Advanced Surrogate Modeling

Kriging Interpolation

Octavio Narváez-Aroche

Mechanical Engineering, PhD candidate

Berkeley Center for Control and Identification

Based on a lecture by Prof. Andrew Packard

November 11th, 2016

Overview

Kriging is a **parametric regression method** widely used in computational simulation, derived from the notion of "minimizing" both the expected value, and variance of the error attained by a linear estimator of a random vector.

- Provides an unbiased surrogate model at on-design samples, which makes it suitable for interpolation.

- The variance of the estimation error can be used for setting bounds on off-design samples.

Optimization Problem

Given random vectors $W \in \mathbb{R}^m$, *and* $Z \in \mathbb{R}^n$ *with known mean* $\mu_Z := E[Z]$, *variance* $\Sigma_Z := E\left[(Z - \mu_Z)(Z - \mu_Z)^\top\right]$, *and covariance* $\Sigma_{ZW} := E\left[(Z - \mu_Z)(W - \mu_W)^\top\right]$, *find* $L \in \mathbb{R}^{m \times n}$, *and* $b \in \mathbb{R}^m$ *such that* $\widehat{W} := LZ + b$ *is the linear estimate of* W *from* Z *that minimizes a measure of* $E\left[\widehat{W} - W\right]$, *and* $E\left[\left(\widehat{W} - W\right)\left(\widehat{W} - W\right)^\top\right]$.

Expected Estimation Error

The expected value of the error with the linear estimator is

$$\begin{aligned} E\left[\hat{W} - W\right] &= E\left[LZ + b - W\right] \\ &= L\left(E\left[Z\right]\right) + b - E\left[W\right] \\ &= L\mu_Z - \mu_W + b \end{aligned} \qquad \text{...(1)}$$

Suppose that we pick L, by plugging

$$b = \mu_W - L\mu_Z \qquad \text{...(2)}$$

into (1), we have $E\left[\widehat{W} - W\right] = 0$.

Variance of Estimation Error

With the value of b in (2) we have

$$
\begin{aligned}
E\left[\left(\hat{W}-W\right)\left(\hat{W}-W\right)^{\top}\right] = & \\
&= E\left[(LZ+b-W)(LZ+b-W)^{\top}\right] \\
&= E\left[(LZ+\mu_W-L\mu_Z-W)(LZ+\mu_W-L\mu_Z-W)^{\top}\right] \\
&= E\left[(L(Z-\mu_Z)-(W-\mu_W))\left((Z-\mu_Z)^{\top}L^{\top}-(W-\mu_W)^{\top}\right)\right] \\
&= L\Sigma_Z L^{\top} - L\Sigma_{ZW} - \Sigma_{ZW}^{\top}L^{\top} + \Sigma_W \\
&=: g(L)
\end{aligned}
\qquad \text{...(3)}
$$

Matrix-valued Completion of Squares

Let $A = A^\top \in \mathbb{R}^{n\times n},\ A \succ 0,\ B \in \mathbb{R}^{n\times m},$ and $C = C^\top \in \mathbb{R}^{m\times m}$.

Define $h\colon \mathbb{R}^{m\times n} \to \mathbb{R}^{m\times m}$ as

$$h(K) := KAK^\top - KB - B^\top K^\top + C.$$

Theorem: There exists $K^\star \in \mathbb{R}^{m\times n}$ such that

$$h(K^\star) \preceq h(K)\ \forall K \in \mathbb{R}^{m\times n}. \qquad \text{...(4)}$$

In fact:

$$K^\star = B^\top A^{-1} \qquad \text{...(5)}$$

$$h(K^\star) = C - B^\top A^{-1} B$$

Linear Minimum-variance Estimator

According to (4), and (5) the matrix $L^\star$ that plugged in (3) achieves $g(L^\star) \preccurlyeq g(L)\ \forall\, L \in \mathbb{R}^{m\times n}$ is

$$L^\star := \Sigma_{ZW}^\top \Sigma_Z^{-1}, \qquad \text{...(6)}$$

with a variance on the estimation error of

$$g\left(L^\star\right) = \Sigma_W - \Sigma_{ZW}^\top \Sigma_Z^{-1} \Sigma_{ZW}. \qquad \text{...(7)}$$

