Notes: Empirical Mode Decomposition & Gaussian Processes

*

Version: June 5, 2019

Abstract. Extension 1, Estimation: Treat each IMF as a separate Gaussian process and then represent the signal using multi-kernel representation of the Gaussian Process.
Extension 2, Forecasting: GP representation does not ensures itself that the predicted function from a given Gaussian process is IMF , that is, it satisfies (I1)-(I2). Therefore, we explore the formulation of IMFs as an analogue of Brownian Bridge.

Contents

1. Gaussian Processes and EMD: IMFs as Gaussian Processes with non-stationary kernels

We treat each IMF as a separate Gaussian process and then represent the signal using multi-kernel representation of the Gaussian Process.

1.1. The Stochastic Representation by Gaussian Processes

Let $s(t)$ for $t \in [0, \infty]$ be a continuous signal which is observed on discrete grid of points in the interval $[0, T]$. We consider not so uncommon situation, when the observed values of $s(t)$ are perturbed by some noise component. The perturbation of the true signal can be either deterministic or stochastic. The deterministic noise might be a results of a chaotic

and therefore the process which we observe is in fact

$$y(t) = s(t) + \epsilon, \text{ for } \epsilon \in \mathcal{N}(0, \sigma^2). \tag{1}$$

Therefore, our observation set consist of the pairs $\{t_n, y_n\}$ where $y_n = y(t_n)$ for $t_n \in [0, T]$. Let us remark that for the same

the realisations of the continuous signal $s(t)$ on the discrete sent of points. W

Let $x(t)$ be a continuous real-valued signal which is an approximation of the signal $s(t)$ based on its observed values. Therefore, $x(t)$ is

and let us define its EMD decomposition into K intrinsic mode functions (IMFs) given by

$$x(t) = \sum_{m=1}^{M} c_m(t) + r_m(t) = \sum_{m=1}^{M} \mathrm{Re}\Big\{ A_m(t) e^{i\theta_m(t)} \Big\} + r_M(t). \tag{2}$$

where r_k is a tendency which does not have much of oscillation and therefore characterize the low frequency tend of $x(t)$. In order to reconstruct the stochastic representation of $x(t)$ given by EMD decomposition we assume that each of the IMFs functions, $c_m(t)$, is a Gaussian process

$$c_k(t) \sim \mathcal{GP}\Big(0, k_m(t, t')\Big), \tag{3}$$

where $k_m(t, t')$ is a positive definite covariance kernel which is parametrized by a set of parameters Ψ_m. The component $r_M(t)$ can be modelled as a Gaussian process itself or one can assume that $x(t)$ is Gaussian Process conditioned on $r_M(t)$, that is

$$x(t)|r_M(t) \sim \mathcal{GP}\Big(r_M(t), k(t, t')\Big). \tag{4}$$

These two approaches provide an unconditional and conditional stochastic representation of $x(t)$, respectively, and determine two different estimators of the out-of-sample forecast for $x(t)$. The later is a more convenient assumption to preserve the monotonicity of the $r_M(t)$ which is a desired property of a residual function in the decomposition in Equation (2). To ensure the function $r_M(t)$ to have only single convexity change, $r_M(t)$ might be extrapolated by a power law which stays monotonic (ie a polynomial up to the second order). Then, the out-of-sample forecast of $x(t)$ would be conditioned on the extrapolation of $r_M(t)$. In order to preserve the monotonicity property of the tendency function $r_M(t)$ in the out-of-sample prediction, the extrapolation from a low order spline representation of $r_M(t)$, which is deterministic, is excepted to behaves better than the forecast from a Gaussian Process since the later would most plausibly wiggle around a trend and, consequently, would loose the monotonicity of $r_M(t)$. In the following work we would like to guarantee the out-of-sample monotonicity of $r_M(t)$ obtained by construction in the in-ample set, and therefore, we chose to work with the conditional representation of $x(t)$ given in Equation (4). TODO: derive the properties of these two estimators..

1.1.1. Denoising Observed Values of the True Signal

We consider the following experiment setup. Let J represents the number of trials in our experiment where we collect the realisations of $y(t)$ on the discrete subsets of the interval $\in [0, T]$, which can be specified by **random or deterministic sub-sampling**. We assume that each trial has a set of N^i samples and N^i varies over trials for $i = 1, \ldots, J$. Let $\mathbf{y}^i$ and $\mathbf{t}^i$ denote N^i-dimensional vectors which represent the N^i observed values in the ith trial and the N_i corresponding time points being a subsample of $[0, T]$, respectively. Given that, $\mathbf{y}^i := y(\mathbf{t}^i) = [y(t^i_1), \ldots, y(t^i_{N^i})]$. **We remark that it is not ensured that for the same time point $t_0 \in [0, T]$, a value of $y(t_0)$ in trial i_1 and a value of $y(t_0)$ in trial i_2 are equal since the definition of $y(t)$ in Equation (1) includes the error term component.**

The sets of the time points for each trial, $\mathbf{t}^i$, can be specified deterministic or be a realisations of the random variable. Regardless of the assumption on the sampling mechanism, the time points collected in the set $\mathbf{t}^i$ can be missing. Therefore, we may distinguish the complete and incomplete cases for the sampling times $\mathbf{t}^i$ and the deterministic or random sampling framework. In the following section we will consider the simplest case, when the elements of $\mathbf{t}^i$ are obtained deterministically and are not missing. Set up notation and cases for the frameworks which we will consider later for subsampling

