Wavelet Decomposition of a gaussian process kernel

Romain RAVAILLE
r.ravaille@univ-st-etienne.fr

Introduction

Nicolas Durrande used an approach based on Gaussian process Regression to analyse data over periodic phenomena by doing a decomposition of a covariance function in two parts, one periodic and the other aperiodic.
As wavelets are the extention of Fourier in the pseudo periodic case, our goal is to take an approach similar to the one Nicolas Durrande took, but with wavelets Riesz basis instead of Fourier basis.

Preliminary work

Considering the wavelet workspace introduced in [1], we can

- Create a Riesz basis by dilatation and translation of a wavelet ψ. We will use the notation: $\psi_{a,b}(x) = \psi((x-b)/a)$.
- Decompose our workspace V_0 in two: $V_0 = V_1 \oplus W_1$ where W_1 is the space from the Riesz basis for a given scaling parameter and V_1 is a residual space.
- Obtain a space $\mathcal{H}_w = \bigoplus_{j=1..n} W_j$ that has all the Riesz basis information by proceeding as above iteratively.

We need to choose our wavelet ψ with respect to some conditions given by [2]:

1. $\mathcal{H}_w$ contains informations on the frequencies we seek to analyse
2. $\mathcal{H}_w$ is a subset of $\mathcal{H}$.

We work on Matern kernels as their RKHS is easy to deal with (a $\text{Matern}_{n/2}$ creates a sobolev space $\mathcal{W}^{n,2}$). We know that $\psi \in \mathbf{L}^2(\mathbb{R})$, $\mathbf{L}^2(\mathbb{R}) = \mathcal{W}^{0,2}$ and that for $0 \leq n_1 \leq n_2$,

$$\mathcal{W}^{n_2,2} \subset \mathcal{W}^{n_1,2} \subset \mathcal{W}^{0,2}. \quad (1)$$

We choose to use the spline wavelets (we will refer to the function as Ψ^m where $m-1$ is the order of the polynomials in Ψ), as it is straightforward to prove that the spline wavelet of order m belongs to $\mathcal{W}^{(m-1),2}$.

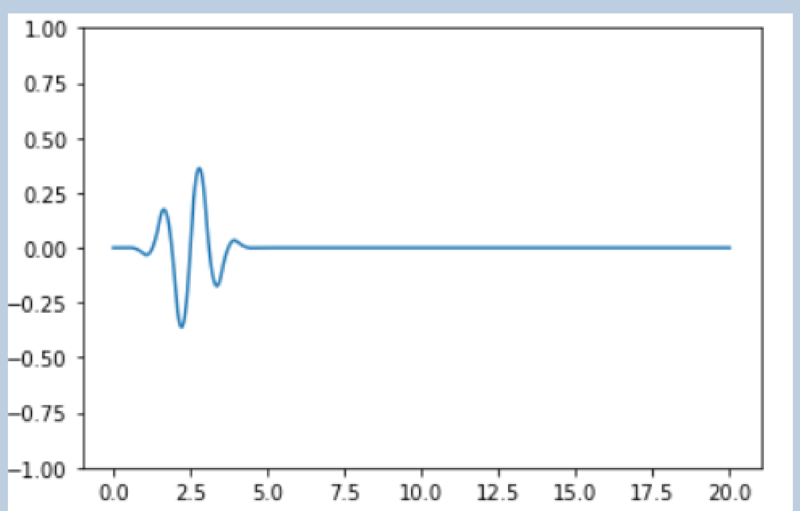

Figure 1: Image of $x \mapsto \Psi^3_{a,b}(x) = \Psi^3((x-b)/a)$ for $a = 1$ and $b = 0$

Algorithm Explanation

Steps explanation of our algorithm

1. Chose B_Ψ: *We choose a set of parameter* $a = \{a_1, a_2 ... a_n\}$ *and* N *to create a wavelet Riesz basis* $B_\Psi = \{\Psi^m_{a,b}\}_{a=a_1...a_n, b=1..N}$ *which enables us to make a time-frequency analysis of the signal. We need to check the correspondance between the chosen wavelet frequency domain and the frequencies present in the signal.*