Using (2), and (6) the linear minimum-variance estimator is

$$\begin{aligned} \hat{W} &= L^* Z + \mu_W - L^* \mu_Z = \Sigma_{ZW}^T \Sigma_Z^{-1} Z + \mu_W - \Sigma_{ZW}^T \Sigma_Z^{-1} \mu_Z \\ &= \Sigma_{ZW}^T \Sigma_Z^{-1} \left(Z - \mu_Z\right) + \mu_W \end{aligned} \qquad \text{...(8)}$$

The Kriging method is a Gaussian process.

Interpolation problem and Kriging

Let $f: \mathbb{R}^n \to \mathbb{R}$.

Given a set of samples $\{x_k, z_k = f(x_k)\}_{k=1}^{N}$, *estimate the value of the function at some new value* x_0.

The Kriging approach:

Model the dependence of z_k on x_k as a stochastic process with known mean, and covariance. Use the samples and the linear estimator in (8) to make a prediction of $f(x_0)$.

Kriging Interpolation

1. Fit the sample points with a (linear, quadratic, etc.) function $\phi: \mathbb{R}^n \to \mathbb{R}$ to compute μ_W, and μ_Z.

2. Consider a hyperparameter σ in the definition of a covariance function $\rho(\sigma, w, z) =: \mathrm{R}_\sigma(w, z)$ for $w, z \in \mathbb{R}^n$ to compute Σ_{WZ}, and Σ_Z.

Kriging Interpolation

3. The linear-minimum variance estimator for $f(x_0)$ becomes:

$$\Sigma_{ZW}^{T}\Sigma_{Z}^{-1}\left(Z-\mu_Z\right)+\mu_W =$$

$$= \begin{bmatrix} R_\sigma\left(x_0,x_1\right) \\ \vdots \\ R_\sigma\left(x_0,x_N\right) \end{bmatrix}^\top \underbrace{\begin{bmatrix} R_\sigma\left(x_1,x_1\right) & R_\sigma\left(x_1,x_2\right) & \cdots & R_\sigma\left(x_1,x_N\right) \\ R_\sigma\left(x_2,x_1\right) & R_\sigma\left(x_2,x_2\right) & \cdots & R_\sigma\left(x_2,x_N\right) \\ \vdots & \vdots & \ddots & \vdots \\ R_\sigma\left(x_N,x_1\right) & R_\sigma\left(x_N,x_2\right) & \cdots & R_\sigma\left(x_N,x_N\right) \end{bmatrix}}_{R :=}^{-1} \begin{bmatrix} z_1-\phi\left(x_1\right) \\ z_2-\phi\left(x_2\right) \\ \vdots \\ z_n-\phi\left(x_N\right) \end{bmatrix} + \phi\left(x_0\right)$$

