
Problèmes numériques lors de la construction d'un modèle de krigeage par processus gaussien

Jean-Marc Martinez

CEA-Saclay, DEN, DM2S, STMF, LGLS F-91191 Gif-sur-Yvette, France.

Séminaire LRC Manon - 29 mars 2013 Saclay

- ▶ Krigeage : modèle prédictif développé initialement en géostatistique par D. G. Krige (thèse, 1951) puis formalisé par G. Matheron (1970)
- ▶ Interprétation probabiliste : les données observées vues comme réalisations d'un processus stochastique
- ▶ On postule des dépendances spatiales \Rightarrow structure de covariance
- ▶ Une mise en oeuvre *pratique* lorsque le processus est supposé gaussien

Les utilisations du krigeage

- ▶ méthode utilisée en **Machine Learning**, régression, classification supervisée (reconnaitances de formes), (C. E. Rasmussen),
- ▶ en simulation numérique, **Design and Analysis of Computer Experiments** (T. J. Santer),
- ▶ en **optimisation** pour la recherche d'extremum, Algorithme EGO, Efficient Global Optimization (D. R. Jones)
- ▶ en **validation** en complétant la modélisation physique à partir de données expérimentales (thèse CEA F. Bachoc, Dir. J. Garnier),
- ▶ **GdR Mascot-Num** cadre alimentant et diffusant la R&D,
- ▶ Outils développés dans le consortium **ReDice** (package R)
- ▶ Librairie **gpLib** (CEA) → Scilab, URANIE (F. Gaudier)

Processus gaussiens

- ▶ Processus gaussien $Y(\mathbf{x})$ indexé sur R^d est complètement spécifié par sa moyenne $\mu(\mathbf{x})$ et sa fonction de covariance $k(\mathbf{x}_1, \mathbf{x}_2)$

$$Y(\mathbf{x}) \sim \mathcal{GP}(\mu(\mathbf{x}), k(\mathbf{x}, \bullet))$$

- ▶ Processus stationnaire lorsque $k(\mathbf{x}_1, \mathbf{x}_2) = k(\mathbf{x}_1 - \mathbf{x}_2)$
- ▶ Exemples : corrélations de Gauss et Matérn

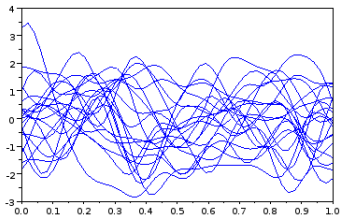
$$k(\mathbf{x}_1, \mathbf{x}_2) = \sigma^2 \exp - \sum_{i=1}^d \left(\frac{\mathbf{x}_{1i} - \mathbf{x}_{2i}}{l_{c_i}} \right)^2$$

$$k(\mathbf{x}_1, \mathbf{x}_2) = \sigma^2 \prod_{i=1}^d \frac{2}{\Gamma(\nu_i)} \left(\sqrt{\nu_i} \frac{\mathbf{x}_{1i} - \mathbf{x}_{2i}}{l_{c_i}} \right)^{\nu_i} K_{\nu_i} \left(2\sqrt{\nu_i} \frac{\mathbf{x}_{1i} - \mathbf{x}_{2i}}{l_{c_i}} \right)$$

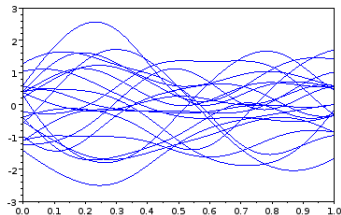
- ▶ Paramètres : la variance du processus σ^2 , les longueurs de corrélation l_{c_i} et les régularités ν_i pour la corrélation de Matérn

Echantillons de 4 processus gaussiens indexés sur R

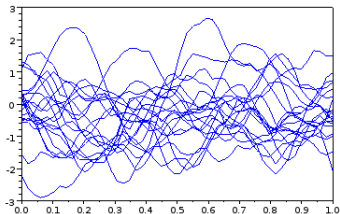
Gauss - longueur de corrélation = 0.1



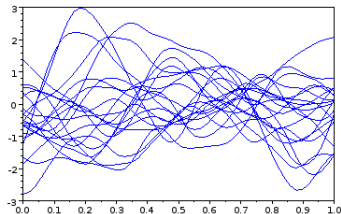
Gauss - longueur de corrélation = 0.3



Matérn 3/2 - longueur de corrélation = 0.2



Matérn 7/2 - longueur de corrélation = 0.2



Modèle de prédiction

- ▶ Vecteur gaussien pour le modèle des observations

$$Y_n = (Y(\mathbf{x}_1), Y(\mathbf{x}_2), \dots, Y(\mathbf{x}_n))^T \sim \mathcal{N}(\boldsymbol{\mu}, K)$$

$$\mu_i = \mu(\mathbf{x}_i) \quad , \quad K(i, j) = k(\mathbf{x}_i - \mathbf{x}_j)$$

