MASCOT-NUM annual meeting

Non-stationary Gaussian process modelling and sequential design of experiments for exploration of high variation regions

22nd March 2017 Institut Henri Poincaré

Sébastien Marmin*†‡
David Ginsbourger
Jean Baccou
Jacques Liandrat

based on joined work with
Idiap Research Institute and *University of Bern
†Institut de Radioprotection et de Sûreté Nucléaire
‡École Centrale de Marseille

Gaussian process model for design of experiments

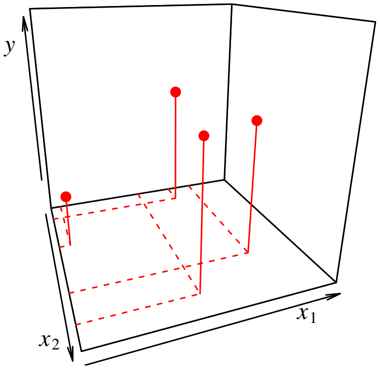
→ Framework

How to approximate $f : D \subset \mathbb{R}^d \to \mathbb{R}$ expensive to evaluate?

Bayesian approach:

- f assumed to be a realization of a Gaussian Process (GP) $Y \sim \mathcal{GP}\left(\mu(\cdot), C(\cdot,\cdot)\right)$.
- n points evaluated $\mathcal{A}_n = (\boldsymbol{y}_{1:n}, X_{1:n})$.
- Model given by the posterior distribution $\mathcal{GP}\left(\mu_n(\cdot), C_n(\cdot,\cdot)\right)$.

Initial design



How to choose the next evaluations $X_{n+1:n+q}$?

Gaussian process model for design of experiments

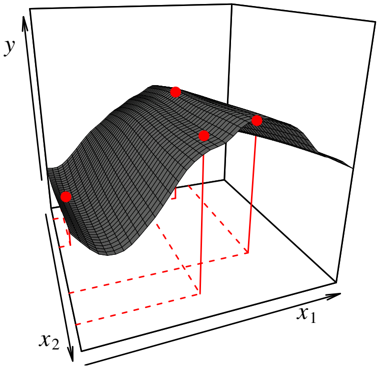
→ Framework

How to approximate $f : D \subset \mathbb{R}^d \to \mathbb{R}$ expensive to evaluate?

Bayesian approach:

- f assumed to be a realization of a Gaussian Process (GP) $Y \sim \mathcal{GP}\left(\mu(\cdot), C(\cdot,\cdot)\right)$.
- n points evaluated $\mathcal{A}_n = (\boldsymbol{y}_{1:n}, X_{1:n})$.
- Model given by the posterior distribution $\mathcal{GP}\left(\mu_n(\cdot), C_n(\cdot,\cdot)\right)$.

Sample path of posterior GP



How to choose the next evaluations $X_{n+1:n+q}$?

Gaussian process model for design of experiments

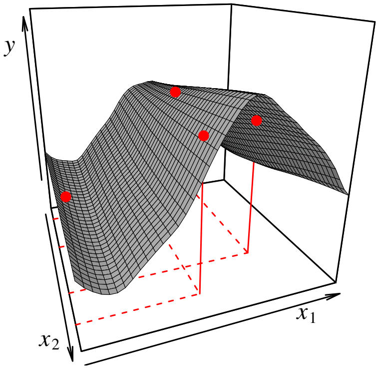
→ Framework

How to approximate $f : D \subset \mathbb{R}^d \to \mathbb{R}$ expensive to evaluate?

Bayesian approach:

- f assumed to be a realization of a Gaussian Process (GP) $Y \sim \mathcal{GP}(\mu(\cdot), C(\cdot,\cdot))$.
- n points evaluated $\mathcal{A}_n = (\boldsymbol{y}_{1:n}, X_{1:n})$.
- Model given by the posterior distribution $\mathcal{GP}(\mu_n(\cdot), C_n(\cdot,\cdot))$.

Sample path of posterior GP



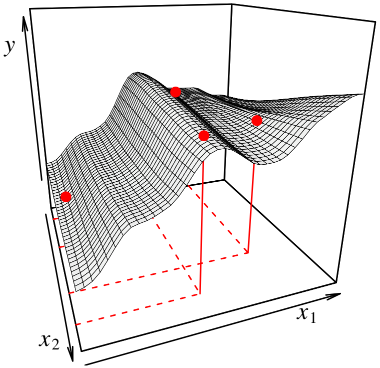
How to choose the next evaluations $X_{n+1:n+q}$?

Gaussian process model for design of experiments

$\rightarrow$ Framework

How to approximate $f : D \subset \mathbb{R}^d \rightarrow \mathbb{R}$ expensive to evaluate?

Sample path of posterior GP



Bayesian approach:

- f assumed to be a realization of a Gaussian Process (GP) $Y \sim \mathcal{GP}(\mu(\cdot), C(\cdot, \cdot))$.
- n points evaluated $\mathcal{A}_n = (\boldsymbol{y}_{1:n}, X_{1:n})$.
- Model given by the posterior distribution $\mathcal{GP}(\mu_n(\cdot), C_n(\cdot, \cdot))$.

