Numerical activities: kriging model

3 hours - no report

The goal is to take in hand the numerical tool offered by kriging approach.

Thus, each student will progressively implement on his/her own a one-dimensional kriging code and investigate basic behavior. Then, depending on own interest and motivation, different aspects can be analysed and comprehend.

1 Random fields, correlation and auto-correlation

The first part is a preliminary investigations about three basic concepts: random field description, correlation, and auto-correlation.

1.1 Random field

1. Create a random field, where the value for each point is considered as an independent random variable.

1.2 Correlation between two random variables

2. Evaluate the correlation between two random variables.

1.3 Correlation structure

3. Investigate the usual analytic laws considered as correlation structures.

1.4 Random field assuming a correlation structure between the points

4. Create various realisations of random fields assuming a correlation structure between the points.

2 Implementation of ordinary kriging model

Let choose a set of observations $S = \{x^i, y^i\}$. Let assume an auto-correlation structure, and an auto-correlation length:

5. Implement the kriging interpolation.
6. Plot the mean of the kriging model.
7. Plot the variance of the kriging model.
8. Plot some realizations of the kriging model.

3 Investigation of the effect of the auto-correlation structure

9. Change the auto-correlation length and investigate the effect of that parameter on the kriging model
10. Change the auto-correlation structure and investigate the effect of that hypothesis on the kriging model

4 Options

4.1 Problem due to very close points

11. Choose two observation points which are very close. What is the influence on the kriging model? Which numerical issue do you encounter?

4.2 Optimisation of the auto-correlation hyper-parameter

12. Choose an auto-correlation structure and evaluate the auto-correlation parameter based on an optimisation problem.

4.3 Variants of ordinary kriging

13. Implement a simple kriging model.
14. Implement a universal kriging model.

4.4 Two-dimensional implementation

15. Implement identically the kriging model for a two-dimensional parametric space.

Summary about ordinary kriging

Among kriging metamodeling, several families, such as simple kriging, ordinary kriging and universal kriging, may be distinguished depending on considered assumptions leading to different complexity levels.
Let consider here ordinary kriging, then, the response surface is estimated as the mean of a stochastic process Y given by

$$Y(x) = \mu + Z(x), \tag{1}$$

with μ the global mean and $Z(x)$ a stationary Gaussian process. The metamodel is based on m observations. Consider a black-box function $\mathcal{M} : \mathbb{X} \to \mathbb{Y}$ between an input $\boldsymbol{x} \in \mathbb{X} \subset \mathbb{R}^n$ and a univariate output $y \in \mathbb{Y} \subset \mathbb{R}$. Furthermore consider some existing samples $\chi = \{\boldsymbol{x}^1, \ldots, \boldsymbol{x}^m\}$ corresponding with a set of training data $S = \{x^i, y^i\}$.
The element ij of the covariance matrix $\boldsymbol{Cov}$ relative to this stochastic process yields

$$\boldsymbol{Cov}_{ij} = cov[Z(\boldsymbol{x}^i), Z(\boldsymbol{x}^j)] = \sigma^2 \boldsymbol{R}_{ij}(\boldsymbol{\theta}) \,, \tag{2}$$

with cov the covariance operator, σ the standard deviation of the stochastic process and $\boldsymbol{R}_{ij}(\boldsymbol{\theta})$ the correlation between outputs corresponding with two samples $\boldsymbol{x}^i$ and $\boldsymbol{x}^j$, defined as the component of the auto-correlation matrix $\boldsymbol{R}$, also named correlation matrix. The correlation function R, which depends on unknown correlation parameters $\boldsymbol{\theta}$, is usually chosen by the user. Correlation parameters are estimated as solution of an optimization problem, and the elements of the correlation matrix read $\boldsymbol{R}_{ij} = R(\boldsymbol{x}^i, \boldsymbol{x}^j, \boldsymbol{\theta})$.

