

Université Toulouse III - Paul Sabatier
M2 MAPI³
ANNÉE UNIVERSITAIRE 2024/2025 - SEMESTRE 1

Rapport de projet - Plan d'expériences
*Modélisation par processus gaussiens avec contraintes
d'inégalités*

ÉTUDIANTS :

Nicolas BOURGEOIS – n° étudiant : 22001342
Mathieu-Anthony BRUISSON – n° étudiant : 22002569
Yan LARROUQUERE – n° étudiant : 22304525
Jean-Pierre MANSOUR – n° étudiant : 22212520
Paul PASQUET – n° étudiant : 22100069
Iyad ZEKHNINI – n° étudiant : 22011122



Table des matières

1	Cadre et idée de la méthode	3
2	Différents types de contraintes	4
2.1	Contraintes de bords	4
2.2	Contraintes sur les dérivées	5
2.3	Contraintes de convexité	5
2.3.1	Cas $D = 1$	6
2.3.2	Cas $D = 2$	6
2.3.3	Un mot sur le cas $D > 2$	6
2.4	Contraintes d'inégalité linéaires	7
3	Méthodes de calcul des espérances conditionnelles	7
3.1	Calcul direct	7
3.1.1	Calculs des matrices de covariances pour les vecteurs tronqués . . .	10
3.2	Techniques d'échantillonnage	12
3.2.1	Échantillonneur de Gibbs basé sur l'algorithme de Robert (1995) . .	12
3.2.2	Extension à la troncature elliptique	13
4	Résultats expérimentaux	13
4.1	Comportements des prédicteurs contraints sur des fonctions 1D	13
4.1.1	Évaluation des Prédicteurs Contraints	14
4.1.2	Impact des contraintes sur la précision (Première fonction Test) . .	14
4.1.3	Influence des contraintes sur le choix du noyau de covariance	14
4.1.4	Impact des contraintes sur la précision (Deuxième fonction Test) . .	15

Table des figures

1	Influence des contraintes sur la précision des prédicteurs	15
2	Variance des estimateurs	15
3	Contraintes de Bornes	16
4	Contraintes de Bornes&Dérivées	16
5	Contraintes de Bornes&Dérivées&Convexité	16

Liste des tableaux

Introduction

Soit $D \in \mathbb{N}^*$. Supposons que l'on ait un code complexe $f : \mathbb{R}^D \rightarrow \mathbb{R}$ avec des vecteurs d'entrée $\mathbf{x} \in \mathbb{R}^D$ et $y = f(\mathbf{x})$ la sortie associée. La fonction f est inconnue et l'évaluer peut s'avérer être extrêmement long, voire même impossible. On peut alors utiliser une approche différente, qui est l'analyse de sensibilité. De manière très générale, l'objectif de l'analyse de sensibilité est de hiérarchiser l'importance des composantes de \mathbf{x} sur les fluctuations de y . Dans ce cadre, on utilise des métamodèles. Les modèles gaussiens en sont un exemple très utilisé. L'objectif de ce projet est d'étudier l'article *Gaussian process modeling with inequality constraints* de DA VEIGA Sébastien et MARREL Amandine, qui propose de perfectionner ces modèles gaussiens, et ainsi d'approfondir ce qui a été vu en cours, en y ajoutant des connaissances supplémentaires sur les sorties, qui peuvent s'exprimer comme des inégalités, et ce, sous forme de contraintes. Ces connaissances peuvent, par exemple, être liées à la monotonie des sorties par rapport à certaines variables d'entrée, ou encore à leur convexité, etc. Une fois l'étude théorique terminée, nous présenterons quelques résultats issus de la mise en pratique de ce qui a été vu.

1 Cadre et idée de la méthode

Dans tout ce qui suit, on notera $X_s = (\mathbf{x}_1 \cdots \mathbf{x}_n)^T$, $Y_s = (y_1 \cdots y_n)^T$ et $F_s = (f(\mathbf{x}_1) \cdots f(\mathbf{x}_n))^T$. Dans le cadre d'un modèle gaussien, la réponse déterministe $y = f(\mathbf{x})$ est considérée comme une réalisation aléatoire $Y(\mathbf{x})$ décomposée de la manière suivante

$$Y(\mathbf{x}) = f_0(\mathbf{x}) + W(\mathbf{x}),$$

avec f_0 la fonction moyenne et W un champ gaussien stationnaire d'espérance 0 et de covariance $C(\tau) = \sigma^2 R(\tau)$, avec R une fonction de corrélation qui dépend uniquement de $\tau = \mathbf{x} - \mathbf{x}'$ et σ^2 la variance du champ en tout point.

Remarque 1 En pratique, on paramètre f_0 :

$$f_0(\mathbf{x}) = \sum_{j=1}^J \beta_j f_j(\mathbf{x}) = F(\mathbf{x})\beta,$$

avec $F(\mathbf{x}) = (f_1(\mathbf{x}) \cdots f_J(\mathbf{x}))$ connue et $\beta = (\beta_1 \cdots \beta_J)$ les paramètres de régression à estimer.

De plus, si on note R_ψ la fonction de corrélation (gaussienne, Matérn, etc.) avec ψ les hyperparamètres associés, alors on a les estimateurs suivants :

$$\hat{\beta} = (F_s R_\psi^{-1} F_s)^{-1} F_s^T R_\psi^{-1} Y_s,$$

$$\hat{\sigma}^2 = \frac{1}{n} \left(Y_s - F_s \hat{\beta} \right)^T R_\psi^{-1} \left(Y_s - F_s \hat{\beta} \right)$$

et

$$\psi^* = \operatorname{argmin}_\psi \hat{\sigma}^2 \det(R_\psi)^{1/n}.$$

Pour effectuer une prédiction en une nouvelle variable $\mathbf{x}^* \in \mathbb{R}^D$, on va s'intéresser au *champ conditionnel*

$$\tilde{Y}(\mathbf{x}^*) = Y(\mathbf{x}^*) | (Y(X_s) = Y_s)$$

Remarque 2 *Il s'agit d'un champ gaussien de moyenne*

$$\tilde{\mu}(\mathbf{x}^*) = F(\mathbf{x}^*)\hat{\beta} + k(\mathbf{x}^*)^T \Sigma_S^{-1} (Y_s - F_s \hat{\beta})$$

et de covariance

$$\tilde{C}(\mathbf{x}, \mathbf{x}') = C(\mathbf{x}, \mathbf{x}') - k(\mathbf{x})^T \Sigma_S^{-1} k(\mathbf{x}'),$$

avec $k(\xi)$ le vecteur de covariance entre ξ et X_s , C la covariance de Y et Σ_S la matrice de covariance de X_S .

