Notes: Empirical Mode Decomposition & Gaussian Processes

*

Version: June 11, 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 **1**
1.1 The Stochastic Representation by Gaussian Processes . . . 1
1.1.1 Continuous signal $S(t)$ as a Gaussian Process . . . 2

2 Multiple Trials Setting: How to Construct $S(t)$ in Noisy Environment **3**
2.1 Review of Denoising Approaches . . . 4

3 Gaussian Process Representation Given Splines Formulation of $S(t)$. **4**

4 Background: Empirical Mode Decomposition **5**

5 Review of Stationary and Non-stationary Kernels **6**

6 Brownian Bridge Analogue to construct IMFs **9**
6.0.1 Brownian Bridge Movement Model . . . 10
6.1 Symmetric Local Extremas of IMFs . . . 11
6.2 Nonsymmetric . . . 11
6.3 Bayesian EMD . . . 11

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 (ie an chaotic system) or stochastic. In fact, the process which we observe is the following

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

The signal $y(t)$ is observed at $t = (t_1 < \cdots < t_N) = \{t_i\}_{i=1:N}$, where the subscripts represent the sampling index times. Therefore, our observation set consists of pairs $\{t_n, y_n\}$ where $y_n = y(t_n)$ for $t_n \in [0, T]$.

For 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 spline representation as in Equation . Having an continuous representation $S(t)$, which is an approximation of $s(t)$, we define the formulation of $S(t)$ given by EMD into M intrinsic mode functions (IMFs) as follows

$$S(t) = \sum_{m=1}^{M} \gamma_m(t) + r(t) = \sum_{m=1}^{M} \mathrm{Re}\Big\{ A_m(t) e^{i\theta_m(t)} \Big\} + r(t). \tag{2}$$

where $r(t)$ represents a tendency which does not have much of oscillation and therefore characterize the low frequency tend of $S(t)$. The background on the EMD decomposition is given in Subsection **??**.

1.1.1. Continuous signal $S(t)$ as a Gaussian Process

Out goal is to obtain a stochastic representation of the continuous signal $S(t)$ by a Gaussian Process. We postulate that each IMFs function, $\gamma_m(t)$, is a Gaussian process

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

where $k_m(t,t')$ is a positive definite covariance kernel which is parametrized by a set of parameters Ψ_m.

Let us assume that we sample $S(t)$, and functions $\gamma_m(t)$ for $m = 1,\ldots,M$ at the N time points $t_1 < \ldots < t_N$. We denote by $\mathbf{t}$ the vector of points t_n for $n = 1,\ldots,N$.

Therefore, given the observations $\gamma_m(\mathbf{t})'' = \big[\gamma_m(t_1),\ldots,\gamma_m(t_N)\big]$, we would like to predict the values of $\gamma_m(t)$ at the argument s that is $c_k(s)$, given the collected information in the observation set. Since $c(t)$ is a Gaussian Process, the random variable $\gamma_m(s)|\gamma_m(\mathbf{t}), \mathbf{t}$ is a Gaussian Process with the conditional mean

$$\mu_m(s) := \mathbb{E}_{\gamma_m(t)|\gamma_m(\mathbf{t}),\mathbf{t}}\big[\gamma_m(s)\big] = \mathbf{k}_m\big(s,\mathbf{t}\big)\mathbf{K}_m\big(\mathbf{t},\mathbf{t}\big)^{-1}\gamma_m(\mathbf{t})$$

and the conditional covariance matrix given by

$$\tilde{k}_m(s,s') := \mathbb{E}_{\gamma_m(t)|\gamma_m(\mathbf{t}),\mathbf{t}}\Big[(\gamma_m(s)-\mu_m(s))(\gamma_m(s')-\mu_m(s'))\Big] = k_m\big(s,s'\big) - \mathbf{k}_m\big(s,\mathbf{t}\big)\mathbf{K}_m\big(\mathbf{t},\mathbf{t}\big)^{-1}\mathbf{k}_m\big(\mathbf{t},s'\big)^T$$

where

$$\mathbf{K}_m(\mathbf{t},\mathbf{t}) := \begin{bmatrix} k_m(t_1,t_1) & k_m(t_1,t_2) & \cdots & k_m(t_1,t_N) \\ k_m(t_2,t_1) & k_m(t_2,t_2) & \cdots & k_m(t_2,t_N) \\ \vdots & \vdots & \ddots & \vdots \\ k_m(t_N^{(i)},t_1) & k_m(t_N,t_2) & \cdots & k_m(t_N,t_N) \end{bmatrix}_{N\times N}$$

and

$$\mathbf{k}_m(s,\mathbf{t}) := \begin{bmatrix} k_m(s,t_1) & k_m(s,t_2) & \cdots & k_m(s,t_N) \end{bmatrix}_{1\times N}.$$