Thus, the idea of kriging metamodeling is to obtain $\widehat{\mathcal{M}}$ the most accurate approximation of $\mathcal{M}$ for any point $\boldsymbol{x}^0$ belonging to $\mathbb{X}$ as the mean of the realizations of the stochastic process defined by Eq.(1) at that point i.e.

$$\mathcal{M}(\boldsymbol{x}^0) \approx \widehat{\mathcal{M}}(\boldsymbol{x}^0) = \mu_{\widehat{Y}(\boldsymbol{x}^0)}. \tag{3}$$

The best linear unbiased predictor for any unobserved value y^0 corresponding with $\boldsymbol{x}^0 \in \mathbb{X}$ yields

$$\mu_{\widehat{Y}(\boldsymbol{x}^0)} = \widehat{\mu} + \boldsymbol{r}_0^T \boldsymbol{R}^{-1}(\boldsymbol{y} - \mathbf{1}\widehat{\mu}), \tag{4}$$

with $\mathbf{1}$ the vector with m components equal to 1 and $\boldsymbol{y}$ the vector gathering the m observation outputs. The prior estimation of the global mean denoted $\widehat{\mu}$ is obtained through a generalized least-square estimate as

$$\widehat{\mu} = (\mathbf{1}^T \boldsymbol{R}^{-1} \mathbf{1})^{-1} \mathbf{1}^T \boldsymbol{R}^{-1} \boldsymbol{y}. \tag{5}$$

It can be seen that it depends on the choice of the auto-correlation structure. The vector $\boldsymbol{r}_0$ collects the cross-correlations between $\boldsymbol{x}^0$ and every sample point as

$$r_{0_i} = R(\boldsymbol{x}^0, \boldsymbol{x}^i, \boldsymbol{\theta}) \quad \text{for } i = 1, \ldots, m. \tag{6}$$

Besides, information about the variance of the metamodel can be extracted for any point $\boldsymbol{x}^0$as

$$\sigma^2_{\widehat{Y}_0(\boldsymbol{x}^0)} = \widehat{\sigma}^2 \left(1 - \boldsymbol{r}_0^T \boldsymbol{R}^{-1} \boldsymbol{r}_0 + \boldsymbol{u}_0^T \left(\mathbf{1}^T \boldsymbol{R}^{-1} \mathbf{1}\right)^{-1} \boldsymbol{u}_0\right), \tag{7}$$

with

$$\boldsymbol{u}_0 = \mathbf{1}^T \boldsymbol{R}^{-1} \boldsymbol{r}_0 - \mathbf{1} \,, \tag{8}$$

and the prior estimation of the global variance which reads

$$\widehat{\sigma}^2 = \frac{1}{m} (\boldsymbol{y} - \mathbf{1}\widehat{\mu})^T \boldsymbol{R}^{-1} (\boldsymbol{y} - \mathbf{1}\widehat{\mu}) \,. \tag{9}$$

Similarly to the prior estimation of the mean, prior estimation of the global variance depends also on the correlation matrix.

Propositions of auto-correlation functions

The power exponential correlation function reads

$$R(\boldsymbol{x}-\boldsymbol{x'},\boldsymbol{\theta}) = \exp\left(-\sum_{i=1}^{n}\left(\frac{|x_i - x'_i|}{\theta_i}\right)^{\nu}\right) \tag{10}$$

with the scale parameters $\{\theta_i > 0, i = 1, \ldots, n\}$ and $0 < \nu \leq 2$. As special cases of this function the exponential ($\nu = 1$) and squared exponential or Gaussian ($\nu = 2$) forms can be considered
Matérn 3/2 auto-correlation is given by

$$R(\boldsymbol{x}-\boldsymbol{x'},\boldsymbol{\theta},\nu=3/2) = \prod_{i=1}^{n}\left(1+\frac{\sqrt{3}|x_i - x'_i|}{\theta_i}\right)\exp\left(-\frac{\sqrt{3}|x_i - x'_i|}{\theta_i}\right) \tag{11}$$

Matérn 3/2 auto-correlation is defined as

$$\begin{aligned} R(\boldsymbol{x}-\boldsymbol{x'},&\boldsymbol{\theta},\nu=5/2) = \\ &\prod_{i=1}^{n}\left(1+\frac{\sqrt{5}|x_i - x'_i|}{l_i}+\frac{5(x_i - x'_i)^2}{3\theta_i^2}\right)\exp\left(-\frac{\sqrt{5}|x_i - x'_i|}{\theta_i}\right). \end{aligned} \tag{12}$$

