

Méthodes de régression à noyau sous contrainte d'équations aux dérivées partielles

Iain Henderson

Thèse dirigée par Pascal Noble et Olivier Roustant

INSA Toulouse/IMT

28 Septembre 2023



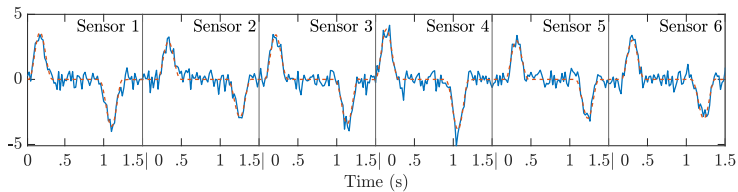
Plan de l'exposé

- 1 Contexte ML et EDP, notions et outils principaux
 - Contexte ML et EDP
 - Régression à noyau, processus gaussiens
 - Notion de solution d'une EDP
- 2 Contraintes physiques sur un processus gaussien
 - Contrainte d'EDP distributionnelle sur un processus aléatoire
 - Régularité Sobolev des processus gaussiens
- 3 Applications "numériques"
 - Covariances pour l'équation des ondes
 - GPR et schémas numériques
- 4 Ouvertures et conclusion

Plan de l'exposé

- 1 Contexte ML et EDP, notions et outils principaux
 - Contexte ML et EDP
 - Régression à noyau, processus gaussiens
 - Notion de solution d'une EDP
- 2 Contraintes physiques sur un processus gaussien
 - Contrainte d'EDP distributionnelle sur un processus aléatoire
 - Régularité Sobolev des processus gaussiens
- 3 Applications "numériques"
 - Covariances pour l'équation des ondes
 - GPR et schémas numériques
- 4 Ouvertures et conclusion

Estimer grâce aux données



Phénomène d'origine

Reconstruction

Formalisation du problème :

- X : paramètres d'entrée; Y : réponses; fonction u , correspondance entre observation $x \in X$ et résultat correspondant $u(x) \in Y$:

Formalisation du problème :

- X : paramètres d'entrée; Y : réponses; fonction u , correspondance entre observation $x \in X$ et résultat correspondant $u(x) \in Y$:

$$u : \begin{cases} X & \longrightarrow Y, \\ x & \longmapsto u(x). \end{cases}$$

Formalisation du problème :

- X : paramètres d'entrée; Y : réponses; fonction u , correspondance entre observation $x \in X$ et résultat correspondant $u(x) \in Y$:

$$u : \begin{cases} X & \longrightarrow Y, \\ x & \longmapsto u(x). \end{cases}$$

- Données :

$$\mathcal{B} = \{(x_1, y_1), \dots, (x_n, y_n)\}, \quad y_i = u(x_i) + \varepsilon_i.$$

Formalisation du problème :

- X : paramètres d'entrée; Y : réponses; fonction u , correspondance entre observation $x \in X$ et résultat correspondant $u(x) \in Y$:

$$u : \begin{cases} X & \longrightarrow Y, \\ x & \longmapsto u(x). \end{cases}$$

- Données :

$$\mathcal{B} = \{(x_1, y_1), \dots, (x_n, y_n)\}, \quad y_i = u(x_i) + \varepsilon_i.$$

A partir de \mathcal{B} , fabriquer $\hat{u} : X \rightarrow Y$, dans l'espoir que " $\hat{u} \simeq u$ ".

- Plusieurs méthodes existent : régression linéaire, régression par réseaux de neurones, régression à noyau...
- Point important : exploiter la structure des données.
→ Exploiter les modèles mathématiques préexistants si possible.
- Méthodes souvent flexibles, adaptées aux problèmes (inverses) mal posés.
- Méthodes qui s'appliquent à de plus en plus de domaines, "récemment" en physique [1].

1. RAISSI, M., PERDIKARIS, P., & KARNIADAKIS, G. (2019). Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *J. Comput. Phys.*, 378, 686-707.

Quelques problèmes (inverses) typiques en physique

- Modèles mathématiques puissants : équations aux dérivées partielles (EDPs).

$$\text{transport : } \begin{cases} \partial_t u + c \partial_x u = 0, \\ u|_{t=0} = u_0. \end{cases} \quad \text{chaleur : } \begin{cases} \partial_t u - D \partial_{xx} u = 0, \\ u|_{t=0} = u_0. \end{cases} \quad (1)$$

- Problèmes typiques : en notant $z = (x, t)$ et étant donné $\mathcal{B} = \{u(z_1), \dots, u(z_n)\}$,
→ Estimer u ? Estimer u_0 ? Estimer c, D ?

Quelques problèmes (inverses) typiques en physique

- Modèles mathématiques puissants : équations aux dérivées partielles (EDPs).

$$\text{transport : } \begin{cases} \partial_t u + c \partial_x u = 0, \\ u|_{t=0} = u_0. \end{cases} \quad \text{chaleur : } \begin{cases} \partial_t u - D \partial_{xx} u = 0, \\ u|_{t=0} = u_0. \end{cases} \quad (1)$$

- Problèmes typiques : en notant $z = (x, t)$ et étant donné $\mathcal{B} = \{u(z_1), \dots, u(z_n)\}$,
→ Estimer u ? Estimer u_0 ? Estimer c, D ?

Ce sont des problèmes d'approximation de fonctions → méthodes de ML ?

Machine learning et modèles physiques

- Combiner modèle et données
 - en pratique, **imposer des lois physiques** dans des méthodes de ML.

Machine learning et modèles physiques

- Combiner modèle et données
→ en pratique, **imposer des lois physiques** dans des méthodes de ML.

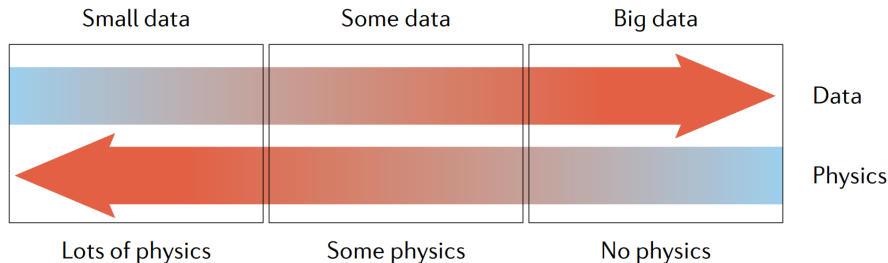
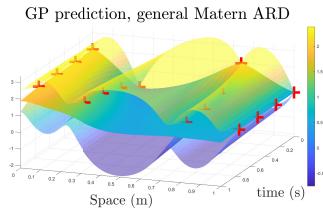
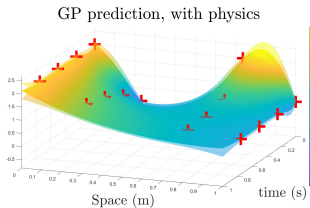
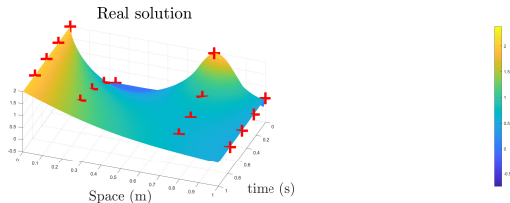


Figure 1 – KARNIADAKIS, G. E., KEVREKIDIS, I. G., LU, L., PERDIKARIS, P., WANG, S., & YANG, L. (2021). Physics-informed machine learning. *Nat. Rev. Phys.*, 3(6), 422-440

- Approches/outils mathématiques a priori **très différents**.

Pourquoi combiner physique et méthode de régression ?

chaleur 1D :
$$\begin{cases} \partial_t u - D \partial_{xx} u = 0, \\ u|_{t=0} = u_0, \quad u(0, t) = 2, \quad u(1, t) = 1/2. \end{cases}$$



Plan de l'exposé

- 1 Contexte ML et EDP, notions et outils principaux
 - Contexte ML et EDP
 - Régression à noyau, processus gaussiens
 - Notion de solution d'une EDP
- 2 Contraintes physiques sur un processus gaussien
 - Contrainte d'EDP distributionnelle sur un processus aléatoire
 - Régularité Sobolev des processus gaussiens
- 3 Applications "numériques"
 - Covariances pour l'équation des ondes
 - GPR et schémas numériques
- 4 Ouvertures et conclusion

Processus aléatoires, processus gaussiens

Soient $(\Omega, \mathcal{F}, \mathbb{P})$ un espace probabilisé et \mathcal{D} un ensemble.

Processus aléatoires, processus gaussiens

Soient $(\Omega, \mathcal{F}, \mathbb{P})$ un espace probabilisé et \mathcal{D} un ensemble.

- $(U(z))_{z \in \mathcal{D}}$ est un processus aléatoire si pour tout z , $U(z)$ est une v.a.r.

Processus aléatoires, processus gaussiens

Soient $(\Omega, \mathcal{F}, \mathbb{P})$ un espace probabilisé et \mathcal{D} un ensemble.

- $(U(z))_{z \in \mathcal{D}}$ est un processus aléatoire si pour tout z , $U(z)$ est une v.a.r.
- $(U(z))_{z \in \mathcal{D}}$ est un processus gaussien si pour tout (z_1, \dots, z_n) , $(U(z_1), \dots, U(z_n))^T \in \mathbb{R}^n$ est un vecteur aléatoire gaussien.

Processus aléatoires, processus gaussiens

Soient $(\Omega, \mathcal{F}, \mathbb{P})$ un espace probabilisé et \mathcal{D} un ensemble.

- $(U(z))_{z \in \mathcal{D}}$ est un processus aléatoire si pour tout z , $U(z)$ est une v.a.r.
- $(U(z))_{z \in \mathcal{D}}$ est un processus gaussien si pour tout (z_1, \dots, z_n) , $(U(z_1), \dots, U(z_n))^T \in \mathbb{R}^n$ est un vecteur aléatoire gaussien.
- La loi d'un processus gaussien est déterminée par

$$m(z) := \mathbb{E}[U(z)], \quad k(z, z') := \text{Cov}(U(z), U(z')).$$
$$(U(z))_{z \in \mathcal{D}} \sim PG(m, k)$$

La fonction k est définie positive : toutes les matrices $(k(z_i, z_j))_{1 \leq i, j \leq n}$ sont semi-définies positives. k est le noyau.

