Mini-Course on *Advanced Stationary Processes Analysis*, VIASM. Part 2: Geostatistics *Chapter 2: Estimation of the Variogram and Kriging*

Jean-Yves Dauxois

INSA - Toulouse Mathematical Institute, Toulouse, FRANCE
jean-yves.dauxois@insa-toulouse.fr

18th to 22th July 2016

Suppose that we have observed the realization of an intrinsic process X at sites: $s_1, \ldots, s_n$ in $S \subset \mathbb{R}^d$.

Problem. The variogram is unknown and we want to estimate it.

The **dissimilarity** between observed values of X at sites s_i and s_j is defined by:

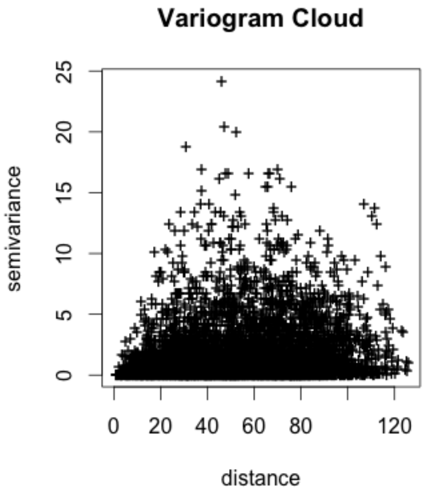
$$\gamma^*_{ij} = \frac{\left(X_{s_i} - X_{s_j}\right)^2}{2} = \gamma^*_{ji}$$

The **Variogram Cloud** is a set of $n(n-1)/2$ points:

$$\mathcal{N}_0 = \left\{ \left(||s_i - s_j||, \gamma^*_{ij} \right); 1 \leq i \neq j \leq n \right\}$$

Remark. We have:

$$\mathbb{E}\gamma^*_{ij} = \gamma(s_i - s_j).$$


Variogram Cloud
semivariance
0
5
10
15
20
25
0
20
40
60
80
120
distance

A nonparametric estimation of the variogram, based on the observations $X_{s_1}, \ldots, X_{s_n}$, is roughly given by the function

$$h \in S \longmapsto \hat{\gamma}(h) = \frac{1}{2N(h)} \sum_{s_i - s_j \simeq h} \left(X_{s_i} - X_{s_j}\right)^2,$$

where

$$N(h) = \text{Card}\{(s_i, s_j); s_i - s_j \simeq h\}.$$

Of course, we need to define more precisely what means: $s_i - s_j \simeq h$.

In the isotropic case, we can consider:

Definition

For given distance $||h|| = r$ and tolerance Δ, the isotropic empirical variogram, based on the observations $X_{s_1}, \ldots, X_{s_n}$, is given by the function

$$h \in S \longmapsto \hat{\gamma}(h) = \gamma(r) = \frac{1}{2N(r)} \sum_{(s_i,s_j);r-\Delta \leq ||s_i - s_j|| \leq r+\Delta} \left(X_{s_i} - X_{s_j}\right)^2,$$

where

$$N(r) = Card\{(s_i, s_j); r - \Delta \leq ||s_i - s_j|| \leq r + \Delta\}.$$

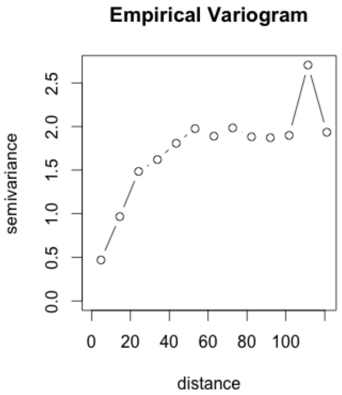


Figure: Isotropic empirical Variogram.

In the anisotropic case, when $S = \mathbb{R}^2$, let us denote r and α the direction of h, i.e. $h = r(\cos(\alpha), \sin(\alpha))$.