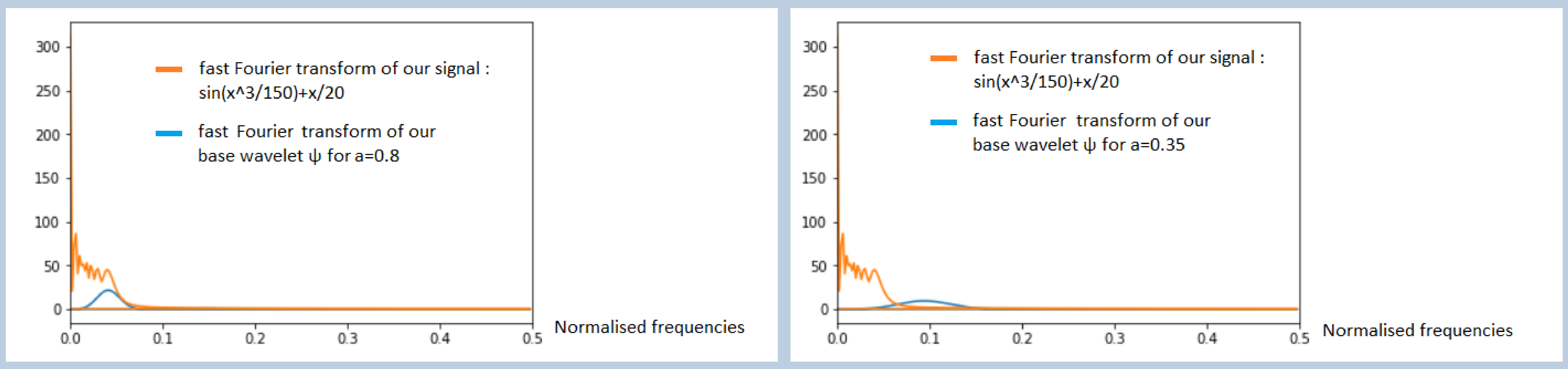

Figure 2: Frequency domain of our test signal and the wavelet $\Psi^3_{a,b}$ for a well chosen a (left) and a bad one (right).

2. Compute $K_w = B^T_\Psi(X) G^{-1} B_\Psi(X)$ and K_r: *we use the formulas provided in [2] to compute* G, *the Gram matrix of the functions in* B_Ψ, K_r *is the residual kernel.*
3. Define $K = K_w + K_r$: *We think the global kernel as the sum of both of the previously defined kernel. We do not assume that both subkernels are a decomposition of* K.
4. Optimize the hyperparameters of K: *We use the already implemented method in the python library GPy to find the max likelihood. The loglikelihood for our gaussian process* f *given points of data at the coordinate* X *will be written as* $\mathcal{L}_K = \mathcal{L}(\sigma_w, \sigma_r, l_w, l_r, X)$, *and is defined by:*

$$\begin{aligned}\mathcal{L}_K &= \log(p(f(x)|x, \sigma_w, \sigma_r, l_w, l_r)) \\ &= -\frac{1}{2} f(X)^T K^{-1} f(X) - \frac{1}{2}\log(\det(K)) - \frac{|X|}{2}\log(2\pi) \\ &= -\frac{1}{2} f(X)^T (B^T_\Psi G^{-1} B_\Psi + K_r)^{-1} f(X) - \frac{1}{2}\log(\det(B^T_\Psi G^{-1} B_\Psi + K_r)) - \frac{|X|}{2}\log(2\pi).\end{aligned}$$

