Physics-informed Gaussian process regression : theory and applications

Iain Henderson, Pascal Noble, Olivier Roustant

Institut de Mathématiques de Toulouse (IMT)
GdR MASCOT-NUM

December 4, 2023

Joint work with...

Figure 1 – Pascal Noble (IMT, PDEs)

Figure 2 – O. Roustant (IMT, ML/UQ)

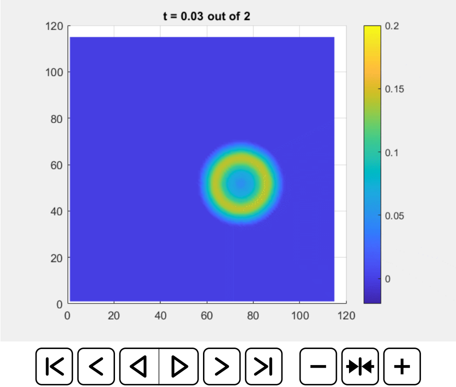
Outline of the talk

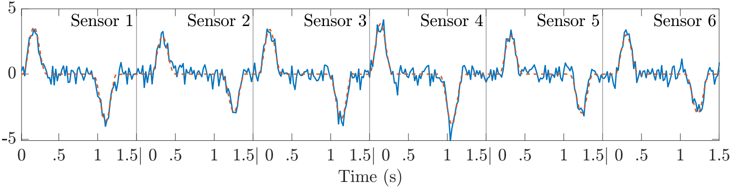
1. ML and PDEs : context, main notions and tools
   - Context : ML, PDEs
   - Kernel regression, Gaussian processes
   - Being a "solution" of a PDE
2. Imposing physical constraints on a GP
   - Distributional PDE constraints on a random process
   - Sobolev regularity of Gaussian processes
3. "Numerical" applications
   - Covariance kernels for the wave equation
   - GPR and numerical schemes
4. Conclusion and perspectives

Outline of the talk

1. ML and PDEs : context, main notions and tools
   - Context : ML, PDEs
   - Kernel regression, Gaussian processes
   - Being a "solution" of a PDE
2. Imposing physical constraints on a GP
   - Distributional PDE constraints on a random process
   - Sobolev regularity of Gaussian processes
3. "Numerical" applications
   - Covariance kernels for the wave equation
   - GPR and numerical schemes
4. Conclusion and perspectives

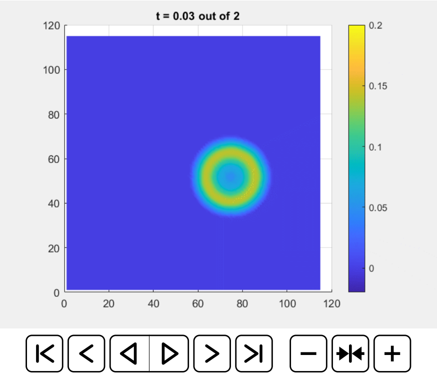
Estimation thanks to data



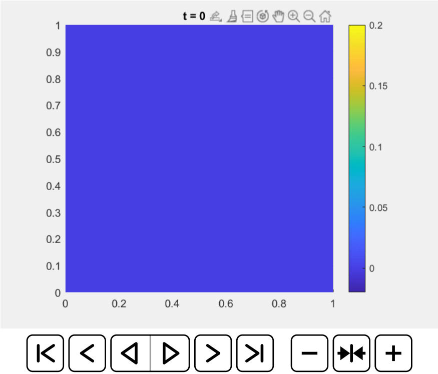


Estimation thanks to data

Original phenomenon



Reconstruction



Statistical learning and function approximation

Formalisation of the problem :

- X : input space ; Y : output space ; function u, maps $x \in X$ to corresponding output $u(x) \in Y$:

Statistical learning and function approximation

Formalisation of the problem :

- X : input space ; Y : output space ; function u, maps $x \in X$ to corresponding output $u(x) \in Y$:

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

Statistical learning and function approximation

Formalisation of the problem :

- X : input space ; Y : output space ; function u, maps $x \in X$ to corresponding output $u(x) \in Y$:

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

- Data :

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

Statistical learning and function approximation

Formalisation of the problem :

- X : input space ; Y : output space ; function u, maps $x \in X$ to corresponding output $u(x) \in Y$:

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

- Data :

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

From $\mathcal{B}$, construct $\hat{u} : X \to Y$, in the hope that "$\hat{u} \simeq u$".

Machine learning and function approximation

- Several methods exist : linear regression, neural network regression, kernel regression...
- Exploiting data structure is critical.
  $\longrightarrow$ Take advantage of mathematical models if available.
- Flexible methods, adapted to ill-posed (inverse) problems.
- More and more application fields, "recently" in physics [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.

Some typical (inverse) problems in physics

- Powerful mathematical models : partial differential equations (PDEs).

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

- Typical problems : denote $z = (x, t)$ and given $\mathcal{B} = \{u(z_1), ..., u(z_n)\}$,

  $\rightarrow$ Estimate u ? Estimate u_0 ? Estimate c, D ?

Some typical (inverse) problems in physics

- Powerful mathematical models : partial differential equations (PDEs).

$$\text{transport}: \begin{cases} \partial_t u + c\partial_x u = 0, \\ u_{|t=0} = u_0. \end{cases} \qquad \text{heat}: \begin{cases} \partial_t u - D\partial_{xx} u = 0, \\ u_{|t=0} = u_0. \end{cases} \tag{1}$$

- Typical problems : denote $z = (x, t)$ and given $\mathcal{B} = \{u(z_1), ..., u(z_n)\}$,

  $\rightarrow$ Estimate u ? Estimate u_0 ? Estimate c, D ?

These are function approximation problems $\rightarrow$ use ML methods ?

Machine learning and models in physics

- Combine model (PDEs) and data (ML, statistics...)
  $\rightarrow$ in practice, impose laws of physics in ML methods.

Machine learning and models in physics

- Combine model (PDEs) and data (ML, statistics...)
  $\rightarrow$ in practice, impose laws of physics in ML methods.

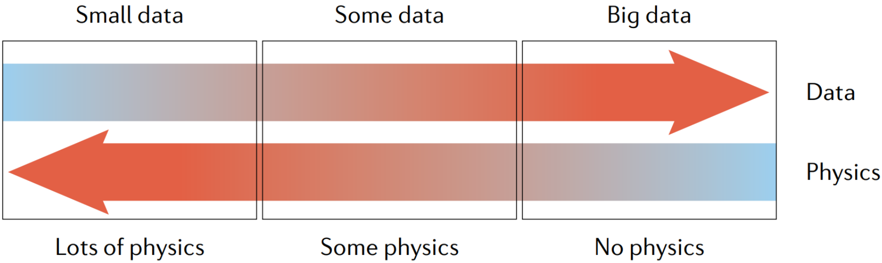
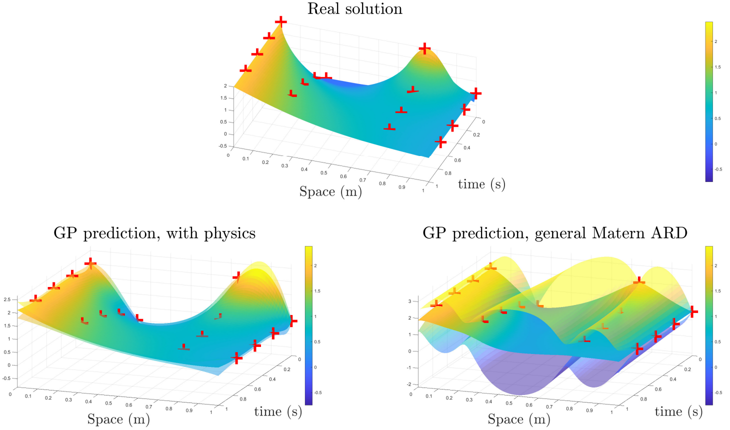


Figure 3 – Karniadakis, G. E., Kevrekidis, I. G., Lu, L., Perdikaris, P., Wang, S., & Yang, L. (2021). Physics-informed machine learning. *Nat. Rev. Phys.*, *3*(6), 422-440

- Approaches/mathematical tools are very different (a priori).