Optimisation of the hyperparameters

Generally, there is no analytical solution for the estimation of the hyperparameters $\boldsymbol{\theta}$. Therefore the correlation matrix consisting of $\sigma^2 R(\bullet,\boldsymbol{\theta})$ is a priori unknown. In order to estimate $\boldsymbol{\theta}$ the application-dependent autocorrelation function has to be chosen. The most common technique for estimating $\boldsymbol{\theta}$ and its variance σ^2 is the so-called maximum likelihood estimation (MLE).
MLE tries to maximize the likelihood of the observations $\boldsymbol{y}$ defined by the multivariate normal probability density function

$$\begin{aligned} L(\boldsymbol{y}|\boldsymbol{\beta},\sigma,\boldsymbol{\theta}) =& \frac{1}{\left((2\pi\sigma^2)^m[\det \boldsymbol{R}(\boldsymbol{\theta})]\right)^{1/2}} \\ & \exp\left[-\frac{1}{2\sigma^2}(\boldsymbol{y}-\boldsymbol{F}\boldsymbol{\beta})^T\boldsymbol{R}(\boldsymbol{\theta})^{-1}(\boldsymbol{y}-\boldsymbol{F}\boldsymbol{\beta})\right], \end{aligned} \tag{13}$$

which only depends on $\sigma^2, \boldsymbol{\beta}$ and $\boldsymbol{\theta}$. Maximizing this quantity is equivalent to minimizing its opposite natural logarithm which yields

$$\begin{aligned} -\log L\left(\boldsymbol{y}|\boldsymbol{\beta},\sigma^2,\boldsymbol{\theta}\right) =& \frac{1}{2\sigma^2}\left(\boldsymbol{y}-\boldsymbol{F}\boldsymbol{\beta}\right)^T\boldsymbol{R}(\boldsymbol{\theta})^{-1}\left(\boldsymbol{y}-\boldsymbol{F}\boldsymbol{\beta}\right)+\frac{m}{2}\log(2\pi) \\ &+\frac{m}{2}\log\left(\sigma^2\right)+\frac{1}{2}\log\left([\det(\boldsymbol{R}(\boldsymbol{\theta}))]\right), \end{aligned} \tag{14}$$

with the maximum likelihood estimates as a function of the hyperparameters given as

$$\begin{aligned} \hat{\boldsymbol{\beta}}(\boldsymbol{\theta}) &= \left(\boldsymbol{F}^T\boldsymbol{R}(\boldsymbol{\theta})\boldsymbol{F}\right)^{-1}\boldsymbol{F}^T\boldsymbol{R}(\boldsymbol{\theta})^{-1}\boldsymbol{y} \\ \hat{\sigma}^2(\boldsymbol{\theta}) &= \frac{1}{m}\left(\boldsymbol{y}-\boldsymbol{F}\hat{\boldsymbol{\beta}}\right)^T\boldsymbol{R}(\boldsymbol{\theta})^{-1}\left(\boldsymbol{y}-\boldsymbol{F}\hat{\boldsymbol{\beta}}\right). \end{aligned} \tag{15}$$

Equation (14) can be expressed to be only dependent on the hyperparameters with

$$\begin{aligned} -\log L\left(\boldsymbol{y}|\boldsymbol{\beta},\sigma^2,\boldsymbol{\theta}\right) &= \frac{m}{2}+\frac{m}{2}\log(2\pi)+\frac{m}{2}\log\left(\hat{\sigma^2}(\boldsymbol{\theta})\right)+\frac{1}{2}\log\left([\det R(\boldsymbol{\theta})]\right) \\ &= \frac{m}{2}\log\left(\psi(\boldsymbol{\theta})\right)+\frac{m}{2}\left(\log(2\pi)+1\right). \end{aligned} \tag{16}$$

Here,

$$\psi(\boldsymbol{\theta}) = \hat{\sigma^2}(\boldsymbol{\theta})[\det R(\boldsymbol{\theta})]^{1/m} \tag{17}$$

is the so-called reduced likelihood function. Eventually the maximum likelihood estimate of the hyperparameters $\boldsymbol{\theta}$ can be calculated by evaluating an auxiliary optimization problem of the form

$$\hat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}^\star} \psi(\boldsymbol{\theta}^\star)\,. \tag{18}$$

Since there is no analytical solution for the optimization problem, it is necessary to use numerical optimization tools. This step is the most challenging for the construction of surrogate models with Kriging because of the multi-modality of the likelihood function. For this reason, during optimization the hyperparameters are constrained.

Variants of ordinary kriging

Considering simple kriging, the random field is still defined by Eq. 1 but the mean μ is assumed by the user. Considering universal kriging the term μ is no more constant but estimated as $F^T\beta$.