Régression par processus gaussiens (GPR)

- Fonction inconnue $z \in \mathcal{D} \mapsto u(z)$, obs. $\mathcal{B} = \{u(z_1), \dots, u(z_n)\}$

Régression par processus gaussiens (GPR)

- Fonction inconnue $z \in \mathcal{D} \mapsto u(z)$, obs. $\mathcal{B} = \{u(z_1), \dots, u(z_n)\}$
- On modélise $z \mapsto u(z)$ par une **trajectoire** d'un processus gaussien $(U(z))_{z \in \mathcal{D}} \sim PG(m, k)$ (*loi a priori*)

Régression par processus gaussiens (GPR)

- Fonction inconnue $z \in \mathcal{D} \mapsto u(z)$, obs. $\mathcal{B} = \{u(z_1), \dots, u(z_n)\}$
- On modélise $z \mapsto u(z)$ par une **trajectoire** d'un processus gaussien $(U(z))_{z \in \mathcal{D}} \sim PG(m, k)$ (*loi a priori*)
- On conditionne la loi de U aux observations \mathcal{B} :
 $V(z) = [U(z) | U(z_1) = u(z_1), \dots, U(z_n) = u(z_n)]$. On obtient

$$V(z) \sim PG(\tilde{m}, \tilde{k}) \quad (\text{loi a posteriori})$$

Régression par processus gaussiens (GPR)

- Fonction inconnue $z \in \mathcal{D} \mapsto u(z)$, obs. $\mathcal{B} = \{u(z_1), \dots, u(z_n)\}$
- On modélise $z \mapsto u(z)$ par une **trajectoire** d'un processus gaussien $(U(z))_{z \in \mathcal{D}} \sim PG(m, k)$ (*loi a priori*)
- On conditionne la loi de U aux observations \mathcal{B} :
 $V(z) = [U(z) | U(z_1) = u(z_1), \dots, U(z_n) = u(z_n)]$. On obtient

$$V(z) \sim PG(\tilde{m}, \tilde{k}) \quad (\text{loi a posteriori})$$

- Prédiction/estimation : $\forall z \in \mathcal{D}$, on prédit $u(z)$ par $\tilde{m}(z)$:
 $\hat{u}(z) = \tilde{m}(z) \simeq u(z)$, incertitude associée $\tilde{k}(z, z) = \text{Var}(V(z))$.

Exemple de GPR en 1D

- Exemple 1D : fonction coûteuse à évaluer, on ne dispose que de 7 valeurs connues de cette fonction

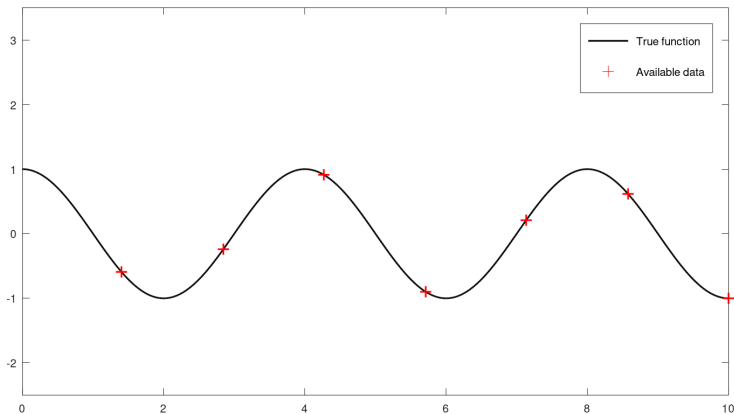


Figure 2 – Fonction à approcher par GPR

Exemple de GPR en 1D

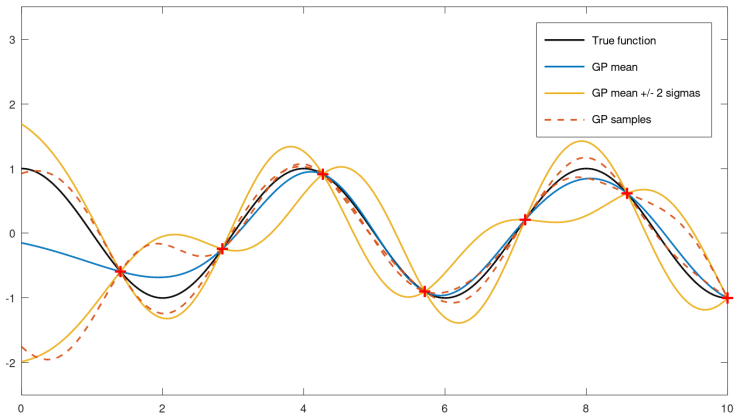


Figure 3 – Régression par processus gaussiens, 7 points d'observation

Plan de l'exposé

- 1 Contexte ML et EDP, notions et outils principaux
 - Contexte ML et EDP
 - Régression à noyau, processus gaussiens
 - Notion de solution d'une EDP
- 2 Contraintes physiques sur un processus gaussien
 - Contrainte d'EDP distributionnelle sur un processus aléatoire
 - Régularité Sobolev des processus gaussiens
- 3 Applications "numériques"
 - Covariances pour l'équation des ondes
 - GPR et schémas numériques
- 4 Ouvertures et conclusion

Transport d'une discontinuité :

$$\begin{cases} \partial_t u + c \partial_x u = 0, \\ u|_{t=0} = u_0. \end{cases} \quad (2)$$

Solution : $u(x, t) = u_0(x - ct)$... Sens de (2) si u_0 discontinue ($u_0 \in L^2$) ?

Solution distributionnelle d'une EDP

Transport d'une discontinuité :

$$\begin{cases} \partial_t u + c \partial_x u = 0, \\ u|_{t=0} = u_0. \end{cases} \quad (2)$$

Solution : $u(x, t) = u_0(x - ct)$... Sens de (2) si u_0 discontinue ($u_0 \in L^2$) ?

$$\begin{aligned} \forall \varphi \in C_c^\infty(\mathbb{R} \times \mathbb{R}_+^*), \quad 0 &= \int_{\mathbb{R} \times \mathbb{R}_+^*} \varphi(x, t) \left(\partial_t u(x, t) + c \partial_x u(x, t) \right) dx dt \\ &= - \int_{\mathbb{R} \times \mathbb{R}_+^*} \left(\partial_t \varphi(x, t) + c \partial_x \varphi(x, t) \right) u(x, t) dx dt. \end{aligned}$$

Solution distributionnelle d'une EDP

Transport d'une discontinuité :

$$\begin{cases} \partial_t u + c \partial_x u = 0, \\ u|_{t=0} = u_0. \end{cases} \quad (3)$$

Solution : $u(x, t) = u_0(x - ct)$... Sens de (3) si u_0 discontinue ?

$$\forall \varphi \in C_c^\infty(\mathbb{R} \times \mathbb{R}_+^*), \quad 0 = - \int_{\mathbb{R} \times \mathbb{R}_+^*} \left(\partial_t \varphi(x, t) + c \partial_x \varphi(x, t) \right) u(x, t) dx dt.$$

Solution distributionnelle d'une EDP

Transport d'une discontinuité :

$$\begin{cases} \partial_t u + c \partial_x u = 0, \\ u|_{t=0} = u_0. \end{cases} \quad (3)$$

Solution : $u(x, t) = u_0(x - ct)$... Sens de (3) si u_0 discontinue ?

$$\forall \varphi \in C_c^\infty(\mathbb{R} \times \mathbb{R}_+^*), \quad 0 = - \int_{\mathbb{R} \times \mathbb{R}_+^*} \left(\partial_t \varphi(x, t) + c \partial_x \varphi(x, t) \right) u(x, t) dx dt.$$

- Notion de dualité.
- Solution faible (Sobolev, 1934), distributions (Schwartz, 1946).

Premières conclusions et questions abordées

Moralité :

- Solution d'une EDP : notion non triviale.
- Utilité/nécessité de l'analyse fonctionnelle.
- Si possible, ne pas supposer la solution continue.

2. GINSBOURGER, D., ROUSTANT, O., & DURRANDE, N. (2016). On degeneracy and invariances of random fields paths with applications in Gaussian process modelling. *Journal of Statistical Planning and Inference*, 170 :117 -128.

Premières conclusions et questions abordées

Moralité :

- Solution d'une EDP : notion non triviale.
- Utilité/nécessité de l'analyse fonctionnelle.
- Si possible, ne pas supposer la solution continue.

Problème cible :

$$\partial_t u = L_\theta(u), \quad \mathcal{B} = \{u(z_1), \dots, u(z_n)\} \quad (4)$$

Imposer les contraintes physiques (EDP) sur le modèle statistique (GPR); identifier les [conséquences sur le noyau](#).

2. GINSBOURGER, D., ROUSTANT, O., & DURRANDE, N. (2016). On degeneracy and invariances of random fields paths with applications in Gaussian process modelling. *Journal of Statistical Planning and Inference*, 170 :117 -128.

Premières conclusions et questions abordées

Moralité :

- Solution d'une EDP : notion non triviale.
- Utilité/nécessité de l'analyse fonctionnelle.
- Si possible, ne pas supposer la solution continue.

Problème cible :

$$\partial_t u = L_\theta(u), \quad \mathcal{B} = \{u(z_1), \dots, u(z_n)\} \quad (4)$$

Imposer les contraintes physiques (EDP) sur le modèle statistique (GPR); identifier les **conséquences sur le noyau**.

Cas d'application :

- EDPs linéaires : $Lu = 0$ (invariance, voir [2])
- régularité Sobolev (énergie).

2. GINSBOURGER, D., ROUSTANT, O., & DURRANDE, N. (2016). On degeneracy and invariances of random fields paths with applications in Gaussian process modelling. *Journal of Statistical Planning and Inference*, 170 :117 -128.

Plan de l'exposé

- 1 Contexte ML et EDP, notions et outils principaux
 - Contexte ML et EDP
 - Régression à noyau, processus gaussiens
 - Notion de solution d'une EDP
- 2 Contraintes physiques sur un processus gaussien
 - Contrainte d'EDP distributionnelle sur un processus aléatoire
 - Régularité Sobolev des processus gaussiens
- 3 Applications "numériques"
 - Covariances pour l'équation des ondes
 - GPR et schémas numériques
- 4 Ouvertures et conclusion

Dérivées partielles en dimension d : notations

Soit $\alpha \in \mathbb{N}^d$ (d -uplet), $\alpha = (\alpha_1, \dots, \alpha_d)$. On note

$$\partial^\alpha := (\partial_{x_1})^{\alpha_1} \dots (\partial_{x_d})^{\alpha_d}. \quad (5)$$

Ordre de dérivation : $|\alpha| := \alpha_1 + \dots + \alpha_d$.

