UPPSALA
UNIVERSITET

Linearly and nonlinearly constrained Gaussian processes

Niklas Wahlström, Uppsala University, Sweden

Joint work with
Carl Jidling, **Philipp Pilar**, **Thomas Schön**, (Uppsala University),
Adrian Wills, **Johannes Hendriks**, **Alexander Gregg**, **Chris Wensrich** (University of Newcastle, Australia),

Lifting Inference with Kernel Embeddings (LIKE22) – Winter school and workshop (online), January 11, 2022

Outline

Aim: Introduce constrained Gaussian process regression and demonstrate it on a few examples.

1. **GP basics**
2. Linear constraints
3. Strain field reconstruction
4. Nonlinear constraints

GP basics

Distribution over functions

$$\begin{bmatrix} f(\mathbf{x}_1) \\ \vdots \\ f(\mathbf{x}_N) \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \mu(\mathbf{x}_1) \\ \vdots \\ \mu(\mathbf{x}_N) \end{bmatrix}, \underbrace{\begin{bmatrix} k(\mathbf{x}_1, \mathbf{x}_1) & \cdots & k(\mathbf{x}_1, \mathbf{x}_N) \\ \vdots & & \vdots \\ k(\mathbf{x}_N, \mathbf{x}_1) & \cdots & k(\mathbf{x}_N, \mathbf{x}_N) \end{bmatrix}}_{K} \right)$$

Gram matrix

Uniquely specified by mean and covariance function

$$\mu(\mathbf{x}_i) = \mathbb{E}\left[f(\mathbf{x}_i)\right]$$
$$k(\mathbf{x}_i, \mathbf{x}_j) = \mathrm{Cov}\left[f(\mathbf{x}_i),\ f(\mathbf{x}_j)\right]$$

Formally

$$f(\mathbf{x}) \sim \mathcal{GP}\left(\mu(\mathbf{x}),\ k(\mathbf{x}, \mathbf{x}')\right)$$

GP basics – prediction

Let

$$y_i = f(\mathbf{x}_i) + \varepsilon, \qquad \varepsilon \sim \mathcal{N}\left(0,\ \sigma^2\right)$$
$$\mathbf{y} = [y_1, y_2 \ldots, y_N]^\mathsf{T}$$

Then

$$\begin{bmatrix} \mathbf{y} \\ f(\mathbf{x}_*) \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \mathbf{0} \\ 0 \end{bmatrix}, \begin{bmatrix} K + \sigma^2 I & \mathbf{k} \\ \mathbf{k}^\mathsf{T} & k(\mathbf{x}_*, \mathbf{x}_*) \end{bmatrix} \right)$$
$$K_{ij} = k(\mathbf{x}_i, \mathbf{x}_j)$$
$$\mathbf{k}_i = k(\mathbf{x}_i, \mathbf{x}_*)$$

and

$$\mathbb{E}\left[f(\mathbf{x}_*)|\mathbf{y}\right] = \mathbf{k}^\mathsf{T}(K + \sigma^2 I)^{-1}\mathbf{y}$$
$$\mathbb{V}\left[f(\mathbf{x}_*)|\mathbf{y}\right] = k(\mathbf{x}_*, \mathbf{x}_*) - \mathbf{k}^\mathsf{T}(K + \sigma^2 I)^{-1}\mathbf{k}$$

