Physics-informed Kriging with applications to inverse problems involving PDEs

Iain Henderson

Advisors : Pascal Noble (IMT), Olivier Roustant (IMT)

INSA Toulouse/IMT

MASCOT-NUM, 04/04/2023

Regression under model contraints

Aim : forecast of some phenomena (physical, oceanography)

- Modelled by some unknown function u
- At our disposal : database $B = \{u(z_1), ..., u(z_n)\}$, perhaps limited.
- Model constraining u : $\frac{1}{c^2}\partial_{tt}^2 u = \partial_{xx}^2 u + f$
- Objective : approximate $u(t,x)$ for all (t,x) (regression)

Regression under model contraints

Aim : forecast of some phenomena (physical, oceanography)

- Modelled by some unknown function u
- At our disposal : database $B = \{u(z_1), ..., u(z_n)\}$, perhaps limited.
- Model constraining u : $\frac{1}{c^2}\partial_{tt}^2 u = \partial_{xx}^2 u + f$
- Objective : approximate $u(t, x)$ for all (t, x) (regression)

Idea : combine data and model (grey box model).

Outline of the talk

1. Constrained Gaussian process regression

2. Physics informed Gaussian processes
   - Distributional formulation of PDEs
   - Sobolev regularity of Gaussian random fields

3. Gaussian process regression for the 3D wave equation
   - GP priors for the 3D wave equation
   - Solving some inverse problems
   - Numerical applications

Outline of the talk

1. Constrained Gaussian process regression

2. Physics informed Gaussian processes
   - Distributional formulation of PDEs
   - Sobolev regularity of Gaussian random fields

3. Gaussian process regression for the 3D wave equation
   - GP priors for the 3D wave equation
   - Solving some inverse problems
   - Numerical applications

Gaussian process regression (GPR)

- Unknown function : $u : \mathcal{D} \to \mathbb{R}$, obs. $B = \{u(z_1), ..., u(z_n)\}$
- Model u as a realization of a Gaussian process $(U_z)_{z \in D} \sim GP(0, k)$.

Gaussian process regression (GPR)

- Unknown function : $u : \mathcal{D} \to \mathbb{R}$, obs. $B = \{u(z_1), ..., u(z_n)\}$
- Model u as a realization of a Gaussian process $(U_z)_{z \in D} \sim GP(0, k)$.
- Condition U on data : $V_z = [U_z | U_{z_1} = u(z_1), ..., U_{z_n} = u(z_n)]$, yielding

$$V_z \sim GP(\tilde{m}, \tilde{k}).$$

  $\tilde{m}$ and $\tilde{k}$ are given by the Kriging formulas.
- $\forall z \in D$, approximate $u(z)$ with $\tilde{m}(z)$: $u(z) \simeq \tilde{m}(z)$ + interpolation.

Gaussian process regression (GPR)

- Unknown function : $u : \mathcal{D} \to \mathbb{R}$, obs. $B = \{u(z_1), ..., u(z_n)\}$
- Model u as a realization of a Gaussian process $(U_z)_{z\in D} \sim GP(0, k)$.
- Condition U on data : $V_z = [U_z | U_{z_1} = u(z_1), ..., U_{z_n} = u(z_n)]$, yielding

$$V_z \sim GP(\tilde{m}, \tilde{k}).$$

  $\tilde{m}$ and $\tilde{k}$ are given by the Kriging formulas.
- $\forall z \in D$, approximate $u(z)$ with $\tilde{m}(z)$: $u(z) \simeq \tilde{m}(z)$ + interpolation.

Why use GPR ?

- Mathematically tractable and interpretable.
- GPR as orthogonal projections in the RKHS
  $\to$ more familiar in the PDE community.

Gaussian process regression (GPR)

- Unknown function : $u : \mathcal{D} \to \mathbb{R}$, obs. $B = \{u(z_1), ..., u(z_n)\}$
- Model u as a realization of a Gaussian process $(U_z)_{z \in D} \sim GP(0, k)$.
- Condition U on data : $V_z = [U_z | U_{z_1} = u(z_1), ..., U_{z_n} = u(z_n)]$, yielding

$$V_z \sim GP(\tilde{m}, \tilde{k}).$$

  $\tilde{m}$ and $\tilde{k}$ are given by the Kriging formulas.
- $\forall z \in D$, approximate $u(z)$ with $\tilde{m}(z)$: $u(z) \simeq \tilde{m}(z)$ + interpolation.

Why use GPR ?

- Mathematically tractable and interpretable.
- GPR as orthogonal projections in the RKHS
  $\to$ more familiar in the PDE community.

Now : understand behaviour of GPR w.r.t. PDEs :

(i) linear, (ii) Sobolev.

Linearly constrained GPR

Assume that $Lu = 0$, L linear. The GPR procedure is adapted to this constraint if $L\tilde{m} = 0$.

Linearly constrained GPR

Assume that $Lu = 0$, L linear. The GPR procedure is adapted to this constraint if $L\tilde{m} = 0$.

Denote $u_{obs} = (u(z_1), ..., u(z_n))$ the data, $K_{ij} := k(z_i, z_j)$ et $k(Z, z)_i := k(z_i, z)$.