Opérateur différentiel linéaire d'ordre n :

$$L = \sum_{|\alpha| \leq n} a_\alpha \partial^\alpha. \quad (6)$$

Formulation distributionnelle d'une EDP linéaire

Soit un opérateur différentiel linéaire sur un ouvert $\mathcal{D} \subset \mathbb{R}^d$:

$$Lu = \sum_{|\alpha| \leq n} a_\alpha \partial^\alpha u.$$

Formulation distributionnelle d'une EDP linéaire

Soit un opérateur différentiel linéaire sur un ouvert $\mathcal{D} \subset \mathbb{R}^d$:

$$Lu = \sum_{|\alpha| \leq n} a_\alpha \partial^\alpha u.$$

Solution forte : $u \in C^n(\mathcal{D})$ et $(Lu)(x) = 0 \quad \forall x \in \mathcal{D}$.

Formulation distributionnelle d'une EDP linéaire

Soit un opérateur différentiel linéaire sur un ouvert $\mathcal{D} \subset \mathbb{R}^d$:

$$Lu = \sum_{|\alpha| \leq n} a_\alpha \partial^\alpha u.$$

Solution forte : $u \in C^n(\mathcal{D})$ et $(Lu)(x) = 0 \quad \forall x \in \mathcal{D}$.

Multiplication par une fonction test $\varphi \in C_c^\infty(\mathcal{D})$ et intégration sur \mathcal{D}

$$\forall \varphi \in C_c^\infty(\mathcal{D}), \int_{\mathcal{D}} Lu(x)\varphi(x)dx = 0. \quad (7)$$

Formulation distributionnelle d'une EDP linéaire

Soit un opérateur différentiel linéaire sur un ouvert $\mathcal{D} \subset \mathbb{R}^d$:

$$Lu = \sum_{|\alpha| \leq n} a_\alpha \partial^\alpha u.$$

Solution forte : $u \in C^n(\mathcal{D})$ et $(Lu)(x) = 0 \quad \forall x \in \mathcal{D}$.

Multiplication par une fonction test $\varphi \in C_c^\infty(\mathcal{D})$ et intégration sur \mathcal{D}

$$\forall \varphi \in C_c^\infty(\mathcal{D}), \int_{\mathcal{D}} Lu(x)\varphi(x)dx = 0. \quad (7)$$

Adjoint formel : $L^*v = \sum_{|\alpha| \leq n} (-1)^{|\alpha|} \partial^\alpha (a_\alpha v)$. Des IPP successives donnent les **solutions distributionnelles** :

$$\forall \varphi \in C_c^\infty(\mathcal{D}), \int_{\mathcal{D}} u(x)L^*\varphi(x)dx = 0. \quad (8)$$

Formulation distributionnelle d'une EDP linéaire

Soit un opérateur différentiel linéaire sur un ouvert $\mathcal{D} \subset \mathbb{R}^d$:

$$Lu = \sum_{|\alpha| \leq n} a_\alpha \partial^\alpha u.$$

Solution forte : $u \in C^n(\mathcal{D})$ et $(Lu)(x) = 0 \quad \forall x \in \mathcal{D}$.

Multiplication par une fonction test $\varphi \in C_c^\infty(\mathcal{D})$ et intégration sur \mathcal{D}

$$\forall \varphi \in C_c^\infty(\mathcal{D}), \int_{\mathcal{D}} Lu(x)\varphi(x)dx = 0. \quad (7)$$

Adjoint formel : $L^*v = \sum_{|\alpha| \leq n} (-1)^{|\alpha|} \partial^\alpha (a_\alpha v)$. Des IPP successives donnent les **solutions distributionnelles** :

$$\forall \varphi \in C_c^\infty(\mathcal{D}), \int_{\mathcal{D}} u(x)L^*\varphi(x)dx = 0. \quad (8)$$

Ne nécessite que $u \in L^1_{loc}(\mathcal{D})$, i.e. $\int_K |u| < +\infty$ pour tout $K \Subset \mathcal{D}$.

Processus aléatoire (gaussien) sous contrainte d'EDP [4]

Un P.A. est *mesurable* si l'application $(\omega, z) \mapsto U(z)(\omega)$ est mesurable.

3. GINSBOURGER, D., ROUSTANT, O., & DURRANDE, N. (2016). On degeneracy and invariances of random fields paths with applications in Gaussian process modelling. *Journal of Statistical Planning and Inference*, 170 :117 -128.

4. HENDERSON, I., NOBLE, P., & ROUSTANT, O. (2023). Characterization of the second order random fields subject to linear distributional PDE constraints. *Bernoulli*, 29(4), 3396-3422.

Processus aléatoire (gaussien) sous contrainte d'EDP [4]

Un P.A. est *mesurable* si l'application $(\omega, z) \mapsto U(z)(\omega)$ est mesurable.

Proposition 1

Soit $\mathcal{D} \subset \mathbb{R}^d$ un ouvert et $L := \sum_{|\alpha| \leq n} a_\alpha \partial^\alpha$, $a_\alpha \in C^{|\alpha|}(\mathcal{D})$. Soit $U = (U(z))_{z \in \mathcal{D}}$ un processus aléatoire *mesurable* centré d'ordre 2 de fonction de covariance k ; supposons que $\sigma : z \mapsto k(z, z)^{1/2} \in L^1_{loc}(\mathcal{D})$. Alors on a équivalence entre

- $\mathbb{P}(L(U) = 0 \text{ au sens des distributions}) = 1$
- $\forall z \in \mathcal{D}, L(k(z, \cdot)) = 0 \text{ au sens des distributions.}$

3. GINSBOURGER, D., ROUSTANT, O., & DURRANDE, N. (2016). On degeneracy and invariances of random fields paths with applications in Gaussian process modelling. *Journal of Statistical Planning and Inference*, 170 :117 -128.

4. HENDERSON, I., NOBLE, P., & ROUSTANT, O. (2023). Characterization of the second order random fields subject to linear distributional PDE constraints. *Bernoulli*, 29(4), 3396-3422.

Processus aléatoire (gaussien) sous contrainte d'EDP [4]

Un P.A. est *mesurable* si l'application $(\omega, z) \mapsto U(z)(\omega)$ est mesurable.

Proposition 1

Soit $\mathcal{D} \subset \mathbb{R}^d$ un ouvert et $L := \sum_{|\alpha| \leq n} a_\alpha \partial^\alpha$, $a_\alpha \in C^{|\alpha|}(\mathcal{D})$. Soit $U = (U(z))_{z \in \mathcal{D}}$ un processus aléatoire *mesurable* centré d'ordre 2 de fonction de covariance k ; supposons que $\sigma : z \mapsto k(z, z)^{1/2} \in L^1_{loc}(\mathcal{D})$. Alors on a équivalence entre

- $\mathbb{P}(L(U) = 0 \text{ au sens des distributions}) = 1$
- $\forall z \in \mathcal{D}, L(k(z, \cdot)) = 0 \text{ au sens des distributions.}$

Généralise un résultat de [3] aux contraintes distributionnelles d'EDP. Cette propriété s'hérite au processus conditionné par GPR.

3. GINSBOURGER, D., ROUSTANT, O., & DURRANDE, N. (2016). On degeneracy and invariances of random fields paths with applications in Gaussian process modelling. *Journal of Statistical Planning and Inference*, 170 :117 -128.

4. HENDERSON, I., NOBLE, P., & ROUSTANT, O. (2023). Characterization of the second order random fields subject to linear distributional PDE constraints. *Bernoulli*, 29(4), 3396-3422.

Exemples de noyaux vérifiant $L(k(z, \cdot)) = 0 \quad \forall z$

Etant donné L , trouver k_L tel que $L(k_L(\cdot, z)) = 0 \quad \forall z$; $\Delta = \sum_{i=1}^d \partial_{x_i}^2$.

Exemples de noyaux vérifiant $L(k(z, \cdot)) = 0 \quad \forall z$

Etant donné L , trouver k_L tel que $L(k_L(\cdot, z)) = 0 \quad \forall z$; $\Delta = \sum_{i=1}^d \partial_{x_i}^2$.

- **Laplace** : $\Delta u = 0$ (MENDES et da COSTA JÚNIOR, 2012), (GINSBOURGER et al., 2016)
- **Chaleur** : $\partial_t - D\Delta u = 0$ (ALBERT et RATH, 2020)
- **Div/Curl** : $\nabla \cdot u = 0, \nabla \times u = 0$ (SCHEUERER et SCHLATHER, 2012), (OWHADI, 2023b)
- **Mécanique des milieux continus** : (JIDLING et al., 2018)
- **Helmholtz** : $-\Delta u = \lambda u$ (ALBERT et RATH, 2020)
- **Maxwell (non)stationnaire** : (WAHLSTROM et al., 2013), (JIDLING et al., 2017), (LANGE-HEGERMANN, 2018)
- **Ondes 3D et transport** : (HENDERSON et al., 2023)
- Voir aussi les "forces latentes" : (ÁLVAREZ et al., 2009), (LÓPEZ-LOPERA et al., 2021)

Exemples de noyaux vérifiant $L(k(z, \cdot)) = 0 \quad \forall z$

Etant donné L , trouver k_L tel que $L(k_L(\cdot, z)) = 0 \quad \forall z$; $\Delta = \sum_{i=1}^d \partial_{x_i}^2$.

