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 constraints

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

Why use GPR?
- Mathematically tractable and interpretable.
- GPR as orthogonal projections in the RKHS → more familiar in the PDE community.

Now: understand behaviour of GPR w.r.t. PDEs:
(i) linear, (ii) Sobolev.
Gaussian process regression (GPR)

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

\( \tilde{m} \) and \( \tilde{k} \) are given by the Kriging formulas.
- \( \forall z \in \mathcal{D} \), approximate \( u(z) \) with \( \tilde{m}(z) : u(z) \sim \tilde{m}(z) + \text{interpolation} \).
Gaussian process regression (GPR)

- Unknown function: \( u : \mathcal{D} \rightarrow \mathbb{R} \), obs. \( B = \{u(z_1), \ldots, u(z_n)\} \)
- Model \( u \) as a realization of a Gaussian process \( (U_z)_{z \in \mathcal{D}} \sim \text{GP}(0, k) \).
- Condition \( U \) on data: \( V_z = [U_z|U_{z_1} = u(z_1), \ldots, U_{z_n} = u(z_n)] \), yielding
  \[
  V_z \sim \text{GP}(\tilde{m}, \tilde{k}).
  \]
  \( \tilde{m} \) and \( \tilde{k} \) are given by the Kriging formulas.
- \( \forall z \in \mathcal{D} \), approximate \( u(z) \) with \( \tilde{m}(z) : u(z) \sim \tilde{m}(z) + \text{interpolation} \).

Why use GPR?

- Mathematically tractable and interpretable.
- GPR as orthogonal projections in the RKHS
  \( \rightarrow \) more familiar in the PDE community.
Gaussian process regression (GPR)

- Unknown function: \( u : \mathcal{D} \rightarrow \mathbb{R} \), obs. \( B = \{ u(z_1), ..., u(z_n) \} \)
- Model \( u \) as a realization of a Gaussian process \((U_z)_{z \in \mathcal{D}} \sim GP(0, k)\).
- 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.
- \( \forall z \in \mathcal{D}, \) approximate \( u(z) \) with \( \tilde{m}(z) : u(z) \sim \tilde{m}(z) + \) interpolation.

Why use GPR?
- Mathematically tractable and interpretable.
- GPR as orthogonal projections in the RKHS → more familiar in the PDE community.

Now: understand behaviour of GPR w.r.t. PDEs:
  \((i)\) linear, \((ii)\) Sobolev.
Assume that $Lu = 0$, $L$ linear. The GPR procedure is adapted to this constraint if $L\tilde{m} = 0$. More generally: incorporate prior knowledge in the GP prior.
Linearly constrained GPR

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

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

$$\tilde{m}(z) = k(Z, z)^T K^{-1} u_{obs} \in \text{Span}(k(z_1, \cdot), ..., k(z_n, \cdot)).$$
Assume that $Lu = 0$, $L$ linear. The GPR procedure is adapted to this constraint if $L\tilde{m} = 0$.

Denote $u_{obs} = (u(z_1), \ldots, u(z_n))$ 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

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

- $L\tilde{m} = 0$ 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.

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

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

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

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

(1)

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

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.

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

(1)

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

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)

→ 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 $\mathbb{R}$. 

I. Henderson (INSA Toulouse/IMT)  
GPR and PDEs  
MASCOT-NUM, 04/04/2023  
9 / 26
Let $\mathcal{D}$ be an open set of $\mathbb{R}$. Note $D$ the single derivative operator.

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

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

(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

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

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

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

\[ \forall \varphi \in C^\infty_c(\mathcal{D}), \int_{\mathcal{D}} Lu(x) \varphi(x) \, dx = \int_{\mathcal{D}} u(x) L^* \varphi(x) \, dx = 0. \]
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,

\[ \forall \varphi \in C^\infty_c(\mathcal{D}), \int_{\mathcal{D}} L u(x) \varphi(x) \, dx = \int_{\mathcal{D}} u(x) L^* \varphi(x) \, dx = 0. \]

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

\[ \forall \varphi \in C^\infty_c(\mathcal{D}), \int_{\mathcal{D}} u(x) L^* \varphi(x) \, dx = 0 \quad (4) \]
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,

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

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

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

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

\[ \int_K |u| < +\infty \quad \text{for all compact set} \quad K \subset \mathcal{D}. \]
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 $a_\alpha \in C^{\alpha}(\mathcal{D})$. Let $U = (U_z)_{z \in \mathcal{D}}$ be a measurable centered second order random field with covariance function $k(z, z')$. Assume that $\sigma : z \mapsto k(z, z)^{1/2} \in L^1_{\text{loc}}(\mathcal{D})$.

