Kernel Packet: An Exact and Scalable Algorithm for Gaussian Process Regression with Matérn Correlations

Haoyuan Chen, Liang Ding and Rui Tuo

Department of Industrial & Systems Engineering, Texas A&M University

Motivation

- Direct inverting the covariance matrix $\mathbf{K}$ is time consuming, because $\mathbf{K}$ is a dense matrix. The main idea of this work is to find an *exact* representation of $\mathbf{K}$ in terms of sparse matrices.
- Suppose K is a one-dimensional kernel. Consider the linear space $\mathcal{K} = \operatorname{span}\{K(\cdot, x_j)\}_{j=1}^n$. The goal is to find another basis for $\mathcal{K}$, denoted as $\{\phi_j\}_{j=1}^n$, satisfying the following properties:
  1. Almost all of the ϕ_j's have *compact supports*.
  2. $\{\phi_j\}_{j=1}^n$ can be obtained from $\{K(\cdot, x_j)\}_{j=1}^n$ via a *sparse linear transformation*, i.e., the matrix defining the linear transform from $\{K(\cdot, x_j)\}_{j=1}^n$ to $\{\phi_j\}_{j=1}^n$ is sparse.

  We assume that the one-dimensional kernel K is a Matérn correlation function, whose spectral density is proportional to $(2\nu/\omega^2 + x^2)^{-(\nu+1/2)}$.

Kernel Packet

1. Given a one-dimensional covariance function K and input points $a_1 < \cdots < a_k$, a non-zero function ϕ is called a *kernel packet (KP) of degree k*, if it admits the representation $\phi(x) = \sum_{j=1}^k A_j K(x, a_j)$, and the support of ϕ is $[a_1, a_k]$.
2. Let $x_1 < \cdots < x_n$ be the input data, and K a Matérn correlation function with a half-integer smoothness. Suppose $n \geq k$. We can construct n functions $\{\phi_j\}_{j=1}^n$, as a subset of $\mathcal{K}$, which form a basis for $\mathcal{K}$, referred to as the *KP basis*.
3. Let K be a one-dimensional Matérn kernel. Our study shows that, under certain conditions, $\mathbf{K}$ admits the following simple factorization:

$$\mathbf{K} = \mathbf{\Phi}\mathbf{A}^{-1}, \qquad (1)$$

where both $\mathbf{\Phi}$ and $\mathbf{A}$ are *banded matrices*, and their bandwidths depend only on the smoothness of K. This factorization is called the **Kernel Packet (KP)** factorization.

Main Theory

Let K be a Matérn correlation with smoothness ν. If ν is a half integer, then K admits a KP with degree $2\nu + 2$. In addition, given $a_1 < \cdots < a_k$, function $\phi_{\mathbf{a}}$ with the form $\phi_{\mathbf{a}}(x) := \sum_{j=1}^k A_j K(x, a_j)$ is a KP if and only if the coefficients A_j's are given by a non-zero solution to $\sum_{j=1}^k A_j a_j^l \exp\{\delta c a_j\} = 0$, with $l = 0, \ldots, (k-3)/2$ and $\delta = \pm 1$.

Figure 1 illustrates that the linear combination of 5 components $\{K(\cdot, a_j)\}_{j=1}^5$ provides a compactly supported KP corresponding to Matérn-3/2 correlation function. Further, it is straightforward to check that, given any $x \in \mathbb{R}$, the vector $\boldsymbol{\phi}(x) = (\phi_1(x), \ldots, \phi_n(x))^T$ has at most $k-1$ non-zero entries. As a result, we can construct a basis for $\mathcal{K}$ satisfying the sparse properties. Figure 2 illustrates a KP basis corresponding to Matérn-3/2 and Matérn-5/2 correlation function with input points $\mathbf{X} = \{0.1, 0.2, \ldots, 1\}$.