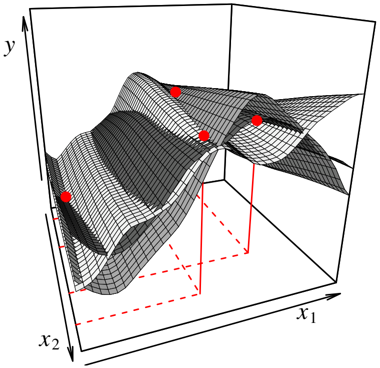
How to choose the next evaluations $X_{n+1:n+q}$?

Gaussian process model for design of experiments

$\rightarrow$ Framework

How to approximate $f : D \subset \mathbb{R}^d \rightarrow \mathbb{R}$ expensive to evaluate?

Sample paths of posterior GP



Bayesian approach:

- f assumed to be a realization of a Gaussian Process (GP) $Y \sim \mathcal{GP}(\mu(\cdot), C(\cdot, \cdot))$.
- n points evaluated $\mathcal{A}_n = (\boldsymbol{y}_{1:n}, X_{1:n})$.
- Model given by the posterior distribution $\mathcal{GP}(\mu_n(\cdot), C_n(\cdot, \cdot))$.

How to choose the next evaluations $X_{n+1:n+q}$?

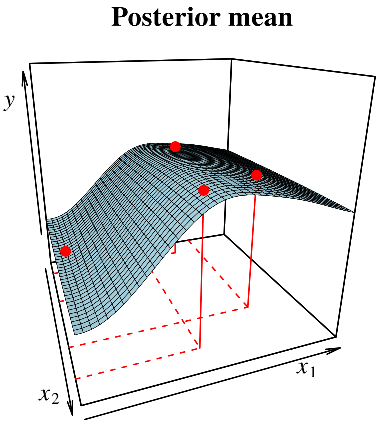
Gaussian process model for design of experiments

→ Framework

How to approximate $f : D \subset \mathbb{R}^d \to \mathbb{R}$ expensive to evaluate?

Bayesian approach:

- f assumed to be a realization of a Gaussian Process (GP) $Y \sim \mathcal{GP}\left(\mu(\cdot), C(\cdot,\cdot)\right)$.
- n points evaluated $\mathcal{A}_n = (\boldsymbol{y}_{1:n}, X_{1:n})$.
- Model given by the posterior distribution $\mathcal{GP}\left(\mu_n(\cdot), C_n(\cdot,\cdot)\right)$.



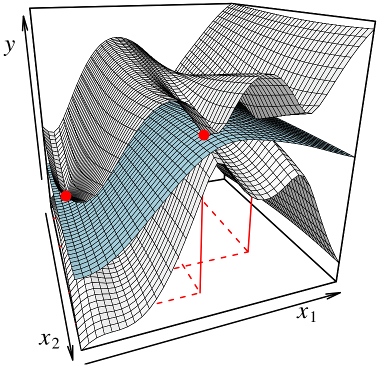
How to choose the next evaluations $X_{n+1:n+q}$?

Gaussian process model for design of experiments

→ Framework

How to approximate $f : D \subset \mathbb{R}^d \to \mathbb{R}$ expensive to evaluate?

Posterior 5%, 50%, 95% quantiles



Bayesian approach:

- f assumed to be a realization of a Gaussian Process (GP) $Y \sim \mathcal{GP}(\mu(\cdot), C(\cdot,\cdot))$.
- n points evaluated $\mathcal{A}_n = (\boldsymbol{y}_{1:n}, X_{1:n})$.
- Model given by the posterior distribution $\mathcal{GP}(\mu_n(\cdot), C_n(\cdot,\cdot))$.

How to choose the next evaluations $X_{n+1:n+q}$?

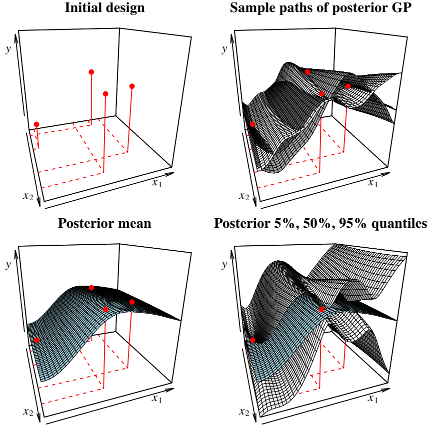
Gaussian process model for design of experiments

→ Framework

How to approximate $f : D \subset \mathbb{R}^d \to \mathbb{R}$ expensive to evaluate?

Bayesian approach:

- f assumed to be a realization of a Gaussian Process (GP) $Y \sim \mathcal{GP}(\mu(\cdot), C(\cdot,\cdot))$.
- n points evaluated $\mathcal{A}_n = (\boldsymbol{y}_{1:n}, X_{1:n})$.
- Model given by the posterior distribution $\mathcal{GP}(\mu_n(\cdot), C_n(\cdot,\cdot))$.



