Constrained Gaussian process modeling

S. Da Veiga* and A. Marrel (CEA)

IFP Energies nouvelles

Outline

- Standard Gaussian process modeling
  - Quick review
  - Towards incorporation of constraints
- Linear constraints for Gaussian processes
  - Theoretical framework
  - Numerical approximations
- Examples on analytical functions
- Conclusions & outlook

Outline

- Standard Gaussian process modeling
  - Quick review
  - Towards incorporation of constraints
- Linear constraints for Gaussian processes
  - Theoretical framework
  - Numerical approximations
- Examples on analytical functions
- Conclusions & outlook

Standard Gaussian process modeling

- **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 $(\mathbf{x}_i, y_i)_{i=1,\ldots,n}$

$$X_s = \left[\mathbf{x}_1^T, \ldots, \mathbf{x}_n^T\right]^T \quad Y_s = [y_1, \ldots, y_n]^T$$

- **Model**
  - Output seen as realization of stationary Gaussian process

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

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

Standard Gaussian process modeling

- Conditioning:

$$\tilde{Y}(\mathbf{x}^*) = Y(\mathbf{x}^*) | Y(X_s) = Y_s$$

  - 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

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

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

Standard Gaussian process modeling

- Why accounting for constraints ?
  - Physics / expected behavior are respected
  - Improved predictions
  - Robustness

Standard Gaussian process modeling

- Problem already studied in nonparametric regression
  - 1D setting
    - Ramsay 2005, Dette and Scheder 2006, Bigot and Gadat 2010
  - Kernel regression
    - Constraints on weights: Hall and Huang 2001, Racine et al. 2009
  - Kriging
    - Data-augmentation: Abrahamsen and Benth 2001
    - Weights: Yoo and Kyriadis 2006
    - Sampling: Michalak 2008, Kleijnen and van Beers 2010

Standard Gaussian process modeling

- Here, we propose to keep the conditional expectation framework

  - Example for bound constraints

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

  - Links 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 for joint distributions in the general case

Standard Gaussian process modeling

- Proposal: discrete-location approximation

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

⇓

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

Standard Gaussian process modeling

- Examples of constrained predictors

  - Bounds $\mathbb{E}\left(\tilde{Y}(\mathbf{x}^*)|\forall i=1,\dots,N,\ a_i \leq \tilde{Y}(\mathbf{x}_i) \leq b_i\right)$

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

  - Convexity $\mathbb{E}\left(\tilde{Y}(x^*)|\forall i=1,\dots,N,\ \frac{\partial^2 \tilde{Y}}{\partial x^2}(x_i) \geq 0\right)$

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

Standard Gaussian process modeling

- Roughly speaking …
  - 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, monotony, …
    - Compute the average to get the new kriging predictor
    - (If desired, the variance yields a measure of accuracy)

Outline

- Standard Gaussian process modeling
  - Quick review
  - Towards incorporation of constraints
- Linear constraints for Gaussian processes
  - Theoretical framework
  - Numerical approximations
- Examples on analytical functions
- Conclusions & outlook

Linear constraints for Gaussian processes

- Relevant constraints …
  - Bounds $\mathbb{E}\left(\tilde{Y}(\mathbf{x}^*) | \forall i = 1, \dots, N,\ a_i \leq \tilde{Y}(\mathbf{x}_i) \leq b_i\right)$
  - Monotony $\mathbb{E}\left(\tilde{Y}(\mathbf{x}^*) | \forall i = 1, \dots, N,\ \frac{\partial \tilde{Y}}{\partial x^j}(\mathbf{x}_i) \geq 0\right)$
  - Convexity $\mathbb{E}\left(\tilde{Y}(x^*) | \forall i = 1, \dots, N,\ \frac{\partial^2 \tilde{Y}}{\partial x^2}(x_i) \geq 0\right)$
  - Conservation $\mathbb{E}\left(\tilde{Y}(\mathbf{x}^*) | \sum_{i=1}^{N} w_i \tilde{Y}(\mathbf{x}_i) \leq M\right)$
- … are all linear constraints for a multivariate normal random vector

Linear constraints for Gaussian processes

- Common framework: truncated normal distribution
  - Given a multivariate normal vector …

$$\mathbf{Z} = (Z_1, \dots, Z_p)$$

$$\phi_{\mu,\Sigma}(\mathbf{z}) = \frac{1}{(2\pi)^{p/2}|\Sigma|^{1/2}} \exp\left(-\frac{1}{2}(\mathbf{z}-\mu)^T\Sigma^{-1}(\mathbf{z}-\mu)\right)$$