- **Laplace** : $\Delta u = 0$ (MENDES et da COSTA JÚNIOR, 2012), (GINSBOURGER et al., 2016)
- **Chaleur** : $\partial_t - D\Delta u = 0$ (ALBERT et RATH, 2020)
- **Div/Curl** : $\nabla \cdot u = 0, \nabla \times u = 0$ (SCHEUERER et SCHLATHER, 2012), (OWHADI, 2023b)
- **Mécanique des milieux continus** : (JIDLING et al., 2018)
- **Helmholtz** : $-\Delta u = \lambda u$ (ALBERT et RATH, 2020)
- **Maxwell (non)stationnaire** : (WAHLSTROM et al., 2013), (JIDLING et al., 2017), (LANGE-HEGERMANN, 2018)
- **Ondes 3D et transport** : (HENDERSON et al., 2023)
- Voir aussi les "forces latentes" : (ÁLVAREZ et al., 2009), (LÓPEZ-LOPERA et al., 2021)

Souvent basé sur des représentations de solutions de $Lu = 0$ de la forme

$$u = Gf.$$

Plan de l'exposé

- 1 Contexte ML et EDP, notions et outils principaux
 - Contexte ML et EDP
 - Régression à noyau, processus gaussiens
 - Notion de solution d'une EDP
- 2 Contraintes physiques sur un processus gaussien
 - Contrainte d'EDP distributionnelle sur un processus aléatoire
 - Régularité Sobolev des processus gaussiens
- 3 Applications "numériques"
 - Covariances pour l'équation des ondes
 - GPR et schémas numériques
- 4 Ouvertures et conclusion

- Fonctionnelles d'énergie (“interpretation physique”) :

- Fonctionnelles d'énergie (“interpretation physique”) :
 - Chaleur : $T(x, t), (x, t) \in \mathbb{R}^d \times \mathbb{R}_+, \partial_t T - \Delta T = 0$.

$$\frac{1}{2} \partial_t \|T(\cdot, t)\|_{L^2}^2 = -\|\nabla T(\cdot, t)\|_{L^2}^2 < 0 \text{ (diffusion)}. \quad (9)$$

- Fonctionnelles d'énergie (“interpretation physique”) :

- Chaleur : $T(x, t), (x, t) \in \mathbb{R}^d \times \mathbb{R}_+, \partial_t T - \Delta T = 0$.

$$\frac{1}{2} \partial_t \|T(\cdot, t)\|_{L^2}^2 = -\|\nabla T(\cdot, t)\|_{L^2}^2 < 0 \quad (\text{diffusion}). \quad (9)$$

- Ondes : $\partial_{tt}^2 u - \Delta u = 0$.

$$\partial_t \left(\|\partial_t u(\cdot, t)\|_{L^2}^2 + \|\nabla u(\cdot, t)\|_{L^2}^2 \right) = 0 \quad (\text{conservation}). \quad (10)$$

Advection : si $\partial_t u + \partial_x u = 0$, alors $\partial_t \|u(\cdot, t)\|_{L^p} = 0$.

- Fonctionnelles d'énergie (“interpretation physique”) :

- Chaleur : $T(x, t), (x, t) \in \mathbb{R}^d \times \mathbb{R}_+, \partial_t T - \Delta T = 0$.

$$\frac{1}{2} \partial_t \|T(\cdot, t)\|_{L^2}^2 = -\|\nabla T(\cdot, t)\|_{L^2}^2 < 0 \quad (\text{diffusion}). \quad (9)$$

- Ondes : $\partial_{tt}^2 u - \Delta u = 0$.

$$\partial_t \left(\|\partial_t u(\cdot, t)\|_{L^2}^2 + \|\nabla u(\cdot, t)\|_{L^2}^2 \right) = 0 \quad (\text{conservation}). \quad (10)$$

Advection : si $\partial_t u + \partial_x u = 0$, alors $\partial_t \|u(\cdot, t)\|_{L^p} = 0$.

→ Norme L^2 des dérivées.

Dérivées à énergie finie et espaces de Sobolev

Quelques fonctions sont "presque" différentiables : $h(x) = \max(0, 1 - |x|)$.

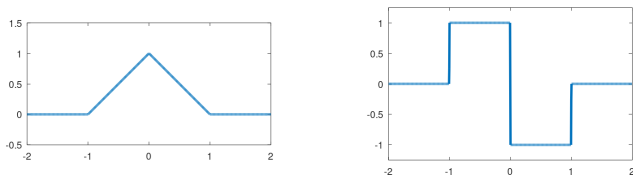


Figure 4 – Gauche : $h(x)$. Droite : $h'(x)$ (...).

Malheureusement, $h' \notin C^0$... mais $h' \in L^2$ (énergie finie)!

Dérivées à énergie finie et espaces de Sobolev

Quelques fonctions sont "presque" différentiables : $h(x) = \max(0, 1 - |x|)$.

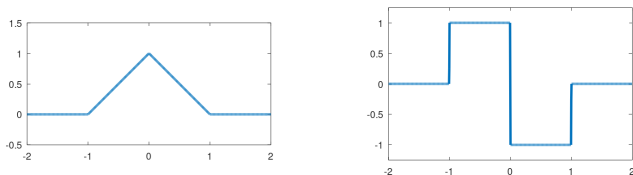


Figure 4 – Gauche : $h(x)$. Droite : $h'(x)$ (...).

Malheureusement, $h' \notin C^0$... mais $h' \in L^2$ (énergie finie)!

Une fonction g est la dérivée faible de h si pour tout $\varphi \in C_c^\infty(\mathbb{R})$,

$$\int_{\mathbb{R}} h(x) \varphi'(x) dx = - \int_{\mathbb{R}} g(x) \varphi(x) dx.$$

Dérivées à énergie finie et espaces de Sobolev

Quelques fonctions sont "presque" différentiables : $h(x) = \max(0, 1 - |x|)$.

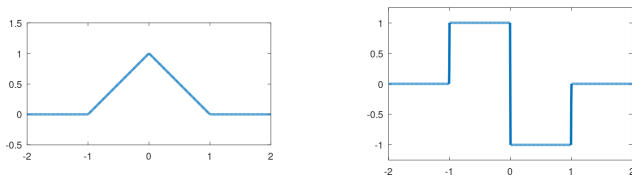


Figure 4 – Gauche : $h(x)$. Droite : $h'(x)$ (...).

Malheureusement, $h' \notin C^0$... mais $h' \in L^2$ (énergie finie)!

Une fonction g est la dérivée faible de h si pour tout $\varphi \in C_c^\infty(\mathbb{R})$,

$$\int_{\mathbb{R}} h(x) \varphi'(x) dx = - \int_{\mathbb{R}} g(x) \varphi(x) dx.$$

On définit ensuite

$$H^1(\mathbb{R}) := \{u \in L^2(\mathbb{R}) : u' \text{ existe au sens faible et } u' \in L^2(\mathbb{R})\},$$

$$H^m(\mathcal{D}) := \{u \in L^2(\mathcal{D}) : \forall |\alpha| \leq m, \partial^\alpha u \text{ existe ASF et } \partial^\alpha u \in L^2(\mathcal{D})\}.$$

Noyau de covariance et régularité du processus

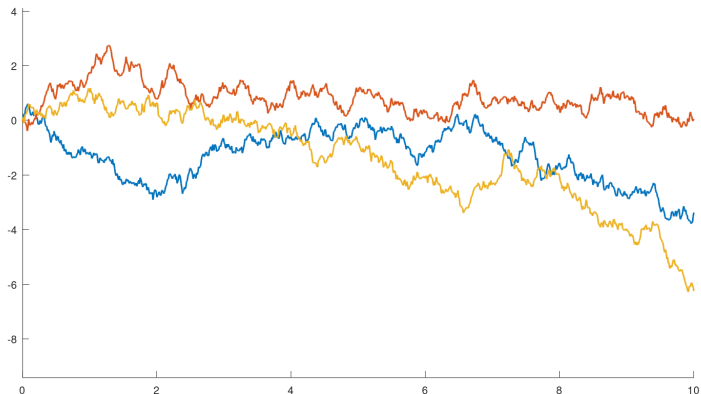


Figure 5 – Mouvement Brownien : $k(x, y) = \min(x, y)$.

Noyau de covariance et régularité du processus

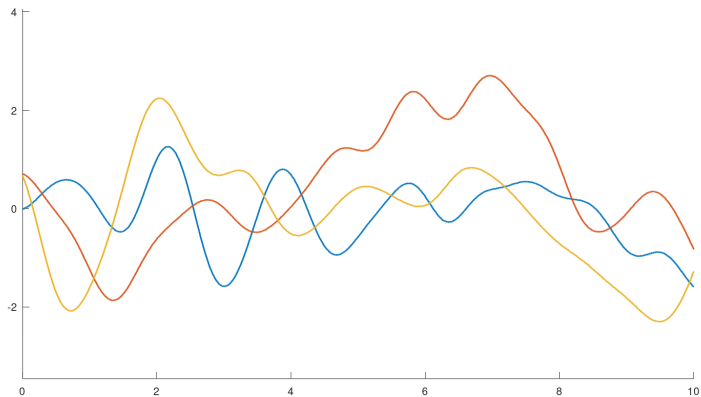


Figure 6 – Gaussien : $k(x, y) = \sigma^2 \exp(-|x - y|^2/2\ell^2)$, $\sigma = 1$, $\ell = 0.5$.

Régularité L^2 d'un processus gaussien [5]

- Critère **intégral** :

$$\mathbb{P}(U \in L^2(\mathcal{D})) = 1 \iff \int_{\mathcal{D}} k(x, x) dx < +\infty. \quad (11)$$

5. BOGACHEV, V. I. (1998). *Gaussian measures*. American Mathematical Soc.

Régularité L^2 d'un processus gaussien [5]

- Critère **intégral** :

$$\mathbb{P}(U \in L^2(\mathcal{D})) = 1 \iff \int_{\mathcal{D}} k(x, x) dx < +\infty. \quad (11)$$

Vient de $\mathbb{E}[\int U(x)^2 dx] = \int \mathbb{E}[U(x)^2] dx = \int k(x, x) dx$.

Régularité L^2 d'un processus gaussien [5]

- Critère **intégral** :

$$\mathbb{P}(U \in L^2(\mathcal{D})) = 1 \iff \int_{\mathcal{D}} k(x, x) dx < +\infty. \quad (11)$$

Vient de $\mathbb{E}[\int U(x)^2 dx] = \int \mathbb{E}[U(x)^2] dx = \int k(x, x) dx$.