In order to specify the distribution of each $c_k(t)$, we collect M paths of $x(t)$. Therefore, we have M collections of of points $\mathbf{t}^{(1)}, \ldots, \mathbf{t}^{(M)}$, each N_i dimensional for $i \in \Big\{1, \ldots, M\Big\}$ and by by $\mathbf{x}^{(i)}$ we denote the values

of $x(t)$ collected in the trail i on the points $t \in \mathbf{t}^{(i)}$. The sets $\mathbf{t}^{(i)}$ can be the same. The EMD decomposition on each of the M replications $x^{(i)}$ gives the following representations

$$\mathbf{x}^{(i)} = \sum_{k=1}^{K} \mathbf{c}_k^{(i)} + \mathbf{r}_K^{(i)} \tag{5}$$

where $\mathbf{c}_k^{(i)}$ is an N_i dimensional vector which represents the observed values of the function $c_k(t)$ at the arguments in $\mathbf{t}^{(i)}$. The same logic applies to the definition of vectors $\mathbf{r}_K^{(i)}$. The vector $\boldsymbol{\mu}_k^{(i)}$ corresponds to the values of the functions $\mu_k(t)$ at the arguments in $\mathbf{t}^{(i)}$, that is, $\boldsymbol{\mu}_k^{(i)} = \mu_k(\mathbf{t}^{(\mathbf{i})})$.

1.1.2. Gaussian Process of IMFs given Splines Formulation of $x(t)$

REMARK: Each element of $\mathbf{c}_k^{(i)}$ or $\mathbf{x}^{(i)}$, that is, some $c_{k,j}^{(i)}$ or $x_j^{(i)}$ is a combination of a spline coefficients fitted for the batch i and being an output of a function $c_k^{(i)}(t)$ and $x^{(i)}(t)$ (the fitted spline in trial i) to the argument point $t_j^{(i)}$.

For the EMD to exist, the underlying signal $x(t)$ needs to be approximated. Such approximation is covered in this work by a natural cubic spline. Each $c_{k,j}^{(i)}$ is also a natural cubic spline, defined as $c_{k,j}^{(i)}(t)$ with the following formulation:

$$c_{k,j}^{(i)}(t) = \begin{cases} s_1(t) = a_1 t^3 + b_1 t^2 + c_1 t + d_1 & \text{for} \quad t \in (t_1^i, t_2^i) \\ s_2(t) = a_2 t^3 + b_2 t^2 + c_2 t + d_2 & \text{for} \quad t \in (t_2^i, t_3^i) \\ \cdots \quad \cdots \quad \cdots \quad \cdots \quad \cdots & \\ s_{N_i}(t) = a_{N_i} t^3 + b_{N_i} t^2 + c_{N_i} t + d_{N_i} & \text{for} \quad t \in (t_{N_i-1}^i, t_{N_i}^i) \end{cases}$$

A shorter version of the above system of equations can be given by:

$$c_{k,j}^{(i)}(t) = \sum_{j=1}^{N_i} \left(a_j t^3 + b_j t^2 + c_j t + d_j \right) \mathbb{1}\left(t \in (t_{j-1}^i, t_j^i) \right) = \sum_{j=1}^{N_i} s_j(t) \mathbb{1}\left(t \in (t_{j-1}^i, t_j^i) \right) \tag{6}$$

where $\mathbb{1}$ represents the indicator function.

Let us start with having observed partially or discretely in time a continuous speech signal through a sample recording. The signal $s(t)$ is observed at $t = (t_1 < \cdots < t_N) = \{t_i\}_{i=1:N}$, where the subscripts represent the sampling index times. For the EMD to exists, the input signal needs to be approximated by a continuous representation; therefore, the discrete signal $s(t)$ is converted back into a continuous analog signal via a natural cubic spline, as given in equation . The natural cubic spline is characterised over time intervals, where the local cubic is expressed in a local time window. The time intervals are structured by points known as knot points; in this paper, such knot points are placed at the sampling times. This gives us a representation of the original signal, identified by $S(t)$ as follows:

$$S(t) = \sum_{i=1}^{N-1} \left(a_i t^3 + b_i t^2 + c_i t + d_i \right) \mathbb{1}\left[t \in [t_{i-1}, t_i] \right], \tag{7}$$

where the spline coefficients will be estimated from the original sample path, such that the representation exactly matches the sample values at these time points, $a_i = S(t_i) = s(t_i)$. We need to construct an analog continuous signal from the discrete one, since our basis decomposition requires a continuous smooth signal for the basis extraction. The number of total convexity changes (oscillations) of the analog signal $S(t)$ corresponds to $K \in \mathbb{N}$ within the time domain t, over which the signal was observed. Note that $S(t)$ is decomposed according to direct extraction of the energy associated with various intrinsic time scales. This is the property that makes it suitable to non linear and non-stationary processes. One may now define the Empirical Mode Decomposition.

Definition 1. *The Empirical Mode Decomposition of signal $S(t)$ is represented by the Intrinsic Mode Functions finite basis expansion given by*

$$S(t) = \sum_{k=1}^{K} c_k(t) + r(t) \tag{8}$$

here the collection of $\{c_k(t)\}$ basis functions are known as the Intrinsic Mode Functions (IMFs) and $r(t)$ represents the final residual (or final tendency) extracted, which has only a single convexity. In general the c_k basis will have k-convexity changes throughout the domain t and furthermore, each IMF satisfies the following mathematical properties:

- ***Oscillation*** *The number of extrema and zero-crossing must either equal or differ at most by one;*

$$\left| \left\{ \frac{dc_k(t)}{dt} = 0 : \quad t \in (t_1, t_N), \frac{dc_k(t)}{dt} \neq 0 \quad \forall t \right\} \right| \in (-1, 1) \tag{9}$$

- ***Local Symmetry*** *The mean value of the envelope defined by the local maxima and the envelope of the local minima is equal to zero.*

$$\frac{M(t) + m(t)}{2} = 0 \tag{10}$$

Note that, in the above representation, $c_k(t)$ is not explicitly expressed in a functional form, as opposed to classical stationary methods where a cosine basis or a wavelet basis function is specified. Here, the basis can take any functional form so long as it satisfies the decomposition relationship and the properties stated on the IMF. A natural way to proceed to represent an IMF is to utilise a smooth, flexible characterisation that can adapt to local non-stationary time structures; we have, again, selected the cubic spline in this work to represent $c_k(t)$.