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

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

Linear constraints for Gaussian processes

- Theoretical results on truncated normal distribution

  - Extensively studied by Tallis

  - Moment generating function

$$m(\mathbf{t}) = \frac{e^T}{\alpha}\Phi(\mathbf{a}-\xi, \mathbf{b}-\xi; \Sigma)$$

$$T = \frac{1}{2}\mathbf{t}^T\Sigma\mathbf{t} \quad \xi = \Sigma\mathbf{t}$$

$$\Phi(\mathbf{u}, \mathbf{v}; \Sigma) = \int_{u_1}^{v_1} \dots \int_{u_p}^{v_p} \phi_{0,\Sigma,\mathbf{a},\mathbf{b}}(\mathbf{z})d\mathbf{z}$$

  - Expectation

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

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

Linear constraints for Gaussian processes

- Numerical approximations
  - Correlation-free formula

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

  - Genz approximation
    - Numerical integration (up to dimension 1000)
  - Sampling point of view
    - Gibbs' sampler (Robert 1995)
    - Fast univariate sampling algorithm

Outline

- Standard Gaussian process modeling
  - Quick review
  - Towards incorporation of constraints
- Linear constraints for Gaussian processes
  - Theoretical framework
  - Numerical approximations
- Examples on analytical functions
- Conclusions & outlook

