GP regression with inequality constraints Adaptive strategies

Sebastien Da Veiga
Joint work with Amandine Marrel (CEA)

ANR Chorus Workshop 30/04/2014

SAFRAN
Snecma

OUTLINE

- **Introduction**
- **GP regression with inequality constraints**
  - Theory
  - Examples
  - Adaptive strategies
- **Conclusion & outlook**

/01/

INTRODUCTION

INTRODUCTION

- **Surrogate models are now commonly used for emulating complex computer codes**
  - UQ, optimization, …
- **Very often, computer codes simulate real physical phenomena, which usually have specific properties**
  - Symmetries
  - Bound constraints (e.g. concentrations between 0 and 1, …)
  - Monotonicity w.r.t. some input variables
  - Solutions of PDEs (e.g. null Laplacian, divergence or curl free, …)
- **It is of great interest to incorporate such constraints in the proxy model**
  - Physics and expected behavior are respected (engineers like that !)
  - Predictions and robustness may be improved

INTRODUCTION

- **Incorporation of bounds and monotonicity constraints have already been studied in nonparametric regression**
  - 1D setting
    - *Ramsay 2005, Bigot and Gadat 2010*
  - Kernel regression
    - *Dette and Scheder 2006*
    - Constraints on weights: *Hall and Huang 2001, Racine et al. 2009*

- **Here, we focus on the GP regression framework**
  - Several recent papers on the topic …
  - … but no full-scale industrial application yet

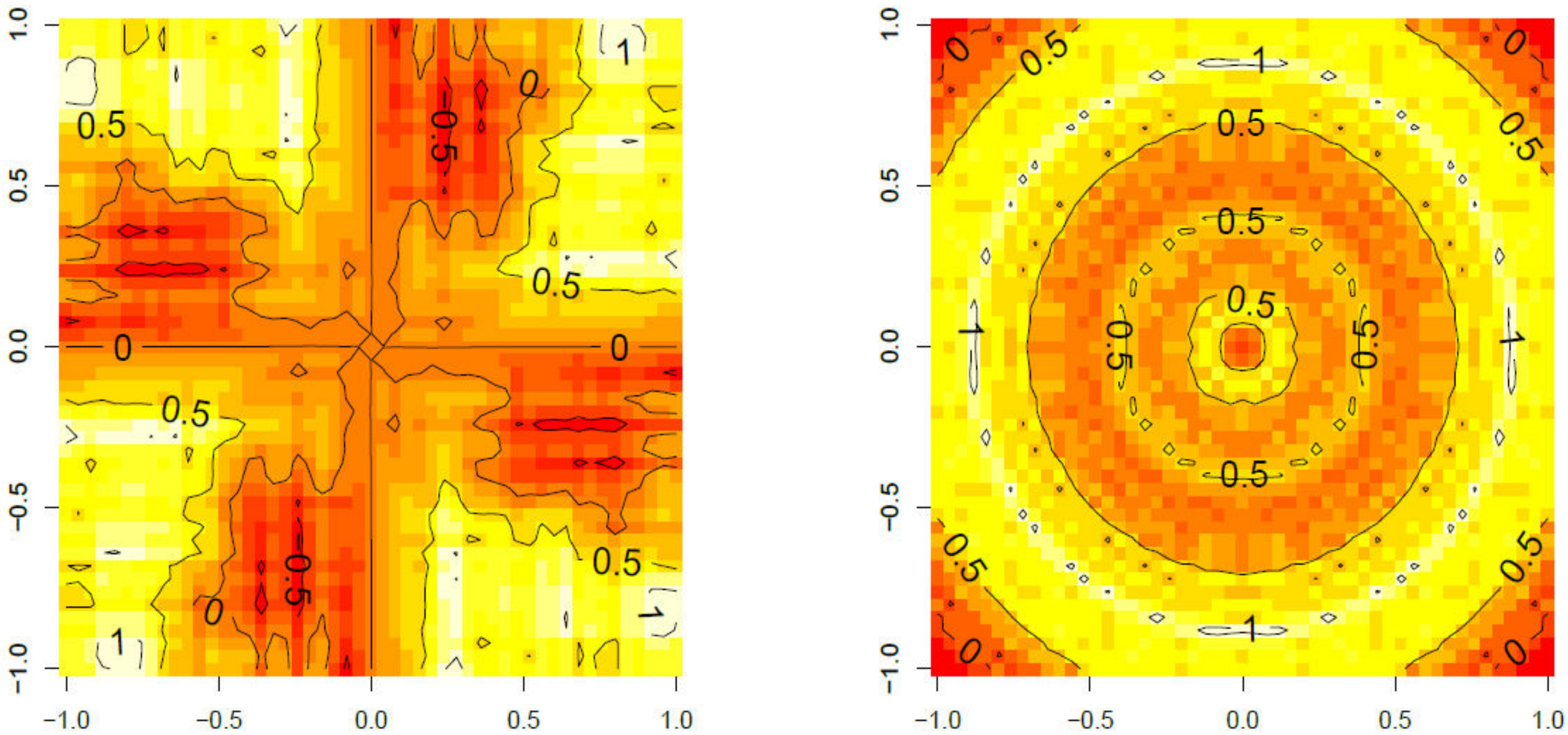
INTRODUCTION

➔ **The GP regression framework is very powerful when considering <u>linear equality</u> constraints**

- Gaussianity + linear constraints make it possible to design adapted covariance functions (kernels)
  - This produces trajectories that intrinsically respect the constraints
- This « simple » remark gave rise to several interesting examples
- General theory recently studied (*Ginsbourger et al. 2013*)

INTRODUCTION

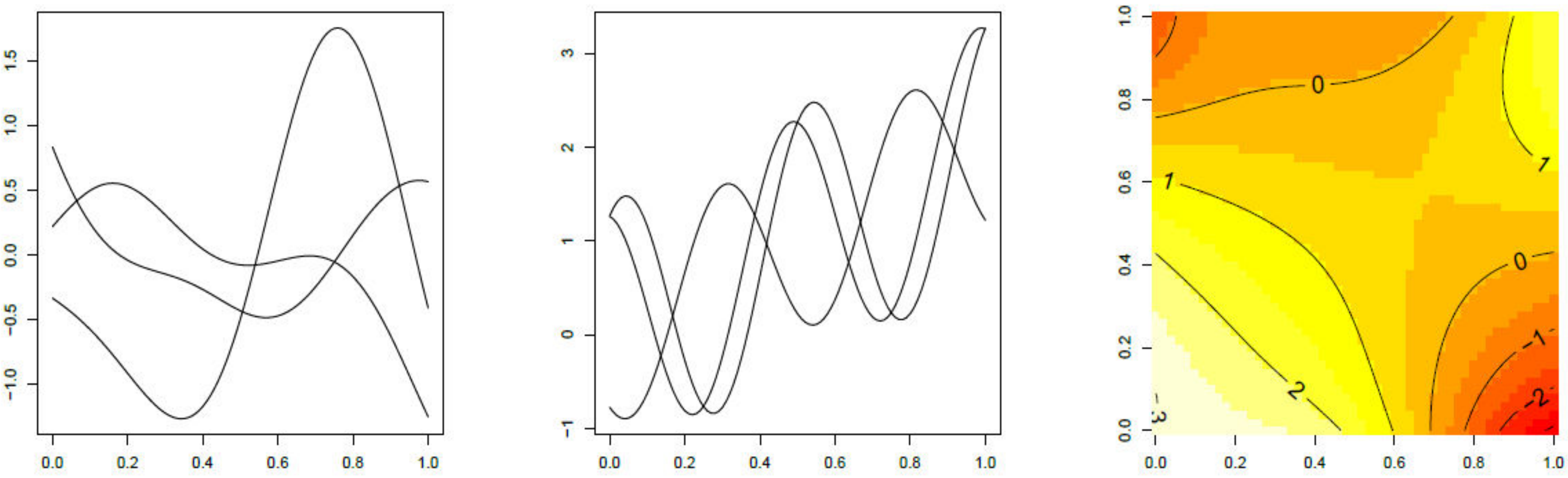
Sample paths of a GP with kernels designed for spatial symmetries



Ginsbourger et al. 2013

INTRODUCTION

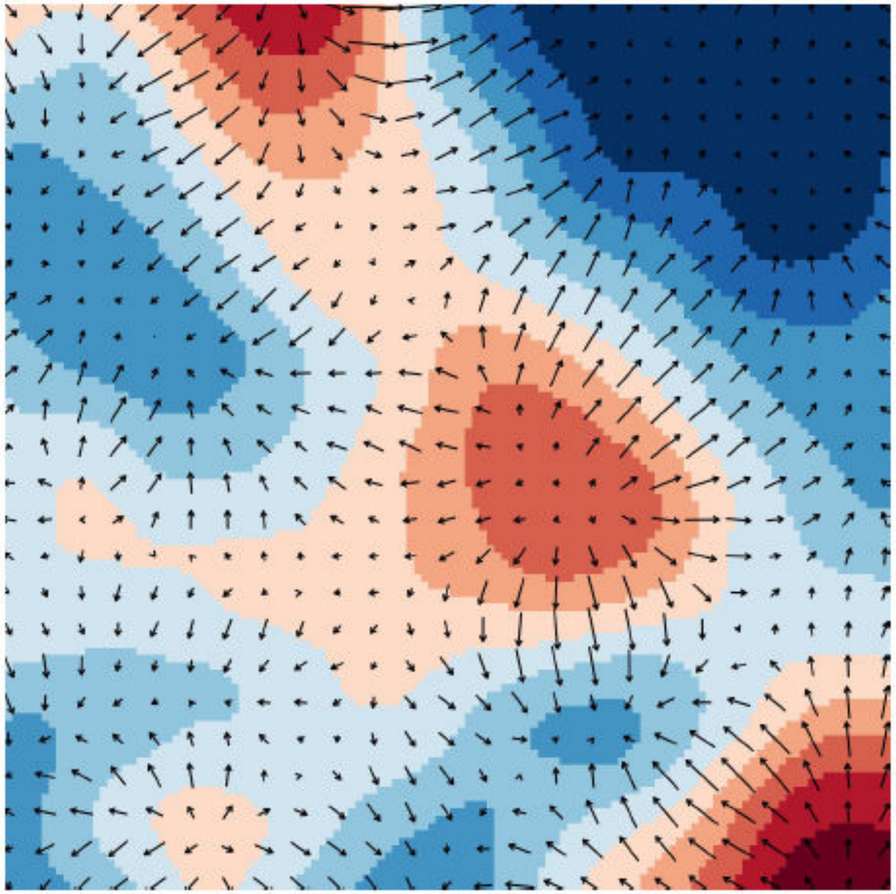
Sample paths of a GP with kernels designed for specific constraints (null integral, solution of ODE and null Laplacian)