$$=: \hat{f}\left(x_0\right)$$

Kriging interpolation

4. Choose the best value for the hyperparameter σ by cross-validation.

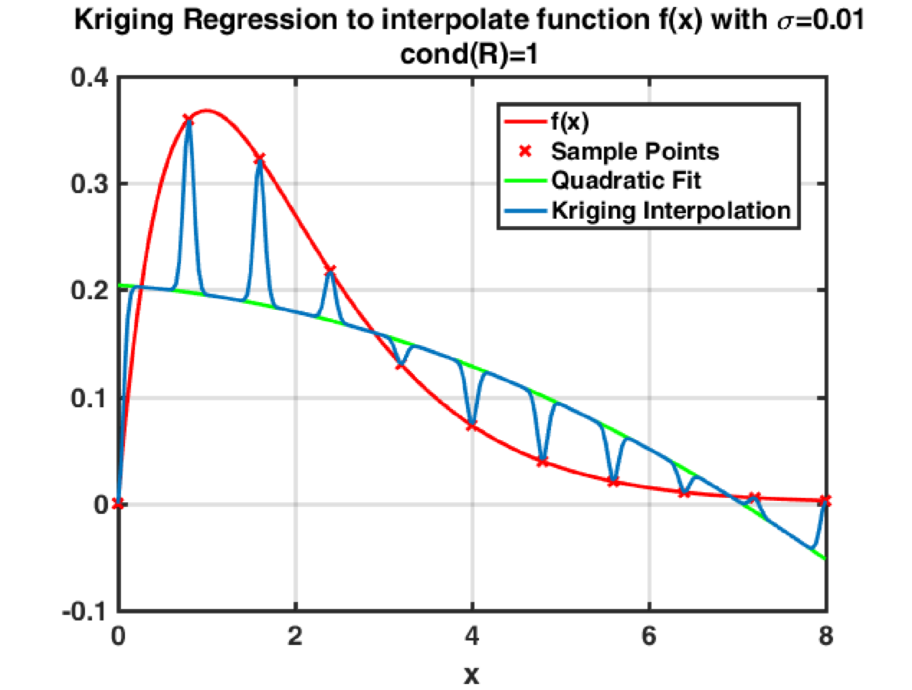
5. Compute upper and lower bounds for the estimation of $f(x_0)$ using the variance of the estimation error in (7).

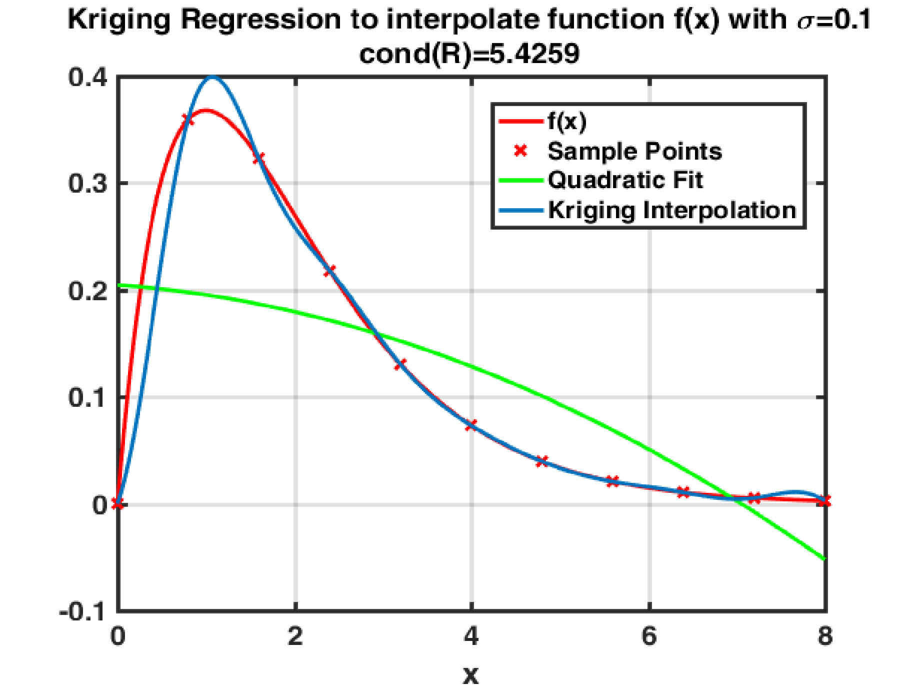
Interpolation Example for $f(x) \coloneqq xe^{-x}$

Code available at https://github.com/ocnaar

Tuning the hyper-parameter $\sigma \in (0, 10]$

$f(x) \coloneqq xe^{-x}$ $\quad \phi(x) \coloneqq a + bx + cx^2$ $\quad R_\sigma(\theta, \eta) \coloneqq e^{\frac{-(\theta-\eta)^2}{\sigma^2}}$ $\quad \hat{f}(x)$