Linearly constrained GPR

Assume that $Lu = 0$, L linear. The GPR procedure is adapted to this constraint if $L\tilde{m} = 0$.

Denote $u_{obs} = (u(z_1), ..., u(z_n))$ the data, $K_{ij} := k(z_i, z_j)$ et $k(Z, z)_i := k(z_i, z)$. Then the Kriging mean is given by

$$\tilde{m}(z) = k(Z, z)^T K^{-1} u_{obs} \in \mathrm{Span}(k(z_1, \cdot), ..., k(z_n, \cdot)).$$

Linearly constrained GPR

Assume that $Lu = 0$, L linear. The GPR procedure is adapted to this constraint if $L\tilde{m} = 0$.

Denote $u_{obs} = (u(z_1), ..., u(z_n))$ the data, $K_{ij} := k(z_i, z_j)$ et $k(Z, z)_i := k(z_i, z)$. Then the Kriging mean is given by

$$\tilde{m}(z) = k(Z, z)^T K^{-1} u_{obs} \in \text{Span}(k(z_1, \cdot), ..., k(z_n, \cdot)).$$

- $L\tilde{m} = 0$ is ensured if $Lk(z, \cdot) = 0$ for all z.
- More generally : incorporate prior knowledge in the GP prior.

The problem with PDE constraints

Examples of physical constraints : positivity, conservation laws... They generally take the form of partial differential equations (PDEs), e.g.

$$Lu := \sum_{|\alpha| \leq n} a_\alpha(x) \partial^\alpha u = 0. \qquad (1)$$

Above : $\alpha = (\alpha_1, ..., \alpha_d) \in \mathbb{N}^d,\ |\alpha| = \alpha_1 + \ldots + \alpha_d,\ \partial^\alpha = \partial_{x_1}^{\alpha_1} \ldots \partial_{x_d}^{\alpha_d}$.

The problem with PDE constraints

Examples of physical constraints : positivity, conservation laws... They generally take the form of partial differential equations (PDEs), e.g.

$$Lu := \sum_{|\alpha| \leq n} a_\alpha(x) \partial^\alpha u = 0. \tag{1}$$

Above : $\alpha = (\alpha_1, ..., \alpha_d) \in \mathbb{N}^d,\ |\alpha| = \alpha_1 + \ldots + \alpha_d,\ \partial^\alpha = \partial_{x_1}^{\alpha_1} \ldots \partial_{x_d}^{\alpha_d}$.

Limited smoothness of solutions of PDEs

Physically meaningful solutions of (1) may not be n times differentiable.

Example : $\partial_t u + c\partial_x u = 0,\ u(\cdot, t = 0) = u_0$.

Solution is $u(x,t) = u_0(x - ct)$, smoothness looks irrelevant !

Worse if nonlinear : solutions may become discontinuous in finite time.

The problem with PDE constraints

Examples of physical constraints : positivity, conservation laws... They generally take the form of partial differential equations (PDEs), e.g.

$$Lu := \sum_{|\alpha|\leq n} a_\alpha(x)\partial^\alpha u = 0. \quad (1)$$

Above : $\alpha = (\alpha_1, ..., \alpha_d) \in \mathbb{N}^d,\ |\alpha| = \alpha_1 + \ldots + \alpha_d,\ \partial^\alpha = \partial_{x_1}^{\alpha_1} \ldots \partial_{x_d}^{\alpha_d}$.

Limited smoothness of solutions of PDEs

Physically meaningful solutions of (1) may not be n times differentiable.

Example : $\partial_t u + c\partial_x u = 0,\ u(\cdot, t = 0) = u_0$.

Solution is $u(x,t) = u_0(x - ct)$, smoothness looks irrelevant !

Worse if nonlinear : solutions may become discontinuous in finite time. The PDE has to be understood in some weakened form (not just a trick)

$\rightarrow$ weak formulations, Sobolev spaces

Outline of the talk

1. Constrained Gaussian process regression
2. Physics informed Gaussian processes
   - Distributional formulation of PDEs
   - Sobolev regularity of Gaussian random fields
3. Gaussian process regression for the 3D wave equation
   - GP priors for the 3D wave equation
   - Solving some inverse problems
   - Numerical applications

Relaxing the definition of derivatives : distributions

Let $\mathcal{D}$ be an open set of $\mathbb{R}$.

Relaxing the definition of derivatives : distributions

Let $\mathcal{D}$ be an open set of $\mathbb{R}$. Note D the single derivative operator.

$$Lu = \sum_{k=1}^{n} a_k D^k u.$$

Relaxing the definition of derivatives : distributions

Let $\mathcal{D}$ be an open set of $\mathbb{R}$. Note D the single derivative operator.

$$Lu = \sum_{k=1}^{n} a_k D^k u.$$

Assume that u is a classical solution to $Lu = 0$: $u \in C^n(\mathcal{D})$ and

$$\forall x \in \mathcal{D}, (Lu)(x) = 0. \tag{2}$$

Let $\mathcal{D}$ be an open set of $\mathbb{R}$. Note D the single derivative operator.

