Kriging for arbitrage-free construction of financial term-structures

Areski Cousin
IRMA, Université de Strasbourg

stresstest2019 : International Workshop on Stress Test and Risk Management
Chaire "Stress Test, RISK Management and Financial Steering"

May 29, 2019

Introduction

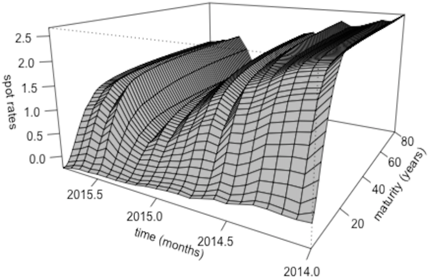
- Areski Cousin, Hassan Maatouk, Didier Rullière
  Kriging of financial term-structure, EJOR, 2016
- Areski Cousin, Djibril Gueye
  Kriging for arbitrage-free construction of volatility surfaces, working paper

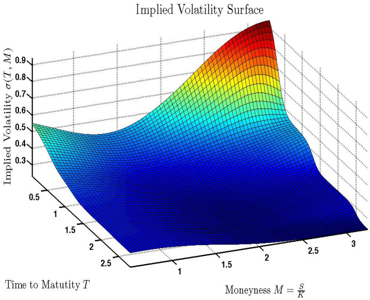
Motivation

- Learn a mapping f representing the evolution of a reference quantity Y as a function of some selected factors or explanatory variables X :

$$Y = f(X) \text{ for } X \in D \subset \mathbb{R}^d$$

- From observations of (input, output) couples : $(x_i, y_i), i = 1, \cdots, n$
- **Examples** : interest rates, default rates, implied volatilities, CVA exposures, mortality rates, surrender rates, computer experiments, any spacial data





Motivation

In risk management applications, this learning/construction problem typically has the following characteristics :

- Incomplete information : the response variable Y is only known or can only be estimated for a small set of input locations
- Indirect observation : the response variable may not be directly observed. (Typical when constructing ZC rate curves based on market quotes of some IR products)
- Noisy measurement : observed data may not be fully reliable (ex : price of illiquid instruments, Monte Carlo estimates, any statistical estimates)
- Known shape constraints : bounds on the response variable, monotonicity, convexity, ...

Motivation

What is kriging ?

- Kriging is a semi-parametric Bayesian estimation method also known as Gaussian Process Regression (or GP)
- It is a particular kernel machine learning method (see Rasmussen and Williams (2006)) but compared to frequentist machine learning techniques (support vector machines, neural network), GP estimates uncertainty
- Kriging also extends spline interpolation to uncertainty quantification : the kriging mean predictor is a spline function (curve in a RKHS with minimum norm, see Wahba (1990) or Bay el al. (2016))
- Implementation infrastructure is mature : R : DiceKriging and constrKriging, Matlab : GPML , Python : GPyTorch

Motivation

Kriging and risk management in the literature

- GP as surrogate model (estimation based on computer experiments) : Liu and Staum (2010), Ludkovski (2018), Ludkovski and Risk (2018), De Spiegeleer et al. (2018), Crépey and Dixon (2019)

- Kriging applied to real-world data (model-free) : Asgharian et al. 2013, Cousin et al. (2016), Ludkowski, Risk, Zail (2018)

- Portfolio optimization : da Barrosa et al. (2016)

- Time-series modelling : Roberts et al. (2013)

Our contributions :

- Show that kriging is a suitable tool for constructing financial term-structures and quantifying uncertainty
- Extend classical kriging to indirect observations, noisy measurements and shape-constraints

Contents

1 The term-structure construction problem

2 Classical kriging

3 Kriging with shape constraints

The term-structure construction problem

1) Compatibility with market data :

- We observe the market quotes $\mathbf{S} = (S_1, \ldots, S_n)$ of n liquidly traded instruments
- $\mathbf{S}$ depend on the value of f at m input locations $X = (x_1, \ldots, x_m)$

The vector of output values $f(X) := (f(x_1), \ldots, f(x_m))^\top$ satisfies a linear system of the form

$$A(\mathbf{S}) \cdot f(X) = \boldsymbol{b}(\mathbf{S}),$$

where

- $A(\mathbf{S})$ is a $n \times m$ real-valued matrix
- $\boldsymbol{b}(\mathbf{S})$ is a n-dimensional column vector

$n < m \Longrightarrow$ indirect and partial information on the curve values at $x_1, \ldots, x_m$

2) No-arbitrage assumption : f is e.g. a decreasing or a convex function

The term-structure construction problem

Example 1 : OIS discount curve

Construction of function $T \to D(t_0, T)$ from $S_i, i = 1, \ldots, n$, where

- S_i : par rate at quotation date t_0 of an OIS with maturity T_i
- $t_1 < \cdots < t_{p_i} = T_i$: fixed-leg payment dates (annual time grid)
- δ_k : year fraction of period (t_{k-1}, t_k)

$$S_i \sum_{k=1}^{p_i - 1} \delta_k D(t_0, t_k) + (S_i \delta_{p_i} + 1)\, D(t_0, T_i) = 1, \quad i = 1, ..., n$$

