline{h}^+(t) H^{-1} \underline{h}(t')$$

so that we need only to prove that

$$-\underline{\Lambda} = \langle \underline{h}, dQ(I-P)h_0 \rangle \tag{2.16}$$

coincides with the second of (2.16). □

To this aim let us adopt a short hand notation like

$$\begin{aligned} \underline{h}(t) &= \underline{L}_{t'} K(t,t') = K(t,\underline{L}) \\ h_0(y) &= L_{0t'} K(t,t') = K(t,L_0) \\ \langle h_0(t), \underline{h}(t) \rangle &= L_{0t'} \underline{L}_{t''} \langle K(t,t'), K(t,t'') \rangle = \\ &= L_{0t'} \underline{L}_{t''} K(t',t'') = K(L_0,\underline{L}) \end{aligned}$$

and so forth.

Then from (2.16)

$$\begin{aligned} &\underline{\Lambda} = -\langle dQ\underline{h}, (I-P)h_0 \rangle \\ &(-dQ\underline{h})(t) = \langle dK(t,t'), \underline{h}(t') \rangle = \\ &= \langle dK(t,t'), K(t',\underline{L}) \rangle = dK(t,\underline{L}) \\ &\left[(I-P)h_0\right](t) = h_0(t) + \\ &-K(t,\underline{L})^+ H^{-1} \langle \underline{h}(t'), h_0(t') \rangle = \\ &= K(t,L_0) - K(t,\underline{L}^+) H^{-1} K(\underline{L},L_0) \ ; \end{aligned}$$

summarizing [4]

$$\begin{aligned} &\underline{\Lambda} = \langle dK(t,\underline{L}), K(t,L_0) + \\ &-K(t,\underline{L}^+) H^{-1} K(\underline{L},L_0) \rangle = \\ &= \langle dK(t,\underline{L}), K(t,L_0) \rangle + \\ &- \left\langle dK(t,\underline{L}), K(t,\underline{L}^+) \right\rangle H^{-1} K(\underline{L},L_0) = \\ &= dK(\underline{L},L_0) - dK(\underline{L},\underline{L}^+) H^{-1} K(\underline{L},L_0) \end{aligned}$$

which indeed coincides with (2.16).

Remark: the meaning of the formula (2.16) is that it allows the computation of the amplitude of the variations $\frac{d\widehat{y}_0}{\|u\|}, d\widehat{Y}_0$, through formulas (2.10), (2.14), thus understanding to what extent they can be considered as small; (of course in one case one has a deterministic norm, in the other one has a mean square value as measure of this amplitude).

To this purpose (2.16) should be compared with the corresponding (squared) estimation error, namely with

$$\begin{aligned} &\|(I-P)h_0\|^2 = \\ &= \|Kh_0\|^2 - \|Ph_0\|^2 = \\ &= K(L_0,L_0) - K(L_0,\underline{L}^+) H^{-1} K(\underline{L},L_0) \ . \end{aligned} \tag{2.17}$$

Remark: (2.17) represents the squared relative error for the deterministic case $\left(\frac{|\widehat{y}_0 - y_0|^2}{\|u\|^2}\right)$ and the absolute mean square error in the stochastic case; this because in the stochastic te information case $K(t,t') = C(t,t')$ includes already information on the amplitude of the signal, while in the deterministic case the $\|u\|$ is completely unknown.

3 Numerical tests and conclusions

In this paragraph we try to illustrate numerically the theoretical results of §2. Since, at least in part, they are based on inequalities we cannot make a very precise test; nevertheless if we take (2.10) and (2.14) in the sense that approximately

$$E\{d\widehat{Y}_0^2\} \cong \|dQ\|^2 \|(S-P)h_0\|^1 \tag{3.1}$$

where,

$$d\widehat{Y}_0 = \widehat{Y}_{01} - \widehat{Y}_{02} \tag{3.2}$$

$\widehat{Y}_{01}$ = estimate with the correct covariance C
$\widehat{Y}_{02}$ = estimate with the modified covariance $C + dC$,

we can find a statement which can be checked numerically.

[4]Note: remember that $K(t,t')$ is symmetric but when t,t' are substituted by vectors the order of the product of vectors and their transpose should be kept.

Namely assume that we have a certain configuration of observables leading to an orthogonal projector P^1; correspondingly we can compute both $\widehat{Y}_{01}^1, \widehat{Y}_{02}^1$ and the mean square variation of the estimator $\widehat{Y}_0$ as

$$(V^1)^2 = E\{d\widehat{Y}_0^{1\,2}\} \tag{3.3}$$

as well as the theoetical estimation error

$$(\mathcal{E}^1)^2 \cong \|dQ\|^2 \|(I - P^1)h_0\|^2 \ . \tag{3.4}$$

To be explicit, hereafter we shall use the upper index 1 or 2 to denote two different configurations and a lower index 1 or 2 to denote prediction with C or $C + dC$ covariances.

Formulas (3.3) and (3.4) refer to configuration 1 and to certain fixed dC, h_0.