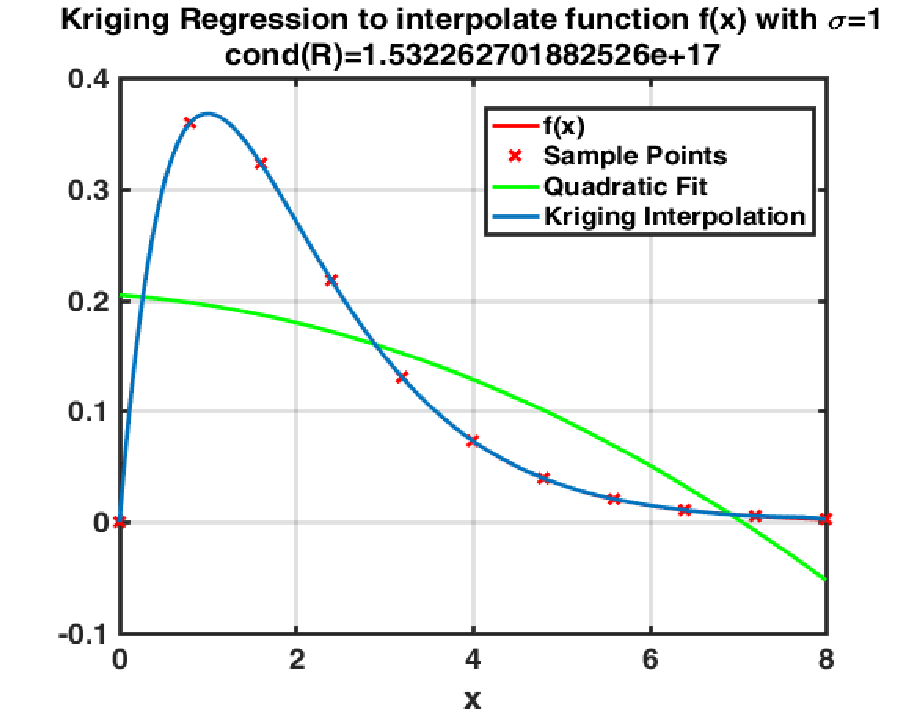


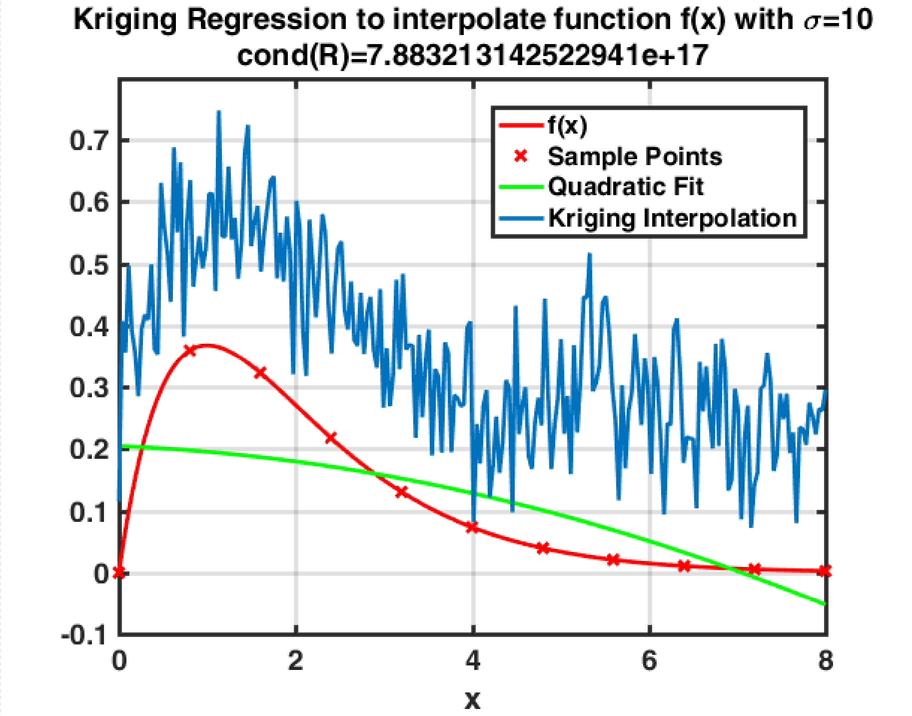


$a, b, c, x, \eta, \theta \in \mathbb{R}$

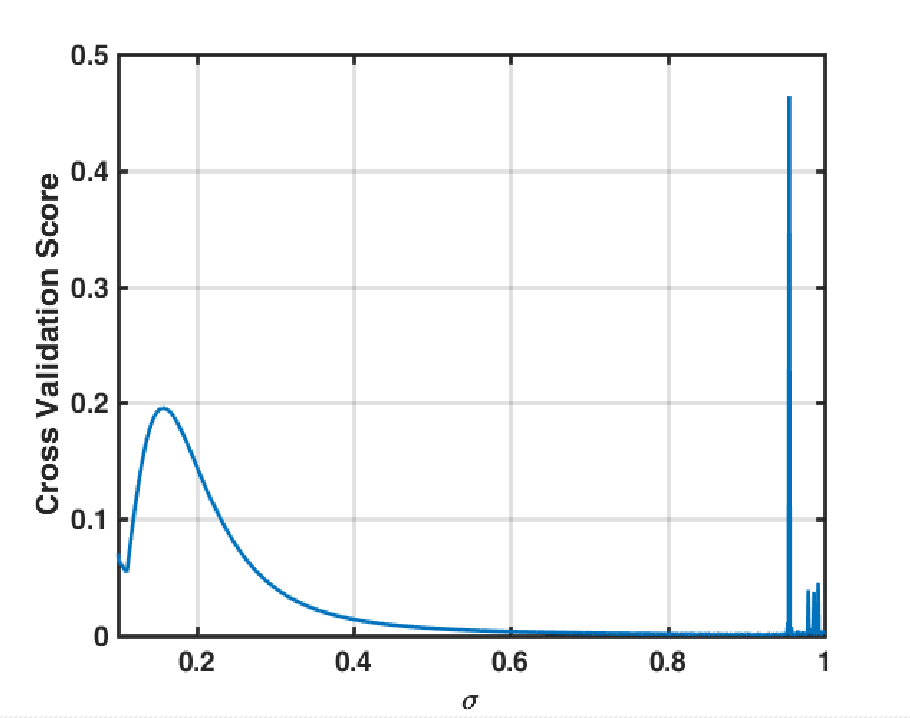
Tuning the hyper-parameter $\sigma \in (0, 10]$