GP basics – linear operator measurements

Linear operator measurements

$$y = \mathcal{L}_{\mathbf{x}} f(\mathbf{x}) + \varepsilon$$

Then

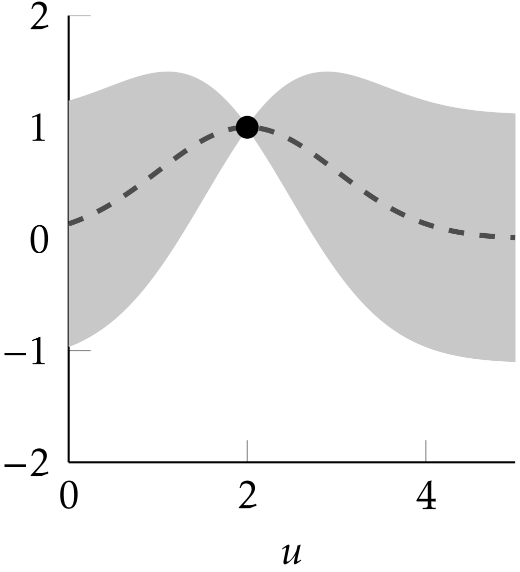
$$\mathbb{E}\left[f(\mathbf{x}_*)|\mathbf{y}\right] = \mathbf{q}^{\mathsf{T}}(Q + \sigma^2 I)^{-1}\mathbf{y}$$
$$\mathbb{V}\left[f(\mathbf{x}_*)|\mathbf{y}\right] = k(\mathbf{x}_*, \mathbf{x}_*) - \mathbf{q}^{\mathsf{T}}(Q + \sigma^2 I)^{-1}\mathbf{q}$$

where

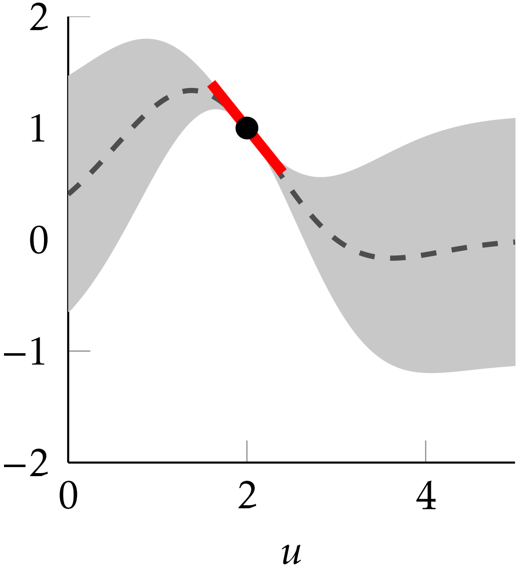
$$Q_{ij} = \mathcal{L}_{\mathbf{x}_i}\mathcal{L}_{\mathbf{x}_j} k(\mathbf{x}_i, \mathbf{x}_j)$$
$$\mathbf{q}_i = \mathcal{L}_{\mathbf{x}_i} k(\mathbf{x}_i, \mathbf{x}_*)$$

Example:

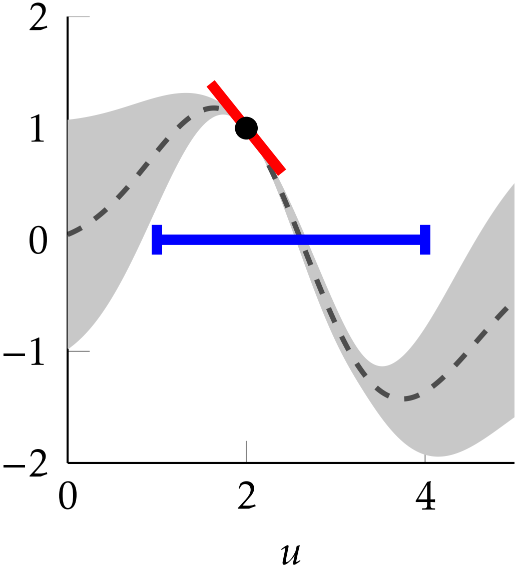
$$y_i = \int_{a_i}^{b_i} f(x)dx \quad \Rightarrow \begin{cases} Q_{ij} = \int_{a_i}^{b_i}\int_{a_j}^{b_j} k(x, x')\, dx'\, dx \\ \mathbf{q}_i = \int_{a_i}^{b_i} k(x, x_*)\, dx \end{cases}$$


2
1
0
−1
−2
0
2
4
u

GP basics – linear operator measurements



GP basics – linear operator measurements



Outline

Multivariate GP – constraint incorporation

Toy Example

Consider a Gaussian process

$$\mathbf{f}(\mathbf{x}) \sim \mathcal{GP}\left(\boldsymbol{\mu}(\mathbf{x}),\ \mathbf{K}(\mathbf{x}, \mathbf{x}')\right)$$

with two-dimensional input and two-dimensional output

$$\mathbf{f}(\mathbf{x}) = \begin{bmatrix} f_1(\mathbf{x}) \\ f_2(\mathbf{x}) \end{bmatrix}, \qquad \mathbf{x} = \begin{bmatrix} x \\ y \end{bmatrix}$$