Given a mathematical representation for the IMFs, we must now proceed to outline the process applied to extract recursively the IMF spline representations. This procedure is known as *sifting*. The first step consists of computing extrema of $S(t)$. By taking the first derivative $S'(t)$ and set it equal to zero, maxima and minima within each interval are calculated, producing the sequence of time points at which maxima and minima of $S(t)$ are located being given by:

$$t_j^* = \left[-\frac{b_j}{3a_j} \pm \sqrt{\frac{{b_j}^2 - 3a_j c_j}{9{a_j}^2}} \right] \mathbb{1}\left[t_j^* \in [t_{i-1}, t_i] \right] \quad j = 1, \ldots, K \tag{11}$$

where $\{t_j^*\}_{j=1:K}$ represents the sequence of extrema and $K << N$. Since maxima and minima always alternate, in 11 the plus refers to the maxima, while the minus to the minima. Consider the case in which the first detected extremum is a maximum and the second one is a minimum; then maxima occur at odd intervals, i.e. t_{2j+1}^*, and minima occur at even intervals, i.e. t_{2j}^*. The second step of sifting builds an upper and lower envelope of $S(t)$ as two natural cubic splines through the sequence of maxima e_2^s and the sequence of minima e_1^S respectively. The upper envelope is given by:

$$S^U(t) = \sum_{j=1}^{K-1} \left(a_{2j+1}t^3 + b_{2j+1}t^2 + c_{2j+1}t + d_{2j+1} \right) \mathbb{1}\left(t \in \left[t_{2j}^*, t_{2j+1}^* \right] \right), \tag{12}$$

such that $S^U(t) = S(t_{2j+1}^*)$ for all $t \in t_{2j+1}^*$. Equivalently, the lower envelope corresponds to:

$$S^L(t) = \sum_{j=1}^{K-1} \left(a_{2j}t^3 + b_{2j}t^2 + c_{2j}t + d_{2j} \right) \mathbb{1}\left(t \in \left[t_{2j-1}^*, t_{2j}^* \right] \right), \tag{13}$$

such that $S^L(t) = S(t_{2j}^*)$ for all $t \in t_{2j}^*$. Next, one utilises these envelopes to construct the mean signal denoted by $m(t)$ and given by

$$m(t) = \left(\frac{S\left(S^U(t)\right) + S\left(S^L(t)\right)}{2} \right) \mathbb{1}\left(t \in [t_1, t_N] \right), \tag{14}$$

which will then be used to compensate the original speech signal $S(t)$ in a recursive fashion, until an IMF is obtained. The procedure is detailed in the following algorithm.

Algorithm 1: EMD Sifting Procedure

Input: Spline $S(t)$ on $[t_1, t_N]$
Output: IMFs basis
repeat
  repeat
    (i) Identify the local extrema of $S(t)$.
    (ii) Calculate the upper envelope $S^U(t)$ and the lower envelope $S^L(t)$ respecting $S^L(t) \leq S(t) \leq S^U(t)$ for all t.
    (iii) Construct a residual time series by calculating the difference between the data and the mean of the upper and lower envelopes $S(t) \leftarrow S(t) - m(t)$.
  until *an IMF $c(t)$ is obtained;*
  Update the signal by subtracting the obtained IMF, $S(t) \leftarrow S(t) - c(t)$.
until *Having obtained a tendency $r(t)$ from the remaining signal has only one convexity in $[t_1, t_N]$.;*

It is often the case that such algorithm does not reach a mean equal to 0; therefore, multiple solutions in the literature have been proposed as stopping criteria of the sifting procedure. For further details see **?**. By exploiting the definition of cubic spline used in the representation of the analog speech signal $S(t)$ and the IMF basis functions, one can obtain a mathematical connection between the coefficients of $S(t)$ and the coefficients of $c_k(t)$ detailed as follows:

1.1.3. Predictive Distribution of IMFs under Uncertainty

Let $N = \sum_{i=1}^{M} N_i$ is an overall number of observed pairs of points $\{x_j^{(i)}, t_j^{(i)}\}$ for $j = 1, \ldots, N_i$ and $i = 1, \ldots, M$. We denote by $\mathbf{t} = \left[\mathbf{t}^{(1)}, \ldots, \mathbf{t}^{(M)}\right]$ an N-dimensional vector which is a collection of all sets of arguments. Let $\mathbf{c}_k = \left[\mathbf{c}_k^{(1)}, \ldots, \mathbf{c}_k^{(M)}\right]$ and $\boldsymbol{\mu}_k = \mu_k(\mathbf{t})$ be N-dimensional vectors.

We define $K_k(\cdot, \cdot)$ as a vector operator such that for two vector $\mathbf{t}^{(i)}$ and $\mathbf{t}^{(j)}$, N_i and N_j-dimensional respectively, it constructs an $N_i \times N_j$ matrix as follows

$$K_k(\mathbf{t}^{(i)}, \mathbf{t}^{(j)}) := \begin{bmatrix} K_k(t_1^{(i)}, t_1^{(j)}) & K_k(t_1^{(i)}, t_2^{(j)}) & \cdots & K_k(t_1^{(i)}, t_{N_j}^{(j)}) \\ K_k(t_2^{(i)}, t_1^{(j)}) & K_k(t_2^{(i)}, t_2^{(j)}) & \cdots & K_k(t_2^{(i)}, t_{N_j}^{(j)}) \\ \vdots & \vdots & \ddots & \vdots \\ K_k(t_{N_i}^{(i)}, t_1^{(j)}) & K_k(t_{N_i}^{(i)}, t_2^{(j)}) & \cdots & K_k(t_{N_i}^{(i)}, t_{N_j}^{(j)}) \end{bmatrix}_{N_i \times N_j}.$$

The distribution of $c_k(t)$ on the observation set $\{\mathbf{c}_k, \mathbf{t}\}$ can be specified under two different assumptions. We may assume the observed values of $c_k(t)$ are noise-free and therefore, the distribution in Equation (3) is valid for this case. In order for this assumption to hold, the M paths of $c_k(t)$ at the same point t_0 should have the same values.