Definition

For given distance r, direction α and tolerances Δ and δ, the anisotropic empirical variogram, based on the observations $X_{s_1}, \ldots, X_{s_n}$, is given by the function

$$h \in S \longmapsto \hat{\gamma}(h) \;=\; \gamma(r,\alpha) = \frac{1}{2N(\Delta,\delta)} \sum_{(s_i,s_j);\nu_{\Delta,\delta}(r,\alpha)} \left(X_{s_i} - X_{s_j}\right)^2,$$

where

$$\nu_{\Delta,\delta}(r,\alpha) = \{v = u(\cos(\beta),\sin(\beta)) \in \mathbb{R}^2; |u-r| \leq \Delta, |\beta-\alpha| \leq \delta\}$$

and

$$N(r) = Card\{(s_i,s_j) \in \nu_{\Delta,\delta}(r,\alpha)\}.$$

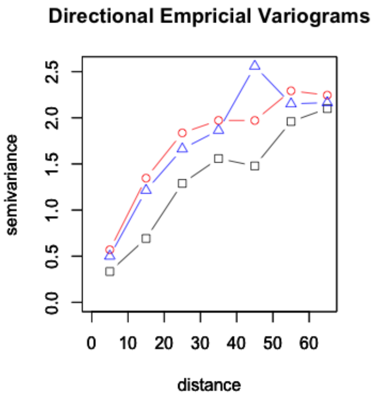


Figure: Directional Variograms: $\alpha \in \{0, \frac{\pi}{4}, \frac{\pi}{2}\}$.

One can show that

$$2\hat{\gamma}(h) = X' A(h) X$$

where

$$X = (X_{s_1}, \ldots, X_{s_n})'$$

and $A(h)$ is the matrix defined by:

$$\begin{aligned} A_{s_i,s_j} &= -\frac{1}{N(h)}, \text{ if } s_i \neq s_j \\ A_{s_i,s_i} &= \frac{N(h)-1}{N(h)}, \text{ otherwise.} \end{aligned}$$

Proposition

If $X \sim N_n(O, \Sigma)$, *then*

$$2\hat{\gamma}(h) \sim \sum_{i=1}^{N(h)} \lambda_i(h)\chi^2_{i1},$$

where the $\lambda_i(h)$ *are the nonzero eigenvectors of the matrix* $A(h)\Sigma$ *and the* χ^2_{i1} *are independent r.v. with a* χ^2_1 *distribution.*
Moreover, we have:

$$\begin{aligned} \mathbb{E}(2\hat{\gamma}(h)) &= Tr(A(h)\Sigma) \\ Var(\hat{\gamma}(h)) &= 2Tr(A(h)\Sigma)^2. \end{aligned}$$

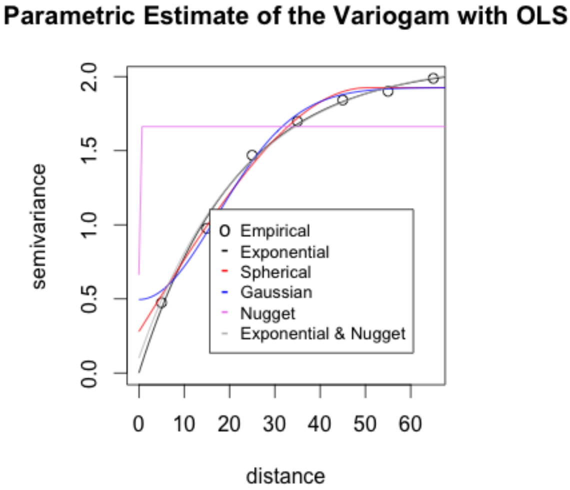
Let us assume a parametric model $\{\gamma(\cdot,\theta); \theta \in \Theta \subset \mathbb{R}^p\}$ for the variogram.

Definition

The Ordinary Least Squares (OLS) estimator of θ is:

$$\hat{\theta}_{OLS} = \arg_{\theta\in\Theta} \min \sum_{k=1}^{K} \left(\hat{\gamma}(h_k) - \gamma(h_k,\theta)\right)^2,$$

where K is the number of classes used for the (isotropic) empirical variogram at the distances $h_1, \dots, h_K$.


Parametric Estimate of the Variogam with OLS
semivariance
0.0
0.5
1.0
1.5
2.0
0
10
20
30
40
50
60
distance
o Empirical
Exponential
Spherical
Gaussian
Nugget
Exponential & Nugget

Definition

The Generalized Least Squares (GLS) estimator of θ is:

$$\hat{\theta}_{GLS} = \arg_{\theta \in \Theta} \min \left(\hat{\gamma}_K - \gamma_K(\theta)\right)' \Sigma^{-1}_{\hat{\gamma}_K}(\theta) \left(\hat{\gamma}_K - \gamma_K(\theta)\right),$$

where

$$\begin{aligned} \hat{\gamma}_K &= (\hat{\gamma}(h_1), \ldots, \hat{\gamma}(h_K))', \\ \gamma_K(\theta) &= (\gamma(h_1, \theta), \ldots, \gamma(h_K, \theta))' \end{aligned}$$

and $\Sigma_{\hat{\gamma}_K}(\theta)$ is the covariance matrix of $\hat{\gamma}_K(\theta)$.