How to choose the next evaluations $X_{n+1:n+q}$?

Gaussian process model for design of experiments

→ Variance-based criteria for sequential design

Next evaluations maximize a criterion function:

$$J_n^{\mathrm{MSE}}(\boldsymbol{x}) = C_n\left(\boldsymbol{x}, \boldsymbol{x}\right),$$

$$J_n^{\mathrm{IMSE}}(\boldsymbol{x}) = -\int_{\boldsymbol{u} \in D} C_{n,\boldsymbol{x}}\left(\boldsymbol{u}, \boldsymbol{u}\right) \mathbf{d}\boldsymbol{u},$$

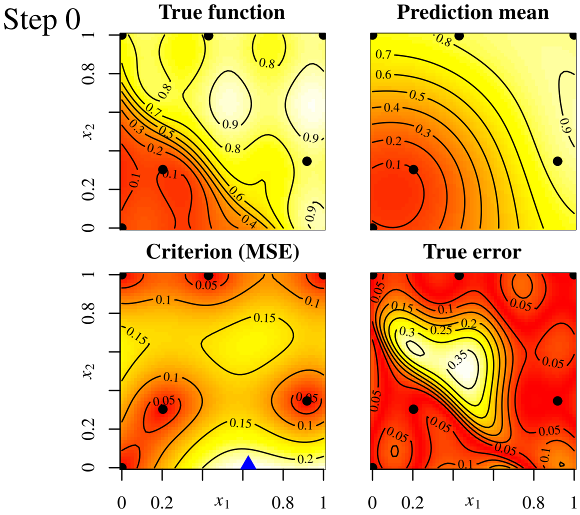
$$J_n^{q\text{-IMSE}}(\underset{\mathbb{R}^{d \times q}}{\underset{\cap}{X}}) = -\int_{\boldsymbol{u} \in D} C_{n,X}\left(\boldsymbol{u}, \boldsymbol{u}\right) \mathbf{d}\boldsymbol{u}.$$

with $C_{n,X}(\boldsymbol{u}, \boldsymbol{v}) = \mathbb{E}\left(\mathrm{cov}\left(Y_{\boldsymbol{u}}, Y_{\boldsymbol{v}} \middle| \boldsymbol{Y}_X\right)\right) = \mathrm{cov}\left(Y_{\boldsymbol{u}}, Y_{\boldsymbol{v}} \middle| \boldsymbol{Y}_X\right).$

→ $\mathrm{cov}\left(Y_{\boldsymbol{u}}, Y_{\boldsymbol{v}} \middle| \boldsymbol{Y}_{\boldsymbol{X}}\right)$ is deterministic : it depends only on X (not on $\boldsymbol{Y}_X$).

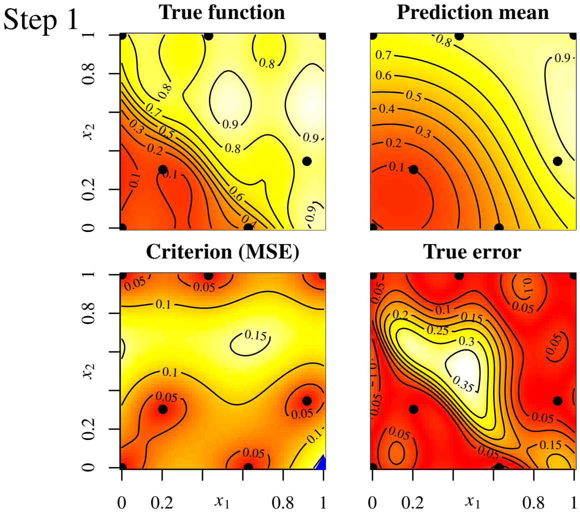
Gaussian process model for design of experiments

→ Sequential design of experiments with the MSE criterion



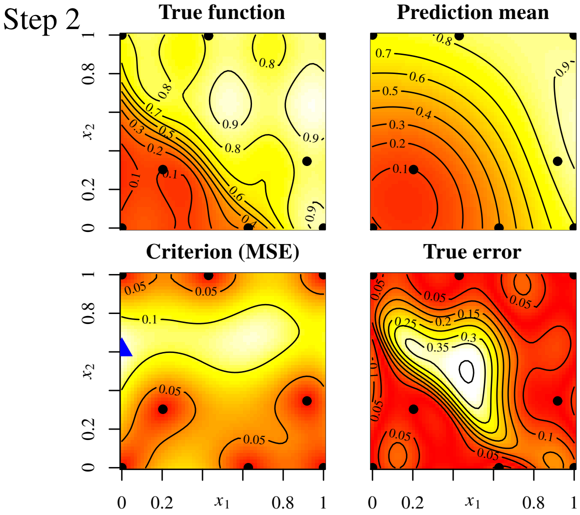
Gaussian process model for design of experiments