- ▶ Modèle intégrant la valeur à prédire en un nouveau point \mathbf{x}_{new}

$$\mathbf{k} = (k(\mathbf{x}_1, \mathbf{x}_{new}), k(\mathbf{x}_2, \mathbf{x}_{new}), \dots, k(\mathbf{x}_n, \mathbf{x}_{new}))^T$$

$$\begin{pmatrix} Y_n \\ Y(\mathbf{x}_{new}) \end{pmatrix} \sim \mathcal{N} \left(\begin{pmatrix} \boldsymbol{\mu} \\ \mu(\mathbf{x}_{new}) \end{pmatrix}, \begin{pmatrix} K & \mathbf{k} \\ \mathbf{k}^T & k(0) \end{pmatrix} \right)$$

- ▶ Conditionnement gaussien \Rightarrow loi de $Y(\mathbf{x}_{new})$ conditionnellement à une réalisation du vecteur gaussien Y_n

$$\mathbb{E}[Y(\mathbf{x}_{new})|Y_n] = \mu(\mathbf{x}_{new}) + \mathbf{k}^T K^{-1}(Y_n - \boldsymbol{\mu})$$

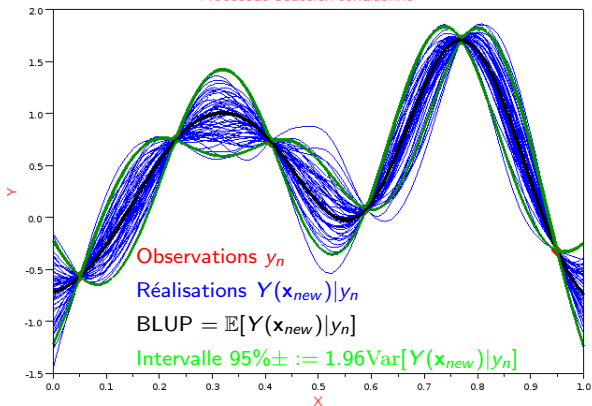
$$\text{Var}[Y(\mathbf{x}_{new})|Y_n] = k(0) - \mathbf{k}^T K^{-1} \mathbf{k}$$

- ▶ BLUP, Best Linear Unbiased Predictor $\hat{Y}(\mathbf{x}_{new}) = \mathbb{E}[Y(\mathbf{x}_{new})|Y_n]$

Processus gaussiens conditionnellement aux observations

Processus $Y(x) \sim (\mu = 0, \sigma^2 = 1 \times \text{Matérn}[lc = 0.3, \nu = 5/2])$

Processus Gaussien conditionné



Construction d'un krigeage

- ▶ Structure général du modèle du processus (sans erreurs de mesure)

$$Y_{obs}(\mathbf{x}) = \underbrace{\mathbf{h}^T(\mathbf{x})\boldsymbol{\beta}}_{\text{moyenne}:=\mu(\mathbf{x})} + \underbrace{Z(\mathbf{x}; \sigma^2, \theta)}_{\text{processus centré}}$$

- ▶ Estimation des paramètres $\boldsymbol{\beta}; \sigma^2, \theta$ par différentes méthodes : maximum de vraisemblance, maximum de la vraisemblance restreinte, par leave one out
- ▶ Dans le cas où il n'y a pas d'erreur de mesures, la matrice de covariance $K_{\theta, \sigma^2} = \sigma^2 K_{\theta}$
- ▶ Estimation des paramètres θ par maximum de vraisemblance à partir des données observées \mathbf{y}_{obs} . **Problème à résoudre :**

$$\hat{\theta} = \arg \min_{\theta} \ln |K_{\theta}| + n \ln [\mathbf{y}_{obs}^T K_{\theta}^{-1} (I - H(H^T K_{\theta}^{-1} H)^{-1} H^T K_{\theta}^{-1}) \mathbf{y}_{obs}]$$

