

ÉCOLE NATIONALE DES PONTS ET CHAUSSÉES

---

Résolution d'une équation aux dérivées  
partielles par processus gaussien

---

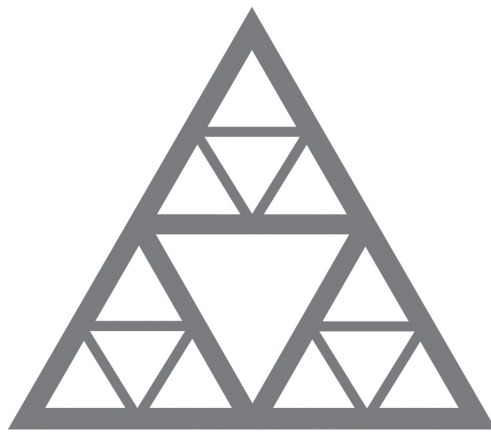
Projet 1A

Par

*Nils Baulier, Raphaël Viard, Salma Kably et Simon Weissberg*

Supervisé par

*Urbain Vaes*



**École des Ponts**

ParisTech

# Table des matières

## 1 Introduction

Historiquement, les processus de régression gaussienne furent introduits par Krige en 1960, cherchant à l'époque à obtenir la distribution d'or dans un milieu, en ne se basant que sur des échantillons. Aujourd'hui, ils permettent par exemple d'étudier la perméabilité radioactive d'un sol. L'approche de résolution d'équations aux dérivées partielles par processus gaussien est une technique d'approximation de la solution par une fonction ayant un caractère aléatoire, permettant par exemple de modéliser l'incertitude sur une donnée du problème. Avant de se lancer dans une telle résolution, il paraît naturel d'abord de simplement essayer d'approcher une fonction connue par un processus gaussien (c'est la régression gaussienne), ce qui sera l'objet principal de ce premier rendu. Cette première étape nous orientera pour ensuite résoudre des équations différentielles linéaires.

## 2 Processus gaussien

### 2.1 Définition

**Définition 1.** *Un processus stochastique est une famille de variables aléatoires  $(X_t)_{t \in T}$  définies sur le même espace de probabilité  $(\Omega, \mathcal{F}, \mathcal{P})$  indexée par  $T$  et à valeurs dans  $S$ . Un processus stochastique peut donc aussi être vu comme une application  $X : \Omega \times T \rightarrow S$  tel que pour tout  $t$  dans  $T$ ,  $X(., t)$  est une variable aléatoire.*

*Un processus gaussien est un processus stochastique  $(X_t)_{t \in T}$  tel que  $\forall n \in \mathbb{N}, \forall (t_1, t_2, \dots, t_n) \in T^n$ ,  $(X_{t_1}, \dots, X_{t_n})$  est un vecteur gaussien.*

**Proposition 1.** *Soit  $(X_t)_{t \in T}$  un processus gaussien. Alors il existe  $\bar{m} : T \rightarrow \mathbb{R}$  et  $\bar{k} : T \times T \rightarrow \mathbb{R}$  définie positive tel que  $\forall n \in \mathbb{N}, \forall (t_1, t_2, \dots, t_n) \in T^n$ ,  $(X_{t_1}, \dots, X_{t_n}) \sim \mathcal{N}((\bar{m}(t_1), \dots, \bar{m}(t_n))^\top, K_{t_1, t_2, \dots, t_n})$  où  $K_{t_1, t_2, \dots, t_n}$  est la matrice de coefficients  $K_{i,j} = \bar{k}(t_i, t_j)$ .*

*Réciproquement, s'il existe un tel  $m$  et un tel  $c$ , alors  $(X_t)_{t \in T}$  est un processus gaussien. On note alors  $(X_t)_{t \in T} \sim \mathcal{GP}(\bar{m}, \bar{k})$  et on appelle  $\bar{m}$  la fonction moyenne et  $\bar{k}$  le noyau de covariance du processus gaussien.*