Assume that we know from the physics that the all samples from the GP prior should obey the constraint

$$\frac{\partial f_1}{\partial x} + \frac{\partial f_2}{\partial y} = 0 \quad \Leftrightarrow \quad \underbrace{\begin{bmatrix} \frac{\partial}{\partial x} & \frac{\partial}{\partial y} \end{bmatrix}}_{\mathscr{F}_{\mathbf{x}}} \mathbf{f}(\mathbf{x}) = 0$$

How can we model the covariance function $\mathbf{K}(\mathbf{x}, \mathbf{x}')$ such that this constraint is guaranteed to be obeyed?

Multivariate GP – constraint incorporation

Assume linear constraints

$$\mathscr{F}_{\mathbf{x}}\mathbf{f}(\mathbf{x}) = \mathbf{0}$$

Let $\mathbf{f}(\mathbf{x}) = \mathscr{G}_{\mathbf{x}}\mathbf{g}(\mathbf{x})$, where $\mathbf{g}(\mathbf{x}) \sim \mathcal{GP}\left(\boldsymbol{\mu}_{\mathbf{g}}(\mathbf{x}),\ \mathbf{K}_{\mathbf{g}}(\mathbf{x}, \mathbf{x}')\right)$

$$\mathbf{f}(\mathbf{x}) = \mathscr{G}_{\mathbf{x}}\mathbf{g}(\mathbf{x}) \sim \mathcal{GP}\left(\mathscr{G}_{\mathbf{x}}\ \boldsymbol{\mu}_{\mathbf{g}}(\mathbf{x}),\ \mathscr{G}_{\mathbf{x}}\mathbf{K}_{\mathbf{g}}(\mathbf{x}, \mathbf{x}')\mathscr{G}_{\mathbf{x}'}^{\mathsf{T}}\right)$$

Then

$$\mathscr{F}_{\mathbf{x}}\mathscr{G}_{\mathbf{x}}\mathbf{g}(\mathbf{x}) = \mathbf{0}$$

Arbitrary $\mathbf{g}(\mathbf{x})$

$$\Rightarrow \mathscr{F}_{\mathbf{x}}\mathscr{G}_{\mathbf{x}} = \mathbf{0}$$

Find $\mathscr{G}_{\mathbf{x}}$

Carl Jidling, Niklas Wahlstöm, Adrian Wills, Thomas B. Schön. **Linearly constrained Gaussian processes**. *Advances in Neural Information Processing Systems (NIPS)*,Long Beach, CA, USA, December, 2017.

Multivariate GP – constraint incorporation

Toy Example (cont.)

We consider the function

$$\mathbf{f}(\mathbf{x}) = \begin{bmatrix} f_1(\mathbf{x}) \\ f_2(\mathbf{x}) \end{bmatrix}, \qquad \mathbf{x} = \begin{bmatrix} x \\ y \end{bmatrix}$$

and the constraint

$$\frac{\partial f_1}{\partial x} + \frac{\partial f_2}{\partial y} = 0 \quad \Leftrightarrow \quad \underbrace{\begin{bmatrix} \frac{\partial}{\partial x} & \frac{\partial}{\partial y} \end{bmatrix}}_{\mathscr{F}_{\mathbf{x}}} \mathbf{f}(\mathbf{x}) = 0$$

Need $\mathscr{G}_{\mathbf{x}}$ such that $\mathscr{F}_{\mathbf{x}}\mathscr{G}_{\mathbf{x}} = \mathbf{0}$. One option is

$$\mathscr{G}_{\mathbf{x}} = \begin{bmatrix} -\frac{\partial}{\partial y} \\ \frac{\partial}{\partial x} \end{bmatrix}$$

since

$$\mathscr{F}_{\mathbf{x}}\mathscr{G}_{\mathbf{x}} = \begin{bmatrix} \frac{\partial}{\partial x} & \frac{\partial}{\partial y} \end{bmatrix} \begin{bmatrix} -\frac{\partial}{\partial y} \\ \frac{\partial}{\partial x} \end{bmatrix} = -\frac{\partial^2}{\partial x \partial y} + \frac{\partial^2}{\partial y \partial x} = 0.$$

