Why Universal Kriging is better than IRFk-Kriging: Estimation of Variograms in the Presence of Trend

Karl Gerald van den Boogaart*& Alexander Brenning †

June 30, 2001

Abstract

Many applications of kriging demand the introduction of a trend or external drift. We can use universal kriging, when we know the variogram of the process. But there are no straight forward methods known to estimate the variogram unbiasedly or consistently in the presence of a trend with unknown parameters. Thus the class of trend surfaces used is often restricted to polynomial trends leading to the application of intrinsic random functions and generalised covariograms. From the theory of intrinsic random functions IRFk we know that it is impossible to estimate the variogram unbiasedly in presence of a polynomial trend. It is shown that this problem comes up only with polynomial and harmonic trend surfaces. Thus polynomial trends are not the best but the worst case to consider. A method based on the statistical technique of identifiable contrasts for fitting the non–stationary variogram models is proposed and applications are presented in example situations. It works unbiasedly and consistently in the presence of internal and external drift and thus overcomes the crucial problem that the variogram could not be estimated unbiasedly in universal Kriging situations. It is shown that the theory of

*Mathematics and Computer Sciences in Geology, Freiberg University of Mining and Technology, FRG

†Friedrich–Alexander–Universität, Institut für Geographie, Kochstr. 4/4, 91054 Erlangen, Germany, e-mail: ali@proforma.de

generalised covariograms and generalised variograms can be further generalised to all sorts of trend surfaces and gets more simple rather than more difficult when we do this.

1 Introduction

The estimation of the variogram is the crucial part of a geostatistical analysis. Especially in the presence of internal or external trend no universal techniques to estimate the variogram are known. The simple approach to remove the trend by some filtering and to estimate the empirical variogram from these filtered data in the usual way leads to a biased estimate and is therefore blamed in many textbooks about kriging [Cressie 1993][Chilès&Delfiner 1999][Goovaerts 1997][Wackernagel 1998]. A modified procedure leading to unbiased estimates along with some theory on identifiability of the variogram is given here.

2 Notation

Let $Z(\mathbf{s})$ denote a random field and $\mathbf{s}_1, \ldots, \mathbf{s}_n$ the locations of our observations of this random field. As usual for universal kriging we assume that the random field consists of the additive superposition of a second order stationary random field $Y(\mathbf{x})$ with mean zero and a deterministic trend $\beta^t \mathbf{f}(\mathbf{s})$ with a known function $\mathbf{f} : S \to I\!R^d$, $\mathbf{f}(\mathbf{s}) = (f_i(\mathbf{s}))_{i=1,\ldots,d}$ and an unknown parameter vector $\beta \in I\!R^d$:

$$Z(\mathbf{x}) = Y(\mathbf{x}) + \beta^t \mathbf{f}(\mathbf{s})$$

For simplification in practical applications let us assume $f_1(\mathbf{s}) \equiv 1$, i.e. the trend contains an unknown overall mean. Using this assumption we can develop the whole theory in terms of the more useful variograms instead of covariograms. Let

$$c(\mathbf{h}) := \mathbf{cov}(Y(\mathbf{x}), Y(\mathbf{x}+\mathbf{h}))$$

$$\gamma(\mathbf{h}) := \text{var}\left[(Y(\mathbf{x}) - Y(\mathbf{x}+\mathbf{h}))^2\right] = c(0) - c(\mathbf{h})$$

denote covariogram and the variogram of $Y(\mathbf{x})$, respectively. It is very important to estimate the variogram γ well, since it is instrumental to calculate the universal kriging prediction for $Z(\mathbf{s})$ given by

$$\hat{Z}(\mathbf{s}) = \begin{pmatrix} \mathbf{z} \\ \mathbb{O}_d \end{pmatrix}^t \begin{pmatrix} \mathbf{\Gamma} & \mathbf{F} \\ \mathbf{F}^t & \mathbb{O}_{d\times d} \end{pmatrix}^{-1} \begin{pmatrix} \boldsymbol{\gamma}(\mathbf{s}) \\ \mathbf{f}(\mathbf{s}) \end{pmatrix}$$

with following notations used throughout this contribution:

$$\begin{aligned}
\mathbf{z} &= (Z(\mathbf{s}_k))_{k=1,\ldots,n} \in I\!R^n \\
\mathbf{F} &= (f_i(\mathbf{s}_k))_{k=1,\ldots,n,\, i=1,\ldots,d} \in I\!R^{n\times d} \\
\mathbf{\Gamma} &= (\gamma(\mathbf{s}_k - \mathbf{s}_l))_{k=1,\ldots,n,\, l=1,\ldots,n} \in I\!R^{n\times n} \\
\mathbf{C} &= (c(\mathbf{s}_k - \mathbf{s}_l))_{k=1,\ldots,n,\, l=1,\ldots,n} \in I\!R^{n\times n} \\
\mathbb{O}_d &:= (0)_{i=1,\ldots,d} \in I\!R^d \\
\mathbb{O}_{d\times d} &:= (0)_{i=1,\ldots,d,\, j=1,\ldots,d} \in I\!R^{d\times d} \\
\boldsymbol{\gamma}(\mathbf{s}) &:= (\gamma(\mathbf{s} - \mathbf{s_k})_{k=1,\ldots,n}) \in I\!R^n \\
\text{and} & \\
\mathbf{y} &= (Y(\mathbf{s}_k))_{k=1,\ldots,n} \in I\!R^n
\end{aligned}$$