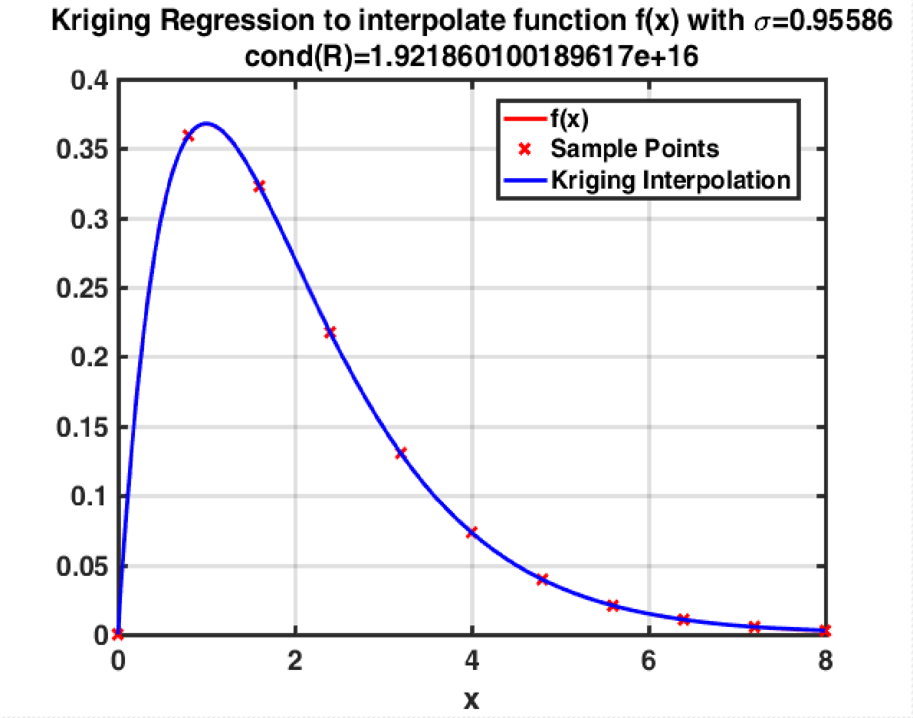
$f(x) := xe^{-x}$ $\quad \phi(x) := a + bx + cx^2$ $\quad R_\sigma(\theta, \eta) := e^{\frac{-(\theta-\eta)^2}{\sigma^2}}$ $\quad \hat{f}(x)$