Une méthode classique pour effectuer une prédiction en \mathbf{x}^* consiste à utiliser l'espérance $\tilde{\mu}(\mathbf{x}^*) = \mathbb{E} [\tilde{Y}(\mathbf{x}^*)]$ de ce champ conditionnel.

Nous arrivons alors au point clé de cette section : l'introduction des contraintes sur les sorties, dans le modèle. Afin de prendre en compte ces connaissances, on va simplement conditionner l'espérance $\tilde{\mu}(\mathbf{x}^*)$:

$$\mathbb{E} [\tilde{Y}(\mathbf{x}^*) \mid \text{contraintes}].$$

2 Différents types de contraintes

Dans cette section, nous allons donner plusieurs exemples de contraintes d'inégalité qu'il est possible de rencontrer. Les contraintes les plus rencontrées sont les contraintes de bords, détaillées dans la sous-section suivante.

2.1 Contraintes de bords

Les contraintes de bord consistent à borner les composantes du vecteur gaussien \tilde{Y} , inférieurement, supérieurement, ou bien les deux à la fois.

Un exemple concret

En chimie, la concentration d'un produit peut, par exemple, être comprise entre 0 et 1.

De manière générale, les contraintes de bord sur les N points $\mathbf{x}_1, \dots, \mathbf{x}_N \in \mathbb{R}^D$ peuvent s'écrire

$$a_i \leq \tilde{Y}(\mathbf{x}_i) \leq b_i, \forall i = 1, \dots, N,$$

où $a_i, b_i \in \mathbb{R}$, pour tout $i \in \{1, \dots, N\}$. Dans ce cas, le prédicteur avec contraintes s'écrit

$$\mathbb{E} [\tilde{Y}(\mathbf{x}^*) \mid \forall i \in \{1, \dots, N\}, a_i \leq \tilde{Y}(\mathbf{x}_i) \leq b_i].$$

2.2 Contraintes sur les dérivées

Dans cette sous-section, on suppose f suffisamment régulière, de sorte que ses dérivées partielles existent, au moins en certains points d'intérêt.

Les contraintes sur les dérivées permettent d'imposer une monotonie des sorties par rapport à certaines variables. Il s'agit donc de contraintes très utiles.

Un exemple concret

En physique, lorsque la vitesse d'un objet doit être croissante en fonction du temps.

Afin d'étudier de telles contraintes, nous allons noter, pour $i = 1, \dots, N$, $\mathbf{x}^i \in \mathbb{R}^D$ les points d'observations. Nous allons ensuite supposer, d'une part, que les prédictions augmentent par rapport à certaines variables x_j , pour $j \in \{1, \dots, D\}$. Les contraintes s'écrivent alors de la manière suivante :

$$\frac{\partial \tilde{Y}}{\partial x_j}(\mathbf{x}^i) \geq 0, \quad \forall i \in \{1, \dots, N\}, \forall j \in \{1, \dots, D\}$$

et on en déduit l'expression du prédicteur avec contraintes :

$$\mathbb{E} \left[\tilde{Y}(\mathbf{x}^*) \mid \forall i \in \{1, \dots, N\}, \frac{\partial \tilde{Y}}{\partial x_j}(\mathbf{x}^i) \geq 0 \right]$$

On peut également considérer que les prédictions diminuent par rapport aux variables x_j , auquel cas les contraintes deviennent

$$\frac{\partial \tilde{Y}}{\partial x_j}(\mathbf{x}^i) \leq 0, \quad \forall i \in \{1, \dots, N\}, \forall j \in \{1, \dots, D\}$$

et le prédicteur

$$\mathbb{E} \left[\tilde{Y}(\mathbf{x}^*) \mid \forall i \in \{1, \dots, N\}, \frac{\partial \tilde{Y}}{\partial x_j}(\mathbf{x}^i) \leq 0 \right].$$

2.3 Contraintes de convexité

Dans cette sous-section, on suppose f suffisamment régulière, de sorte que ses dérivées partielles d'ordre 2 existent, au moins en certains points d'intérêt.

Comme leur nom l'indique, ces contraintes ont pour objectif de prendre en compte la convexité ou bien de concavité en certains points. Pour ce type de contraintes, il est nécessaire de considérer la dimension D de l'espace dans lequel vivent les vecteurs d'entrée. On se limitera aux cas $D = 1$ et $D = 2$.

Un exemple concret

En économie, les fonctions de coût sont souvent supposées convexes, reflétant des rendements décroissants.