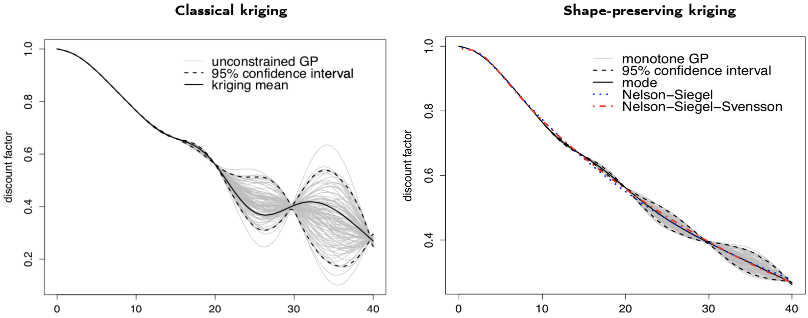
where $D(t_0, T)$ is the OIS discount factor with maturity T

The arbitrage-free curve $T \to D(t_0, T)$ is decreasing and $D(t_0, t_0) = 1$

The term-structure construction problem

Example 1 : OIS discount curve

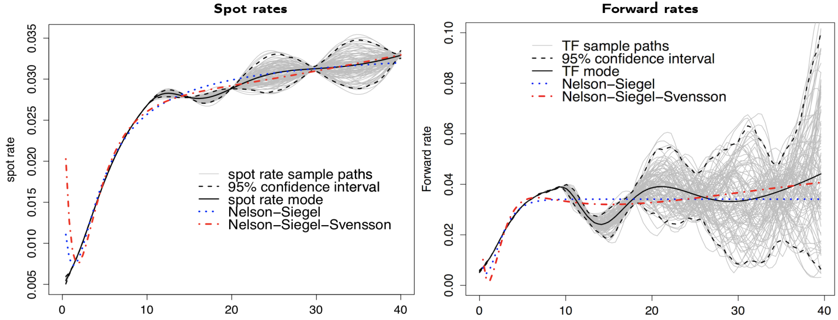
- Data : quoted swap rates as of June 3, 2010, for OIS with maturities $1y, \ldots, 10y, 15y, 20y, 30y, 40y$
- Classical kriging (left) vs kriging with monotonicity constraint (right)



The term-structure construction problem

Example 1 : OIS discount curve

- Corresponding spot rate and forward rate curves
- Monotonic kriging - GP prior with Matérn 5/2 kernel - no noise



The term-structure construction problem

Example 2 : Default rates implied from CDS spreads

- S_i : CDS spread at time t_0 with maturity T_i
- $t_1 < \cdots < t_{p_i} = T_i$: trimestrial premium payment dates, δ_k : year fraction of period (t_{k-1}, t_k)
- $D(t_0, T)$ is the discount factor associated with maturity date T
- R : expected recovery rate of the reference entity

$$S_i \sum_{k=1}^{p_i} \delta_k D(t_0, t_k) Q(t_0, t_k) = -(1 - R) \int_{t_0}^{T_i} D(t_0, u) dQ(t_0, u)$$

where $T \to Q(t_0, T)$ is the $\mathcal{F}_{t_0}$-conditional (risk-neutral) survival distribution of the reference entity, i.e.,

$$Q(t_0, T) = \mathbb{Q}(\tau > T \mid \mathcal{F}_{t_0})$$

The term-structure construction problem

Example 2 : Default rates implied from CDS spreads (cont.)

Using an integration by parts, the survival function $u \to Q(t_0, u)$ satisfies a linear relation :

$$S_i \sum_{k=1}^{p_i} \delta_k D(t_0, t_k) Q(t_0, t_k) + (1 - R) D(t_0, T_i) Q(t_0, T_i)$$
$$+ (1 - R) \int_{t_0}^{T_i} f(t_0, u) D(t_0, u) Q(t_0, u) du = 1 - R, \; i = 1, \ldots, n$$

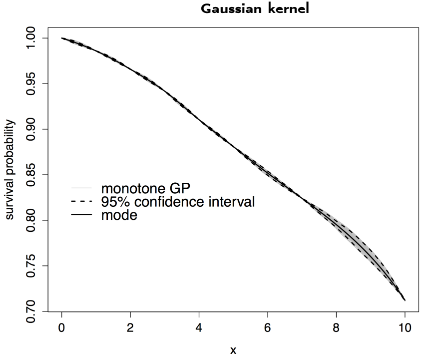
where $f(t_0, u)$ is the instantaneous forward (discount) rate associated with maturity date u.

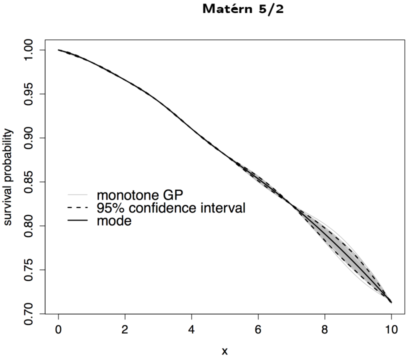
As a survival function, $T \to Q(t_0, T)$ **shall be decreasing and such that** $Q(t_0, t_0) = 1$

The term-structure construction problem

Example 2 : Default rates implied from CDS spreads (cont.)