$$Lu = \sum_{k=1}^{n} a_k D^k u.$$

Assume that u is a classical solution to $Lu = 0$: $u \in C^n(\mathcal{D})$ and

$$\forall x \in \mathcal{D}, (Lu)(x) = 0. \tag{2}$$

Multiply (2) by $\varphi \in C_c^\infty(\mathcal{D})$ and integrate over $\mathcal{D}$:

$$\forall \varphi \in C_c^\infty(\mathcal{D}), \int_{\mathcal{D}} Lu(x)\varphi(x)dx = 0. \quad \text{(smooth local averages)} \tag{3}$$

Relaxing the definition of linear PDEs : distributions

Integration by parts on (3) :

$$\int_{\mathcal{D}} D^k u(x)\varphi(x)dx = (-1)^k \int_{\mathcal{D}} u(x)D^k\varphi(x)dx.$$

Relaxing the definition of linear PDEs : distributions

Integration by parts on (3) :

$$\int_{\mathcal{D}} D^k u(x)\varphi(x)dx = (-1)^k \int_{\mathcal{D}} u(x)D^k\varphi(x)dx.$$

Define $L^* v = \sum_{k=1}^{n} a_k(-1)^k D^k v$ (formal adjoint).

Relaxing the definition of linear PDEs : distributions

Integration by parts on (3) :

$$\int_{\mathcal{D}} D^k u(x)\varphi(x)dx = (-1)^k \int_{\mathcal{D}} u(x) D^k \varphi(x) dx.$$

Define $L^* v = \sum_{k=1}^{n} a_k (-1)^k D^k v$ (formal adjoint). Then,

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

Relaxing the definition of linear PDEs : distributions

Integration by parts on (3) :

$$\int_{\mathcal{D}} D^k u(x)\varphi(x)dx = (-1)^k \int_{\mathcal{D}} u(x) D^k \varphi(x)dx.$$

Define $L^* v = \sum_{k=1}^{n} a_k (-1)^k D^k v$ (formal adjoint). Then,

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

A function u is a solution to the PDE $Lu = 0$ in the distributional sense if

$$\forall \varphi \in C_c^\infty(\mathcal{D}), \int_{\mathcal{D}} u(x) L^*\varphi(x)dx = 0 \quad (4)$$

Relaxing the definition of linear PDEs : distributions

Integration by parts on (3) :

$$\int_{\mathcal{D}} D^k u(x)\varphi(x)dx = (-1)^k \int_{\mathcal{D}} u(x) D^k \varphi(x)dx.$$

Define $L^* v = \sum_{k=1}^{n} a_k (-1)^k D^k v$ (formal adjoint). Then,

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

A function u is a solution to the PDE $Lu = 0$ in the distributional sense if

$$\forall \varphi \in C_c^\infty(\mathcal{D}), \int_{\mathcal{D}} u(x) L^*\varphi(x)dx = 0 \tag{4}$$

One only requires that $u \in L^1_{loc}(\mathcal{D})$ to make sense of (4), i.e.

$$\int_K |u| < +\infty \quad \text{for all compact set} \quad K \subset \mathcal{D}.$$

Random fields under linear distributional PDE constraints

Proposition 1 (H. et al. [2023, to appear])

Let $\mathcal{D} \subset \mathbb{R}^d$ be an open set and let $L := \sum_{|\alpha| \leq n} a_\alpha \partial^\alpha$ with $a_\alpha \in \mathcal{C}^{|\alpha|}(\mathcal{D})$. Let $U = \big(U_z\big)_{z \in \mathcal{D}}$ be a mesurable centered second order random field with covariance function $k(z, z')$. Assume that $\sigma : z \longmapsto k(z,z)^{1/2} \in L^1_{loc}(\mathcal{D})$.

1) Then $\mathbb{P}(\{\omega \in \Omega : U(\omega) \in L^1_{loc}(\mathcal{D})\}) = 1$ and $L^1_{loc}(\mathcal{D})$ and $k(z, \cdot) \in L^1_{loc}(\mathcal{D})$ for all $z \in \mathcal{D}$.

2) The following statements are equivalent :

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

Random fields under linear distributional PDE constraints

Proposition 1 (H. et al. [2023, to appear])

Let $\mathcal{D} \subset \mathbb{R}^d$ be an open set and let $L := \sum_{|\alpha| \leq n} a_\alpha \partial^\alpha$ with $a_\alpha \in \mathcal{C}^{|\alpha|}(\mathcal{D})$. Let $U = \left(U_z\right)_{z \in \mathcal{D}}$ be a mesurable centered second order random field with covariance function $k(z, z')$. Assume that $\sigma : z \longmapsto k(z,z)^{1/2} \in L^1_{loc}(\mathcal{D})$.

1) Then $\mathbb{P}(\{\omega \in \Omega : U(\omega) \in L^1_{loc}(\mathcal{D})\}) = 1$ and $L^1_{loc}(\mathcal{D})$ and $k(z, \cdot) \in L^1_{loc}(\mathcal{D})$ for all $z \in \mathcal{D}$.

2) The following statements are equivalent :

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

This generalizes a result from Ginsbourger et al. [2016] to distributional PDE constraints. This property is inherited on conditioned GPs.