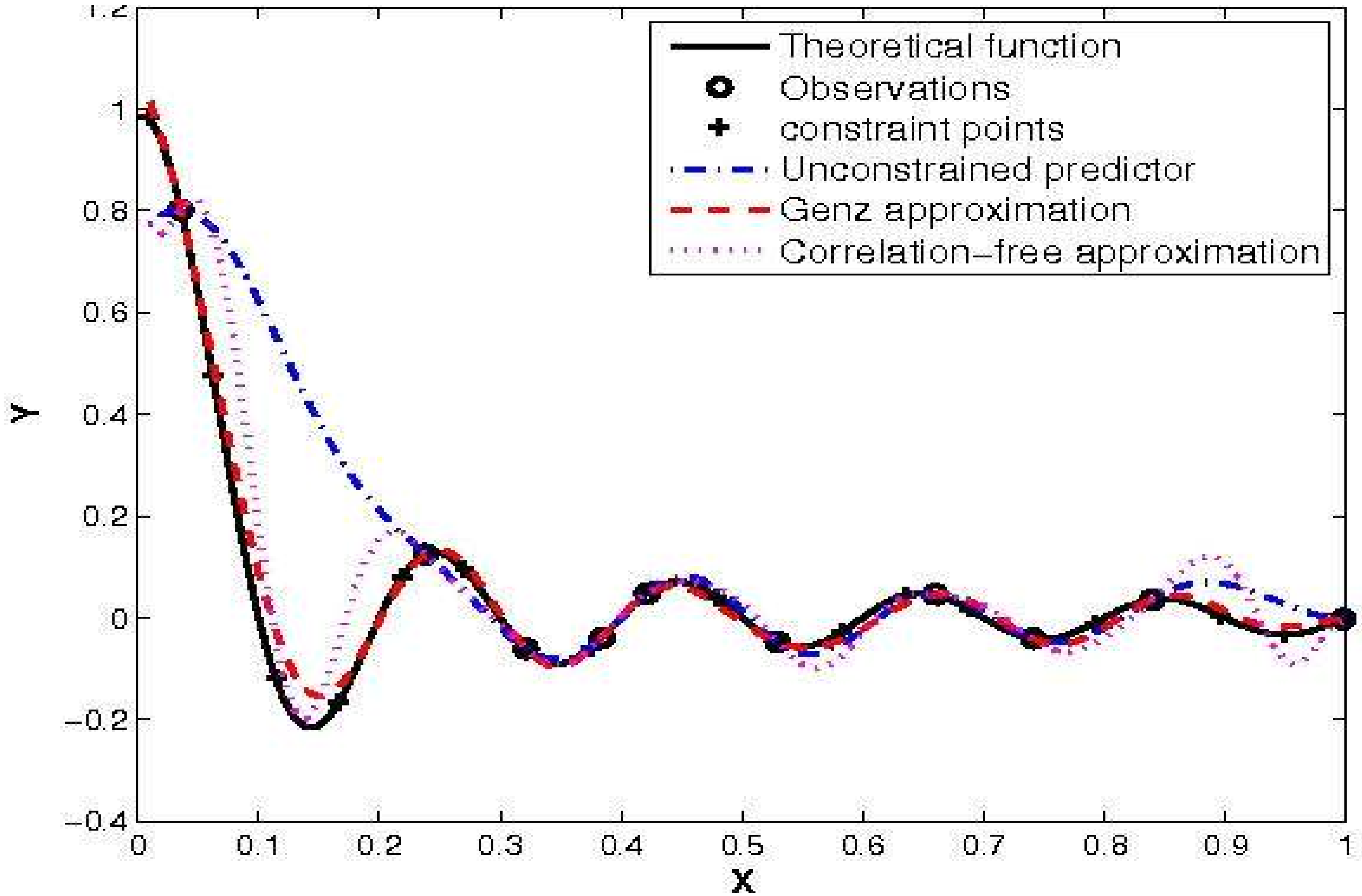
Examples on analytical functions

- 1D stationary function

$$f(x) = \frac{\sin(10\pi x)}{10\pi x}$$

  - n = 10 observations (at random)
  - Gaussian correlation function
  - Bound constraints: positive / negative
  - N = 20 constraint locations (equally spaced)

Examples on analytical functions



Q2 = 0.021 (uncons)

Q2 = 0.83 (correlation-free)

Q2 = 0.98 (Genz / sampling)

Examples on analytical functions

Gaussian covariance, bound constraints

n	N	Unconstrained	Genz	Sampling	Correlation-free
n = 10	N = 20	0.30 +/- 0.46	0.53 +/- 0.36	0.47 +/- 0.39	0.42 +/- 0.29
n = 10	N = 30	0.24 +/- 0.43	0.41 +/- 0.39	0.39 +/- 0.43	0.41 +/- 0.30
n = 15	N = 20	0.58 +/- 0.53	0.72 +/- 0.36	0.72 +/- 0.36	0.69 +/- 0.34
n = 15	N = 30	0.62 +/- 0.56	0.80 +/- 0.34	0.77 +/- 0.37	0.74 +/- 0.33
n = 20	N = 20	0.93 +/- 0.28	0.96 +/- 0.12	0.96 +/- 0.12	0.91 +/- 0.28
n = 20	N = 30	0.91 +/- 0.34	0.95 +/- 0.14	0.95 +/- 0.15	0.93 +/- 0.18
n = 25	N = 20	0.99 +/- 0.03	0.99 +/- 0.03	0.99 +/- 0.03	0.99 +/- 0.03
n = 25	N = 30	0.96 +/- 0.29	0.99 +/- 0.05	0.99 +/- 0.05	0.98 +/- 0.10

Matérn 3/2 covariance, bound constraints

n	N	Unconstrained	Genz	Sampling	Correlation-free
n = 10	N = 20	0.46 +/- 0.61	0.60 +/- 0.41	0.46 +/- 0.58	0.45 +/- 0.33
n = 10	N = 30	0.29 +/- 0.46	0.52 +/- 0.33	0.32 +/- 0.39	0.47 +/- 0.28
n = 15	N = 20	0.60 +/- 0.40	0.72 +/- 0.32	0.70 +/- 0.34	0.68 +/- 0.31
n = 15	N = 30	0.55 +/- 0.45	0.74 +/- 0.34	0.65 +/- 0.42	0.69 +/- 0.31
n = 20	N = 20	0.73 +/- 0.39	0.83 +/- 0.29	0.80 +/- 0.33	0.81 +/- 0.30
n = 20	N = 30	0.79 +/- 0.33	0.85 +/- 0.25	0.84 +/- 0.26	0.83 +/- 0.25
n = 25	N = 20	0.82 +/- 0.38	0.87 +/- 0.27	0.87 +/- 0.27	0.86 +/- 0.28
n = 25	N = 30	0.89 +/- 0.28	0.92 +/- 0.20	0.92 +/- 0.20	0.91 +/- 0.21