- CDS spreads for protection maturities 1y, 2y, 3y, 4y, 5y, 7y, 10y
- Russian sovereign debt, quotes as of 06/01/2005
- Monotonic kriging - GP without noise

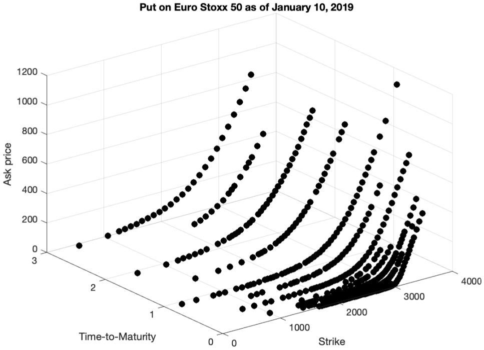




The term-structure construction problem

Example 3 : Volatility surface

We observe at t_0, a series of put option prices $f(x_i) = P((K_i, T_i))$ for different characteristics $x_i = (K_i, T_i)$, $i = 1, \ldots, n$.



The term-structure construction problem

Example 3 : Volatility surface

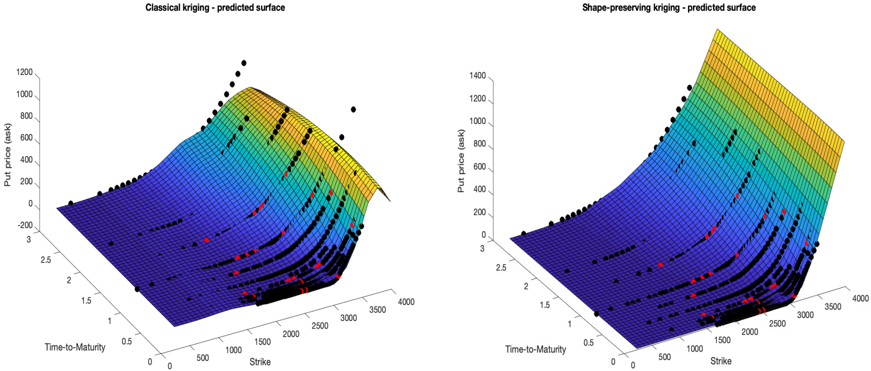
The put price surface $(K, T) \rightarrow P(K, T)$ is free of static arbitrage if

- $K \rightarrow P(K, T)$ is a convex function such that $P(0, T) = 0$ and $\frac{\partial P}{\partial K}(0, T) = 0$, for any $T \geq 0$
- $T \rightarrow P(K, T)$ is a non-decreasing function, for any $K \geq 0$
- $P(K, 0) = (K - S_0)^+$ where S_0 is the spot price.

The term-structure construction problem

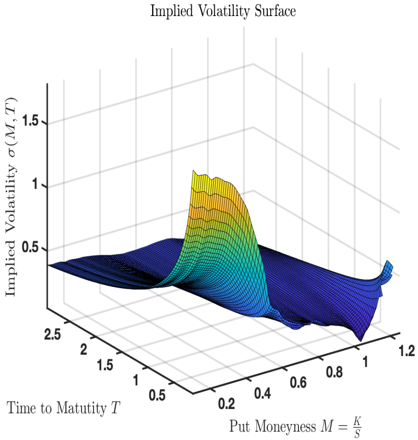
Example 3 : Volatility surface

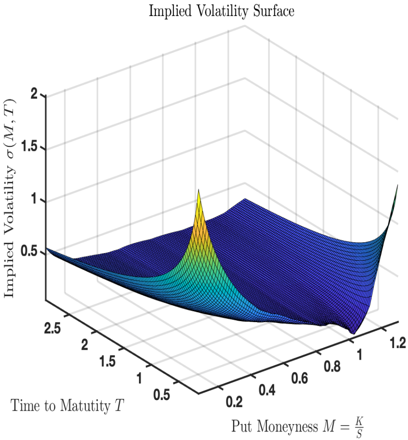
- Data : Euro Stoxx 50 Put prices as of January 10, 2019
- 5% of the data used (red points)
- Classical kriging (left) vs kriging with no-arbitrage constraints (right)



The term-structure construction problem

Example 3 : Volatility surface

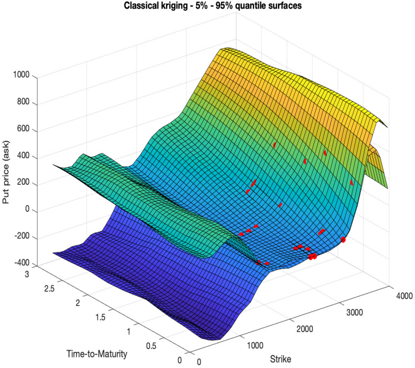


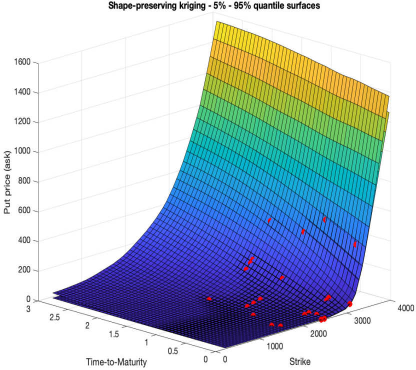


The term-structure construction problem

Example 3 : Volatility surface

- 5% and 95% estimated quantiles of the fitted GP
- Classical kriging (left) vs kriging with no-arbitrage constraints (right)