3 Discussion in the literature

Various methods have be discussed in the literature to estimate γ or to work around this problem. One could think of a simple approach: Estimate β, remove the trend and estimate the variogram from the residuals. However, we can find warnings that this approach is wrong since it leads to a biased estimation and largely depends on the method of estimation of β (see [Cressie 1993],[Chilès&Delfiner 1999], [Goovaerts 1997]). The same warning is given for a variogram estimated from reestimated residuals based on the estimated variogram.

The idea to remove the trend, to do ordinary kriging for the residuals and to add the trend surface afterwards given in [Goovaerts 1997] solves the problem in so far it only needs the variogram of the residuals, which can be estimated from the residuals, but curing this problem we get the side-effect of loosing the crucial assumption of stationarity, since the residuals are in general not stationary.

For translation invariant trend models, especially for polynomial trend functions, we can replace the variogram by the generalised covariogram of an intrinsic random function [Cressie 1993][Chilès&Delfiner 1999]. These functions can be estimated from the data. The problems of this approach are the difficult handling of generalised covariograms, which are defined as equivalence classes of functions, and the limitation to special trend surfaces.

Another alternative is to estimate the variogram from pairs uneffected by the trend [Goovaerts 1997]. Then the limitation is given by the difficulty for the user of the method to find such pairs.

The general ideas of ML (maximum likelihood) or REML (restricted maximum likelihood) can be easily applied to any trend. However we would need distributional assumptions and can only estimate the parameters of the model and not empirical variograms.

An alternate approach based on the method of moments is given in this paper.

4 Comparison of empirical and theoretical residual variogram

The central idea of this approach is to look at the relationship of the covariance structure of arbitrary residuals and the covariance structure of Y and Z. With $\mathbf{F}$ given in section 2 define:

$$\mathbf{P} = 1 - \mathbf{F}(\mathbf{F}^t\mathbf{F})^{-1}\mathbf{F}^t$$

The matrix $\mathbf{P}$ satisfies $\mathbf{PF} = 0$, and additionally $\mathbf{P}^2 = \mathbf{P} = \mathbf{P}^t$. It is emphasised that all conclusions given in the next section hold for any matrix $\mathbf{P}$ with the property $\mathbf{PF} = 0$. The most important corollary of $\mathbf{PF} = 0$ is :

$$\mathbf{Pz} = \mathbf{P}(\mathbf{y} + \mathbf{F}\beta) = \mathbf{Py}$$

and thus

$$\mathrm{E}[\mathbf{Pz}] = \mathbf{P}\mathrm{E}[y] = \mathbf{P}0 = 0 \tag{1}$$

We can use it to calculate the variance–covariance matrix of $\mathbf{Pz}$

$$\mathbf{Var}(\mathbf{Pz}) \;=\; \mathbf{P}\mathbf{Var}(\mathbf{y})\mathbf{P}^t = \mathbf{PCP^t} = -\mathbf{P}\Gamma\mathbf{P^t} \tag{2}$$

since $\mathbf{P}\Gamma\mathbf{P}^t = \mathbf{P}(c(0)\mathbb{1}_{n\times n} - \mathbf{C})\mathbf{P^t} = -\mathbf{PCP^t}$ due to $f_1(\mathbf{s}) \equiv 1$. From eq. (1) and (2) it follows:

$$\mathrm{E}[\mathbf{Pzz}^t\mathbf{P}^t] = -\mathbf{P}\Gamma\mathbf{P}^t$$

Thus it could be a good idea to minimise the squared difference of $\mathbf{Pzz}^t\mathbf{P}^t$ and $-\mathbf{P}\Gamma\mathbf{P}^t$ to estimate the variogram. It follows from the general principle of minimum contrast estimators,[Witting 1995].

Let $\gamma_\theta(\mathbf{h})$, $\theta \in I\!R^p$, denote a variogram model, which is pointwise two times partially differentiable in θ, and let

$$\Gamma(\theta) = (\gamma_\theta(\mathbf{s}_k - \mathbf{s}_l))_{k=1,\ldots,n,\, l=1,\ldots,n} \in I\!R^{n\times n}$$

An estimation for the parameter θ is defined in the following way:

$$\hat{\theta} = \underset{\theta}{\operatorname{argmin}} \|\mathbf{Pzz}^t\mathbf{P}^t + \mathbf{P}\Gamma(\theta)\mathbf{P}^t\|^2 \text{ with } \|(a_{ij})\|^2 := \sum_{ij} a_{ij}^2 \tag{3}$$

Later we will show that this is an estimation with good properties.