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

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

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

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

- Laplace : $\Delta u = 0$ Mendes and da Costa Júnior [2012], Ginsbourger et al. [2016]
- Heat : $\partial_t - D\Delta u = 0$ Albert and Rath [2020]
- Div/Curl : $\nabla \cdot u = 0$, $\nabla \times u = 0$ Scheuerer and Schlather [2012],Owhadi [2023b]
- Continuum mechanics : Jidling et al. [2018]
- Helmholtz : $-\Delta u = \lambda u$ Albert and Rath [2020]
- (Non)stationary Maxwell : Wahlstrom et al. [2013], Jidling et al. [2017],Lange-Hegermann [2018]
- 3D wave equation, transport : H. et al. [2023, to appear]
- 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 s.t. $L(k_L(z, \cdot)) = 0 \ \forall z$; $\Delta = \sum_{i=1}^{d} \partial^2_{x_i x_i}$.

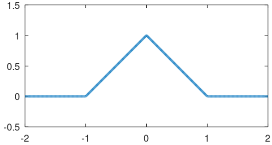
- Laplace : $\Delta u = 0$ Mendes and da Costa Júnior [2012], Ginsbourger et al. [2016]
- Heat : $\partial_t - D\Delta u = 0$ Albert and Rath [2020]
- Div/Curl : $\nabla \cdot u = 0$, $\nabla \times u = 0$ Scheuerer and Schlather [2012],Owhadi [2023b]
- Continuum mechanics : Jidling et al. [2018]
- Helmholtz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

Always based on representations of solutions of $Lu = 0$ of the form

$$u = Gf \quad \text{(Green's function/impulse response)}$$

Relaxing the definition of derivatives : Sobolev spaces

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



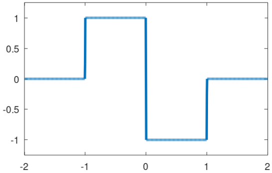


Figure 1 – Left : $h(x)$. Right : $h'(x)$ (hopefully).

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

Relaxing the definition of derivatives : Sobolev spaces

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

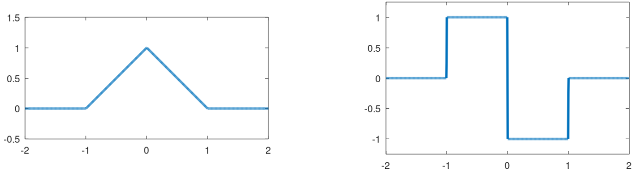


Figure 1 – Left : $h(x)$. Right : $h'(x)$ (hopefully).

Unfortunately, $h' \notin C^0$... but $h' \in L^2$ (finite energy) !
A function $g \in L^1_{loc}(\mathbb{R})$ is the weak derivative of h if for all $\varphi \in C_c^\infty(\mathbb{R})$,

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

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

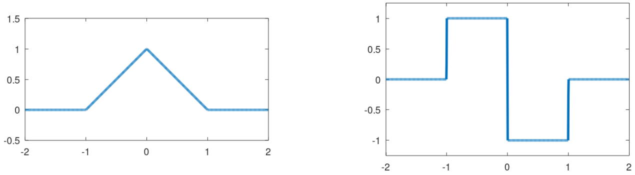


Figure 1 – Left : $h(x)$. Right : $h'(x)$ (hopefully).

Unfortunately, $h' \notin C^0$... but $h' \in L^2$ (finite energy) !
A function $g \in L^1_{loc}(\mathbb{R})$ is the weak derivative of h if for all $\varphi \in C_c^\infty(\mathbb{R})$,

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

We then define

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

Sobolev regularity of Gaussian random fields

Proposition 2 (H. [2022])

Let $(U_z)_{z\in\mathcal{D}} \sim GP(0,k)$ *be a measurable GP, we have equivalence between*
(i) $\mathbb{P}(\{\omega \in \Omega : U(\omega) \in H^m(\mathcal{D})\}) = 1$

Sobolev regularity of Gaussian random fields

Proposition 2 (H. [2022])

Let $(U_z)_{z \in \mathcal{D}} \sim GP(0, k)$ *be a measurable GP, we have equivalence between*
(i) $\mathbb{P}(\{\omega \in \Omega : U(\omega) \in H^m(\mathcal{D})\}) = 1$
(ii) For all $|\alpha| \leq m$, $\partial^{\alpha,\alpha} k \in L^2(\mathcal{D} \times \mathcal{D})$ *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, $Tr(\mathcal{E}_k^\alpha) = \int_{\mathcal{D}} \partial^{\alpha,\alpha} k(x,x) dx < +\infty.$

Sobolev regularity of Gaussian random fields

Proposition 2 (H. [2022])

Let $(U_z)_{z\in\mathcal{D}} \sim GP(0,k)$ *be a measurable GP, we have equivalence between*
(i) $\mathbb{P}(\{\omega \in \Omega : U(\omega) \in H^m(\mathcal{D})\}) = 1$
(ii) For all $|\alpha| \leq m$, $\partial^{\alpha,\alpha} k \in L^2(\mathcal{D}\times\mathcal{D})$ *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) There exists $(\phi_n) \subset L^2(\mathcal{D})$ *such that* $k(x,y) = \sum_n \phi_n(x)\phi_n(y)$ *in* $L^2(\mathcal{D}\times\mathcal{D})$. *Moreover, if* $|\alpha| \leq m$, *then* $\phi_n \in H^m(\mathcal{D})$ *and*

