Nonlinear Kriging, potentialities and drawbacks

K. G. van den Boogaart
TU Bergakademie Freiberg, Germany; *boogaart@grad.tu-freiberg.de*

Motivation

Kriging is known to be the best linear prediction to conclude from observations of a random field at individual points to the values of the random field at other locations. When the random field is jointly Gaussian, simple kriging is also the best prediction (in the sense of mean squared error), since for joint normal distributions the linear prediction is equivalent to the conditional expectation. However in our days kriging goes far beyond the prediction of gaussian random fields with known mean and known covariance function from point measurements without any further information. Many specialized theories have been developed to overcome the limitations of simple kriging.

- *Unknown or instationary mean*
  Ordinary kriging and IRFk theory replace optimality of the kriging predictior by conditional optimality to overcome unknown parameters in the distribution model. Alternatively one could apply Baysian Kriging.
- *Unknown covariance function*
  Kriging is mostly based on estimated covariance functions or variograms and thus looses its optimality property towards near optimal solutions. The problem how precise the variogram must be estimated to make kriging superior to deterministic interpolators seems to stay without rigerous solution.
  The unknown variogram is again a problem of unknown parameters in the distribution model leading to loose of optimality. Baysian kriging based on normal distributions provides a solution here. An alternative way would be the definition of a conditional optimality, as for unknown trend parameters.
  A general method of optimal prediction in case of unknown variogram would be desirable.
- *Additional knowledge: e.g. known linear differential equations*
  Some processes in nature are governed by partial differential equations. Linear differential equations for the process correspond to conditions on the covariance function and the trend model (Boogaart, 2002) and are can be used e.g. in hydrology (Chiles and Delfiner 1999), deformation and gravity analysis. However most interesting differential equations are nonlinear and can only be handled approximately by linearization.
- *Non gaussian distributions are predicted best with nonlinear methods*
  For non gaussian random fields the linear prediction is in general not optimal. Special generalisations of kriging to special nonlinear predictors, all based on transformations of the individual values have been proposed: lognormal kriging, indicator kriging, disjunctive kriging.
- *Kriging of functionals*
  Often the quantity of interest is not the random function itself, but a functional such as blockmeans, gradients and deformation, exceeding of thresholds, total costs, connectivity or direction of water flow. For some of these specialized methods based on kriging of linear functionals, on linearization, on probability kriging or, on simulation methods.
- *Kriging based on functionals*
  Sometimes only functionals of a random field can be observed: The plain is below the deeps of the drilling, the Cd-content is below detection limit, the slope, a potential difference in gravity and many more. A good interpolation method should make use of these indirect observations. kriging of linear functionals and its special cases such as gradient kriging allow the use of observed linear functionals. Special publications can be found on special nonlinear functionals such as indicators.

Thus we started out from simple kriging and extended the theory to more and more sophisticated applications. Let us now ask the question the other way round: How can we calculate the optimal predictor for any general (i.e. linear or nonlinear) functional F_0 of a random field $Z(\mathbf{x})$ based on any set of observed functionals $F_1,..., F_n$, when the distribution of the random field is given up to some parameters and what are the implication of a known general partial differential equations of the random field? This theory should be outlined here. The given mathematical foundation is by no means complete or general, since the used arguments do not exist in many cases due to the missing compactness of the corresponding operators. The precise domain of validity is to be established later. Further this should not be seen as a theory to be used for itself in full generality but as system to generate the specialisation most appropriate for the current application. Indeed most - if not all - of the known generalizations of kriging are indeed special cases of this new general framework. The presentation of the theory here may only be seen as a brief sketch without references or proofs.

A new foundation: The moment generating functional

The variogram describes the second order structure of Z and is therefore only useful to compute the variance of linear combinations. In this nonlinear theory we will replace the variogram (or covariance function) with a mathematical object containing all the information about all moments at once:
The moment generating functional $Z^{\#}$ of Z. The moment generating functional $Z^{\#}$ is a distribution dependent nonlinear mapping of linear functionals L, replacing the increments of IRFk, to a real number and is defined as (cf. Rinne 1997):

$$Z^{\#}(L) := E[\ \exp(LZ)\ ]$$

It does not exist for all L just as the covariance function is meaningful only for special increments. Thus analog to generalized covariance functions we will define generalized moment generating functionals as equivalence classes of functionals only giving finite values for specific “increments” L. Strong stationarity of Z corresponds to translation invariance of $Z^{\#}$:

$$Z^{\#}(L_h)=Z^{\#}(L)$$

Where $L_hZ := Y(L(\cdot-h))$ is a translated functional. The name "moment generating functional" originates from the fact that the uncentered moments can be calculated from it taking derivatives of the functional and evaluating them in 0.