Why combine physics and ML methods ?

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



Outline of the talk

1. ML and PDEs : context, main notions and tools
   - Context : ML, PDEs
   - Kernel regression, Gaussian processes
   - Being a “solution” of a PDE
2. Imposing physical constraints on a GP
   - Distributional PDE constraints on a random process
   - Sobolev regularity of Gaussian processes
3. “Numerical” applications
   - Covariance kernels for the wave equation
   - GPR and numerical schemes
4. Conclusion and perspectives

Random processes, Gaussian processes

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and $\mathcal{D}$ a set.

Random processes, Gaussian processes

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and $\mathcal{D}$ a set.

- $(U(z))_{z \in \mathcal{D}}$ is a random process if for all z, $U(z)$ is a (real-valued) random variable.

Random processes, Gaussian processes

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and $\mathcal{D}$ a set.

- $(U(z))_{z \in \mathcal{D}}$ is a random process if for all z, $U(z)$ is a (real-valued) random variable.

- $(U(z))_{z \in \mathcal{D}}$ is a Gaussian process if for all $(z_1, ..., z_n)$, $(U(z_1), ..., U(z_n))^T \in \mathbb{R}^n$ is a Gaussian random vector.

Random processes, Gaussian processes

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and $\mathcal{D}$ a set.

- $(U(z))_{z\in\mathcal{D}}$ is a random process if for all z, $U(z)$ is a (real-valued) random variable.

- $(U(z))_{z\in\mathcal{D}}$ is a Gaussian process if for all $(z_1, ..., z_n)$, $(U(z_1), ..., U(z_n))^T \in \mathbb{R}^n$ is a Gaussian random vector.
- The law of a Gaussian process is determined by

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

The function k is positive definite : all matrices of the form $(k(z_i, z_j))_{1\leq i,j\leq n}$ are symmetric positive semi-definite. k is the kernel.

Gaussian process regression (GPR)

- Unknown function $z \in \mathcal{D} \longmapsto u(z)$, obs. $\mathcal{B} = \{u(z_1), ..., u(z_n)\}$

Gaussian process regression (GPR)

- Unknown function $z \in \mathcal{D} \longmapsto u(z)$, obs. $\mathcal{B} = \{u(z_1), ..., u(z_n)\}$
- Model $z \longmapsto u(z)$ as a realization of a Gaussian process $(U(z))_{z \in \mathcal{D}} \sim GP(m, k)$ (*prior* distribution)

Gaussian process regression (GPR)

- Unknown function $z \in \mathcal{D} \longmapsto u(z)$, obs. $\mathcal{B} = \{u(z_1), ..., u(z_n)\}$
- Model $z \longmapsto u(z)$ as a realization of a Gaussian process $(U(z))_{z \in \mathcal{D}} \sim GP(m, k)$ (*prior* distribution)
- Condition the distribution of U on obs. $\mathcal{B}$: "$V(z) = [U(z) | U(z_1) = u(z_1), ..., U(z_n) = u(z_n)]$". We obtain

$$V(z) \sim GP(\tilde{m}, \tilde{k}) \quad \text{(\textit{posterior} distribution)}$$

Gaussian process regression (GPR)

- Unknown function $z \in \mathcal{D} \longmapsto u(z)$, obs. $\mathcal{B} = \{u(z_1), ..., u(z_n)\}$
- Model $z \longmapsto u(z)$ as a realization of a Gaussian process $(U(z))_{z \in \mathcal{D}} \sim GP(m, k)$ (*prior* distribution)
- Condition the distribution of U on obs. $\mathcal{B}$: "$V(z) = [U(z) | U(z_1) = u(z_1), ..., U(z_n) = u(z_n)]$". We obtain

$$V(z) \sim GP(\tilde{m}, \tilde{k}) \quad \text{(\textit{posterior} distribution)}$$

- Prediction/estimation : $\forall z \in \mathcal{D}$, we estimate $u(z)$ with $\tilde{m}(z)$: $\hat{u}(z) = \tilde{m}(z) \simeq u(z)$, associated uncertainty $\tilde{k}(z, z) = \text{Var}(V(z))$.

GPR in 1D

- 1D example : expensive function, only 7 data points available.

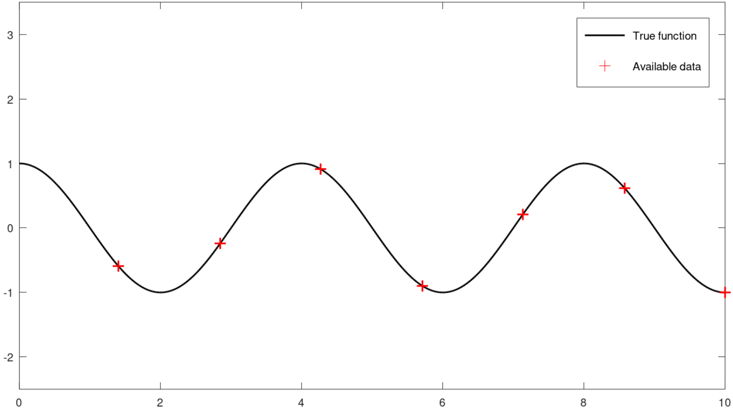


Figure 4 – Function to be approximated using GPR

GPR in 1D

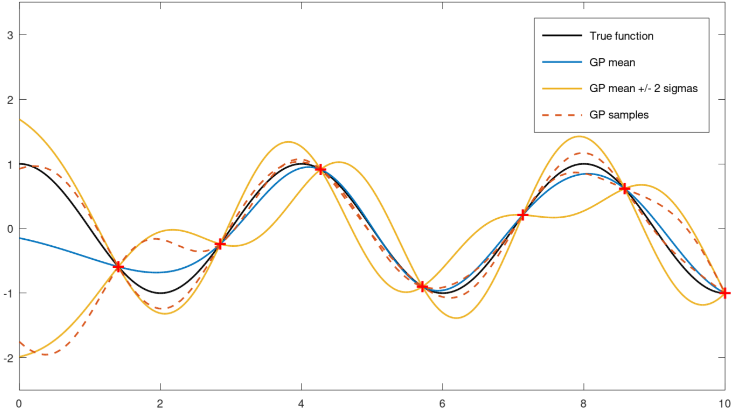


Figure 5 – GPR, 7 data points

Outline of the talk

1. ML and PDEs : context, main notions and tools
   - Context : ML, PDEs
   - Kernel regression, Gaussian processes
   - Being a “solution” of a PDE
2. Imposing physical constraints on a GP
   - Distributional PDE constraints on a random process
   - Sobolev regularity of Gaussian processes
3. “Numerical” applications
   - Covariance kernels for the wave equation
   - GPR and numerical schemes
4. Conclusion and perspectives

Distributional solution of a PDE

Transport of a discontinuity :

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

Solution given by $u(x,t) = u_0(x-ct)$... Meaning of (2) if u_0 is discontinuous (e.g. only $u_0 \in L^2$) ?

Distributional solution of a PDE

Transport of a discontinuity :

$$\begin{cases} \partial_t u + c\partial_x u = 0, \\ u_{|t=0} = u_0. \end{cases} \tag{2}$$

Solution given by $u(x,t) = u_0(x-ct)$... Meaning of (2) if u_0 is discontinuous (e.g. only $u_0 \in L^2$) ?