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

Sobolev regularity of Gaussian random fields

Proposition 2 (H. [2022])

Let $(U_z)_{z\in\mathcal{D}} \sim GP(0,k)$ *be a measurable GP, we have equivalence between*
(i) $\mathbb{P}(\{\omega \in \Omega : U(\omega) \in H^m(\mathcal{D})\}) = 1$
(ii) 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) There exists $(\phi_n) \subset L^2(\mathcal{D})$ *such that* $k(x,y) = \sum_n \phi_n(x)\phi_n(y)$ *in* $L^2(\mathcal{D}\times\mathcal{D})$. *Moreover, 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) $RKHS(k) \subset H^m(\mathcal{D})$ *and the imbedding* $\mathcal{I} : RKHS(k) \to H^m(\mathcal{D})$ *is Hilbert-Schmidt with* $||\mathcal{I}||_{HS}^2 = \sum_{|\alpha|\le m} Tr(\mathcal{E}_k^\alpha)$.

Sobolev regularity of Gaussian random fields : case $W^{m,p}, 1 < p < +\infty, m \in \mathbb{N}$

Proposition 3 (H. [2022])

Let $(U_z)_{z\in\mathcal{D}} \sim GP(0,k)$ *be a measurable GP, we have equivalence between*
(i) $\mathbb{P}(\{\omega \in \Omega : U(\omega) \in W^{m,p}(\mathcal{D})\}) = 1$
(ii) For all $|\alpha| \leq m$, $\partial^{\alpha,\alpha} k \in L^p(\mathcal{D}\times\mathcal{D})$ *and the integral operator* $\mathcal{E}_k^\alpha$

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

is symmetric, nonnegative and nuclear : there exists $(\phi_n^\alpha) \subset L^p(\mathcal{D})$ *such that* $\partial^{\alpha,\alpha} k(x,y) = \sum_n \phi_n^\alpha(x)\phi_n^\alpha(y)$ *dans* $L^p(\mathcal{D}\times\mathcal{D})$ *verifying*

$$\sum_{n=0}^{+\infty} ||\phi_n^\alpha||_p^2 < +\infty \quad (+\textit{refinements if } 1 \leq p \leq 2)$$

(iii) For all $|\alpha| \leq m$, $\partial^{\alpha,\alpha} k \in L^p(\mathcal{D}\times\mathcal{D})$ *and* $\int_{\mathcal{D}} \partial^{\alpha,\alpha} k(x,x)^{p/2} dx < +\infty$.

Outline of the talk

1. Constrained Gaussian process regression
2. Physics informed Gaussian processes
   - Distributional formulation of PDEs
   - Sobolev regularity of Gaussian random fields
3. Gaussian process regression for the 3D wave equation
   - GP priors for the 3D wave equation
   - Solving some inverse problems
   - Numerical applications

GP priors for the 3D wave equation (H. et al. [2023])

Consider the 3D wave equation ($\Delta := \partial^2_{xx} + \partial^2_{yy} + \partial^2_{zz}$)

$$\begin{cases} Lu & = \frac{1}{c^2}\partial^2_{tt}u - \Delta u = \Box u = 0, \quad (x,t) \in \mathbb{R}^3 \times \mathbb{R}^+ \\ u(x,0) & = u_0(x), \quad \partial_t u(x,0) = v_0(x). \end{cases} \tag{5}$$

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

$$u(x,t) = (F_t * v_0)(x) + (\dot{F}_t * u_0)(x). \tag{6}$$

GP priors for the 3D wave equation (H. et al. [2023])

Consider the 3D wave equation ($\Delta := \partial^2_{xx} + \partial^2_{yy} + \partial^2_{zz}$)

$$\begin{cases} Lu &= \frac{1}{c^2}\partial^2_{tt}u - \Delta u = \Box u = 0, \quad (x,t) \in \mathbb{R}^3 \times \mathbb{R}^+ \\ u(x,0) &= u_0(x), \quad \partial_t u(x,0) = v_0(x). \end{cases} \tag{5}$$

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

$$u(x,t) = (F_t * v_0)(x) + (\dot{F}_t * u_0)(x). \tag{6}$$

Assume that u_0 and v_0 are unknown $\rightarrow u_0 \sim GP(0, k_u)$ and $v_0 \sim GP(0, k_v)$, assumed independant. The u given by (6) is a centered GP with covariance function

$$k((x,t),(x',t')) = [(F_t \otimes F_{t'}) * k_v](x,x') + [(\dot{F}_t \otimes \dot{F}_{t'}) * k_u](x,x'). \tag{7}$$

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

Estimation of physical parameters and initial conditions

- Reconstruction of initial conditions : the Kriging 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

- Reconstruction of initial conditions : the Kriging 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 informations u_0 and v_0.
  Example : initial condition u_0 with compact support yield the prior over u_0

$$k_u(x, x') = k_u^0(x, x') \mathbb{1}_{B_R(x_0, R)}(x) \mathbb{1}_{B(x_0, R)}(x'). \tag{8}$$

  Hence, $(x_0, R) \in \theta_u$. Likewise for v_0 (We can also encode symetries).
  $\rightarrow$ these can be estimated via negative log marginal likelihood minimization.

