On Inductive Biases for Gaussian Processes from Differential Algebra

Markus Lange-Hegermann

Institut für industrielle Informationstechnik - inIT
Department of Electrical Engineering and Computer Science, TH OWL

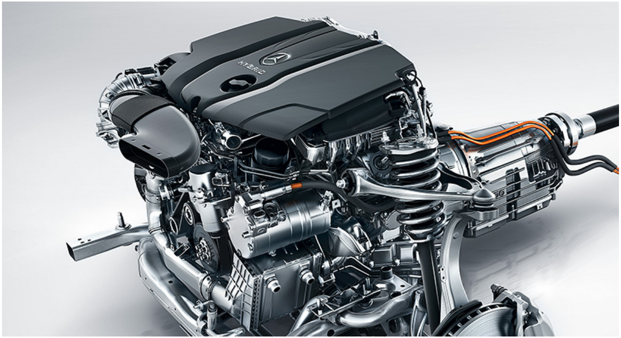
Leipzig, 2022/06/08
Workshop on Differential Algebra
Max Planck Institute for Mathematics in the Sciences

Various Thanks

- Organizer for giving me this opportunity
- Intersting people and interesting discussion
- Bogdan, Marc, and Rida for commenting Ehrenpreis-Palamodov
- Meeting old friends in person
- Heather, for bringing data into differential algebra
- Shiva and Werner, for stressing the importance of function spaces

Processes Combustion Calibration Diesel Engine

measure engine ⟶ math model ⟶ optimize



http://www.mercedes-benz.com.au

Table of Contents

- Probability
  - Or: what are Gaussian processes?
- Differential algebra
  - Or: can we bring data into differential algebra?

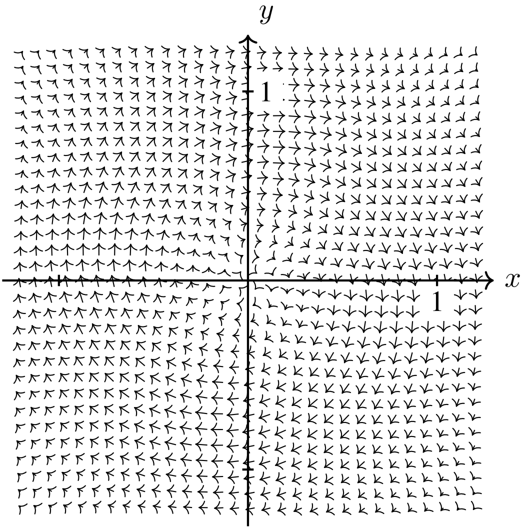
Maxwell's Equations

The operator matrix

$$A := \begin{bmatrix} 0 & -\partial_z & \partial_y & \partial_t & 0 & 0 & 0 & 0 & 0 & 0 \\ \partial_z & 0 & -\partial_x & 0 & \partial_t & 0 & 0 & 0 & 0 & 0 \\ -\partial_y & \partial_x & 0 & 0 & 0 & \partial_t & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \partial_x & \partial_y & \partial_z & 0 & 0 & 0 & 0 \\ -\partial_t & 0 & 0 & 0 & -\partial_z & \partial_y & -1 & 0 & 0 & 0 \\ 0 & -\partial_t & 0 & \partial_z & 0 & -\partial_x & 0 & -1 & 0 & 0 \\ 0 & 0 & -\partial_t & -\partial_y & \partial_x & 0 & 0 & 0 & -1 & 0 \\ \partial_x & \partial_y & \partial_z & 0 & 0 & 0 & 0 & 0 & 0 & -1 \end{bmatrix}$$

acts on 3 components electrical field, 3 components magnetic (pseudo-)field, 3 components electric current and a component electric flux.

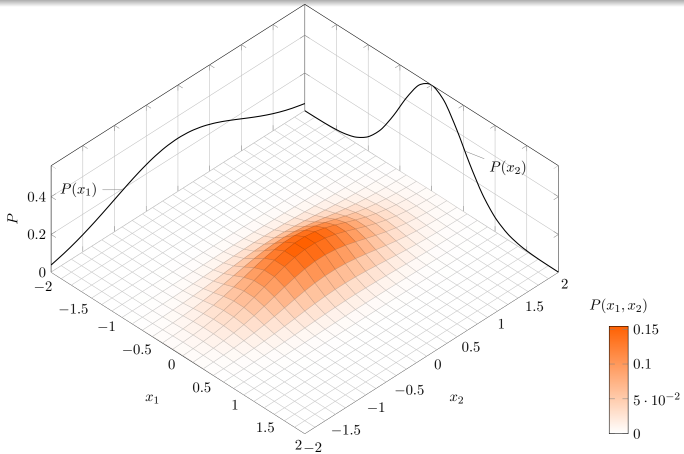
(constants := 1)



y
1
x
1

Gaussian Distributions and Processes

Gaussian Distribution on $\mathbb{R}^n$



$$\text{Density: } \frac{(2\pi)^{-\frac{1}{2}n}}{\sqrt{\det \Sigma}} \exp\left(-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)\right)$$

Why is the Gaussian distribution so ubiquitous?

Gaussian Distribution: Properties Viewed by a Bayesian

Theorem (first two moments/cumulants describe everything.)

The **Gaussian distribution maximizes the entropy** among all probability distributions on $\mathbb{R}^n$ with fixed mean and (co)variance.

Maximum entropy prior (Jaynes)

Known/suspected mean and (co)variance: take **Gaussian prior**.

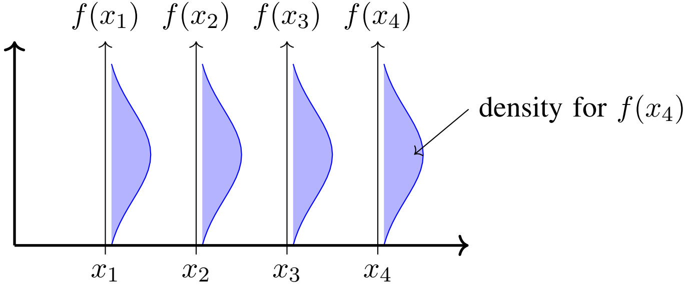
Corollaries (colloquial)

- uncorrelated $\Longrightarrow$ independent.
- Central limit theorem (iid random variables (finite mean and variance) average to a Gaussian).
- Closed under marginal distributions: drop the marginalized part
- Closed under conditional distributions:
  $\mu_{x_1|x=a} = \mu_{x_1} + \Sigma_{x_1,x}\Sigma_{x,x}^{-1}(a-\mu_x)$
  $\Sigma_{x_1,x_2|x=a} = \Sigma_{x_1,x_2} - \Sigma_{x_1,x}\Sigma_{x,x}^{-1}\Sigma_{x,x_2}$
- Sampling is possible: diagonalize covariance

Gaussian Processes

Idea

Assume **Gaussian function values** of the regression function f.
Marginalization: only consider finitely many function evaluations.