In order to relax this assumption, we may introduce a zero-mean Gaussian noise ϵ_t with variance σ_k which adds a degree of perturbation to the observed values of $c_k(t)$. This assumption allows unequal values of $c_k(t)$ at the same argument but results is

$$c_k(t) \sim \mathcal{GP}\Big(\mu_k(t), K_k(t, t') + \sigma_k\Big), \tag{15}$$

We may chose a different distribution for the error term but for the convenience of the notation and derivations, we will assume ϵ_t to be Gaussian. In the reminder of this manuscript we assume that the observed values of $c_k(t)$ are noisy, if not otherwise specified. The derivations for the noise-free case are analogous but committing the additional variance component.

Therefore, given the observation set $\{\mathbf{c}_k, \mathbf{t}\}$, we would like to estimate the values of $c_k(t)$ at the arguments in N_0-dimensional vector $\mathbf{s}$, that is $c_k(\mathbf{s}$, given the collected information in the observation set. Given the model in Equation (15), the random pair $\big(c_k(\mathbf{t}), c_k(\mathbf{s})\big)$ has the following distribution

$$\begin{bmatrix} c_k(\mathbf{t}) \\ c_k(\mathbf{s}) \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \mu_k(\mathbf{t}) \\ \mu_k(\mathbf{s} \end{bmatrix}, \begin{bmatrix} K_k(\mathbf{t},\mathbf{t}) + \sigma_k \mathbb{I}_N & K_k(\mathbf{t},\mathbf{s}) \\ K_k(\mathbf{s},\mathbf{t}) & K_k(\mathbf{s},\mathbf{s}) \end{bmatrix} \right) \tag{16}$$

Given the formulation of the formulation of the conditional distribution of two Gaussian random variables, the predictive distribution of $c_k(t)$ on a new set of points $\mathbf{s}$, which is conditioned on the observed information and assuming that there is an observation error, is Gaussian with the conditional mean

$$\mathbb{E}_{c_k(t)|\mathbf{c}_k,\mathbf{t}}\big[c_k(\mathbf{s})\big] = \mu_k(\mathbf{s}) + K_k\big(\mathbf{s},\mathbf{t}\big)\Big(K_k\big(\mathbf{t},\mathbf{t}\big) + \sigma_k^2 \mathbf{I}_N\Big)^{-1}\big(\mathbf{c}_k - \mu_k(\mathbf{t})\big)$$

and the conditional covariance matrix given by

$$\mathbf{Cov}_{c_k(t)|\mathbf{c}_k,\mathbf{t}}\big[c_k(\mathbf{s})\big] = K_k\big(\mathbf{s},\mathbf{s}\big) - K_k\big(\mathbf{s},\mathbf{t}\big)\Big(K_k\big(\mathbf{t},\mathbf{t}\big) + \sigma_k^2 \mathbf{I}_N\Big)^{-1} K_k\big(\mathbf{t},\mathbf{s}\big)$$

TODO: explain this concept by using a priori sample and a posteriori sample plots on a simple Gaussian kernel with zero as a mean.

1.1.4. Kernel Choice

Based on Bochner's theorem, the Fourier transfor of a continious shift-invariant positive definite kernel $K(x, x')$ is a proper probability distribution function $\pi(\omega)$, assuming that $K(x, x')$ is properly scaled, that is

$$K(x, x') = \int \pi(\omega) e^{i\omega^T (x - x')}\, d\omega = \mathbb{E}_\omega\big[\phi_\omega(x)\phi_\omega(x')^*\big] \tag{17}$$

for $\phi_\omega(x) = e^{j\omega^T x} = r\big(\cos(\omega x) + i\sin(\omega x)\big)$. The density of ω is denoted by spectral density.

The non-stationary kernel can be characterised by a spectral density $\pi(\omega, \omega')$ such that

$$K(x, x') = \int\int \pi(\omega) e^{i\omega^T x - i\omega',\, ^T x'}\, d\omega d\omega' = \mathbb{E}_\omega\big[\phi_\omega(x)\phi_\omega(x')^*\big] \tag{18}$$

TODO: produce some plots about the kernel choice for IMFS, plots like from the Turners presentation - elipsoids, a priori generated sample, a posteriori distribution given a few points