Multikernel Representation of $S(t)$

The tendency component $r(t)$ can be modelled as a Gaussian Process itself or one can assume that $S(t)$ is a Gaussian Process conditioned on $r(t)$, that is

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

where $k(t,t')$ is a function of the kernels $k_m(t,t')$ for $m \in \{1,\ldots,M\}$ These two approaches provide an unconditional and conditional stochastic representation of $S(t)$, respectively, and determine two different estimators of the out-of-sample forecast for $S(t)$. The later is a more convenient assumption to preserve the monotonicity of the $r(t)$ which is a desired property of a residual function in the decomposition in Equation (2).

To ensure the function $r(t)$ to have only single convexity change, $r(t)$ 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 $S(t)$ would be conditioned on the extrapolation of $r(t)$. In order to preserve the monotonicity property of the tendency function $r(t)$ in the out-of-sample prediction, the extrapolation from a low order spline representation of $r(t)$, 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 $r(t)$. In the following work we would like to guarantee the out-of-sample monotonicity of $r(t)$ 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..

Given the Gaussian Process model of the $\gamma_m(t)$ in Equation (3), the distribution of $S(t)$ can be formulated as a uniform mixture of Gaussian Processes with different kernels. If we assume that the processes $\gamma_m(t)$ are independent, then, the stochastic representation of $S(t)$ from Equation (4) can be formulated as follows

$$S(t)|r(t) \sim GP\left(r(t); \sum_{m=1}^{M} k_m(t,t')\right) \tag{5}$$

If we denote by $k(t,t') := \sum_{m=1}^{M} k_m(t,t')$, then predictive distribution of $S(t)$ is given by

$$\mu(s) := \mathbb{E}_{S(t)|r(t),\mathbf{s},\mathbf{t}}[S(\mathbf{s})] = r(\mathbf{s}) + \sum_{m=1}^{M} \mu_m(s) \tag{6}$$

and the covariance matrix given by

$$\tilde{k}(s,s') := \mathbb{E}_{S(t)|S(\mathbf{t}),\mathbf{t}}\Big[(S_m(s) - \mu(s))(S(s') - \mu(s'))\Big] = \sum_{m=1}^{M} \tilde{k}(s,s') \tag{7}$$

If the processes of $\gamma_m(t)$ are not independent, the Gram matrix of the model for $S(t)$ contain additional elements which provide the correlation structure between different IMFs

$$s(t)|r(t) \sim \mathcal{GP}\left(r(t); \sum_{m=1}^{M} k_m(t,t') + 2 \sum_{m_1,m_2=1,m_1<m_2}^{M} k_{m1,m2}(t,t')\right) \tag{8}$$

where $k_{m1,m2}(t,t')$ defines the dependence structure between $\gamma_{m_1}(t)$ and $\gamma_{m_2}(t)$.

2. Multiple Trials Setting: How to Construct $S(t)$ in Noisy Environment

We consider the following experiment setup. Let J represents the number of trials in our experiment which characterize the set of sample that we collected, ie. 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 $t_0 \in [0,T]$, that the 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

2.1. Review of Denoising Approaches

1. median filter
2. smoothing splines
3. L1 trend filter
4. Exponential moving average (EMA) / a weighted moving average (WMA)

3. Gaussian Process Representation Given Splines Formulation of $S(t)$.

As remarked in Subsection **??**, the EMD procedure required that the underlying signal has a continuous formulation. For the EMD to exist, the underlying signal $S(t)$ needs to be approximated. Such approximation is covered in this work by a natural cubic spline representation of $S(t)$.

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:

$$S(t) = \sum_{i=1}^{N-1} \left(a_i t^3 + b_i t^2 + c_i t + d_i \right) \mathbb{1}\left[t \in [t_{i-1}, t_i] \right], \tag{9}$$

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 EMD defined in Section 4 of the signal $S(t)$ as in Equation (2).