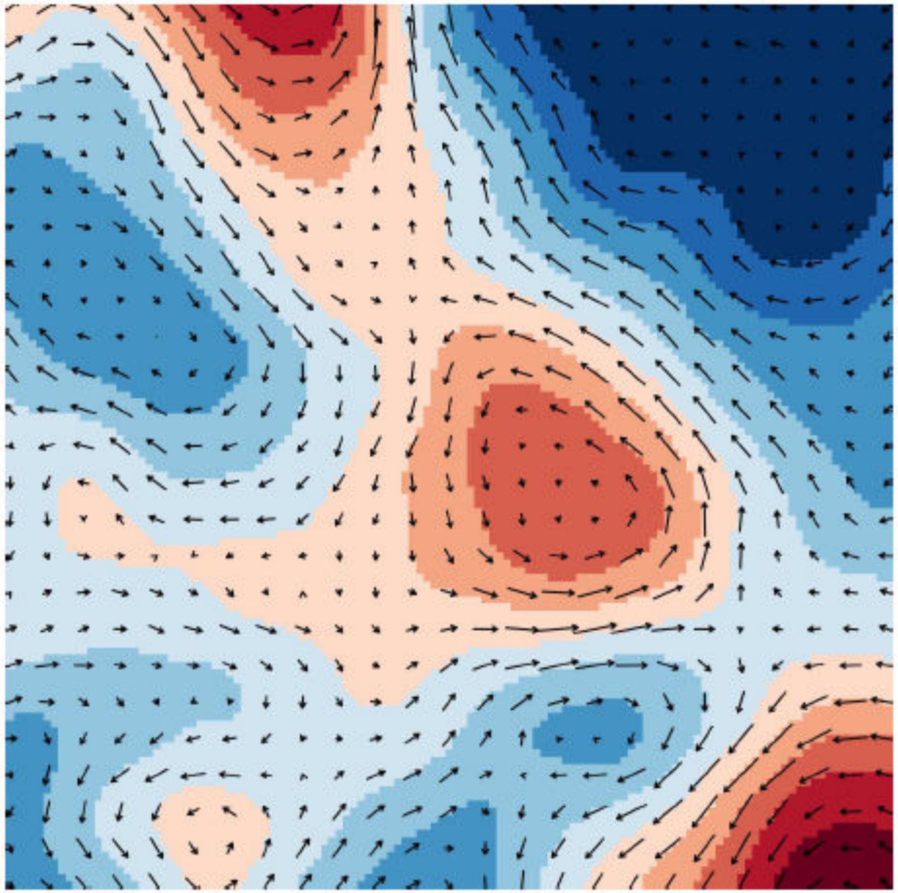


Ginsbourger et al. 2013

INTRODUCTION

Sample paths of a 2D-GP with kernels for curl-free and divergence free fields





Scheuerer and Schlather 2012

SAFRAN Snecma

INTRODUCTION

➔ **The GP regression framework is very powerful when considering linear equality constraints**

➔ **However, inequality constraints cannot be handled so easily**

- This included bound and monotonicity constraints
- But also bounds on integrals or divergence/curl

➔ **Previous work on GP regression with inequality constraints**

- Monotonicity
  - Data-augmentation: *Abrahamsen and Benth 2001*
  - Weights: *Yoo and Kyriadis 2006*
  - Sampling: *Michalak 2008, Kleijnen and van Beers 2010*
  - Constrained posterior distribution: *Riihimaki and Vehtari 2010, Wang and Berger 2011*
  - Expansion on a dedicated basis + constraints on weights: *Mattouk 2014*
- Any linear inequality constraints
  - Expectation of truncated normal distributions: *Da Veiga and Marrel 2012*

/02/

GP regression with inequality constraints

SAFRAN
Snecma

STANDARD GP REGRESSION

➔ Notations

- Computer code $f : \mathbb{R}^D \to \mathbb{R}$
- Inputs $\mathbf{x} = (x^1, \ldots, x^D) \in \mathbb{R}^D$
- Output $y = f(\mathbf{x})$
- Observations $(\mathbf{x}_i, y_i)_{i=1,\ldots,n}$ $X_s = [\mathbf{x}_1^T, \ldots, \mathbf{x}_n^T]^T$ $Y_s = [y_1, \ldots, y_n]^T$

➔ Model: Output seen as realization of stationary Gaussian process

$$Y(\mathbf{x}) = f_0(\mathbf{x}) + Z(\mathbf{x})$$

$$f_0(\mathbf{x}) = \sum_{j=1}^{J} \beta_j f_j(\mathbf{x}) = F(x)\beta \qquad C(\boldsymbol{\tau}) = \sigma^2 R(\boldsymbol{\tau})$$

➔ Conditioning

- MLE estimates

$$\hat{\beta} = (F_s R\psi^{-1} F_s)^{-1} F_s^T R_\psi^{-1} Y_s \qquad \widehat{\sigma^2} = \frac{1}{n}(Y_s - F_s\hat{\beta})^T R_\psi^{-1} (Y_s - F_s\hat{\beta}) \qquad \psi^* = \arg\min_{\psi} \widehat{\sigma^2} \det(R_\psi)^{\frac{1}{n}}$$

- Predictor

$$\tilde{\mu} = \mathbb{E}\left(\tilde{Y}(\mathbf{x}^*)\right)$$

$$\tilde{\mu} = F(\mathbf{x}^*)\hat{\beta} + k(\mathbf{x}^*)^T \Sigma_S^{-1} \left(Y_s - F_s\hat{\beta}\right)$$

GP REGRESSION WITH INEQUALITY CONSTRAINTS

➔ **To incorporate the constraints, we propose to keep the conditional expectation framework**

- Predictions are equal to the expectation of the GP (conditioned at the observations) given that it respects the inequality constraints

➔ **For example, the corresponding predictor for bound constraints may be**

$$\mathbb{E}\left(\tilde{Y}(\mathbf{x}^*)|\forall \mathbf{x} \in I, a \leq \tilde{Y}(\mathbf{x}) \leq b\right)$$

- Note the link with with extrema of random fields …

$$\mathbb{E}\left(\tilde{Y}(\mathbf{x}^*)|\min_{\mathbf{x}\in I} \tilde{Y}(\mathbf{x}) \geq a, \max_{\mathbf{x}\in I} \tilde{Y}(\mathbf{x}) \leq b\right)$$

- … but no tractable formula exists for joint distributions in the general case

GP REGRESSION WITH INEQUALITY CONSTRAINTS

➔ **We thus propose a discrete-location approximation:**

$$\mathbb{E}\left(\tilde{Y}(\mathbf{x}^*)|\forall \mathbf{x} \in I, a \leq \tilde{Y}(\mathbf{x}) \leq b\right)$$

⬇

$$\mathbb{E}\left(\tilde{Y}(\mathbf{x}^*)|\forall i = 1, \ldots, N,\ a \leq \tilde{Y}(\mathbf{x}_i) \leq b\right)$$

- Same approximation in *Riihimaki and Vehtari 2010, Wang and Berger 2011*

➔ **This generalizes easily to other constraints**

$$\mathbb{E}\left(\tilde{Y}(\mathbf{x}^*)|\forall i = 1, \ldots, N,\ \frac{\partial \tilde{Y}}{\partial x^j}(\mathbf{x}_i) \geq 0\right)$$

Monotonicity

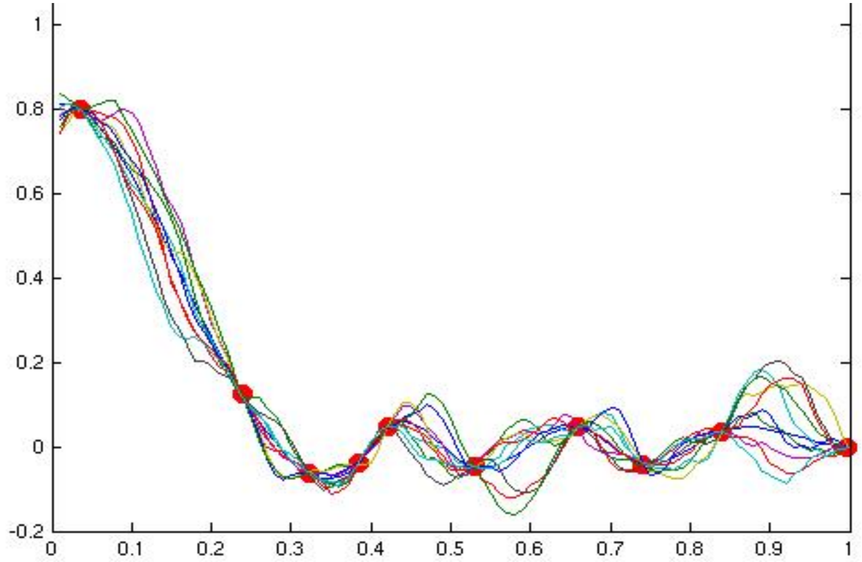
$$\mathbb{E}\left(\tilde{Y}(\mathbf{x}^*)|\sum_{i=1}^{N} w_i \tilde{Y}(\mathbf{x}_i) \leq M\right)$$

Conservation

GP REGRESSION WITH INEQUALITY CONSTRAINTS

Standard framework:

- Take all trajectories which interpolate the observations
- Compute the average to get the kriging predictor
- (If desired, the variance yields a measure of accuracy)

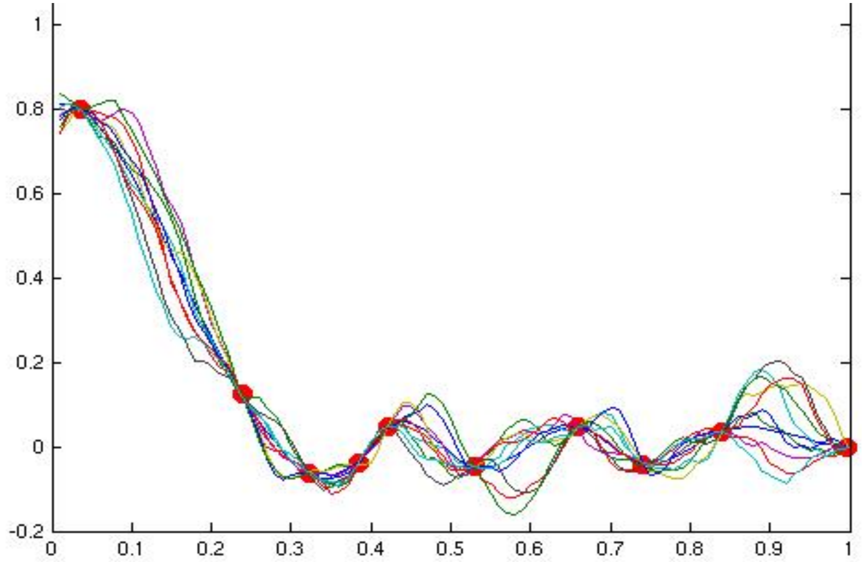


SAFRAN Snecma

GP REGRESSION WITH INEQUALITY CONSTRAINTS

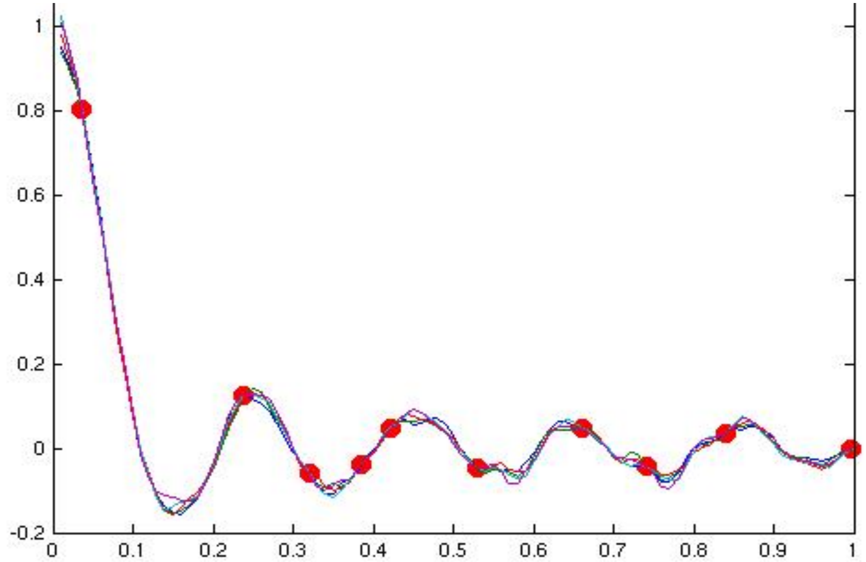
Standard framework:

- Take all trajectories which interpolate the observations
- Compute the average to get the kriging predictor
- (If desired, the variance yields a measure of accuracy)



Here:

- Take all trajectories which interpolate the observations
- Select those which respect the constraints of bounds, monotonicity, …
- Compute the average to get the new kriging predictor
- (If desired, the variance yields a measure of accuracy)



SAFRAN Snecma

GP REGRESSION WITH INEQUALITY CONSTRAINTS

➔ **But how can we compute such expectations ?**

➔ **This is where the linearity assumption comes into play**

- Bounds, monotonicity, integral, divergence/curl constraints are linear w.r.t. the output
- The GP obtained by stacking the output and the quantities related to the constraints is then a GP too
- The problem reduces to compute moments of a multivariate normal vector subject to linear equality constraints
  - Truncated normal distribution

GP REGRESSION WITH INEQUALITY CONSTRAINTS

➔ The truncated multivariate normal distribution

- Given a multivariate normal vector …

$$\mathbf{Z} = (Z_1, \ldots, Z_p) \qquad \phi_{\mu,\Sigma}(\mathbf{z}) = \frac{1}{(2\pi)^{p/2}\det(\Sigma)^{1/2}} \exp\left(-\frac{1}{2}(\mathbf{z}-\mu)^T\Sigma^{-1}(\mathbf{z}-\mu)\right)$$

- … its truncated version has the following p.d.f.

$$\phi_{\mu,\Sigma,\mathbf{a},\mathbf{b}}(\mathbf{z}) = \begin{cases} \frac{\phi_{\mu,\Sigma}(\mathbf{z})}{\mathbb{P}(\mathbf{a}\leq\mathbf{Z}\leq\mathbf{b})}, & \text{for } \mathbf{a} \leq \mathbf{z} \leq \mathbf{b}, \\ 0, & \text{otherwise.} \end{cases}$$

- Its expectation is given by

$$\mathbb{E}(Z_i | \mathbf{a} \leq \mathbf{Z} \leq \mathbf{b}) = \mu + \sum_{k=1}^{p} \sigma_{ik} \left(F_k(a_k) - F_k(b_k)\right)$$

$$F_i(z) = \int_{a_1}^{b_1} \cdots \int_{a_{i-1}}^{b_{i-1}} \int_{a_{i+1}}^{b_{i+1}} \cdots \int_{a_p}^{b_p} \phi_{\mu,\Sigma,\mathbf{a},\mathbf{b}}(z_1, \ldots, z_{i-1}, z, z_{i+1}, \ldots, z_p) dz_1 \ldots dz_{i-1} dz_{i+1} \ldots dz_p$$

- Other formulas for the covariance, linear and elliptical constraints available since the 60's (*Tallis 61, Tallis 63, Tallis 65*)

GP REGRESSION WITH INEQUALITY CONSTRAINTS

- **The truncated multivariate normal distribution**

$$\mathbb{E}(Z_i|\mathbf{a} \leq \mathbf{Z} \leq \mathbf{b}) = \mu + \sum_{k=1}^{p} \sigma_{ik} \left(F_k(a_k) - F_k(b_k)\right)$$

- **Available formulas involve Gaussian integrals with dimensionality equal to the number of points where we impose the constraints**

GP REGRESSION WITH INEQUALITY CONSTRAINTS

- **The truncated multivariate normal distribution**

$$\mathbb{E}(Z_i|\mathbf{a} \leq \mathbf{Z} \leq \mathbf{b}) = \mu + \sum_{k=1}^{p} \sigma_{ik} \left(F_k(a_k) - F_k(b_k)\right)$$

- **Available formulas involve Gaussian integrals with dimensionality equal to the number of points where we impose the constraints**

- **We thus need efficient approximations when this number is large (as it should be !)**
  - Genz numerical approximation of Gaussian integrals (*Genz 92*)
    - Cholesky decomposition + QMC integration: up to 1000 points
  - Sampling from a truncated Gaussian
    - Gibbs sampler (*Geweke 91, Robert 95*) + fast univariate sampler: up to 1000 points
  - Correlation-free formula (« crude » covariance tapering)

$$\mathbb{E}(Z_1|a_1 \leq Z_1 \leq b_1) = \mu_1 + \frac{\phi(\frac{a_1-\mu_1}{\sigma_{11}}) - \phi(\frac{b_1-\mu_1}{\sigma_{11}})}{\Phi(\frac{b_1-\mu_1}{\sigma_{11}}) - \Phi(\frac{a_1-\mu_1}{\sigma_{11}})}\sigma_{11}$$

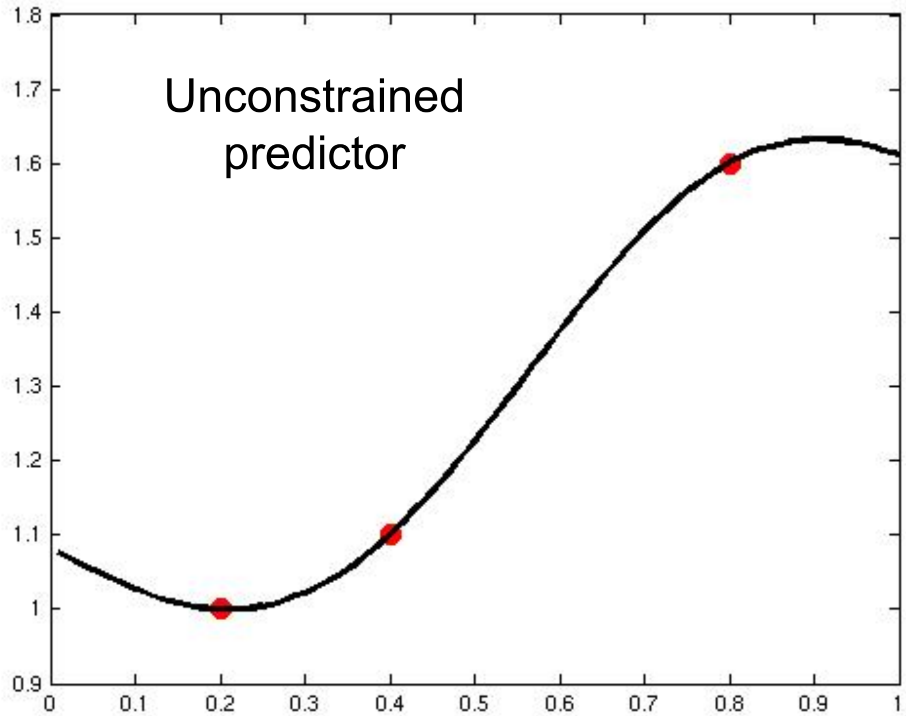
GP REGRESSION WITH INEQUALITY CONSTRAINTS

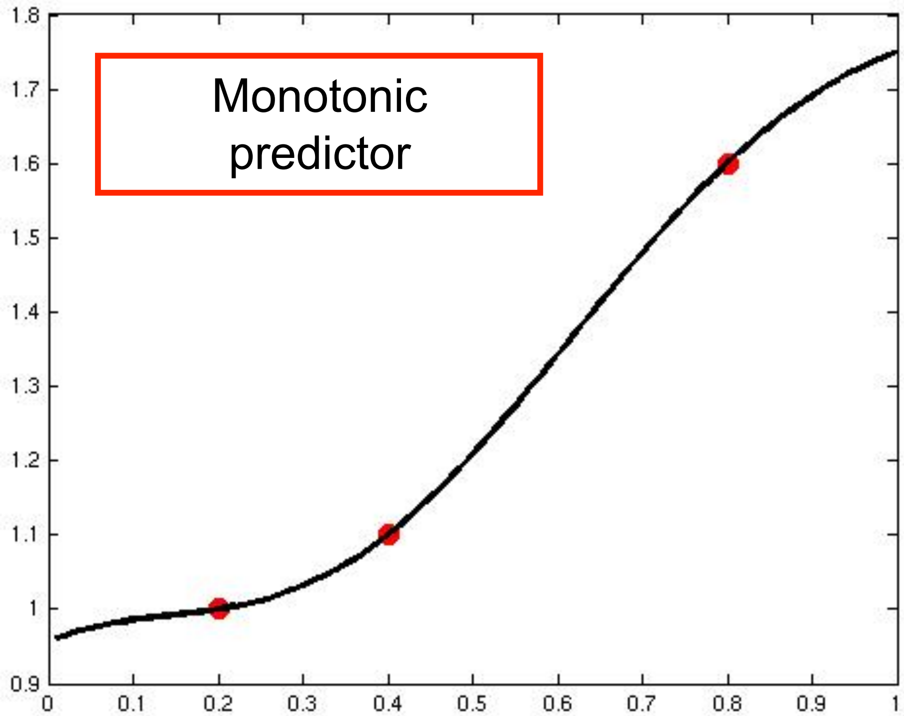
➔ In practice

- Train the standard GP surrogate on the observations: Ycond
- Set up you constraints
  - Compute the full covariance matrix of Ycond and Zconst where Zconst is the GP on the quantity which must be constrained (Y, its derivatives, its integral, …)
  - Select the constraint points (e.g. equally spaced on a grid, or optimized LHS)
- Compute the expectation of the conditioned GP at the constraint points subject to truncation
- The final predictor is obtained by further conditioning Ycond given that Zconst is equal to the above expectation (*Kotz et al. 2000*)