Numerical application

Restrictive framework

Expensive convolutions (4D) $\rightarrow$ we assume radial symmetry over the initial conditions (explicit convolutions)

- Numerical resolution (finite differences in $[0,1]^3$) of the wave equation with $v_0 = 0$ and

$$u_0(x) = A\mathbb{1}_{[R_1,R_2]}(|x - x_0^*|)\left(1 + \cos\left(\frac{2\pi(|x - x_0^*| - \frac{R_1+R_2}{2})}{R_2 - R_1}\right)\right).$$

- Database generation : scattered sensors in $[0,1]^3$ (LHS).
  $B = \{u(x_i, t_j) + \epsilon_{ij}, 1 \leq i \leq N_C, 1 \leq j \leq N_T\}, N_C = 30,\ N_T = 75.$

Numerical application

Restrictive framework

Expensive convolutions (4D) $\rightarrow$ we assume radial symmetry over the initial conditions (explicit convolutions)

- Numerical resolution (finite differences in $[0,1]^3$) of the wave equation with $v_0 = 0$ and

$$u_0(x) = A\mathbb{1}_{[R_1,R_2]}(|x - x_0^*|)\left(1 + \cos\left(\frac{2\pi(|x - x_0^*| - \frac{R_1+R_2}{2})}{R_2 - R_1}\right)\right).$$

- Database generation : scattered sensors in $[0,1]^3$ (LHS).
$B = \{u(x_i,t_j) + \epsilon_{ij}, 1 \leq i \leq N_C, 1 \leq j \leq N_T\}, N_C = 30, \ N_T = 75.$
- Kriging with

$$k_u(x,x') = k_{5/2}(x - x') \times \mathbb{1}_{B_R(x_0,R)}(x)\mathbb{1}_{B(x_0,R)}(x').$$

Data visualization

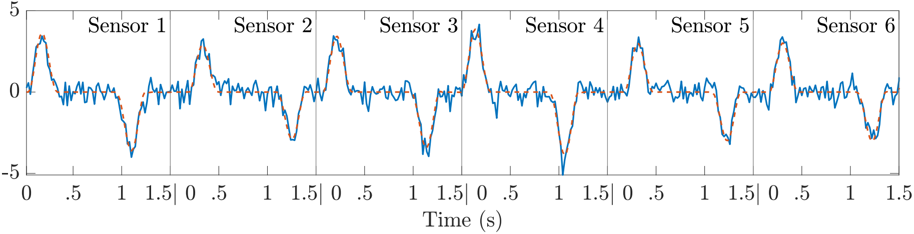


Figure 2 – Examples of captured signals. Red : noiseless. Blue : noisy.

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

Initial condition reconstruction

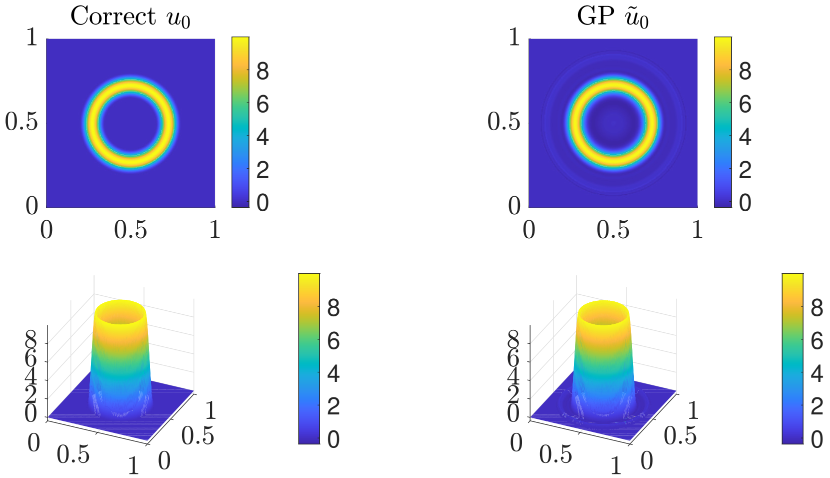


Figure 3 – True u_0 (left column) vs GPR u_0 (right column). 15 sensors were used. The images correspond to slices at $z = 0.5$.

Point source localization

Cas where $u_0 \equiv 0$ and the source v_0 is almost a Dirac mass at x_0^* : we use the kernels

$$k_v^R(x,x') = k_v(x,x') \frac{\mathbb{1}_{B(x_0,R)}(x)}{4\pi R^3/3} \frac{\mathbb{1}_{B(x_0,R)}}{4\pi R^3/3} \tag{9}$$

$$k((x,t),(x',t')) = [(F_t \otimes F_{t'}) * k_v^R](x,x') \tag{10}$$

with $R \ll 1$. Hyperparameters of k : (θ_v, x_0, R, c) We fix θ_v, R et c at the "right values" : $\mathcal{L}(\theta) = \mathcal{L}(x_0)$.