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

2) The following statements are equivalent:
   - $\mathbb{P}(\{ \omega \in \Omega : L(U(\omega)) = 0 \text{ in the sense of distributions} \}) = 1$,
   - $\forall z \in \mathcal{D}, L(k(z, \cdot)) = 0 \text{ in the sense of distributions}$.
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 $a_\alpha \in C^{\alpha}(\mathcal{D})$. Let $U = (U_z)_{z \in \mathcal{D}}$ be a measurable centered second order random field with covariance function $k(z, z')$. Assume that $\sigma : z \mapsto k(z, z)^{1/2} \in L^1_{\text{loc}}(\mathcal{D})$.

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

2) The following statements are equivalent:
   - $\mathbb{P}(\{\omega \in \Omega : L(U(\omega)) = 0 \text{ in the sense of distributions} \}) = 1$,
   - $\forall z \in \mathcal{D}, L(k(z, \cdot)) = 0 \text{ in the sense of distributions}$.

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

- **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

$$u = Gf$$  (Green’s function/impulse response)
Some functions are "almost" differentiable: \( h(x) = \max(0, 1 - |x|) \).

Unfortunately, \( h' \not\in C^0 \) ... but \( h' \in L^2 \) (finite energy)!

**Figure 1** – Left: \( h(x) \). Right: \( h'(x) \) (hopefully).
Relaxing the definition of derivatives: Sobolev spaces

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

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 $\varphi \in C_c^\infty(\mathbb{R})$,

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

Figure 1 – Left: $h(x)$. Right: $h'(x)$ (hopefully).
Relaxing the definition of derivatives: Sobolev spaces

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

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{figure1.png}
\caption{Left: \( h(x) \). Right: \( h'(x) \) (hopefully).}
\end{figure}

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

A function \( g \in L^1_{\text{loc}}(\mathbb{R}) \) is the weak derivative of \( h \) if for all \( \varphi \in C^\infty_c(\mathbb{R}) \),

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

We then define

\[
H^1(\mathbb{R}) := \{ u \in L^2(\mathbb{R}) : u' \text{ exists in the weak sense and } u' \in L^2(\mathbb{R}) \},
\]

\[
H^m(\mathcal{D}) := \{ u \in L^2(\mathcal{D}) : \forall |\alpha| \leq m, \partial^\alpha u \text{ exists ITWS and } \partial^\alpha u \in L^2(\mathcal{D}) \}.
\]
Proposition 2 (H. [2022])

Let \((U_z)_{z \in \mathcal{D}} \sim \text{GP}(0, k)\) be a measurable GP, we have equivalence between

(i) \(\mathbb{P}(\{\omega \in \Omega : U(\omega) \in H^m(\mathcal{D})\}) = 1\)
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, \(Tr(\mathcal{E}_k^\alpha) = \int_{\mathcal{D}} \partial^{\alpha,\alpha} k(x, x)dx < +\infty\).
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, \(\text{Tr}(\mathcal{E}_k^\alpha) = \int_{\mathcal{D}} \partial^{\alpha,\alpha} k(x, x)dx < +\infty\).