- Critère **spectral/Mercer** : soit $\mathcal{E}_k : L^2(\mathcal{D}) \rightarrow L^2(\mathcal{D})$ l'opérateur

$$(\mathcal{E}_k f)(x) := \int k(x, y) f(y) dy. \quad (12)$$

5. BOGACHEV, V. I. (1998). *Gaussian measures*. American Mathematical Soc.

Régularité L^2 d'un processus gaussien [5]

- Critère **intégral** :

$$\mathbb{P}(U \in L^2(\mathcal{D})) = 1 \iff \int_{\mathcal{D}} k(x, x) dx < +\infty. \quad (11)$$

Vient de $\mathbb{E}[\int U(x)^2 dx] = \int \mathbb{E}[U(x)^2] dx = \int k(x, x) dx$.

- Critère **spectral/Mercer** : soit $\mathcal{E}_k : L^2(\mathcal{D}) \rightarrow L^2(\mathcal{D})$ l'opérateur

$$(\mathcal{E}_k f)(x) := \int k(x, y) f(y) dy. \quad (12)$$

Si $\int k(x, x) dx < +\infty$, alors (...) il existe $(\psi_n) \subset L^2$ t.q.

$$k(x, y) = \sum_{n=0}^{+\infty} \psi_n(x) \psi_n(y) \quad \text{dans } L^2(\mathcal{D} \times \mathcal{D}) \quad (\text{"Mercer"}). \quad (13)$$

Régularité L^2 d'un processus gaussien [7]

Donne (formellement)

$$\int k(x, x) dx = \int \sum_{n=0}^{+\infty} \psi_n(x)^2 dx = \sum_{n=0}^{+\infty} \int \psi_n(x)^2 dx \quad (14)$$

$$= \sum_{n=0}^{+\infty} \|\psi_n\|_2^2 = \sum_{n=0}^{+\infty} \lambda_n = \text{Tr}(\mathcal{E}_k) < +\infty \quad (\text{trace finie}). \quad (15)$$

6. DRISCOLL, M. F. (1973). The reproducing kernel Hilbert space structure of the sample paths of a Gaussian process. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete*, 26, 309-316.

7. BOGACHEV, V. I. (1998). *Gaussian measures*. American Mathematical Soc.

Régularité L^2 d'un processus gaussien [7]

Donne (formellement)

$$\int k(x, x) dx = \int \sum_{n=0}^{+\infty} \psi_n(x)^2 dx = \sum_{n=0}^{+\infty} \int \psi_n(x)^2 dx \quad (14)$$

$$= \sum_{n=0}^{+\infty} \|\psi_n\|_2^2 = \sum_{n=0}^{+\infty} \lambda_n = \text{Tr}(\mathcal{E}_k) < +\infty \quad (\text{trace finie}). \quad (15)$$

- **Injection du RKHS** : Si $\int k(x, x) dx < +\infty$, alors on a $\text{RKHS}(k) \subset L^2(\mathcal{D})$, et en notant \mathcal{I} le plongement associé, $\mathcal{I}\mathcal{I}^*(= \mathcal{E}_k)$ est à trace fini (“Driscoll” [6]).

6. DRISCOLL, M. F. (1973). The reproducing kernel Hilbert space structure of the sample paths of a Gaussian process. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete*, 26, 309-316.

7. BOGACHEV, V. I. (1998). *Gaussian measures*. American Mathematical Soc.

Régularité H^m d'un processus gaussien, $m \in \mathbb{N}$

Proposition 2 (H., 2022, "Sobolev regularity of GRFs")

Soit $(U(z))_{z \in \mathcal{D}} \sim PG(0, k)$ un PG *mesurable*, on a équivalence entre
(i) $\mathbb{P}(U \in H^m(\mathcal{D})) = 1$

Régularité H^m d'un processus gaussien, $m \in \mathbb{N}$

Proposition 2 (H., 2022, "Sobolev regularity of GRFs")

Soit $(U(z))_{z \in \mathcal{D}} \sim PG(0, k)$ un PG *mesurable*, on a équivalence entre

(i) $\mathbb{P}(U \in H^m(\mathcal{D})) = 1$

(ii) Pour tout $|\alpha| \leq m$, $\partial^{\alpha, \alpha} k \in L^2(\mathcal{D} \times \mathcal{D})$ et l'opérateur intégral \mathcal{E}_k^α

$$\mathcal{E}_k^\alpha : L^2(\mathcal{D}) \rightarrow L^2(\mathcal{D}), \quad \mathcal{E}_k^\alpha f(x) = \int_{\mathcal{D}} \partial^{\alpha, \alpha} k(x, y) f(y) dy,$$

est à trace finie, avec $Tr(\mathcal{E}_k^\alpha) = \int_{\mathcal{D}} \partial^{\alpha, \alpha} k(x, x) dx < +\infty$.

Régularité H^m d'un processus gaussien, $m \in \mathbb{N}$

Proposition 2 (H., 2022, "Sobolev regularity of GRFs")

Soit $(U(z))_{z \in \mathcal{D}} \sim PG(0, k)$ un PG *mesurable*, on a équivalence entre

(i) $\mathbb{P}(U \in H^m(\mathcal{D})) = 1$

(ii) Pour tout $|\alpha| \leq m$, $\partial^{\alpha, \alpha} k \in L^2(\mathcal{D} \times \mathcal{D})$ et l'opérateur intégral \mathcal{E}_k^α

$$\mathcal{E}_k^\alpha : L^2(\mathcal{D}) \rightarrow L^2(\mathcal{D}), \quad \mathcal{E}_k^\alpha f(x) = \int_{\mathcal{D}} \partial^{\alpha, \alpha} k(x, y) f(y) dy,$$

est à trace finie, avec $\text{Tr}(\mathcal{E}_k^\alpha) = \int_{\mathcal{D}} \partial^{\alpha, \alpha} k(x, x) dx < +\infty$.

(iii) Il existe $(\phi_n) \subset L^2(\mathcal{D})$ telle que $k(x, y) = \sum_n \phi_n(x) \phi_n(y)$ dans $L^2(\mathcal{D} \times \mathcal{D})$. De plus, si $|\alpha| \leq m$, alors $\phi_n \in H^m(\mathcal{D})$ et

$$\text{Tr}(\mathcal{E}_k^\alpha) = \sum_{n=0}^{+\infty} \|\partial^\alpha \phi_n\|_2^2 < +\infty.$$

Régularité H^m d'un processus gaussien, $m \in \mathbb{N}$

Proposition 2 (H., 2022, "Sobolev regularity of GRFs")

Soit $(U(z))_{z \in \mathcal{D}} \sim PG(0, k)$ un PG *mesurable*, on a équivalence entre

(i) $\mathbb{P}(U \in H^m(\mathcal{D})) = 1$

(ii) Pour tout $|\alpha| \leq m$, $\partial^{\alpha, \alpha} k \in L^2(\mathcal{D} \times \mathcal{D})$ et l'opérateur intégral \mathcal{E}_k^α

$$\mathcal{E}_k^\alpha : L^2(\mathcal{D}) \rightarrow L^2(\mathcal{D}), \quad \mathcal{E}_k^\alpha f(x) = \int_{\mathcal{D}} \partial^{\alpha, \alpha} k(x, y) f(y) dy,$$

est à trace finie, avec $\text{Tr}(\mathcal{E}_k^\alpha) = \int_{\mathcal{D}} \partial^{\alpha, \alpha} k(x, x) dx < +\infty$.

(iii) Il existe $(\phi_n) \subset L^2(\mathcal{D})$ telle que $k(x, y) = \sum_n \phi_n(x) \phi_n(y)$ dans $L^2(\mathcal{D} \times \mathcal{D})$. De plus, si $|\alpha| \leq m$, alors $\phi_n \in H^m(\mathcal{D})$ et

$$\text{Tr}(\mathcal{E}_k^\alpha) = \sum_{n=0}^{+\infty} \|\partial^\alpha \phi_n\|_2^2 < +\infty.$$

(iv) On a $\text{RKHS}(k) \subset H^m(\mathcal{D})$; en notant \mathcal{I} le plongement associé, on a $\text{Tr}(\mathcal{I}\mathcal{I}^*) = \sum_{|\alpha| \leq m} \text{Tr}(\mathcal{E}_k^\alpha) < +\infty$.

Plan de l'exposé

- 1 Contexte ML et EDP, notions et outils principaux
 - Contexte ML et EDP
 - Régression à noyau, processus gaussiens
 - Notion de solution d'une EDP
- 2 Contraintes physiques sur un processus gaussien
 - Contrainte d'EDP distributionnelle sur un processus aléatoire
 - Régularité Sobolev des processus gaussiens
- 3 Applications "numériques"
 - Covariances pour l'équation des ondes
 - GPR et schémas numériques
- 4 Ouvertures et conclusion

GPR et équation des ondes ([8])

Ondes en 3D homogène : $\Delta := \partial_{xx}^2 + \partial_{yy}^2 + \partial_{zz}^2$

$$\begin{cases} Lu &= \frac{1}{c^2} \partial_{tt}^2 u - \Delta u = \square u = 0, & (x, t) \in \mathbb{R}^3 \times \mathbb{R}^+ \\ u(x, 0) &= u_0(x), \quad \partial_t u(x, 0) = v_0(x). \end{cases} \quad (16)$$

8. H., I., NOBLE, P., & ROUSTANT, O. (to appear). Covariance models and Gaussian process regression for the wave equation. Application to related inverse problems. *J. Comput. Phys.*

GPR et équation des ondes ([8])

Ondes en 3D homogène : $\Delta := \partial_{xx}^2 + \partial_{yy}^2 + \partial_{zz}^2$

$$\begin{cases} Lu &= \frac{1}{c^2} \partial_{tt}^2 u - \Delta u = \square u = 0, & (x, t) \in \mathbb{R}^3 \times \mathbb{R}^+ \\ u(x, 0) &= u_0(x), \quad \partial_t u(x, 0) = v_0(x). \end{cases} \quad (16)$$

Représentation de u (Krichhoff) : $F_t = \sigma_{ct}/4\pi c^2 t$ et $\dot{F}_t = \partial_t F_t$

$$u(x, t) = (F_t * v_0)(x) + (\dot{F}_t * u_0)(x). \quad (17)$$

8. H., I., NOBLE, P., & ROUSTANT, O. (to appear). Covariance models and Gaussian process regression for the wave equation. Application to related inverse problems. *J. Comput. Phys.*

GPR et équation des ondes ([8])

Ondes en 3D homogène : $\Delta := \partial_{xx}^2 + \partial_{yy}^2 + \partial_{zz}^2$

$$\begin{cases} Lu &= \frac{1}{c^2} \partial_{tt}^2 u - \Delta u = \square u = 0, & (x, t) \in \mathbb{R}^3 \times \mathbb{R}^+ \\ u(x, 0) &= u_0(x), \quad \partial_t u(x, 0) = v_0(x). \end{cases} \quad (16)$$

Représentation de u (Krichhoff) : $F_t = \sigma_{ct}/4\pi c^2 t$ et $\dot{F}_t = \partial_t F_t$

$$u(x, t) = (F_t * v_0)(x) + (\dot{F}_t * u_0)(x). \quad (17)$$

On suppose u_0 et v_0 inconnues $\rightarrow u_0 \sim PG(0, k_u)$ et $v_0 \sim PG(0, k_v)$, indépendants. u donné par (17) est alors un PG centré de noyau

$$k((x, t), (x', t')) = [(F_t \otimes F_{t'}) * k_v](x, x') + [(\dot{F}_t \otimes \dot{F}_{t'}) * k_u](x, x'). \quad (18)$$

Le noyau k vérifie $\square k((x, t), \cdot) = 0$ pour tout $(x, t) \in \mathbb{R}^3 \times \mathbb{R}_+$.