Contents

1 The term-structure construction problem

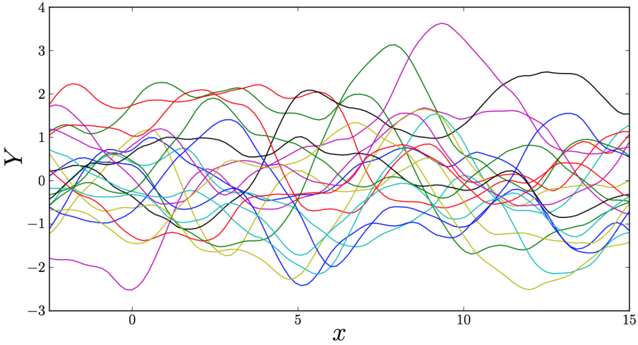
2 Classical kriging

3 Kriging with shape constraints

Classical kriging

Estimation of the unknown function f using Bayesian statistics

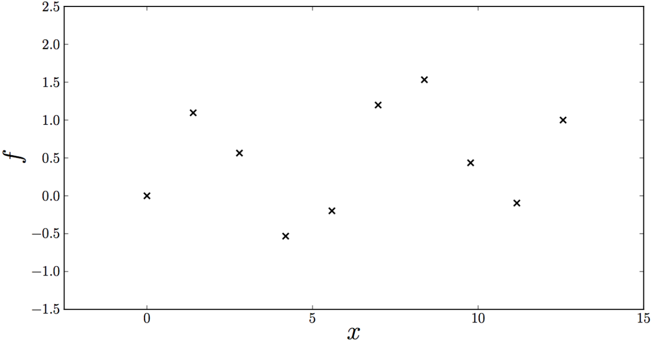
Our first belief in f is given as a Gaussian process prior Y



Classical kriging

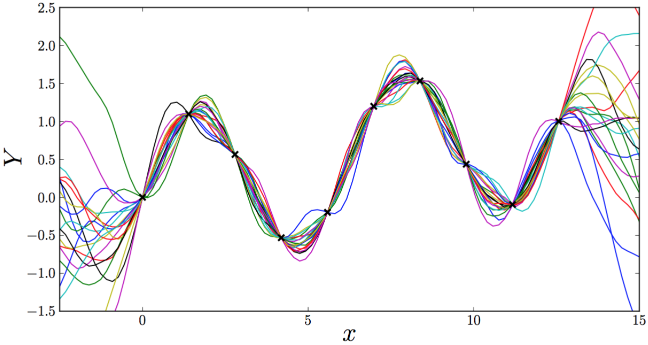
The function f is known at some input points $x^1, \dots, x^n$:

$$f\left(x^1\right) = y^1, \dots, f\left(x^n\right) = y^n.$$



Classical kriging

This belief is updated given that $Y(x_1) = y_1, \ldots, Y(x_n) = y_n$



Source : presentation of N. Durrande

Classical kriging

Definition : Gaussian process (GP) or Gaussian random field

A Gaussian process is a collection of random variables, any finite number of which have (consistent) joint Gaussian distributions.

A Gaussian process $\left(Y(x), x \in \mathbb{R}^d\right)$ is characterized by its mean function

$$\mu : x \in \mathbb{R}^d \longrightarrow \mathbb{E}(Y(x)) \in \mathbb{R}.$$

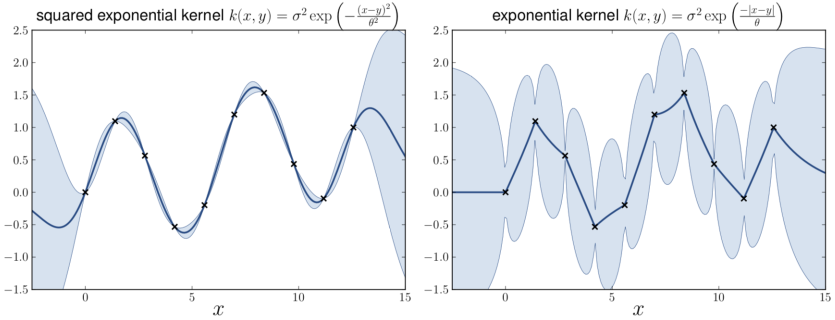
and its covariance function

$$K : (x, x') \in \mathbb{R}^d \times \mathbb{R}^d \longrightarrow \mathrm{Cov}(Y(x), Y(x')) \in \mathbb{R}.$$

1D kriging kernel	$K(x, x')$	Class
Gaussian	$\sigma^2 \exp\left(-\frac{(x-x')^2}{2\theta^2}\right)$	$\mathcal{C}^\infty$
Matérn 5/2	$\sigma^2 \left(1 + \frac{\sqrt{5}\lvert x-x'\rvert}{\theta} + \frac{5(x-x')^2}{3\theta^2}\right) \exp\left(-\frac{\sqrt{5}\lvert x-x'\rvert}{\theta}\right)$	$\mathcal{C}^2$
Matérn 3/2	$\sigma^2 \left(1 + \frac{\sqrt{3}\lvert x-x'\rvert}{\theta}\right) \exp\left(-\frac{\sqrt{3}\lvert x-x'\rvert}{\theta}\right)$	$\mathcal{C}^1$
Exponential	$\sigma^2 \exp\left(-\frac{\lvert x-x'\rvert}{\theta}\right)$	$\mathcal{C}^0$

Classical kriging

Changing the kernel K has a huge impact on the model



Source : presentation of N. Durrande

Classical kriging - indirect observations with noise

Assume that f is known up to solving a linear equality system with measurement errors :

$$A \cdot f(X) + \varepsilon = \boldsymbol{b}. \tag{1}$$

where

- A is a given $n \times m$ matrix
- $X = (x_1, \ldots, x_m)^\top \in \mathbb{R}^{m \times d}$
- $f(X) = (f(x_1), \ldots, f(x_m))^\top \in \mathbb{R}^m$
- $\boldsymbol{b} \in \mathbb{R}^n$
- ε is zero-mean Gaussian noise in $\mathbb{R}^n$ with covariance matrix Σ_{noise}
- ε is assumed to be independent of the GP Y