5 Derivatives of the objective function

For further considerations we need the derivatives of the objective function $\|\mathbf{Pzz}^t\mathbf{P}^t + \mathbf{P}\Gamma(\theta)\mathbf{P}^t\|^2$, which will help to proof the properties of the estimator and to calculate the estimate $\hat{\theta}$ for θ.

$$\begin{aligned}
&\frac{d}{d\theta}\|\mathbf{Pzz}^t\mathbf{P}^t + \mathbf{P}\Gamma(\theta)\mathbf{P}^t\|^2 = \\
=\ & \frac{d}{d\theta}\operatorname{tr}\left[(\mathbf{Pzz}^t\mathbf{P}^t + \mathbf{P}\Gamma(\theta)\mathbf{P}^t)^t(\mathbf{Pzz}^t\mathbf{P}^t + \mathbf{P}\Gamma(\theta)\mathbf{P}^t)\right] = \\
=\ & \frac{d}{d\theta}\operatorname{tr}\left[\mathbf{Pzz}^t\mathbf{P}^t\mathbf{Pzz}^t\mathbf{P}^t + \mathbf{Pzz}^t\mathbf{P}^t\mathbf{P}\Gamma(\theta)\mathbf{P}^t + \mathbf{P}\Gamma(\theta)\mathbf{P}^t\mathbf{Pzz}^t\mathbf{P}^t + (\mathbf{P}\Gamma(\theta)\mathbf{P}^t)^2\right] \\
=\ & \left[2\operatorname{tr}\left(\left(\mathbf{P}\Gamma(\theta)\mathbf{P}^t + \mathbf{Pzz}^t\mathbf{P}^t\right)\mathbf{P}\frac{d\Gamma(\theta)}{d\theta_m}\mathbf{P}^t\right)\right]_{m=1,\ldots,p} \\
=\ & \left[2\operatorname{tr}\left(\left(\Gamma(\theta) + \mathbf{zz^t}\right)\mathbf{P}\frac{d\Gamma(\theta)}{d\theta_m}\mathbf{P}^t\right)\right]_{m=1,\ldots,p}
\end{aligned}$$

$$\begin{aligned}
&\frac{d^2}{d\theta^2}\|\mathbf{Pzz}^t\mathbf{P}^t + \mathbf{P}\Gamma(\theta)\mathbf{P}^t\|^2 = \\
=\ & \frac{d}{d\theta}\left[2\operatorname{tr}\left(\left(\Gamma(\theta) + \mathbf{zz^t}\right)\mathbf{P}\frac{d\Gamma(\theta)}{d\theta_m}\mathbf{P}^t\right)\right]_{m=1,\ldots,p} \\
=\ & \left[2\operatorname{tr}\left(\left(\Gamma(\theta) + \mathbf{zz^t}\right)\mathbf{P}\frac{d^2\Gamma(\theta)}{d\theta_m d\theta_l}\mathbf{P}^t + \frac{d\Gamma(\theta)}{d\theta_l}\mathbf{P}\frac{d\Gamma(\theta)}{d\theta_m}\mathbf{P}^t\right)\right]_{m=1,\ldots,p,\, l=1,\ldots,p}
\end{aligned} \tag{4}$$

$-\Gamma$ can be replaced by any matrix $\mathbf{C}$ representing a generalised covariogram.

6 Variogram models with linear parameterisation

Let us first consider the case of a class of variogram models which is linear in its parameters:

$$\gamma(\mathbf{h}) = \sum_{m=1}^{p} \theta_m \gamma_m(\mathbf{h}), \quad \theta = (\theta_1, \ldots, \theta_p) \in \Theta \subset (I\!R_0^+)^d \tag{5}$$

where $\gamma_1, \ldots \gamma_n$ are conditionally negative definite functions. This is a model of possible variograms. To simplify the equations we define

$$\mathbf{\Gamma}_m = (\gamma_m(\mathbf{s}_i - \mathbf{s}_j))_{i,j=1,\ldots,n} \in I\!R^{n\times n}$$

and get:

$$\begin{aligned} & \frac{d}{d\theta}\|\mathbf{P}\mathbf{z}\mathbf{z}^t\mathbf{P}^t + \mathbf{P}\Gamma(\theta)\mathbf{P}^t\|^2 = \\ = \; & \left(2\mathrm{tr}\left(\left(\Gamma(\theta) + \mathbf{z}\mathbf{z}^{\mathbf{t}}\right)\mathbf{P}\Gamma_m\mathbf{P}^t\right)\right)_{m=1,\ldots,p} \end{aligned}$$

$$\begin{aligned} & \frac{d^2}{d\theta^2}\|\mathbf{P}\mathbf{z}\mathbf{z}^t\mathbf{P}^t - \mathbf{P}\Gamma(\theta)\mathbf{P}^t\|^2 = \\ = \; & \left(2\mathrm{tr}\left(\mathbf{P}\Gamma_l\mathbf{P}\Gamma_m\mathbf{P}^t\right)\right)_{l=1,\ldots,p,\, m=1,\ldots,p} =: \mathbf{A} \end{aligned}$$