8. H., I., NOBLE, P., & ROUSTANT, O. (to appear). Covariance models and Gaussian process regression for the wave equation. Application to related inverse problems. *J. Comput. Phys.*

Estimation de paramètres physiques et de conditions initiales

- Reconstruction de conditions initiales : la moyenne de GPR vérifie

□ $\tilde{m} = 0$. Donc

$$\tilde{m}(\cdot, t = 0) \simeq u_0, \quad \partial_t \tilde{m}(\cdot, t = 0) \simeq v_0$$

Estimation de paramètres physiques et de conditions initiales

- Reconstruction de conditions initiales : la moyenne de GPR vérifie $\square \tilde{m} = 0$. Donc

$$\tilde{m}(\cdot, t = 0) \simeq u_0, \quad \partial_t \tilde{m}(\cdot, t = 0) \simeq v_0$$

- Le noyau k est paramétré par c, θ_u et θ_v ; θ_u et θ_v peuvent contenir des informations physiques sur u_0 et v_0 .

Exemple : condition initiale à support compact donnent le modèle

$$k_u(x, x') = k_u^0(x, x') \mathbb{1}_{B(x_0, R)}(x) \mathbb{1}_{B(x_0, R)}(x'), \quad (19)$$

donne $(x_0, R) \in \theta_u$. De même pour v_0 (on peut aussi encoder des symétries).

→ estimation possible avec la [vraisemblance marginale](#).

Cadre restrictif

Convolutions coûteuses (4D) \rightarrow on se place dans le cadre de la symétrie radiale (convolutions explicites).

- Résolution numérique de l'équation des ondes dans $[0, 1]^3$, $v_0 = 0$ et

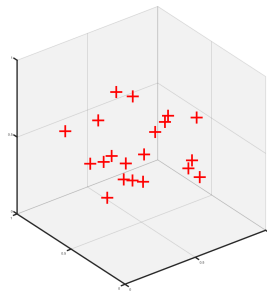
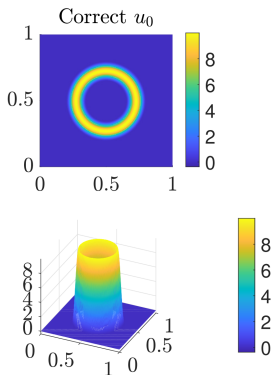


Figure 7 – Positions des capteurs

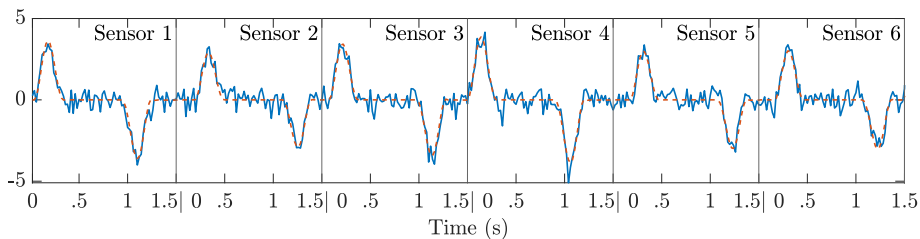


Figure 8 – Exemples de signaux capturés. Rouge : signal non bruité. Bleu : signal bruité.

Reconstruction de condition initiale et paramètres de position

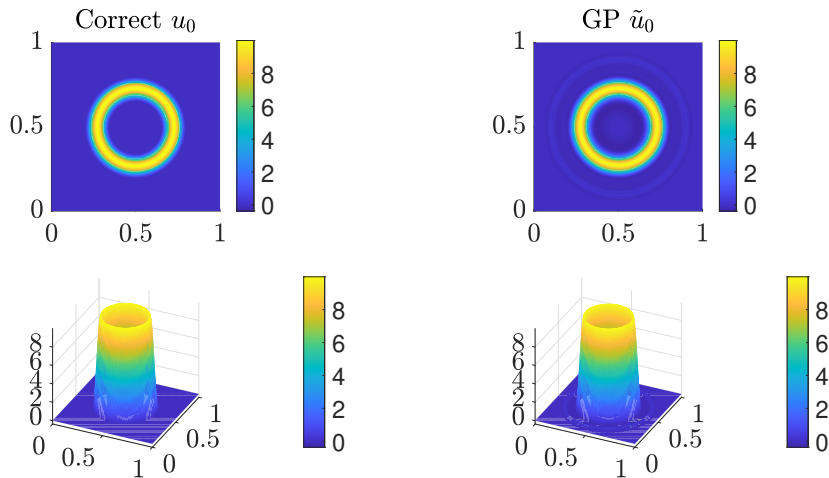


Figure 9 – Vraie u_0 (colonne de gauche) vs GPR u_0 (colonne de droite). 15 senseurs sont utilisés. Les images correspondent à des coupes 3D pour $z = 0.5$.

Localisation de source ponctuelle : $R \ll 1$

Cas où $u_0 \equiv 0$ et la source v_0 est quasi ponctuelle en x_0^* : on utilise les noyaux

$$k_v^R(x, x') = k_v(x, x') \frac{\mathbb{1}_{B(x_0, R)}(x)}{4\pi R^3/3} \frac{\mathbb{1}_{B(x_0, R)}(x')}{4\pi R^3/3}, \quad (20)$$

$$k((x, t), (x', t')) = [(F_t \otimes F_{t'}) * k_v^R](x, x'), \quad (21)$$

avec $R \ll 1$. Hyperparamètres de k : (θ_v, x_0, R, c) On fixe θ_v, R et c aux "bonnes valeurs" : vraisemblance marginale $\mathcal{L}(\theta) = \mathcal{L}(x_0)$, $x_0 \in \mathbb{R}^3$.

Question : comportement de $x_0 \mapsto \mathcal{L}(x_0)$?

Minimiser la vraisemblance marginale négative ≡ localisation GPS [9]

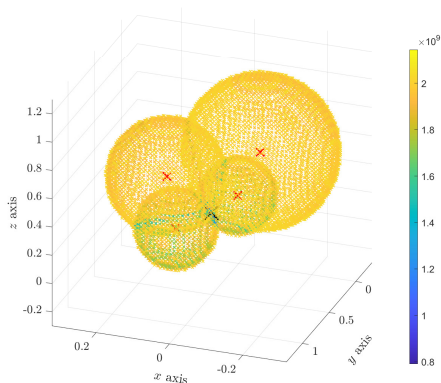


Figure : $x_0 \mapsto \mathcal{L}(x_0)$.

Valeurs affichées :
uniquement celles
 $\leq 2.035 \times 10^9$.

× : localisation des
capteurs.

× : localisation de la
source.

Etude + preuve :
article.

9. H., I., NOBLE, P., & ROUSTANT, O. (to appear). Covariance models and Gaussian process regression for the wave equation. Application to related inverse problems. *J. Comput. Phys.*

Plan de l'exposé

- 1 Contexte ML et EDP, notions et outils principaux
 - Contexte ML et EDP
 - Régression à noyau, processus gaussiens
 - Notion de solution d'une EDP
- 2 Contraintes physiques sur un processus gaussien
 - Contrainte d'EDP distributionnelle sur un processus aléatoire
 - Régularité Sobolev des processus gaussiens
- 3 Applications "numériques"
 - Covariances pour l'équation des ondes
 - GPR et schémas numériques
- 4 Ouvertures et conclusion

Problématique : résolution numérique d'une EDP. Advection linéaire :

$$\partial_t u + c \partial_x u = 0.$$

Grille $G = \{(i\Delta x, n\Delta t), i \in \mathbb{Z}, n \in \mathbb{N}\}$. Pour construire $u_i^n \simeq u(i\Delta x, n\Delta t)$, on peut résoudre

$$\frac{u_i^{n+1} - u_i^n}{\Delta t} + c \frac{u_{i+1}^n - u_i^n}{\Delta x} = 0,$$

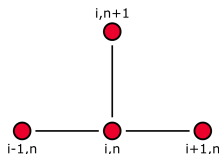
soit, avec $\lambda = c\Delta x/\Delta t$,

$$u_i^{n+1} = (1 + \lambda)u_i^n - \lambda u_{i+1}^n.$$

GPR sur une grille

- Fonction $u(x, t)$ à estimer sur une grille.
- A partir de $(u_{i-1}^n, u_i^n, u_{i+1}^n)$, estimations des valeurs de u en les points

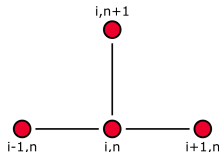
$((i-1)\Delta x, n\Delta t)$, $(i\Delta x, n\Delta t)$, $((i+1)\Delta x, n\Delta t)$,
estimer $u(i\Delta x, (n+1)\Delta t)$.



GPR sur une grille

- Fonction $u(x, t)$ à estimer sur une grille.
- A partir de $(u_{i-1}^n, u_i^n, u_{i+1}^n)$, estimations des valeurs de u en les points

$((i-1)\Delta x, n\Delta t)$, $(i\Delta x, n\Delta t)$, $((i+1)\Delta x, n\Delta t)$,
estimer $u(i\Delta x, (n+1)\Delta t)$.