Algorithm idea – toy example

Step 1: Assume that $\mathscr{G}_{\mathbf{x}}$ contains the same operators as $\mathscr{F}_{\mathbf{x}}$

$$\mathscr{G}_{\mathbf{x}} = \begin{bmatrix} \gamma_{11} & \gamma_{12} \\ \gamma_{21} & \gamma_{22} \end{bmatrix} \begin{bmatrix} \frac{\partial}{\partial x} \\ \frac{\partial}{\partial y} \end{bmatrix}$$

Step 2: Expand

$$\mathscr{F}_{\mathbf{x}}\mathscr{G}_{\mathbf{x}} = \begin{bmatrix} \frac{\partial}{\partial x} & \frac{\partial}{\partial y} \end{bmatrix} \begin{bmatrix} \gamma_{11} & \gamma_{12} \\ \gamma_{21} & \gamma_{22} \end{bmatrix} \begin{bmatrix} \frac{\partial}{\partial x} \\ \frac{\partial}{\partial y} \end{bmatrix}$$

$$= \gamma_{11} \frac{\partial^2}{\partial x^2} + (\gamma_{12} + \gamma_{21}) \frac{\partial^2}{\partial x \partial y} + \gamma_{22} \frac{\partial^2}{\partial y^2}$$

Algorithm idea – toy example

Step 3: We need

$$\begin{cases} \gamma_{11} &= 0 \\ \gamma_{12} &= -\gamma_{21} \\ \gamma_{22} &= 0 \end{cases}$$

Step 4: Choosing $\gamma_{21} = 1$, we get

$$\mathcal{G}_{\mathbf{x}} = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} \frac{\partial}{\partial x} \\ \frac{\partial}{\partial y} \end{bmatrix} = \begin{bmatrix} -\frac{\partial}{\partial y} \\ \frac{\partial}{\partial x} \end{bmatrix}$$

No solution? Retry with higher order operators!

Even more formal treatment based on polynomial rings and Gröbner basis theory is published in

Markus Lange-Hegermann. **Algorithmic Linearly Constrained Gaussian Processes**, *Advances in Neural Information Processing Systems (NeurIPS)*, Montreal, 2018.

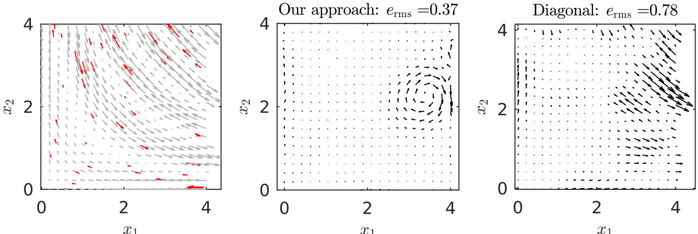
Simulation experiment - toy example

If we choose $k_{\mathbf{g}}(\mathbf{x}, \mathbf{x}') = \sigma_f^2 e^{-\frac{\|\mathbf{x}-\mathbf{x}'\|^2}{2l^2}}$ we get

$$\mathbf{K}(\mathbf{x}, \mathbf{x}') = \mathcal{G}_{\mathbf{x}} \mathcal{G}_{\mathbf{x}'}^{\mathsf{T}} k_{\mathbf{g}}(\mathbf{x}, \mathbf{x}') = \begin{bmatrix} -\frac{\partial}{\partial y} \\ \frac{\partial}{\partial x} \end{bmatrix} \begin{bmatrix} -\frac{\partial}{\partial y} & \frac{\partial}{\partial x} \end{bmatrix} k_{\mathbf{g}}(\mathbf{x}, \mathbf{x}')$$

