Notes: Empirical Mode Decomposition & Gaussian Processes

*

Version: June 3, 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. IMFs as Gaussian Processes

Let $x(t)$ be a continuous real-valued signal and let us define its EMD decomposition into K intrinsic mode functions (IMFs) given by

$$x(t) = \sum_{k=1}^{K} c_k(t) + r_K(t) = \sum_{k=1}^{K} \mathrm{Re}\left\{ A_k(t) e^{i\theta_k(t)} \right\} + r_K(t). \tag{1}$$

We assume that each $c_k(t)$ is a Gaussian process such that

$$c_k(t) \sim \mathcal{GP}\Big(\mu_k(t), K_k(t,t')\Big), \tag{2}$$

where $\mu_k(t)$ is a mean function of the process and $K_k(t,t')$ is a positive definite covariance kernel which are parametrized with a set of parameters φ_k and Ψ_k, respectively. Therefore, the moments of $c_k(t)$ are given in the form of the following one dimensional and two dimensional mappings

$$\mathbb{E}_{c_k(t)|\varphi_k}\left[c_k(t)\right] = \mu_k(t),$$
$$\mathbb{E}_{c_k(t)|\Psi_k}\left[\left(c_k(t)-\mu_k(t)\right)\left(c_k(t')-\mu_k(t')\right)\right] = K_k(t,t')$$

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 $i \in \left\{1,\ldots,M\right\}$ 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

$$\mathbf{x}^{(i)} = \sum_{k=1}^{K} \mathbf{c}_k^{(i)} + \mathbf{r}_K^{(i)} \tag{3}$$

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, $\boldsymbol{\mu}_k^{(i)} = \mu_k(\mathbf{t}^{(\mathbf{i})})$.

1.1.1. 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 $t_j^{(i)}$.

1.1.2. 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 (2) 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

$$c_k(t) \sim \mathcal{GP}\Big(\mu_k(t), K_k(t,t') + \sigma_k\Big), \tag{4}$$

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 (4), the random pair $(c_k(\mathbf{t}), c_k(\mathbf{s}))$ has the following distribution

$$\begin{bmatrix} c_k(\mathbf{t}) \\ c_k(\mathbf{s}) \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \mu_k(\mathbf{t}) \\ \mu_k(\mathbf{s} \end{bmatrix}, \begin{bmatrix} K_k(\mathbf{t},\mathbf{t}) + \sigma_k \mathbb{I}_N & K_k(\mathbf{t},\mathbf{s}) \\ K_k(\mathbf{s},\mathbf{t}) & K_k(\mathbf{s},\mathbf{s}) \end{bmatrix} \right) \tag{5}$$

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

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

and the conditional covariance matrix given by

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

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

1.1.3. 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

$$K(x, x') = \int \pi(\omega) e^{i\omega^T (x - x')} \, d\omega = \mathbb{E}_\omega \left[\phi_\omega(x) \phi_\omega(x')^* \right] \tag{6}$$

for $\phi_\omega(x) = e^{j\omega^T x} = r(\cos(\omega x) + i \sin(\omega x)$.