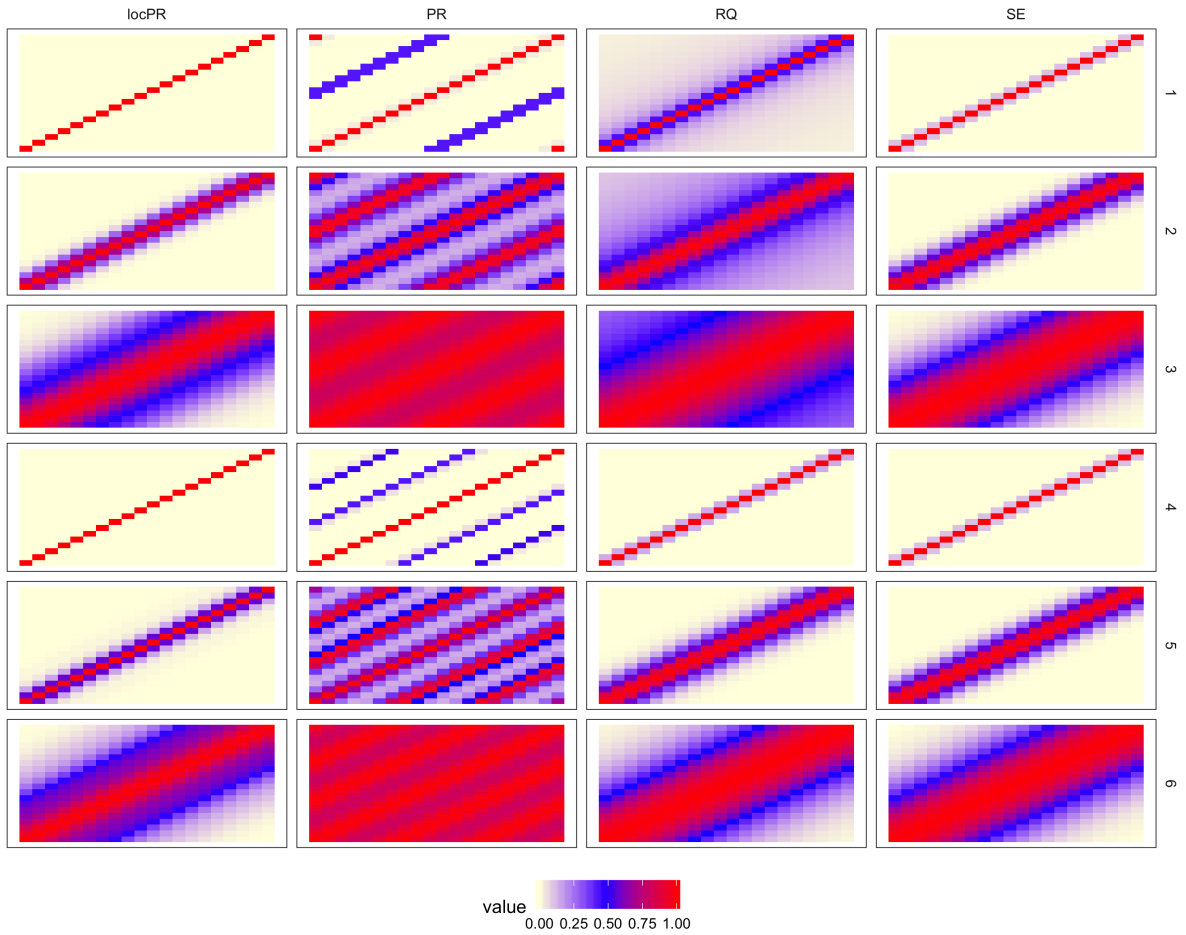

Figure 1. Stationary kernels under 6 different sets of hyper-parameters.