$$E\left[(LZ)^k\right]==\frac{d^k}{d\varepsilon}Z^{\#}(\varepsilon L)|_{\varepsilon=0}$$

Nonlinear increments and its transforms

With more effort the moment generating functional can be used to calculate the moments of nonlinear functionals g of Z. We define the moment transform $g^{\#}$ of functionals g based on a Taylor series expansion of the functional g into multilinear operators A_i:

$$gZ=\sum_{i=0}^{\infty}A_i(Z_1,\ldots,Z_i)$$

with functionals A_i multilinear in i components. Unfortunately a Taylor series expansion only exists for analytical operators. Analytical operators are transformed to

$$g^{\#}:=\sum_{i=0}^{\infty}A_i\left(\frac{\partial}{\partial L},\ldots,\frac{\partial}{\partial L}\right)$$

which is a linear differential operator which maps the moment generating functional of Z to the moment generating functionals of the nonlinear statistic gZ represented by g:

$$(gZ)^{\#} = g^{\#}Z^{\#},$$

Thus the moment generating functional and thus the moments of gZ can be calculated from $Z^{\#}$. An analogues transformation $G^{\#}$ can be defined for linear operators G. The only difference is that here the A_i are multilinear operators in i-variables rather then only functionals. Correspondingly we get:

$$(GZ)^{\#} = G^{\#}Z^{\#}$$

The application of $g^{\#}$ on the moment generating functional corresponds to the evaluation of the trend function

$$E[\ gZ\ ] = g^{\#}Z^{\#}\ |_{L=0}$$

Comparing the Taylor series we find the simple relation:

$$(g_1(Z)g_2(Z))^{\#} = g_1^{\#}g_2^{\#}Z^{\#}$$

$$E[\ g_1(Z)g_2(Z)\ ] = g_1^{\#}g_2^{\#}Z^{\#}|_{L=0} = g_2^{\#}\ g_1^{\#}Z^{\#}|_{L=0}$$

which corresponds to evaluation of the product moment $E[\ Z(g_1),Z(g_2)\ ] = c(g_1,g_2)$ in case of simple kriging (μ=0) for two locations (rather than functionals). $Z^{\#}$ incorporates all moment informations at once.

Nonlinear differential equations

Let us assume that every realization of Z solves a (linear or nonlinear) differential equation

$$GZ \equiv 0$$

with some analytical differential operator G. The moment generating function $(GZ)^{\#} = G^{\#}Z^{\#}$ of the constant 0 is equivalent to 0 and thus the differential equation implies the constraint

$$G^{\#}Z^{\#} \equiv 0$$

to the moment generating functional. This generalizes the restrictional effect of linear differential equations on the variogram and the trend (Boogaart, 2002):

$$LZ = k(x) \Leftrightarrow Lc(x,y) \equiv 0, Lf_i(x) = 0, Lf_0(x) = k(x)$$

An obviously the other way round, when $(GZ)^{\#} = G^{\#}Z^{\#} \equiv 0$ holds we also have $GZ \equiv 0$ in distribution. Thus the implications of nonlinear differential equations can be fully reproduced by this presentation just as linear differential equations are by linear kriging. However the prediction will in general not solve the nonlinear differential equation.

Ansatz for nonlinear Kriging

The general problem of nonlinear kriging can be formulated as follows: The random field Z is observed by an observation operator $G = (g_1,\ldots,g_n)^t$ consisting of a set of (nonlinear) observation functionals $g_1,\ldots,g_n$ resulting in observations $GZ = (g_1Z,..,g_nZ)^t$. The predictor (kriging weights) corresponds to an operator $\hat{g}$, which needs to by applied to GZ. The best predictor $\hat{g}$ is defined by the usual expected mean squared error loose criterion:

$$E\left[(\hat{g}GZ - gZ)^2\right] \rightarrow \min$$

which can be found by the variation formulation

$$\frac{d}{dL} E\left[(\hat{g}GZ(x) + LGZ(x) - gZ(x))^2\right] = \Big|_{L=0} 0$$

which simplifies to

$$E\left[\hat{g}GZ(x)GZ\right] = E[gZGZ]$$

This on the other hand shows up as a linear problem in the transformed formulation:

$$\hat{g}^{\#}G^{\#}G^{\#}Z^{\#} = g^{\#}G^{\#}Z^{\#}$$

Where now $\hat{g}^{\#}$ and thus by backtransformation $\hat{g}$ can be found by solving the linear operator equation within the image space of the $^{\#}$ transform. This equation is analog to the simple kriging equation

$$C\lambda = c$$

with $\hat{g}^{\#}$ corresponding to λ, $G^{\#}G^{\#}Z^{\#}$ to $C=(c(x_i,x_j))_{ij}$ and $g^{\#}G^{\#}Z$ to c. A central problem is to understand the meaning of $G^{\#}G^{\#}Z^{\#}$. Thus the optimal predictor can more or less be found by solving a linear equation. However the equation is an operator equation and we need methods to find a numerical solution.