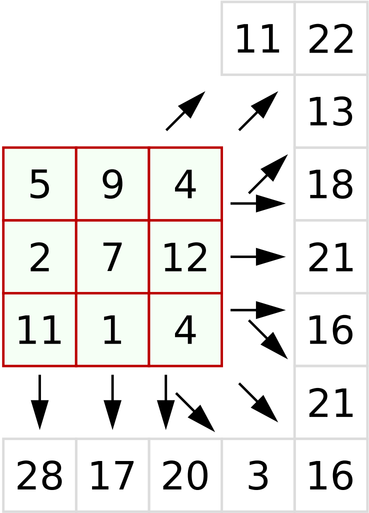
$$= \sigma_f^2 e^{-\frac{\|\mathbf{x}-\mathbf{x}'\|^2}{2l^2}} \left(\left(\frac{\mathbf{x} - \mathbf{x}'}{l} \right) \left(\frac{\mathbf{x} - \mathbf{x}'}{l} \right)^{\mathsf{T}} - \left(1 - \frac{\|\mathbf{x} - \mathbf{x}'\|^2}{l^2} \right) I_2 \right)$$

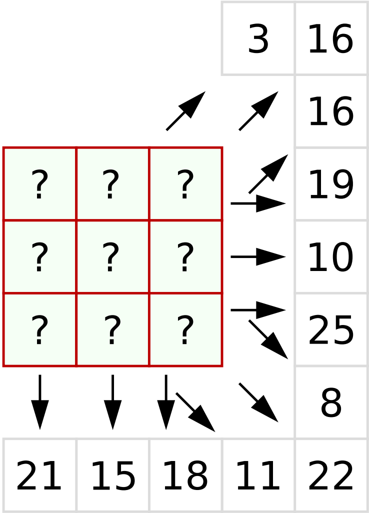
Below we have simulated a field which we know fulfills the constraint



Outline

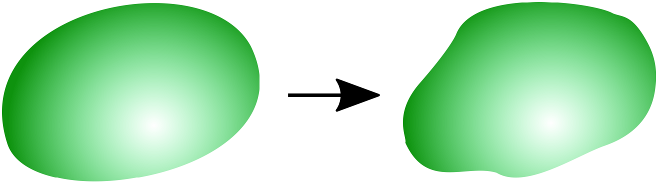
Tomography intuition

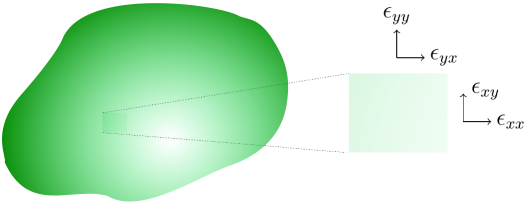




Strain field reconstruction

Deformed object

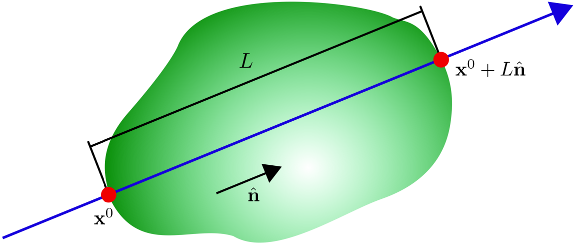




Reconstruct the *strain tensor*

$$\boldsymbol{\epsilon}(\mathbf{x}) = \begin{bmatrix} \epsilon_{xx}(\mathbf{x}) & \epsilon_{xy}(\mathbf{x}) \\ \epsilon_{xy}(\mathbf{x}) & \epsilon_{yy}(\mathbf{x}) \end{bmatrix}$$

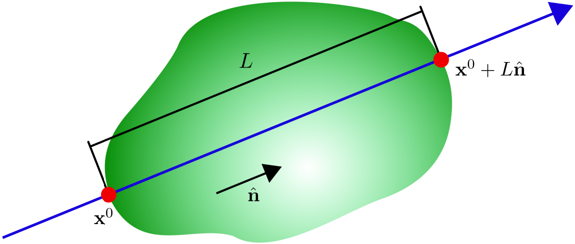
Strain field reconstruction



$$y = \frac{1}{L}\int_0^L \hat{\mathbf{n}}^\mathsf{T}\boldsymbol{\epsilon}(\mathbf{x}^0 + s\hat{\mathbf{n}})\hat{\mathbf{n}}\, ds + \varepsilon$$

$$\hat{\mathbf{n}} = \begin{bmatrix} n_x \\ n_y \end{bmatrix}, \qquad \varepsilon \sim \mathcal{N}(0, \sigma^2)$$

