Regularity dependence of the rate of convergence of the BLUP in a noisy kriging framework

Loic Le Gratiet [†] [‡], Josselin Garnier [†]

[‡] CEA, DAM, DIF, F-91297 Arpajon, France
[†] Université Paris Diderot 75205 Paris Cedex 13

Abstract

Kriging-based approximation is a useful tool to approximate the output of complex functions given noisy observations. Our objective is to determine the rates of convergence of the Best Linear Unbiased Predictor (BLUP) when the number of observations is large in a kriging framework.

Introduction

The goal is to build a surrogate model of a function $f(x)$ given noisy observations of it. In a kriging context, we suppose that $f(x)$ is a realization of a Gaussian process $Z(x)$ with known mean and known covariance kernel $k(x,y)$. We denote the ns noisy observations $(y_i = f(x_i) + \varepsilon_i)_{i=1,\dots,ns}$ with $\varepsilon_i \sim \mathcal{N}(0, n\tau)$. The BLUP of $f(x)$ is:

$$\hat{f}(x) = k(x)^T (K + n\tau I)^{-1} y^{ns}$$

where $y^{ns} = (y_i)_{i=1,\dots,ns}$, $k(x)^T = k(x, D)$ and $K = k(D, D)$. Its Mean Squared Error (MSE) - also called kriging variance - is:

$$\sigma^2(x) = k(x,x) - k^T(x)(K + n\tau I)^{-1} k(x)$$

Theorem (Convergence of the MSE)

Let us consider $Z(x)$ a Gaussian random field with known mean and with covariance kernel $k(x,y) \in \mathcal{C}^0(Q \times Q)$, Q a compact subspace of $\mathbb{R}^d$. Let us consider $D \subset Q$ an experimental design set constituting by ns independent random points $(x_i)_{1 \le i \le ns}$ sampled with the probability measure $\mu(x)$ supported on Q. If we consider the eigenvalues $(\lambda_p)_{p \ge 0}$ sorted in decreasing order and the corresponding eigenfunctions $(\phi_p(x))_{p \ge 0}$ of the Hilbert-Schmidt's integral operator $T_{\mu,k}$:

$$(T_{\mu,k} f)(x) = \int_Q k(x,y) f(y)\, d\mu(y)$$

Then, for non-degenerate kernel, we have the following convergence in probability when $n \to \infty$:

$$\sigma^2(x) \to \sum_{p \ge 0} \left(\frac{\tau \lambda_p}{\tau + s\lambda_p} \right) \phi_p(x)^2$$

Furthermore, for degenerate kernel, i.e. with a finite number $\bar{p}$ of non zero eigenvalues, the convergence is almost sure.

Proposition (Convergence of the IMSE_μ)

With the same assumptions as in the previous theorem, for non-degenerate kernel, we have the following convergence in probability when $n \to \infty$:

$$\mathrm{IMSE}_\mu = \int_Q \sigma^2(x)\, d\mu(x) \to \sum_{p \ge 0} \left(\frac{\tau \lambda_p}{\tau + s\lambda_p} \right)$$

Furthermore, for degenerate kernel, the convergence is almost sure.

Applications: rates of convergence

- For degenerate kernels the IMSE_μ decreases as s^{-1}.
- For a fractional Brownian kernel (FBk) with Hurst parameter H, we have $\lambda_p \sim p^{-(2H+1)}$, when $p \gg 1$ [Bronski (2003)]. Therefore, the IMSE_μ decreases as $s^{\frac{1}{2H+1}-1}$.
- For a d-D Gaussian kernel (Gk), we have $\lambda_p \lesssim \exp\left(-p^{\frac{1}{d}}\right)$, when $p \gg 1$. Therefore, the IMSE_μ decay is bounded by $s^{-1}\log^d(s)$.
- For a d-D tensorised Matèrn kernel (Mk) with regularity parameter ν, we have $\lambda_p \sim p^{-2\nu}\log(1+p)^{2(d-1)\nu}$, when $p \gg 1$ [Pusev (2011)]. Therefore, the IMSE_μ decreases as $s^{\frac{1}{2\nu}-1}\log^{d-1}(s)$.