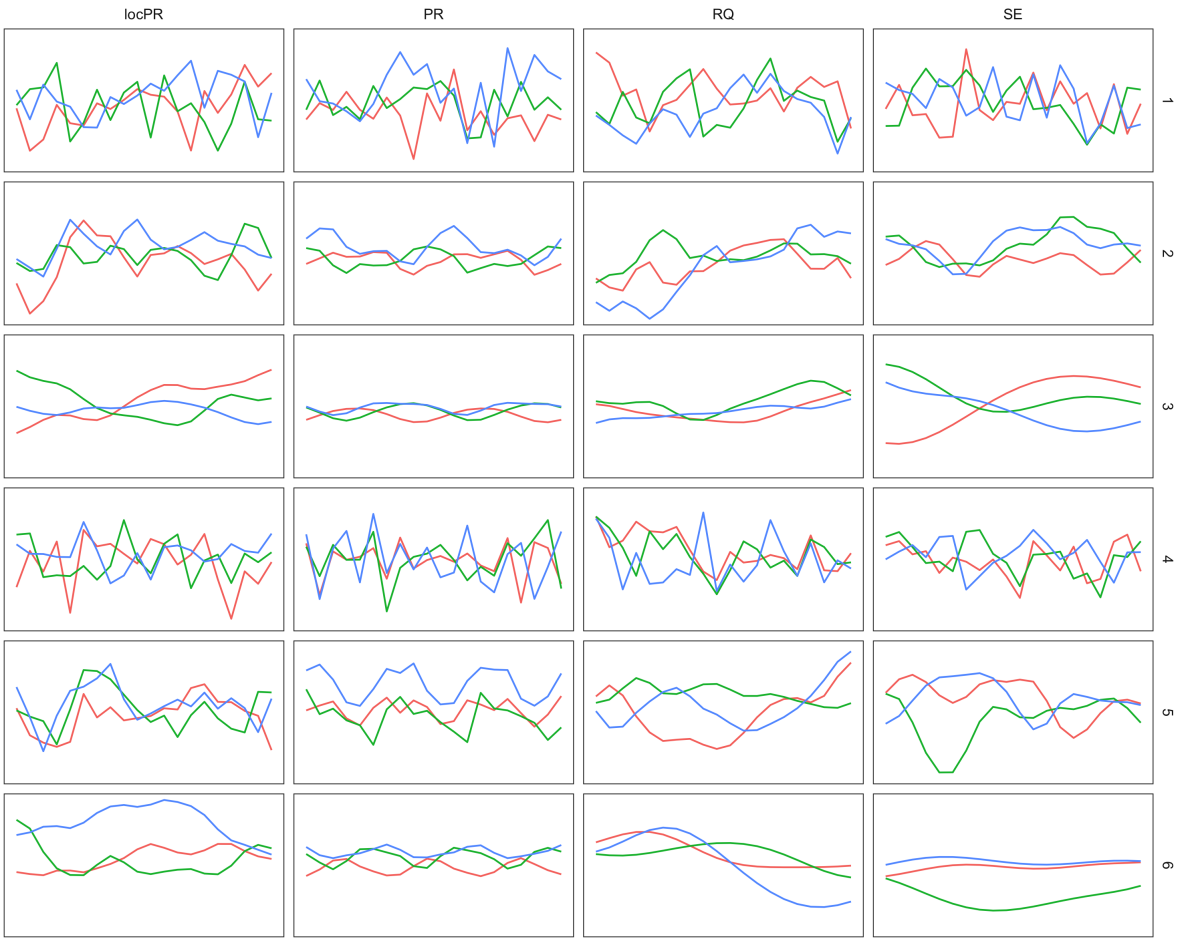

Figure 2. The 3 path of the signal c_k simulated from the a priori distribution in Equation (3) under different stationary kernel assumptions (columns wise) and for 6 different sets of hyper-parameters.

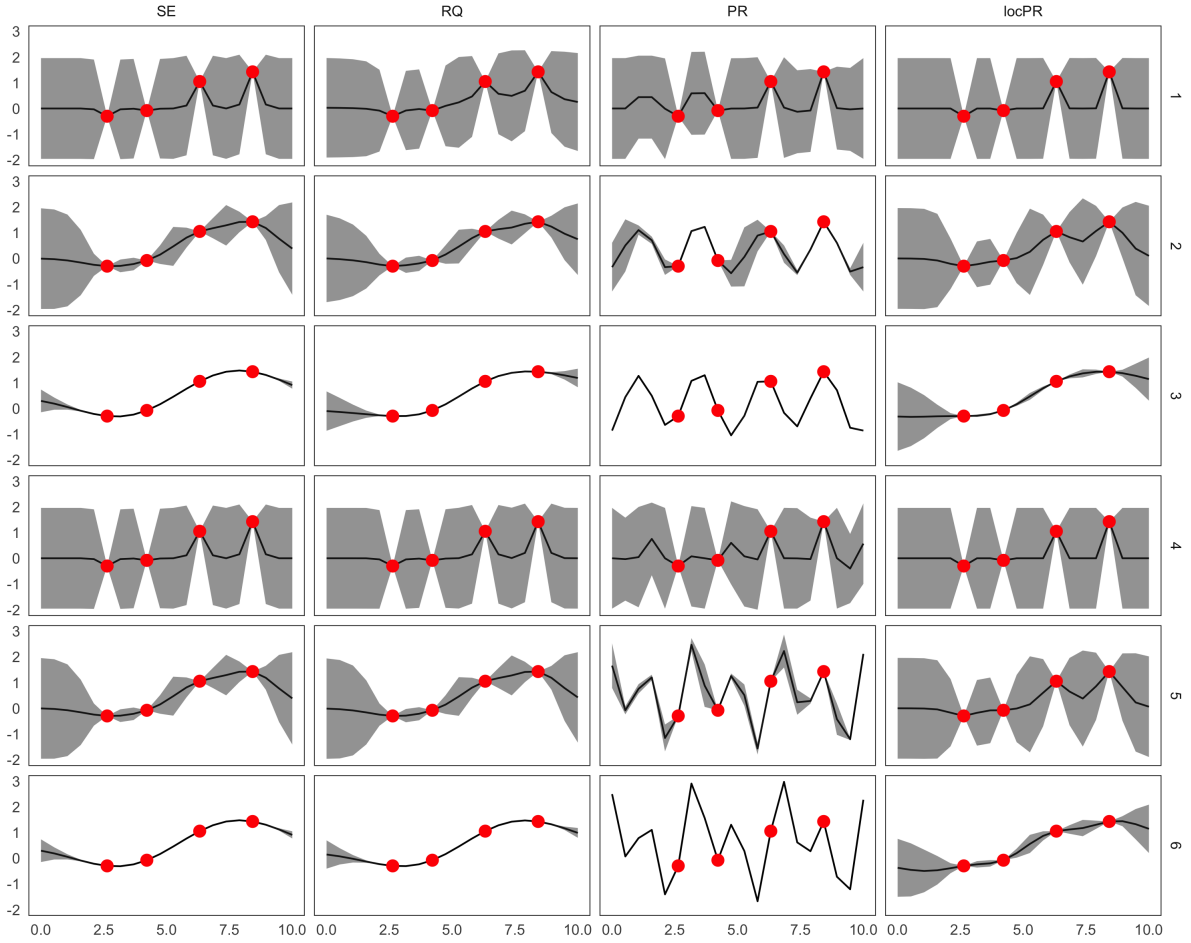

Figure 3. The predictive conditional mean of the IMF with the confidence intervals under noise-free assumption for different stationary kernel assumptions (columns wise) given 6 different sets of hyper-parameters. The red dots correspond to the observed values of the signal