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

Distributional solution of a PDE

Tansport of a discontinuity :

$$\begin{cases} \partial_t u + c\partial_x u = 0, \\ u_{|t=0} = u_0. \end{cases} \tag{3}$$

Solution : $u(x,t) = u_0(x-ct)$... Meaning of (2) if u_0 is discontinuous ?

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

Distributional solution of a PDE

Tansport of a discontinuity :

$$\begin{cases} \partial_t u + c\partial_x u = 0, \\ u_{|t=0} = u_0. \end{cases} \tag{3}$$

Solution : $u(x,t) = u_0(x - ct)$... Meaning of (2) if u_0 is discontinuous ?

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

- Duality.
- Weak solution (Sobolev, 1934), theory of distributions (Schwartz, 1946).

Some conclusions

Moral of the story :

- Solutions of a PDE : non trivial definition.
- Functional analysis is useful/required.
- If possible, avoid continuity assumptions over target function.

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.

Some conclusions

Moral of the story :

- Solutions of a PDE : non trivial definition.
- Functional analysis is useful/required.
- If possible, avoid continuity assumptions over target function.

Problem for today :

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

Impose laws of physics (PDE) on statistical model (GPR) ; identify consequences on the kernel.

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.

Some conclusions

Moral of the story :

- Solutions of a PDE : non trivial definition.
- Functional analysis is useful/required.
- If possible, avoid continuity assumptions over target function.

Problem for today :

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

Impose laws of physics (PDE) on statistical model (GPR) ; identify consequences on the kernel.

Application cases :

- Linear PDEs : $Lu = 0$ (invariance, see [2])
- Sobolev regularity (energy).

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.


EVERYONE CALM DOWN
I GOT THIS!

Outline of the talk

1. ML and PDEs : context, main notions and tools
   - Context : ML, PDEs
   - Kernel regression, Gaussian processes
   - Being a "solution" of a PDE
2. Imposing physical constraints on a GP
   - Distributional PDE constraints on a random process
   - Sobolev regularity of Gaussian processes
3. "Numerical" applications
   - Covariance kernels for the wave equation
   - GPR and numerical schemes
4. Conclusion and perspectives

Partial derivatives in d dimensions : notations

Soit $\alpha \in \mathbb{N}^d$ ($d-$tuple), $\alpha = (\alpha_1, ..., \alpha_d)$. Denote

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

Order of differentiation : $|\alpha| := \alpha_1 + ... + \alpha_d$.

Linear differential operator of order n :

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

Distributional formulation of a linear PDE

Let L be a linear differential operator over an open set $\mathcal{D} \subset \mathbb{R}^d$:

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

Distributional formulation of a linear PDE

Let L be a linear differential operator over an open set $\mathcal{D} \subset \mathbb{R}^d$:

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

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

Distributional formulation of a linear PDE

Let L be a linear differential operator over an open set $\mathcal{D} \subset \mathbb{R}^d$:

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

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

Multiply by a test function $\varphi \in C_c^\infty(\mathcal{D})$ and integrate over $\mathcal{D}$

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

Distributional formulation of a linear PDE

Let L be a linear differential operator over an open set $\mathcal{D} \subset \mathbb{R}^d$:

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

Strong solution : $u \in C^n(\mathcal{D})$ et $(Lu)(x) = 0 \;\; \forall x \in \mathcal{D}$.

Multiply by a test function $\varphi \in C_c^\infty(\mathcal{D})$ and integrate over $\mathcal{D}$

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

Formal adjoint : $L^* v = \sum_{|\alpha| \leq n} (-1)^{|\alpha|} \partial^\alpha (a_\alpha v)$. Successive IBP yield distributional solutions :

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

Distributional formulation of a linear PDE

Let L be a linear differential operator over an open set $\mathcal{D} \subset \mathbb{R}^d$:

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

Strong solution : $u \in C^n(\mathcal{D})$ et $(Lu)(x) = 0 \;\; \forall x \in \mathcal{D}$.

Multiply by a test function $\varphi \in C_c^\infty(\mathcal{D})$ and integrate over $\mathcal{D}$

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

Formal adjoint : $L^* v = \sum_{|\alpha| \leq n} (-1)^{|\alpha|} \partial^\alpha (a_\alpha v)$. Successive IBP yield distributional solutions :

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

Only requires that $u \in L^1_{loc}(\mathcal{D})$, i.e. $\int_K |u| < +\infty$ for all $K \Subset \mathcal{D}$.

PDE constrained random (Gaussian) processes [4]

Given $U = (U(z))_{z\in\mathcal{D}}$ a RF, when does $\mathbb{P}(L(U) = 0) = 1$ hold ?

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

PDE constrained random (Gaussian) processes [4]

Given $U = (U(z))_{z \in \mathcal{D}}$ a RF, when does $\mathbb{P}(L(U) = 0) = 1$ hold ?
U is *measurable* if $(\omega, z) \mapsto U(z)(\omega)$ is measurable.

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

PDE constrained random (Gaussian) processes [4]

Given $U = (U(z))_{z \in \mathcal{D}}$ a RF, when does $\mathbb{P}(L(U) = 0) = 1$ hold ?
U is *measurable* if $(\omega, z) \mapsto U(z)(\omega)$ is measurable.

Proposition 1

Let $\mathcal{D} \subset \mathbb{R}^d$ be an open set, and $L := \sum_{|\alpha| \leq n} a_\alpha \partial^\alpha$, $a_\alpha \in \mathcal{C}^{|\alpha|}(\mathcal{D})$. Let $U = \big(U(z)\big)_{z \in \mathcal{D}}$ be a second order centered measurable random field, with covariance function k ; assume that $\sigma : z \longmapsto k(z,z)^{1/2} \in L^1_{loc}(\mathcal{D})$. Then the following statements are equivalent.

- $\mathbb{P}(L(U) = 0$ *in the distributional sense*$) = 1$
- $\forall z \in \mathcal{D}, L(k(z, \cdot)) = 0$ *in the distributional sense.*

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

PDE constrained random (Gaussian) processes [4]

Given $U = (U(z))_{z\in\mathcal{D}}$ a RF, when does $\mathbb{P}(L(U) = 0) = 1$ hold ?
U is *measurable* if $(\omega, z) \mapsto U(z)(\omega)$ is measurable.

Proposition 1

Let $\mathcal{D} \subset \mathbb{R}^d$ be an open set, and $L := \sum_{|\alpha|\leq n} a_\alpha \partial^\alpha$, $a_\alpha \in \mathcal{C}^{|\alpha|}(\mathcal{D})$. Let $U = \big(U(z)\big)_{z\in\mathcal{D}}$ be a second order centered measurable random field, with covariance function k ; assume that $\sigma : z \longmapsto k(z,z)^{1/2} \in L^1_{loc}(\mathcal{D})$. Then the following statements are equivalent.

- $\mathbb{P}(L(U) = 0$ *in the distributional sense*$) = 1$
- $\forall z \in \mathcal{D}, L(k(z,\cdot)) = 0$ *in the distributional sense.*

Generalizes a result from [3] to distributional PDE constraints. This property is inherited to conditioned GPs (for GPR purposes).

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

Examples of kernels verifying $L(k(z,\cdot)) = 0 \ \ \forall z$

Given L, find k_L such that $L(k_L(\cdot, z)) = 0 \ \forall z$; $\Delta = \sum_{i=1}^d \partial^2_{x_i x_i}$.

Examples of kernels verifying $L(k(z,\cdot)) = 0 \;\; \forall z$

Given L, find k_L such that $L(k_L(\cdot, z)) = 0 \; \forall z$; $\Delta = \sum_{i=1}^{d} \partial^2_{x_i x_i}$.