La GPR est linéaire en les données :

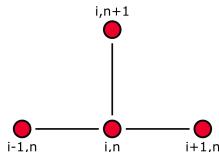
$$u_i^{n+1} = \tilde{m}(i\Delta x, (n+1)\Delta t) = a_{-1}u_{i-1}^n + a_0u_i^n + a_1u_{i+1}^n, \quad (22)$$

où $(a_{-1}, a_0, a_1)^T$ donné par les formules de GPR.

GPR sur une grille

- Fonction $u(x, t)$ à estimer sur une grille.
- A partir de $(u_{i-1}^n, u_i^n, u_{i+1}^n)$, estimations des valeurs de u en les points

$((i-1)\Delta x, n\Delta t), (i\Delta x, n\Delta t), ((i+1)\Delta x, n\Delta t),$
estimer $u(i\Delta x, (n+1)\Delta t)$.



La GPR est linéaire en les données :

$$u_i^{n+1} = \tilde{m}(i\Delta x, (n+1)\Delta t) = a_{-1}u_{i-1}^n + a_0u_i^n + a_1u_{i+1}^n, \quad (22)$$

où $(a_{-1}, a_0, a_1)^T$ donné par les formules de GPR.

L'équation (22) est-elle un "schéma numérique" ? Pour quelle EDP ? Quel noyau utiliser ? Propriétés ?

Proposition 3

Si $k((x, t), (x', t')) = k_0(x - ct, x' - ct')$, $k_0(h) = k_\nu(h/\ell)$, Matérn d'ordre $\nu > 0$, alors quand $\ell \rightarrow +\infty$, on obtient

- Si $\nu = 1/2$, *upwind* :

$$a_{-1} = (\lambda + |\lambda|)/2, \quad a_0 = 1 - |\lambda|, \quad a_1 = (-\lambda + |\lambda|)/2.$$

- Si $\nu \geq 2$, *Lax-Wendroff* :

$$a_{-1} = (\lambda + \lambda^2)/2, \quad a_0 = 1 - \lambda^2, \quad a_1 = (-\lambda + \lambda^2)/2.$$

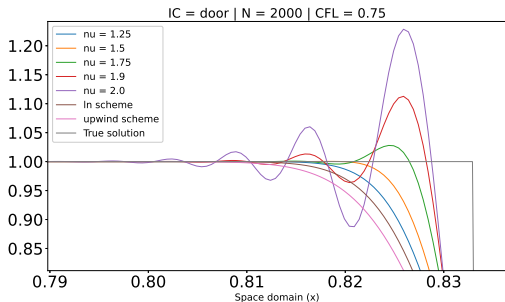
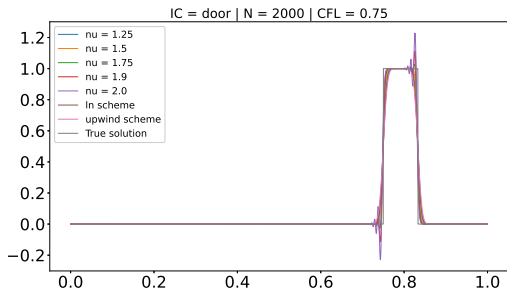
- Si $1 < \nu < 2$, *splines fractionnaires* :

$$a_{-1} = \frac{\lambda + D(\nu, \lambda)}{2}, \quad a_0 = 1 - D(\nu, \lambda), \quad a_1 = \frac{-\lambda + D(\nu, \lambda)}{2},$$

avec

$$D(\nu, \lambda) := \frac{|\lambda + 1|^{2\nu} + |\lambda - 1|^{2\nu} - 2|\lambda|^{2\nu} - 2}{4^\nu - 4}.$$

Propriétés qualitatives



Etude de convergence

$$E_{\nu,2}(\Delta x) := \left(\Delta x \sum_{j=1}^N |(u_{\text{num},\nu})_j - (u_T)_j|^2 \right)^{1/2} = \|u_{\text{num},\nu} - u_T\|_2.$$

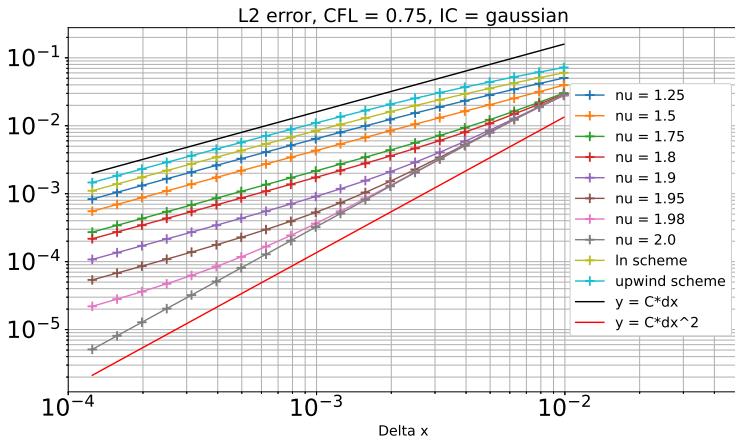


Figure 10 – Erreur ℓ^2 , $\Delta x \mapsto E_{\nu,2}(\Delta x)$

GPR, GPs sous contraintes physiques

- Contrainte d'EDP linéaire [10]
- Contrainte d'énergie : $H^m, W^{m,p}$ [11]

→ Des CNS sous des hypothèses minimales.

Applications “numériques” :

- Application à l'équation des ondes [12]
- Résultats préliminaires sur liens entre DF et GPR.

10. HENDERSON, I., NOBLE, P., & ROUSTANT, O. (2023). Characterization of the second order random fields subject to linear distributional PDE constraints. *Bernoulli*, 29(4), 3396-3422.

11. H., I. (2022). *Sobolev regularity of Gaussian random fields* [working paper or preprint].

12. H., I., NOBLE, P., & ROUSTANT, O. (to appear). Covariance models and Gaussian process regression for the wave equation. Application to related inverse problems. *J. Comput. Phys.*

- Formulations variationnelles \rightarrow Sobolev fractionnaires.
- Analyse d'erreur en norme Sobolev pour la GPR [13] [14].
- Ondes 3D : problématiques computationnelles.
- Sortir du cas gaussien ? Processus α -stables [15].

Mathématiques variées : EDP, probabilités, statistiques, analyse fonctionnelle, analyse numérique...

-
13. CHEN, Y., HOSSEINI, B., OWHADI, H., & STUART, A. M. (2021). Solving and learning nonlinear PDEs with Gaussian processes. *Journal of Computational Physics*, 447, 110668.
 14. BATLLE, P., CHEN, Y., HOSSEINI, B., OWHADI, H., & STUART, A. M. (2023). Error Analysis of Kernel/GP Methods for Nonlinear and Parametric PDEs.
 15. SULLIVAN, T. J. (2017). Well-posed Bayesian inverse problems and heavy-tailed stable quasi-Banach space priors. *Inverse Probl. Imaging*, 11(5), 857-874.

Merci de votre attention !

Contact : iain.pl.henderson@gmail.com

$$\begin{aligned} & [(F_t \otimes F_{t'}) * k_v](x, x') \\ &= \frac{\operatorname{sgn}(tt')}{16c^2 rr'} \sum_{\varepsilon, \varepsilon' \in \{-1, 1\}} \varepsilon \varepsilon' K_v((r + \varepsilon ct)^2, (r' + \varepsilon' c|t'|)^2) \end{aligned}$$

$$\begin{aligned} & [(\dot{F}_t \otimes \dot{F}_{t'}) * k_u](x, x') \\ &= \frac{1}{4rr'} \sum_{\varepsilon, \varepsilon' \in \{-1, 1\}} (r + \varepsilon ct)(r' + \varepsilon' c|t'|) k_u((r + \varepsilon ct)^2, (r' + \varepsilon' c|t'|)^2) \end{aligned}$$

Un équivalent déterministe : la version RKHS

Soit $k : D \times D \rightarrow \mathbb{R}$ une fonction définie positive. On définit H_k comme

$$H_k := \left\{ \sum_{i=1}^{+\infty} a_i k(z_i, \cdot) \text{ où } (a_i) \subset \mathbb{R}, (z_i) \subset D \text{ et } \sum_{i,j=1}^{+\infty} a_i a_j k(z_i, z_j) < +\infty \right\}$$

muni du produit scalaire

$$\left\langle \sum_{i=1}^{+\infty} a_i k(x_i, \cdot), \sum_{j=1}^{+\infty} b_j k(y_j, \cdot) \right\rangle := \sum_{i,j=1}^{+\infty} a_i b_j k(x_i, y_j)$$

La fonction k vérifie les propriétés de reproduction suivantes

$$\langle k(z, \cdot), k(z', \cdot) \rangle = k(z, z') \text{ et } \langle k(z, \cdot), f \rangle = f(z) \forall f \in H_k$$

Un équivalent déterministe : la version RKHS

Un RKHS est exactement un espace de Hilbert de fonctions $D \rightarrow \mathbb{R}$ tel que pour tout $z \in D$, la forme linéaire $f \mapsto f(z)$ est continue (ex : $H^{s+d/2}$, $s > 0$).

Soit maintenant le problème d'interpolation régularisé

$$\inf_{v \in \mathcal{H}_k} \|v\|_{\mathcal{H}_k} \quad \text{s.t.} \quad v(z_i) = u(z_i) \quad \forall i \in \{1, \dots, n\}$$

Alors $v = \tilde{m}$. De plus, $\tilde{m} = p_F(u)$ où $F := \text{Span}(k(z_1, \cdot), \dots, k(z_n, \cdot))$. De même, $\tilde{k}(z, \cdot) = P_{F^\perp}(k(z, \cdot))$. On a alors l'estimation

$$|u(z) - \tilde{m}(z)| \leq \tilde{k}(z, z)^{1/2} \|u\|_{H_k}$$

Différence avec la GPR : la GPR donne accès à une mesure de probabilité $\rightarrow \mathbb{P}(\sup_z |V_z| \leq \varepsilon)$, $\mathbb{P}(\|V\|_E \geq M)$... + la vraisemblance marginale.

N.B. : terrain d'échange entre deux communautés.