EXAMPLES

Simple incorporation of monotonicity on 100 equally-spaced constraint points

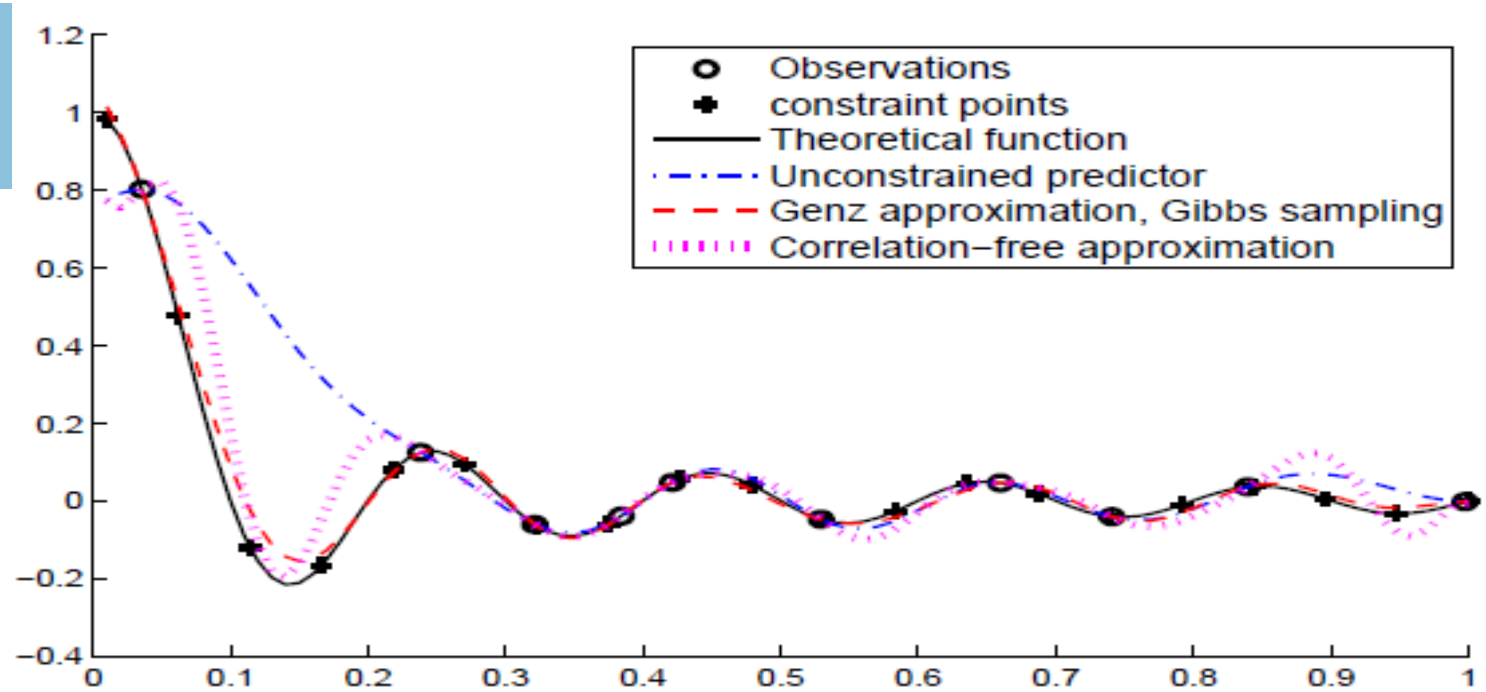




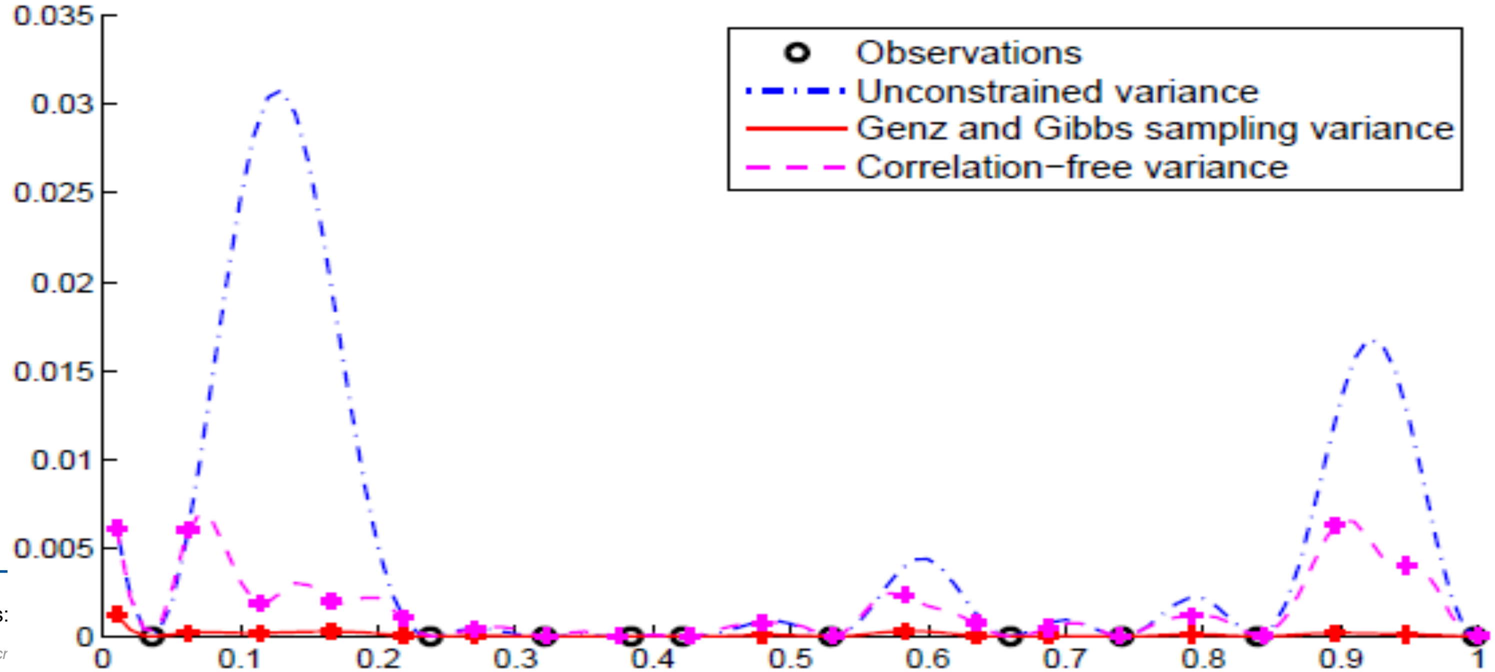
SAFRAN Snecma

EXAMPLES

Bounded predictor



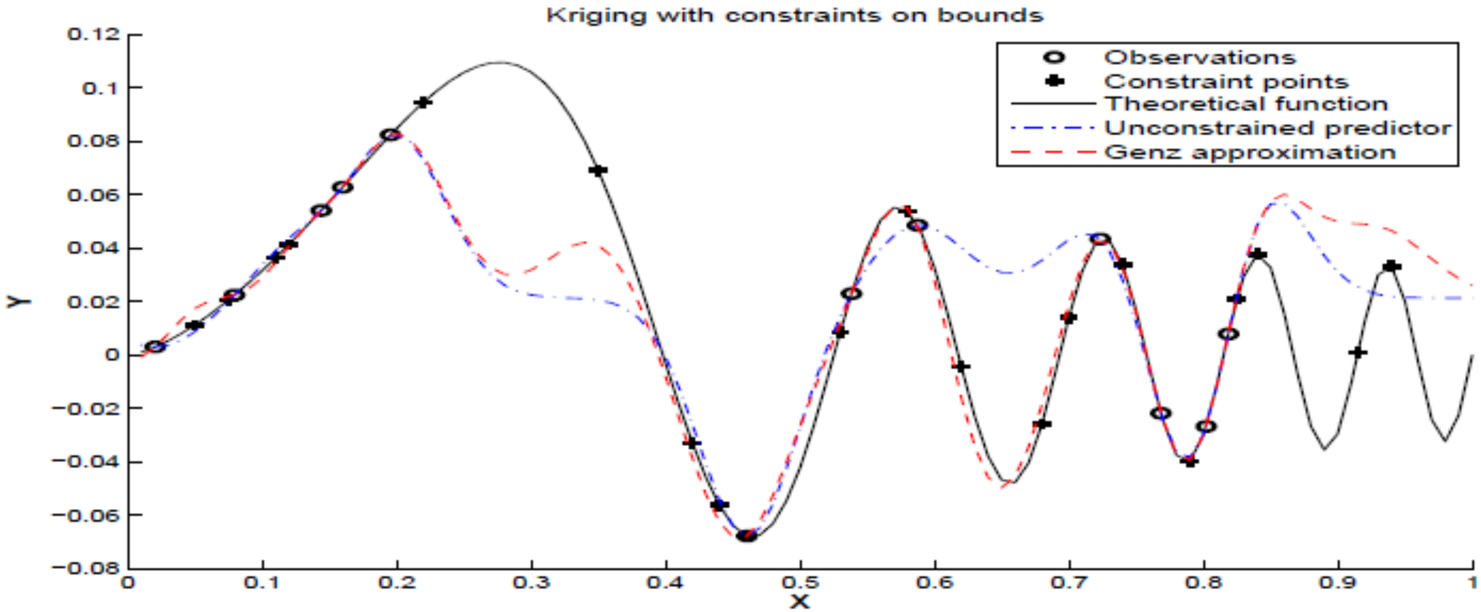
Prediction variance



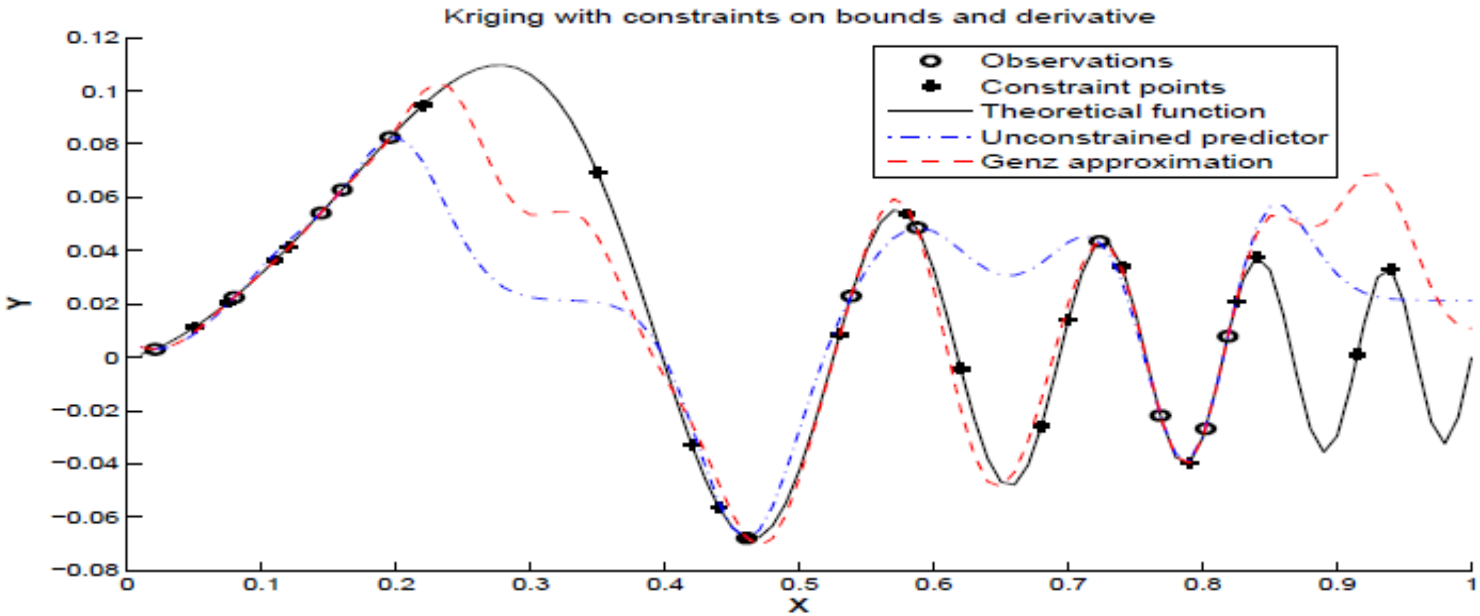
Da Veiga & Marrel 2012

EXAMPLES

Bounded predictor



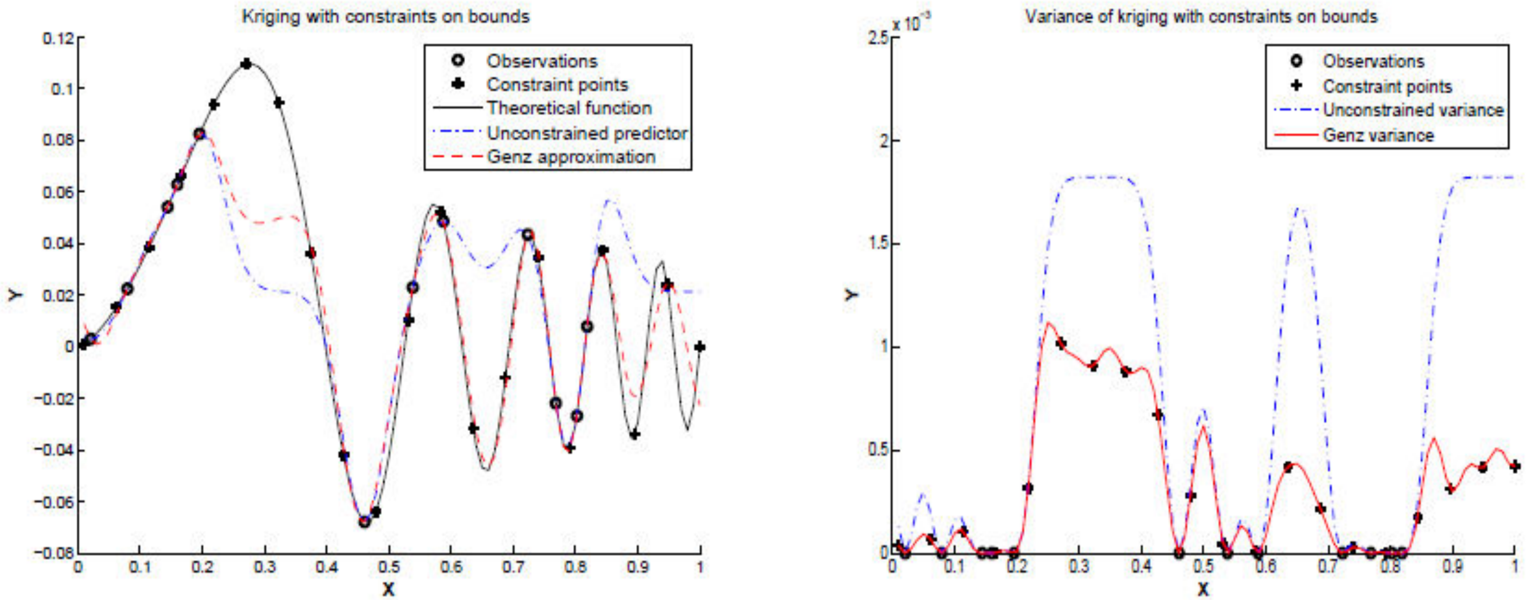
Bounded predictor with bounded derivative



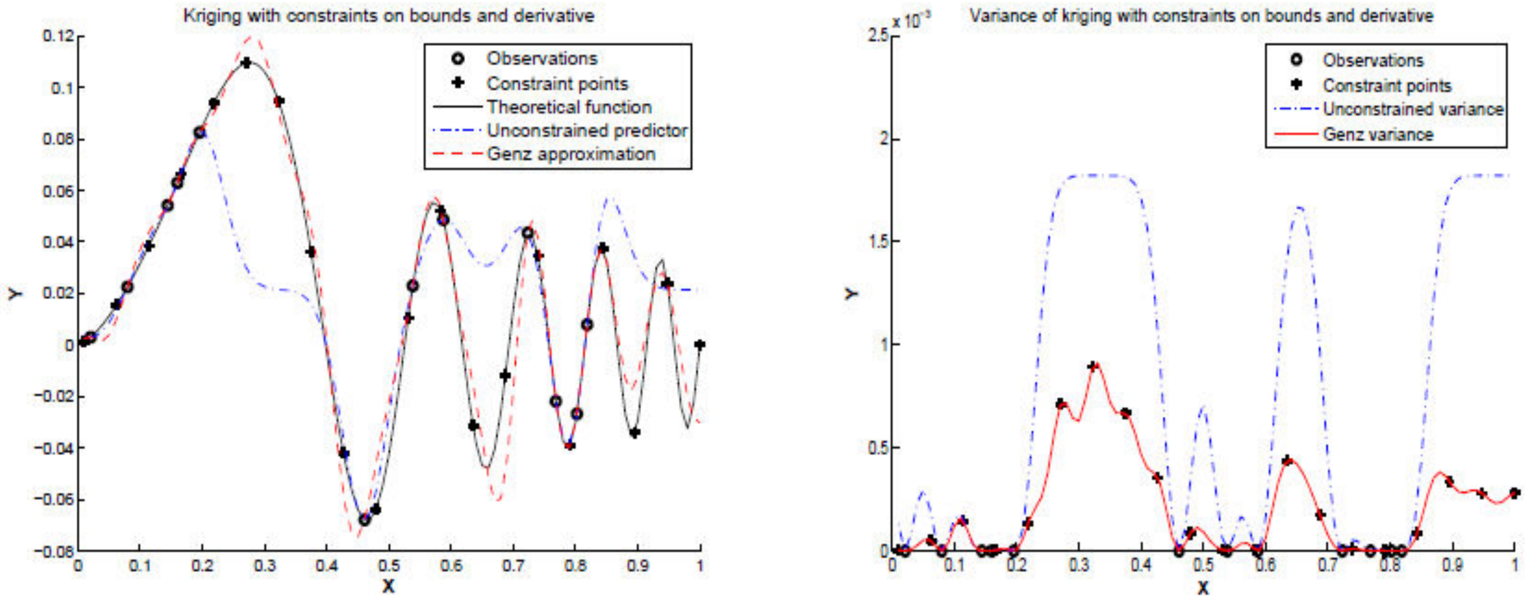
Da Veiga & Marrel 2012

EXAMPLES

Bounded predictor



Bounded predictor with bounded derivative



Da Veiga & Marrel 2012

SAFRAN Snecma

GP REGRESSION WITH INEQUALITY CONSTRAINTS

➔ **Additional results available in our paper**

- Extensive 1D studies with several kernels
- One 2D example

➔ **But efficient generalization to higher dimensional problems is not so easy**

- From a theoretical perspective, no change in the formulas
- However, « spanning » the subset where we impose constraints will necessitate much more constraint points in the discrete-location approximation
  - Genz numerical integration and sampling cannot be used with several thousands of constraints

SAFRAN Snecma

GP REGRESSION WITH INEQUALITY CONSTRAINTS

➔ Additional results available in our paper

- Extensive 1D studies with several kernels
- One 2D example

➔ But efficient generalization to higher dimensional problems is not so easy