Given the spline representation of $S(t)$, each IMF $\gamma_{m,j}^{(i)}$ can be obtained as a natural cubic spline, defined as $\gamma_{m,j}^{(i)}(t)$ with the following formulation:

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

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

$$\gamma_{m,j}^{(i)}(t) = \sum_{j=1}^{N_i} \left(a_j t^3 + b_j t^2 + c_j t + d_j \right) \mathbb{1}\left(t \in (t_{j-1}^i, t_j^i) \right) = \sum_{j=1}^{N_i} s_j(t) \mathbb{1}\left(t \in (t_{j-1}^i, t_j^i) \right) \tag{10}$$

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

Note that, in the above representation, $\gamma_m(t)$ 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 $\gamma_m(t)$.

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 signal $S(t)$ in a recursive fashion, until an IMF is obtained. The procedure is detailed in the following algorithm.

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:

Proposition 1. *The k-th extracted IMF denoted as c_k can be expressed as a cubic spline whose coefficients are a linear combination of the spline coefficients of $S(t)$ and the coefficients of the mean envelopes of the $k-1$ IMFs previously extracted, i.e.*

$$c_D = c_0 - \sum_{i=0}^{Q-1} m_{D,i} = \sum_{i=1}^{N-1} \left(a_i^Q t^3 + b_i^Q t^2 + c_i^Q t + d_i^Q \right) \mathbb{1}\left(t \in [t_{i-1}, t_i] \right) \tag{15}$$

where c_0 is $S(t)$ and $m_{D,i} = \frac{S(e_1^{h_{D,i}}) + S(e_2^{h_{D,i}})}{2}$ and Q is the number of sifting steps required to extract the IMF itslef.

Note that our notation $c(t)$ changed in this proposition by becoming c for a clearer proofs. The proof is provided in the appendix **??**.

4. Background: 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*

$$S(t) = \sum_{k=1}^{K} c_k(t) + r(t) \tag{16}$$

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;*

$$\left| \left\{ \frac{d\gamma_m(t)}{dt} = 0 : \quad t \in (t_1, t_N), \frac{d\gamma_m(t)}{dt} \neq 0 \quad \forall t \right\} \right| \in (-1, 1) \tag{17}$$

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

$$\frac{M(t) + m(t)}{2} = 0 \tag{18}$$

5. Review of Stationary and Non-stationary Kernels

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

$$K(x, x') = \int \pi(\omega) e^{i\omega^T (x - x')} \, d\omega = \mathbb{E}_\omega \left[\phi_\omega(x) \phi_\omega(x')^* \right] \tag{19}$$

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.