Classical kriging - indirect observations with noise

- $X = (x_1, \ldots, x_m)^\top \in \mathbb{R}^{m \times d}$: some design points
- $\boldsymbol{b} = (b_1, \ldots, b_n)^\top \in \mathbb{R}^n$: right-hand side of the linear system
- $Y(X) = (Y(x_1), \ldots, Y(x_m))$: vector composed of Y at point $\boldsymbol{X}$

The conditional process is still a Gaussian Process

Let Y be a GP with mean μ and covariance function K. The conditional process $Y \mid AY(X) + \varepsilon = \boldsymbol{b}$ is a GP with mean function

$$\eta(x) = \mu(x) + (A\boldsymbol{k}(x))^\top \left(A\mathbb{K}A^\top + \Sigma_{noise}\right)^{-1} (\boldsymbol{b} - A\boldsymbol{\mu}), \quad x \in \mathbb{R}^d$$

and covariance function $\tilde{K}$ given by

$$\tilde{K}(x, x') = K(x, x') - (A\boldsymbol{k}(x))^\top \left(A\mathbb{K}A^\top + \Sigma_{noise}\right)^{-1} A\boldsymbol{k}(x'), \quad x, x' \in \mathbb{R}^d$$

where $\boldsymbol{\mu} = \mu(X) = (\mu(x_1), \ldots, \mu(x_m))^\top$, $\mathbb{K}$ is the covariance matrix of $Y(X)$, $\boldsymbol{k}(x) = (K(x, x_1), \ldots, K(x, x_m))^\top$

Contents

1. The term-structure construction problem
2. Classical kriging
3. Kriging with shape constraints

Shape-preserving kriging

New formulation of the problem : estimation of an unknown real-valued function $f : [0,1]^d \to \mathbb{R}$ given that

$$\begin{cases} A \cdot f(X) + \varepsilon = \boldsymbol{b} \\ f \in \mathcal{M} \end{cases}$$

where $\mathcal{M}$ is a convex set of functions satisfying some shape property.

For instance, $\mathcal{M}$ can be :

- $\mathcal{M}_0^d := \{f \in \mathcal{C}([0,1]^d, \mathbb{R}) \mid y_{\text{min}} \leq f(x) \leq y_{\text{max}},\ \forall x \in D\}$
- $\mathcal{M}_1^1 := \{f \in \mathcal{C}([0,1], \mathbb{R}) \mid f \text{ is non-decreasing}\}$
- $\mathcal{M}_2^1 = \{f \in \mathcal{C}([0,1], \mathbb{R}) \mid f \text{ is convex}\}$
- $\mathcal{M}_{12}^2 = \{f \in \mathcal{C}([0,1]^2, \mathbb{R}) \mid f \text{ is non-decreasing in } x \text{ and convex in } y\}$

Shape-preserving kriging

Main issues :

- The posterior process is not Gaussian anymore.
- The shape condition is usually infinite-dimensional.

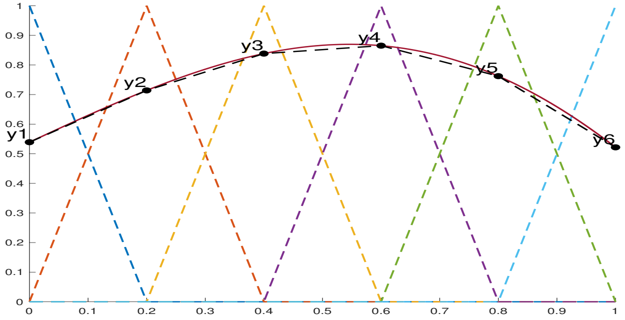
Proposed solutions :

- We construct a finite-dimensional approximation of Y for which the shape condition is easy to check.
- We consider the mode of the posterior distribution (as opposed to the posterior mean) as a new response surface estimator
- Hyper-parameters are estimated using MLE

Finite-dimensional approximation of GP (1d case)

As in Maatouk and Bay (2014), Cousin et al. (2016), López et al. (2018), we rely on basis function approximation.

- Input domain D is discretized on a regular subdivision $u_0 < \ldots < u_N$ with a constant mesh δ.
- For each u_i, we consider hat functions $\phi_i(x) := \max\left(1 - \frac{|x-u_i|}{\delta}, 0\right)$
- Y is approximated on D by $Y^N(x) = \sum_{i=0}^N Y(u_i)\phi_i(x)$



Finite-dimensional approximation of GP (1d case)

Proposition

Let Y be a zero-mean GP with covariance function K and almost surely continuous paths.