Strain field reconstruction



Vectorised form

$$y = \frac{1}{L}\int_0^L \vec{\mathbf{n}}^{\mathsf{T}} \mathbf{f}(\mathbf{x}^0 + s\hat{\mathbf{n}})\, ds + \varepsilon$$

$$\mathbf{f}(\mathbf{x}) = \begin{bmatrix} f_{xx}(\mathbf{x}) \\ f_{xy}(\mathbf{x}) \\ f_{yy}(\mathbf{x}) \end{bmatrix} = \begin{bmatrix} \epsilon_{xx}(\mathbf{x}) \\ \epsilon_{xy}(\mathbf{x}) \\ \epsilon_{yy}(\mathbf{x}) \end{bmatrix}, \qquad \vec{\mathbf{n}} = \begin{bmatrix} n_x^2 \\ 2n_x n_y \\ n_y^2 \end{bmatrix}$$

Strain field reconstruction – prediction

Put a GP on $\mathbf{f}(\mathbf{x})$

$$\mathbf{f}(\mathbf{x}) \sim \mathcal{GP}\left(\mathbf{0},\ \mathbf{K}(\mathbf{x}, \mathbf{x}')\right)$$

As before

$$\mathbb{E}\left[\mathbf{f}(\mathbf{x}_*)|\mathbf{y}\right] = Q_*^\mathsf{T}(Q + \sigma^2\mathbf{I})^{-1}\mathbf{y}$$

$$\mathrm{Cov}\left[\mathbf{f}(\mathbf{x}_*)|\mathbf{y}\right] = \mathbf{K}(\mathbf{x}_*, \mathbf{x}_*) - Q_*^\mathsf{T}(Q + \sigma^2\mathbf{I})^{-1}Q_*$$

$$Q_{ij} = \frac{1}{L_i L_j}\int_0^{L_i}\int_0^{L_j} \vec{\mathbf{n}}_i^\mathsf{T}\mathbf{K}(\mathbf{x}_i^0 + s\hat{\mathbf{n}}_i, \mathbf{x}_j^0 + t\hat{\mathbf{n}}_j)\vec{\mathbf{n}}_j\, dt\, ds$$

$$(Q_*)_i = \frac{1}{L_i}\int_0^{L_i} \vec{\mathbf{n}}_i^\mathsf{T}\mathbf{K}(\mathbf{x}_i^0 + s\hat{\mathbf{n}}_i, \mathbf{x}_*)\, ds$$

Strain field reconstruction – covariance model

Since $\mathbf{f}(\mathbf{x})$ is multivariate, the covariance function is a *matrix*

$$\mathbf{K}(\mathbf{x}, \mathbf{x}') = \begin{bmatrix} k_{11}(\mathbf{x}, \mathbf{x}') & k_{12}(\mathbf{x}, \mathbf{x}') & k_{13}(\mathbf{x}, \mathbf{x}') \\ k_{21}(\mathbf{x}, \mathbf{x}') & k_{22}(\mathbf{x}, \mathbf{x}') & k_{23}(\mathbf{x}, \mathbf{x}') \\ k_{31}(\mathbf{x}, \mathbf{x}') & k_{32}(\mathbf{x}, \mathbf{x}') & k_{33}(\mathbf{x}, \mathbf{x}') \end{bmatrix}$$

How should we select $\mathbf{K}(\mathbf{x}, \mathbf{x}')$?

Strain field reconstruction – constraint incorporation

A physical strain field must satisfy the *equilibrium constraints*

$$\begin{aligned} 0 &= \frac{\partial f_{xx}(\mathbf{x})}{\partial x} + (1-\nu)\frac{\partial f_{xy}(\mathbf{x})}{\partial y} + \nu\frac{\partial f_{yy}(\mathbf{x})}{\partial x} \\ 0 &= \nu\frac{\partial f_{xx}(\mathbf{x})}{\partial y} + (1-\nu)\frac{\partial f_{xy}(\mathbf{x})}{\partial x} + \frac{\partial f_{yy}(\mathbf{x})}{\partial y} \end{aligned}$$