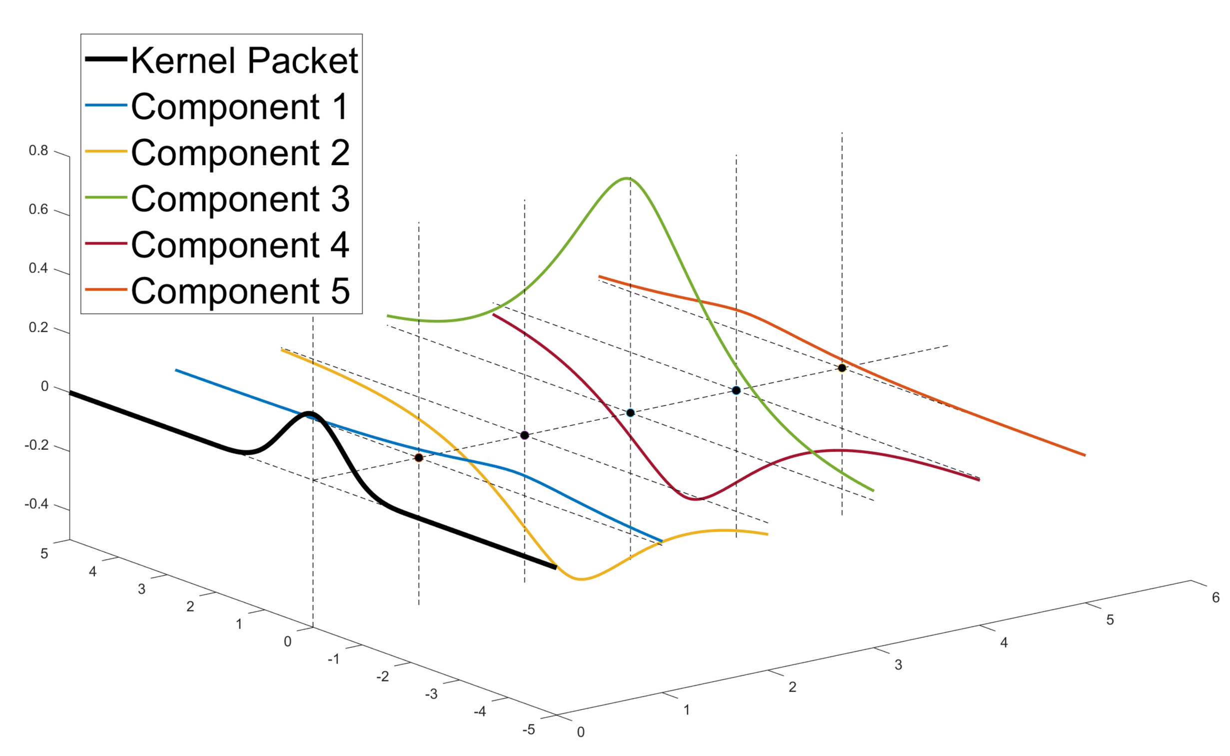

Figure 1: The addition of $A_j K(\cdot, x_j)$ (colored lines, without compact supports) leads to a KP (black line, with a compact support).