2.3.1 Cas $D = 1$

Dans le but de traiter le cas unidimensionnel, nous allons noter $\mathcal{E}_{\text{conv}} = \{x_1, \dots, x_n\}$ l'ensemble des points en lesquels la fonction doit être convexe. La convexité s'exprime par positivité de la dérivée seconde par rapport aux variables concernées, de sorte que le prédicteur avec contraintes s'écrive sous la forme

$$\mathbb{E} \left[\tilde{Y}(x^*) \mid \forall x_i \in \mathcal{E}_{\text{conv}}, \frac{\partial^2 \tilde{Y}}{\partial x^2}(x_i) \geq 0 \right]$$

On peut également considérer le cas de contraintes de concavité. Nous notons $\mathcal{E}_{\text{conc}} = \{x_1, \dots, x_n\}$ l'ensemble des points en lesquels la fonction doit être concave. Le prédicteur avec contraintes s'écrit cette fois-ci sous la forme

$$\mathbb{E} \left[\tilde{Y}(x^*) \mid \forall x_i \in \mathcal{E}_{\text{conc}}, \frac{\partial^2 \tilde{Y}}{\partial x^2}(x_i) \leq 0 \right]$$

2.3.2 Cas $D = 2$

De la même manière que pour traiter le cas unidimensionnel, le cas bidimensionnel nécessite de définir un espace de points en lesquels la fonction doit être convexe : nous le noterons ici $\mathcal{E}_{\text{conv}} = \{(x_1, x'_1), \dots, (x_n, x'_n)\}$. Dans ce cas, étudier simplement le signe d'une seule dérivée partielle n'a plus de sens, et il faut alors réécrire le prédicteur avec contraintes sous la forme

$$\mathbb{E} \left[\tilde{Y}(x^*) \mid \forall (x_i, x'_i) \in \mathcal{E}_{\text{conv}}, \nabla_{\mathbf{x}^* \mathbf{x}^*}^2 \tilde{Y}(x_i, x'_i) \text{ SDP} \right],$$

où $\nabla_{\mathbf{x}^* \mathbf{x}^*}^2$ désigne la matrice Hessienne de \tilde{Y} en x^* et SDP signifie "Semi-Définie Positive". Comme pour la dimension 1, nous pouvons considérer des contraintes de concavité en un ensemble de points $\mathcal{E}_{\text{conc}} = \{(x_1, x'_1), \dots, (x_n, x'_n)\}$, en écrivant le prédicteur comme suit :

$$\mathbb{E} \left[\tilde{Y}(x^*) \mid \forall (x_i, x'_i) \in \mathcal{E}_{\text{conc}}, \nabla_{\mathbf{x}^* \mathbf{x}^*}^2 \tilde{Y}(x_i, x'_i) \text{ SDN} \right],$$

où SDN signifie "Semi-Définie Négative".

Remarque 3 *Pour pouvoir prendre en compte de telles contraintes, il faut que :*

- *f soit deux fois différentiable ;*
- *et qu'il existe une fonction de covariance C telle que $\partial^4 C(\tau) / \partial \tau_j^4$ existe et soit finie pour tout $1 \leq j \leq D$.*

2.3.3 Un mot sur le cas $D > 2$

Les prédicteurs donnés dans toute cette sous-section **peuvent être étendus à des dimensions supérieures** : la matrice Hessienne en chaque point de contrainte doit respecter des conditions de convexité ou de concavité.

2.4 Contraintes d'inégalité linéaires

Nous allons traiter un dernier type de contraintes : les contraintes d'inégalité linéaires. On prend par exemple une contrainte de *conservation d'une quantité*, de la forme

$$\int_{\Omega} \tilde{Y}(\mathbf{x}) d\mathbf{x} \leq M,$$

avec M une constante réelle et Ω une partie quelconque de \mathbb{R}^D . Cependant, cette intégrale n'a pas de formulation analytique. Il est en revanche possible de l'approcher **linéairement** de la manière suivante :

$$\int_{\Omega} \tilde{Y}(\mathbf{x}) d\mathbf{x} \approx \sum_{i=1}^N w_i \tilde{Y}(\mathbf{x}_i).$$

On obtient par conséquent le prédicteur avec contraintes ci-dessous :

$$\mathbb{E} \left[\tilde{Y}(\mathbf{x}^*) \mid \sum_{i=1}^N w_i \tilde{Y}(\mathbf{x}_i) \leq M \right]$$

Un exemple concret

En logistique, la charge totale transportée par un véhicule ne doit pas dépasser une capacité maximale M .

3 Méthodes de calcul des espérances conditionnelles

Jusqu'à présent, nous avons vu comment introduire des contraintes d'inégalité à l'intérieur d'un prédicteur dans le cadre d'un modèle gaussien. Nous obtenons donc des espérances conditionnelles, et se pose alors le problème du calcul de telles quantités.

Dans cette section, nous allons présenter deux méthodes permettant de calculer ces prédicteurs avec contraintes. Nous nous limiterons en revanche au cas de contraintes de bords, décrites en 2.1. On considère, pour cela, un vecteur gaussien $\mathbf{Z} := (Z_1, \dots, Z_D) \sim \mathcal{N}_D(\mu, \Sigma)$ de densité $\phi_{\mu, \Sigma}(\mathbf{z})$.

3.1 Calcul direct

Remarque 4 Soient $n \in \mathbb{N}^*$, $\mathbf{u} = (u_1, \dots, u_n)$, $\mathbf{v} = (v_1, \dots, v_n) \in \mathbb{R}^n$.

Dans la suite, on notera $\mathbf{u} \leq \mathbf{v}$ pour signifier que

$$\forall i \in \{1, \dots, n\}, u_i \leq v_i.$$

On considère deux vecteurs $\mathbf{a}, \mathbf{b} \in \mathbb{R}^D$ qui contiennent les bornes des composantes de \mathbf{Z} . Les contraintes s'écrivent alors $\mathbf{a} \leq \mathbf{Z} \leq \mathbf{b}$. La densité de probabilité de \mathbf{Z} conditionnellement à la contrainte $(\mathbf{a} \leq \mathbf{Z} \leq \mathbf{b})$ est donnée par

$$\phi_{\mu, \Sigma, \mathbf{a}, \mathbf{b}} = \begin{cases} \frac{\phi_{\mu, \Sigma}(\mathbf{z})}{\mathbb{P}(\mathbf{a} \leq \mathbf{Z} \leq \mathbf{b})} & \text{si } \mathbf{a} \leq \mathbf{Z} \leq \mathbf{b}, \\ 0 & \text{sinon.} \end{cases}$$

La proposition suivante, que nous allons démontrer, donne une formule explicite du prédicteur $\mathbb{E}[\mathbf{Z} | \mathbf{a} \leq \mathbf{Z} \leq \mathbf{b}]$.