$\mathbf{A}$ is regular if and only if:

$$\sum_{m=1}^{p} \theta_m \mathbf{P}\Gamma_m\mathbf{P}^t \neq 0 \quad \forall\theta \in I\!R^p \setminus \{0\}$$

We call this an identifiability condition, since the parameter is fully identifiable from the second order properties of one realisation of $Z(\mathbf{s})$ if and only if this condition holds. This follows from lemma 1 below. Note that this condition does not depend on the choice of $\mathbf{P}$.

If this condition holds, then the second derivative is constant and positive definite since for all $\theta \neq 0$ it holds

$$\theta^t\mathbf{A}\theta = \left(\sum_{m=1}^{p} \theta_m \mathbf{P}\Gamma_m\mathbf{P}^t\right)^2 > 0$$

and thus we can obtain the optimum from the linear equation:

$$\hat{\theta} = -\mathbf{A}^{-1} \left[2\mathrm{tr}\left((\mathbf{P}\mathbf{z}\mathbf{z}^{\mathbf{t}}\mathbf{P})\Gamma_m \right) \right]_{m=1,\dots,p}$$

Taking the expectation yields:

$$\begin{aligned} E[\hat{\theta}] &= -\mathbf{A}^{-1} \left[2\mathrm{tr}\left(-(\mathbf{P}\Gamma\mathbf{P})\Gamma_m \right) \right]_{m=1,\dots,p} \\ &= \mathbf{A}^{-1} \left[2\mathrm{tr}\left(\left(\sum_{l=1}^{p} \theta_l \Gamma_l \right) \mathbf{P}\Gamma_m \mathbf{P}^t \right) \right]_{m=1,\dots,p} \\ &= \mathbf{A}^{-1} \left[2\mathrm{tr}\left(\mathbf{P}\Gamma_l \mathbf{P}\Gamma_m \mathbf{P}^t \right) \right]_{m=1,\dots,p,\, l=1,\dots,p} \theta \\ &= \mathbf{A}^{-1}\mathbf{A}\theta = \theta \end{aligned}$$

Thus $\hat{\theta}$ is an unbiased estimator for θ.

7 Consistency of the linear variogram estimator

To determine the variance of a variogram estimator we need assumptions concerning the moments of fourth order of the process and the measurement plan. We assume a measurement plan $\mathbf{s}_1, \dots, \mathbf{s}_n$, $n = 1, \dots, \infty$.

$$\mu_{ijkl} := \mathbf{cov}(Y(\mathbf{s}_i)Y(\mathbf{s}_j), Y(\mathbf{s}_l)Y(\mathbf{s}_m))$$

For some constant C we assume:

$$|\mu_{ijkl}| \leq \tfrac{1}{2}(D_{ik}D_{jl} + D_{il}D_{jk}) \tag{6}$$

where the maximum eigenvalue of the symmetric matrices $\mathbf{D} = [D_{ij}]_{i,j}$ is bounded by a constant C

$$\frac{\|\mathbf{D}\mathbf{v}\|}{\|\mathbf{v}\|} < C \in I\!R \tag{7}$$

This is true for all random fields reaching independence within a finite range and regular measurement plans increasing proportional to n. Further we assume:

$$\left\| \sum_{l=1}^{p} \theta_l P C_l P^t \right\|^2 \geq \frac{1}{c} \|\theta\|^2 n, \quad \forall \theta \in I\!R^p \setminus \{0\} \tag{8}$$

i.e. the average amount of new information about the covariogram per new observation does not fall under some limit. It is a condition on the combination of trend, measurement plan and variogram model.

To proof the consistency of $\hat{\theta}$ we need a precise understanding of its structure. Let $a : I\!R^{n\times n} \to I\!R^{n^2}$,

$$a((b_{ij})_{ij}) = (b_{((k-1) \bmod n)+1,\lfloor k/n+1\rfloor})_{k=1,\ldots,n^2}$$

denote an embedding of the $I\!R^{n\times n}$ matrix space into the $I\!R^{n^2}$ vector space. With:

$$\mathbf{M} := [\theta_l a(PC_lP^t)_k]_{k,l} \in I\!R^{n^2\times p}$$

we can rewrite $\hat{\theta}$ as:

$$\hat{\theta} = (\mathbf{M}^t\mathbf{M})^{-1}\mathbf{M}^t a(\mathbf{z}\mathbf{z}^t) = \mathbf{M}^- a(\mathbf{z}\mathbf{z}^t)$$

where M^- denotes the Moore–Penrose–inverse of $\mathbf{M}$. From condition (8) we conclude that $\mathbf{M}$ has full row rank and that the smallest singular value of $\mathbf{M}$ different from 0 is not smaller than $\sqrt{\frac{n}{c}}$, and thus the largest singular value of $\mathbf{M}^-$ is not larger than $\sqrt{\frac{c}{n}}$.