- Laplace : $\Delta u = 0$ (Mendes et da Costa Júnior, 2012), (Ginsbourger et al., 2016)
- Heat : $\partial_t - D\Delta u = 0$ (Albert et Rath, 2020)
- Div/Curl free : $\nabla \cdot u = 0$, $\nabla \times u = 0$ (Scheuerer et Schlather, 2012),(Owhadi, 2023b)
- Linear continuum mechanics : (Jidling et al., 2018)
- Helmholtz : $-\Delta u = \lambda u$ (Albert et Rath, 2020)
- (non)stationary Maxwell : (Wahlstrom et al., 2013), (Jidling et al., 2017),(Lange-Hegermann, 2018)
- 3D wave equation, transport : (H. et al., 2023b)
- See also “latent forces” : (Álvarez et al., 2009),(López-Lopera et al., 2021)

Examples of kernels verifying $L(k(z,\cdot)) = 0 \ \forall z$

Given L, find k_L such that $L(k_L(\cdot,z)) = 0 \ \forall z$; $\Delta = \sum_{i=1}^d \partial^2_{x_i x_i}$.

- Laplace : $\Delta u = 0$ (Mendes et da Costa Júnior, 2012), (Ginsbourger et al., 2016)
- Heat : $\partial_t - D\Delta u = 0$ (Albert et Rath, 2020)
- Div/Curl free : $\nabla \cdot u = 0$, $\nabla \times u = 0$ (Scheuerer et Schlather, 2012),(Owhadi, 2023b)
- Linear continuum mechanics : (Jidling et al., 2018)
- Helmholtz : $-\Delta u = \lambda u$ (Albert et Rath, 2020)
- (non)stationary Maxwell : (Wahlstrom et al., 2013), (Jidling et al., 2017),(Lange-Hegermann, 2018)
- 3D wave equation, transport : (H. et al., 2023b)
- See also "latent forces" : (Álvarez et al., 2009),(López-Lopera et al., 2021)

Often based on representations of solutions of $Lu = 0$ (+BC) of the form

$$u = Gf.$$

Outline of the talk

1. ML and PDEs : context, main notions and tools
   - Context : ML, PDEs
   - Kernel regression, Gaussian processes
   - Being a "solution" of a PDE
2. Imposing physical constraints on a GP
   - Distributional PDE constraints on a random process
   - Sobolev regularity of Gaussian processes
3. "Numerical" applications
   - Covariance kernels for the wave equation
   - GPR and numerical schemes
4. Conclusion and perspectives

Sobolev spaces and PDEs

- Energy functionals ("physical interpretation") :

Sobolev spaces and PDEs

- Energy functionals ("physical interpretation") :
  - Heat : $T(x,t), (x,t) \in \mathbb{R}^d \times \mathbb{R}_+$, $\partial_t T - \Delta T = 0$.

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

Sobolev spaces and PDEs

- Energy functionals ("physical interpretation") :
  - Heat : $T(x,t), (x,t) \in \mathbb{R}^d \times \mathbb{R}_+$, $\partial_t T - \Delta T = 0$.

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

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

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

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

Sobolev spaces and PDEs

- Energy functionals ("physical interpretation") :
  - Heat : $T(x,t), (x,t) \in \mathbb{R}^d \times \mathbb{R}_+$, $\partial_t T - \Delta T = 0$.

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

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

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

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

$\rightarrow$ L^2 norms of derivatives.

Finite energy derivatives and Sobolev spaces

Some functions are "almost" differentiable : $h(x) = \max(0, 1 - |x|)$.

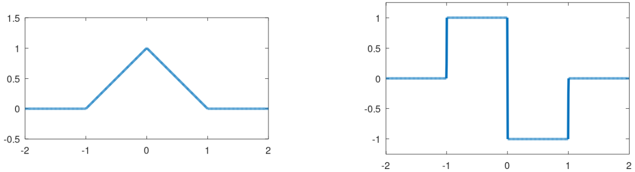
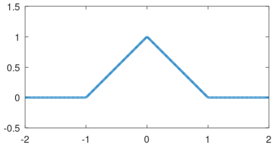


Figure 6 – Left : $h(x)$. Right : $h'(x)$ (hopefully).

Unfortunately, $h' \notin C^0$... but $h' \in L^2$ (finite energy) !

Finite energy derivatives and Sobolev spaces

Some functions are "almost" differentiable : $h(x) = \max(0, 1 - |x|)$.



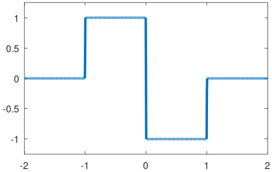


Figure 6 – Left : $h(x)$. Right : $h'(x)$ (hopefully).

Unfortunately, $h' \notin C^0$... but $h' \in L^2$ (finite energy) !
A function g is the weak derivative of h if for all $\varphi \in C_c^\infty(\mathbb{R})$,

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

Finite energy derivatives and Sobolev spaces

Some functions are "almost" differentiable : $h(x) = \max(0, 1 - |x|)$.

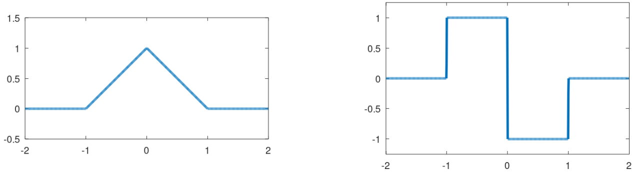


Figure 6 – Left : $h(x)$. Right : $h'(x)$ (hopefully).

Unfortunately, $h' \notin C^0$... but $h' \in L^2$ (finite energy) !
A function g is the weak derivative of h if for all $\varphi \in C_c^\infty(\mathbb{R})$,

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

Define then

$$H^1(\mathbb{R}) := \{u \in L^2(\mathbb{R}) : u' \text{ exists in the weak sense and } u' \in L^2(\mathbb{R})\},$$
$$H^m(\mathcal{D}) := \{u \in L^2(\mathcal{D}) : \forall\ |\alpha| \leq m, \partial^\alpha u \text{ exists ITWS and } \partial^\alpha u \in L^2(\mathcal{D})\}.$$

Covariance kernel and regularity of the GP

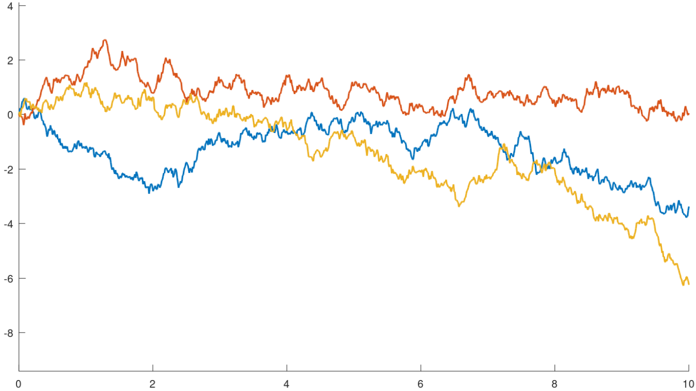


Figure 7 – Brownian motion : $k(x, y) = \min(x, y)$.