Definition: **Gaussian process**

A distribution on functions s.t. the evaluations $f(x_1), \ldots, f(x_n)$ at any $x_1, \ldots, x_n$ are (jointly) Gaussian.

Characterizing Gaussian Processes

Gaussian distribution		Gaussian process
1D $\mathcal{N}(\mu, \sigma^2)$	finite dimensional $\mathcal{N}(\mu, \Sigma)$	$\mathcal{GP}(\mu(x), k(x_1, x_2))$
mean μ	mean vector μ	mean function $\mu(x)$
variance σ^2	covariance matrix Σ	covariance function $k(x_1, x_2)$
higher moments/cumulants irrelevant/zero		

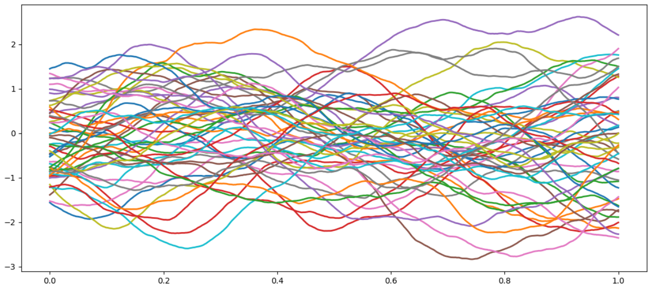
Set mean function to the constant zero function (normalize data).

It remains to...

...encode information in the covariance function.

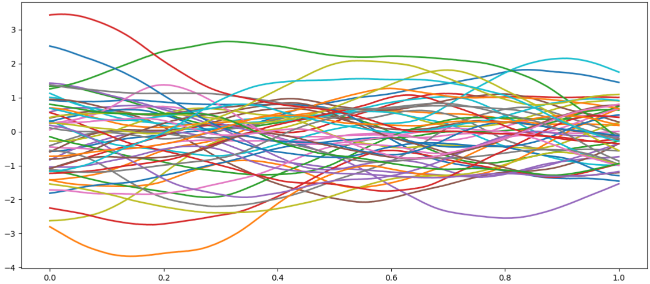
Covariance: Interdependence of Function Evaluations

C^1: continuously differentiable



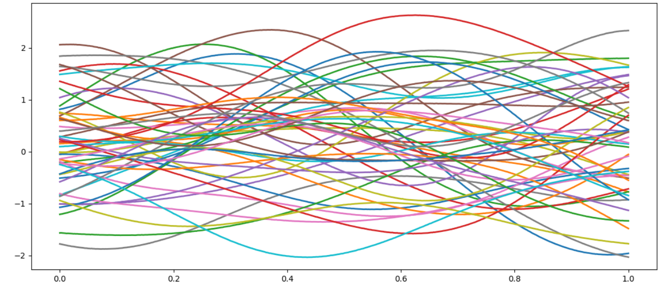
Covariance: Interdependence of Function Evaluations

C^2: twice continuously differentiable



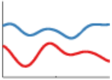
Covariance: Interdependence of Function Evaluations

C^∞: smooth

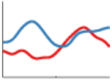


Kernel Cookbook

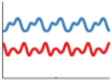
squared exponential $$\sigma^2 \exp\left(-\frac{1}{2}\frac{(x-x')^2}{\ell^2}\right)$$



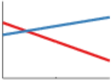
rational quadratic $$\sigma^2 \left(1+\frac{1}{2\alpha}\frac{(x-x')^2}{\ell^2}\right)^{-\alpha}$$



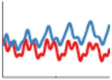
periodic $$\sigma^2 \exp\left(-2\frac{\sin^2\left(\frac{\pi}{p}|x-x'|\right)}{\ell^2}\right)$$



linear $$a^2+b^2xx'$$

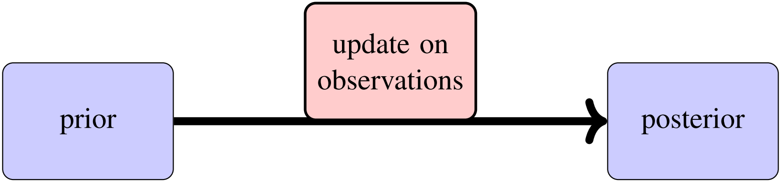


local periodic $$\sigma^2 \exp\left(-2\frac{\sin^2\left(\frac{\pi}{p}|x-x'|\right)}{\ell^2}-\frac{1}{2}\frac{(x-x')^2}{\ell^2}\right)$$



David Duvenaud, *Kernel Cookbook*, `http://www.cs.toronto.edu/~duvenaud/cookbook/`

Bayesian approach



(Due to their computational simplicity: GPs are the standard functional prior in Bayesian ML&Stats.)

Gaussian Process Regression: Math

Reminder: Gaussian process $g = \mathcal{GP}(\mu, k)$

A distribution on $\mathbb{R}^d \to \mathbb{R}^\ell$ s.t. $g(x_1), \dots, g(x_n)$ are Gaussian.
Data structure: $\mu : \mathbb{R}^d \to \mathbb{R}^\ell$ and $k : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}^{\ell \times \ell}_{\geq 0}$.

Regression model

Assume $\mu = 0$. Condition on $\{(x_i, y_i) \in \mathbb{R}^{1\times(d+\ell)} \mid i = 1, \dots, n\}$.

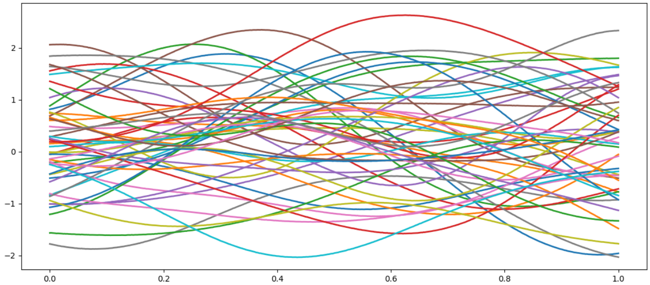
$$\mathcal{GP}\Big(\begin{aligned} x &\mapsto \mathbf{y}\mathbf{k}(\mathbf{X},\mathbf{X})^{-1}\mathbf{k}(\mathbf{X},\mathbf{x}), \\ (x,x') &\mapsto \mathbf{k}(\mathbf{x},\mathbf{x}') - \mathbf{k}(\mathbf{x},\mathbf{X})\mathbf{k}(\mathbf{X},\mathbf{X})^{-1}\mathbf{k}(\mathbf{X},\mathbf{x}') \end{aligned} \Big) .$$

$$k(X,X) = \begin{bmatrix} k(x_1,x_1) & \dots \\ \vdots & \ddots \end{bmatrix} \in \mathbb{R}^{\ell n \times \ell n}_{\geq 0},$$

$k(x,X) = \begin{bmatrix} k(x,x_1) & \dots \end{bmatrix} \in \mathbb{R}^{\ell \times \ell n}$, and $y = \begin{bmatrix} y_1 & \dots \end{bmatrix} \in \mathbb{R}^{1 \times \ell n}$.