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

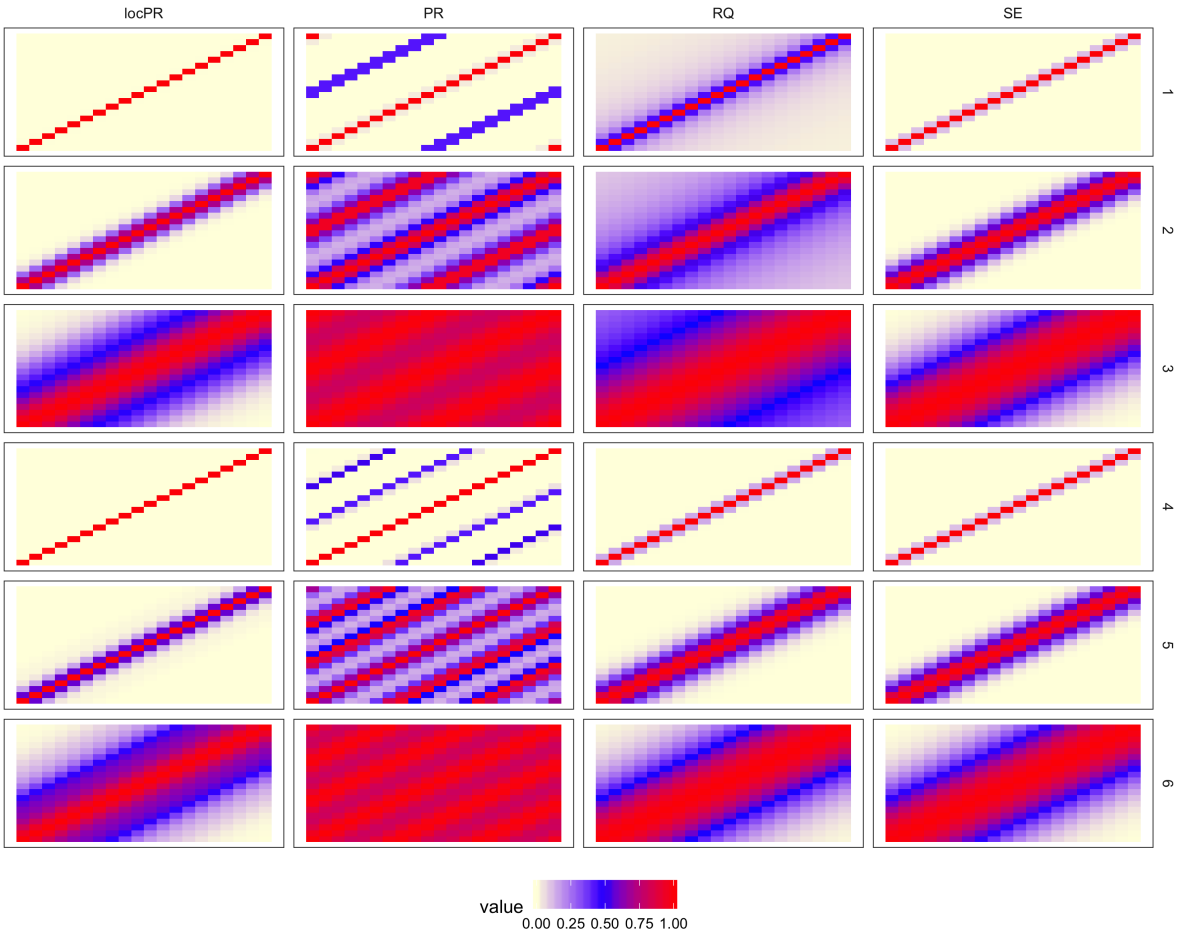

Figure 1. Stationary kernels under 6 different sets of hyper-parameters.