→ Sequential design of experiments with the MSE criterion



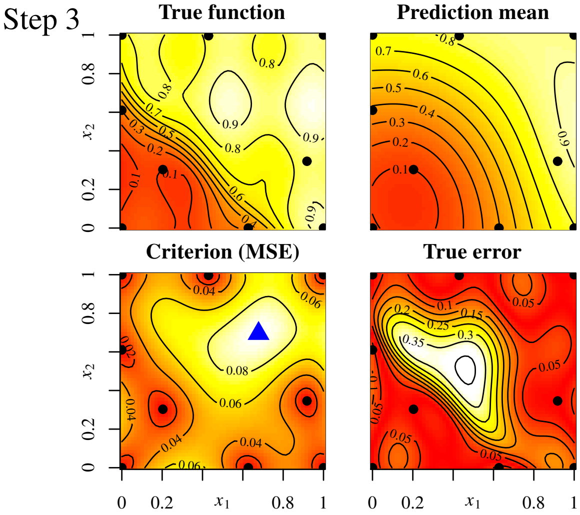
Gaussian process model for design of experiments

→ Sequential design of experiments with the MSE criterion



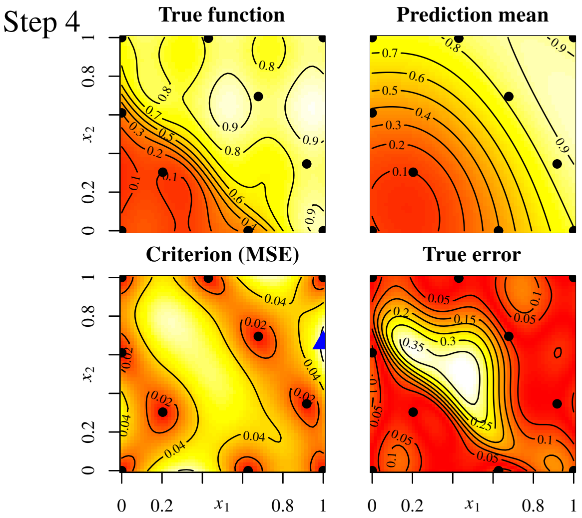
Gaussian process model for design of experiments

→ Sequential design of experiments with the MSE criterion



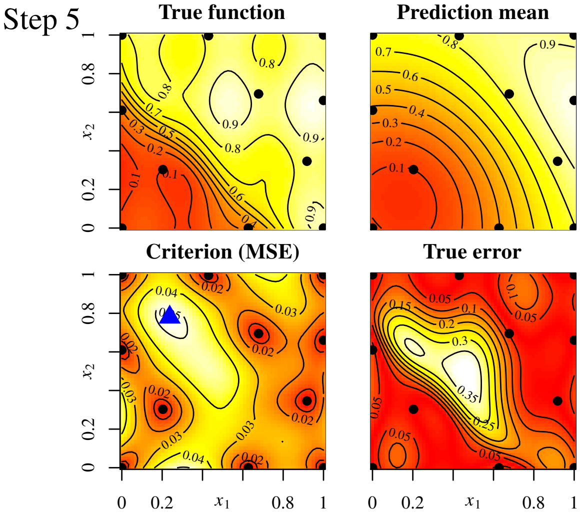
Gaussian process model for design of experiments

→ Sequential design of experiments with the MSE criterion



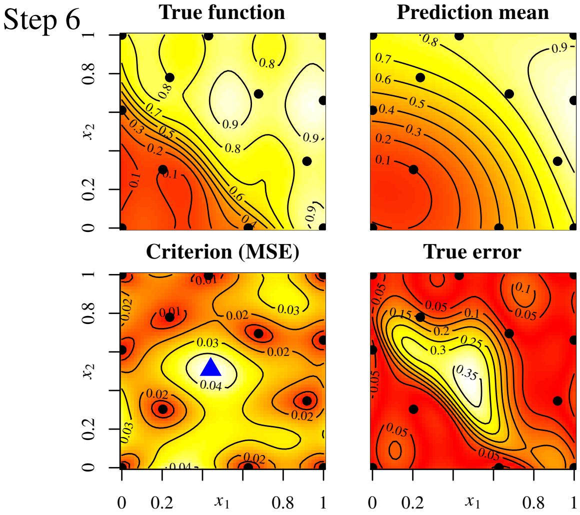
Gaussian process model for design of experiments

→ Sequential design of experiments with the MSE criterion



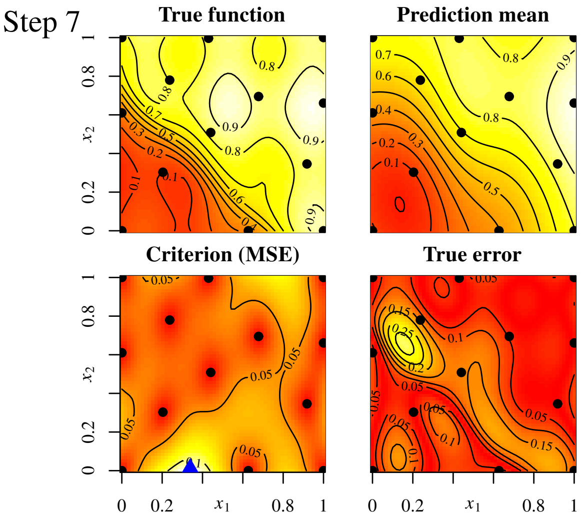
Gaussian process model for design of experiments

→ Sequential design of experiments with the MSE criterion



Gaussian process model for design of experiments

→ Sequential design of experiments with the MSE criterion



Variance-based criteria fill the space but do not focus on a specific region...
→ How to deal with functions exhibiting heterogeneous variations?