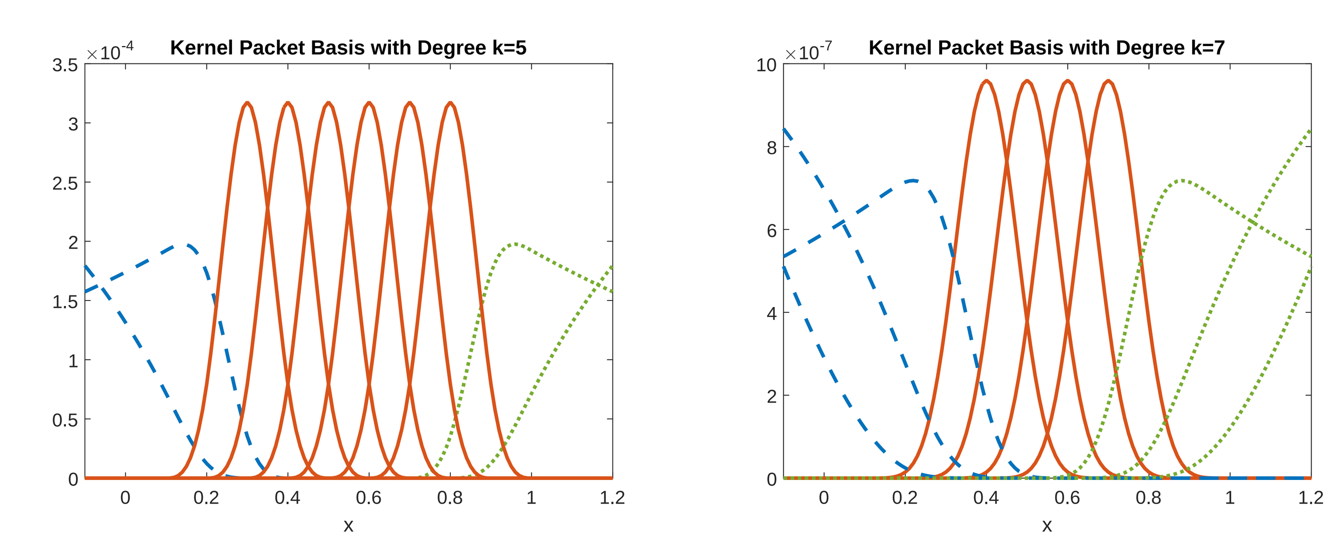

Figure 2: KP basis functions corresponding to Matérn-3/2 (left) and Matérn-5/2 (right) correlation function with input points $\mathbf{X} = \{0.1, 0.2, \ldots, 1\}$. The KPs, left-sided KPs, and the right-sided KPs are plotted in orange, blue, and green lines, respectively.

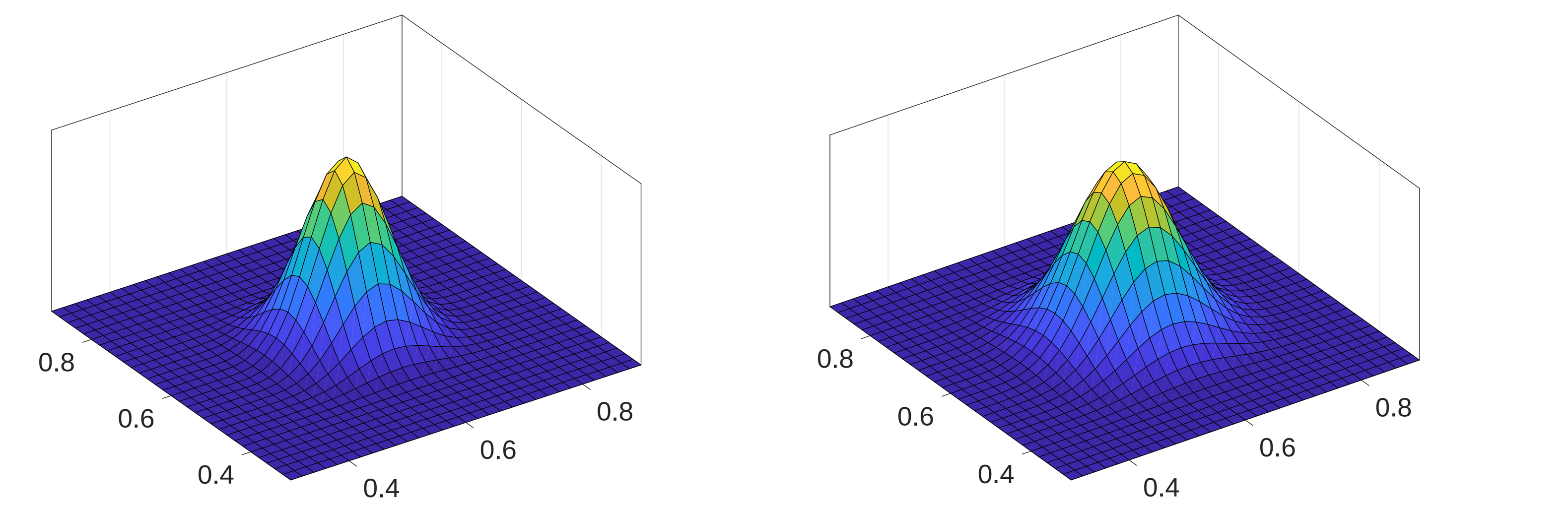

Figure 3: product KP basis functions $\phi_{(.4\ .5\ .6\ .7\ .8)}(x_1)\phi_{(.4\ .5\ .6\ .7\ .8)}(x_2)$ corresponding to Matérn-3/2 (left) and $\phi_{(.3\ .4\ .5\ .6\ .7\ .8\ .9)}(x_1)\phi_{((.3\ .4\ .5\ .6\ .7\ .8\ .9)}(x_2)$ corresponding to Matérn-5/2 (right) correlation function.