ANR Cos

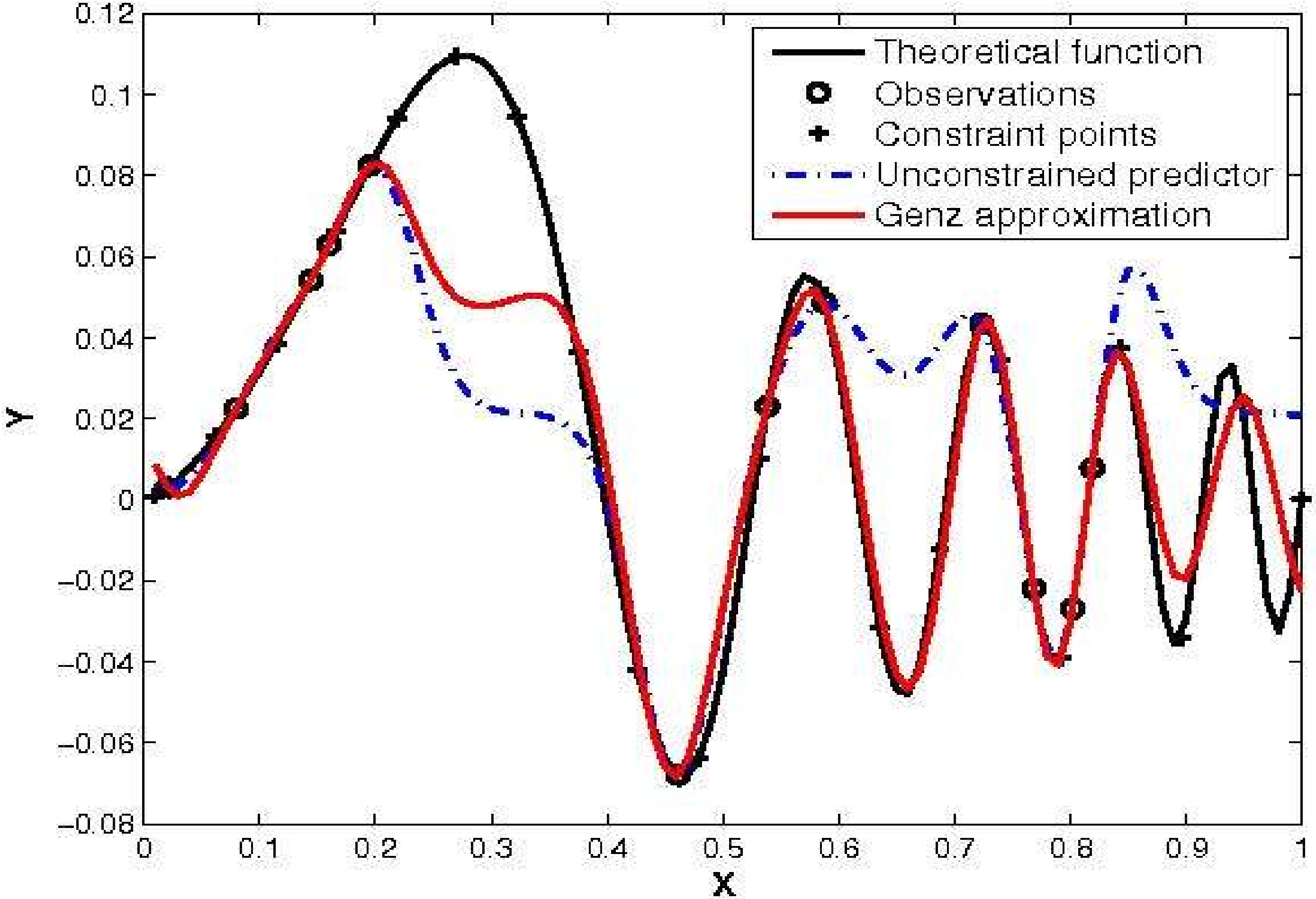
Examples on analytical functions

- 1D non-stationary function

$$f(x) = \frac{\sin(10\pi x^{5/2})}{10\pi x}$$

  - n = 15 observations (at random)
  - Gaussian correlation function
  - Bound constraints: positive / negative
  - N = 20 constraint locations (equally spaced)

Examples on analytical functions



Q2 = 0.43
(uncons)

Q2 = 0.84
(Genz / sampling)