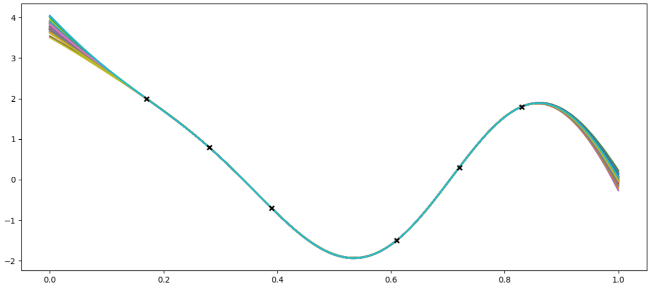
Include Measurement Data

C^∞: smooth



Include Measurement Data

C^∞, conditioned on data



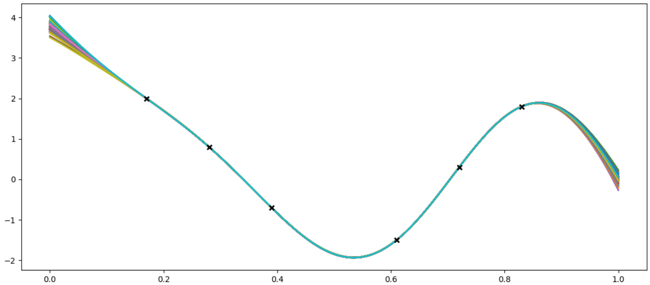
Gaussian Process Regression: Noise

- Take a maximum entropy prior on the behavior unexplained by g:
  Add Gaussian white **noise** ε (works well enough if noise is not strictly Gaussian).
- Replace covariance $k(X,X)$ by $k(X,X)+\mathrm{var}(\varepsilon)I_{\ell n}$.
  (more variance in data, no new correlations)
- Posterior:

$$\mathcal{GP}\Big(\quad x\mapsto y(k(X,X)+\mathrm{var}(\varepsilon)I_{\ell n})^{-1}k(x,X)^T,$$
$$(x,x')\mapsto k(x,x')-k(x,X)(k(X,X)+\mathrm{var}(\varepsilon)I_{\ell n})^{-1}k(x',X)^T\Big)$$

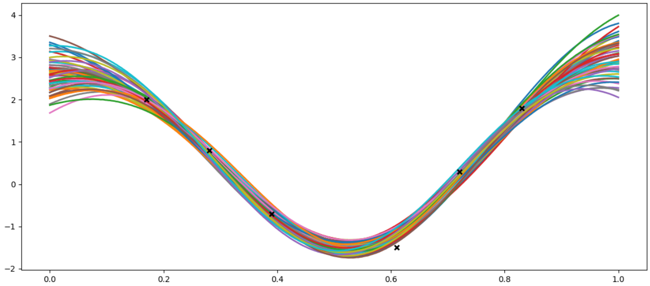
Include Measurement Data

C^∞, conditioned on data



Include Measurement Data

C^∞, conditioned on noisy data



Gaussian Process Regression: Hyperparameters

- Hyperparameters in the priors:
  - length scales ℓ
  - signal variance σ
  - noise ε
  - period p
  - etc.
- **Optimal hyperparameters**: optimize the (log-)likelihood.

$$\log p(y|X) = -\underbrace{\frac{1}{2}y^T K^{-1} y}_{\text{data fit}} - \underbrace{\frac{1}{2}\log(\det(K))}_{\text{model complexity}} - \frac{n}{2}\log 2\pi$$

  Computable via linear Algebra (including gradients)
- Hyperparameters in GPs are interpretable and learnable
  E.g. learn a period in your data or amount of noise.

RKHS

Reproducing Kernel Hilbert Spaces (RKHS)

A Hilbert space of functions s.t. the evaluation functionals are continuous.

- Continuity (or even differentiability) of the model evaluation is typically required for model training.
- Hence, most ML-models can be described by an RKHS.

RKHS

Let $g = \mathcal{GP}(0, k)$.

The $x \mapsto k(x_i, x)$ for $x_i \in \mathbb{R}^d$ generate the pre-Hilbert space $\mathcal{H}^0(g)$ with scalar product $\langle k(x_i, -), k(x_j, -)\rangle := k(x_i, x_j)$.

The closure $\mathcal{H}(g)$ of $\mathcal{H}^0(g)$ w.r.t. $\langle \cdot, \cdot \rangle$ is the **Reproducing Kernel Hilbert Space (RKHS)** of g.

Theorem (Moore–Aronszajn)

Any RKHS is of this form, i.e. has a so-called **reproducing kernel** k. Hence, there is a 1-1-correspondence: covariance functions $\leftrightarrow$ RKHS.

$\mathcal{H}^0(g)$ is the space of posterior mean functions.

In many settings, the RKHS $\mathcal{H}(g)$ is the Cameron-Martin Space of the Gaussian measure induced by g.

Support and Realizations

No Gaussian measure on $\mathcal{H}(g)$ if it is infinite dimensional.
GP g induces a Gaussian measure on a space of functions $\mathcal{F} \hookleftarrow \mathcal{H}(g)$

(e.g., abstract Weiner space) under mild assumptions on the topology of $\mathcal{F}$, e.g. $\mathcal{F}$ Fréchet.

The following three sets are identical under similar mild assumptions:

1. The support of this measure.
2. The realizations (samples) of g.
3. The closure $\overline{\mathcal{H}(g)}$ of $\mathcal{H}(g)$ in $\mathcal{F}$.

E.g. (1)⇔(3) holds for all Radon Gaussian measures ($\Leftarrow$ $\mathcal{F}$ locally compact $\Leftarrow$ $\mathcal{F}$ Fréchet$\Leftarrow$ $\mathcal{F}$ Banach).

E.g. (1)⇔(2) holds almost surely anyway, and strictly holds for a continuous modification of g.

Moral

Knowing $\mathcal{H}(g)$ means knowing g.

Support and Realizations

Trivial example

The linear covariance function $k(t,t') = f(t) \cdot f(t')$ induces a GP with realizations equal to the space $\mathcal{H}(k) = \mathbb{R} \cdot (t \mapsto f(t))$.

Non-trivial example

The squared exponential covariance function

$$k(t,t') = \exp\left(-\frac{1}{2}(t-t')^2\right)$$

induces a GP with realizations dense (Fréchet topology) in $C^\infty(\mathbb{R},\mathbb{R})$.

Sums

Theorem

Let $g_1 = (0, k_1)$ and $g_2 = (0, k_2)$ GPs and $g = (0, k_1 + k_2)$. Then,

$$\mathcal{H}(g) = \mathcal{H}(g_1) + \mathcal{H}(g_2).$$

(For a suitable choice of the scalar product in the sum.)