Definition

The Weighted Least Squares (GLS) estimator of θ is:

$$\hat{\theta}_{WLS} = \arg_{\theta\in\Theta} \min \sum_{k=1}^{K} \frac{N(h_k)}{\gamma^2(h_k,\theta)} \left(\hat{\gamma}(h_k) - \gamma(h_k,\theta)\right)^2 .$$

Assuming a specific distribution for the spatial process X, as for example a gaussian distribution, one can use the Maximum Likelihood technic to estimate the parameter θ of the variogram.

In order to select Variogram Models, one can use:

- **Cross Validation**. Put aside each observation (e.g. s_i), estimate it with Kriging ($\hat{X}_{s_i}(-s_i)$) and select the model according to one of the following criterions:

$$\begin{aligned}
\text{Bias} &= \frac{1}{n}\sum_{i=1}^{n}\left(X_{s_i} - \hat{X}_{s_i}(-s_i)\right) \\
\text{MSE} &= \frac{1}{n}\sum_{i=1}^{n}\left(X_{s_i} - \hat{X}_{s_i}(-s_i)\right)^2 \\
\text{MSNE} &= \frac{1}{n}\sum_{i=1}^{n}\frac{\left(X_{s_i} - \hat{X}_{s_i}(-s_i)\right)^2}{\hat{\sigma}^2_{s_i}},
\end{aligned}$$

  where $\sigma^2_{s_i}$ is the kriging variance.

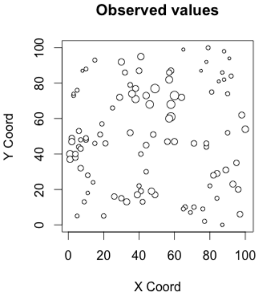
- **Parametric Bootstrap.**

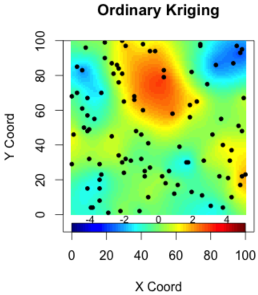
Suppose that we have observed the realization of an intrinsic process X at sites: $s_1, \ldots, s_n$ in $S \subset \mathbb{R}^d$, i.e. we have observed $X_{s_1}, \ldots, X_{s_n}$.

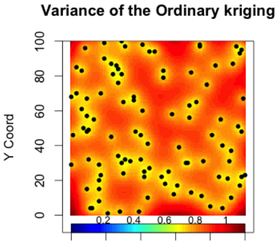
Problem. We want to estimate the value of X at a site s_0 where we don't have the observation of its value. One says that we will do the **prediction** of X_{s_0}.

Hypothesis. We assume in the following that the stationary covariance function C (in case of a second order stationary process) or the variogram γ (in case of an Intrinsic process) is known.

If this is not the case, we will first estimate (empirically or parametrically) the Covariance function or the variogram.


Observed values
Y Coord
0
20
40
60
80
100
X Coord
0
20
40
60
80
100


Ordinary Kriging
Y Coord
0
20
40
60
80
100
X Coord
0
20
40
60
80
100
-4
-2
0
2
4


Variance of the Ordinary kriging
Y Coord
0
20
40
60
80
100
0.2
0.4
0.6
0.8
1

Kriging

The idea of the Kriging is to predict X_{s_0} with an estimator $\hat{X}_{s_0}$ of the form

$$\hat{X}_{s_0} = \sum_{i=1}^{n} \lambda_i(s_0) X_{s_i} + \lambda_0(s_0),$$

where the parameters $\lambda_0(s_0), \ldots, \lambda_n(s_0)$ are reals such that the Mean Square Error (MSE)

$$\text{MSE}(s_0) = \mathbb{E}(\hat{X}_{s_0} - X_{s_0})^2$$

is minimum.

Notation. Let us write $\lambda_i = \lambda_i(s_0)$, but do not forget that these parameters usually depend on the point s_0 of prediction!

Notations

Let us write:

- $X_0 = X_{s_0}$,
- $X = (X_{s_1}, \ldots, X_{s_n})' \in \mathbb{R}^n$,
- $\Sigma_X =$ the covariance matrix of the random vector X,
- $\sigma_0^2 = \mathrm{Var}(X_0)$
- and $c = (\mathrm{Cov}(X_{s_1}, X_0), \ldots, \mathrm{Cov}(X_{s_n}, X_0))' \in \mathbb{R}^n$.