- From a theoretical perspective, no change in the formulas
- However, « spanning » the subset where we impose constraints will necessitate much more constraint points in the discrete-location approximation
  - Genz numerical integration and sampling cannot be used with several thousands of constraints

- Our idea is to use the correlation induced among the constraint points (and with the observations)
  - It is not necessary to place constraint points where the predictor has a high probability to respect the constraints (e.g. close to another constraint point, or where the prediction variance is very low)

GP REGRESSION WITH INEQUALITY CONSTRAINTS

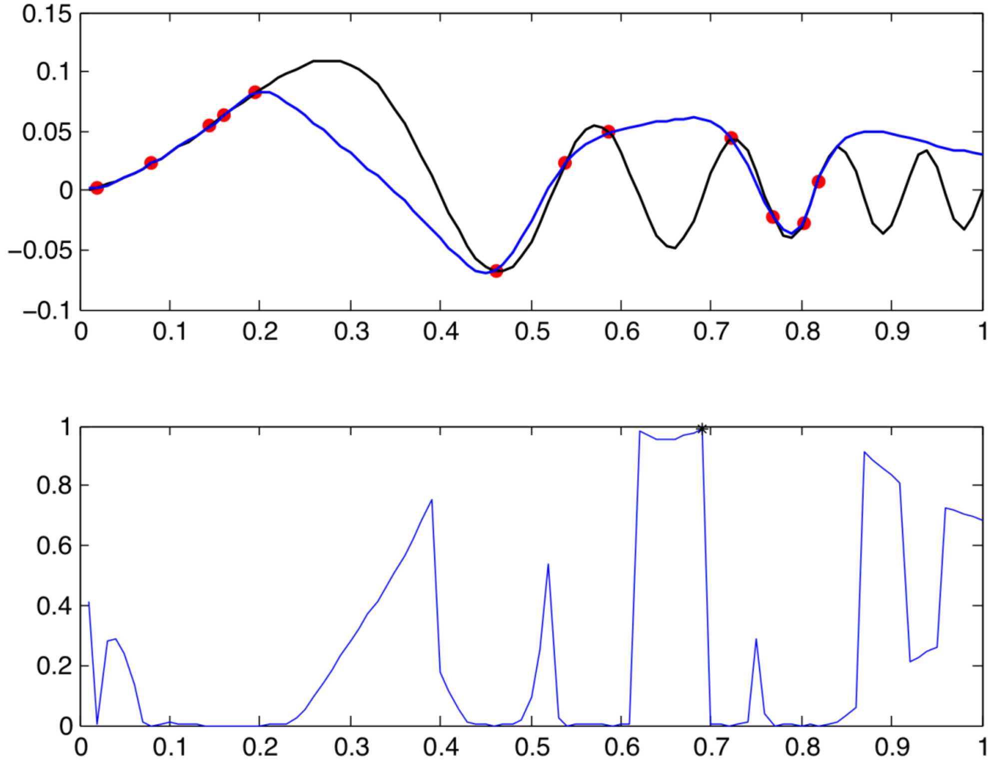
➔ **This motivates the design of an adaptive strategy for choosing the constraints locations**

- In the GP framework, it is straightforward to compute the probability that the GP does not respect the constraints at any location
- Constraint points are thus added one at a time, at locations where this probability is the highest

SAFRAN Snecma

GP REGRESSION WITH INEQUALITY CONSTRAINTS

- **This motivates the design of an adaptive strategy for choosing the constraints locations**
  - In the GP framework, it is straightforward to compute the probability that the GP does not respect the constraints at any location
  - Constraint points are thus added one at a time, at locations where this probability is the highest