- The finite-dimensional process $Y^N(\cdot) = \sum_{i=0}^N Y(u_i)\phi_i(\cdot)$ uniformly converges to Y on D as $N \to \infty$, almost surely.
- $Y^N(x) = \mathbf{\Phi}(x)\boldsymbol{\xi}$ where $\boldsymbol{\xi} := (Y(u_0), \ldots, Y(u_N))^\top$ is a zero-mean Gaussian vector with covariance matrix Γ^N such that $\Gamma^N_{i,j} = K(u_i, u_j)$

Shape-preserving conditions :

- Y^N takes values on $[y_{\min}, y_{\max}]$ if and only if $y_{\min} \leq \xi_i \leq y_{\max}$
- Y^N is non-decreasing on D if and only if $\xi_{i+1} \geq \xi_i$
- Y^N is convex on D if and only if $\xi_{i+2} - \xi_{i+1} \geq \xi_{i+1} - \xi_i$
- $\ldots$

Finite-dimensional approximation of GP (2d case)

- $D = [0,1]^2$ is discretized on a $(N_x + 1) \times (N_t + 1)$ regular grid with knots (u_i, v_j), $i = 1, \ldots, N_x$, $j = 1, \ldots, N_t$.
- For each knot (u_i, v_j), we consider tensor product basis functions

$$\phi_{i,j}(x,t) := \max\left(1 - \frac{|x - u_i|}{\delta_x}, 0\right) \max\left(1 - \frac{|t - v_j|}{\delta_t}, 0\right)$$

- Y is approximated on D by

$$Y^N(x,t) = \sum_{i=0}^{N_x} \sum_{j=0}^{N_t} Y(u_i, v_j) \phi_{i,j}(x,t)$$

- $N = (N_x + 1)(N_t + 1)$ is the number of knots

Finite-dimensional approximation of GP (2d case)

Proposition

Let Y be a zero-mean GP with covariance function K and with almost surely continuous paths.

- The finite-dimensional process Y^N uniformly converges to Y on D as $N_x \to \infty$ and $N_t \to \infty$, almost surely.
- $Y^N(x) = \mathbf{\Phi}(x)\boldsymbol{\xi}$ where $\boldsymbol{\xi} := (Y(u_0, v_0), Y(u_0, v_1), \ldots, Y(u_{N_x}, v_{N_t}))^\top$ is a zero-mean Gaussian vector with $N \times N$ covariance matrix Γ^N such that $\Gamma^N = K((u_{i_1}, v_{j_1}), (u_{i_2}, v_{j_2}))$.

Shape-preserving conditions :

- Y^N is bounded on $[y_{\min}, y_{\max}]$ if and only if $y_{\min} \leq \xi_{i,j} \leq y_{\max}$
- $Y^N(x,t)$ is non-decreasing function of x if and only if $\xi_{i+1,j} \geq \xi_{i,j}$
- $Y^N(x,t)$ is a convex function of x if and only if $\xi_{i+2,j} - \xi_{i+1,j} \geq \xi_{i+1,j} - \xi_{i,j}$
- $\ldots$

Kriging under shape-preserving conditions

Consider a zero-mean GP prior Y with covariance function K and N-dimensional approximation Y^N.

Kriging the unknown function f boils down to finding the conditional distribution of Y^N given

$$\begin{cases} A \cdot Y^N(X) + \varepsilon = \boldsymbol{b} \\ Y^N \in \mathcal{M} \end{cases}$$

This is equivalent to finding the distribution of the truncated Gaussian vector $\boldsymbol{\xi} \sim \mathcal{N}(0, \Gamma^N)$ given that

$$\begin{cases} A \cdot \boldsymbol{\Phi}(X) \cdot \boldsymbol{\xi} + \varepsilon = \boldsymbol{b} \\ \boldsymbol{\xi} \in \mathcal{C}_{ineq} \end{cases}$$

where $\mathcal{C}_{ineq}$ is a set of linear inequality constraints.

Estimation of hyper-parameters

- We consider d-dimensional anisotropic stationary kernels :

$$K(x, x') = \sigma^2 \prod_{i=1}^{d} K_i(x_i - x_i'; \theta_i)$$

  where K_i is stationary kernel : Gaussian, Matérn 5/2, Matérn 3/2, Exponential.

- Homoscedastic noise : $\Sigma_{noise} = \sigma_{noise} \mathbb{I}_n$
- **Hyper-parameters** : $\boldsymbol{p} = (\sigma, \theta_1, \ldots, \theta_d, \sigma_{noise})$

Estimation of hyper-parameters

Following López-Lopera et al (2017), we consider two MLE approaches

- **Unconditional likelihood** : Find $\boldsymbol{p}$ that maximizes the Gaussian likelihood $\mathbb{P}\left(A \cdot \boldsymbol{\Phi}(X) \cdot \boldsymbol{\xi} + \boldsymbol{\varepsilon} = \boldsymbol{b} \mid \boldsymbol{p}\right)$ or log-likelihood

$$\mathcal{L}_N(\boldsymbol{p}) := -\frac{n}{2}\log(2\pi) - \frac{1}{2}\log|C| - \frac{1}{2}\boldsymbol{b}^\top C^{-1}\boldsymbol{b}$$

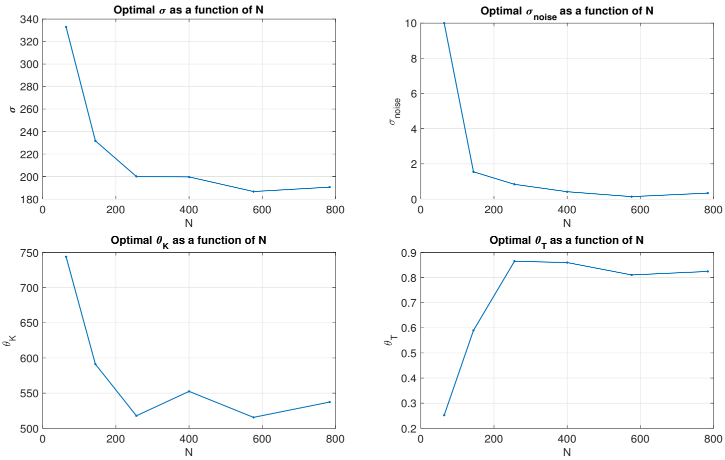
where $C := A\boldsymbol{\Phi}(X)\Gamma^N(\boldsymbol{p})\boldsymbol{\Phi}(X)A^\top + \Sigma_{noise}(\boldsymbol{p})$