- Explain an effect as a sum of two causes.
  - E.g. smooth plus periodic.
  - E.g. use a summand for “unexplained behavior”.

Products

Theorem

Let $g_1 = (0, k_1)$ and $g_2 = (0, k_2)$ GPs and $g = (0, k_1 \cdot k_2)$. Then,

$$\mathcal{H}(g) = \mathcal{H}(g_1) \otimes \mathcal{H}(g_2).$$

The Hilbert space $\otimes$ is the completion of the vector space $\otimes$. The fun begins when tensoring the spaces of realizations.

- All causes are needed for an effect.
  - E.g. locally periodic behavior.

Questions?

Questions?

Gaussian processes and linear differential equations

Linear Systems (in the sense of linear algebra)

For $\mathcal{F} = C^\infty(\mathbb{R}, \mathbb{R})$ and $A = \begin{bmatrix} 2 & -3 \end{bmatrix}$ consider

$$\mathrm{sol}_{\mathcal{F}}(A) := \left\{ \begin{bmatrix} f_1(x) \\ f_2(x) \end{bmatrix} \in \mathcal{F}^{2\times 1} \;\middle|\; A \cdot \begin{bmatrix} f_1(x) \\ f_2(x) \end{bmatrix} = 0 \right\}$$

Linear Systems (in the sense of linear algebra)

For $\mathcal{F} = C^\infty(\mathbb{R}, \mathbb{R})$ and $A = \begin{bmatrix} 2 & -3 \end{bmatrix}$ consider

$$\mathrm{sol}_{\mathcal{F}}(A) := \left\{ \begin{bmatrix} f_1(x) \\ f_2(x) \end{bmatrix} \in \mathcal{F}^{2\times 1} \,\middle|\, A \cdot \begin{bmatrix} f_1(x) \\ f_2(x) \end{bmatrix} = 0 \right\}$$

Use $B = \begin{bmatrix} 3 \\ 2 \end{bmatrix}$ as parametrization:

$$\mathrm{sol}_{\mathcal{F}}(A) = B \cdot \mathcal{F} = \{ B \cdot f(x) \mid f(x) \in \mathcal{F} \}$$

Linear Systems (in the sense of linear algebra)

For $\mathcal{F} = C^\infty(\mathbb{R},\mathbb{R})$ and $A = \begin{bmatrix} 2 & -3 \end{bmatrix}$ consider

$$\mathrm{sol}_{\mathcal{F}}(A) := \left\{ \begin{bmatrix} f_1(x) \\ f_2(x) \end{bmatrix} \in \mathcal{F}^{2\times 1} \;\middle|\; A \cdot \begin{bmatrix} f_1(x) \\ f_2(x) \end{bmatrix} = 0 \right\}$$

Use $B = \begin{bmatrix} 3 \\ 2 \end{bmatrix}$ as parametrization:

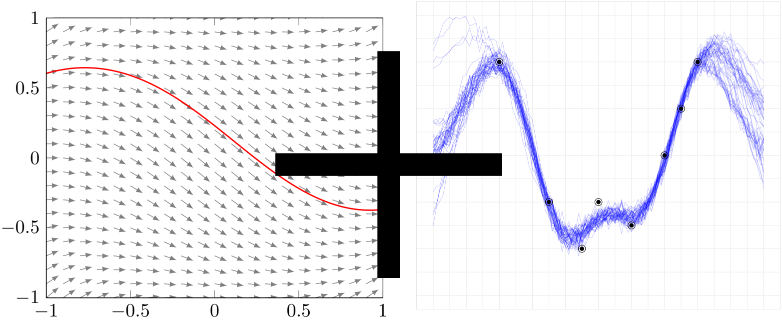
$$\mathrm{sol}_{\mathcal{F}}(A) = B \cdot \mathcal{F} = \{ B \cdot f(x) \mid f(x) \in \mathcal{F} \}$$

Taking a GP prior $g = \mathcal{GP}(0, k)$ for $f \in \mathcal{F}$ yields a GP prior

$$B_* g := \mathcal{GP}(0, BkB^T) = \mathcal{GP}\left(0, \begin{bmatrix} 9k & 6k \\ 6k & 4k \end{bmatrix}\right)$$

for $\mathrm{sol}_{\mathcal{F}}(A)$.

Combination of Gaussian Processes with Operator Equations



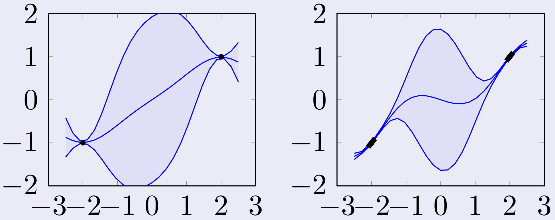
- Combine **strict, global information** from differential equations with **noisy, local information** from observations.
- Incorporate justified assumptions: use the **full information** of the observations for a precise regression model.

The class of GPs is closed under linear operators under mild assumptions.

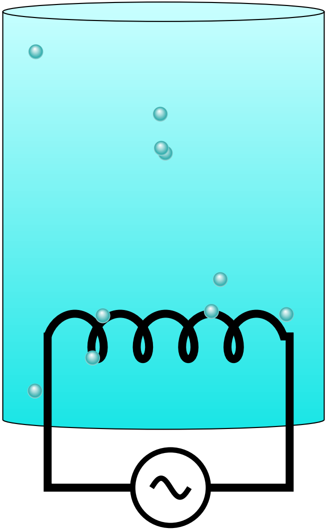
Example

Use $B = \begin{bmatrix} 1 \\ \partial_x \end{bmatrix}$ as parametrization and $g = \mathcal{GP}(0, k)$:

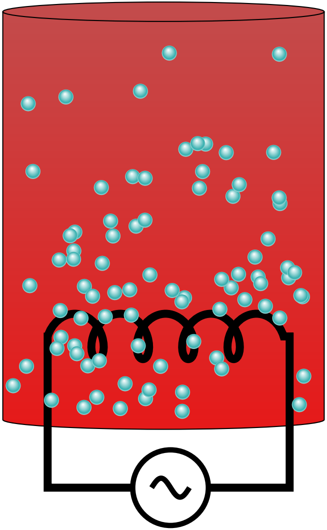
$$\rightsquigarrow B_* g := \mathcal{GP}(0, BkB'^T) = \mathcal{GP}\left(0, \begin{bmatrix} k(x,x') & \frac{\partial}{\partial x'}k(x,x') \\ \frac{\partial}{\partial x}k(x,x') & \frac{\partial^2}{\partial x \partial x'}k(x,x') \end{bmatrix}\right)$$