Proposition 1 *Pour tout $i \in \{1, \dots, D\}$, la i -ème composante du prédicteur avec contraintes est donnée par*

$$\mathbb{E}[Z_i | \mathbf{a} \leq \mathbf{Z} \leq \mathbf{b}] = \mu_i + \sum_{j=1}^D \sigma_{ij} (f_j(a_j) - f_j(b_j)),$$

avec

$$f_j(\mathbf{z}) = \int_{a_1}^{b_1} \cdots \int_{a_{j-1}}^{b_{j-1}} \int_{a_{j+1}}^{b_{j+1}} \cdots \int_{a_D}^{b_D} \phi_{\mu, \Sigma, \mathbf{a}, \mathbf{b}}(z_1, \dots, z_{j-1}, z, z_{j+1}, \dots, z_D) dz_1 \dots dz_{j-1} dz_{j+1} \dots dz_D$$

$$\text{et } \sigma_{ij} = (\Sigma)_{ij}.$$

Preuve. On suppose (sans perte de généralité) que $\mu = 0$. Ainsi, la densité de \mathbf{Z} devient :

$$\phi_{\Sigma}(z) = \frac{1}{(2\pi)^{D/2} |\Sigma|^{1/2}} \exp \left(-\frac{1}{2} \mathbf{z}^{\top} \Sigma^{-1} \mathbf{z} \right).$$

On note de plus $\alpha = \mathbb{P}(\mathbf{a} \leq \mathbf{Z} \leq \mathbf{b})$. Le moment générateur s'exprime alors sous la forme :

$$m(\mathbf{t}) = \mathbb{E}[\exp(\langle \mathbf{t}, \mathbf{Z} \rangle)] = \frac{\exp(\frac{1}{2} \mathbf{t}^{\top} \Sigma \mathbf{t})}{\alpha} \Phi(\mathbf{a} - \xi, \mathbf{b} - \xi; \Sigma),$$

où $\xi = \Sigma \mathbf{t}$ et $\Phi(\mathbf{u}, \mathbf{v}; \Sigma) = \int_{u_1}^{v_1} \cdots \int_{u_D}^{v_D} \phi_{\Sigma}(\mathbf{z}) d\mathbf{z}$.

L'espérance conditionnelle $\mathbb{E}[Z_i | \mathbf{a} \leq \mathbf{Z} \leq \mathbf{b}]$ est obtenue en dérivant $m(\mathbf{t})$ par rapport à t_i , puis en évaluant au point $\mathbf{t} = 0$:

$$\mathbb{E}[Z_i | \mathbf{a} \leq \mathbf{Z} \leq \mathbf{b}] = \left. \frac{\partial m(\mathbf{t})}{\partial t_i} \right|_{\mathbf{t}=0}.$$

Dérivons alors la fonction m :

$$\frac{\partial m}{\partial t_i}(\mathbf{t}) = \frac{1}{\alpha} \left[\frac{\partial}{\partial t_i} \exp \left(\frac{1}{2} \mathbf{t}^{\top} \Sigma \mathbf{t} \right) \right] \Phi(\mathbf{a} - \xi, \mathbf{b} - \xi; \Sigma) + \frac{1}{\alpha} \exp \left(\frac{1}{2} \mathbf{t}^{\top} \Sigma \mathbf{t} \right) \frac{\partial}{\partial t_i} \Phi(\mathbf{a} - \xi, \mathbf{b} - \xi; \Sigma)$$

Nous allons d'abord dériver la fonction g définie par $g(\mathbf{t}) = \exp(h(\mathbf{t}))$ où $h(\mathbf{t}) := \frac{1}{2} \mathbf{t}^{\top} \Sigma \mathbf{t}$. Par la formule de la dérivée d'une fonction composée, on a

$$\frac{\partial g}{\partial t_i} = \exp(h(\mathbf{t})) \frac{\partial h}{\partial t_i}.$$

Commençons par calculer $\partial_{t_i} h(\mathbf{t})$. On développe le produit matriciel :

$$h(\mathbf{t}) = \frac{1}{2} \sum_{j=1}^D \sum_{k=1}^D t_j \sigma_{jk} t_k$$

Il vient alors

$$\frac{\partial h}{\partial t_i} = \frac{1}{2} \sum_{j=1}^D \sum_{k=1}^D \sigma_{jk} (t_k \delta_{ji} + t_j \delta_{ki}).$$

En simplifiant, on obtient :

$$\frac{\partial h}{\partial t_i} = \frac{1}{2} \sum_{k=1}^D \sigma_{ik} t_k + \frac{1}{2} \sum_{j=1}^D t_j \sigma_{ji}.$$

Et, puisque Σ est symétrique ($\sigma_{ik} = \sigma_{ki}$), on peut regrouper les termes, ce qui donne

$$\frac{\partial h}{\partial t_i} = \sum_{j=1}^D \sigma_{ij} t_j.$$

On obtient ainsi l'expression de la dérivée partielle recherchée :

$$\frac{\partial g}{\partial t_i} = \exp\left(\frac{1}{2} \mathbf{t}^\top \Sigma \mathbf{t}\right) \sum_{j=1}^D \sigma_{ij} t_j.$$

Or, pour $\mathbf{t} = 0$, cette dérivée s'annule. On aboutit alors à

$$\begin{aligned} \mathbb{E}[Z_i \mid \mathbf{a} \leq \mathbf{Z} \leq \mathbf{b}] &= \left. \frac{\partial m(\mathbf{t})}{\partial t_i} \right|_{\mathbf{t}=0} \\ &= \frac{1}{\alpha} \exp\left(\frac{1}{2} \mathbf{t}^\top \Sigma \mathbf{t}\right) \left. \frac{\partial}{\partial t_i} \Phi(\mathbf{a} - \xi, \mathbf{b} - \xi; \Sigma) \right|_{\mathbf{t}=0} \\ &= \frac{1}{\alpha} \left. \frac{\partial}{\partial t_i} \Phi(\mathbf{a} - \xi, \mathbf{b} - \xi; \Sigma) \right|_{\mathbf{t}=0} \end{aligned}$$