Let

$$V := \mathbf{Var}(a(\mathbf{z}\mathbf{z}^t)) = (\mathbf{cov}(a(\mathbf{z}\mathbf{z}^t)_k, a(\mathbf{z}\mathbf{z}^t)_l))_{kl} = (\mu_{ijkl})_{ij,kl} \in I\!R^{n^2\times n^2}$$

From equation (6) we get[1] $V \leq_L \mathbf{D} \otimes \mathbf{D}$ and thus

$$\|\mathbf{Var}(\hat{\theta})\| = \|M^- V M^{-t}\| \leq \sqrt{\frac{c}{n}} C^2 \sqrt{\frac{c}{n}} = \frac{c\,C^2}{n}$$

Thus we derive weak consistency, and the usual $\sqrt{n}$-convergence rate for $\hat{\theta}$.

8 The empirical variogram corrected for trend

We could try to use these results to estimate the empirical variogram in the presence of trend. We just have to specify the variogram model as a step function of the form

$$\gamma_\theta(\mathbf{h}) := \begin{cases} 0, & 0 \\ \theta_k, & h_{k-1} < \|\mathbf{h}\| \leq h_k \end{cases}, \quad 0 = h_0 < \ldots < h_p = \infty$$

[1] "$\leq_L$" denotes Loewner matrix half ordering: $V \leq_L B \Leftrightarrow \forall \mathbf{v} : \mathbf{v}^t V \mathbf{v} \leq \mathbf{v}^t B \mathbf{v}$ and $\otimes$ denotes the tensor product

where the h_k represent limits of distance classes. When fitting this step function we could hope to get a corrected empirical variogram using the general unbiasedness and consistency theorems, but we find the serious problem of trade off between estimation variance and estimation bias due to the model misspecification. The model is misspecified because the true variogram is not a step function and this introduces a bias in the parameter estimation. Thus we need a general theorem limiting the bias in the estimation of misspecified models.

Theorem 1 *Assume $\gamma(\mathbf{h})$ to be the true variogram of $Y(\mathbf{s})$ and*

$$\gamma_\theta(\mathbf{h}) = \sum_{m=1}^{p} \theta_m \gamma_m(\mathbf{h}), \quad \theta = (\theta_1, \ldots, \theta_p) \in W$$

a misspecified model for γ such that

$$\exists\, \theta_0\ \exists\, \varepsilon : \sum_{i,j} (\gamma_{\theta_0}(\mathbf{x_i} - \mathbf{x}_j) - \gamma(\mathbf{x}_i - \mathbf{x}_j))^2 \leq n\varepsilon \tag{9}$$

then

$$(E[\hat{\theta} - \theta_0])^2 \leq \frac{\varepsilon}{\frac{1}{n} \inf\limits_{\|\theta\|=1} \| \sum_k \theta_k \mathbf{P}\Gamma_k \mathbf{P}^t \|^2} \quad (\leq c\varepsilon)$$

The interpretation of this theorem is that only misspecification of the variogram for large distances can lead to infinite ??? unbounded ??? bias in the estimation. Misspecification at small distances do not vanish but they are well bounded if $\frac{1}{n} \inf\limits_{\|\theta\|=1} \| \sum_k \theta_k \mathbf{P}\Gamma_k \mathbf{P}^t \|^2$ is large. This term is closely related to the identifiability condition and we can therefore interpret it as a bound for the estimation error for "very" identifiable parameters.
Proof: $\hat{\theta}$ can be written in the form:

$$\hat{\theta} = \mathbf{M}^{-} a(\mathbf{z}\mathbf{z}^t)$$

$$E[\hat{\theta} - \theta_0] = \mathbf{M}^{-} \left(a(\Gamma) - a(\Gamma(\theta_0)) \right)$$

from equation (9) we get:

$$\|a(\Gamma) - a(\Gamma(\theta_0))\|^2 \leq n\varepsilon$$

Since $\inf\limits_{\|\theta\|=1} \| \sum_k \theta_k \mathbf{P}\Gamma_k \mathbf{P}^t \|$ is the smallest singular value of $\mathbf{M}$ different from 0, $\left(\inf\limits_{\|\theta\|=1} \| \sum_k \theta_k \mathbf{P}\Gamma_k \mathbf{P}^t \| \right)^{-1}$ is the largest singular value of $\mathbf{M}^-$ and thus:

$$\|M^- \left(a(\mathbf{P}\Gamma\mathbf{P}^t) - a(\mathbf{P}\Gamma(\theta_0)\mathbf{P}^t) \right) \|^2 \leq \frac{n\varepsilon}{\inf\limits_{\|\theta\|=1} \| \sum_k \theta_k \mathbf{P}\Gamma_k \mathbf{P}^t \|^2}$$

q.e.d.