Similarly we can repeat the same computation with another configuration 2, thus deriving the corresponding $(V^2)^2$ and $(\mathcal{E}^2)^2$. Consequently, under our hypothesis, we can claim that

$$\frac{(V^1)^2}{(V^2)^2} \cong \frac{(\mathcal{E}^1)^2}{(\mathcal{E}^2)^2} \tag{3.5}$$

and this is of course easy to verify.

Let us examine two examples.

Example 1: here we have use purely simulated data and evaluation functionals only for a process in 1D.

More precisely, assume that $u(t)$ is a stochastic process with zero mean and covariance

$$C(\tau) = 4e^{-0{,}1|\tau|} \ ; \tag{3.6}$$

the wrong (or modified) covariance function for our experiment is

$$C(\tau) + \delta C(\tau) = 0.4\ e^{-0{,}5|\tau|} \tag{3.7}$$

Now we define as configuration 1 of the observation vector $\underline{Y}$ and as functionals to be predicted Y_0 precisely

$$\underline{Y} = \left| \begin{matrix} u(-2) \\ u(-1) \\ u(1) \\ u(2) \end{matrix} \right| \qquad Y_0 = u(0) \tag{3.8}$$

Of course form $\underline{Y}$ we can perform the prediction (with 1 upper index)

$$\widehat{u}^1(0) = \widehat{Y}_{01}^1 = [C(2)C(1)C(1)C(2)] \cdot \tag{3.9}$$

$$\cdot \begin{bmatrix} C(0) & C(1) & C(3) & C(4) \\ C(1) & C(0) & C(1) & C(3) \\ C(3) & C(1) & C(0) & C(1) \\ C(4) & C(3) & C(1) & C(0) \end{bmatrix}^{-1} \begin{bmatrix} u(-2) \\ u(-1) \\ u(1) \\ u(2) \end{bmatrix}$$

If in (3.9) we use $C + dC$ instead of C, we derive the modified estimator $\widehat{Y}_{02}^1$ and we can compute the variation $d\widehat{Y}_0^1 = \widehat{Y}_{01}^1 - \widehat{Y}_{02}^1$.

Now we have independently sampled our process producing 10 vectors and 10 scalars

$$\left(\underline{Y}_i^1, Y_i^1 \qquad i = 1, \dots 10\right)$$

and then we computed the sample variance

$$(V^1) = E\{(d\widehat{Y}_0^1)\} \cong \frac{1}{10} \sum_{i=1}^{10} \left(\widehat{Y}_{01i}^1 - \widehat{Y}_{02i}^1\right)^2 \tag{3.10}$$

as well as the sample estimation error

$$(\mathcal{E}^1)^2_{emp} \cong \frac{1}{10} \sum_{i=1}^{10} \left(\widehat{Y}_{01i}^1 - \widehat{Y}_{0i}^1\right)^2 \tag{3.11}$$

which in any way has to be compared with the theoretical expression

$$(\mathcal{E}^1)^2 = C(0) - \sum_{i,k} C(\tau_i) C^{(-1)}(\tau_i \tau_k) C(\tau_k) \tag{3.12}$$

The same procedure adopted for configuration 1 has been repeated for configuration 2 consisting of

$$\begin{aligned} &\tau_i = -2,\ -1.5,\ -1,\ -0.5,\ 0.5,\ 1,\ 1.5,\ 2 \\ &\tau_0 = 0 \\ &\tau_i = \text{measurement points} \\ &\tau_0 = \text{prediction point.} \end{aligned}$$

the three indexes $(V^2)^2, (\mathcal{E}^2)^2_{emp}$ and $(\mathcal{E}^2)^2$ have also been recomputed obtaining the results shown in Table 3.1.

	Conf. 1	Conf. 2	Ratios
$\mathcal{E}$	0.63	0.45	1.40
$\mathcal{E}_{emp}$	0.59	0.54	1.09
V	0.26	0.069	3.77

Table 3.1

As we see the experiment provides ratios of the same order of magnitude but not really equal one to the other. It is opinion of the authors that this reflectes the fact that we are approximating an inequality with an equality, which might be too crude a hypothesis.

Example 2: in this case a more realistic example has been worked out from a classical geodetic problem, namely the geoid estimation from gravity anomalies where instead of using many independent samples we have rather computed many sample points $\Delta g(P_{ik})$ and predicted the geoid at the same points $\widehat{N}(P_{ik})$ for one model gravity field only. The model gravity field used here is EGM96 from degree 8 to degree 180. Also the covariances used in the experiment are the true ones

$$C(\psi) = \left(\frac{\mu}{R}\right)^2 \sum_{n=8}^{180} C_n P_n(\cos\psi) \qquad (3.13)$$

$$C_n = \sum_{m=-n}^{m} T_{nm}^2 \qquad n = z_1 \ldots 180$$

$$T_{nm} = \text{exact EGM coefficients}$$

and two modified covariances

$$C(\psi) + dC(\psi) = \left(\frac{\mu}{R}\right)^2 \sum_{n=M}^{180} C_n P_n(\cos\psi) \qquad (3.14)$$

where C_n are as in (3.13) and M is taken once as $M = 2$ (lower index 2) and once as M=4 (lower index 3).