Examples on analytical functions

Matérn 3/2 covariance, bound constraints

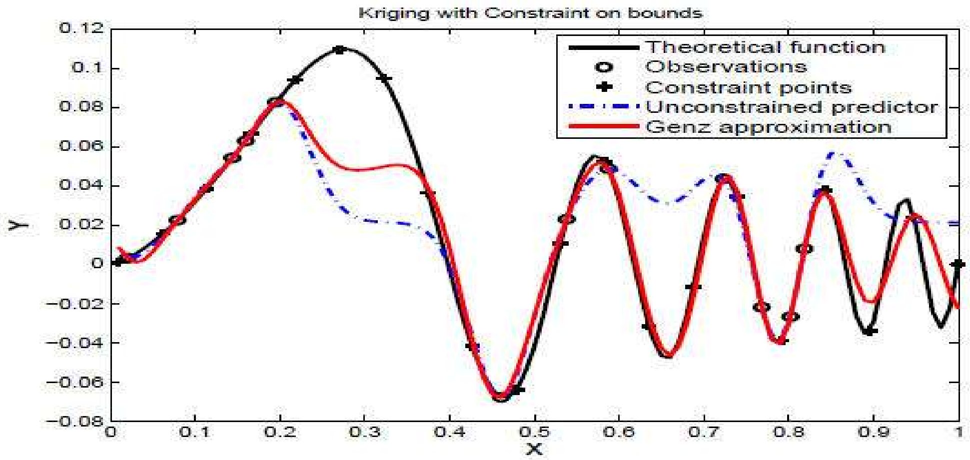
n	N	Unconstrained	Genz	Sampling	Correlation-free
n = 10	N = 20	0.43 +/- 0.29	0.80 +/- 0.11	0.74 +/- 0.19	0.78 +/- 0.12
n = 10	N = 30	0.44 +/- 0.27	0.83 +/- 0.10	0.76 +/- 0.24	0.79 +/- 0.12
n = 10	N = 50	0.47 +/- 0.22	0.83 +/- 0.13	0.75 +/- 0.21	0.79 +/- 0.10
n = 10	N = 100	0.47 +/- 0.24	0.77 +/- 0.17	0.74 +/- 0.20	0.80 +/- 0.10
n = 20	N = 20	0.74 +/- 0.15	0.91 +/- 0.06	0.89 +/- 0.07	0.90 +/- 0.06
n = 20	N = 30	0.77 +/- 0.12	0.93 +/- 0.05	0.91 +/- 0.08	0.91 +/- 0.06
n = 20	N = 50	0.75 +/- 0.20	0.94 +/- 0.05	0.89 +/- 0.19	0.91 +/- 0.07
n = 20	N = 100	0.75 +/- 0.14	0.93 +/- 0.07	0.90 +/- 0.12	0.92 +/- 0.06
n = 30	N = 20	0.87 +/- 0.10	0.96 +/- 0.02	0.95 +/- 0.03	0.95 +/- 0.03
n = 30	N = 30	0.89 +/- 0.09	0.96 +/- 0.02	0.95 +/- 0.07	0.96 +/- 0.04
n = 30	N = 50	0.88 +/- 0.09	0.97 +/- 0.02	0.97 +/- 0.02	0.96 +/- 0.02
n = 30	N = 100	0.87 +/- 0.16	0.96 +/- 0.14	0.94 +/- 0.15	0.95 +/- 0.04
n = 40	N = 20	0.94 +/- 0.05	0.97 +/- 0.01	0.97 +/- 0.02	0.97 +/- 0.02
n = 40	N = 30	0.94 +/- 0.06	0.98 +/- 0.01	0.98 +/- 0.01	0.98 +/- 0.01
n = 40	N = 50	0.95 +/- 0.06	0.98 +/- 0.01	0.97 +/- 0.04	0.98 +/- 0.02
n = 40	N = 100	0.94 +/- 0.06	0.99 +/- 0.01	0.98 +/- 0.01	0.98 +/- 0.01
n = 50	N = 20	0.97 +/- 0.04	0.98 +/- 0.01	0.98 +/- 0.01	0.98 +/- 0.01
n = 50	N = 30	0.97 +/- 0.04	0.98 +/- 0.01	0.98 +/- 0.01	0.98 +/- 0.01
n = 50	N = 50	0.97 +/- 0.05	0.99 +/- 0.01	0.99 +/- 0.01	0.98 +/- 0.01
n = 50	N = 100	0.97 +/- 0.05	0.99 +/- 0.01	0.99 +/- 0.02	0.99 +/- 0.01