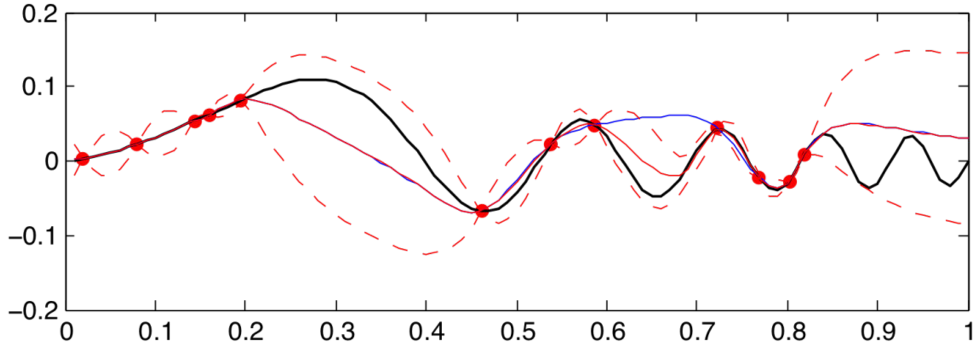


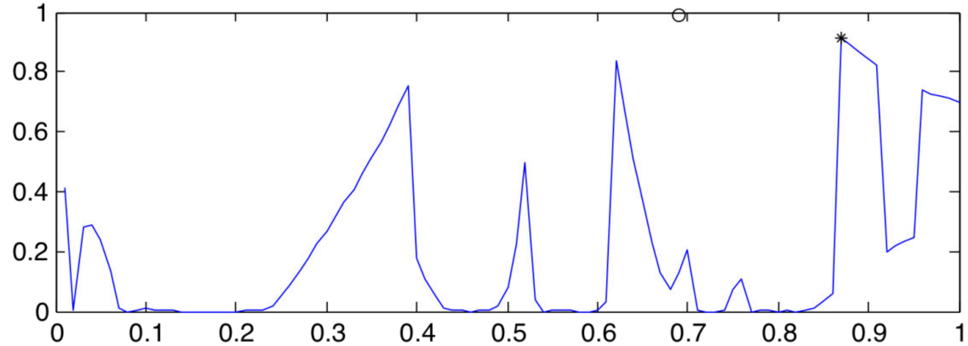
SAFRAN
Snecma

GP REGRESSION WITH INEQUALITY CONSTRAINTS

➔ **This motivates the design of an adaptive strategy for choosing the constraints locations**

- In the GP framework, it is straightforward to compute the probability that the GP does not respect the constraints at any location
- Constraint points are thus added one at a time, at locations where this probability is the highest



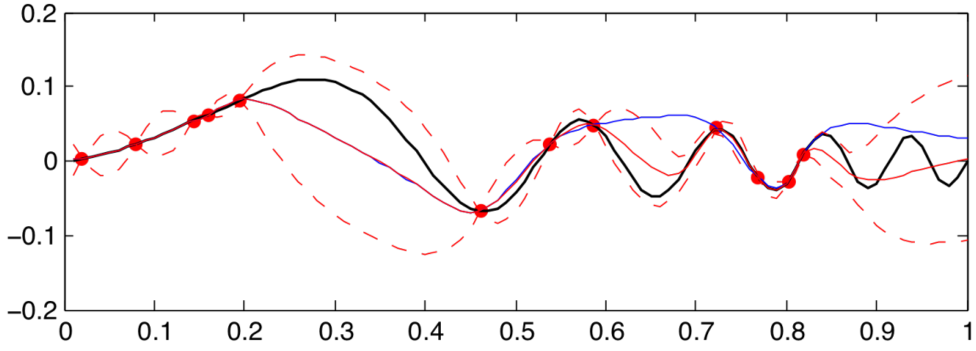


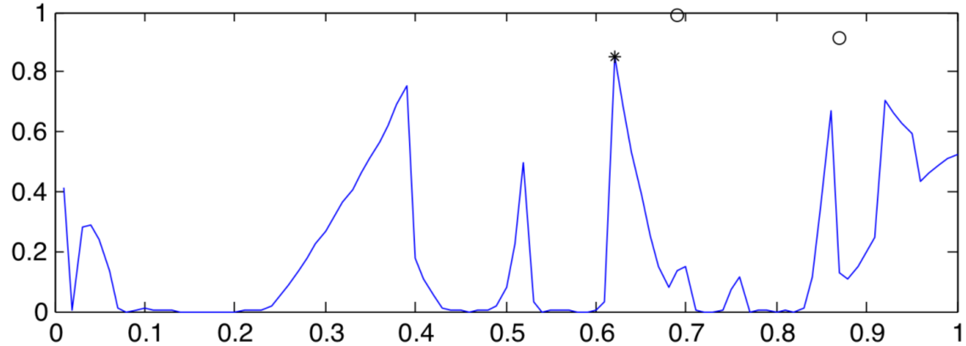
SAFRAN Snecma

GP REGRESSION WITH INEQUALITY CONSTRAINTS

➔ **This motivates the design of an adaptive strategy for choosing the constraints locations**

- In the GP framework, it is straightforward to compute the probability that the GP does not respect the constraints at any location
- Constraint points are thus added one at a time, at locations where this probability is the highest



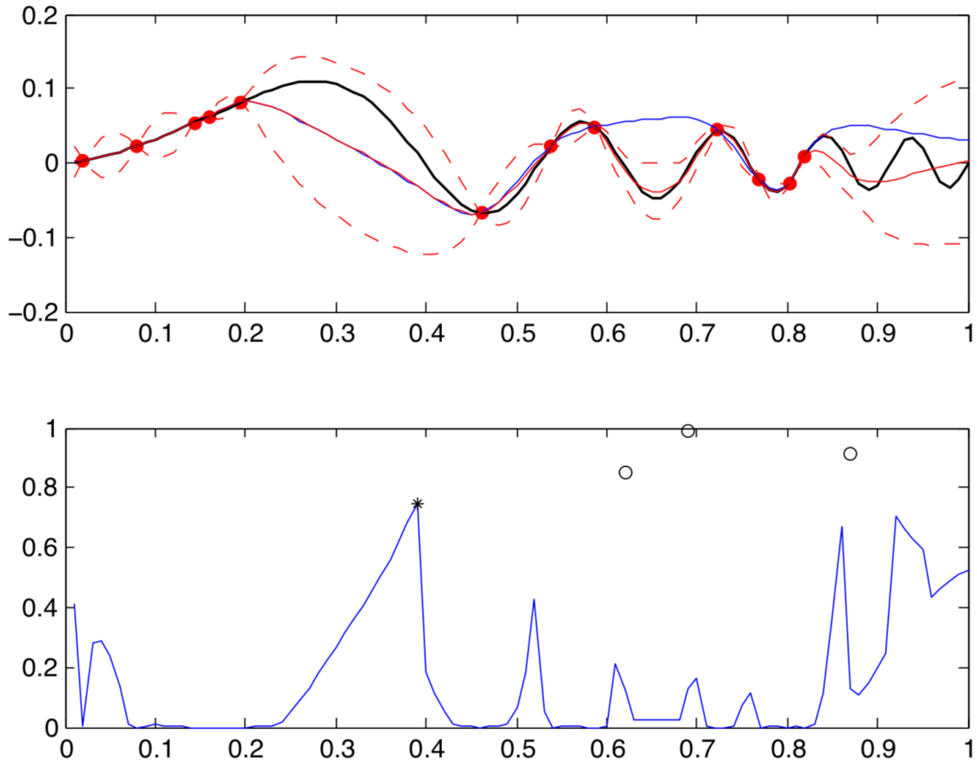


SAFRAN Snecma

GP REGRESSION WITH INEQUALITY CONSTRAINTS

➔ **This motivates the design of an adaptive strategy for choosing the constraints locations**

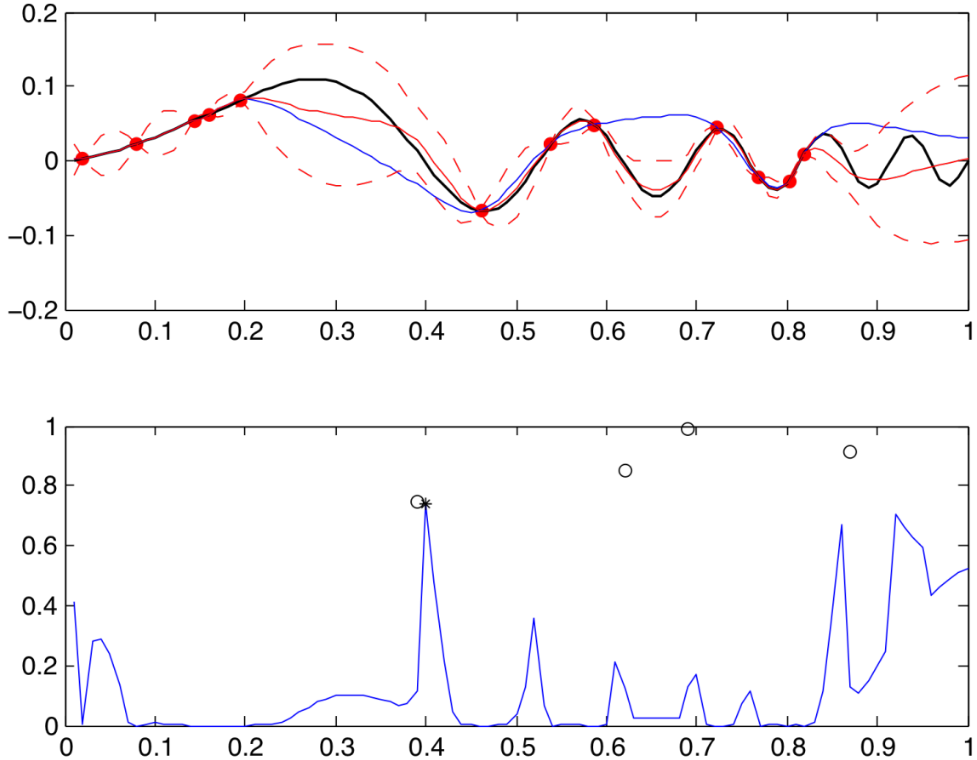
- In the GP framework, it is straightforward to compute the probability that the GP does not respect the constraints at any location
- Constraint points are thus added one at a time, at locations where this probability is the highest



SAFRAN Snecma

GP REGRESSION WITH INEQUALITY CONSTRAINTS

- **This motivates the design of an adaptive strategy for choosing the constraints locations**
  - In the GP framework, it is straightforward to compute the probability that the GP does not respect the constraints at any location
  - Constraint points are thus added one at a time, at locations where this probability is the highest