Results

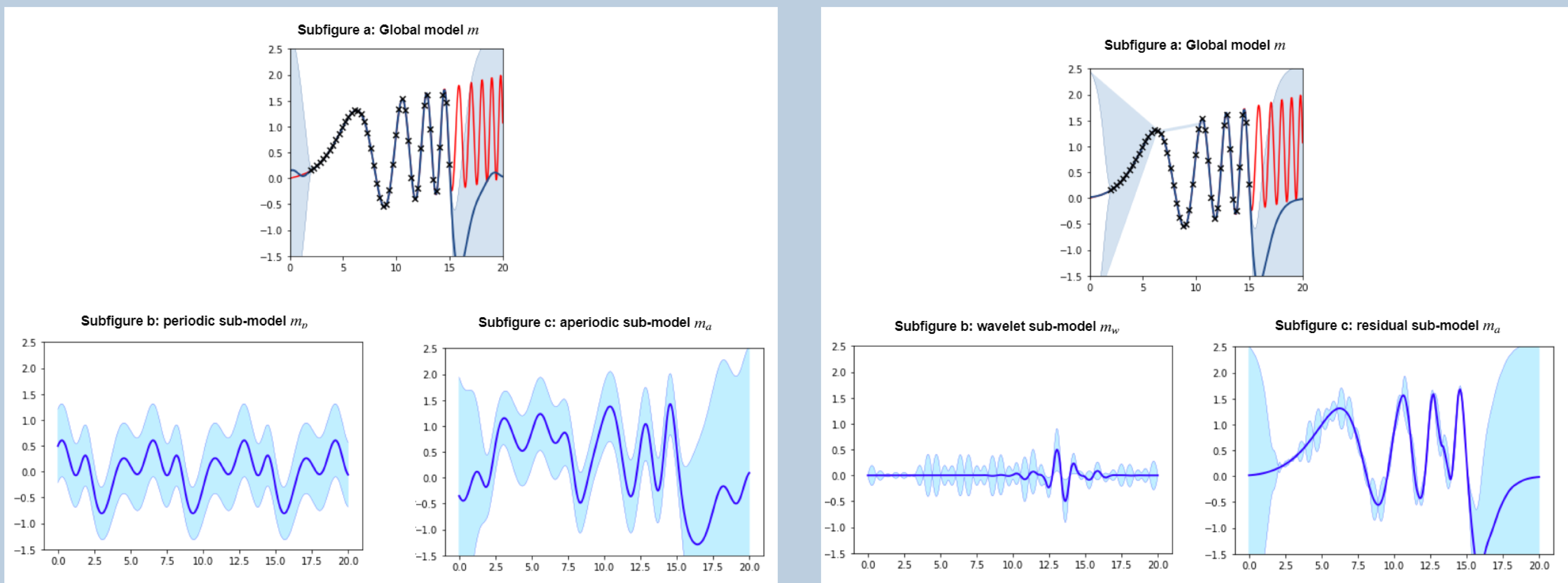

Figure 3: Comparison of Durrande algorithm response to a chirp (left) and my results (right)

As of now, restrictions on the time (and memory) it takes to compute G^{-1} only enable us the use of one a and the wavelet Riesz basis only has 10 members.
By interpreting the right part of Fig. 3, there is an interesting component for our chosen a somewhere on the time axis, but the method still can not separate the whole chirp from the linear function as only one scaling of our wavelet is not enough. Our Wavelet kernel is leaving the information he does not have in the residual kernel.

Conclusion

- Absence of a computation-efficient method to obtain G. We can not say that our method provides a solution to model a chirp right now.
- Some encouraging results: able to separate the information we have in our wavelet Riesz basis from the rest.

References

[1] Stephane Mallat. (1998) a wavelet tour of signal processing, academic press.

[2] Nicolas Durrande, James Hensman, Magnus Rattray, and Neil D. Lawrence. Gaussian process models for periodicity detection, (2016) peerj computer science 2:e50.

Acknowledgements

I would like to express my special thanks to Xavier Bay, Laurence Grammont and the APSSE group for their work and help that was of great assistance in the understanding of these subjects. Moreover, this work wouldn't be nearly as advanced as it is without the help of Olivier Alata and Julian Tugaut, I thank them grealy for their advices.