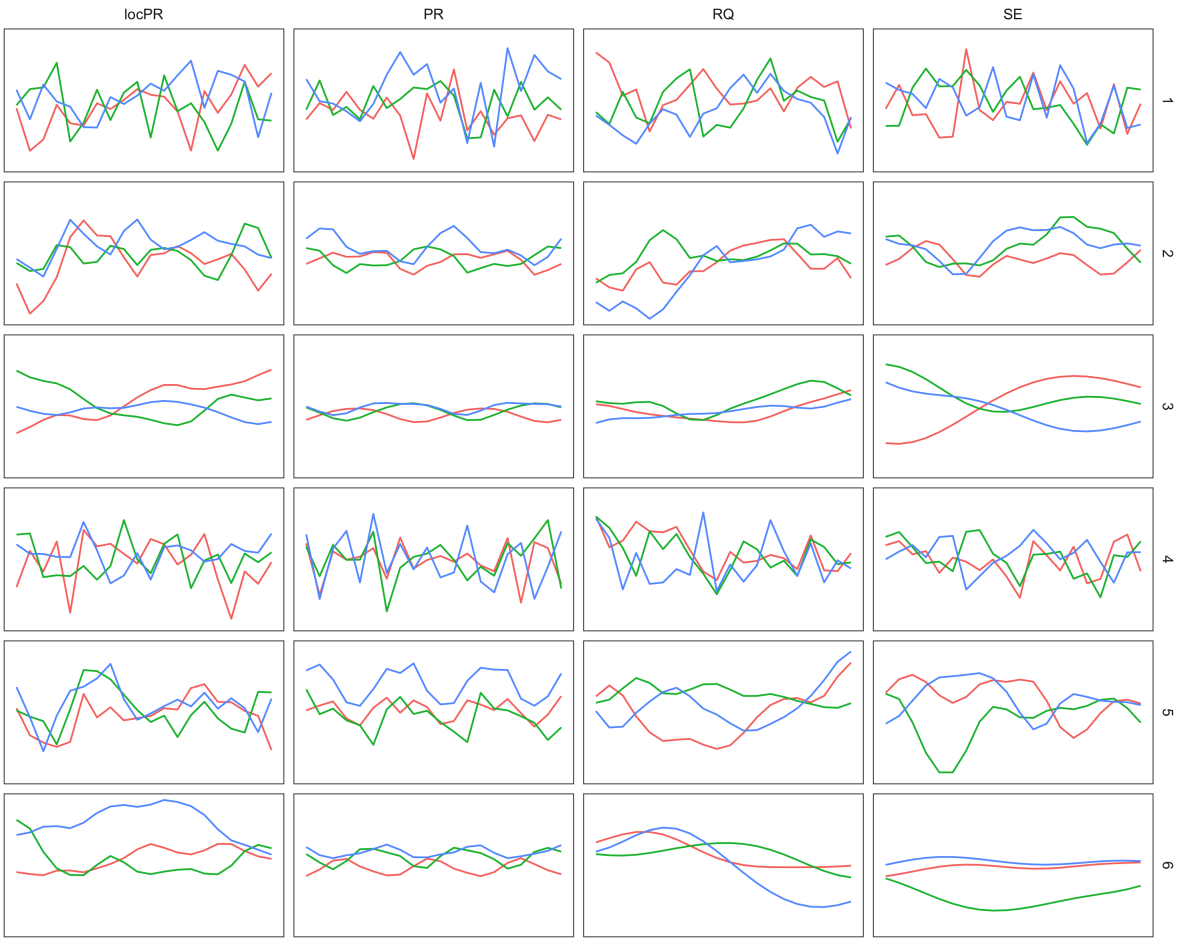

Figure 2. The 3 path of the signal c_k simulated from the a priori distribution in Equation (3) under different stationary kernel assumptions (columns wise) and for 6 different sets of hyper-parameters.