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

Minimize negative marginal likelihood ≡ GPS localization

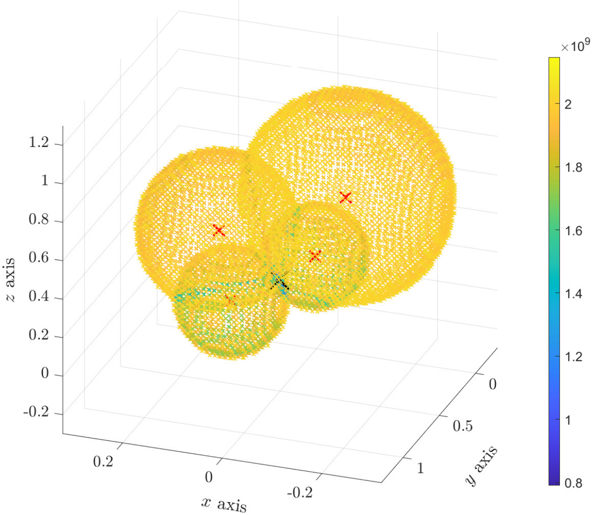


Figure : negative log marginal likelihood.

Display values : less than 2.035×10^9.

× : sensor locations.

× : source location.

See H. et al. [2023] for study/proofs.

Conclusion and perspectives

Some overall conclusions :

- GPR : at the intersection of machine learning, statistical and Bayesian approaches and functional analysis.
- Very explicit links can be established in between the different approaches and mathematical tools.

Conclusion and perspectives

Some overall conclusions :

- GPR : at the intersection of machine learning, statistical and Bayesian approaches and functional analysis.
- Very explicit links can be established in between the different approaches and mathematical tools.

Some research perspectives :

- Insert the Sobolev regularity results in the analysis of GPR for PDEs.
- Current research : draw links between numerical methods for PDEs (finite elements, finite differences) and some GPR regimes.

Thank you for your attention !

Contact : henderso@insa-toulouse.fr

References I

C. G. Albert and K. Rath. Gaussian process regression for data fulfilling linear differential equations with localized sources. *Entropy*, 22(2), 2020. ISSN 1099-4300. URL https://www.mdpi.com/1099-4300/22/2/152.

Yifan Chen, Bamdad Hosseini, Houman Owhadi, and Andrew M. Stuart. Solving and learning nonlinear PDEs with Gaussian processes. *Journal of Computational Physics*, 447 :110668, 2021. ISSN 0021-9991. doi : https://doi.org/10.1016/j.jcp.2021.110668. URL https://www.sciencedirect.com/science/article/pii/S0021999121005635.

D. Ginsbourger, O. Roustant, and N. Durrande. On degeneracy and invariances of random fields paths with applications in Gaussian process modelling. *Journal of Statistical Planning and Inference*, page 170 :117 – 128, 2016.

Iain H. Sobolev regularity of Gaussian random fields. working paper or preprint, October 2022. URL https://hal.science/hal-03769576.

References II

Iain H., Pascal Noble, and Olivier Roustant. Wave equation-tailored Gaussian process regression with applications to related inverse problems. working paper or preprint, January 2023. URL `https://hal.science/hal-03941939`.

Iain H., Pascal Noble, and Olivier Roustant. Characterization of the second order random fields subject to linear distributional PDE constraints. *Bernoulli*, 2023, to appear. URL `https://hal.science/hal-03770715`.

C. Jidling, N. Wahlström, A. Wills, and T. B. Schön. Linearly constrained Gaussian processes. In I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc., 2017. URL `https://proceedings.neurips.cc/paper/2017/file/71ad16ad2c4d81f348082ff6c4b20768-Paper.pdf`.

C. Jidling, J. Hendriks, N. Wahlstrom, A. Gregg, T. Schon, C. Wensrich, and A. Wills. Probabilistic modelling and reconstruction of strain. *Nuclear Instruments & Methods in Physics Research Section B-beam Interactions With Materials and Atoms*, 436 :141–155, 2018.

M. Lange-Hegermann. Algorithmic linearly constrained Gaussian processes. In S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 31. Curran Associates, Inc., 2018. URL `https://proceedings.neurips.cc/paper/2018/file/68b1fbe7f16e4ae3024973f12f3cb313-Paper.pdf`.

A. F. López-Lopera, N. Durrande, and M. Álvarez. Physically-inspired Gaussian process models for post-transcriptional regulation in drosophila. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 18 :656–666, 2021.

References IV

Fábio Macêdo Mendes and Edson Alves da Costa Júnior. Bayesian inference in the numerical solution of Laplace's equation. *AIP Conference Proceedings*, 1443(1) :72–79, 2012. doi : 10.1063/1.3703622. URL https://aip.scitation.org/doi/abs/10.1063/1.3703622.