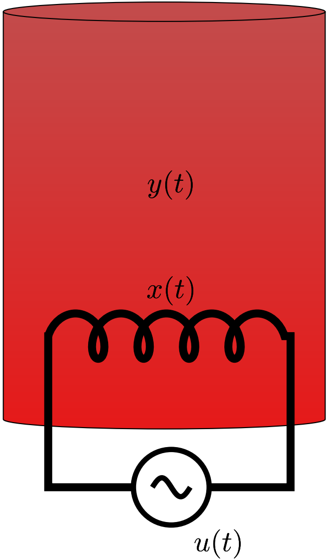
Heating System



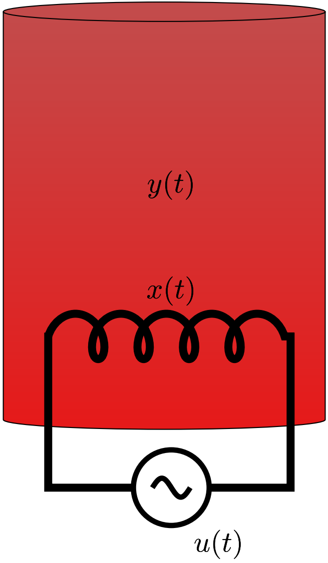
Heating System



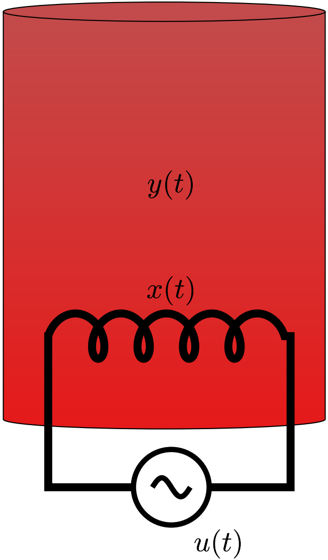
Heating System



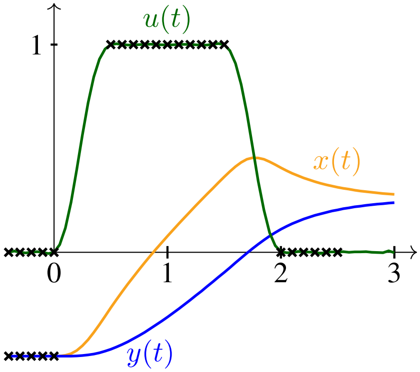
Heating System



$$\partial_t x(t) = -(x(t) - y(t)) + u(t)$$
$$\partial_t y(t) = +(x(t) - y(t))$$



$$\partial_t x(t) = -(x(t) - y(t)) + u(t)$$
$$\partial_t y(t) = +(x(t) - y(t))$$



Smith Normal Form

Smith normal form

Given a matrix A (over a PID), there are invertible matrices S and T s.t.

$$SAT = D$$

where D is a matrix with non-zero entries only on the diagonal.

(D can be made unique by demanding that each diagonal entry divides the next one.)

Computable in polynomial time (as long as the PID is Euclidean), even in parallel (NC^2).

Using the Smith normal form

$$Af = 0 \Leftrightarrow SAT\underbrace{T^{-1}f}_{=:h} = 0$$

$$\Leftrightarrow Dh = 0$$

If we get a GP prior for $h = T^{-1}f$, we have one for $f = Th$.

Prior for h

$$\begin{bmatrix} 1 & & & \\ & \partial_t - 1 & & \\ & & \partial_t^2 + 1 & 0 \end{bmatrix} \cdot \begin{bmatrix} h_1(t) \\ h_2(t) \\ h_3(t) \\ h_4(t) \end{bmatrix} = 0$$

Since we can easily solve such ODEs:

$h_1(t) = 0$	$k_1(t_1, t_2) = 0$
$h_2(t) = c \cdot \exp(t)$	$k_2(t_1, t_2) = \exp(t_1)\exp(t_2)$
$h_3(t) = c_1 \sin(t) + c_2 \cos(t)$	$k_3(t_1, t_2) = \cos(t_1 - t_2)$
$h_4(t)$ arbitrary (smooth)	$k_4(t_1, t_2) = \exp(-\frac{1}{2}(t_1 - t_2)^2)$

joint with Andreas Besginow.

Gaussian Processes and Linear Operator Matrices

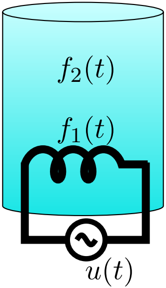
Let $T \in R^{\ell \times m}$ and $g = \mathcal{GP}(\mu, k)$.
Define the pushforward GP $T_* g$ by applying T to the realizations of g.

Lemma

Assume that T commutes w.r.t. expectation of the relevant measures.

- (Pushforward is again a Gaussian process)
  $T_* g = \mathcal{GP}(T\mu(x), Tk(x, x')(T')^T)$ where T' operates on x'.
- (Realizations behave reasonable)
  For $g = \mathcal{GP}(0, k)$ with zero mean function, $\mathcal{H}(T_* g) = T\mathcal{H}(g)$.

Heating System



Add parameters a and b quantifying heat exchange:

$$f_1'(t) = -a \cdot (f_1(t) - f_2(t)) + u(t)$$
$$f_2'(t) = -b \cdot (f_2(t) - f_1(t))$$

(Training data from a solution, $a = 3, b = 1$).

- Model reconstructs a and b with error $< 2.8\%$ (data without noise) resp. $< 5.3\%$ (data with 1% noise), 10 training runs.
- Satisfies ODEs ($a = 3$, $b = 1$) with median error of $5\mathrm{e}{-3}$, trained on noisy data, despite approximate a and b in model.

joint with Andreas Besginow.

Simple Control System

Time dependent system $\partial_t x(t) = t^3 u(t)$.

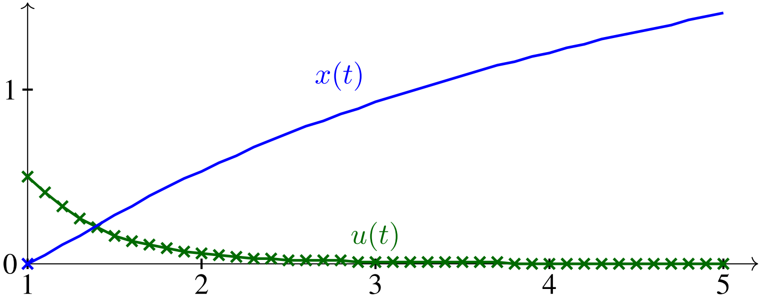
(We need the Jacobson form instead of the Smith form, since we are over a Weyl algebra)

Simple Control System

Time dependent system $\partial_t x(t) = t^3 u(t)$.

(We need the Jacobson form instead of the Smith form, since we are over a Weyl algebra)

Set an input $u(t)$ to **influence** a state $x(t)$.
Set $x(1) = 0$ and $u(t) = \frac{1}{t^4+1}$ for $t \in \{1, \frac{11}{10}, \frac{12}{10}, \dots, 5\}$.