- ▶ Puis estimation de la variance et des paramètres de la moyenne :

$$\hat{\sigma}^2 = \frac{1}{n} \mathbf{y}_{obs}^T K_{\hat{\theta}}^{-1} (I - H(H^T K_{\hat{\theta}}^{-1} H)^{-1} H^T K_{\hat{\theta}}^{-1}) \mathbf{y}_{obs}$$

$$\hat{\boldsymbol{\beta}} = (H^T K_{\hat{\theta}}^{-1} H)^{-1} H^T K_{\hat{\theta}}^{-1} \mathbf{y}_{obs}$$

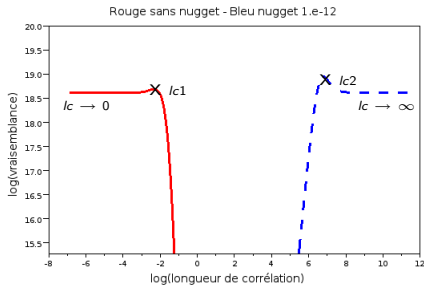
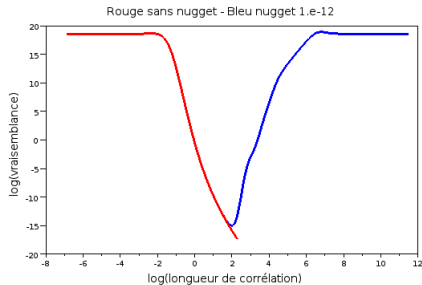
- ▶ Estimation des paramètres nécessite le calcul du déterminant et de l'inverse de la matrice de covariance K , matrice symétrique définie positive
- ▶ Problème : taille et conditionnement de K
- ▶ Exemple : processus sur R , corrélation gaussienne, longueur de corrélation $lc = 0.2$, nombre d'observations $n = 20$, conditionnement(K) $\simeq 1.55e11$, déterminant(K) $\simeq 1.27e - 61$.
- ▶ Et pour $n = 40$, la matrice K n'est pas *numériquement* définie positive (Scilab 5.4.0, décomposition de Cholesky basée sur les routines Lapack DPOTRF)

Effet nugget

- ▶ Ce qui se fait : amélioration du conditionnement de la matrice de corrélation en ajoutant une petite valeur (*nugget*) sur la diagonale de la matrice de covariance
- ▶ MAIS : analyse de l'effet *nugget* (I. Andrianakis, P. G. Challenor, Computational Statistics and Data Analysis, december 2012)
 1. par effet *nugget*, équivalence entre la vraisemblance (restreinte) pour des longueurs de corrélation nulle et infini
 2. la solution obtenue sur l'émulateur du code peut être *polluée* par l'effet *nugget* même si celui - ci est de l'ordre de la précision machine

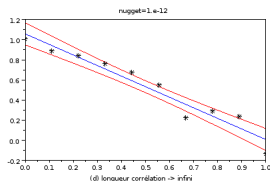
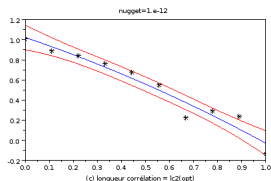
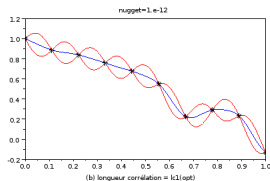
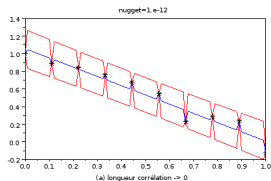
Exemple de l'effet nugget (1/2)

Processus sur R , 10 points, Matérn $7/2$. Evolution du log de la vraisemblance (restreinte) en fonction de la longueur de corrélation (rouge sans nugget, en bleu nugget = 10^{-12} , effet à partir de $\ln(lc) > 2$)



Exemple de l'effet nugget (2/2)

Prédictions des modèles de krigeage pour $lc \rightarrow 0, lc1, lc2, lc \rightarrow \infty$:



⇒ Grande variabilité des prédicteurs pour des vraisemblances *voisines* pour les 4 longueurs de corrélation retenues

Conclusions

- ▶ Estimation des paramètres du modèle de krigeage nécessite le traitement de matrices pleines définies positives
- ▶ Problèmes posés par la dimension, la taille et le conditionnement de la matrice de covariance
- ▶ La stabilisation *numérique* par *nugget* peut perturber les solutions obtenues sur les modèles de krigeage
- ▶ la recherche des paramètres devrait être initialisée avec de faibles valeurs des longueurs de corrélation où avec un faible degré de régularité (matrice à diagonale dominante)
- ▶ Pour des matrices pleines, s.d.p., existe-t-il des pré-conditionneurs ?
- ▶ Apport éventuel des méthodes itératives ?
- ▶ Décomposition Cholesky → en valeurs singulières ?