- Take a non-stationary model,
- define a criterion that intensifies exploration in high variation regions.

Non-stationary Gaussian process models

Input space warping:

- Non-linear input space warping $T : D \to T(D) \subset \mathbb{R}^d$ creates non-stationary GP [Sampson and Guttorp, 1992].

$$C(\boldsymbol{x}, \boldsymbol{x}') = k(T(\boldsymbol{x}), T(\boldsymbol{x}')).$$

- T can be decomposed with basis functions [Gibbs, 1997].
  $\rightarrow$ Many parameters to estimate.
- T can be simplified as a tensor product $T(\boldsymbol{x}) = (T_i(x_i))_{i=1,\dots,d}$ [Xiong et al., 2007].
  $\rightarrow$ irrelevant when high variations occur along unknown non-canonical directions.

Treed Gaussian Process (TGP):

- different GPs are independently build on several partitions of the input space [Gramacy, 2005].
  $\rightarrow$ not a GP model.

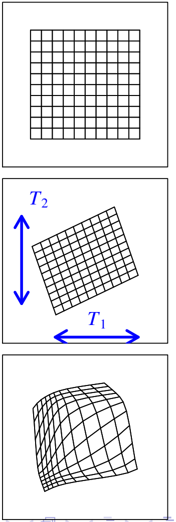
Warped Multiple Index Gaussian Process (WaMI-GP)

Definition

$$C(\boldsymbol{x}, \boldsymbol{x}') = k(T(\boldsymbol{x}), T(\boldsymbol{x}')), \text{ with } T(\boldsymbol{x}) = \left(T_i \left(\boldsymbol{a}_i^\top \boldsymbol{x}; \boldsymbol{\tau}_i \right) \right)_{i=1,\ldots,p}. \tag{1}$$

Composition of warping T

1. Linear transformation of input space,
   $\boldsymbol{x} \to A\boldsymbol{x}$, $A \in \mathbb{R}^{p \times d}$;
   $\to$ reduces dimension ($p \leq d$),
   $\to$ changes canonical axis.
2. Axial warping, $A\boldsymbol{x} \to (T_i((A\boldsymbol{x})_i; \boldsymbol{\tau}_i))_{i=1,\ldots,d}$
   $\to$ produces non-stationarity.



Parameters

$A = (\boldsymbol{a}_i)_{i=1,\ldots,p}$ and $(\boldsymbol{\tau}_i)_{i=1,\ldots,p}$.

Proposed covariance kernel family

→ Example

- Stationary isotropic

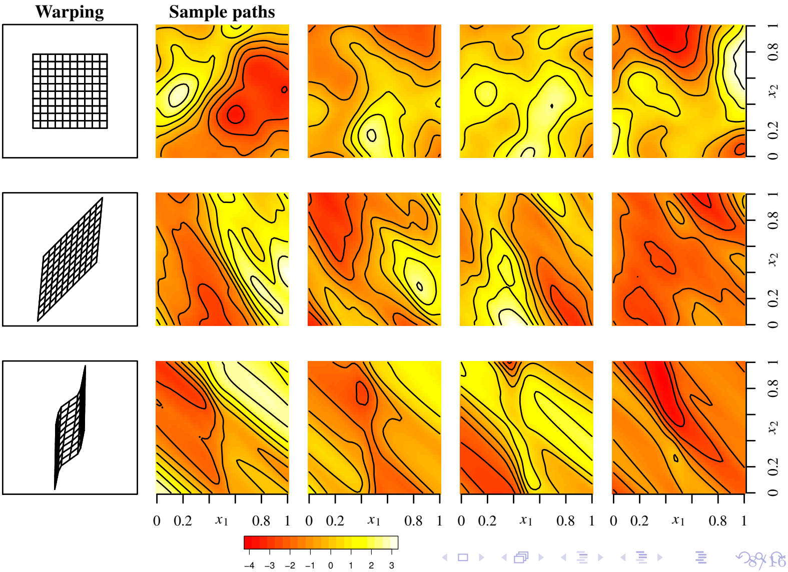
  $k(\boldsymbol{x}, \boldsymbol{x}')$.

- Stationary anisotropic

  $k(A\boldsymbol{x}, A\boldsymbol{x}')$.

- WaMI

  $k(T(\boldsymbol{x}), T(\boldsymbol{x}'))$.