These can be written as

$$\mathbf{0} = \underbrace{\begin{bmatrix} \frac{\partial}{\partial x} & (1-\nu)\frac{\partial}{\partial y} & \nu\frac{\partial}{\partial x} \\ \nu\frac{\partial}{\partial y} & (1-\nu)\frac{\partial}{\partial x} & \frac{\partial}{\partial y} \end{bmatrix}}_{\mathscr{F}_{\mathbf{x}}} \mathbf{f}(\mathbf{x}) = \begin{bmatrix} \boldsymbol{c}_1{}^{\mathsf{T}} \\ \boldsymbol{c}_2{}^{\mathsf{T}} \end{bmatrix} \mathbf{f}(\mathbf{x})$$

Strain field reconstruction – constraint incorporation

We get

$$\boldsymbol{\mathcal{G}}_{\mathbf{x}} = \boldsymbol{c}_1 \times \boldsymbol{c}_2 = \begin{bmatrix} \frac{\partial^2}{\partial y^2} - \nu \frac{\partial^2}{\partial x^2} \\ -(1+\nu)\frac{\partial^2}{\partial x \partial y} \\ \frac{\partial^2}{\partial x^2} - \nu \frac{\partial^2}{\partial y^2} \end{bmatrix}$$

Hence

$$\mathbf{f}(\mathbf{x}) = \boldsymbol{\mathcal{G}}_{\mathbf{x}} \varphi(\mathbf{x})$$

The scalar $\varphi(\mathbf{x})$ is the *Airy stress function*. Now let

$$\varphi(\mathbf{x}) \sim \mathcal{GP}\left(0,\ k_\varphi(\mathbf{x}, \mathbf{x}')\right)$$

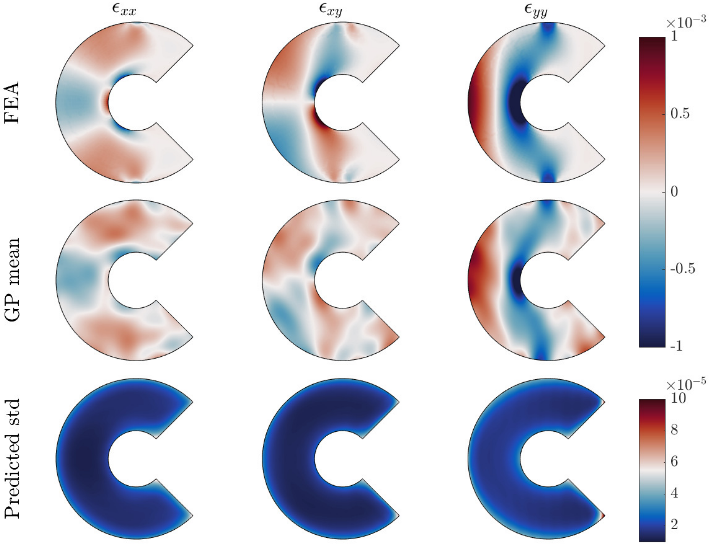
Then

$$\mathbf{f}(\mathbf{x}) \sim \mathcal{GP}\left(\mathbf{0},\ \boldsymbol{\mathcal{G}}_{\mathbf{x}} \boldsymbol{\mathcal{G}}_{\mathbf{x}'}^{\top} k_\varphi(\mathbf{x}, \mathbf{x}')\right)$$

Note

$$y = \mathcal{L}_{\mathbf{x}}[\boldsymbol{\mathcal{G}}_{\mathbf{x}} \varphi(\mathbf{x})] + \varepsilon$$

Strain field reconstruction – experimental results



Carl Jidling, Johannes Hendriks, Niklas Wahlström, Alexander Gregg, Thomas B. Schön, Chris Wensrich, Adrian Wills. **Probabilistic modelling and reconstruction of strain**, *Nuclear instruments and methods in physics research section B*, 436:141-155, 2018.

Outline

Nonlinearly constrained Gaussian processes- idea

Question: What can we do if we have nonlinear constraints?

We focus on sum-constrained Gaussian processes

$$\mathcal{F}[\mathbf{f}(\mathbf{x})] = \sum_i a_i h_i(f_i(\mathbf{x})) = C,$$

where i indexes the outputs of the GP and where $h_i(\cdot)$ is a non-linear function.

Idea: Transform the outputs $f_i' = h_i(f_i)$. The constraint will then be linear

$$\mathcal{F}[\mathbf{f}'(\mathbf{x})] = \sum a_i f_i'(\mathbf{x}) = C(\mathbf{x}),$$

Let f_i' be the output of the GP and train it on transform data $y_i' = h_i(y_i)$.