Σ_X, σ_0^2 and c are supposed to be known (or estimated).

Definition

The Kriging is called **Simple Kriging** *if the mean function m of the random field X is known.*

We know that we can write:

$$\mathsf{MSE}(s_0) = \mathbb{E}(\hat{X}_0 - X_0)^2 = \mathrm{Var}(X_0 - \hat{X}_0) + \mathbb{E}^2(X_0 - \hat{X}_0).$$

The variance being invariant by translation, we can first choose λ_0 such that

$$\mathbb{E}(X_0 - \hat{X}_0) = 0.$$

We obtain:

$$\lambda_0 = m(s_0) - \sum_{i=1}^{n} \lambda_i m(s_i)$$

which leads to

$$\hat{X}_0 = m(s_0) + \sum_{i=1}^{n} \lambda_i (X_{s_i} - m(s_i)).$$

Thus, our problem reduces to predict the value of the mean zero process $Y = X - m$ at s_0 by an estimator of the form

$$\hat{Y}_0 = \sum_{i=1}^{n} \lambda_i Y_{s_i}.$$

Simple Kriging

The Simple Kriging supposes that the random field X is mean zero and that we look for a predictor

$$\hat{X}_0 = \sum_{i=1}^{n} \lambda_i X_{s_i} = \lambda' X, \text{ where } \lambda = (\lambda_1, \ldots, \lambda_n)',$$

of X_0 which minimizes the MSE.

Proposition (Simple Kriging)

The linear prediction of X_0 which minimizes the MSE when X is mean zero is

$$\hat{X}_0 = c'\Sigma_X^{-1}X.$$

We say that this estimator is **BLUP: Best Linear Unbiased Predictor**.

The variance of the error of prediction (also called Simple Kriging Variance) is:

$$\sigma^2_{SK}(s_0) = \sigma^2_0 - c'\Sigma_X^{-1}X.$$

Proposition (Reminder)

Let $\mathbb{Y} = (\mathbb{Y}_1, \mathbb{Y}_2)' \in \mathbb{R}^{p+q}$ *be a two blocs gaussian random vector with mean* $\mu = (\mu_1, \mu_2)'$ *and covariance matrix*

$$\Sigma_{\mathbb{Y}} = \begin{pmatrix} \Sigma_{\mathbb{Y}_1} & \Sigma_{\mathbb{Y}_1\mathbb{Y}_2} \\ \Sigma_{\mathbb{Y}_2\mathbb{Y}_1} & \Sigma_{\mathbb{Y}_2} \end{pmatrix}.$$

The conditional distribution of $\mathbb{Y}_1$*, given* $\mathbb{Y}_2 = y_2$ *is gaussian with*

$$\begin{aligned} \mathbb{E}(\mathbb{Y}_1|\mathbb{Y}_2) &= \mu_1 + \Sigma_{\mathbb{Y}_1\mathbb{Y}_2}\Sigma_{\mathbb{Y}_2}^{-1}(\mathbb{Y}_2 - \mu_2) \\ \Sigma_{\mathbb{Y}_1|\mathbb{Y}_2} &= \Sigma_{\mathbb{Y}_1} - \Sigma_{\mathbb{Y}_1\mathbb{Y}_2}\Sigma_{\mathbb{Y}_2}^{-1}\Sigma_{\mathbb{Y}_2\mathbb{Y}_1}. \end{aligned}$$

Proposition

If X is a mean zero gaussian random field, then the Simple Kriging $\hat{X}_0$ is such that

$$\hat{X}_0 = \mathbb{E}(X_0 | X_{s_1}, \ldots, X_{s_n}).$$

Consequences.

- $X_0 - \hat{X}_0$ and $\hat{X}_0$ are independent.
- The conditional distribution of X_0, given $\hat{X}_0 = x$, is gaussian with mean x and variance σ^2_{SK}.