- **Conditional likelihood** : Find $\boldsymbol{p}$ that maximizes the conditional probability $\mathbb{P}\left(A \cdot \boldsymbol{\Phi}(X) \cdot \boldsymbol{\xi} + \boldsymbol{\varepsilon} = \boldsymbol{b} \mid \boldsymbol{\xi} \in \mathcal{C}_{ineq}, \boldsymbol{p}\right)$ or the log-likelihood

$$\mathcal{L}_{N,cond}(\boldsymbol{p}) := \mathcal{L}_N(\boldsymbol{p}) + \log\mathbb{P}(\boldsymbol{\xi} \in \mathcal{C}_{ineq} \mid A\cdot\boldsymbol{\Phi}(X)\cdot\boldsymbol{\xi} + \boldsymbol{\varepsilon} = \boldsymbol{b}) - \log\mathbb{P}(\boldsymbol{\xi} \in \mathcal{C}_{ineq})$$

Estimation of hyper-parameters

Convergence of optimal parameter as a function of N (number of basis functions)



Mode estimator

We define the (a posteriori) most probable response surface and measurement noises as

$$\begin{cases} M_K^N(x) := \boldsymbol{\Phi}(x) \cdot (c_1^*, \ldots, c_N^*)^\top, \; x \in D \\ \boldsymbol{e}^* := (e_1^*, \ldots, e_n^*)^\top \end{cases}$$

where $(\boldsymbol{c}^*, \boldsymbol{e}^*)$ is the mode of the truncated Gaussian vector $(\boldsymbol{\xi}, \boldsymbol{\varepsilon})$ given the constraints, defined as solution of

$$\max_{\boldsymbol{c}, \boldsymbol{e}} \mathbb{P} \left(\boldsymbol{\xi} \in [\boldsymbol{c}, \boldsymbol{c} + d\boldsymbol{c}], \boldsymbol{\varepsilon} \in [\boldsymbol{e}, \boldsymbol{e} + d\boldsymbol{e}] \mid A \cdot \boldsymbol{\Phi}(X) \cdot \boldsymbol{\xi} + \boldsymbol{\varepsilon} = \boldsymbol{b}, \; \boldsymbol{\xi} \in \mathcal{C}_{ineq} \right).$$

The mode $(\boldsymbol{c}^*, \boldsymbol{e}^*)$ is solution of a quadratic problem

$$\min_{A \cdot \boldsymbol{\Phi}(X) \cdot \boldsymbol{c} + \boldsymbol{e} = \boldsymbol{b}, \; \boldsymbol{c} \in \mathcal{C}_{ineq}} \left(\boldsymbol{c}^\top (\Gamma^N)^{-1} \boldsymbol{c} + \boldsymbol{e}^\top \Sigma_{noise}^{-1} \boldsymbol{e} \right)$$

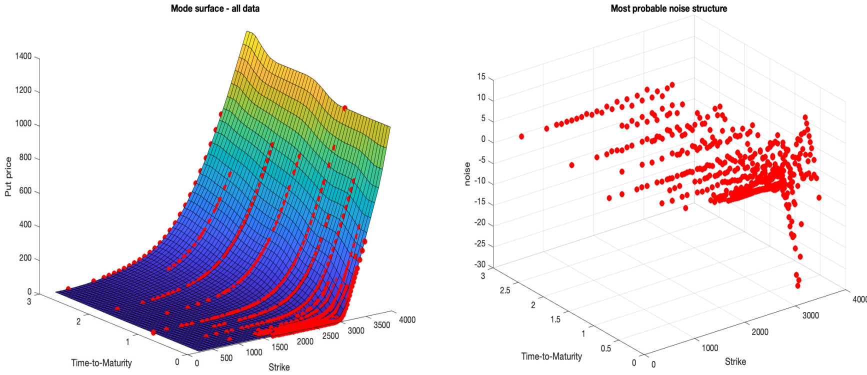
Mode estimator

The mode estimator has several advantages (over alternative estimators) :

- It satisfies the constraints on the entire domain D
- It is easy to compute as the solution of a quadratic optimisation problem
- It corresponds to the maximum a posteriori estimator in the sense of Bayesian statistics
- As N tends to infinity, the limit of M_K^N corresponds to a constrained spline that depends on K (Bay et al., 2016)

Mode estimator

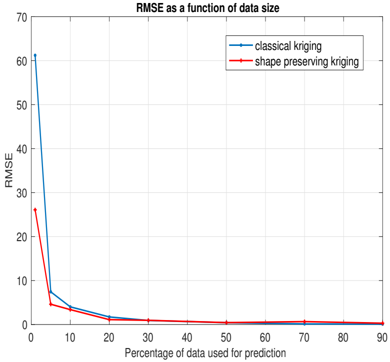
- Data : Euro Stoxx 50 Put prices as of January 10, 2019
- Fitted Gaussian kernel using uncond. MLE, all data used
- Most probable surface (left) vs most probable noise values (right)



Mode estimator - prediction accuracy

RMSE as a function of data size

- We construct a series of data subsets with increasing number of points
- We apply classical kriging and shape-preserving kriging on these subsets
- For each data size, we compute average RMSE wrt the original data set.