Toy Example (harmonic oscillator)

Consider the harmonic oscillator

$$E = E_{\mathrm{pot}}(t) + E_{\mathrm{kin}}(t) = kz(t)^2/2 + mv(t)^2/2,$$

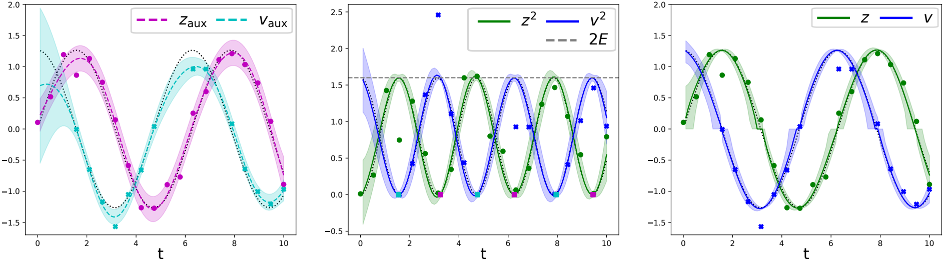
E_{pot} and E_{kin} denote potential and kinetic energy, respectively.

We assume that the displacement from the rest position z and the velocity v are the outputs of a multitask GP, whereas the time t is the input.

$$\mathbf{f}(t) = \begin{bmatrix} z(t) \\ v(t) \end{bmatrix}$$

There we have $a_1 = k/2$, $a_2 = m/2$, $h_1(z) = z^2$, $h_2(v) = v^2$ and $C = E$.

Nonlinearly constrained Gaussian processes - toy example



Left: Results for unconstrained GP

Middle: Results for transformed output learned by the constrained GP

Right: The back transformed output. The results for the unconstrained GP are used to recover the signs.

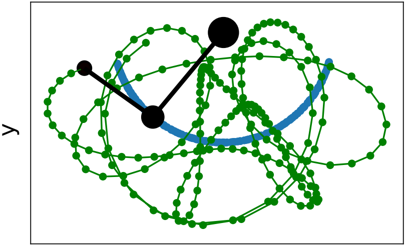
Nonlinearly constrained Gaussian processes - Double pendulum (real data)

Real data example (double pendulum)

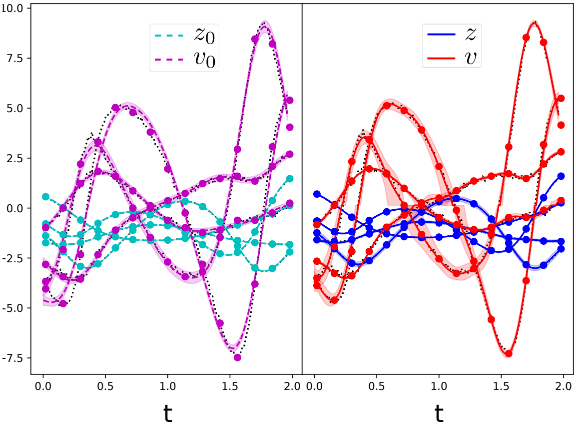
We model both positions z_x, z_y and velocities v_x, v_y of the two masses, (i.e. 8 outputs), while at the same time respecting the law of energy conservation

$$E = m_b g z_{by} + m_g g z_{gy} + \frac{m_b}{2}\left(v_{bx}^2 + v_{by}^2\right) + \frac{m_g}{2}\left(v_{gx}^2 + v_{gy}^2\right),$$

Indices b and g refer to blue and green pendulum, respectively.



Nonlinearly constrained Gaussian processes - Double pendulum (real data) - Results



Left: Results for unconstrained GP **Right:** Results for constrained GP

Philipp Pilar, Carl Jidling, Thomas B. Schön, Niklas Wahlström. **Incorporating sum constraints into multitask Gaussian processes**, *To be submitted*, 2022.

Conclusions

- Linear transformations are easily incorporated
- Physical laws can be built into the model
- Promising results on real data experiments
- The idea can also be extended to a nonlinear constraints

The combination of model driven physical knowledge and data driven flexibility is promising