Numerical Experiments

We test our algorithm on the following deterministic function:

$$f(\mathbf{x}) = \sin(12\pi x_1) + \sin(12\pi x_2), \quad \mathbf{x} \in (0,1)^2.$$

Samples of f are collected from a level-η full grid design: $\mathbf{X}_\eta^{\mathsf{FG}} = \times_{j=1}^2 \{2^{-\eta}, 2 \cdot 2^{-\eta}, \ldots, 1 - 2^{-\eta}\}$ with $\eta = 5, 6, \cdots, 13$. We sample 1000 i.i.d. test points uniformly from $(0,1)^2$ for each experimental trial. Figure 4 compares the MSE and the computational time of all algorithms, both under logarithmic scales, for sample sizes 2^{2j}, $j = 5, 6, \ldots 13$.

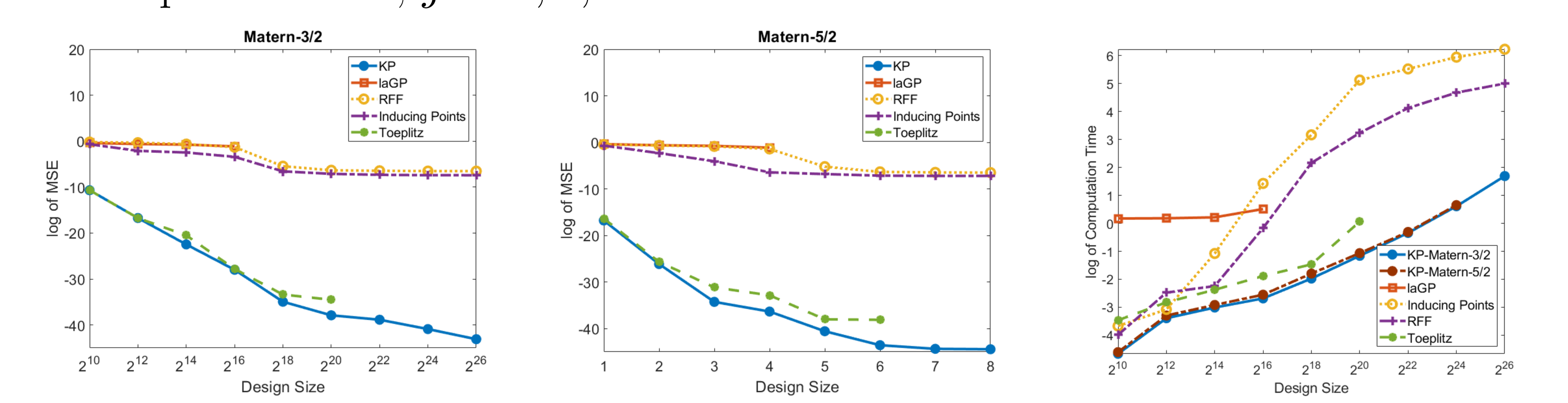

Figure 4: Logarithm of MSE for predictions with Matérn-3/2 correlation function (left) and Matérn-5/2 correlation function (middle) and logarithm of averaged computational time (right). The laGP uses the Gaussian covariance family in both the left and the middle figure. No results are shown for the cases when a runtime error occurs or the prediction error ceases to improve.

Conclusions

In this work, we propose a rapid and exact algorithm for one-dimensional Gaussian process regression under Matérn correlations with half-integer smoothness. The proposed algorithm only requires $\mathcal{O}(\nu^3 n)$ operations and $\mathcal{O}(\nu n)$ storage.