The three covariances are represented in Fig. 3.1.

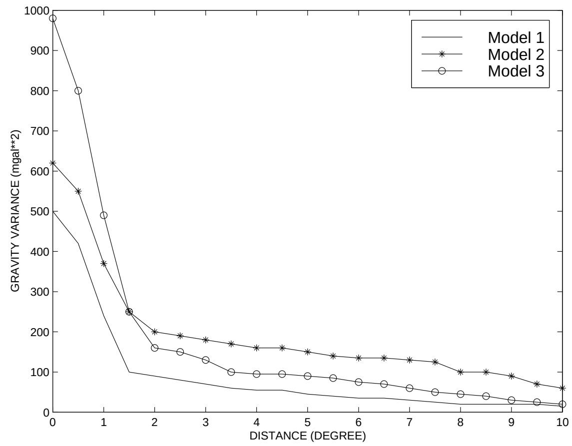

Figure 3.1: The three model covariances

Note also here that in (3.14) the modification of C is obtained by including more degree variances than in (3.13); in fact without this rather drastic change in the covariance the different prediction $\widehat{N}_1(P_{ik}), \widehat{N}_2(P_{ik})$ were so close to one another that the differences become physically insignificant.

In this case the two configurations correspond to two regular grids on the sphere, namely

$$\text{Conf.1} \rightarrow 10^\circ \times 10^\circ \text{grid}$$
$$\text{Conf.2} \rightarrow 5^\circ \times 5^\circ \text{grid.}$$

For each of the two grids and each of the three covariance functions (3.13), (3.14) with $M = 2, 4$, from a synthetic data set $\Delta g(P_{ik})$ the corresponding predictions of geoid undulations have been performed through well-known formulas

$$\gamma \widehat{N}_{ik} = \sum_{\ell,m,j,n} C_{T\Delta g}(P_{ik}, P_{\ell m}) C^{(-1)}_{\Delta g \Delta g}(P_{\ell m}, P_{jn}) \Delta g_{jn} \qquad (3.15)$$

$(\gamma \cong \mu/R^2)$

thus obtaining two series of three sets of predicted values $\widehat{N}_1(P_{ik}), \widehat{N}_2(P_{ik}), \widehat{N}_3(P_{ik})$ (we skip here the upper index 1 or 2).

As it is standard the covariances and cross covariances on (3.15) are related to $C(\psi)$ by

$$C_{T\Delta g}(\psi) = R\left(\frac{\mu}{R^2}\right)^2 \sum_{n=M}^{180} (n-1) C_n P_n(\cos\psi) \tag{3.16}$$

$$C_{\Delta g \Delta g}(\psi) = \left(\frac{\mu}{R^2}^2\right) \sum_{n=M}^{180} (n-1)^2 C_n P_n(\cos\psi) \tag{3.17}$$

with $M = 8, 2, 4$ respectively in the three cases.

The results are summarized in Table 3.2 where in particular the mean square differences

$$(V_{12})^2 = \frac{1}{N} \sum_{(i,k)} \left[\widehat{N}_2(P_{ik}) - \widehat{N}_1(P_{ik})\right]^2$$

$$(V_{13})^2 = \frac{1}{N} \sum_{(i,k)} \left[\widehat{N}_3(P_{ik}) - \widehat{N}_1(P_{ik})\right]^2$$

are reported in units of m^2,

	Conf. 1	Conf. 2	Ratios
$\mathcal{E}$	4.40	3.41	1.29
$\mathcal{E}_{emp}$	4.40	3.16	1.39
V_{12}	6.50	5.50	1.18
V_{13}	4.60	3.50	1.31

Table 3.2

As we can see the ratios are again in a very reasonable agreement with our hypotesis, the differences being of the same order as the difference between the ratios of empirical values and theoretical values of $\mathcal{E}^2$.

The conclusion is that the analysis performed is confirmed and the dependence of the solution of collocation on the kernel/covariance choice is not very critical particularly where the density of measurement points is high.

References

[Lauritzen, 1973] Lauritzen S.: 1973, The probabilistic background of some statistical methods in physical geodesy. Danish Geodetic Inst., no 48, Copenhagen 1973.

[Rozanov, 1982] Rozanov Y.A.: 1982, Markov random fields. Springer Verlag 1982.

[Sansò, 1986] Sansò F.: 1986, Statistical methods in physical geodesy. Lecture notes in earth sciences, vol 7,"Mathematical and numerical techniques in physical geodesy". H. Suenkel ed, Springer Verlag, 1986

[Sansò, Venuti, 1998] F. Sansò, G. Venuti: 1998, White noise stochastic BVPs and Cimmino's theory. Hotine-Marussi Symp. Trento, September 1998.

[Tscherning, 1977] Tscherning C.C.: 1977, A note on choice of norm when using collocation for the computation of the approximation to the anomalous potential. Bul. Geod. vol 51, no 2, 1977.

[Yosida, 1978] Yosida K.: 1978, Functional Analysis. Springer Verlag, 1978.