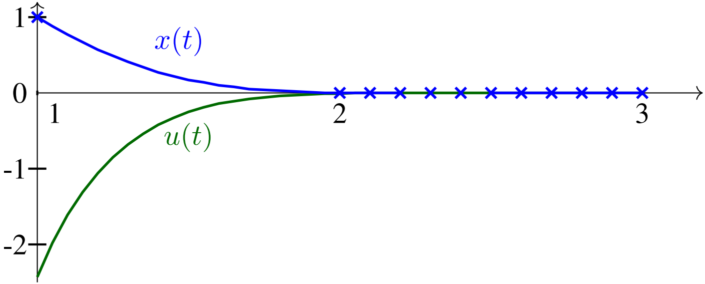
Model: $x(5) \approx 1.436537$, close to $\int_1^5 \frac{t^3}{t^4+1}\,\mathrm{d}t \approx 1.436551$.

Simple Control System

Time dependent system $\partial_t x(t) = t^3 u(t)$.

(We need the Jacobson form instead of the Smith form, since we are over a Weyl algebra)

Prescribe a state $x(t)$. Automatically **construct** an input $u(t)$.



Assumptions for PDEs

Let R be an $\mathbb{R}$-algebra, a *ring of linear operators*, and $\mathcal{F}$ an R-module of functions $\mathbb{R}^d \to \mathbb{R}$ with topology. Assume:

1. We can compute with operators
2. Functions yield enough solutions
3. Gaussian processes describe functions

4. Compatible operators and topology
5. Compatible Gaussian processes and topology

6. Compatible Gaussian processes and operators

Assumptions for PDEs

Let R be an $\mathbb{R}$-algebra, a *ring of linear operators*, and $\mathcal{F}$ an R-module of functions $\mathbb{R}^d \to \mathbb{R}$ with topology. Assume:

1. We can compute with operators: R allows a Gröbner bases.
2. Functions yield enough solutions: $\mathcal{F}$ is an injective R-module.
3. Gaussian processes describe functions: There is a scalar $g = \mathcal{GP}(0, k)$ s.t. its RKHS $\mathcal{H}(g)$ is dense in $\mathcal{F}$ and its set of realizations is contained (a.s.) in $\mathcal{F}$.
4. Compatible operators and topology: R acts continuously on $\mathcal{F}$.
5. Compatible Gaussian processes and topology: GPs in $\mathcal{F}$ are $1:1$ with Gaussian measures on $\mathcal{F}$ w.r.t. the Borel σ-algebra.
6. Compatible Gaussian processes and operators: the operation of R on $\mathcal{H}(g)$ commutes with expectation (g induces measure).

Assumptions for PDEs

Let R be an $\mathbb{R}$-algebra, a *ring of linear operators*, and $\mathcal{F}$ an R-module of functions $\mathbb{R}^d \to \mathbb{R}$ with topology. Assume:

1. We can compute with operators: R allows a Gröbner bases.
2. Functions yield enough solutions: $\mathcal{F}$ is an injective R-module.
3. Gaussian processes describe functions: There is a scalar $g = \mathcal{GP}(0,k)$ s.t. its RKHS $\mathcal{H}(g)$ is dense in $\mathcal{F}$ and its set of realizations is contained (a.s.) in $\mathcal{F}$.
4. Compatible operators and topology: R acts continuously on $\mathcal{F}$.
5. Compatible Gaussian processes and topology: GPs in $\mathcal{F}$ are $1:1$ with Gaussian measures on $\mathcal{F}$ w.r.t. the Borel σ-algebra.
6. Compatible Gaussian processes and operators: the operation of R on $\mathcal{H}(g)$ commutes with expectation (g induces measure).

Theorem

Assumptions hold for $R = \mathbb{R}[\partial_{x_1}, \ldots, \partial_{x_d}]$, $\mathcal{F} = C^\infty(\mathbb{R}^d, \mathbb{R})$ with Fréchet topology, and g with SE covariance.

Assumptions for PDEs

Let R be an $\mathbb{R}$-algebra, a *ring of linear operators*, and $\mathcal{F}$ an R-module of functions $\mathbb{R}^d \to \mathbb{R}$ with topology. Assume:

1. We can compute with operators: R allows a Gröbner bases.
2. Functions yield enough solutions: $\mathcal{F}$ is an injective R-module.
3. Gaussian processes describe functions: There is a scalar $g = \mathcal{GP}(0, k)$ s.t. its RKHS $\mathcal{H}(g)$ is dense in $\mathcal{F}$ and its set of realizations is contained (a.s.) in $\mathcal{F}$.
4. Compatible operators and topology: R acts continuously on $\mathcal{F}$.
5. Compatible Gaussian processes and topology: GPs in $\mathcal{F}$ are $1:1$ with Gaussian measures on $\mathcal{F}$ w.r.t. the Borel σ-algebra.
6. Compatible Gaussian processes and operators: the operation of R on $\mathcal{H}(g)$ commutes with expectation (g induces measure).

Proposition

Assumptions hold for $R = \mathbb{R}(t)\langle\partial_t\rangle$, $\mathcal{F} = C^\infty(D, \mathbb{R})$ with Fréchet topology, g with SE covariance and $D \subseteq \mathbb{R}$.

Assumptions for PDEs

Let R be an $\mathbb{R}$-algebra, a *ring of linear operators*, and $\mathcal{F}$ an R-module of functions $\mathbb{R}^d \to \mathbb{R}$ with topology. Assume:

1. We can compute with operators: R allows a Gröbner bases.
2. Functions yield enough solutions: $\mathcal{F}$ is an injective R-module.
3. Gaussian processes describe functions: There is a scalar $g = \mathcal{GP}(0, k)$ s.t. its RKHS $\mathcal{H}(g)$ is dense in $\mathcal{F}$ and its set of realizations is contained (a.s.) in $\mathcal{F}$.
4. Compatible operators and topology: R acts continuously on $\mathcal{F}$.
5. Compatible Gaussian processes and topology: GPs in $\mathcal{F}$ are $1:1$ with Gaussian measures on $\mathcal{F}$ w.r.t. the Borel σ-algebra.
6. Compatible Gaussian processes and operators: the operation of R on $\mathcal{H}(g)$ commutes with expectation (g induces measure).

Remark

Assumptions hold for $R = \mathbb{R}[x_1, \ldots, x_n]$, $\mathcal{F} = C^\infty(D, \mathbb{R})$ with Fréchet topology, g with SE covariance and $D \subseteq \mathbb{R}$.

Assumptions for PDEs

Let R be an $\mathbb{R}$-algebra, a *ring of linear operators*, and $\mathcal{F}$ an R-module of functions $\mathbb{R}^d \to \mathbb{R}$ with topology. Assume:

1. We can compute with operators: R allows a Gröbner bases.
2. Functions yield enough solutions: $\mathcal{F}$ is an injective R-module.
3. Gaussian processes describe functions: There is a scalar $g = \mathcal{GP}(0, k)$ s.t. its RKHS $\mathcal{H}(g)$ is dense in $\mathcal{F}$ and its set of realizations is contained (a.s.) in $\mathcal{F}$.
4. Compatible operators and topology: R acts continuously on $\mathcal{F}$.
5. Compatible Gaussian processes and topology: GPs in $\mathcal{F}$ are $1 : 1$ with Gaussian measures on $\mathcal{F}$ w.r.t. the Borel σ-algebra.
6. Compatible Gaussian processes and operators: the operation of R on $\mathcal{H}(g)$ commutes with expectation (g induces measure).