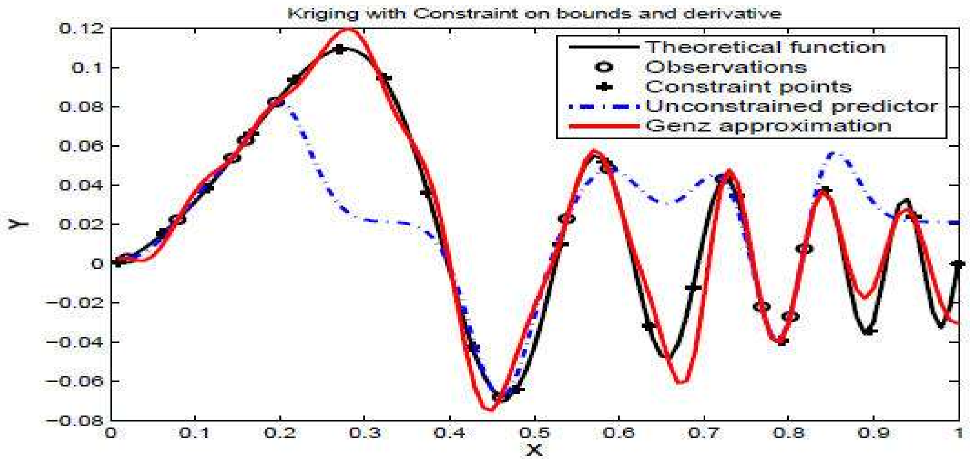
Examples on analytical functions

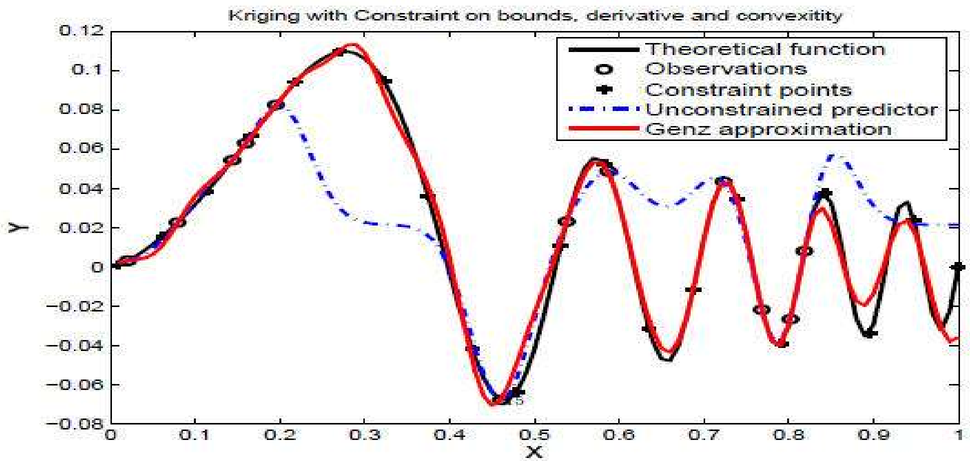
- 1D non-stationary function

$$f(x) = \frac{\sin(10\pi x^{5/2})}{10\pi x}$$

  - n = 15 observations (at random)
  - Gaussian correlation function
  - Several constraints
    - Bounds (positive / negative)
    - 1^{st} order derivative (increasing / decreasing)
    - 2^{nd} order derivative (convex / concave)
  - N = 20 constraint locations (equally spaced)







Q2 = 0.43
(uncons)

Q2 = 0.84
(bound constraints)

Q2 = 0.95
(bounds + deriv.)

Q2 = 0.98
(bounds + deriv. + convexity)