Important remark

Classical results about Monte-Carlo convergence give that the variance decay as s^{-1} whatever the dimension. Nevertheless, for non-degenerate kernels we are in infinite dimension. We observe that in this case the convergence is slower than s^{-1}. Furthermore, for degenerate kernel, we are in finite dimension and the IMSE decay as s^{-1} (*i.e.* the classical Monte-Carlo convergence).

Numerical illustrations

We illustrate here the IMSE_μ convergence for different models. We consider $ns = 200, 400, \dots, 2000$ and $n\tau = 1$.

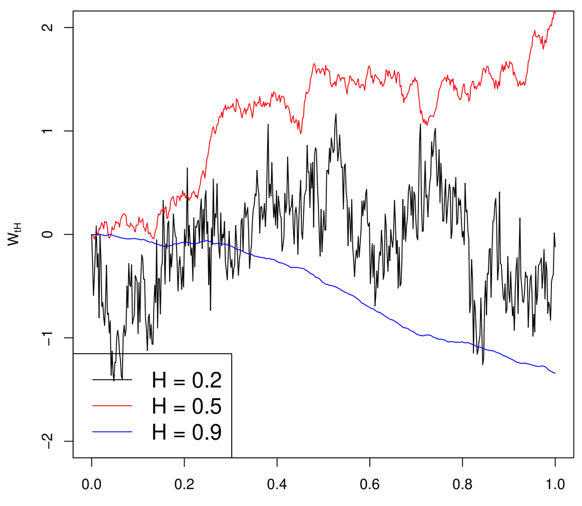

Figure: Realizations of fractional Brownian motions.

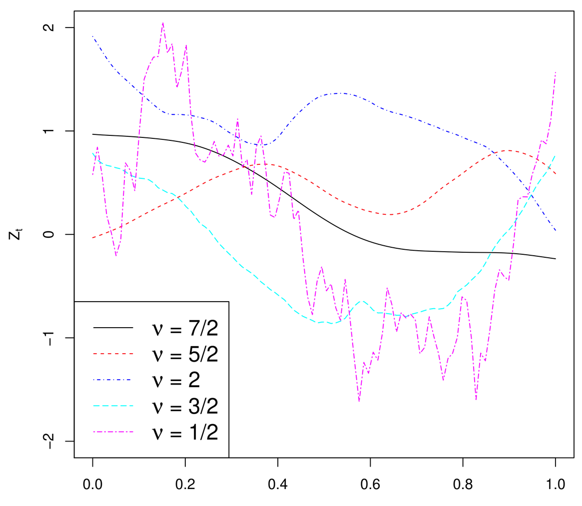

Figure: Realizations of Gaussian processes with Matèrn-ν kernels.

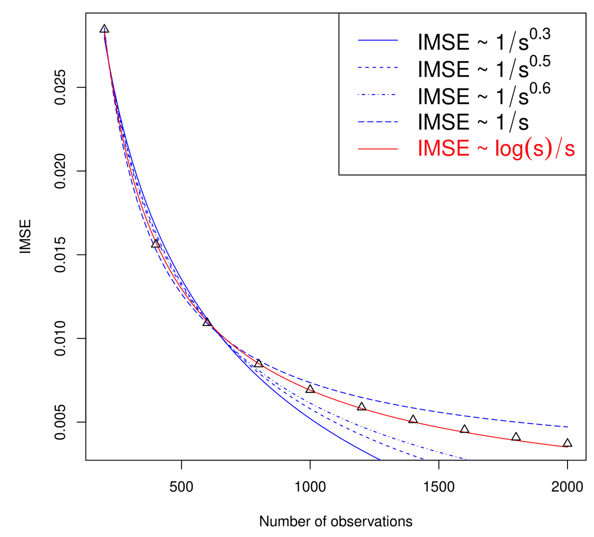

Figure: Convergence rate for 1-D Gk.