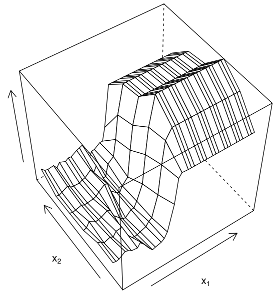


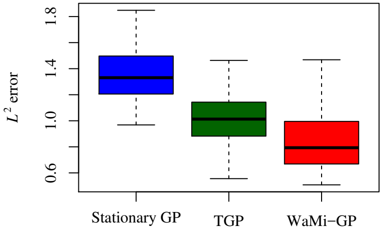
Proposed covariance kernel family

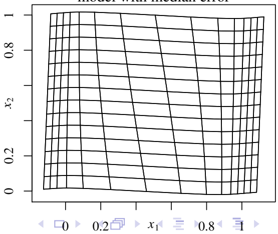
→ Results on a mechanical test case

- test case in dimension $d = 2$,
- $n = 20$ initial evaluations (optimized Latin hypercube design),
- comparison between stationary GP, TGP, WaMI-GP,
- 100 repetitions of the calculation from different maximin LHS designs optimized,
- parameter estimations by maximum likelihood (R package `kergp`).

Response studied (IRSN XPER platform)







Proposed covariance kernel family

→ Properties

- Strict positive definiteness, Appendix 1
- mean-squared differentiability, Appendix 2
- sample path differentiability. Appendix 3

Sampling criteria

→ Gradient random field

- ∇Y is a vector-valued GP with $\forall \boldsymbol{x}, \boldsymbol{x}' \in D$

$$\mathbb{E}\left(\nabla Y(\boldsymbol{x}) \middle| \mathcal{A}_n\right) = \nabla \mu_n(\boldsymbol{x}), \text{ and}$$

$$\operatorname{cov}\left(\frac{\partial}{\partial x_i} Y(\boldsymbol{x}), \frac{\partial}{\partial x_j} Y(\boldsymbol{x}') \middle| \mathcal{A}_n\right) = \left.\frac{\partial^2}{\partial t_i \partial t_j'} C_n(\boldsymbol{t}, \boldsymbol{t}')\right|_{\boldsymbol{t}=\boldsymbol{x}, \boldsymbol{t}'=\boldsymbol{x}'}.$$

→ Proposed criteria

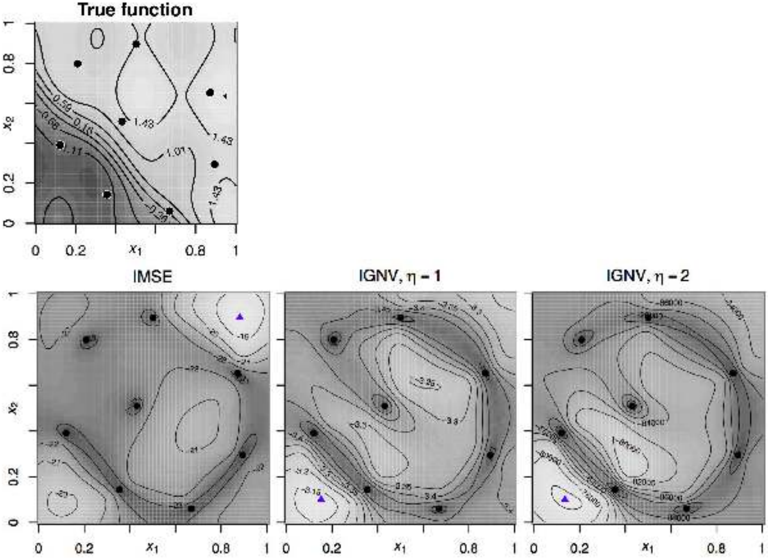
- Gradient Norm (power η) Variance (GNV):

$$J_n^{\mathrm{GNV},\eta}(\boldsymbol{x}) = \operatorname{var}\left(||\nabla Y_{\boldsymbol{x}}||^{\eta} \middle| \mathcal{A}_n\right).$$

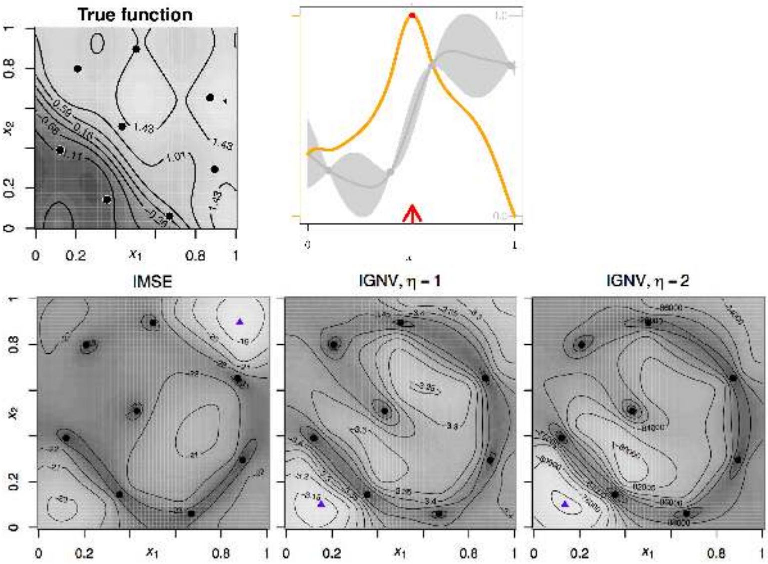
- The corresponding integrated form:

$$J_n^{\mathrm{IGNV},\eta}(\underset{\mathbb{R}^{d\times q}}{\underset{\cap}{X}}) = -\int_{\boldsymbol{u}\in D} \mathbb{E}\left(\operatorname{var}\left(||\nabla Y_{\boldsymbol{u}}||^{\eta} \middle| \mathcal{A}_n, \boldsymbol{Y}_X\right)\right) \mathbf{d}\boldsymbol{u}.$$

Sampling criteria



Sampling criteria



Sampling criteria

→ Calculation of $J_n^{\mathrm{GNV},\eta}(\boldsymbol{x})$ and $\mathbb{E}\left(\mathrm{var}\left(||\nabla Y_{\boldsymbol{u}}||^{\eta} | \mathcal{A}_n, \boldsymbol{Y}_X\right)\right)$:

- $\eta = 1$, semi-analytic (requires an reduced integral quadrature),
- $\eta = 2$, analytic.

Proposition

Let $\boldsymbol{x} \in D$ and denote by $(\lambda_i(\boldsymbol{x}))_{1 \leq i \leq d}$ the eigenvalues of $\nabla^2 c_n(\boldsymbol{x}, \boldsymbol{x})$. Then, the GNV(2) *criterion can be written as follows:*

$$J_n^{\mathrm{GNV},\eta=2}(\boldsymbol{x}) = 4\ \nabla m_n(\boldsymbol{x})^{\top} \nabla^2 c_n(\boldsymbol{x}, \boldsymbol{x}) \nabla m_n(\boldsymbol{x}) + 2 \sum_{i=1}^{d} \lambda_i(\boldsymbol{x})^2,$$

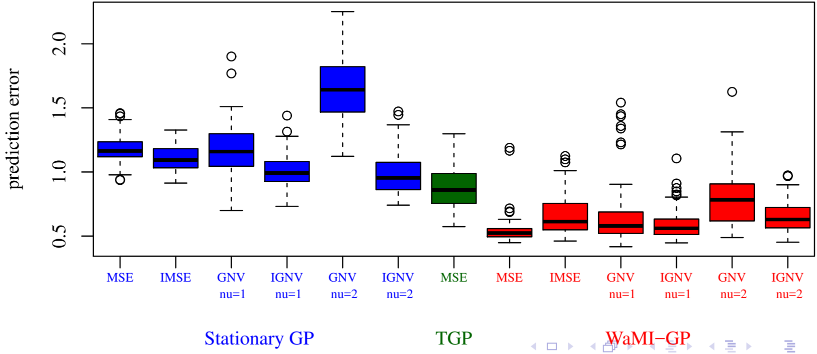
where $\lambda_{i,\boldsymbol{x}}(\boldsymbol{u})$ are the eigenvalues of $\nabla^2 c_{n,\boldsymbol{x}}(\boldsymbol{u}, \boldsymbol{u})$.
Furthermore, the GNV(1) *criterion can be expanded as follows:*

$$J_n^{\mathrm{GNV},\eta=1}(\boldsymbol{x}) = ||\nabla m_n(\boldsymbol{x})||^2 + \mathrm{tr}\left(\nabla^2 c_n(\boldsymbol{x}, \boldsymbol{x})\right) - \mathbb{E}\left(||Y_{\boldsymbol{x}}||\ |\mathcal{A}_n\right)^2.$$

Application

→ IRSN test case

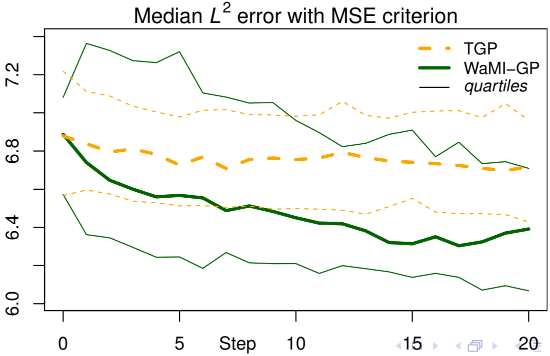
- test case in dimension $d = 2$,
- $n = 20$ initial evaluations (optimized Latin hypercube design),
- comparison between several sequential design strategies under stationary GP, TGP, WaMI-GP,
- 100 repetitions of the calculation from different initial designs,
- parameter estimation by maximum likelihood (R package `kergp`).
- sequential sampling of $r = 10$ points with the different criteria.



Application

→ NASA fuild mechanics test case

- test case in dimension $d = 3$,
- $n = 50$ initial evaluations (optimized Latin hypercube design),
- comparison between several sequential design strategies under stationary GP, TGP, WaMI-GP,
- 50 repetitions of the calculation from different initial designs,
- parameter estimation by maximum likelihood (R package **kergp**).
- sequential sampling of $r = 20$ points with the different criteria.