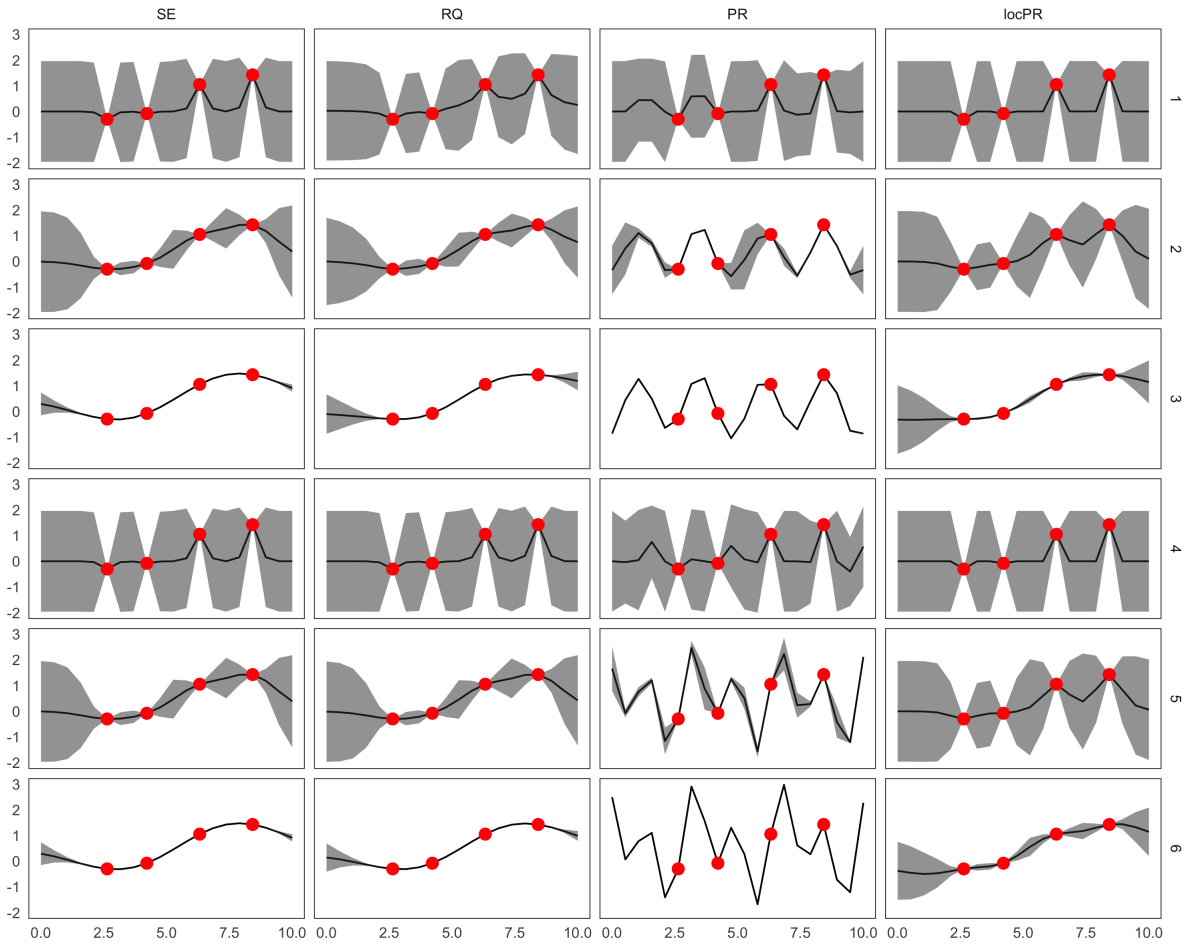

Figure 3. The predictive conditional mean of the IMF with the confidence intervals under noise-free assumption for different stationary kernel assumptions (columns wise) given 6 different sets of hyper-parameters. The red dots correspond to the observed values of the signal

Non-stationary Kernels

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

$$k(t,t') = \int\int \pi(\omega,\omega') e^{2\pi i \omega t - \omega' t'} \, d\omega d\omega' = \mathbb{E}_{\omega,\omega'}\left[\phi_{\omega,\omega'}(x)\phi_{\omega,\omega'}(x')^*\right] \tag{20}$$

where $\pi(\omega)$ is a spectral density of the kernel k over frequencies ω and ω'. For a scalar inputs x and x' the density $\pi(\omega, \omega')$ corresponds to bivariate distribution.

1. Non-stationary generalisation of the Squared Exponential kernel (GSE)

$$k_{GSE}(t,t') = \sigma(t)\sigma(t')\left(\frac{2\mu(t)\mu(t')}{\mu(t)^2+\mu(t')^2}\right)^{-\frac{1}{2}} \exp\left\{-\frac{(t-t')^2}{\mu(t)^2+\mu(t')^2}\right\}$$

for

$$\log\sigma(t) \sim \mathcal{GP}\left(\mu_\sigma, k_\sigma(t,t')\right)$$
$$\log\mu(t) \sim \mathcal{GP}\left(\mu_\mu, k_\mu(t,t')\right)$$

2. Spectral Mixture Kernel by (**?**) which is formulated as follows

$$k_{SM}(t,t') = \sum_{q=1}^{Q} w_q^2 \exp\Big\{ -2\pi^2 \tilde{\mathbf{t}}\boldsymbol{\Sigma}_q \tilde{\mathbf{t}}^T \Big\} \boldsymbol{\phi}_{\mu_q,\mu_q'}(t) \boldsymbol{\phi}_{\mu_q,\mu_q'}(t') \tag{21}$$

where

$$\tilde{\mathbf{t}} = \begin{bmatrix} t \\ -t' \end{bmatrix} \text{ and } \boldsymbol{\phi}_{\mu_q,\mu_q'}(t) = \begin{bmatrix} \cos(2\pi\mu_q t) + \cos(2\pi\mu_q' t) \\ \sin(2\pi\mu_q t) + \sin(2\pi\mu_q' t) \end{bmatrix}.$$

The spectral density of the kernel k_{SM} is defined by a weighted mixture

$$\pi(\omega,\omega') = \sum_{q=1}^{Q} w_q^2 \pi_q(\omega,\omega) \tag{22}$$

where $\pi_q(\omega,\omega)$ is a sum of bivariate normal densities with two dimension mean vectors are equal to the the eight combinations of the two element permutations of the set $\{\mu_q,\mu_q'\}$ and $\{-\mu_q,-\mu_q',\}$, and the covariance matrix $\boldsymbol{\Sigma}_q$.

3. Generalized Spectral Mixture Kernel by (**?**) with Gibbs kernel formulations which is formulated as follows

$$k_{GSM}(t,t') = \sum_{q=1}^{Q} w_q(t) w_q(t') \Big(\frac{2 l_g(t) l_q(t')}{l_g(t)^2 + l_q(t')^2} \Big)^{-\frac{1}{2}} \exp\Big\{ -\frac{(t-t')^2}{l_q(t)^2 + l_q(t')^2} \Big\} \cos\Big(2\pi(\mu_q(t)t - \mu_q(t')t') \Big) \tag{23}$$

for

$$\begin{aligned} \log l_q(t) &\sim \mathcal{GP}\Big(\mu_l, k_l(t,t')\Big) \\ \log it\mu_q(t) &\sim \mathcal{GP}\Big(\mu_\mu, k_\mu(t,t')\Big) \\ \log w_q(t) &\sim \mathcal{GP}\Big(\mu_w, k_w(t,t')\Big) \end{aligned}$$

The spectral density of the kernel k_{GSM} is ...