(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

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

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

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

(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

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

(iv) \(\text{RKHS}(k) \subset H^m(\mathcal{D})\) and the imbedding \(\mathcal{I} : \text{RKHS}(k) \rightarrow H^m(\mathcal{D})\) is Hilbert-Schmidt with \(\|\mathcal{I}\|_{HS}^2 = \sum_{|\alpha| \leq m} \text{Tr}(\mathcal{E}_k^\alpha)\).
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

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

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

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

(5)

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

\[
u(x, t) = (F_t * v_0)(x) + (\dot{F}_t * u_0)(x).
\]

(6)
Consider the 3D wave equation \((\Delta := \partial^2_{xx} + \partial^2_{yy} + \partial^2_{zz})\)

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

(5)

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

\[
u(x, t) = (F_t * v_0)(x) + (\dot{F}_t * u_0)(x).
\]

(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 independent. The \(u\) given by (6) is a centered GP with covariance function

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

(7)

The kernel \(k\) verifies \(\Box k((x, t), \cdot) = 0\) for all \((x, t) \in \mathbb{R}^3 \times \mathbb{R}^+\).
Reconstruction of initial conditions: the Kriging mean verifies $\hat{m} = 0$. Hence,

$$\hat{m}(\cdot, t = 0) \simeq u_0, \quad \partial_t \hat{m}(\cdot, t = 0) \simeq v_0.$$
Reconstruction of initial conditions: the Kriging mean verifies $\tilde{m} = 0$. 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 $c, \theta_u$ and $\theta_v$; $\theta_u$ and $\theta_v$ may contain physical informations $u_0$ and $v_0$.

Example: initial condition $u_0$ with compact support yield the prior over $u_0$

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

Hence, $(x_0, R) \in \theta_u$. Likewise for $v_0$ (We can also encode symettries). These can be estimated via negative log marginal likelihood minimization.
Numerical application

Restrictive framework

Expensive convolutions (4D) → 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).

  $$B = \{u(x_i, t_j) + \epsilon_{ij}, 1 \leq i \leq N_C, 1 \leq j \leq N_T\}, \quad N_C = 30, \quad N_T = 75.$$
Numerical application

Restrictive framework

Expensive convolutions (4D) → 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).

$$B = \{u(x_i, t_j) + \epsilon_{ij}, 1 \leq i \leq N_C, 1 \leq j \leq N_T\}, \quad N_C = 30, \quad N_T = 75.$$

- Kriging with

$$k_u(x, x') = k_{5/2}(x - x') \times \mathbb{1}_{B_R(x_0, R)}(x) \mathbb{1}_{B(x_0, R)}(x').$$
Figure 2 – Examples of captured signals. Red : noiseless. Blue : noisy.
### Table 1 – Hyperparameter estimation and relative errors

<table>
<thead>
<tr>
<th>$N_{\text{sensors}}$</th>
<th>3</th>
<th>5</th>
<th>10</th>
<th>15</th>
<th>20</th>
<th>25</th>
<th>30</th>
<th>Target</th>
</tr>
</thead>
<tbody>
<tr>
<td>$</td>
<td>\hat{x}_0 - x_0^*</td>
<td>$</td>
<td>0.204</td>
<td>0.003</td>
<td>0.004</td>
<td>0.008</td>
<td>0.003</td>
<td>0.004</td>
</tr>
<tr>
<td>$\hat{R}$</td>
<td>0.386</td>
<td>0.432</td>
<td>0.462</td>
<td>0.431</td>
<td>0.414</td>
<td>0.471</td>
<td>0.452</td>
<td>0.25</td>
</tr>
<tr>
<td>$</td>
<td>\hat{c} - c^*</td>
<td>$</td>
<td>0.084</td>
<td>0.004</td>
<td>0.005</td>
<td>0.005</td>
<td>0.006</td>
<td>0.001</td>
</tr>
<tr>
<td>$\hat{\sigma}_{\text{noise}}^2$</td>
<td>0.917</td>
<td>0.879</td>
<td>0.93</td>
<td>0.99</td>
<td>0.361</td>
<td>0.988</td>
<td>0.377</td>
<td>0.2025</td>
</tr>
<tr>
<td>$\hat{\ell}$</td>
<td>0.02</td>
<td>0.02</td>
<td>0.025</td>
<td>0.02</td>
<td>0.035</td>
<td>0.024</td>
<td>0.032</td>
<td>$\sim$ 0.05</td>
</tr>
<tr>
<td>$\hat{\sigma}^2$</td>
<td>2.367</td>
<td>3.513</td>
<td>4.903</td>
<td>3.168</td>
<td>4.446</td>
<td>4.619</td>
<td>4.79</td>
<td>Unknown</td>
</tr>
<tr>
<td>$e_{1,\text{rel}}^u$</td>
<td>1.275</td>
<td>0.157</td>
<td>0.128</td>
<td>0.168</td>
<td>0.11</td>
<td>0.103</td>
<td>0.248</td>
<td>0</td>
</tr>
<tr>
<td>$e_{2,\text{rel}}^u$</td>
<td>1.056</td>
<td>0.095</td>
<td>0.082</td>
<td>0.124</td>
<td>0.088</td>
<td>0.064</td>
<td>0.213</td>
<td>0</td>
</tr>
<tr>
<td>$e_{\infty,\text{rel}}^u$</td>
<td>1.037</td>
<td>0.132</td>
<td>0.128</td>
<td>0.198</td>
<td>0.136</td>
<td>0.101</td>
<td>0.321</td>
<td>0</td>
</tr>
</tbody>
</table>
Figure 3 – True $u_0$ (left column) vs GPR $u_0$ (right column). 15 sensors were used. The images correspond to slices at $z = 0.5$. 
Cas where $u_0 \equiv 0$ and the source $v_0$ is almost a Dirac mass at $x_0^*$: we use the kernels