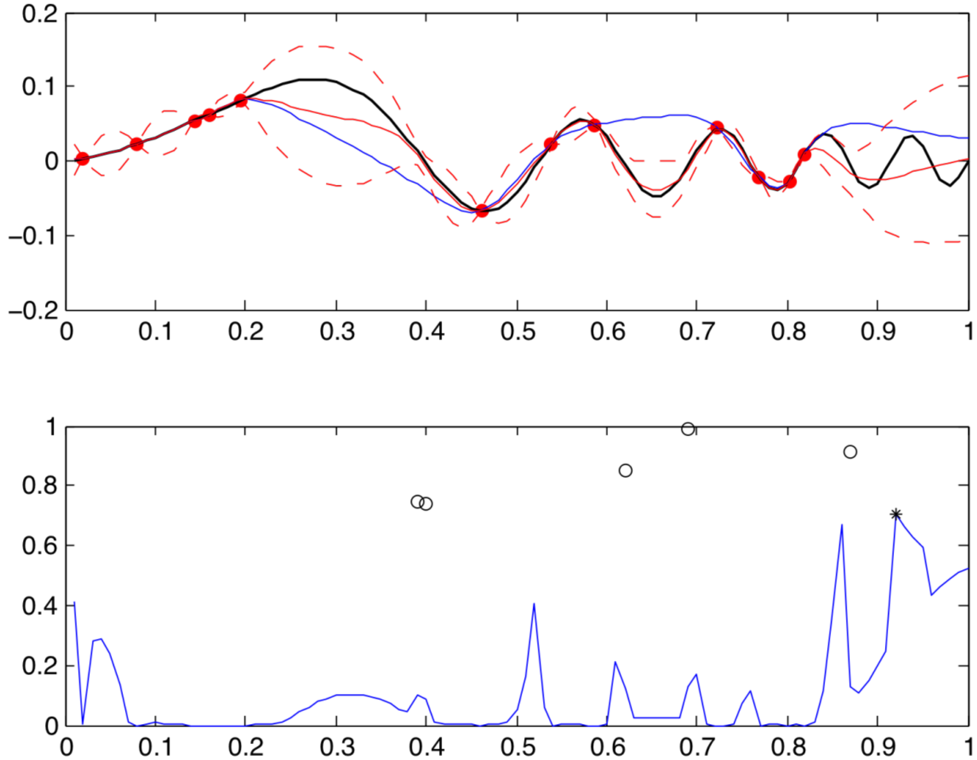


SAFRAN Snecma

GP REGRESSION WITH INEQUALITY CONSTRAINTS

➔ **This motivates the design of an adaptive strategy for choosing the constraints locations**

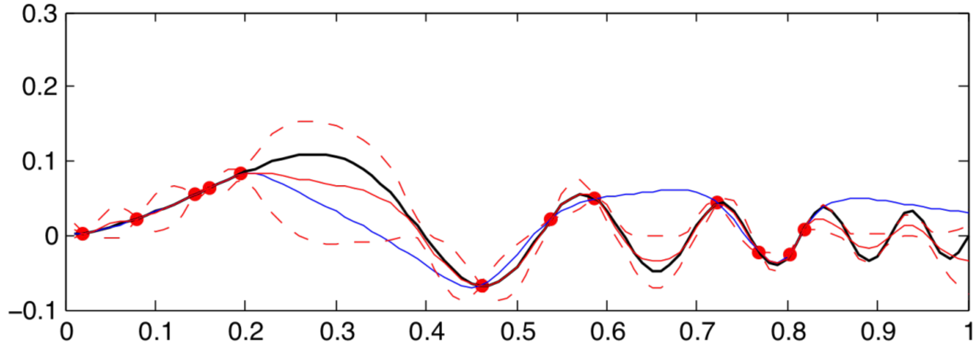
- In the GP framework, it is straightforward to compute the probability that the GP does not respect the constraints at any location
- Constraint points are thus added one at a time, at locations where this probability is the highest

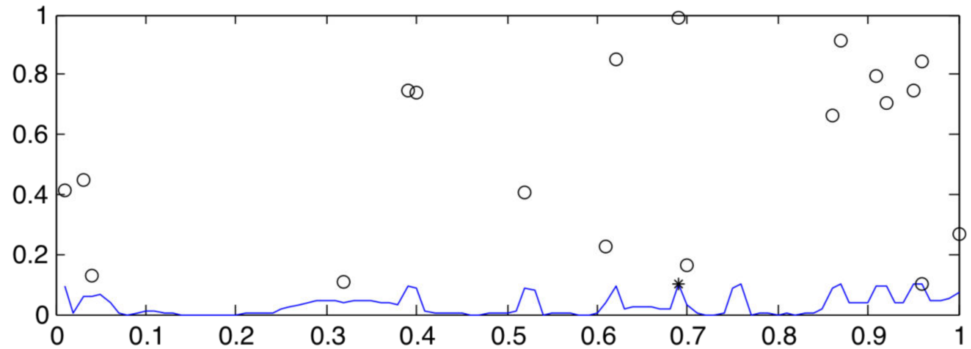


GP REGRESSION WITH INEQUALITY CONSTRAINTS

➔ **This motivates the design of an adaptive strategy for choosing the constraints locations**

- In the GP framework, it is straightforward to compute the probability that the GP does not respect the constraints at any location
- Constraint points are thus added one at a time, at locations where this probability is the highest



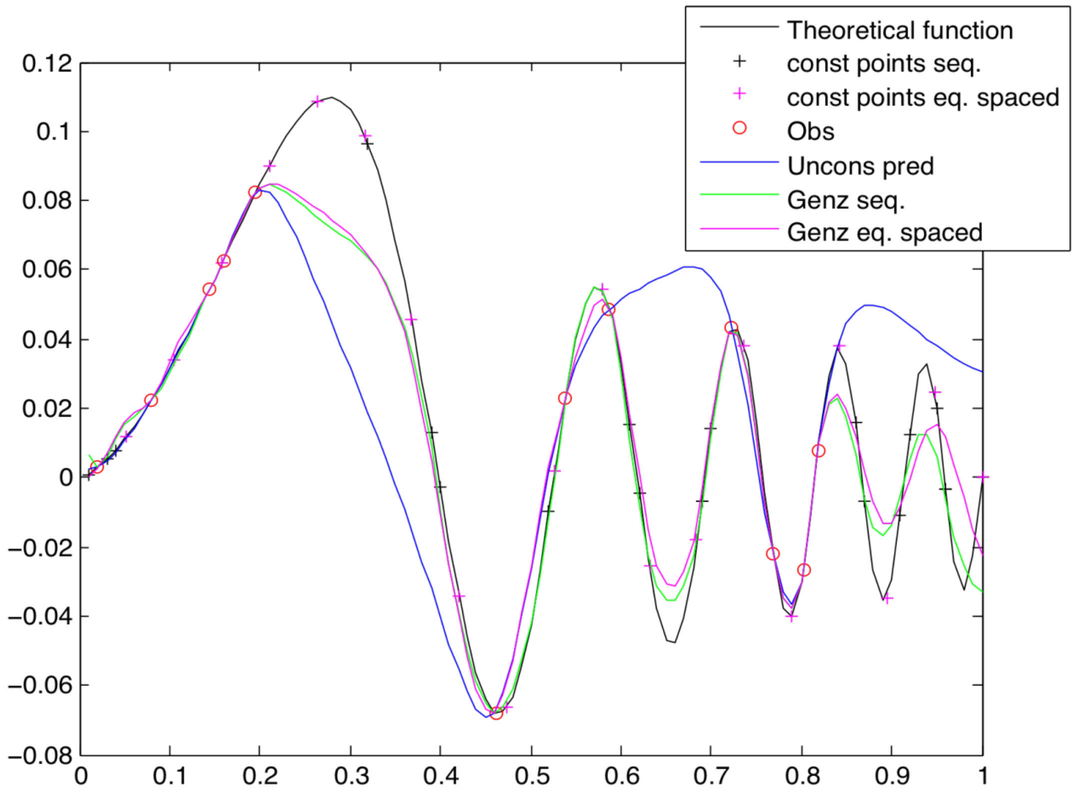


SAFRAN
Snecma

GP REGRESSION WITH INEQUALITY CONSTRAINTS

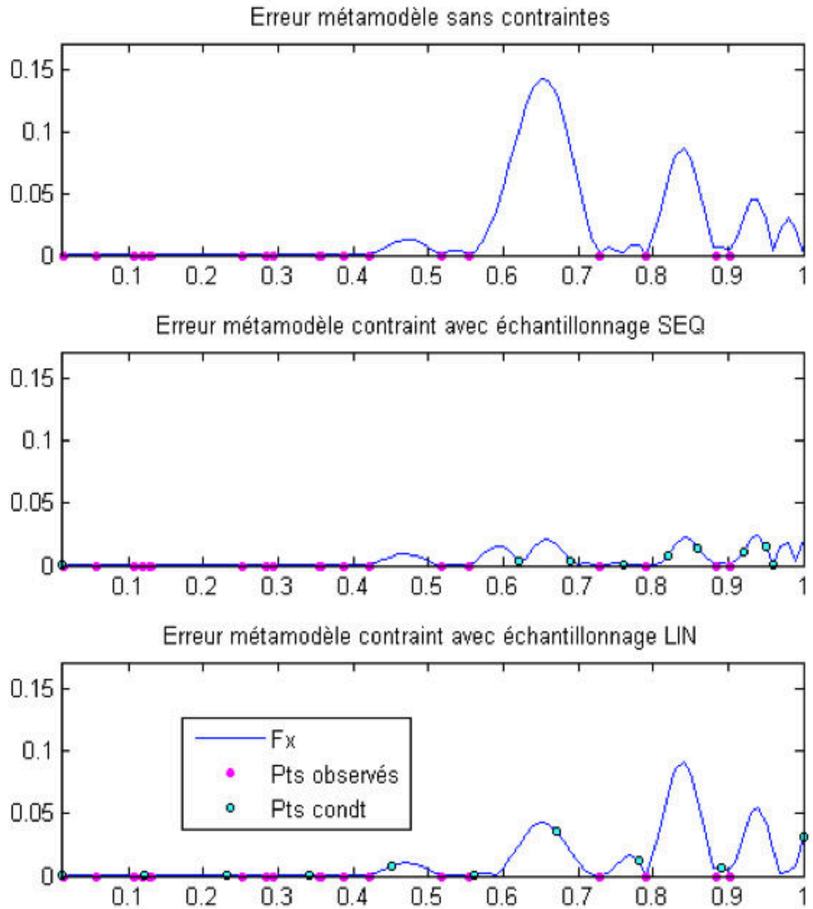
➔ **This motivates the design of an adaptive strategy for choosing the constraints locations**

- In the GP framework, it is straightforward to compute the probability that the GP does not respect the constraints at any location
- Constraint points are thus added one at a time, at locations where this probability is the highest