### 2.2 Modélisation d'une trajectoire d'un processus gaussien

## 3 Interpolation gaussienne

### 3.1 Régression gaussienne

**Définition 2.** *Soit  $\chi$  un ensemble non vide et  $f : \chi \rightarrow \mathbb{R}$  une fonction. Soit  $n \in \mathbb{N}$  et  $(x_i, y_i)_{i=1}^n \subset \chi \times \mathbb{R}$  tel que*

$$y_i = f(x_i) + \xi_i, \quad i = 1, \dots, n$$

*Les  $(x_i, y_i)$  sont les points que l'on cherche à interpoler et les  $\xi_i$  sont des variables aléatoires qui représentent le "bruit", ou alors l'incertitude/variabilité des réponses que l'on peut avoir pour une même entrée.*

Le problème de la régression est d'estimer la fonction  $f$  en se basant (\*). La fonction  $f$  est alors appelée fonction de régression.

La régression gaussienne est d'imposer  $f \sim \mathcal{GP}(m, k)$  et  $\xi_i \sim \mathcal{N}(0, \sigma^2)$ ,  $i = 1, \dots, n$ . Ici,  $m$  et  $k$  doivent être choisis de manière à représenter des connaissances ou des hypothèses sur la fonction de régression.

**Lemme 1.** Soit  $a \in \mathbb{R}^d$  et  $b \in \mathbb{R}$  des vecteurs gaussiens tels que

$$\begin{bmatrix} a \\ b \end{bmatrix} \sim \mathcal{N} \left( \begin{bmatrix} \mu_a \\ \mu_b \end{bmatrix} \begin{bmatrix} A & C \\ C^T & B \end{bmatrix} \right)$$

où  $\mu_a$  et  $\mu_b$  sont les vecteurs moyennes,  $A$  et  $B$  les matrices de covariances et  $C \in \mathbb{R}^{n \times m}$ . Alors la distribution conditionnelle de  $b$  sachant  $a$  est donnée par

$$b|a \sim \mathcal{N}(\mu_b + C^T A^{-1}(a - \mu_a), B - C^T A^{-1}C)$$

**Théorème 1.** Supposons  $f$  solution du problème de régression gaussienne. Alors

$$f|Y \sim \mathcal{GP}(\bar{m}, \bar{k}),$$

où  $\bar{m} : \chi \rightarrow \mathbb{R}$  et  $\bar{k} : \chi \times \chi \rightarrow \mathbb{R}$  sont données par

$$\bar{m}(x) = m(x) + k_{xX}(k_{XX} + \sigma^2 I_n)^{-1}(Y - m_X), \quad x \in \chi$$

$$\bar{k}(x, x') = k(x, x') - k_{xX}(k_{XX} + \sigma^2 I_n)^{-1}k_{Xx'}, \quad x, x' \in \chi$$

où  $k_{XX} = k_{xX}^T = (k(x_1, x), \dots, k(x_n, x))^T$  et  $m_X^T = (m(x_1), \dots, m(x_n))^T$

### 3.2 Espace à noyau reproduisant

Une façon de retrouver l'expression de la moyenne obtenu dans le Théorème 1 est de d'utiliser un espace à noyau reproduisant.

**Définition 3.** Soit  $\chi$  un ensemble non vide. Un noyau défini positif est une fonction symétrique  $k : \chi \times \chi \rightarrow \mathbb{R}$  tel que  $\forall n \in \mathbb{N}, (c_1, \dots, c_n) \in \mathbb{R}^n$  et  $(x_1, \dots, x_n) \in \chi^n$

$$\sum_{i=1}^n \sum_{j=1}^n c_i c_j k(x_i, x_j) \geq 0.$$

**Remarque 1.** En reprenant les notations de I.a., on observe que le noyau de covariance  $c$  peut être vu comme un noyau défini positif sur  $T \times T$ .