Definition

The Kriging is called **Ordinary Kriging** *if the mean function m of X is constant but unknown.*

We still want to predict X_0 with an estimator $\hat{X}_0$ of the form

$$\hat{X}_0 = \sum_{i=1}^{n} \lambda_i X_{s_i} + \lambda_0,$$

which minimizes the MSE

$$\text{MSE}(s_0) = \mathbb{E}(\hat{X}_0 - X_0)^2.$$

Ordinary Kriging for **second order stationary processes**

Let us denote by a_0 the value of the constant mean function m. We have:

$$\mathbb{E}(\hat{X}_0 - X_0) = \lambda_0 + a_0 \left(\sum_{i=1}^{n} \lambda_i - 1 \right).$$

Thus, in order to get an unbiased linear estimator of X_0, the parameters have to be such that

$$\lambda_0 = 0 \text{ and } \sum_{i=1}^{n} \lambda_i = 1.$$

Thus we have to solve the problem:

$$\left\{ \begin{array}{l} \arg_{(\lambda_1,\ldots,\lambda_n)} \min \left(\sigma_0^2 - 2\lambda' c + \lambda' \Sigma_X \lambda\right) \\ \text{subject to } \sum_{i=1}^n \lambda_i = 1 \end{array} \right. .$$

Using the Lagrange multiplier, one has to solve the system of $n+1$ equations and $n+1$ unknown parameters $(\lambda_1, \ldots, \lambda_n, \mu)$:

$$\left\{ \begin{array}{l} \Sigma_X \lambda + \mu \mathbb{1}_n = c \\ \sum_{i=1}^n \lambda_i = 1 \end{array} \right. ,$$

where $\mathbb{1}_n = (1, \ldots, 1)' \in \mathbb{R}^n$.

Proposition (Ordinary Kriging for second order stationary processes)

Suppose that X is second order stationary process with constant but unknown mean function.
The unbiased linear prediction of X_0 which minimizes the MSE is

$$\hat{X}_0 = \lambda' X,$$

where

$$\lambda = \Sigma_X^{-1} \left(c - \frac{c' \Sigma_X^{-1} \mathbb{1}_n - 1}{\mathbb{1}_n' \Sigma_X^{-1} \mathbb{1}_n} \mathbb{1}_n \right).$$

The variance of the error of prediction (also called Ordinary Kriging Variance) is:

$$\sigma_{OK}^2 = \sigma_0^2 - \lambda' c - \frac{c' \Sigma_X^{-1} \mathbb{1}_n - 1}{\mathbb{1}_n' \Sigma_X^{-1} \mathbb{1}_n}.$$

Ordinary Kriging for **intrinsic processes**

Since an intrinsic process doesn't have necessarily moments of order 1 and 2, we first have to check that we can keep our aim to find a Best (in terms of MSE) Linear Unbiased Predictor.

In fact, it appears that, if we still assume

$$\lambda_0 = 0 \text{ and } \sum_{i=1}^{n} \lambda_i = 1,$$

then the error of prediction $\hat{X}_0 - X_0$ has moment of order one and two.

After some calculus, we can show that:

$$\mathrm{MSE}(s_0) = \mathbb{E}(\hat{X}_0 - X_0)^2 = 2\lambda' v - \lambda' \Gamma_X \lambda,$$

where

$$v = \begin{pmatrix} \gamma(s_0 - s_1) \\ \vdots \\ \gamma(s_0 - s_n) \end{pmatrix} \text{ and } \Gamma_X = (\gamma(s_i - s_j))_{1 \leq i,j \leq n}.$$

Thus, we have to solve the problem:

$$\left\{ \begin{array}{l} \arg_{(\lambda_1,\ldots,\lambda_n)} \min \left(2\lambda' v - \lambda' \Gamma_X \lambda\right) \\ \text{subject to } \sum_{i=1}^{n} \lambda_i = 1 \end{array} \right. .$$

Using the Lagrange multiplier, one has to solve the system of $n+1$ equations and $n+1$ unknown parameters $(\lambda_1, \ldots, \lambda_n, \mu)$:

$$\left\{ \begin{array}{l} \Gamma_X \lambda = v + \mu \mathbb{1}_n \\ \lambda' \mathbb{1}_n = 1 \end{array} \right. .$$

Proposition (Ordinary Kriging for intrinsic processes)

Suppose that X is an intrinsic process with constant but unknown mean function.
The unbiased linear prediction of X_0 which minimizes the MSE is

$$\hat{X}_0 = \lambda' X,$$

where

$$\lambda = \Gamma_X^{-1}\left(v + \frac{1 - v'\Gamma_X^{-1}\mathbb{1}_n}{\mathbb{1}_n'\Gamma_X^{-1}\mathbb{1}_n}\mathbb{1}_n\right).$$