Covariance kernel and regularity of the GP

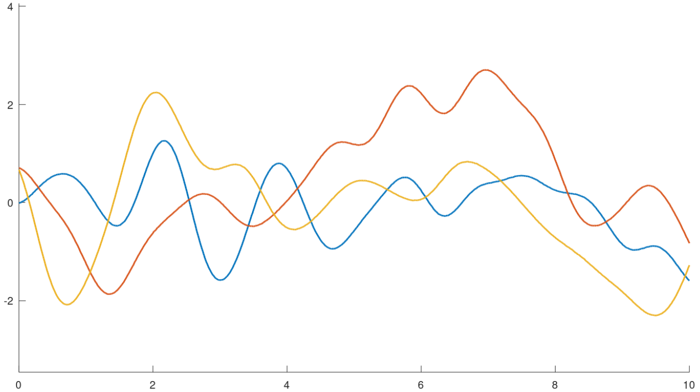


Figure 8 – Gaussian (SE) : $k(x, y) = \sigma^2 \exp(-|x - y|^2/2\ell^2), \sigma = 1, \ell = 0.5.$

H^m ($W^{m,p}$) regularity of a Gaussian process, $m \in \mathbb{N}$ [5]

Proposition 2

Let $(U(z))_{z\in\mathcal{D}} \sim GP(0,k)$ *be a measurable GP, there is an equiv. between*

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

5. H., I. (2024). Sobolev regularity of Gaussian random fields. *J. Func. Anal., 286(3)*, Paper No. 110241.

H^m ($W^{m,p}$) regularity of a Gaussian process, $m \in \mathbb{N}$ [5]

Proposition 2

Let $(U(z))_{z\in\mathcal{D}} \sim GP(0,k)$ be a measurable GP, there is an equiv. between

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

(ii) (Spectral) For all $|\alpha| \le m$, $\partial^{\alpha,\alpha} k \in L^2(\mathcal{D}\times\mathcal{D})$ and the integral operator $\mathcal{E}_k^\alpha$

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

is trace class, with $Tr(\mathcal{E}_k^\alpha) = \int_{\mathcal{D}} \partial^{\alpha,\alpha} k(x,x)dx < +\infty$.

5. H., I. (2024). Sobolev regularity of Gaussian random fields. *J. Func. Anal., 286(3)*, Paper No. 110241.

H^m ($W^{m,p}$) regularity of a Gaussian process, $m \in \mathbb{N}$ [5]

Proposition 2

Let $(U(z))_{z\in\mathcal{D}} \sim GP(0,k)$ be a measurable GP, there is an equiv. between
(i) (Sobolev) $\mathbb{P}(U \in H^m(\mathcal{D})) = 1$

(ii) (Spectral) For all $|\alpha| \le m$, $\partial^{\alpha,\alpha} k \in L^2(\mathcal{D}\times\mathcal{D})$ and the integral operator $\mathcal{E}_k^\alpha$

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

is trace class, with $Tr(\mathcal{E}_k^\alpha) = \int_{\mathcal{D}} \partial^{\alpha,\alpha} k(x,x)dx < +\infty$.

(iii) (Mercer) There exists $(\phi_n) \subset L^2(\mathcal{D})$ s.t. $k(x,y) = \sum_n \phi_n(x)\phi_n(y)$ in L^2. If $|\alpha| \le m$, then $\phi_n \in H^m(\mathcal{D})$ and $Tr(\mathcal{E}_k^\alpha) = \sum_{n=0}^{+\infty} \|\partial^\alpha \phi_n\|_2^2 < +\infty$.

5. H., I. (2024). Sobolev regularity of Gaussian random fields. *J. Func. Anal., 286(3)*, Paper No. 110241.

H^m ($W^{m,p}$) regularity of a Gaussian process, $m \in \mathbb{N}$ [5]

Proposition 2

Let $(U(z))_{z\in\mathcal{D}} \sim GP(0,k)$ be a measurable GP, there is an equiv. between

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

(ii) (Spectral) For all $|\alpha| \le m$, $\partial^{\alpha,\alpha} k \in L^2(\mathcal{D}\times\mathcal{D})$ *and the integral operator* $\mathcal{E}_k^\alpha$

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

is trace class, with $Tr(\mathcal{E}_k^\alpha) = \int_{\mathcal{D}} \partial^{\alpha,\alpha} k(x,x) dx < +\infty$.

(iii) (Mercer) There exists $(\phi_n) \subset L^2(\mathcal{D})$ *s.t.* $k(x,y) = \sum_n \phi_n(x)\phi_n(y)$ *in* L^2. *If* $|\alpha| \le m$, *then* $\phi_n \in H^m(\mathcal{D})$ *and* $Tr(\mathcal{E}_k^\alpha) = \sum_{n=0}^{+\infty} \|\partial^\alpha \phi_n\|_2^2 < +\infty$.

(iv) (Driscoll) $RKHS(k) \subset H^m(\mathcal{D})$, *and denoting* $\mathcal{I}$ *the associated imbedding, we have* $Tr(\mathcal{I}\mathcal{I}^*) = \sum_{|\alpha|\le m} Tr(\mathcal{E}_k^\alpha) < +\infty$.

5. H., I. (2024). Sobolev regularity of Gaussian random fields. *J. Func. Anal., 286(3)*, Paper No. 110241.

Outline of the talk

1. ML and PDEs : context, main notions and tools
   - Context : ML, PDEs
   - Kernel regression, Gaussian processes
   - Being a "solution" of a PDE
2. Imposing physical constraints on a GP
   - Distributional PDE constraints on a random process
   - Sobolev regularity of Gaussian processes
3. "Numerical" applications
   - Covariance kernels for the wave equation
   - GPR and numerical schemes
4. Conclusion and perspectives

GPR and the wave equation [6]

3D homogeneous wave equation : $\Delta := \partial^2_{xx} + \partial^2_{yy} + \partial^2_{zz}$

$$\begin{cases} Lu & = \frac{1}{c^2}\partial^2_{tt}u - \Delta u = \Box u = 0, \quad (x,t) \in \mathbb{R}^3 \times \mathbb{R}^+ \\ u(x,0) & = u_0(x), \quad \partial_t u(x,0) = v_0(x). \end{cases} \tag{11}$$

6. H., I., Noble, P., & Roustant, O. (2023a). Covariance models and Gaussian process regression for the wave equation. Application to related inverse problems. *Journal of Computational Physics*, *494*, Paper No. 112519.

GPR and the wave equation [6]

3D homogeneous wave equation : $\Delta := \partial^2_{xx} + \partial^2_{yy} + \partial^2_{zz}$

$$\begin{cases} Lu & = \frac{1}{c^2}\partial^2_{tt} u - \Delta u = \Box u = 0, \quad (x,t) \in \mathbb{R}^3 \times \mathbb{R}^+ \\ u(x,0) & = u_0(x), \quad \partial_t u(x,0) = v_0(x). \end{cases} \tag{11}$$

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

$$u(x,t) = (F_t * v_0)(x) + (\dot{F}_t * u_0)(x). \tag{12}$$

6. H., I., Noble, P., & Roustant, O. (2023a). Covariance models and Gaussian process regression for the wave equation. Application to related inverse problems. *Journal of Computational Physics*, *494*, Paper No. 112519.

GPR and the wave equation [6]

3D homogeneous wave equation : $\Delta := \partial^2_{xx} + \partial^2_{yy} + \partial^2_{zz}$

$$\begin{cases} Lu & = \frac{1}{c^2}\partial^2_{tt}u - \Delta u = \Box u = 0, \quad (x,t) \in \mathbb{R}^3 \times \mathbb{R}^+ \\ u(x,0) & = u_0(x), \qquad \partial_t u(x,0) = v_0(x). \end{cases} \tag{11}$$

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

$$u(x,t) = (F_t * v_0)(x) + (\dot{F}_t * u_0)(x). \tag{12}$$

Assume that u_0, v_0 are unknown $\rightarrow u_0 \sim GP(0, k_u)$ and $v_0 \sim GP(0, k_v)$, independant. u given by (12) is a centered GP, its kernel is

$$k((x,t),(x',t')) = [(F_t \otimes F_{t'}) * k_v](x,x') + [(\dot{F}_t \otimes \dot{F}_{t'}) * k_u](x,x'). \tag{13}$$

The kernel k verifies $\Box k((x,t),\cdot) = 0$ for all $(x,t) \in \mathbb{R}^3 \times \mathbb{R}_+$.