Conclusion

For sampling functions with heterogeneous variations:

- We propose a new class of non-stationary GP models (WaMI-GP).
- In a generic GP framework (notably with stationary kernels), we define and provide analytical formulas for gradient-based criteria, GNV and IGNV.

Numerical applications:

- On a first test case, WaMI-GP reduces prediction errors compared to stationary GP and TGP. With a stationary model, we observe better performance of the IGNV criterion compared to classical variance-based criteria. Overall best performance is obtained with WaMI-GP combined with MSE.
- On a second test-case in higher dimension, TGP offers better performances at the initial design stage but is outperformed by WaMI-GP along sequential design based on the MSE criterion.

Currently under development:

- Definition and derivation of new criteria,
- open source R implementation,
- parameter estimation procedures for higher dimension cases.

References

Gibbs, M. (1997).
Bayesian Gaussian processes for regression and classification.
PhD thesis, University of Cambridge.

Gramacy, R. B. (2005).
Bayesian treed Gaussian process models.
PhD thesis, University of California Santa Cruz.

Marmin, S., Ginsbourger, D., Baccou, J., Perales, F., and Liandrat, J. (2015).
Processus gaussiens déformés pour l'apprentissage de zones instationnaires.
In *47èmes journées de la Société Franaise de Statistique, Lille, France.*
http://papersjds15.sfds.asso.fr/submission_125.pdf.

Perales, F., Dubois, F., Monerie, Y., Piar, B., and Stainier, L. (2010).
A NonSmooth Contact Dynamics-based Multi-domain Solver. Code coupling (Xper) and application to fracture.
European Journal of Computational Mechanics, 19:389–417.

Roustant, O., Ginsbourger, D., and Deville, Y. (2012).
DiceKriging, DiceOptim: Two R packages for the analysis of computer experiments by Kriging-Based Metamodelling and Optimization.
Journal of Statistical Software, 51 (1):1–55.

Sampson, P. and Guttorp, P. (1992).
Nonparametric Estimation of Nonstationary Spatial Covariance Structure.
Journal of the American Statistical Association, 87(417):108–119.

Xiong, Y., Chen, W., Apley, D., and Ding, X. (2007).
A non-stationary covariance-based kriging method for metamodelling in engineering design.
International Journal for Numerical Methods in Engineering, 71(6):733–756.

Thank you

Appendix

Proposition

Positive definiteness: *Assume that k is strictly positive definite, that the $T_i(\cdot; \boldsymbol{\tau}_i)$ are injective and that the rank of A is equal to d. Then the WaMI kernel of equation (1) is strictly positive definite.*

Appendix

Proposition

Mean-squared differentiability. *The centred Gaussian process with the covariance c defined in (1) is mean-squared differentiable in all canonical direction under the following conditions:*

- $T_i(\cdot;\boldsymbol{\tau}_i)$, $i = 1, \dots, q$, *have regularity* C^1 *on* $\mathbb{R}$. *We write* $T_i'(\cdot;\boldsymbol{\tau}_i)$ *the univariate and continuous derivatives.*
- $\forall j, j' \in \{1, \dots, q\}, \forall \boldsymbol{u} \in \mathbb{R}^q$, $\left.\frac{\partial^2 k(\boldsymbol{v},\boldsymbol{v}')}{\partial v_j \partial v'_{j'}}\right|_{(\boldsymbol{u},\boldsymbol{u})}$ *exists and is finite.*

Appendix

Proposition

Sample path differentiability. *With the same assumptions on* $T_i(\cdot;\boldsymbol{\tau}_i)$*'s and* k *as in the previous proposition, and assuming in addition that*

- D *is compact,*
- *there exist* $C_0, \eta_0, \varepsilon_0 > 0$ *such that* $\forall j, j' \in \{1, \ldots, q\}$*, and* $\forall \boldsymbol{u}, \boldsymbol{u}' \in \mathbb{R}^q$, $||\boldsymbol{u} - \boldsymbol{u}'|| < \varepsilon_0$*, we have*
  $$\left.\frac{\partial^2 k(\boldsymbol{v},\boldsymbol{v}')}{\partial v_j \partial v'_{j'}}\right|_{(\boldsymbol{u},\boldsymbol{u})} + \left.\frac{\partial^2 k(\boldsymbol{v},\boldsymbol{v}')}{\partial v_j \partial v'_{j'}}\right|_{(\boldsymbol{u}',\boldsymbol{u}')} - 2\left.\frac{\partial^2 k(\boldsymbol{v},\boldsymbol{v}')}{\partial v_j \partial v'_{j'}}\right|_{(\boldsymbol{u},\boldsymbol{u}')} \leq \frac{C_0}{|\ln ||\boldsymbol{u}-\boldsymbol{u}'|||^{1+\eta_0}},$$

then the covariance c *gives rise to a centred Gaussian Process possessing a version with differentiable sample paths.*