Examples on analytical functions

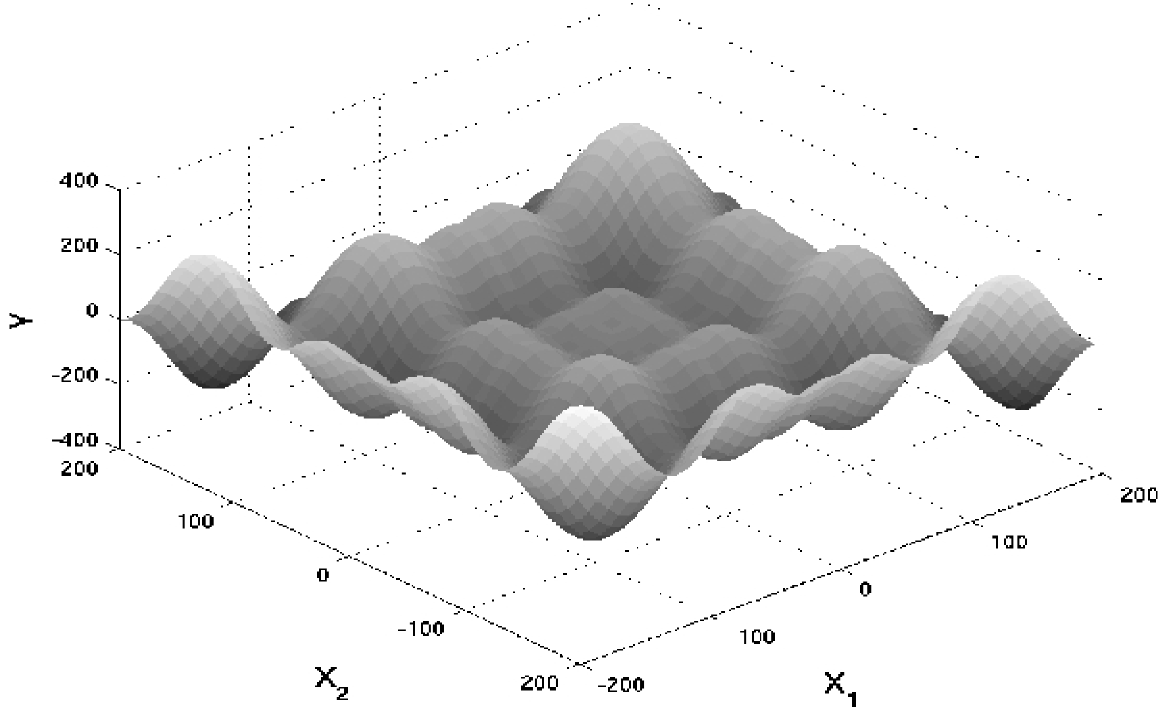
- Successive incorporation of constraints (n=15, N=20)

Constraints	Gaussian	Matérn
No	0.62 +/- 0.20	0.63 +/- 0.19
Bounds only	0.79 +/- 0.19	0.88 +/- 0.06
Derivatives only	0.80 +/- 0.19	0.77 +/- 0.11
Bounds and derivatives	0.80 +/- 0.23	0.91 +/- 0.06
Bounds and derivatives and convexity	0.85 +/0.19	×

Examples on analytical functions

- 2D Schwefel's function

$$f(x^1, x^2) = -x^1 \sin(\sqrt{|x^1|}) - x^2 \sin(\sqrt{|x^2|})$$



Examples on analytical functions

		Gaussian covariance, bounds			Matérn 3/2 covariance, bounds		
n	N	Unconstrained	Genz	Correlation-free	Unconstrained	Genz	Correlation-free
n = 50	N = 100	0.29 +/- 0.08	0.57 +/- 0.07	0.56 +/- 0.06	0.32 +/- 0.09	0.60 +/- 0.10	0.60 +/- 0.07
n = 50	N = 225	0.31 +/- 0.09	0.58 +/- 0.14	0.61 +/- 0.19	0.32 +/- 0.09	0.69 +/- 0.10	0.69 +/- 0.06
n = 50	N = 400	0.29 +/- 0.09	0.41 +/- 0.16	0.60 +/- 0.21	0.32 +/- 0.10	0.63 +/- 0.11	0.68 +/- 0.04
n = 100	N = 100	0.52 +/- 0.07	0.68 +/- 0.04	0.68 +/- 0.04	0.64 +/- 0.07	0.78 +/- 0.03	0.77 +/- 0.03
n = 100	N = 225	0.52 +/- 0.08	0.69 +/- 0.09	0.70 +/- 0.11	0.65 +/- 0.07	0.81 +/- 0.05	0.79 +/- 0.05
n = 100	N = 400	0.52 +/- 0.07	0.61 +/- 0.10	0.64 +/- 0.44	0.64 +/- 0.08	0.78 +/- 0.05	0.80 +/- 0.03
n = 150	N = 100	0.66 +/- 0.06	0.76 +/- 0.04	0.76 +/- 0.04	0.82 +/- 0.05	0.88 +/- 0.03	0.87 +/- 0.03
n = 150	N = 225	0.67 +/- 0.05	0.74 +/- 0.07	0.76 +/- 0.06	0.83 +/- 0.06	0.88 +/- 0.03	0.87 +/- 0.03
n = 150	N = 400	0.66 +/- 0.05	0.69 +/- 0.06	0.77 +/- 0.05	0.83 +/- 0.04	0.87 +/- 0.03	0.88 +/- 0.02
n = 200	N = 100	0.75 +/- 0.04	0.81 +/- 0.03	0.81 +/- 0.03	0.92 +/- 0.03	0.94 +/- 0.02	0.94 +/- 0.03
n = 200	N = 225	0.75 +/- 0.04	0.80 +/- 0.05	0.81 +/- 0.05	0.92 +/- 0.03	0.93 +/- 0.02	0.93 +/- 0.02
n = 200	N = 400	0.74 +/- 0.05	0.76 +/- 0.05	0.81 +/- 0.05	0.92 +/- 0.s03	0.93 +/- 0.03	0.94 +/- 0.02
n = 300	N = 100	0.84 +/- 0.05	0.87 +/- 0.04	0.87 +/- 0.04	0.98 +/- 0.02	0.99 +/- 0.01	0.99 +/- 0.01
n = 300	N = 225	0.84 +/- 0.05	0.85 +/- 0.03	0.86 +/- 0.03	0.98 +/- 0.02	0.98 +/- 0.02	0.98 +/- 0.02
n = 300	N = 400	0.84 +/- 0.04	0.85 +/- 0.04	0.88 +/- 0.04	0.98 +/- 0.02	0.98 +/- 0.02	0.99 +/- 0.01