The variance of the error of prediction (also called Ordinary Kriging Variance) is:

$$\sigma^2_{OK} = \lambda' v - \frac{1 - v'\Gamma_X^{-1}\mathbb{1}_n}{\mathbb{1}_n'\Gamma_X^{-1}\mathbb{1}_n}.$$

Definition

The Kriging is called **Universal Kriging** *if the mean function m of X is of the form:*

$$m(s) = \sum_{l=0}^{L} \beta_l f_l(s),$$

where $f_0, f_1, \ldots, f_L$ *are known functions and* $\beta_0, \beta_1, \ldots, \beta_L$ *are unknown real parameters.*

In other words, in case of Universal Kriging, we assume that

$$X_s = \sum_{l=0}^{L} \beta_l f_l(s) + \varepsilon_s,$$

where ε is a mean zero second order stationary (or intrinsic) process.

This model can be written like $X = Z\beta + \varepsilon$, where:

$$X = \begin{pmatrix} X_{s_1} \\ \vdots \\ X_{s_n} \end{pmatrix}, \ Z = \begin{pmatrix} f_0(s_1) & \dots & f_L(s_1) \\ \vdots & & \vdots \\ f_0(s_n) & \dots & f_L(s_n) \end{pmatrix},$$

$$\beta = \begin{pmatrix} \beta_0 \\ \vdots \\ \beta_L \end{pmatrix} \text{ and } \varepsilon = \begin{pmatrix} \varepsilon_{s_1} \\ \vdots \\ \varepsilon_{s_n} \end{pmatrix}.$$

In order to get an unbiased linear predictor, the parameters $\lambda_1, \dots, \lambda_n$ have to satisfy the **universality condition**: $\lambda' Z = z_0'$.

Universal Kriging for **second order stationary processes**

Here we have to solve the problem:

$$\left\{ \begin{array}{l} \arg_{(\lambda_1,\ldots,\lambda_n)} \min \, (\sigma_0^2 - 2\lambda' c + \lambda' \Sigma_X \lambda) \\ \text{subject to } \lambda' Z = z_0' \end{array} \right. .$$

Using the Lagrange multiplier, one has to solve the system of $n + L + 1$ equations and $n + L + 1$ unknown parameters $(\lambda_1, \ldots, \lambda_n)$ and $\mu \in \mathbb{R}^{L+1}$:

$$\left\{ \begin{array}{l} \Sigma_X \lambda = c - Z\mu \\ Z' \lambda = z_0 \end{array} \right. .$$

Proposition (Universal Kriging for second order stationary processes)

Suppose that X is second order stationary process with mean function of the form: $m(s) = \sum_{l=0}^{L} \beta_l f_l(s)$.
The unbiased linear prediction of X_0 which minimizes the MSE is

$$\hat{X}_0 = \left(c'\Sigma_X^{-1} - \left(Z'\Sigma_X^{-1}c - z_0 \right)' \left(Z'\Sigma_X^{-1}Z \right)^{-1} Z'\Sigma_X^{-1} \right) X.$$

The variance of the error of prediction (also called Universal Kriging Variance) is:

$$\sigma_{OK}^2 = \sigma_0^2 - c'\Sigma_X^{-1}c + \left(Z'\Sigma_X^{-1}c - z_0 \right)' \left(Z'\Sigma_X^{-1}Z \right)^{-1} \left(Z'\Sigma_X^{-1}c - z_0 \right).$$

Universal Kriging for **intrinsic processes**

One can show that the universality condition ensures here also that $\hat{X}_0 = \lambda' X$ is unbiased. If, in addition, the first baseline function f_0 is the constant function equal to 1, i.e. $f_0 \equiv 1$, then the variance of $\hat{X}_0 - X_0$ exists and we have:

$$\mathrm{MSE}(s_0) = 2\lambda' v - \lambda' \Gamma_X \lambda.$$

The problem to solve is:

$$\begin{cases} \arg_{(\lambda_1,\ldots,\lambda_n)} \min \left(2\lambda' v - \lambda' \Gamma_X \lambda\right) \\ \text{subject to } \lambda' Z = z_0' \end{cases} .$$

Proposition (Universal Kriging for intrinsic processes)

Suppose that X is an intrinsic process with mean function of the form: $m(s) = \sum_{l=0}^{L} \beta_l f_l(s)$.
The unbiased linear prediction of X_0 which minimizes the MSE is

$$\hat{X}_0 = \left(v + Z \left(Z' \Gamma_X^{-1} Z \right)^{-1} \left(z_0 - Z' \Gamma_X^{-1} v \right) \right)' \Gamma_X^{-1} X.$$