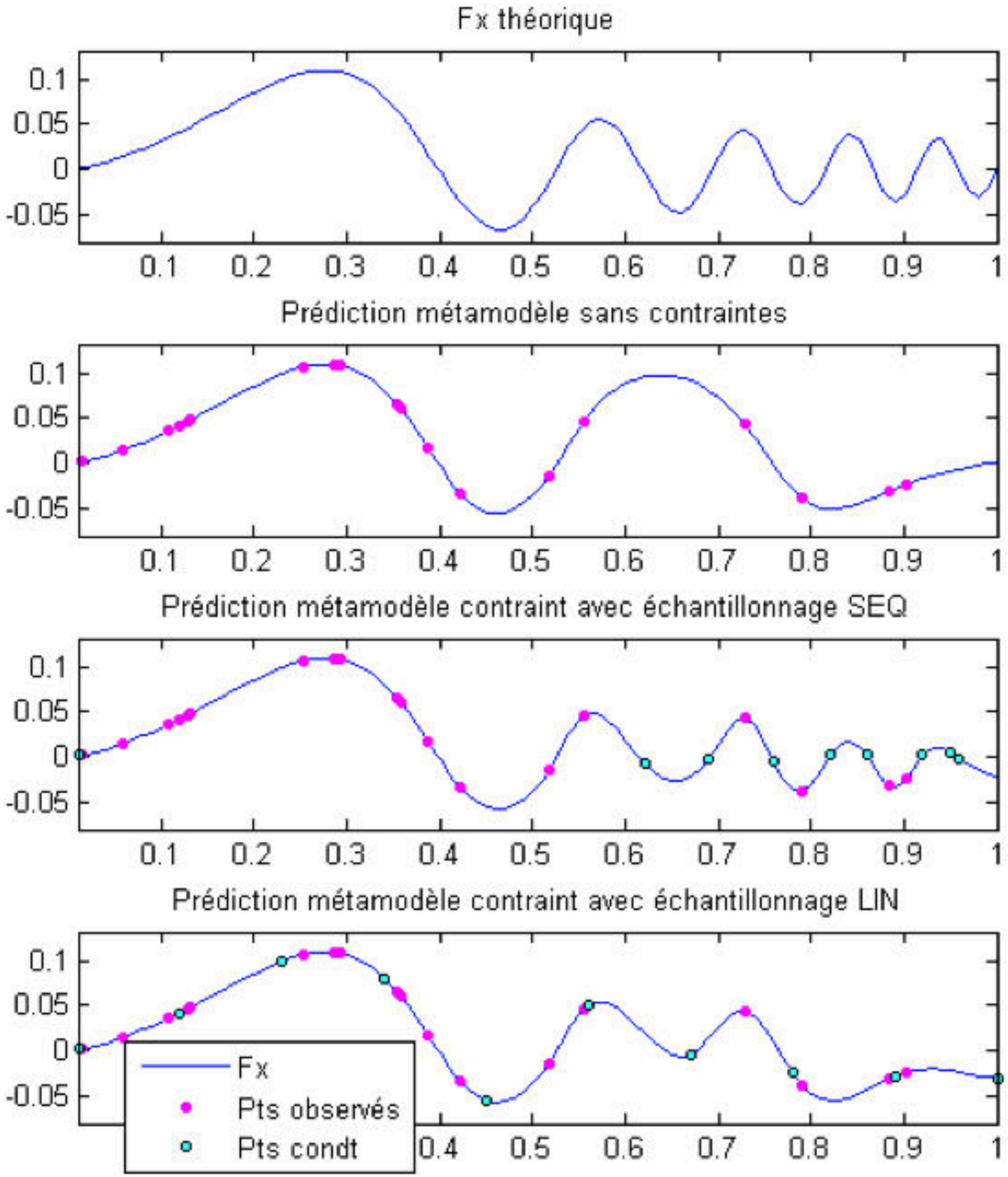


SAFRAN Snecma

EXAMPLES



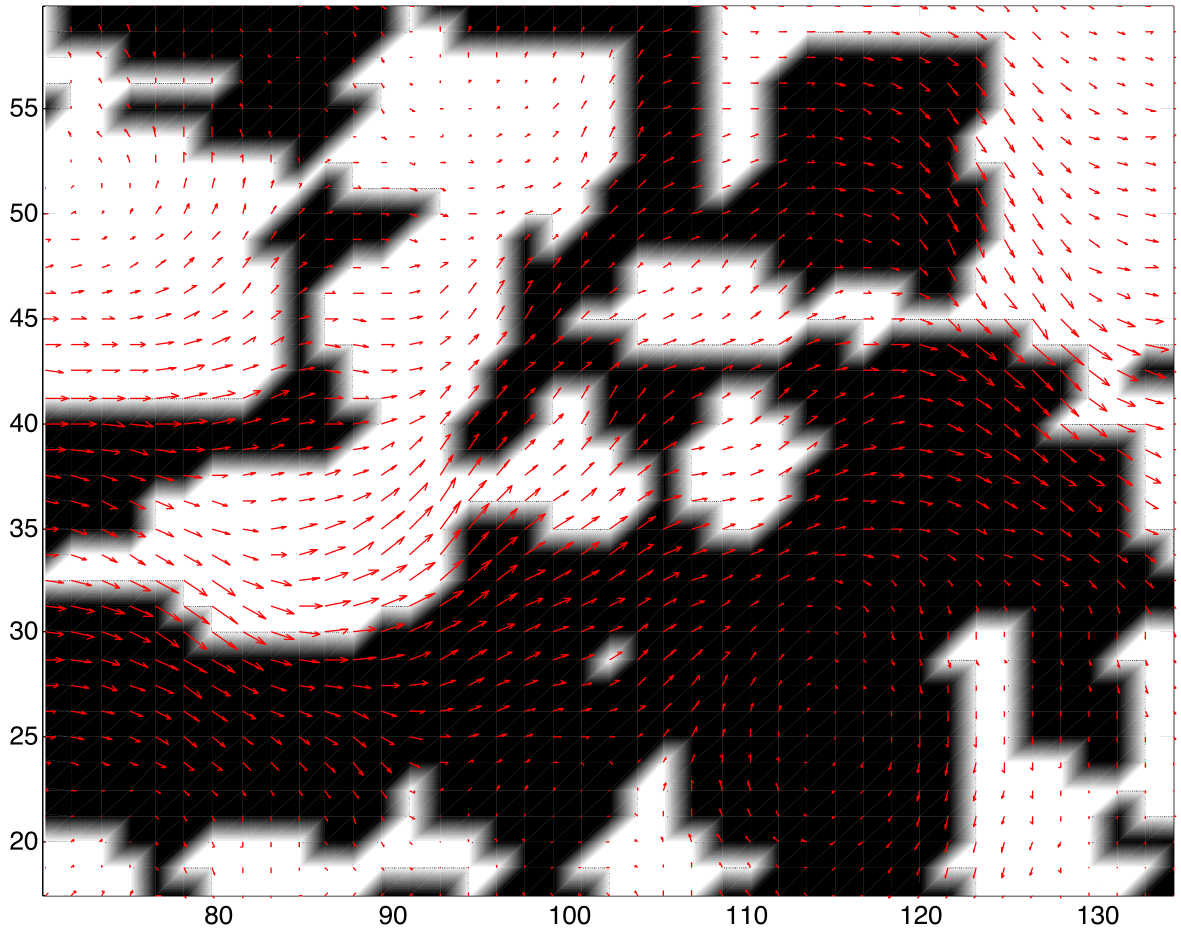
Erreur métamodèle sans contraintes
Erreur métamodèle contraint avec échantillonnage SEQ
Erreur métamodèle contraint avec échantillonnage LIN
Fx
Pts observés
Pts condt



Fx théorique
Prédiction métamodèle sans contraintes
Prédiction métamodèle contraint avec échantillonnage SEQ
Prédiction métamodèle contraint avec échantillonnage LIN
Fx
Pts observés
Pts condt

EXAMPLES

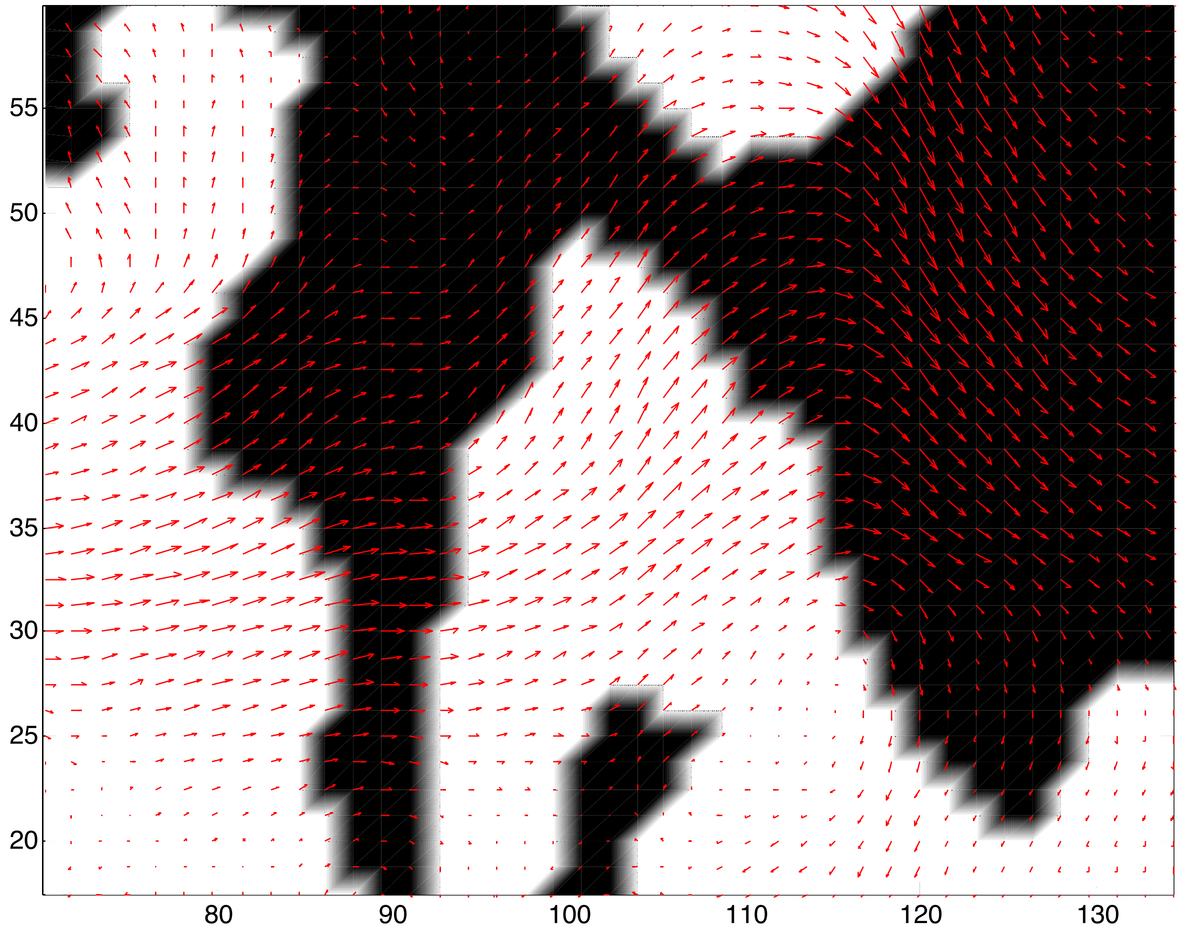
2D-GP predictor with constraints on the sign of the curl



SAFRAN
Snecma

EXAMPLES

2D-GP predictor with constraints on the sign of the divergence



GP REGRESSION WITH INEQUALITY CONSTRAINTS

- **Current tests on 5D challenging function with monotonicity w.r.t. one input variable**
  - Adaptive strategy performs very well

- **Computational trick**
  - Instead of using Genz n times, find the constraint locations with the correlation-free formula (no cost)
  - Once the locations are found, the final prediction is performed with Genz
  - Results seem to indicate that we have almost no lost of prediction accuracy

- **Paper to be submitted soon**

/03/

CONCLUSION & OUTLOOK

SAFRAN
Snecma

INTRODUCTION

- **Theoretical framework to incorporate any linear inequality constraints in GP regression**
  - Truncated normal distribution + approximation formulas for moments

- **From a practical point of view, high-dimensional problems can be accommodated with an adaptive strategy**
  - Even in low-dimensional examples, it is more efficient to choose the constraint locations sequentially
  - The correlation-free trick heavily accelerates the search

- **For challenging applications, advanced computational tools will certainly be necessary**
  - Machine learning methods may be of great help, with adaptation
    - Incomplete Choleshy decomposition (*Bach and Jordan 2002*)
    - Random Kitchen Sink (*Rahimi and Recht 2007, 2008*)

SAFRAN Snecma