Il reste donc à calculer $\partial_{t_i} \Phi(\mathbf{a} - \xi, \mathbf{b} - \xi; \Sigma)$. Par définition de Φ , on peut écrire

$$\frac{\partial}{\partial t_i} \Phi(\mathbf{a} - \xi, \mathbf{b} - \xi; \Sigma) = \frac{\partial}{\partial t_i} \left(\int_{a_1 - \xi_1}^{b_1 - \xi_1} \cdots \int_{a_D - \xi_D}^{b_D - \xi_D} \phi_\Sigma(z_1, \dots, z_D) dz_1 \cdots dz_D \right)$$

Par le théorème de Fubini, on a pour tous $u, v \in \mathbb{R}^D$,

$$\begin{aligned} \Phi(\mathbf{u}, \mathbf{v}, \Sigma) &= \int_{u_1}^{v_1} \cdots \int_{u_D}^{v_D} \phi_\Sigma(x_1, \dots, x_D) dx_1 \cdots dx_D \\ &= \int_{u_i}^{v_i} \left(\int_{u_1}^{v_1} \cdots \int_{u_{i-1}}^{v_{i-1}} \int_{u_{i+1}}^{v_{i+1}} \cdots \int_{u_D}^{v_D} \phi_\Sigma(x_1, \dots, x_D) dx_1 \cdots dx_{i-1} dx_{i+1} \cdots dx_D \right) dx_i \\ &= \int_{u_i}^{v_i} \mathbb{P}_i(x_i) dx_i \end{aligned}$$

Ainsi, en notant $\mathbf{a}^* = \mathbf{a} - \Sigma \mathbf{t}$ et $\mathbf{b}^* = \mathbf{b} - \Sigma \mathbf{t}$, il vient :

$$\Phi(\mathbf{a}^*, \mathbf{b}^*; \Sigma) = \int_{a_1^*}^{b_1^*} \cdots \int_{a_D^*}^{b_D^*} \phi_\Sigma(x_1, \dots, x_D) dx_1 \cdots dx_D.$$

Comme $\frac{d}{dt} \left[\int_{a(t)}^{b(t)} f(x, t) dx \right] = b'(t) f(b(t), t) - a'(t) f(a(t), t) + \int_{a(t)}^{b(t)} \partial_t f(x, t) dx$ et que

$$\begin{cases} \frac{d}{dt} a_j^* = \frac{d}{dt_i} \left[- \sum_{k=1}^D \sigma_{jk} t_k \right] = -\sigma_{ji} \\ \frac{d}{dt} b_j^* = \frac{d}{dt_i} \left[- \sum_{k=1}^D \sigma_{jk} t_k \right] = -\sigma_{ji} \end{cases},$$

alors

$$\begin{aligned}\frac{\partial \Phi}{\partial t_i}(\mathbf{a}^*, \mathbf{b}^*; \Sigma) &= -\sigma_{1i} \int_{a_2^*}^{b_2^*} \cdots \int_{a_D^*}^{b_D^*} \phi_\Sigma(a_1^*, x_2, \dots) dx_2 \cdots dx_D \\ &\quad + \sigma_{1i} \int_{a_2^*}^{b_2^*} \cdots \int_{a_D^*}^{b_D^*} \phi_\Sigma(b_1^*, x_2, \dots) dx_2 \cdots dx_D \\ &\quad + \int_{a_1^*}^{b_1^*} \frac{d}{dt_i} \left[\int_{a_2^*}^{b_2^*} \cdots \int_{a_D^*}^{b_D^*} \phi_\Sigma(x_1, \dots, x_D) dx_2 \cdots dx_D \right] dx_1\end{aligned}$$

De manière analogue, on peut montrer par récurrence que

$$\frac{\partial \Phi}{\partial t_i}(\mathbf{a}^*, \mathbf{b}^*; \Sigma) = \sum_{k=1}^D \sigma_{ik} (\mathbb{P}_k(a_k^*) - \mathbb{P}_k(b_k^*)).$$

Ainsi,

$$\left. \frac{\partial m(\mathbf{t})}{\partial t_i} \right|_{\mathbf{t}=0} = \frac{1}{\alpha} \left. \frac{\partial \Phi}{\partial t_i}(\mathbf{a}^*, \mathbf{b}^*; \Sigma) \right|_{\mathbf{t}=0} = \sum_{k=1}^D \sigma_{ik} \left(\frac{\mathbb{P}_k(a_k^*)}{\alpha} - \frac{\mathbb{P}_k(b_k^*)}{\alpha} \right).$$

Or, pour tout $x^* \in [a_k, b_k]$,

$$\begin{aligned}\frac{\mathbb{P}_k(x^*)}{\alpha} &= \int_{a_1}^{b_1} \cdots \int_{a_D}^{b_D} \frac{\phi_\Sigma(x_1, \dots, x^*, \dots, x_D)}{\alpha} dx_1 \cdots dx_D \\ &= \int_{a_1}^{b_1} \cdots \int_{a_D}^{b_D} \phi_{\Sigma, \mathbf{a}, \mathbf{b}}(x_1, \dots, x^*, \dots, x_D) dx_1 \cdots dx_D \\ &= f_k(x^*)\end{aligned}$$

On conclut :

$$\mathbb{E}[Z_i \mid \mathbf{a} \leq \mathbf{Z} \leq \mathbf{b}] = \sum_{k=1}^D \sigma_{ik} (f_k(a_k) - f_k(b_k)).$$

Il suffit ensuite de traduire par μ pour le cas non centré, ce qui achève la preuve. \square

Remarque 5 En dimension $D = 1$, cette formule s'écrit

$$\mathbb{E}[Z_1 | a_1 \leq Z_1 \leq b_1] = \mu_1 + \sigma_1 \frac{\phi(\tilde{a}_1) - \phi(\tilde{b}_1)}{\Phi(\tilde{b}_1) - \Phi(\tilde{a}_1)} \quad (1)$$