Remark

Assumptions hold for $R = \mathbb{R}[\sigma_1, \ldots, \sigma_n]$, $\mathcal{F} = C^\infty(\mathbb{R}^n, \mathbb{R})$ with Fréchet topology and g with SE covariance, where $\sigma_i(x_j) = x_i + \delta_{ij}$.

Assumptions for PDEs

Let R be an $\mathbb{R}$-algebra, a *ring of linear operators*, and $\mathcal{F}$ an R-module of functions $\mathbb{R}^d \to \mathbb{R}$ with topology. Assume:

1. We can compute with operators: R allows a Gröbner bases.
2. Functions yield enough solutions: $\mathcal{F}$ is an injective R-module.
3. Gaussian processes describe functions: There is a scalar $g = \mathcal{GP}(0, k)$ s.t. its RKHS $\mathcal{H}(g)$ is dense in $\mathcal{F}$ and its set of realizations is contained (a.s.) in $\mathcal{F}$.
4. Compatible operators and topology: R acts continuously on $\mathcal{F}$.
5. Compatible Gaussian processes and topology: GPs in $\mathcal{F}$ are $1:1$ with Gaussian measures on $\mathcal{F}$ w.r.t. the Borel σ-algebra.
6. Compatible Gaussian processes and operators: the operation of R on $\mathcal{H}(g)$ commutes with expectation (g induces measure).

Theorem

Under the above assumptions, there exists GP priors with realizations dense in controllable system, i.e. systems with vector potentials.

Proof / Algorithm (Differential Algebra, Malgrange)

Let $M = \operatorname{coker}(R^{\ell'} \xrightarrow{A} R^{\ell})$ a torsionless R-module for $A \in R^{\ell' \times \ell}$.

- Compute $\hom_R(M, R)$ and a free hull $\hom_R(M, R) \overset{B}{\twoheadleftarrow} R^{\ell'' \times 1}$.
  This gives the embedding $\hom_R(\hom_R(M, R), R) \overset{B}{\hookrightarrow} R^{1 \times \ell''}$.
  - Gröbner: $B := r - \ker(A)$.
- For M torsionless, i.e. $M \to \hom_R(\hom_R(M, R), R)$ monic:

$$M \overset{I}{\hookrightarrow} \hom_R(\hom_R(M, R), R) \overset{B}{\hookrightarrow} R^{1 \times \ell''} .$$

  - Gröbner: Does $l - \ker(B)$ reduce to zero w.r.t. the rows of A?
- Apply the (exact, since $\mathcal{F}$ injective, (2)) functor $\hom_R(-, \mathcal{F})$:

$$\hom_R(M, \mathcal{F}) \xtwoheadleftarrow{\quad B \quad} \mathcal{F}^{\ell'' \times 1}$$

- Parametrize solutions by the Noether-Malgrange isomorphism

$$\mathrm{sol}_{\mathcal{F}}(A) \cong \hom_R(M, \mathcal{F}) : f \mapsto (e_i \mapsto f_i)$$

Proof / Properties (Functional Analysis and Probability)

- By (3) we have a GP on $\mathcal{F}$ and hence also a GP g with realizations dense in $\mathcal{F}^{\ell''\times 1}$.
- Topology (continuity, denseness) implies properties of GPs (5).
- $\hom_R(M,\mathcal{F}) \overset{B}{\twoheadleftarrow} \mathcal{F}^{\ell''\times 1}$ is epic, continuous (4) and commutes with expactation (6 & Lemma above). Hence, the realizations of $B_* g$ are dense in $\mathrm{sol}_{\mathcal{F}}(A) \cong \hom_R(M,\mathcal{F})$. $\square$

More is possible in principle (e.g. as in the case of ODEs) if both

- certain Ext's vanish and
- we can construct certain base covariances.

Or might be possible (much more speculative) to use the

- Ehrenpreis-Palamodov theorem.

Maxwell's Equations

The operator matrix

$$A := \begin{bmatrix} 0 & -\partial_z & \partial_y & \partial_t & 0 & 0 & 0 & 0 & 0 & 0 \\ \partial_z & 0 & -\partial_x & 0 & \partial_t & 0 & 0 & 0 & 0 & 0 \\ -\partial_y & \partial_x & 0 & 0 & 0 & \partial_t & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \partial_x & \partial_y & \partial_z & 0 & 0 & 0 & 0 \\ -\partial_t & 0 & 0 & 0 & -\partial_z & \partial_y & -1 & 0 & 0 & 0 \\ 0 & -\partial_t & 0 & \partial_z & 0 & -\partial_x & 0 & -1 & 0 & 0 \\ 0 & 0 & -\partial_t & -\partial_y & \partial_x & 0 & 0 & 0 & -1 & 0 \\ \partial_x & \partial_y & \partial_z & 0 & 0 & 0 & 0 & 0 & 0 & -1 \end{bmatrix}$$

acts on 3 components electrical field, 3 components magnetic (pseudo-)field, 3 components electric current and a component electric flux.

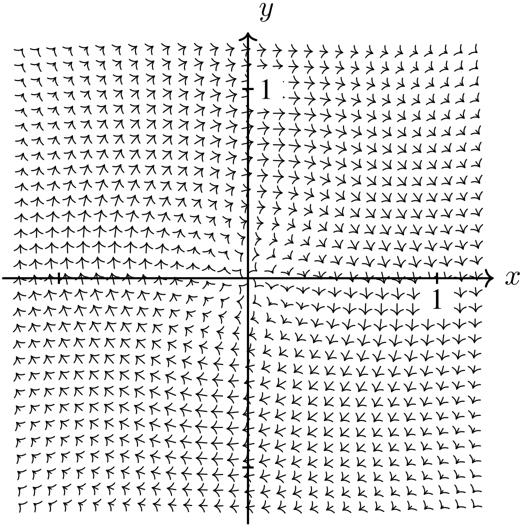
(constants := 1)

Maxwell's Equations

Electrical potential and magnetic potentials parametrize the solutions.

$$B := \begin{bmatrix} \partial_x & \partial_t & 0 & 0 \\ \partial_y & 0 & \partial_t & 0 \\ \partial_z & 0 & 0 & \partial_t \\ 0 & 0 & \partial_z & -\partial_y \\ 0 & -\partial_z & 0 & \partial_x \\ 0 & \partial_y & -\partial_x & 0 \\ -\partial_t\partial_x & \partial_y^2 + \partial_z^2 - \partial_t^2 & -\partial_y\partial_x & -\partial_z\partial_x \\ -\partial_t\partial_y & -\partial_y\partial_x & \partial_x^2 + \partial_z^2 - \partial_t^2 & -\partial_z\partial_y \\ -\partial_t\partial_z & -\partial_z\partial_x & -\partial_z\partial_y & \partial_x^2 + \partial_y^2 - \partial_t^2 \\ \partial_x^2 + \partial_y^2 + \partial_z^2 & \partial_t\partial_x & \partial_t\partial_y & \partial_t\partial_z \end{bmatrix}$$

Maxwell's Equations



The Koszul Complex

The matrix $A = \begin{bmatrix} x_1 & x_2 & x_3 \end{bmatrix}$ yields tangents of a sphere.