Sampling finite-dimensional GP with shape constraints

First remark that the distribution of $\boldsymbol{\xi}$ given $A \cdot \boldsymbol{\Phi}(X) \cdot \boldsymbol{\xi} + \varepsilon = \boldsymbol{b}$ is multinormal $\mathcal{N}(\boldsymbol{\mu}_{cond}, \Sigma_{cond})$ where

$$\begin{cases} \boldsymbol{\mu}_{cond} = \Gamma^N B^\top \left(B \Gamma^N B^\top + \Sigma_{noise} \right)^{-1} \boldsymbol{b} \\ \Sigma_{cond} = \Gamma^N - \Gamma^N B^\top \left(B \Gamma^N B^\top + \Sigma_{noise} \right)^{-1} B \Gamma^N \end{cases}$$

with $B = A \cdot \boldsymbol{\Phi}(X)$.

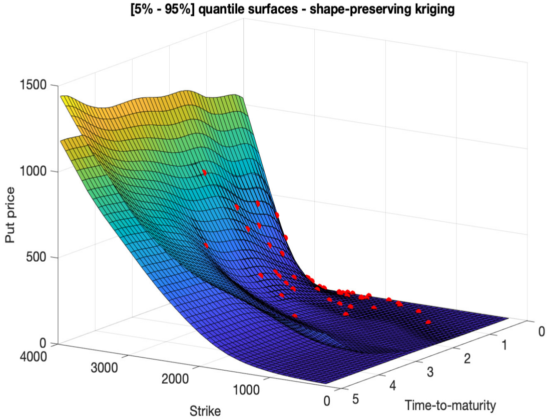
Following López-Lopera et al (2017), we consider the Hamiltonian Monte Carlo method introduced by Pakman and Paninski (2013) for sampling truncated multivariate Gaussians :

$$\mathcal{TN}\left(\boldsymbol{\mu}_{cond}, \Sigma_{cond}, \mathcal{C}_{ineq}\right)$$

MCMC initialized using the mode estimator since it satisfies the inequality constraints .

Sampling finite-dimensional GP with shape constraints

- We extrapolate the GP in T direction (adding 2 years)
- 5% and 95% estimated pointwise quantiles



Thanks for your attention.

References I

Asgharian, H., Hess, W., and Liu, L. (2013).
A spatial analysis of international stock market linkages.
Journal of Banking & Finance, 37(12) :4738–4754.

Bay, X., Grammont, L., and Maatouk, H. (2016).
Generalization of the Kimeldorf-Wahba correspondence for constrained interpolation.

Cousin, A., Maatouk, H., and Rullière, D. (2016).
Kriging of financial term-structures.
European J. Oper. Res., 255(2) :631–648.

Cressie, N. (1990).
The origins of kriging.
Math. Geol., 22(3) :239–252.

da Barrosa, M. R., Salles, A. V., and Ribeiro, C. d. O. (2016).
Portfolio optimization through kriging methods.
Applied Economics, 48(50) :4894–4905.

De Spiegeleer, J., Madan, D. B., Reyners, S., and Schoutens, W. (2018).
Machine learning for quantitative finance : fast derivative pricing, hedging and fitting.
Quantitative Finance, 18(10) :1635–1643.

References II

Dixon, M. F. and Crépey, S. (2018).
Multivariate gaussian process regression for derivative portfolio modeling : Application to cva.

Liu, M. and Staum, J. (2010).
Stochastic kriging for efficient nested simulation of expected shortfall.
Journal of Risk, 12(3) :3.

López-Lopera, A. F., Bachoc, F., Durrande, N., and Roustant, O. (2018).
Finite-dimensional gaussian approximation with linear inequality constraints.
SIAM/ASA Journal on Uncertainty Quantification, 6(3) :1224–1255.

Ludkovski, M. (2018).
Kriging metamodels and experimental design for bermudan option pricing.
Journal of Computational Finance, 22(1) :37–77.

Ludkovski, M. and Risk, J. (2018).
Sequential design and spatial modeling for portfolio tail risk measurement.
SIAM Journal of Financial Mathematics.

Ludkovski, M., Risk, J., and Zail, H. (2018).
Gaussian process models for mortality rates and improvement factors.
ASTIN Bulletin : The Journal of the IAA, 48(3) :1307–1347.

References III

Maatouk, H. and Bay, X. (2014).
Gaussian Process Emulators for Computer Experiments with Inequality Constraints.
in revision SIAM/ASA J. Uncertainty Quantification.

Pakman, A. and Paninski, L. (2014).
Exact hamiltonian monte carlo for truncated multivariate gaussians.
Journal of Computational and Graphical Statistics, 23(2) :518–542.

Roberts, S., Osborne, M., Ebden, M., Reece, S., Gibson, N., and Aigrain, S. (2013).
Gaussian processes for time-series modelling.
Philosophical Transactions of the Royal Society A : Mathematical, Physical and Engineering Sciences, 371(1984) :20110550.

Wahba, G. (1990).
Spline models for observational data, volume 59.
Siam.

Williams, C. K. and Rasmussen, C. E. (2006).
Gaussian processes for machine learning.
the MIT Press, 2(3) :4.