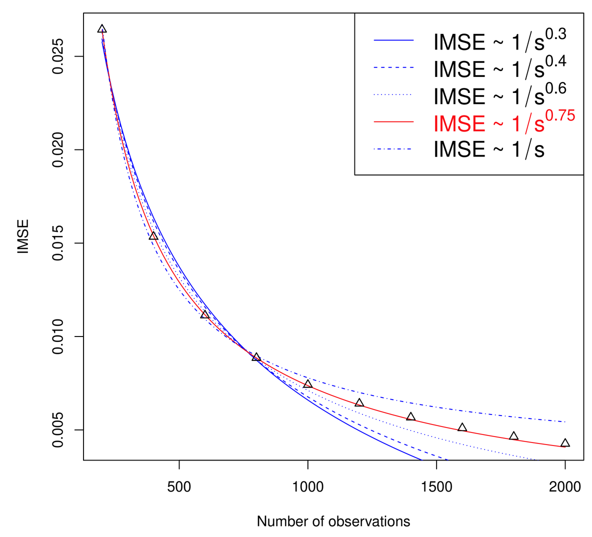

Figure: Convergence rate for 1-D Mk with $\nu = 2$.

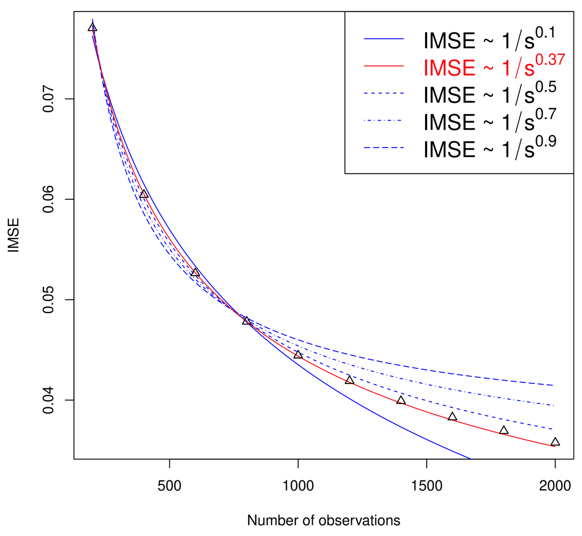

Figure: Convergence rate for FBk with H=0.3.

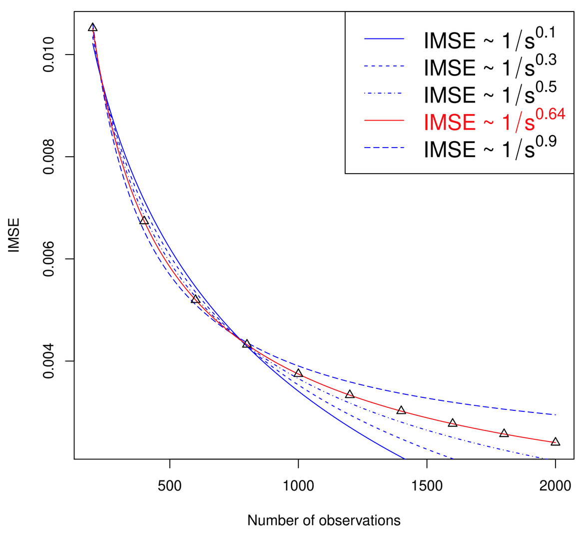

Figure: Convergence rate for FBk with H=0.9.

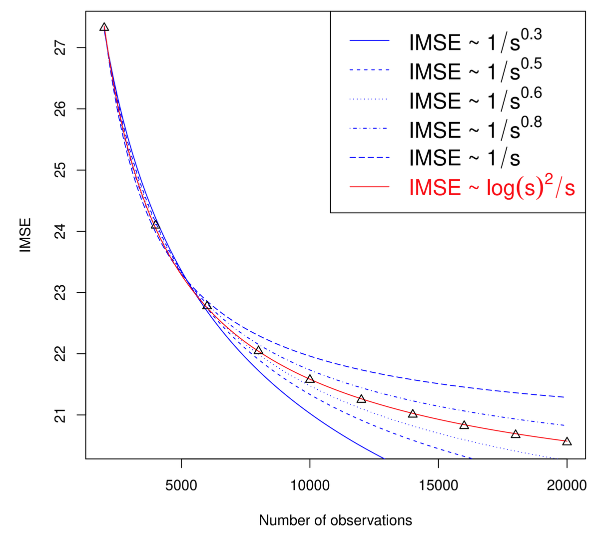

Figure: Convergence rate for 2-D Gk.