3.1.1 Calculs des matrices de covariances pour les vecteurs tronqués

D'un point de vue calculatoire, si toutes les variables ne sont pas tronquées il est possible de mettre des bornes infinies dans les intégrales et de se ramener à un calcul de moindre dimension. Une autre méthode consiste à réduire la complexité du problème en supposant, sans perte de généralité, que seules les variables Z_1, \dots, Z_k avec $k < D$ sont

tronquées. On suppose sans perte de généralité que $\mu = 0$ et note la matrice de covariance Σ de Z non tronquée comme suit :

$$\Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}$$

Ici Σ_{11} correspond à la matrice de covariance des variables non tronquées Z_1, \dots, Z_k , Σ_{22} correspond à la matrice de covariance des variables Z_{k+1}, \dots, Z_D et Σ_{12} et Σ_{21} sont les covariances croisés. le but de cette partie est de montrer la proposition suivante :

Proposition 2 *Si la moyenne ν_1 et la covariance Γ_{11} du vecteur tronqué (Z_1, \dots, Z_k) sont données, alors la moyenne ν_2 du vecteur non contraint (Z_{k+1}, \dots, Z_D) est donnée par :*

$$\nu_2 = \Sigma_{21} \Sigma_{11}^{-1} \nu_1 \quad (2)$$

et sa matrice de covariance Γ_{22} est donnée par :

$$\Gamma_{22} = \Sigma_{22} - \Sigma_{21}(\Sigma_{11}^{-1} - \Sigma_{11}^{-1} \Gamma_{11} \Sigma_{11}^{-1}) \Sigma_{12} \quad (3)$$

Preuve Montrons d'abord (2). Notons $w := (f_i(a_i) - f_i(b_i))_{1 \leq i \leq k}$. On remarque qu'on a $\nu_2 = \Sigma_{21} w$. Or comme $w = \Sigma_{11}^{-1} \Sigma_{11} w$ on peut écrire que $\nu_2 = \Sigma_{21} \Sigma_{11}^{-1} \Sigma_{11} w$. Or $\Sigma_{11} w = \nu_1$ donc en réinjectant dans ν_2 on a (2).

Montrons maintenant (3). Notons $Z_{k+1 \rightarrow D}$ le vecteur aléatoire de tailles $D - k$ dont les composantes sont Z_{k+1}, \dots, Z_D . On a :

$$\begin{aligned} \Gamma_{22} &= \mathbb{E}(\mathbb{E}(Z_{k+1 \rightarrow D} Z_{k+1 \rightarrow D}^T | a_1 \leq Z_1 \leq b_1, \dots, a_k \leq Z_k \leq b_k)) \\ &\quad - \mathbb{E}(\mathbb{E}(Z_{k+1 \rightarrow D} | a_1 \leq Z_1 \leq b_1, \dots, a_k \leq Z_k \leq b_k)) \mathbb{E}(\mathbb{E}(Z_{k+1 \rightarrow D} | a_1 \leq Z_1 \leq b_1, \dots, a_k \leq Z_k \leq b_k))^T \\ &= \mathbb{E}(Z_{k+1 \rightarrow D} Z_{k+1 \rightarrow D}^T) - \mathbb{E}(Z_{k+1 \rightarrow D}) \mathbb{E}(Z_{k+1 \rightarrow D})^T \end{aligned}$$

Comme on suppose que $\mu = 0$, le premier terme est égal à Σ_2 . Reste à calculer le second terme. On a par la formule (2) :

$$\begin{aligned} \mathbb{E}(Z_{k+1 \rightarrow D}) \mathbb{E}(Z_{k+1 \rightarrow D})^T &= \nu_2 \nu_2^T \\ &= \Sigma_{21} \Sigma_{11}^{-1} \nu_1 \nu_1^T \Sigma_{11} \Sigma_{21}^T \\ &= \Sigma_{21} \Sigma_{11}^{-1} \nu_1 \nu_1^T \Sigma_{11}^{-1} \Sigma_{12} \end{aligned}$$

Or on a :

$$\Gamma_{11} = \mathbb{E} \left(\begin{pmatrix} Z_1 \\ \vdots \\ Z_k \end{pmatrix} (Z_1 \dots Z_k) \right) - \nu_1 \nu_1^T$$

Comme $\mu = 0$, le premier terme de la somme ci-dessus vaut Σ_{11} , on en déduit donc que :

$$\nu_1 \nu_1^T = \Sigma_{11} - \Gamma_{11}$$

Ainsi en réinjectant dans $\mathbb{E}(Z_{k+1 \rightarrow D})\mathbb{E}(Z_{k+1 \rightarrow D})^T$ on a :

$$\begin{aligned}\mathbb{E}(Z_{k+1 \rightarrow D})\mathbb{E}(Z_{k+1 \rightarrow D})^T &= \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{11} - \Gamma_{11}\Sigma_{11}^{-1}\Sigma_{12} \\ &= \Sigma_{21}(\Sigma_{11}^{-1} - \Sigma_{11}^{-1}\Gamma_{11}\Sigma_{11}^{-1})\Sigma_{21}\end{aligned}$$

Ainsi on a :

$$\Gamma_{22} = \Sigma_{22} - \Sigma_{21}(\Sigma_{11}^{-1} - \Sigma_{11}^{-1}\Gamma_{11}\Sigma_{11}^{-1})\Sigma_{21}$$

Ce qui achève la preuve de (3) et de la proposition.

En pratique, si l'on se donne un processus Gaussien \hat{Y} contraint en N points x_1, \dots, x_N , pour prédire \hat{Y} en x^* on peut calculer la moyenne tronquée de $(\hat{Y}(x_1), \dots, \hat{Y}(x_N))$ et la formule 2 de la proposition ci-dessus permet de calculer la moyenne du nouveau vecteur à prédire. On peut également généraliser la proposition ci-dessus à d'autres types de contraintes telles que des contraintes linéaires.