4. Sparse Spectrum Kernel by (**?**)

6. 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

$$W(t) \sim \text{GP}(0, K(t,t')) \tag{24}$$

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

$$B(t) = W(t) - \frac{t}{T} W(T) \tag{25}$$

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 $B(t_0) = a$ and $B(_t1) = b$ is a solution to the following SDE system of equations

$$\begin{cases} dB(t) = dW(t) \\ B(t_0) = a \\ B(t_1) = b \end{cases} \text{, for } t_0 \leq t \leq t_1 \tag{26}$$

Therefore, the Brownian Bridge

and after calculations results in the form

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

and therefore it is a Gaussian Process

$$B(t) \sim \mathrm{GP}\Big(a + (b-a)\frac{t}{T}, K(t,t') - \frac{tt'}{T}\Big) \tag{28}$$

for $0 \leq t \leq T$

6.0.1. Brownian Bridge Movement Model

Let $W(t)$ denote the Brownian Motion such that

$$W(t) \sim \mathcal{GP}, \big(0, k(t,t')\big) \tag{29}$$

with the kernel function $k(t,t') = \min\{t,t'\}$ and $t \in [0,T]$. Let us define the sequence of N points $0 \leq t_1 < t_2 < \ldots < t_N \leq < T$ such that we require that the process $W(t)$ had the fixed values at that points $W(t_i) = a_i$ for $i \in \{1,\ldots,N\}$. Theretofore, we are looking for a Gaussian process for $t \in [0,T]$ model of a conditional variable defined as

$$B(t) := W(t)|W(t_1) = a_1, W(t_2) = a_2, \ldots, \ldots W(t_N) = a_N \tag{30}$$

Let $\mathbf{W} = \big[W(t_1), W(t_2), \ldots, \ldots W(t_N)\big]$ and $\mathbf{t} = \big[t_1, t_2, \ldots, t_N\big]$ be N-dimensional vectors. Since $W(t)$ is a Gaussian Process, the random variable $W(t)|W(t_1), W(t_2), \ldots, \ldots W(t_N)$ is also a Gaussian Process with the conditional mean

$$\mu(t) := \mathbb{E}_{W(t)|W(t_1),W(t_2),\ldots,\ldots W(t_N)}\big[W(t)\big] = \mathbf{k}\big(t,\mathbf{t}\big)^T\mathbf{K}\big(\mathbf{t},\mathbf{t}\big)^{-1}\mathbf{W} \tag{31}$$

and the covariance function

$$\big(t,t'\big) = \mathbb{E}_{W(t)|W(t_1),W(t_2),\ldots,\ldots W(t_N)}\Big[\big[W(t)-\mu(t)\big]\big[W(t')-\mu(t')\big]\big]\Big] = k(t,t') - \mathbf{k}\big(t,\mathbf{t}\big)^T\mathbf{K}\big(\mathbf{t},\mathbf{t}\big)^{-1}\mathbf{k}\big(t',\mathbf{t}\big) \tag{32}$$

where

$$\mathbf{K}(\mathbf{t},\mathbf{t}) := \begin{bmatrix} t_1 & t_1 & t_1 & \ldots & t_1 \\ t_1 & t_2 & t_2 & \ldots & t_2 \\ t_1 & t_2 & t_3 & \ldots & t_3 \\ \vdots & \vdots & \ddots & \vdots & \\ t_1 & t_2 & t_3 & \vdots & t_N \end{bmatrix}_{N\times N} \tag{33}$$

and

$$\mathbf{k}(t,\mathbf{t}) := \begin{bmatrix} \min(t,t_1) \\ \min(t,t_2) \\ \vdots \\ \min(t,t_N) \end{bmatrix}_{N\times 1} \tag{34}$$

6.1. 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 $t \in [\tau_i, \tau_{i+1}$

6.2. Nonsymmetric

6.3. 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