Note that even the classical empirical variogram is biased due to misspecification of the model, but we can prove that the bias is bounded by 1ε representing the fact that the expectation of the variogram estimator at a lag $\mathbf{h}$ we can take as a value of any of the lags in the corresponding bin. The bias with the trend corrected estimation has a more complicated structure, but it is larger only by the factor:

$$\frac{n}{\inf\limits_{\|\theta\|=1} \| \sum_k \theta_k \mathbf{P}\Gamma_k \mathbf{P}^t \|^2}$$

9 General parameterisation

Most variogram models used in practice are not of the from (5), e.g. the exponential variogram model:

$$\gamma(\mathbf{h}) = \theta_1 \left(1 - e^{-\theta_2 \|\mathbf{h}\|} \right), \theta_1, \theta_2 \geq 0$$

Let us now consider the problem of fitting a variogram model not linear in the parameter θ, using the estimator:

$$\hat{\theta} = \operatorname*{argmin}_{\theta} \|\mathbf{P}\mathbf{z}\mathbf{z}^t\mathbf{P}^t + \mathbf{P}\Gamma(\theta)\mathbf{P}^t\|^2 \tag{10}$$

To guarantee the existence of the (not necessarily unique) global optimum we restrict the parameters to a closed set. Since the Hessian–matrix of the objective function:

$$v(\theta) := \|\mathbf{P}\mathbf{z}\mathbf{z}^t\mathbf{P}^t + \mathbf{P}\Gamma(\theta)\mathbf{P}^t\|^2$$

given in equation (4) is not always positive definite it is not trivial to find the global optimum, but we can use algorithms of global optimisation. Good starting values can be found by inspection of the corrected empirical variogram given in section 8.

10 Second–order identifiability

The difference of two variograms $\gamma(\mathbf{h})$ and $\tilde{\gamma}(\mathbf{h})$ is visible in the second order properties of observations at $\mathbf{s}_1, \ldots, \mathbf{s}_n$ of one realisation of a random field $Z(\mathbf{s})$ with random trend, if and only if the identifiability condition:

$$\|\Delta(\theta, \theta_0)\|^2 \neq 0$$

holds. More precisely this is formulated in the following lemma.

Lemma 1 (Second–order identifiability) *For two random fields $Y(\mathbf{s})$ and $\tilde{Y}(\mathbf{s})$ as in chapter 2 with variograms $\gamma(\mathbf{h})$ and $\tilde{\gamma}(\mathbf{h})$ there exist two fields $Z(\mathbf{s}) = \beta^t \mathbf{f}(\mathbf{s}) + Y(\mathbf{s})$ and $\tilde{Z}(\mathbf{s}) = \tilde{\beta}^t \mathbf{f}(\mathbf{s}) + \tilde{Y}(\mathbf{s})$ with $\mathbf{f}(\mathbf{s})$ as in chapter 2 such that $\mathbf{C} = \tilde{\mathbf{C}}$ if and only if*

$$\mathbf{P}(\Gamma - \tilde{\Gamma})\mathbf{P}^t = 0$$

Proof:
1) show $\mathbf{C} = \tilde{\mathbf{C}} \Rightarrow \mathbf{P}(\Gamma - \tilde{\Gamma})\mathbf{P}^t = 0$

$$\mathbf{P}(\Gamma - \tilde{\Gamma})\mathbf{P}^t = \mathbf{P}\Gamma\mathbf{P}^t - \mathbf{P}\tilde{\Gamma}\mathbf{P}^t = -\mathbf{P}\mathbf{C}\mathbf{P}^t + \mathbf{P}\tilde{\mathbf{C}}\mathbf{P}^t = 0$$

2) show $\mathbf{P}(\Gamma - \tilde{\Gamma})\mathbf{P}^t = 0 \;\Rightarrow\; \exists Z, \tilde{Z}$ such that $\ldots$
Let $\mathbf{F}^-$ denote the Moore–Penrose–Inverse of $\mathbf{F}$ such that: $\mathbf{F}\mathbf{F}^-\mathbf{F} = \mathbf{F}$. Set

$$\beta := -\mathbf{F}^-(1 - \mathbf{P})\mathbf{y}, \quad \tilde{\beta} := -\mathbf{F}^-(1 - \mathbf{P})\tilde{\mathbf{y}},$$

$$\mathbf{z} = -\mathbf{F}\mathbf{F}^-(1 - \mathbf{P})\mathbf{y} + \mathbf{y} = -(1 - \mathbf{P})\mathbf{y} + \mathbf{y} = \mathbf{P}\mathbf{y}$$

since $\mathrm{im}(1 - \mathbf{P}) = \mathrm{im}\,\mathbf{F}$. Analogously we get $\tilde{\mathbf{z}} = \mathbf{P}\tilde{\mathbf{y}}$ and thus:

$$E[Z(\mathbf{s}_i)] = 0 = E[\tilde{Z}(\mathbf{s}_i)]$$

and

$$\mathbf{Var}(\mathbf{z}) = \mathbf{Var}(\mathbf{P}\mathbf{y}) = -\mathbf{P}\Gamma\mathbf{P} = -\mathbf{P}\tilde{\Gamma}\mathbf{P} = \mathbf{Var}(\mathbf{P}\tilde{\mathbf{y}}) = \mathbf{Var}(\tilde{\mathbf{z}})$$