3.2 Techniques d'échantillonnage

Une approche alternative pour approximer les espérances conditionnelles repose sur les techniques d'échantillonnage. En pratique, si nous sommes capables de générer des échantillons du vecteur tronqué $\mathbf{Z} \sim \mathcal{N}_D(\mu, \Sigma)$ sous les contraintes $\mathbf{a} \leq \mathbf{Z} \leq \mathbf{b}$, alors les moments empiriques obtenus à partir de ces échantillons permettent d'obtenir des approximations du prédicteur avec contraintes.

La simulation de variables gaussiennes tronquées repose principalement sur des méthodes de type MCMC (Monte Carlo par Chaînes de Markov), en particulier l'échantillonneur de Gibbs. Ce type d'algorithme génère des échantillons en mettant à jour séquentiellement chaque composante du vecteur \mathbf{Z} selon sa loi conditionnelle univariée.

Dans ce qui suit, nous détaillerons l'algorithme de Gibbs avec le rejet de Robert, que nous utiliserons dans nos simulations.

3.2.1 Échantillonneur de Gibbs basé sur l'algorithme de Robert (1995)

Un échantillonneur de Gibbs met à jour séquentiellement chaque composante d'un vecteur selon sa loi conditionnelle univariée. Ici, on considère une distribution normale standard $\mathcal{N}(0, 1)$ tronquée dans l'intervalle $[a, b]$. L'algorithme de rejet proposé est le suivant :

1. Générer un nombre uniforme z dans $[a, b]$.
2. Calculer :

$$g(z) = \begin{cases} \exp(-z^2/2), & \text{si } 0 \in [a, b] \\ \exp((b^2 - z^2)/2), & \text{si } b < 0 \\ \exp((a^2 - z^2)/2), & \text{si } 0 < a \end{cases}$$

3. Générer un nombre uniforme u dans $[0, 1]$ et accepter z si $u \leq g(z)$, sinon recommencer à l'étape 1.

Cette technique assure un taux d'acceptation élevé et est utilisée au sein d'un échantillonneur de Gibbs pour simuler des vecteurs gaussiens tronqués.

L'algorithme complet a une complexité de $\mathcal{O}(D^3 + Dq)$, où q est la taille de l'échantillon souhaité. Le terme $\mathcal{O}(D^3)$ correspond à la décomposition de Cholesky requise par le Gibbs sampler, tandis que les appels à l'échantillonneur univarié ont une complexité $\mathcal{O}(Dq)$.

3.2.2 Extension à la troncature elliptique

Dans le cas d'une troncature elliptique, l'échantillonneur de Gibbs peut toujours être utilisé car il s'adapte aux ensembles de troncature convexes. Cependant, une alternative plus efficace repose sur un rejet multidimensionnel.

Soit $\mathbf{X} \sim \mathcal{N}_D(0, I)$ et soit Y une variable indépendante de \mathbf{X} , suivant la loi :

$$f_Y(y) \propto \exp(-y/2)y^{\frac{D}{2}-1}\mathbf{1}_{(y>r)}.$$

Alors, la variable :

$$\mathbf{Z} = \sqrt{Y} \frac{\mathbf{X}}{\|\mathbf{X}\|}$$

suit une loi normale tronquée en dehors de la sphère de rayon r .

L'idée principale est d'exploiter cette propriété pour déterminer la plus grande sphère contenue dans l'ellipsoïde de troncature et la plus petite sphère englobante, puis d'utiliser un algorithme de rejet rapide.

4 Résultats expérimentaux

4.1 Comportements des prédictors contraints sur des fonctions 1D

Considérons la fonction

$$f_1(x) = \frac{\sin(10\pi x)}{10\pi x}, \quad x \in [0, 1]$$

Nous supposons que n observations $(x_i, y_i = f_1(x_i))_{i=1, \dots, n}$ sont disponibles, où les x_i sont échantillonnés selon une distribution uniforme sur $[0, 1]$. Ces observations sont utilisées pour construire le champ conditionnel $\tilde{Y}(x)$ et l'espérance associée :

$$\tilde{\mu}(x^*) = \mathbb{E}(\tilde{Y}(x^*))$$

introduite en section 2.1. La fonction moyenne $f_0(x)$ est supposée constante et les hyperparamètres sont estimés par maximum de vraisemblance.

De plus, on sait que f_1 est positif sur l'intervalle $[\frac{2m}{10}, \frac{2m+1}{10}]$, et négatif sur $[\frac{2m+1}{10}, \frac{2m+2}{10}]$ pour $m = 0, \dots, 4$. Par la suite, on considère le prédictor contraint

$$\mathbb{E}(\tilde{Y}(x^*) | \forall i = 1, \dots, N, a_i \leq \tilde{Y}(x_i) \leq b_i)$$

de l'équation (3.7), calculé en choisissant N points uniformément sur $[0, 1]$ avec des valeurs a_i et b_i définies selon les intervalles mentionnés. Son approximation est obtenue via Genz, Gibbs et la méthode sans corrélation.

4.1.1 Évaluation des Prédicteurs Contraints

Pour évaluer la précision des prédicteurs, on utilise le coefficient Q^2 :

$$Q^2(Y, \tilde{Y}) = 1 - \frac{\sum_{i=1}^{n_{\text{test}}} (Y_i - \tilde{Y}_i)^2}{\sum_{i=1}^{n_{\text{test}}} (Y_i - \bar{Y})^2}$$

où Y représente les observations réelles, \hat{Y} leur moyenne empirique et \tilde{Y} les valeurs prédites.

4.1.2 Impact des contraintes sur la précision (Première fonction Test)

Les tests avec $n = 10$ observations et $N = 20$ contraintes montrent :

- **Prédicteur non contraint** : ne reproduit pas bien f_1 sur la partie gauche, faute d'observations. Comportement similaire à droite, menant à un Q^2 faible (0.82).
- **Méthodes contraintes** :
 - **Approximation sans corrélation** : variations incorrectes aux extrémités, mais $Q^2 = 0.83$, légèrement supérieur au modèle non contraint.
 - **Genz et Gibbs** : reproduction extrêmement précise avec $Q^2 \approx 0.98$, et variance significativement réduite.