Houman Owhadi. Do ideas have shape ? idea registration as the continuous limit of artificial neural networks. *Physica D : Nonlinear Phenomena*, 444 :133592, 2023a. ISSN 0167-2789. doi : https://doi.org/10.1016/j.physd.2022.133592. URL https://www.sciencedirect.com/science/article/pii/S0167278922002962.

Houman Owhadi. Gaussian process hydrodynamics, 2023b.

C. E. Rasmussen and C.K.I. Williams. *Gaussian Processes for Machine Learning*. the MIT Press, 2006. ISBN 026218253X. URL www.{G}aussianProcess.org/gpml.

M. Scheuerer and M. Schlather. Covariance models for divergence-free and curl-free random vector fields. *Stochastic Models*, 28 :433 – 451, 2012.

References V

N. Wahlstrom, M. Kok, T. B. Schön, and F. Gustafsson. Modeling magnetic fields using Gaussian processes. *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 3522–3526, 2013.

Holger Wendland. *Scattered data approximation*, volume 17. Cambridge university press, 2004.

M. Álvarez, D. Luengo, and N. D. Lawrence. Latent force models. In D. van Dyk and M. Welling, editors, *Proceedings of the Twelth International Conference on Artificial Intelligence and Statistics*, volume 5 of *Proceedings of Machine Learning Research*, pages 9–16, Hilton Clearwater Beach Resort, Clearwater Beach, Florida USA, 16–18 Apr 2009. PMLR. URL https://proceedings.mlr.press/v5/alvarez09a.html.

GPR and neural networks

- Some Gaussian processes as limits of one layer, infinite neurons NN (Rasmussen and Williams [2006], Section 4.2.3).
- NN as a kernel method with a kernel learnt from data (Owhadi [2023a] ; Mallat, collège de France).
- GPR : "only" current alternative to (physics informed) neural networks (PINNs), see Chen et al. [2021] for a discussion.

Radial symmetry formulas

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

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

Details on F_t and $\dot{F}_t$

$\longrightarrow F_t = \sigma_{ct}/4\pi c^2 t$ means that

Details on F_t and $\dot{F}_t$

$\longrightarrow F_t = \sigma_{ct}/4\pi c^2 t$ means that

$$\int_{\mathbb{R}^3} f(x)F_t(dx) = \frac{t}{4\pi}\int_0^{2\pi}\int_0^{\pi} f(ct\gamma)\sin\theta d\theta d\varphi = \frac{t}{4\pi}\int_{S(0,1)} f(ct\gamma)d\Omega$$

where γ is the unit length vector $\gamma = (\sin\theta\cos\varphi, \sin\theta\sin\varphi, \cos\theta)^T$.

Details on F_t and $\dot{F}_t$

$\longrightarrow$ $F_t = \sigma_{ct}/4\pi c^2 t$ means that

$$\int_{\mathbb{R}^3} f(x) F_t(dx) = \frac{t}{4\pi} \int_0^{2\pi} \int_0^{\pi} f(ct\gamma) \sin\theta d\theta d\varphi = \frac{t}{4\pi} \int_{S(0,1)} f(ct\gamma) d\Omega$$

where γ is the unit length vector $\gamma = (\sin\theta\cos\varphi, \sin\theta\sin\varphi, \cos\theta)^T$.

$\longrightarrow$ Convolution between functions and measures :

$$(f * g)(x) = \int_{\mathbb{R}^3} g(x-y) f(y) dy \qquad (\mu * g)(x) = \int_{\mathbb{R}^3} g(x-y) \mu(dy)$$

Details on F_t and $\dot{F}_t$

$\longrightarrow$ $F_t = \sigma_{ct}/4\pi c^2 t$ means that

$$\int_{\mathbb{R}^3} f(x)F_t(dx) = \frac{t}{4\pi}\int_0^{2\pi}\int_0^{\pi} f(ct\gamma)\sin\theta d\theta d\varphi = \frac{t}{4\pi}\int_{S(0,1)} f(ct\gamma)d\Omega$$

where γ is the unit length vector $\gamma = (\sin\theta\cos\varphi, \sin\theta\sin\varphi, \cos\theta)^T$.

$\longrightarrow$ Convolution between functions and measures :

$$(f * g)(x) = \int_{\mathbb{R}^3} g(x-y)f(y)dy \qquad (\mu * g)(x) = \int_{\mathbb{R}^3} g(x-y)\mu(dy)$$

$\longrightarrow$ $\dot{F}_t = \partial_t F_t$ means that

$$\begin{aligned}\langle \dot{F}_t, f\rangle &= \partial_t \int f(x)dF_t(x)\\ &= \frac{1}{4\pi}\int_{S(0,1)} f(ct\gamma)d\Omega + \frac{c}{4\pi}\int_{S(0,1)} \nabla f(ct\gamma)\cdot\gamma d\Omega\end{aligned}$$

Extension to non linear PDEs

- Non linear constraints on $k(z, \cdot)$: not realistic (+ GP interpretation not valid).
- Alternative : in Chen et al. [2021], the nonlinear PDE constraint in applied pointwise on $\tilde{m}$: modification of the RKHS optimization problem as

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

  Generalizes an approach desribed in Wendland [2004].
- Coupling of this approach with strict linear constraints : Owhadi [2023b] (div/curl/périodicity).