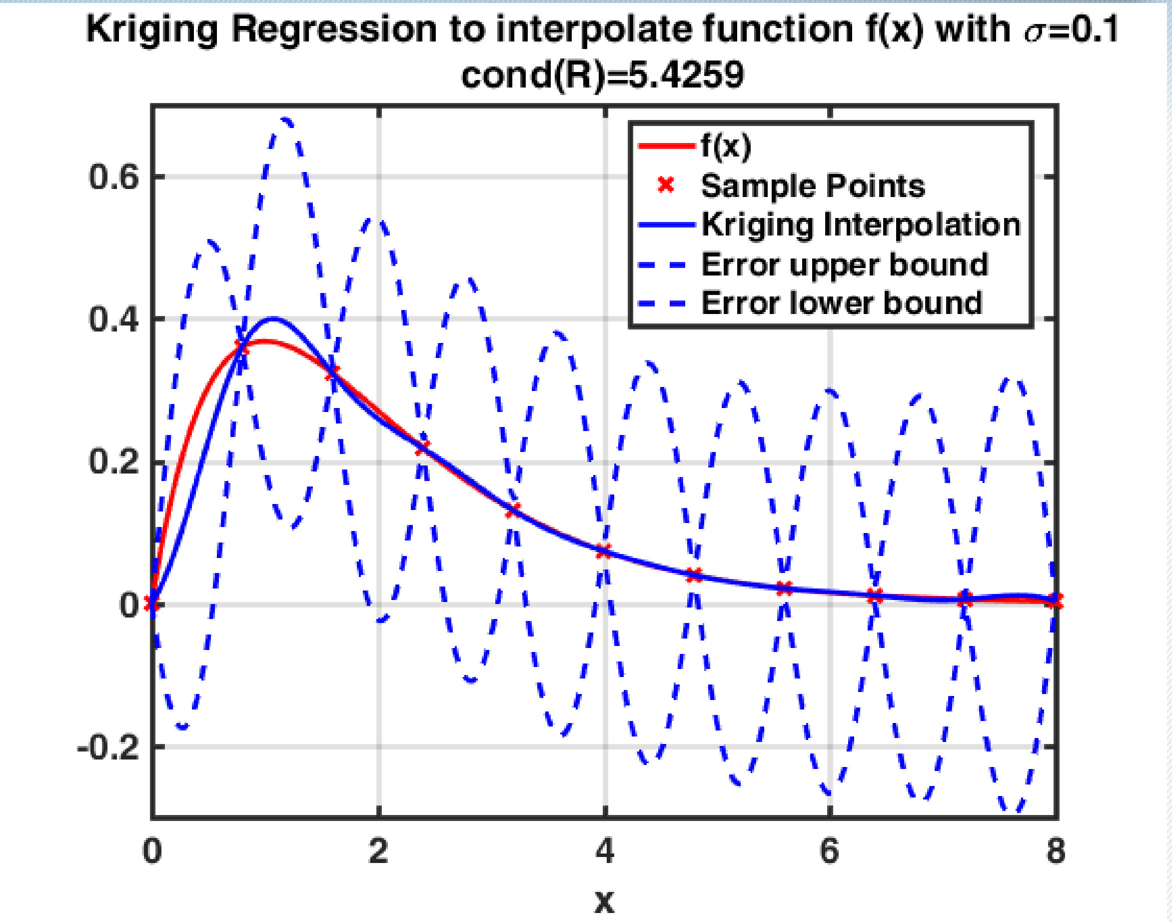


Choice of best σ using cross-validation





Bounds from variance of estimation error



Reference

[1] Stephen Boyd, Lieven Vandenberghe, 2004. *Convex Optimization*. Cambridge University Press, p. 111.