6. H., I., Noble, P., & Roustant, O. (2023a). Covariance models and Gaussian process regression for the wave equation. Application to related inverse problems. *Journal of Computational Physics*, *494*, Paper No. 112519.

Estimation of physical parameters and initial conditions

- Initial condition reconstruction : the GPR mean verifies $\square \tilde{m} = 0$. Hence

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

Estimation of physical parameters and initial conditions

- Initial condition reconstruction : the GPR mean verifies $\Box\tilde{m} = 0$. Hence

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

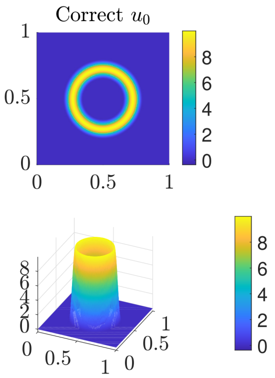
- Parameters of the PDE may also be estimated with GPR : celerity c, source position, source size...

  $\rightarrow$ can be estimated using marginal likelihood (standard in GPR).

Restrictive framework

Expensive convolutions (4D) $\rightarrow$ radial symmetry framework (explicit convolutions).

- Numerical solution of the wave equation in $[0,1]^3$, $v_0 = 0$ and



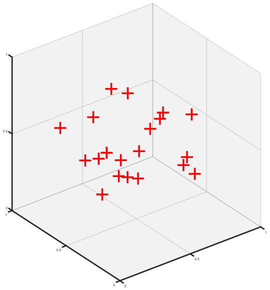


Figure 9 – Sensor positions

Data visulatization

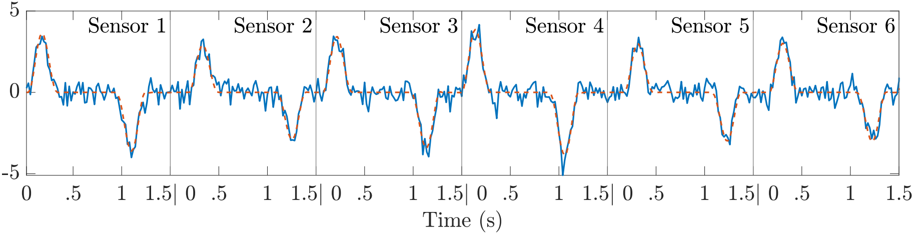


Figure 10 – Examples of captured signals. Red : noiseless signal. Blue : noisy signal.

Reconstruction of initial conditions and position parameters

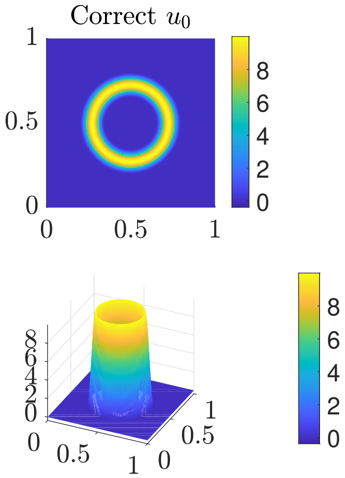


Figure 11 – True u_0 (left column) vs GPR u_0 (right column). 15 sensors are used. Images correspond to 3D slices at $z = 0.5$.

Outline of the talk

1. ML and PDEs : context, main notions and tools
   - Context : ML, PDEs
   - Kernel regression, Gaussian processes
   - Being a "solution" of a PDE
2. Imposing physical constraints on a GP
   - Distributional PDE constraints on a random process
   - Sobolev regularity of Gaussian processes
3. "Numerical" applications
   - Covariance kernels for the wave equation
   - GPR and numerical schemes
4. Conclusion and perspectives

The finite difference method

We seek the numerical solution of a given PDE. Linear transport :

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

Grid $G = \{(i\Delta x, n\Delta t), i \in \mathbb{Z}, n \in \mathbb{N}\}$. To construct $u_i^n \simeq u(i\Delta x, n\Delta t)$, me may solve

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

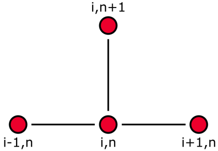
i.e., with $\lambda = c\Delta x/\Delta t$,

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

GPR on a grid

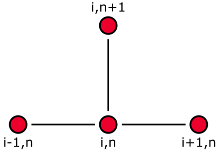
- Function $u(x,t)$ to be estimated on a grid.
- From $(u_{i-1}^n, u_i^n, u_{j+1}^n)$, estimate the values of u at

$$((i-1)\Delta x, n\Delta t),\ (i\Delta x, n\Delta t),\ ((i+1)\Delta x, n\Delta t),$$
$$\text{estimate } u(i\Delta x, (n+1)\Delta t).$$



GPR on a grid

- Function $u(x,t)$ to be estimated on a grid.
- From $(u^n_{i-1}, u^n_i, u^n_{j+1})$, estimate the values of u at



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

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

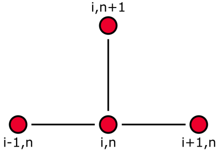
GPR is linear w.r.t. data :

$$u_i^{n+1} = \tilde{m}(i\Delta x, (n+1)\Delta t) = a_{-1}u^n_{i-1} + a_0 u^n_i + a_1 u^n_{i+1}, \tag{14}$$

where $(a_{-1}, a_0, a_1)^T$ given by GPR formulas.

GPR on a grid

- Function $u(x,t)$ to be estimated on a grid.
- From $(u_{i-1}^n, u_i^n, u_{j+1}^n)$, estimate the values of u at



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

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

GPR is linear w.r.t. data :

$$u_i^{n+1} = \tilde{m}(i\Delta x, (n+1)\Delta t) = a_{-1}u_{i-1}^n + a_0 u_i^n + a_1 u_{i+1}^n, \qquad (14)$$

where $(a_{-1}, a_0, a_1)^T$ given by GPR formulas.

Is equation (14) a "numerical scheme" ? For what PDE ? What kernel to use ? Properties ?

GPR and FDM : a first result [7], [8]

Proposition 3

If $k((x,t),(x',t')) = k_0(x-ct, x'-ct')$, $k_0(h) = k_\nu(h/\ell)$, *Matérn of order* $\nu > 0$, *then when* $\ell \to +\infty$, *we obtain*

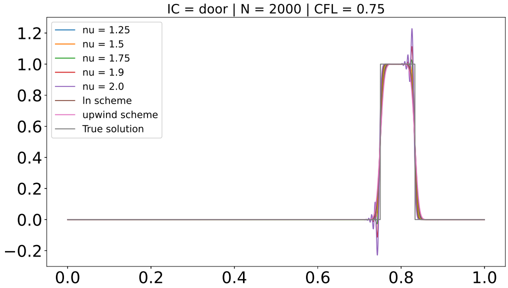
- *if* $\nu = 1/2$, *upwind :*
  $a_{-1} = (\lambda + |\lambda|)/2, \quad a_0 = 1 - |\lambda|, \quad a_1 = (-\lambda + |\lambda|)/2.$
- *if* $\nu \geq 2$, *Lax-Wendroff :*
  $a_{-1} = (\lambda + \lambda^2)/2, \quad a_0 = 1 - \lambda^2, \quad a_1 = (-\lambda + \lambda^2)/2.$
- *if* $1 < \nu < 2$, *fractional splines :*

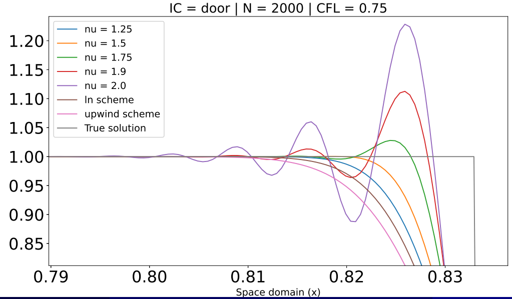
$$a_{-1} = \frac{\lambda + D(\nu,\lambda)}{2}, \quad a_0 = 1 - D(\nu,\lambda), \quad a_1 = \frac{-\lambda + D(\nu,\lambda)}{2},$$

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

7. H., I. (2023). *PDE constrained kernel regression methods* (thèse de doct.). INSA Toulouse.
8. Micchelli, C. A., & Miranker, W. L. (1974). Asymptotically optimal approximation in fractional Sobolev spaces and the numerical solution of differential equations. *Numer. Math., 22*, 75-87.