4.1.3 Influence des contraintes sur le choix du noyau de covariance

Pour valider ces résultats, on répète l'expérience 100 fois avec les noyaux gaussien et Matérn 3/2 :

- Le Matérn 3/2 offre de bonnes prédictions mais des trajectoires différentiables uniquement au premier ordre.
- Lorsque n est faible (10 ou 15), les contraintes améliorent la précision et réduisent l'écart-type.
- Pour n élevé, leur impact devient négligeable et $N = 20$ contraintes suffisent.
- La covariance gaussienne est plus performante que Matérn 3/2 dans ce cas.

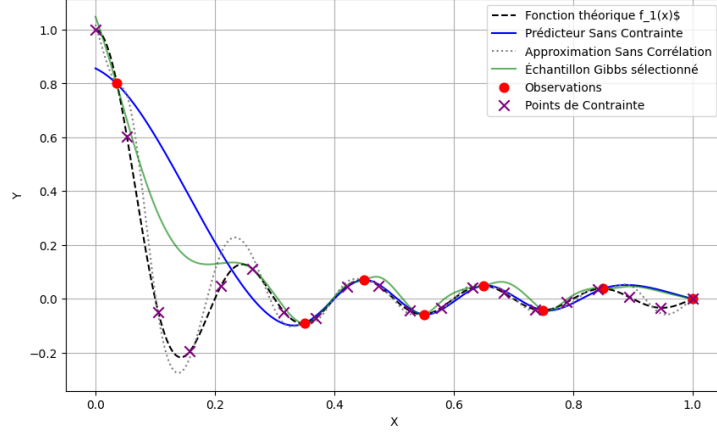


FIGURE 1 – Influence des contraintes sur la précision des prédicteurs

Et nous pouvons aussi voir que l'estimateur de Gibbs nous donne une variance très réduite par rapport à l'estimateur non-contraint

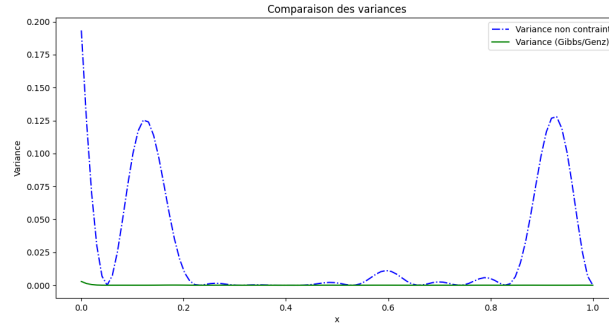


FIGURE 2 – Variance des estimateurs

Considérons maintenant la fonction f_2 définie par :

$$f_2(x) = \frac{\sin(10\pi x^{5/2})}{10\pi x}, \quad x \in [0, 1]$$

Cette fonction présente une fréquence variable rendant son approximation plus complexe. Les observations $(x_i, y_i = f_2(x_i))_{i=1, \dots, n}$ sont échantillonnées uniformément. Nous utilisons des contraintes de bornes et des contraintes sur les dérivées pour améliorer les prédictions, et nous appliquons les méthodes de Genz, Gibbs et la formule sans corrélation.

4.1.4 Impact des contraintes sur la précision (Deuxième fonction Test)

Nous testons les contraintes suivantes :

- **Bornes seules** : amélioration modérée, avec $Q^2 = 0.79$ (gaussienne) et 0.88 (Matérn 3/2).
- **Dérivées seules** : résultats inférieurs aux contraintes de bornes seules, avec $Q^2 = 0.80$ (gaussienne) et 0.77 (Matérn 3/2).
- **Bornes et dérivées** : amélioration nette, atteignant $Q^2 = 0.80$ (gaussienne) et 0.91 (Matérn 3/2).

- **Bornes, dérivées et convexité** : applicable uniquement avec la covariance gaussienne, avec un gain supplémentaire ($Q^2 = 0.85$).

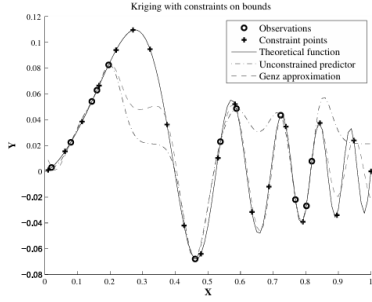


FIGURE 3 – Contraintes de Bornes

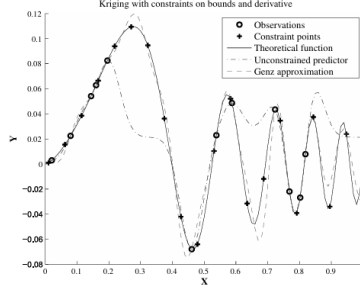


FIGURE 4 – Contraintes de Bornes&Dérivées

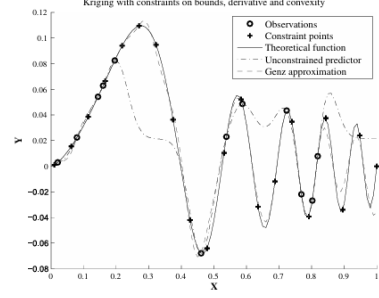


FIGURE 5 – Contraintes de Bornes&Dérivées&Convexité

Conclusion

Les éléments connus concernant les sorties d'un modèle gaussien peuvent être intégrés sous forme de contraintes dans le prédicteur. Cela permet d'obtenir des espérances conditionnelles, qu'il est nécessaire de calculer ou d'approcher par différentes méthodes. Nous avons présenté ici deux de ces méthodes et les résultats qu'elles génèrent : un calcul direct basé sur des formules bien établies, ainsi que l'utilisation de techniques d'échantillonnage.