1.2. Estimation of the Static Parameters

1.2.1. MLE Estimation of the Static Parameters in Gaussian Processes Models

In the following subsection we derive the MLE estimator of the vectors of parameters φ_k and Ψ_k. Given the model in Equation (15), the loglikelihood of the the observation set $\{\mathbf{c}_k, \mathbf{t}\}$ is the following

$$l_k\left(\mathbf{c}_k, \mathbf{t}, \varphi_k, \Psi_k\right) = -\frac{N}{2}\log 2\pi - \frac{1}{2}\log|\mathbf{K}_k + \sigma_k^2\mathbb{I}_N| - \frac{1}{2}\mathbf{v}_k^T\left(\mathbf{K}_k + \sigma_k^2\mathbb{I}_N\right)^{-1}\mathbf{v}_k \tag{19}$$

where $\mathbf{v}_k = \mathbf{c}_k - \mu_k$ and $\mathbf{K}_k$ denotes a $N \times N$ Gram matrix defined as

$$\mathbf{K}_k := K_k(\mathbf{t},\mathbf{t}) = \begin{bmatrix} K_k(\mathbf{t}^{(1)},\mathbf{t}^{(1)}) & K_k(\mathbf{t}^{(1)},\mathbf{t}^{(2)}) & \cdots & K_k(\mathbf{t}^{(1)},\mathbf{t}^{(M-1)}) & K_k(\mathbf{t}^{(1)},\mathbf{t}^{(M)}) \\ K_k(\mathbf{t}^{(2)},\mathbf{t}^{(1)}) & K_k(\mathbf{t}^{(2)},\mathbf{t}^{(2)}) & \cdots & K_k(\mathbf{t}^{(2)},\mathbf{t}^{(M-1)}) & K_k(\mathbf{t}^{(2)},\mathbf{t}^{(M)}) \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ K_k(\mathbf{t}^{(M-1)},\mathbf{t}^{(1)}) & K_k(\mathbf{t}^{(M-1)},\mathbf{t}^{(2)}) & \cdots & K_k(\mathbf{t}^{(M-1)},\mathbf{t}^{(M-1)}) & K_k(\mathbf{t}^{(M-1)},\mathbf{t}^{(M)}) \\ K_k(\mathbf{t}^{(M)},\mathbf{t}^{(1)}) & K_k(\mathbf{t}^{(M)},\mathbf{t}^{(2)}) & \cdots & K_k(\mathbf{t}^{(M)},\mathbf{t}^{(M-1)}) & K_k(\mathbf{t}^{(M)},\mathbf{t}^{(M)}) \end{bmatrix}_{N\times N},$$

If the sets of points $\mathbf{t}^{(i)}$ are the same and equal to $\mathbf{t}^*$, the vector $\mathbf{t}$ is constructed by stacking $\mathbf{t}^*$ by M times. Then the formulation of the likelihood simplifies to

$$l_k\Big(\mathbf{c}_k, \mathbf{t}^*, \varphi_k, \Psi_k\Big) = -\frac{N}{2}\log 2\pi - \frac{M}{2}\log|K_k(\mathbf{t}^*, \mathbf{t}^*) + \sigma_k^2\mathbb{I}_{N_*}| - \frac{1}{2}\sum_{i=1}^{M}\mathbf{v}^{(i)\ T}\Big(K_k(\mathbf{t}^*, \mathbf{t}^*) + +\sigma_k^2\mathbb{I}_{N_*}\Big)^{-1}\mathbf{v}^{(i)}) \tag{20}$$

Under the formulation of the loglikelihood in Equation (19), the static parameters of the model in Equation (15) can be estimated by solving the system of equations given by

$$\nabla l_k\Big(\mathbf{c}_k, \mathbf{t}, \varphi_k, \Psi_k\Big) = \mathbf{0} \tag{21}$$

where $\nabla l_k\Big(\mathbf{c}_k, \mathbf{t}, \varphi_k, \Psi_k\Big)$ denotes the gradient of the loglikelihood with respect to the vector of static parameters given by

$$\begin{aligned}
\frac{\partial l_k\Big(\mathbf{c}_k, \mathbf{t}, \varphi_k, \Psi_k\Big)}{\partial \varphi_k} &= \frac{1}{2} = \mathbf{c}_k\Big(\mathbf{K}_k + \sigma_k^2\mathbb{I}_N\Big)^{-1}\mathbf{v}_k\frac{\partial \mu_k(\mathbf{t})}{\partial \varphi_k} \\
\frac{\partial l_k\Big(\mathbf{c}_k, \mathbf{t}, \varphi_k, \Psi_k\Big)}{\partial \Psi_k} &= \frac{1}{2}\,\mathrm{Tr}\left\{\left(\Big(\mathbf{K}_k + \sigma_k^2\mathbb{I}_N\Big)^{-1}\mathbf{v}_k\mathbf{v}_k^T\Big(\mathbf{K}_k + \sigma_k^2\mathbb{I}_N\Big)^{-1} - \Big(\mathbf{K}_k + \sigma_k^2\mathbb{I}_N\Big)^{-1}\right)\frac{\partial \mathbf{K}_k}{\partial \Psi_k}\right\} \\
\frac{\partial l_k\Big(\mathbf{c}_k, \mathbf{t}, \varphi_k, \Psi_k\Big)}{\partial \sigma_k^2} &= \frac{1}{2}\,\mathrm{Tr}\left\{\Big(\mathbf{K}_k + \sigma_k^2\mathbb{I}_N\Big)^{-1}\mathbf{v}_k\mathbf{v}_k^T\Big(\mathbf{K}_k + \sigma_k^2\mathbb{I}_N\Big)^{-1} - \Big(\mathbf{K}_k + \sigma_k^2\mathbb{I}_N\Big)^{-1}\right\}
\end{aligned}$$

1.2.2. Kernel Alignment

1.2.3. Estimators of the Static Parameters given Splines Formulation of $x(t)$

1.3. Multikernel Representation of the Signal

1.3.1. Assuming Independence of IMFS

Given the Gaussian Process model of the $c_k(t)$, the distribution of $x(t)$ can be formulated as a uniform mixture of Gaussian Processes with different kernels. Again, we can either assume that the observed values are or are not perturbed by a noise. In the following derivation we assume that the model of the $x(t)$ includes additional term corresponding to the zero mean Gaussian noise with variance σ^2, that is

$$x(t) = \sum_{k=1}^{K} c_k(t) + r_K(t) + \epsilon \tag{22}$$

and results in the following distribution of $x(t)$

$$x(t) \sim\ GP\Big(r_K(t) + \sum_{k=1}^{K}\mu_k(t); \sum_{k=1}^{K}K_k(t,t') + \sigma^2\Big) \tag{23}$$