Qualitative properties





Quantitative properties

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

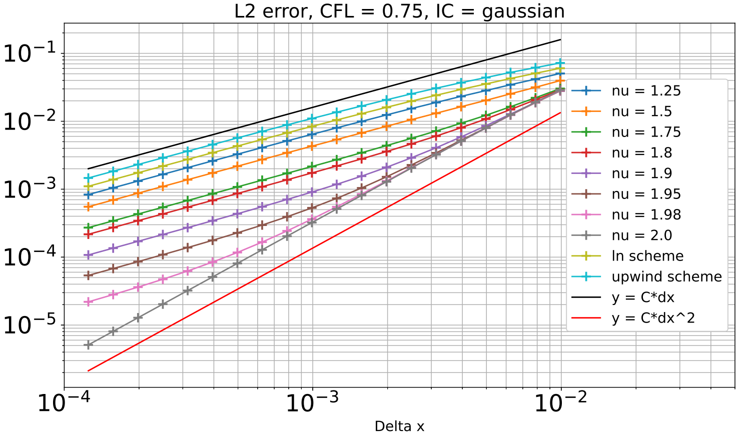


Figure 12 – ℓ^2 error, $\Delta x \mapsto E_{\nu,2}(\Delta x)$

Retrospective

GPR, GPs constrained by physical laws

- Linear distributional PDE constraints [9]
- Energy constraints : H^m, $W^{m,p}$ [10]

→ Necessary and sufficient conditions outside of continuity assumptions.

"Numerical" applications :

- Wave equation and related inverse problems [11]
- First result between GPR and FDM.

---

9. H., I., Noble, P., & Roustant, O. (2023b). Characterization of the second order random fields subject to linear distributional PDE constraints. *Bernoulli*, *29*(4), 3396-3422.
10. H., I. (2024). Sobolev regularity of Gaussian random fields. *J. Func. Anal.*, *286(3)*, Paper No. 110241.
11. H., I., Noble, P., & Roustant, O. (2023a). Covariance models and Gaussian process regression for the wave equation. Application to related inverse problems. *Journal of Computational Physics*, *494*, Paper No. 112519.

Perspectives

- Non linear PDEs : [12]
- Error analysis of GPR using Sobolev norms [13].
- 3D wave equation : computational issues (convolutions) [14].

12. Chen, Y., Hosseini, B., Owhadi, H., & Stuart, A. M. (2021). Solving and learning nonlinear PDEs with Gaussian processes. *Journal of Computational Physics*, *447*, 110668.
13. Batlle, P., Chen, Y., Hosseini, B., Owhadi, H., & Stuart, A. M. (2023). Error Analysis of Kernel/GP Methods for Nonlinear and Parametric PDEs.
14. Tick, J., Pulkkinen, A., Lucka, F., Ellwood, R., Cox, B. T., Kaipio, J. P., Arridge, S. R., & Tarvainen, T. (2018). Three dimensional photoacoustic tomography in Bayesian framework. *The Journal of the Acoustical Society of America*, *144*(4), 2061-2071.

Thank you for your attention !

Contact : iain.pl.henderson@gmail.com

Physical parameter estimation

N_{sensors}	3	5	10	15	20	25	30	Target
$\lvert \hat{x}_0 - x_0^* \rvert$	0.204	0.003	0.004	0.008	0.003	0.004	0.015	0
$\hat{R}$	0.386	0.432	0.462	0.431	0.414	0.471	0.452	0.25
$\lvert \hat{c} - c^* \rvert$	0.084	0.004	0.005	0.005	0.006	0.001	0.004	0
$\hat{\sigma}^2_{\text{noise}}$	0.917	0.879	0.93	0.99	0.361	0.988	0.377	0.2025
$\hat{\ell}$	0.02	0.02	0.025	0.02	0.035	0.024	0.032	~ 0.05
$\hat{\sigma}^2$	2.367	3.513	4.903	3.168	4.446	4.619	4.79	Unknown
$e^{\text{u}}_{1,\text{rel}}$	1.275	0.157	0.128	0.168	0.11	0.103	0.248	0
$e^{\text{u}}_{2,\text{rel}}$	1.056	0.095	0.082	0.124	0.088	0.064	0.213	0
$e^{\text{u}}_{\infty,\text{rel}}$	1.037	0.132	0.128	0.198	0.136	0.101	0.321	0

Table 1 – Hyperparameter estimation and relative errors

Point source localization : $R \ll 1$

Case where $u_0 \equiv 0$ and the source v_0 is $\sim$ point source at x_0^* : use the kernels

$$k_v^R(x,x') = k_v(x,x')\frac{\mathbb{1}_{B(x_0,R)}(x)}{4\pi R^3/3}\frac{\mathbb{1}_{B(x_0,R)}(x')}{4\pi R^3/3}, \tag{15}$$

$$k((x,t),(x',t')) = [(F_t \otimes F_{t'}) * k_v^R](x,x'), \tag{16}$$

with $R \ll 1$. Hyperparameters of k : $\theta = (\theta_v, x_0, R, c)$. Fix all hyperparemeters to "correct values" except for x_0 : marginal likelihood is $\mathcal{L}(\theta) = \mathcal{L}(x_0),\ x_0 \in \mathbb{R}^3$.

Question : behaviour of $x_0 \mapsto \mathcal{L}(x_0)$?

Marginal likelihood estimation ≡ GPS localization [15]

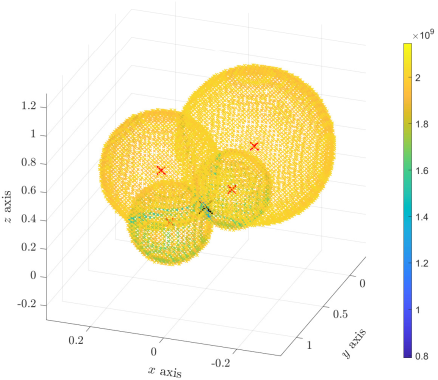


Figure : $x_0 \mapsto \mathcal{L}(x_0)$.

Displayed values : such that $\leq 2.035 \times 10^9$.

× : location of sensors.

× : true location of source.

15. H., I., Noble, P., & Roustant, O. (2023a). Covariance models and Gaussian process regression for the wave equation. Application to related inverse problems. *Journal of Computational Physics*, *494*, Paper No. 112519.

Estimation of physical parameters and initial conditions

- Initial condition reconstruction : the GPR mean verifies $\Box \tilde{m} = 0$. Hence

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

Estimation of physical parameters and initial conditions

- Initial condition reconstruction : the GPR mean verifies $\Box\tilde{m} = 0$. Hence

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

- The kernel k is parametrized by c, θ_u and θ_v ; θ_u and θ_v may contain physical data w.r.t. u_0 and v_0.
  Example : compactly supported initial condition is modelled by

$$k_u(x, x') = k_u^0(x, x')\mathbb{1}_{B(x_0,R)}(x)\mathbb{1}_{B(x_0,R)}(x'), \tag{17}$$

  gives that $(x_0, R) \in \theta_u$. Likewise for v_0 (we can also encode symmetries).
  $\rightarrow$ can be estimated using marginal likelihood.