**Définition 4.** Soit  $\chi$  un ensemble non vide. Un espace de Hilbert  $(\mathcal{H}_k, \langle \cdot, \cdot \rangle_{\mathcal{H}_k})$  de fonctions de  $\chi$  est appelé un espace à noyau reproduisant de noyau reproduisant  $k$  si les conditions suivantes sont satisfaites :

$$(i) \forall x \in \chi, k(\cdot, x) \in \mathcal{H}_k;$$

$$(ii) \forall x \in \chi, \forall f \in \mathcal{H}_k, f(x) = \langle f, k(\cdot, x) \rangle_{\mathcal{H}_k}$$

**Remarque 2.** (ii) équivaut à dire que l'évaluation est continue sur l'espace des fonctions de  $\mathcal{H}_k$

### 3.3 Méthode des moindres carrés

Nous souhaitons interpoler l'ensemble de points  $(x_1, y_1), \dots, (x_n, y_n)$ . Nous nous intéressons d'abord à résoudre le problème

$$\hat{f} = \operatorname{argmin} \left( \frac{1}{n} \sum_{i=1}^n L(x_i, y_i, f(x_i)) + \lambda \|f\|_{\mathcal{H}_k}^2 \mid f \in \mathcal{H}_k \right)$$

où  $L : \mathcal{X} \times \mathbb{R} \times \mathbb{R} \rightarrow \mathbb{R}^+$  est une fonction de perte et  $\lambda > 0$  une constante de régularisation.  $L$  joue sur la différence entre les points  $y_i$  et les  $f(x_i)$ . Nous allons ici nous intéresser à l'estimateur des moindres carrés  $L(x, y, y') = (y - y')^2$  donc au problème suivant :

$$\hat{f} = \operatorname{argmin} \left( \frac{1}{n} \sum_{i=1}^n (f(x_i) - y_i)^2 + \lambda \|f\|_{\mathcal{H}_k}^2 \mid f \in \mathcal{H}_k \right)$$

**Théorème 2.** *La solution est unique, et est donnée par*

$$\hat{f}(x) = k_{xX} (k_{XX} + n\lambda I_n)^{-1} Y$$

*Démonstration.* Notons  $h : f \in \mathcal{H}_k \rightarrow \mathbb{R}$  la fonction  $h(f) = \frac{1}{n} \sum_{i=1}^n (f(x_i) - y_i)^2 + \lambda \|f\|_{\mathcal{H}_k}^2$ , qu'on cherche donc à minimiser.

Posons  $\mathcal{H}_0 = \operatorname{vect}(\{k(\cdot, x_i)\}_{i=1, \dots, n})$ .  $\mathcal{H}_0$  est un sous espace vectoriel de  $\mathcal{H}_k$ . Ainsi, on peut considérer  $\mathcal{H}_0^\perp$  supplémentaire orthogonal de  $\mathcal{H}_0$ .

Soit  $f^*$  solution optimale du problème. Alors il existe  $f_0^* \in \mathcal{H}_0$  et  $g^* \in \mathcal{H}_0^\perp$  tel que  $f^* = f_0^* + g^*$

Ainsi,  $\|f^*\|_{\mathcal{H}_k}^2 = \|f_0^*\|_{\mathcal{H}_k}^2 + \|g^*\|_{\mathcal{H}_k}^2$  (théorème de pythagore)

et  $f^*(x_i) = \langle f, k(\cdot, x_i) \rangle_{\mathcal{H}_k} = \langle f_0^*, k(\cdot, x_i) \rangle_{\mathcal{H}_k} + \langle g^*, k(\cdot, x_i) \rangle_{\mathcal{H}_k} = \langle f_0^*, k(\cdot, x_i) \rangle_{\mathcal{H}_k}$  car  $g^* \mathcal{H}_0^\perp$

donc finalement  $h(f^*) = h(f_0^*) + \|g^*\|_{\mathcal{H}_k}^2$  où  $\|g^*\|_{\mathcal{H}_k}^2 > 0$  si  $f^*$  n'appartient pas à  $\mathcal{H}_0$ , ce qui montre que  $f_0^*$  est une meilleure solution au problème, ce qui n'est pas car  $f$  est par hypothèse la solution optimale de ce dernier. Donc la solution du problème si elle existe est un élément de  $\mathcal{H}_0$ .