$$k^R_v(x, x') = k_v(x, x') \frac{1_B(x_0, R)(x)}{4\pi R^3/3} \frac{1_B(x_0, R)}{4\pi R^3/3}$$

(9)

$$k((x, t), (x', t')) = [(F_t \otimes F_{t'}) \ast k^R_v](x, x')$$

(10)

with $R \ll 1$. Hyperparameters of $k: (\theta_v, x_0, R, c)$ We fix $\theta_v, R$ 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 $\equiv$ GPS localization

Figure: negative log marginal likelihood.

Display values: less than $2.035 \times 10^9$.

$\times$: sensor locations.

$\times$: 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


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


Houman Owhadi. Gaussian process hydrodynamics, 2023b.


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

\[
[(F_t \otimes F_{t'}) \ast k_v](x, x') = \frac{\text{sgn}(tt')}{16c^2rr'} \sum_{\varepsilon, \varepsilon' \in \{-1, 1\}} \varepsilon \varepsilon' K_v((r + \varepsilon ct)^2, (r' + \varepsilon' c|t'|)^2)
\]

\[
[(\dot{F}_t \otimes \dot{F}_{t'}) \ast k_u](x, x') = \frac{1}{4rr'} \sum_{\varepsilon, \varepsilon' \in \{-1, 1\}} (r + \varepsilon ct)(r' + \varepsilon' c|t'|)k_u((r + \varepsilon ct)^2, (r' + \varepsilon' c|t'|)^2)
\]
Details on $F_t$ and $\dot{F}_t$

$\rightarrow F_t = \sigma_{ct}/4\pi c^2 t$ means that
Details on $F_t$ and $\dot{F}_t$

\[ \rightarrow F_t = \sigma_{ct}/4\pi c^2 t \] 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 $\gamma$ is the unit length vector $\gamma = (\sin \theta \cos \varphi, \sin \theta \sin \varphi, \cos \theta)^T$. 
Details on $F_t$ and $\dot{F}_t$

$\rightarrow F_t = \sigma_{ct}/4\pi c^2 t$ 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 $\gamma$ is the unit length vector $\gamma = (\sin \theta \cos \varphi, \sin \theta \sin \varphi, \cos \theta)^T$.

$\rightarrow$ Convolution between functions and measures:

$$(f \ast g)(x) = \int_{\mathbb{R}^3} g(x - y)f(y)dy \quad (\mu \ast g)(x) = \int_{\mathbb{R}^3} g(x - y)\mu(dy)$$
Details on $F_t$ and $\dot{F}_t$

$\rightarrow F_t = \sigma_{ct}/4\pi c^2 t$ 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 $\gamma$ is the unit length vector $\gamma = (\sin \theta \cos \varphi, \sin \theta \sin \varphi, \cos \theta)^T$.

$\rightarrow$ Convolution between functions and measures:

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

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

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

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

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