A word on GPR and neural networks

- Some GPs as limits of infinitely wide NNs ([16], Section 4.2.3).
- NN regression is a kernel regression method with a kernel learnt from data ([17] ; Mallat, collège de France).
- GPR : "competes" with NNs (PINNs in particular), cf [18] for a 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.

Radial symmetry formulas

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

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

L^2 regularity of a Gaussian process [19]

- Integral criterion :

$$\mathbb{P}\big(U \in L^2(\mathcal{D})\big) = 1 \iff \int_{\mathcal{D}} k(x,x)dx < +\infty. \tag{18}$$

19. Bogachev, V. I. (1998). *Gaussian measures*. American Mathematical Soc.

L^2 regularity of a Gaussian process [19]

- Integral criterion :

$$\mathbb{P}\big(U \in L^2(\mathcal{D})\big) = 1 \iff \int_{\mathcal{D}} k(x,x)dx < +\infty. \tag{18}$$

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

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

L^2 regularity of a Gaussian process [19]

- Integral criterion :

$$\mathbb{P}\big(U \in L^2(\mathcal{D})\big) = 1 \iff \int_{\mathcal{D}} k(x,x)dx < +\infty. \tag{18}$$

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

- Spectral/Mercer-type criterion : let $\mathcal{E}_k : L^2(\mathcal{D}) \to L^2(\mathcal{D})$ be

$$(\mathcal{E}_k f)(x) := \int k(x,y)f(y)dy. \tag{19}$$

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

L^2 regularity of a Gaussian process [19]

- Integral criterion :

$$\mathbb{P}\big(U \in L^2(\mathcal{D})\big) = 1 \iff \int_{\mathcal{D}} k(x,x)dx < +\infty. \tag{18}$$

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

- Spectral/Mercer-type criterion : let $\mathcal{E}_k : L^2(\mathcal{D}) \to L^2(\mathcal{D})$ be

$$(\mathcal{E}_k f)(x) := \int k(x,y)f(y)dy. \tag{19}$$

Si $\int k(x,x)dx < +\infty$, then (...) there exists $(\psi_n) \subset L^2$ s.t.

$$k(x,y) = \sum_{n=0}^{+\infty} \psi_n(x)\psi_n(y) \quad \text{in} \quad L^2(\mathcal{D}\times\mathcal{D}) \quad (\text{"Mercer"}). \tag{20}$$

19. Bogachev, V. I. (1998). *Gaussian measures*. American Mathematical Soc.

L^2 regularity of a Gaussian process [21]

Yields (formally)

$$\int k(x,x)dx = \int \sum_{n=0}^{+\infty} \psi_n(x)^2 dx = \sum_{n=0}^{+\infty} \int \psi_n(x)^2 dx \tag{21}$$

$$= \sum_{n=0}^{+\infty} \|\psi_n\|_2^2 = \sum_{n=0}^{+\infty} \lambda_n = \mathrm{Tr}(\mathcal{E}_k) < +\infty \quad \text{(trace class)}. \tag{22}$$

20. 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.
21. BOGACHEV, V. I. (1998). *Gaussian measures*. American Mathematical Soc.

L^2 regularity of a Gaussian process [21]

Yields (formally)

$$\int k(x,x)dx = \int \sum_{n=0}^{+\infty} \psi_n(x)^2 dx = \sum_{n=0}^{+\infty} \int \psi_n(x)^2 dx \tag{21}$$

$$= \sum_{n=0}^{+\infty} \|\psi_n\|_2^2 = \sum_{n=0}^{+\infty} \lambda_n = \mathrm{Tr}(\mathcal{E}_k) < +\infty \quad \text{(trace class)}. \tag{22}$$

- Imbedding of the RKHS : if $\int k(x,x)dx < +\infty$, then $RKHS(k) \subset L^2(\mathcal{D})$, and denoting $\mathcal{I}$ the associated imbedding, $\mathcal{I}\mathcal{I}^*(= \mathcal{E}_k)$ is trace class ("Driscoll" [20]).

20. 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.
21. Bogachev, V. I. (1998). *Gaussian measures*. American Mathematical Soc.

Un équivalent déterministe : la version RKHS

Soit $k : D \times D \to \mathbb{R}$ 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

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

Extension to nonlinear PDEs

- Nonlinear constraints on $k(z, \cdot)$: not realistic (+ invalid GP model).
- Alternative : in [22], the nonlinear PDE constraint is only imposed pointwise on $\tilde{m}$: modification of the optimization problem in $RKHS(k)$:

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

  Generalizes an approach described in [23].
- Can be coupled with strict linear constraints : [24] (div/curl/periodicity).

22. Chen, Y., Hosseini, B., Owhadi, H., & Stuart, A. M. (2021). Solving and learning nonlinear PDEs with Gaussian processes. *Journal of Computational Physics*, *447*, 110668.
23. Wendland, H. (2004). *Scattered data approximation*. Cambridge university press.
24. Owhadi, H. (2023b). Gaussian Process Hydrodynamics.

Un équivalent déterministe : la version RKHS

Un RKHS est exactement un espace de Hilbert de fonctions $D \to \mathbb{R}$ 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

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

Alors $v = \tilde{m}$. De plus, $\tilde{m} = p_F(u)$ où $F := \text{Span}(k(z_1, \cdot), ..., k(z_n, \cdot))$. 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 $\to \mathbb{P}(\sup_z |V_z| \leq \varepsilon)$, $\mathbb{P}(||V||_E \geq M)$... + 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 $k : \mathcal{D} \times \mathcal{D} \to \mathbb{R}$

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

- $H_k = \overline{\text{Vect}(k(z,\cdot), z \in \mathcal{D})}$, $\|\sum_{i=1}^n a_i k(z_i,\cdot)\|_k^2 = \sum_{i,j=1}^n a_i a_j k(z_i, z_j)$.
- 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 :

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

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 $k : \mathcal{D} \times \mathcal{D} \to \mathbb{R}$

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

- $H_k = \overline{\text{Vect}(k(z,\cdot), z \in \mathcal{D})}$, $\| \sum_{i=1}^n a_i k(z_i,\cdot)\|_k^2 = \sum_{i,j=1}^n a_i a_j k(z_i, z_j)$.
- 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 :

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

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

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 $u_{obs} = (u(z_1), ..., u(z_n)) \in \mathbb{R}^n$ 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

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

Régularité $W^{m,p}$ d'un GP, $m \in \mathbb{N}, p \in (1, +\infty)$

Proposition 4 (H., 2024)

Soit $(U(z))_{z\in\mathcal{D}} \sim GP(0,k)$ *un GP 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* $\mathcal{E}_k^\alpha$

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

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*

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

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

Why kernel methods ?

- interprétation bayésienne → processus gaussiens [25].
- interprétation RKHS → analyse fonctionnelle [26].
- “standard” si relativement peu de données [27].
- une façon d’étudier les réseaux de neurones [28].

25. Rasmussen, C. E., & Williams, C. (2006). *Gaussian Processes for Machine Learning*. The MIT Press.
26. Wendland, H. (2004). *Scattered data approximation*. Cambridge university press.
27. Gramacy, R. B. (2020). *Surrogates: Gaussian process modeling, design, and optimization for the applied sciences*. CRC press.
28. Belkin, M., Ma, S., & Mandal, S. (2018). To Understand Deep Learning We Need to Understand Kernel Learning. *35th ICML*, 541-549.

Formules de GPR

On note $u_{obs} = (u(z_1), ..., u(z_n))$ 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

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

- 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 GP permettent de définir un prior sur E.