Il suffit alors de résoudre le problème sur  $\mathcal{H}_0$  (espace de dimension finie), ce qui aboutit à la formule annoncée.  $\square$

**Proposition 2.** *Soit  $k$  un noyau défini positif sur un ensemble  $\mathcal{X}$  et définissons  $X := (x_1, \dots, x_n) \in \mathcal{X}^n$  et  $Y := (y_1, \dots, y_n) \in \mathbb{R}^n$ . Alors  $\bar{m} = \hat{f}$  si  $\sigma^2 = n\lambda$  où*

*$\bar{m}$  est la fonction moyenne de la regression gaussienne sur  $(X, Y)$ , le prateur du processus gaussien  $\{ \sim \mathcal{GP}(0, k) \}$  en considérant le bruit  $\xi_1, \dots, \xi_n \sim \mathcal{N}(0, \sigma^2)$*

*$\hat{f}$  est la solution de la régression des moindres carrées de  $(X, Y)$  basé sur l'espace à noyau reproduisant associé à  $k$ .*

### 3.4 Interpolation

## 4 Equations aux dérivées partielles linéaires

### 4.1 Définition

**Définition 5.** *Soit  $\mathcal{L} : \mathbb{U} \rightarrow \mathbb{U}$  un opérateur linéaire. On appelle équation linéaire sur  $\mathbb{U}$  une équation de la forme  $\mathcal{L}[u] = f$*

**Proposition 3.** *Les equations aux dérivées partielles sont des equations linéaires sur  $\mathbb{U}$  où ici  $\mathbb{U}$  est un espace de fonctions suffisamment régulière pour que l'opérateur différentiel linéaire considéré y soit défini.*

## 4.2 Conditionner sur un opérateur linéaire

**Théorème 3.** *Soit  $u \sim \mathcal{GP}(m, k)$ ,  $\mathcal{L} : \mathbb{U} \rightarrow \mathbb{R}$  un opérateur linéaire et une variable gaussienne indépendante  $\epsilon \sim \mathcal{N}(\mu, \Sigma)$ . Alors*

$$u | (\mathcal{L}[u] + \epsilon = y) \sim \mathcal{GP}(m^{u|y}, k^{u|y})$$

où la moyenne et la covariance sont données par

$$m^{u|y}(x) = m(x) + \mathcal{L}[k(., x)]^T (\mathcal{L}k\mathcal{L} + \Sigma)^{-1} (y - (\mathcal{L}[m] + \mu))$$

$$k^{u|y} = k(x_1, x_2) + \mathcal{L}[k(., x_1)]^T (\mathcal{L}k\mathcal{L} + \Sigma)^{-1} \mathcal{L}[k(., x_2)]$$

où  $(\mathcal{L}k\mathcal{L})_{ij} := \mathcal{L}[x \mapsto \mathcal{L}[k(x, .)]_j]_i$

## 4.3 Convergence

Il est judicieux de prouver que notre solution approchée de l'EDP converge vers la solution réelle.

## 4.4 Equation de Poisson

## 4.5 Equation d'advection diffusion

# 5 Equation aux dérivées partielles non linéaires

Soit  $d \geq 1$  et  $\Omega$  un ouvert borné de  $\mathbb{R}^d$

Considérons l'équation non linéaire vérifiée par  $u^*$  :

$$\begin{aligned} \Delta u^*(x) + \tau u^*(x) &= f(x), & \forall x \in \Omega \\ u^*(x) &= g(x), & x \in \partial\Omega \end{aligned} \tag{1}$$

avec  $f : \Omega \rightarrow \mathbb{R}$ ,  $g : \partial\Omega \rightarrow \mathbb{R}$  et  $\tau : \mathbb{R} \rightarrow \mathbb{R}$  sont des fonctions continues données. On suppose que  $f, g$  et  $\tau$  sont choisis de manière à ce que l'EDP ait une solution unique.

On considère un noyau symétrique, défini positif  $K : \Omega \times \Omega \rightarrow \mathbb{R}$  avec  $\Omega := \Omega \cup \partial\Omega$ . Soit  $U$  le RKHS associé à  $K$  et  $\|\cdot\|$ , la norme associée au RKHS. Soit  $1 \leq M_\Omega \leq M \leq \infty$  et fixons  $M$  points dans  $\Omega$  tel que  $x_1, \dots, x_{M_\Omega} \in \Omega$  et  $x_{M_\Omega+1}, \dots, x_M \in \partial\Omega$ . La méthode suivante approche la solution  $u^*$  de (??) avec