Espace de Hilbert à noyau reproduisant (RKHS)

Matrice $M \in \mathbb{R}^{d \times d}$ semi-définie positive

\longleftrightarrow produit scalaire sur \mathbb{R}^d , $\langle u, v \rangle_M = \langle u, Mv \rangle$

Fonction définie positive $k : \mathcal{D} \times \mathcal{D} \rightarrow \mathbb{R}$

$\longleftrightarrow H_k$, espace de Hilbert de fonctions $\mathcal{D} \rightarrow \mathbb{R}$

- $H_k = \overline{\text{Vect}(k(z, \cdot), z \in \mathcal{D})}$, $\| \sum_{i=1}^n a_i k(z_i, \cdot) \|_k^2 = \sum_{i,j=1}^n a_i a_j k(z_i, z_j)$.
- H_k vérifie $\langle k(z, \cdot), f \rangle_k = f(z)$, $\langle k(z, \cdot), k(z', \cdot) \rangle_k = k(z, z')$.
- RPG \equiv projection orthogonale de u dans H_k , le RKHS de k :

$$\tilde{m} = p_F(u), \quad F = \text{Vect}(k(z_1, \cdot), \dots, k(z_n, \cdot)).$$

Espace de Hilbert à noyau reproduisant (RKHS)

Matrice $M \in \mathbb{R}^{d \times d}$ semi-définie positive

\longleftrightarrow produit scalaire sur \mathbb{R}^d , $\langle u, v \rangle_M = \langle u, Mv \rangle$

Fonction définie positive $k : \mathcal{D} \times \mathcal{D} \rightarrow \mathbb{R}$

$\longleftrightarrow H_k$, espace de Hilbert de fonctions $\mathcal{D} \rightarrow \mathbb{R}$

- $H_k = \overline{\text{Vect}(k(z, \cdot), z \in \mathcal{D})}$, $\| \sum_{i=1}^n a_i k(z_i, \cdot) \|_k^2 = \sum_{i,j=1}^n a_i a_j k(z_i, z_j)$.
- H_k vérifie $\langle k(z, \cdot), f \rangle_k = f(z)$, $\langle k(z, \cdot), k(z', \cdot) \rangle_k = k(z, z')$.
- $\text{RPG} \equiv$ projection orthogonale de u dans H_k , le RKHS de k :

$$\tilde{m} = p_F(u), \quad F = \text{Vect}(k(z_1, \cdot), \dots, k(z_n, \cdot)).$$

- $H^{s+d/2}(\mathcal{D})$, $\mathcal{D} \subset \mathbb{R}^d$ est un RKHS.

Un mot sur la GPR et les réseaux de neurones

- Certains processus gaussiens comme limite de réseaux de neurones à une couche avec une infinité de neurones ([16], Section 4.2.3).
- La régression par réseaux de neurones comme une GPR avec un noyau appris des données ([17]; Mallat, collège de France).
- GPR : “seul” concurrent actuel aux réseaux de neurones (informés par la physique, PINNs), cf [18] pour une discussion.

16. RASMUSSEN, C. E., & WILLIAMS, C. (2006). *Gaussian Processes for Machine Learning*. The MIT Press.

17. OWHADI, H. (2023a). Do ideas have shape? Idea registration as the continuous limit of artificial neural networks. *Physica D: Nonlinear Phenomena*, 444, 133592.

18. CHEN, Y., HOSSEINI, B., OWHADI, H., & STUART, A. M. (2021). Solving and learning nonlinear PDEs with Gaussian processes. *Journal of Computational Physics*, 447, 110668.

Fonction de vraisemblance marginale

Souvent, noyau paramétré : $k = k_\theta, \theta \in \Theta \subset \mathbb{R}^p$.

$$k_{(\sigma^2, \ell)}(x, y) = \sigma^2 \exp(-|x - y|^2 / 2\ell^2)$$

$\theta = (\sigma^2, \ell) \in \mathbb{R}_+^* \times \mathbb{R}_+^*$. σ^2 : variance ; ℓ : longueur caractéristique.

- On peut considérer la (densité de) probabilité d'obtenir les données observées $u_{obs} = (u(z_1), \dots, u(z_n)) \in \mathbb{R}^n$ sachant une valeur de θ : en notant $K_{\theta ij} = k_\theta(z_i, z_j)$ alors

$$p(u_{obs} | \theta) = \frac{1}{(2\pi)^{n/2} \det K_\theta^{1/2}} e^{-\frac{1}{2} u_{obs}^T K_\theta^{-1} u_{obs}}$$

- On pose $\mathcal{L}(\theta) := -\log p(u_{obs} | \theta)$ et on cherche à résoudre

$$\theta^* = \arg \min_{\theta \in \Theta} \mathcal{L}(\theta)$$

Extension aux EDPs non linéaires

- Contraintes non linéaires sur $k(z, \cdot)$: pas réaliste (+ interprétation PG non valide).
- Alternative : dans [19], la contrainte d'EDP non linéaire est uniquement appliquée ponctuellement sur \tilde{m} : modification du problème d'optimisation dans le RKHS en

$$\inf_{v \in \mathcal{H}_k} \|v\|_{\mathcal{H}_k} \quad \text{s.t.} \quad \mathcal{N}(v(z_i), \nabla v(z_i), \dots) = \ell_i \quad \forall i \in \{1, \dots, n\}$$

Généralise une approche décrite dans [20].

- Couplage de cette approche avec des contraintes linéaires strictes : [21] (div/curl/périodicité).

-
19. CHEN, Y., HOSSEINI, B., OWHADI, H., & STUART, A. M. (2021). Solving and learning nonlinear PDEs with Gaussian processes. *Journal of Computational Physics*, 447, 110668.
 20. WENDLAND, H. (2004). *Scattered data approximation*. Cambridge university press.
 21. OWHADI, H. (2023b). Gaussian Process Hydrodynamics.

Proposition 4 (H., 2022)

Soit $(U(z))_{z \in \mathcal{D}} \sim GP(0, k)$ un PG mesurable, il y a équivalence entre

(i) $\mathbb{P}(U \in W^{m,p}(\mathcal{D})) = 1$

(ii) Pour tout $|\alpha| \leq m$, $\partial^{\alpha,\alpha} k \in L^p(\mathcal{D} \times \mathcal{D})$ et l'opérateur \mathcal{E}_k^α

$$\mathcal{E}_k^\alpha : L^q(\mathcal{D}) \rightarrow L^p(\mathcal{D}), \quad \mathcal{E}_k^\alpha f(x) = \int_{\mathcal{D}} \partial^{\alpha,\alpha} k(x, y) f(y) dy$$

est symétrique, positif et nucléaire : il existe $(\phi_n^\alpha) \subset L^p(\mathcal{D})$ telle que

$\partial^{\alpha,\alpha} k(x, y) = \sum_n \psi_n^\alpha(x) \psi_n^\alpha(y)$ dans $L^p(\mathcal{D} \times \mathcal{D})$ avec

$$\sum_{n=0}^{+\infty} \|\psi_n^\alpha\|_p^2 < +\infty \quad (+\text{raffinement si } 1 \leq p \leq 2)$$

(iii) Pour tout $|\alpha| \leq m$, $\partial^{\alpha,\alpha} k \in L^p(\mathcal{D} \times \mathcal{D})$, $\int_{\mathcal{D}} [\partial^{\alpha,\alpha} k(x, x)]^{p/2} dx < +\infty$.

Pourquoi les méthodes de régression à noyau ?

Parmi les méthodes de ML, focalisation sur les méthodes à noyau :

- interprétation bayésienne → processus gaussiens [22].
- interprétation RKHS → analyse fonctionnelle [23].
- “standard” si relativement peu de données [24].
- une façon d'étudier les réseaux de neurones [25].

22. RASMUSSEN, C. E., & WILLIAMS, C. (2006). *Gaussian Processes for Machine Learning*. The MIT Press.

23. WENDLAND, H. (2004). *Scattered data approximation*. Cambridge university press.

24. GRAMACY, R. B. (2020). *Surrogates: Gaussian process modeling, design, and optimization for the applied sciences*. CRC press.

25. BELKIN, M., MA, S., & MANDAL, S. (2018). To Understand Deep Learning We Need to Understand Kernel Learning. *35th ICML*, 541-549.

On note $u_{obs} = (u(z_1), \dots, u(z_n))$ les données, $K_{ij} := k(z_i, z_j)$ et $k(Z, z)_i := k(z_i, z)$. Alors la moyenne et la covariance a posteriori sont données par

$$\begin{cases} \tilde{m}(z) &= k(Z, z)^T K^{-1} u_{obs} \in \text{Span}(k(z_1, \cdot), \dots, k(z_n, \cdot)), \\ \tilde{k}(z, z') &= k(z, z') - k(Z, z)^T K^{-1} k(Z, z'). \end{cases}$$

- On remplace parfois K par $K + \lambda I$, $\lambda > 0$: régression ridge.

On note $u_{obs} = (u(z_1), \dots, u(z_n))$ les données, $K_{ij} := k(z_i, z_j)$ et $k(Z, z)_i := k(z_i, z)$. Alors la moyenne et la covariance a posteriori sont données par

$$\begin{cases} \tilde{m}(z) &= k(Z, z)^T K^{-1} u_{obs} \in \text{Span}(k(z_1, \cdot), \dots, k(z_n, \cdot)), \\ \tilde{k}(z, z') &= k(z, z') - k(Z, z)^T K^{-1} k(Z, z'). \end{cases}$$

- On remplace parfois K par $K + \lambda I$, $\lambda > 0$: régression ridge.
- Supposons que $Lu = 0$, L linéaire. k est adapté à cette contrainte si $L\tilde{m} = 0$, i.e. $Lk(z, \cdot) = 0$ pour tout z .
- $K \sim$ matrice de Gram, $k(Z, z) \sim$ vecteur de produits scalaires.

L'estimation bayésienne en une slide

- Quantité inconnue $u \in E$ ($E \subset \mathbb{R}^p$ par exemple) + base de données B
- On *modélise* u comme aléatoire : $u \sim \pi$, π mesure de proba sur E , c'est le prior. B est aussi vue aléatoire.
- On conditionne le prior sur B : $\pi_B = \pi(\cdot | B)$ est le posterior
- π_B permet d'estimer u par $\hat{u} = \mathbb{E}[u|B]$.
En général : contraction du posterior vers une masse de Dirac si B est suffisamment riche.

Les processus gaussiens permettent ce type d'approche lorsque u est une fonction. E est alors un espace de fonctions et les PG permettent de définir un prior sur E .