11 Relation to intrinsic random functions

An alternate approach for the estimation of covariograms in the presence of trend are intrinsic random functions of order k (IRFk) defined e.g. in [Cressie 1993][Chilès&Delfiner 1999]. The trend used in IRFk exploits a stationary trend model, i.e. it holds:

$$\forall\, \alpha\; \exists\, \beta(\mathbf{h})\; \forall\, \mathbf{s} : \alpha^t \mathbf{f}(\mathbf{s}) = \beta(\mathbf{h})^t \mathbf{f}(\mathbf{s}+\mathbf{h})$$

Thus for $\gamma_{mk}(\mathbf{s}_i, \mathbf{s}_j) = f_m(\mathbf{s}_i) f_k(\mathbf{s}_j)$ and any configuration of measurement locations it holds:

$$\operatorname{im} \Gamma_{mk} \subset \operatorname{im} \mathbf{F}$$

where

$$\Gamma_m = (\gamma_m(\mathbf{s}_i, \mathbf{s}_j))_{ij} = (f_m(\mathbf{s}_i) f_k(s_j))_{ij}$$

and thus

$$\mathbf{P}\Gamma_m \mathbf{P}^t = 0$$

Note that for symmetric Γ_m this is equivalent to:

$$\mathbf{P}\Gamma_m = 0$$

Thus as a consequence of the theory given here, we can estimate $\gamma(\mathbf{h})$ and $c(\mathbf{h})$ only up the equivalence class given by the symmetric functions (i.e. $\gamma(\mathbf{s}_i, \mathbf{s}_j) = \gamma(\mathbf{s}_j, \mathbf{s}_i)$ in $\langle \gamma_{mk}(\mathbf{s}_i, \mathbf{s}_j) : m, k = 1, \ldots, p \rangle$. The same result was given for the generalised covariances in $IRFk$–theory, where these symmetric cross products are the even polynomials up to degree $2n$. Note that we have an implicit extension of the IRFk theory to all stationary trend models (e.g. harmonic trend functions) from the pure polynomial trend function in the classical IRFk theory. This paper can only deal with surfaces, which have a stationary covariogram. Thus for polynomial trend functions the class of processes described by IRFk is larger than the class of processes discussed in this paper.

On the other hand we get an interesting result for continuous trend models which are not translation invariant. Define the translation invariant part $I_{\mathbf{f}}$ of the model as:

$$I_{\mathbf{f}} := \bigcap_{\mathbf{h} \in I\!R^d} \{\alpha^t \mathbf{f}(\cdot + \mathbf{h}) : \alpha \in I\!R^p\}$$

Then there exist a measuring plan, such that $c(\mathbf{h})$ and $\gamma(\mathbf{h})$ are identifiable only up to differences in

$$J_{\mathbf{f}} := \langle f_1(\mathbf{x})f_2(\mathbf{y}) + f_2(\mathbf{x})f_1(\mathbf{y}) : f_1, f_2 \in I_{\mathbf{f}} \rangle$$

More precisely: When two covariance functions $c_1(\mathbf{h})$, $c(\mathbf{h}9$ have a difference in $J_{\mathbf{f}}$:

$$(c_1 - c_2)(\mathbf{h}) \in J_f$$

then the corresponding $\mathbf{Pz}$ have the same second order properties and we can not distinguish between these two processes covariance functions from our observation. However this is not necessary, since both lead to the same kriging weights and the same kriging errors. Whenever

$$(c_1 - c_2)(\mathbf{h}) \notin J_f$$

we can make a measurement plan such that we can distinguish between these two and estimate the variogram.

Thus for the purpose of kriging it is sufficient to find a covariogram which only differs in by an element of J_f from the true covariogram c.

We now proof that we always choose such a plan of measurement locations: Let $\gamma_o(\mathbf{h}) = (\gamma_1 - \gamma_2)(\mathbf{h})$ denote such a modelling direction not in $J_{\mathbf{f}}$. Since $\gamma_o(\mathbf{h})$ is not in $J_{\mathbf{f}}$ there exists an $\mathbf{x}$ such that $\gamma_m(\cdot + \mathbf{h}) \notin \{\alpha^t \mathbf{f}(\cdot) : \alpha \in I\!R^p\}$ and thus we get $\mathbf{s}_1 = \mathbf{h}, \ldots, \mathbf{s}_n$ such that the first column of Γ_o is not in the image of $\mathbf{F}$ and thus $\mathbf{P}\Gamma_o \neq 0$. This means that changes of the covariance by adding any multiple of Γ_m are identifiable.