Conclusions

- New theoretical framework for incorporating constraints in Gaussian process modeling
  - Three dedicated numerical approximations

- Results on analytical functions are promising
  - Improved predictivity when the number of simulations is small
  - More robust
  - Predictions are consistent with respect to physics

- Paper submitted to special issue of "Annales de la faculté des sciences de Toulouse"

Outlook

- Constrained estimation of hyperparameters
- Theoretical and numerical study on constraint locations
  - Current result to be refined: very few uniform locations seem to be sufficient
- Convergence proof to justify discrete-location approximation

References

Abrahamsen P. and Benth F.E. (2001). Kriging with inequality constraints. Mathematical Geology, 33(6):719–744.

Azaïs J.-M. and Wschebor M. (2009). Level sets and extrema of random processes and fields. New York: Wiley.

Bigot J. and Gadat S. (2010). Smoothing under diffeomorphic constraints with homeomorphic splines. SIAM Journal on Numerical Analysis, 48(1):224–243.

Dette, H. and Scheder, R. (2006). Strictly monotone and smooth nonparametric regression for two or more variables. The Canadian Journal of Statistics, 34(44):535–561.

Fernandez P.J., Ferrari P.A. and Grynberg S.P. (2007). Perfectly random sampling of truncated multinormal distributions. Adv. in Appl. Probab., 39(4):973–990.

Genz A. and Bretz F. (2009). Computation of Multivariate Normal and t Probabilities. Lecture Notes in Statistics, Vol. 195, Springer-Verlag, Heidelberg.

Griffiths W. (2002). A Gibbs' sampler for the parameters of a truncated multivariate normal distribution. Working Paper, http://ideas.repec.org/p/mlb/wpaper/856.html.

Hall P. and Huang L.-S. (2001). Nonparametric kernel regression subject to monotonicity constraints. The Annals of Statistics, 29(3):624–647.

Hazelton M.L. and Turlach B.A. (2011). Semiparametric regression with shape-constrained penalized splines. Computational Statistics and Data Analysis, 55:2871–2879.

Kleijnen J.P.C. and van Beers, W.C.M. (2010). Monotonicity-preserving bootstrapped Krigingmetamodels for expensive simulations. Working Paper, http://www.tilburguniversity.edu/research/institutes-and-researchgroups/center/staff/kleijnen/monotone Kriging.pdf.

Michalak A.M. (2008). A Gibbs sampler for inequality-constrained geostatistical interpolation and inverse modeling. Water Resour. Res., 44, W09437, doi:10.1029/2007WR006645.

Racine J.S., Parmeter C.F. and Du P. (2009). Constrained nonparametric kernel regression: Estimation and inference. Working Paper, http:/economics.ucr.edu/spring09/Racine paper for 5 8 09.pdf.

Ramsay J.O. and Silverman B.W. (2005). Functional Data Analysis. Springer Series in Statistics, Springer-Verlag.

Robert C.P. (1995). Simulation of truncated normal variables. Statistics and Computing, 5:121–125.

Tallis G.M. (1961). The moment generating function of the truncated multinormal distribution. Journal of the Royal Statistical Society, Series B, 23(1):223–229.

Tallis G.M. (1963). Elliptical and radial truncation in normal populations. Ann. Math. Statist., 34:940–944.

Tallis G.M. (1965). Plane truncation in normal populations. Journal of the Royal Statistical Society, Series B, 27(2):301–307.

Yoo E.-H. and Kyriakidis P.C. (2006). Area-to-point Kriging with inequality-type data. Journal of Geographical Systems, 8(4):357.