The scalar σ^2 can be estimated by MLE of $x(t)$, given its M realization, $\mathbf{x}^{(i)}$, formed into a vector $\mathbf{x} = \big[\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(M)}\big]$. If we denote by $K(t,t') := \sum_{k=1}^{K}K_k(t,t')$ a vector operator similarly defined as $K_k(t,t')$ and by $\mu(t) = r_K(t) + \sum_{k=1}^{K}\mu_k(t)$, then the log-likelihood of the model

$$l\big(\mathbf{x}, \mathbf{t}, \sigma^2\big) = -\frac{N}{2}\log 2\pi - \frac{1}{2}\log|K(\mathbf{t},\mathbf{t}) + \sigma^2\mathbb{I}_N| - \frac{1}{2}(\mathbf{x} - \mu(\mathbf{t}))^T\Big(K(\mathbf{t},\mathbf{t}) + \sigma^2\mathbb{I}_N\Big)^{-1}(\mathbf{x} - \mu(\mathbf{t})) \tag{24}$$

with corresponding gradient

$$\frac{\partial l(\mathbf{x},\mathbf{t},\sigma^2)}{\partial \sigma^2} = \frac{1}{2}\,\mathrm{Tr}\left\{\left(K(\mathbf{t},\mathbf{t})+\sigma^2\mathbb{I}_N\right)^{-1}(\mathbf{x}-\mu(\mathbf{t}))(\mathbf{x}-\mu(\mathbf{t}))^T\left(K(\mathbf{t},\mathbf{t})+\sigma^2\mathbb{I}_N\right)^{-1} - \left(K(\mathbf{t},\mathbf{t})+\sigma^2\mathbb{I}_N\right)^{-1}\right\}$$

The predictive distribution of $x(t)$ is given by

$$\mathbb{E}_{x(t)|\mathbf{t}}[x(\mathbf{s})] = \sum_{k=1}^{K}\mathbb{E}_{c_k(t)|\mathbf{t}}[c_k(\mathbf{s})]$$

and the covariance matrix given by

$$\mathbf{Cov}_{x(t)|\mathbf{t}}[x(\mathbf{s})] = \sum_{k=1}^{K}\mathbf{Cov}_{c_k(t)|\mathbf{t}}[c_k(\mathbf{s})] + \sigma^2$$

1.3.2. Correlation of IMFS

If the GP of c_k are not independent, the Gram matrix of the model for $x(t)$ would contain additional elements which provide the correlation structure between different IMFs

$$x(t) \sim\ GP\left(r_K(t) + \sum_{k=1}^{K} m_k(t); \sum_{k=1}^{K} K_k(t,t') + 2\sum_{k_1,k_2=1,k_1<k_2}^{K} K_{k1,k2}(t,t') + \sigma^2\right) \tag{25}$$

where $K_{k1,k2}(t,t')$ defines the dependence structure between $c_{k_1}(t)$ and $c_{k_2}(t)$

2. Brownian Bridge Analogue to construct IMFs

GP representation does not ensures itself that the predicted function from a given Gaussian process is IMF , that is, it satisfies (I1)-(I2). Therefore, we explire the following approches

Weiner process is a zero mean non-stationary Gaussian Process with the kernel $K(t,t') = min(t,t')$, that is

$$W(t) \sim \mathrm{GP}(0, K(t,t')) \tag{26}$$

The Brownian Bridge for $t \in [0,T]$ is defined as

$$B(t) = W(t) - \frac{t}{T}W(T) \tag{27}$$

Therefore, it is also the Gaussian Process which is zero mean and has the covariance kernel equals to

$$\begin{aligned}\mathbf{Cov}(B(t),B(s)) &= \mathbf{Cov}(W(t),W(s)) - \frac{s}{T}\mathbf{Cov}(W(t),W(T)) - \frac{t}{T}\mathbf{Cov}(W(s),W(T)) + \frac{st}{T^2}\mathbf{Cov}(W(T),W(T))\\ &= K(t,s) - \frac{s}{T}K(t,T) - \frac{t}{T}K(T,s) + \frac{ts}{T^2}K(T,T)\\ &= min(t,t') - \frac{ts}{T}\end{aligned}$$

The described process $B(t)$ satisfies that $B(0) = B(T) = 0$. The Brownian bridge which statisfied $B(0) = a$ and $B(T) = b$ is a solution to the following SDE system of equations

$$\begin{cases} dB(t) = dW(t) \\ B(0) = a \qquad \text{, for } 0 \le t \le T \\ B(T) = b \end{cases} \tag{28}$$

and after calculations results in the form

$$B(t) = a + (b-a)\frac{t}{T} + W(t) - \frac{t}{T}W(T) \tag{29}$$

and therefore it is a Gaussian Process

$$B(t) \sim \mathrm{GP}\Big(a + (b-a)\frac{t}{T}, K(t,t') - \frac{tt'}{T}\Big) \tag{30}$$

for $0 \leq t \leq T$

2.1. Symmetric Local Extremas of IMFs

On every time internal there is a Brownian bridge or constrained Brownian bridge which starts and end from local extrema which are $x^{min}(t) = -x^{max}(t)$ for $t \in [\tau_i, \tau_{i+1}$

2.2. Nonsymmetric

2.3. Bayesian EMD

1. Construct a set of functions in Bayesian setting to have a IMF representation with restricted posterior (what needs to be satisfied on maxima and minima and how to ensure it) 2. Analogous of Brownian Bridge IMFs in Bayesian setting

Berger's optimal theory. Books on smoothing