When we restrict ourselves to functions with the property $\gamma(\mathbf{x}, \mathbf{y}) = \gamma_S(\mathbf{x} - \mathbf{y})$ we can rewrite $J_{\mathbf{f}}$ as

$$J_{\mathbf{f}S} = \left\langle f_1(\mathbf{x})f_2(\mathbf{x} + \mathbf{h}) + f_2(\mathbf{x})f_1(\mathbf{x} + \mathbf{h}) : f_1, f_2 \in I_{\mathbf{f}}, \mathbf{x} \in I\!R^d \right\rangle$$

12 Local trend surfaces

Sometimes we do not believe to know a model for the global trend, but we want to use a local trend model, which is only valid in a prespecified searching distance of the kriging estimator[Goovaerts 1997]. Since the covariogram estimator proposed in this paper relies on the global validity of the trend we have to modify it for considerations of local trend. The general idea of the estimator is that we filter the trend using a known projection matrix $\mathbf{P}$ and compare every individual resulting covariance with the covariance calculated for such a transformed quantity from the covariogram model. This idea can be generalised. We only need to decide for every pair up to which distance $h_{\max}$ points should be taken into account for trend filtering. This results in a projection matrix that we can use to project a subset of the measurements only and to calculate the covariance of the resulting vectors. The corresponding pair can still be identified since we have only removed the trend estimated from its local neighbours. Then we can still compare this now locally filtered covariance with the appropriate theoretical one calculated with the following modified transformations:

$$\begin{aligned}
N^{ij}_{kl} &:= \begin{cases} \delta_{kl} & , \|\mathbf{s_i} - \mathbf{s_j}\| < h_{\max} \wedge \|\mathbf{s_i} - \mathbf{s_k}\| < h_{\max} \wedge \|\mathbf{s_j} - \mathbf{s_k}\| < h_{\max} \\ 0 & , \text{otherwise} \end{cases} \\
\mathbf{N}^{ij} &:= (N^{ij}_{kl})_{k=1,\ldots,n,\; l=1,\ldots,n} \\
\mathbf{P}^{ij} &:= 1 - \mathbf{N}^{ij}\mathbf{F}(\mathbf{F}^t{\mathbf{N}^{ij}}^t\mathbf{N}^{ij}\mathbf{F})^{-1}\mathbf{F}^t{\mathbf{N}^{ij}}^t \\
\hat{\theta} &:= \underset{\theta}{\operatorname{argmin}} \sum_{ij} \left(\mathbf{e}_i^t\mathbf{P}^{ij}(\mathbf{z}\mathbf{z}^t + \Gamma(\theta)){\mathbf{P}^{ij}}^t\mathbf{e}_j\right)^2
\end{aligned}$$

13 Conclusion

Even in situations of linear trend models it is possible to calculate empirical variograms and to fit variogram models consistently and unbiasedly. In the situation of universal kriging the variogram is fully identifiable up to a linear space depending on the translation invariant part of the trend. Two functions that cannot be distinguished by the variogram estimation procedure lead to the same kriging weights and errors. It it is therefore irrelevant which candidate we choose. This is very similar to what we have with intrinsic random functions, but it applies to any internal or external trend model and not only to very specific internal trend models as with intrinsic random functions. In practical applications we can, but need not, consider the generalised covariograms, because we can also fit ordinary variogram models to trended surfaces.

Thus universal kriging based on a universal method of variogram estimation is more flexible and more intuitive than kriging of intrinsic random functions.

References

[Boogaart 1999] BOOGAART, K.G. (1999): New possibility for modelling variograms in complex geology, to appear in *Proc. of StatGIS 1999*

[Cressie 1993] CRESSIE, N.A.C. (1993): *Statistics for Spatial Data (rev. ed.)*: J. Wiley & Sons

[Chilès&Delfiner 1999] CHILÈS, J.-P., DELFINER, P. (1999): *Geostatistics: Modelling Spatial Uncertainty*: Wiley

[Goovaerts 1997] GOOVAERTS, P (1997): *Geostatistics for Natural Resource Evaluation*, Oxford University Press, New York

[Wackernagel 1998] WACKERNAGEL, HANS (1998): *Multivariate Geostatistics, An Introduction With Applications*, 2nd, completely revised edition, Springer Verlag, Berlin

[Witting 1995] WITTING, H., U. MÜLLER-FUNK (1995): *Mathematische Statistik. 2. Asymptotische Statistik: parametrische Modelle und nichtparametrische Funktionale*, B.G. Teubner, Stuttgart