Prediction Error

Like with linear kriging we can calculate the prediction error for our optimal predictor from the moment generating functional:

$$E[(\hat{g}GZ - gZ)^2] = ((\hat{g}G)^{\#} - g^{\#})^2 Z^{\#}\Big|_0$$

Weakly optimal predictors

Linear differential equations have week solutions for test spaces and reduced candidate spaces. It is interesting to study these suboptimal solutions for some reasons:

- It will be necessary to calculate the optimal predictor numerically, since we are most often not able to solve the nonlinear kriging equation analytically.

- Linear kriging is special weak solution of the nonlinear kriging equation. Test functions corresponding to linear functionals and linear candidate spaces correspond to the linear kriging predictor.
- We can restrict the problem to "simple to calculate" subclasses of predictors (Such as polynomials, indicators, disjunctive transforms, ...), resulting in different types of nonlinear kriging.
- And last but not least: When we use the moment generating function of an estimated or inferred structure, we may want to disregard special properties of the structure which are unsure. Thus we want to reduce the usage of the structure to some of its functionals.

Prediction with structure determination

Introduction

The central problem of kriging is that the structure represented by $Z^{\#}$ is a-priori unknown und thus we can never give good predictors unless we have enough data to infer the true structure very well. However in realistic situations parts of the structure are always unknown and need to be inferred from the observations. Structural parameters such as the sill and the range correspond to unknown functionals of $FZ^{\#}$. There are two basic ideas to work around this problem: Baysian and classical.

Baysian approach

The more simple situation is having a Baysian prior to the structural parameters. Baysian formulation of interpolation with unknown structural parameters implies a prior on the structural parameters. We can use expectation of the momenten generating function under the prior as moment generating function of the double experiment randomly choosing the structural parameters first and then choosing a random structure. Thus the Baysian Problem can be handle like the standard problem with known structure and needs no further provisions and Baysian kriging is just a special case of nonlinear kriging. The $Z^{\#}$ used is defined by the $Z^{\#}$ expected under the prior distribution of the unknown model parameters.

Classical approach: Disregarding unknown structures

The other idea is to assume the structural parameters $FZ^{\#}$ totally unknown. In this case we don't know some (nonlinear) functionals of the measure. However, when we know some (minimal) sufficient statistics $s_1,\ldots,s_u$ for that parameters (which we normally do not observe), we just need to assume that we don't know anything about the distribution of these, because normally the unknown parameters allow any probability distribution for these and when we don't know anything about the distribution of these, we impose no knowledge about the parameters. For sure we can then not find a globally optimal solution without reducing our selves to test functions and candidate functions with moment transforms orthogonal to $s_1^{\#},\ldots,s_u^{\#}$. The unbiasedness condition in ordinary kriging is a special case of this technique, since the candidate and test spaces are reduced to directions orthogonal to the mean. However a similar technique can be applied to unknown variogram parameters in case of nonlinear kriging.

Conclusions

Nonlinear kriging yields a framework for many problems of kriging

- Optimal prediction in case of unknown parameters means prediction disregarding special functionals of Z^{*}. A special case is the IRFk theory. But the theory can also be applied to unknown variogram parameters.
- Nonlinear kriging shows how to use and predict any functional of the random field and constructively shows which moments of the field need to be known, when estimated with given candidate and test spaces.
- Nonlinear kriging provides the tools to introduce any physical law described by differential equations into a prediction problem.
- Using the linear finite element basis as candidate and test spaces and boundary conditions as data, all necessary informations on $Z^{\#}$ are provided by a linear partial differential equation

resulting in the finite element method. Thus nonlinear kriging is a joint generalization of finite elements and kriging and provides the means to combine both methods in future applications.

- Although always providing the optimal predictor, nonlinear kriging is to demanding for real applications, but special choices of candidate and test spaces provide many interesting special cases including most of the known generalizations of kriging. It can help to find the best generalization to be used in a specific application and serve as a framework to compare different types of kriging.

References

Cressie, N.C., 1993: Spatial Statistics, second revised edition, New York a.o.L.
Chiles, J.-P., P. Delfiner, 1999: Geostatistics, Modeling Spatial Uncertainty, Wiley, New York
Rinne, H., 1997, Taschenbuch der Statistik, 2.,überarb. und erw., Harry Deutsch, Thun