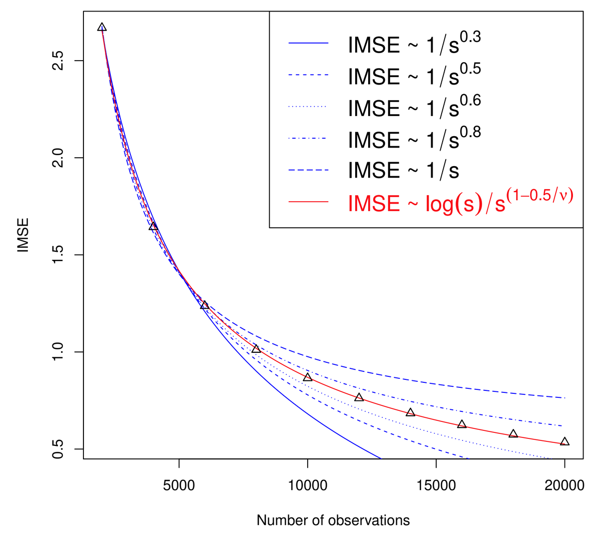

Figure: Convergence rate for 2-D Mk with $\nu = \frac{5}{2}$.

Key steps for the Proof of the Theorem

According to the Mercer theorem, $k(x,y)$ can be written as :

$$k(x,y) = \sum_{p \ge 0} \lambda_p \phi_p(x) \phi_p(y)$$

1. The degenerate case

For a degenerate kernel, the number $\bar{p}$ of non zero eigenvalues is finite. If we denote $\Lambda = \mathrm{diag}(\lambda_i)_{1 \le i \le \bar{p}}$ and :

$$\Phi(X) = \begin{pmatrix} \phi(x_1) \\ \vdots \\ \phi(x_{ns}) \end{pmatrix} \qquad \phi(x) = (\phi_1(x), \dots, \phi_{\bar{p}}(x))$$

we have the following expression for $\sigma^2(x)$:

$$\sigma^2(x) = \phi(x) \left(\frac{\Phi(X)^T \Phi(X)}{n\tau} + \Lambda^{-1} \right) \phi(x)^T$$

The points x_i being independent and identically distributed according to the measure $\mu(x)$, we then have by the strong law of large numbers $\frac{1}{n}\sum_{i=1}^{ns} \phi_p(x_i)\phi_{p'}(x_i) \to s\delta_{p=p'}$ when $n \to \infty$. Therefore, when $n \to \infty$:

$$\sigma^2(x) \to \sum_{p \le \bar{p}} \left(\frac{\tau \lambda_p}{\tau + s\lambda_p} \right) \phi_p(x)^2$$

This proof is presented in [Rasmussen et al. (2006)], [Picheny (2009)] and [Opper et al. (1999)]. A proof in 1-D for non-degenerate kernel is given in [Ritter (1996)] but cannot be extended to higher dimension.

2. Upper bound for $\sigma^2(x)$

If we denote $\sigma^2_{LUP}(x)$ the MSE of a Linear Unbiased Predictor (LUP) and $\sigma^2(x)$ the MSE of the BLUP, we have:

$$\sigma^2(x) \le \sigma^2_{LUP}(x)$$

The idea is to find a LUP so that its MSE is a tight upper bound of $\sigma^2(x)$. We take the LUP $k(x)^T A y^{ns}$ with A the $ns \times ns$ matrix:

$$A = L^{-1} + \sum_{k=1}^{q} (-1)^k (L^{-1}M)^k L^{-1}$$

with q a finite integer, L and M defined by:

$$L = n\tau I + \sum_{p < p*} \lambda_p [\phi_p(x_i)\phi_p(x_j)]_{1 \le i,j \le ns}$$

$$M = \sum_{p > p*} \lambda_p [\phi_p(x_i)\phi_p(x_j)]_{1 \le i,j \le ns}$$

and $p*$ such that $s\lambda_{p*} < \tau$. We hence have :

$$\sigma^2_{LUP}(x) = k(x,x) - k(x)^T L^{-1} k(x) - \sum_{i=1}^{2q+1} (-1)^i k(x)^T (L^{-1}M)^i L^{-1} k(x)$$