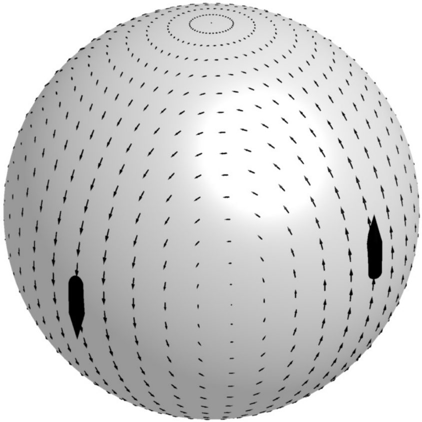
Parametrized by $B = \begin{bmatrix} 0 & x_3 & -x_2 \\ -x_3 & 0 & x_1 \\ x_2 & -x_1 & 0 \end{bmatrix}$.

Covariance function for tangential fields on the sphere:

$$\begin{bmatrix} -y_1y_2 - z_1z_2 & y_1x_2 & z_1x_2 \\ x_1y_2 & -x_1x_2 - z_1z_2 & z_1y_2 \\ x_1z_2 & y_1z_2 & -x_1x_2 - y_1y_2 \end{bmatrix} \cdot k$$

The Koszul Complex

Smooth field, conditioned at **4 points** at the equator, neighboring tangent vectors point into opposed directions (north/south).



Intersecting two Koszul Complexes

The matrix $A = \begin{bmatrix} \partial_1 & \partial_2 & \partial_3 \end{bmatrix}$ represents the divergence and its kernel is the rotation $B = \begin{bmatrix} 0 & \partial_3 & -\partial_2 \\ -\partial_3 & 0 & \partial_1 \\ \partial_2 & -\partial_1 & 0 \end{bmatrix}$.

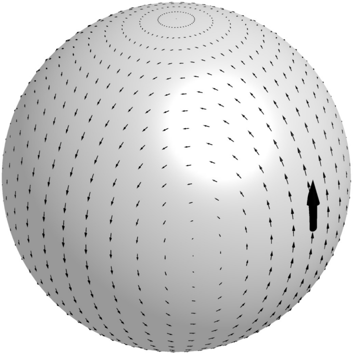
Intersecting parametrizations

We can intersect parametrizations via a pullback under suitable assumptions

Intersecting two Koszul Complexes

Intersection of tangent fields with divergence free fields.

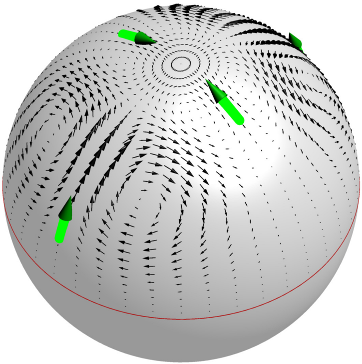
Data: 2 points opposed at the equator with tangents pointing north:



Parametrization of Dirichlet boundary conditions

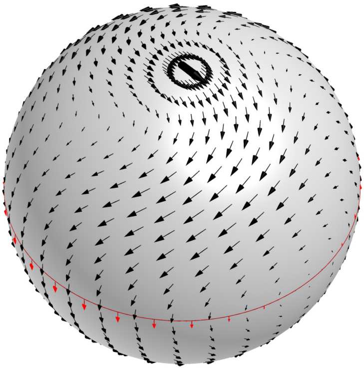
Functions vanishing on hyperplane $x_3 = 0$: $\langle x_3 \rangle \trianglelefteq \mathcal{F} = C^\infty(\mathbb{R}^d, \mathbb{R})$.

Intersect parametrizations via pullback.

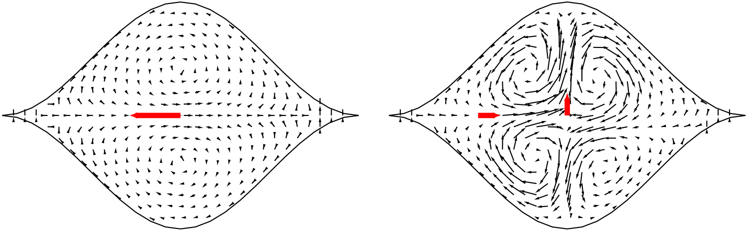


Inhomogeneous Boundary Conditions

Smooth divergence free fields f on the sphere and inhomogeneous boundary condition $f_3(x_1, x_2, 0) = x_2$.
Take particular solution $\mu = \begin{bmatrix} 0 & -x_3 & x_2 \end{bmatrix}^T$ as mean.



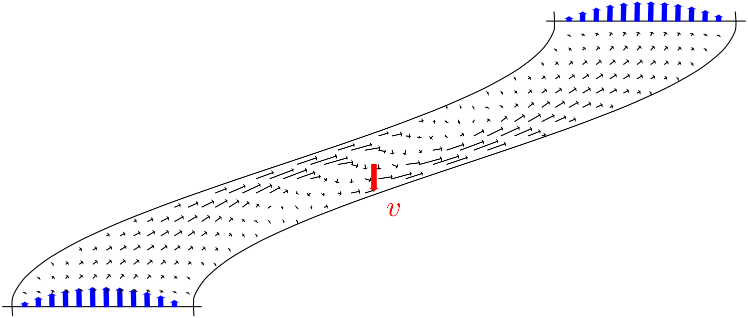
Analytic Boundary



Divergence-free fields bounded by $y^2 - \sin(x)^4$.

with Daniel Robertz.

Analytic Boundary with Analytic Boundary Conditions



Left&right boundary: zero flow.
Bottom resp. top: flow in resp. out.

with Daniel Robertz.

Message

Differential algebra and data

- GPs play nice with linear operators (in particular PDEs)
- Can be used to learn/understand systems
- Can be used as a very, *very* strong inductive bias
- Combines differential algebra with data

Summary

Thx!

Questions?

References:
On boundary conditions parametrized by analytic functions (1801.09197)
Linearly Constrained Gaussian Processes with Boundary Conditions (2002.00818)
Algorithmic Linearly Constrained Gaussian Processes (2205.03185)

Funded in part by the MKW.NRW through the project GAIA in the graduate School DataNinja

dataninja.nrw

Ministerium für
Kultur und Wissenschaft
des Landes Nordrhein-Westfalen