Using the Woodbury formula, the strong law of large numbers and the continuity of the inverse operator, we obtain the following almost sure convergence when $n \to \infty$:

$$k(x)^T L^{-1} k(x) \to \sum_{p < p*} \frac{s\lambda_p^2}{s\lambda_p + \tau} \phi_p(x)^2 + \frac{s}{\tau} \sum_{p > p*} \lambda_p^2 \phi_p(x)^2$$

Note that we can use the strong law of large numbers since $p*$ is finite and independent of n. Then, using the Markov inequality and the equality $\sum_{p \ge 0} \lambda_p \phi_p(x)^2 = \sigma^2$, we obtain the following convergence in probability:

$$k(x)^T \left(L^{-1}M\right)^i L^{-1} k(x) \to \left(\frac{s}{\tau}\right)^{i+1} \sum_{p > p*} \lambda_p^{i+2} \phi_p(x)^2$$

Note that we cannot use the strong law of large numbers because of the infinite sum in M. Finally, by considering the asymptotic $q \to \infty$ and the inequality $s\lambda_{p*} < \tau$, we obtain the following convergence in probability:

$$\limsup_{n \to \infty} \sigma^2(x) \le \sum_{p \ge 0} \left(\frac{\tau \lambda_p}{\tau + s\lambda_p} \right) \phi_p(x)^2$$

3. Lower bound for $\sigma^2(x)$

Let us consider the Karhunen-Loève decomposition of $Z(x)$:

$$Z(x) = \sum_{p \ge 0} Z_p \sqrt{\lambda_p} \phi_p(x)$$

If we denote $a_i(x)$ the coefficient of the BLUP associate to $Z(x)$, we have:

$$\sigma^2(x) = \sum_{p \ge 0} \lambda_p \left(\phi_p(x) - \sum_{i=1}^{n} a_i(x) \phi_p(x_i) \right)^2$$

For a fixed $\bar{p}$, the following inequality holds:

$$\sigma^2(x) \ge \mathbb{E}\left[\left(\tilde{Z}(x) - \sum_{i=1}^{n} a_i(x) \tilde{Z}(x_i) \right)^2 \right]$$

with $\tilde{Z}(x) = \sum_{p \le \bar{p}} Z_p \sqrt{\lambda_p} \phi_p(x)$. A fortiori, denoting by $\tilde{\sigma}^2(x)$ the MSE of the BLUP of $\tilde{Z}(x)$, we have $\sigma^2(x) \ge \tilde{\sigma}^2(x)$. Since $\tilde{Z}(x)$ has degenerate kernel, $\forall \bar{p} > 0$ we know that the MSE of its BLUP converges almost surely as $n \to \infty$. By considering the limit $\bar{p} \to \infty$, we obtain:

$$\liminf_{n \to \infty} \sigma^2(x) \ge \sum_{p \ge 0} \left(\frac{\tau \lambda_p}{\tau + s\lambda_p} \right) \phi_p(x)^2$$

□

Bibliography

- Opper, M. & Vivarelli, F., 1999, *General Bounds on Bayes Errors for Regression with Gaussian Processes*. Advances in Neural Information Processing Systems 11, 302-308. MIT Press.
- Picheny, V., 2009, *Improving Accuracy and Compensating for uncertainty in Surrogate Modeling*. Ecole Nationale Supérieure des Mines de Saint Etienne, 2009.
- Pusev R.S., 2011, *Small Deviation Asymptotics for Matèrn Processes and Fields under Weighted Quadratic Norm*. Theory Probab. Appl. **55**, No. 1, 164-172.
- Rasmussen, C. E. & Williams, C. K. I., 2006, *Gaussian Processes for Machine Learning*. the MIT Press.
- Ritter, K., 1996, *Almost Optimal Differentiation Using Noisy Data*. Journal of approximation theory 86, 293-309.
- Bronski, J.C, 2003, *Asymptotics of Karhunen-Loeve Eigenvalues and Tight Constants for Probability Distributions of Passive Scalar Transport*. Commun. Math. Phys. **238**, 563-582.