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

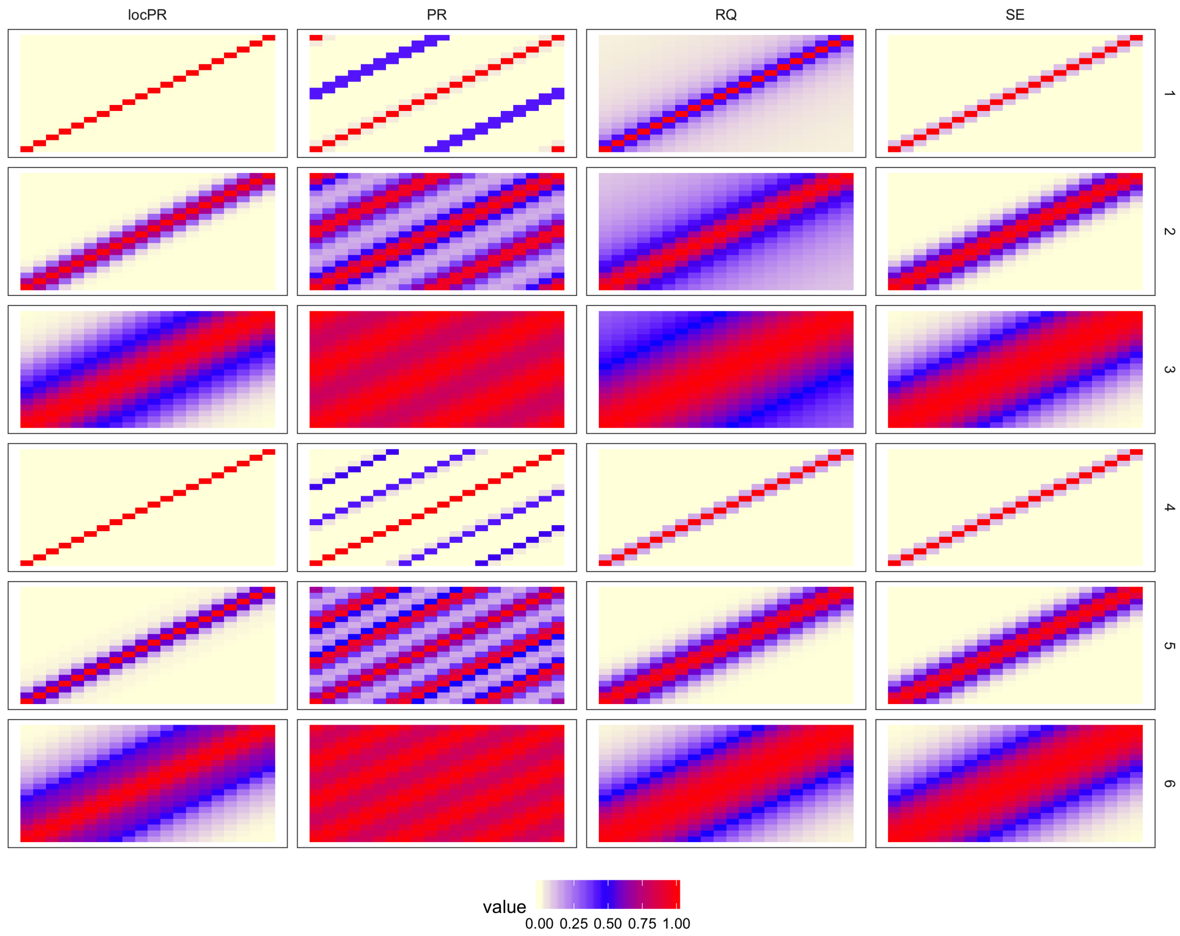

Figure 1. Stationary kernels under 6 different sets of hyper-parameters.

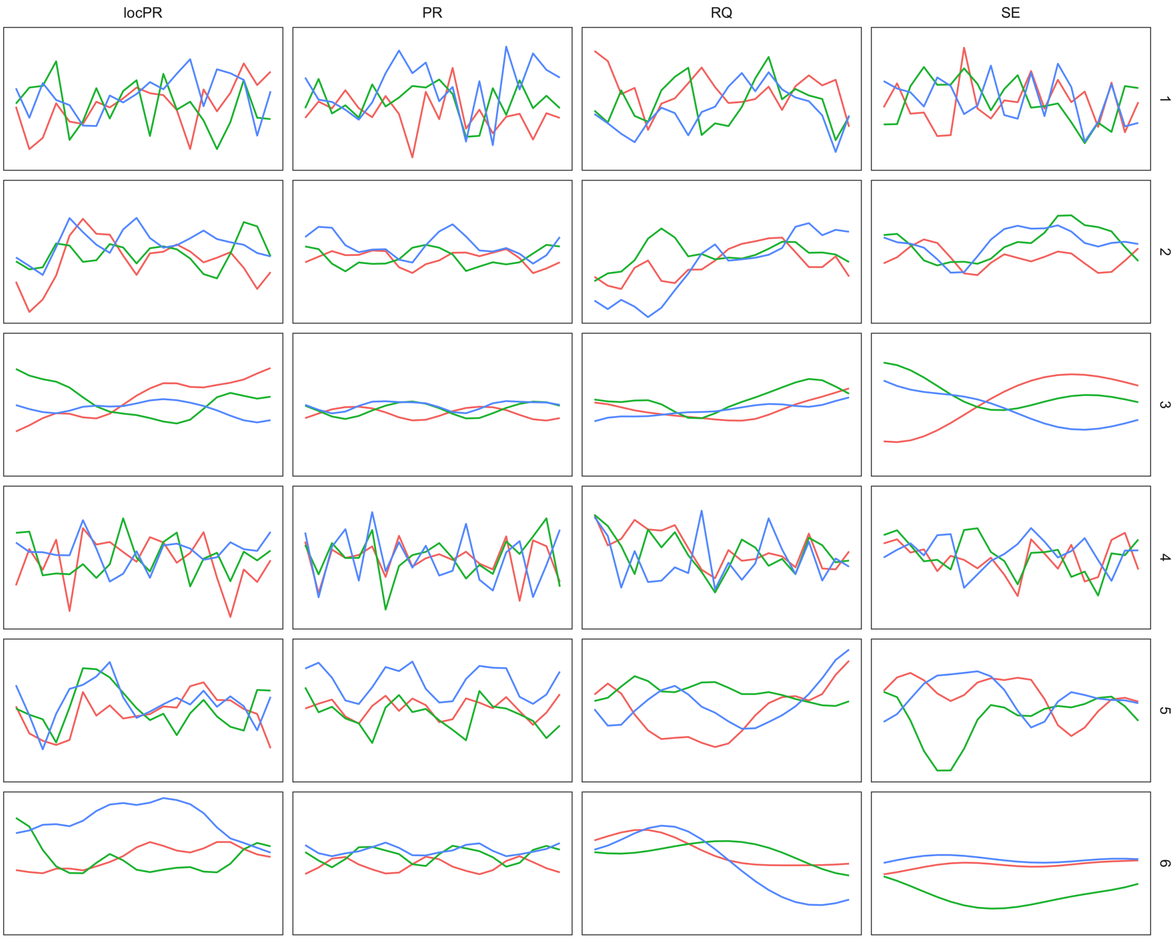

Figure 2. The 3 path of the signal c_k simulated from the a priori distribution in Equation (2) under different stationary kernel assumptions (columns wise) and for 6 different sets of hyper-parameters.

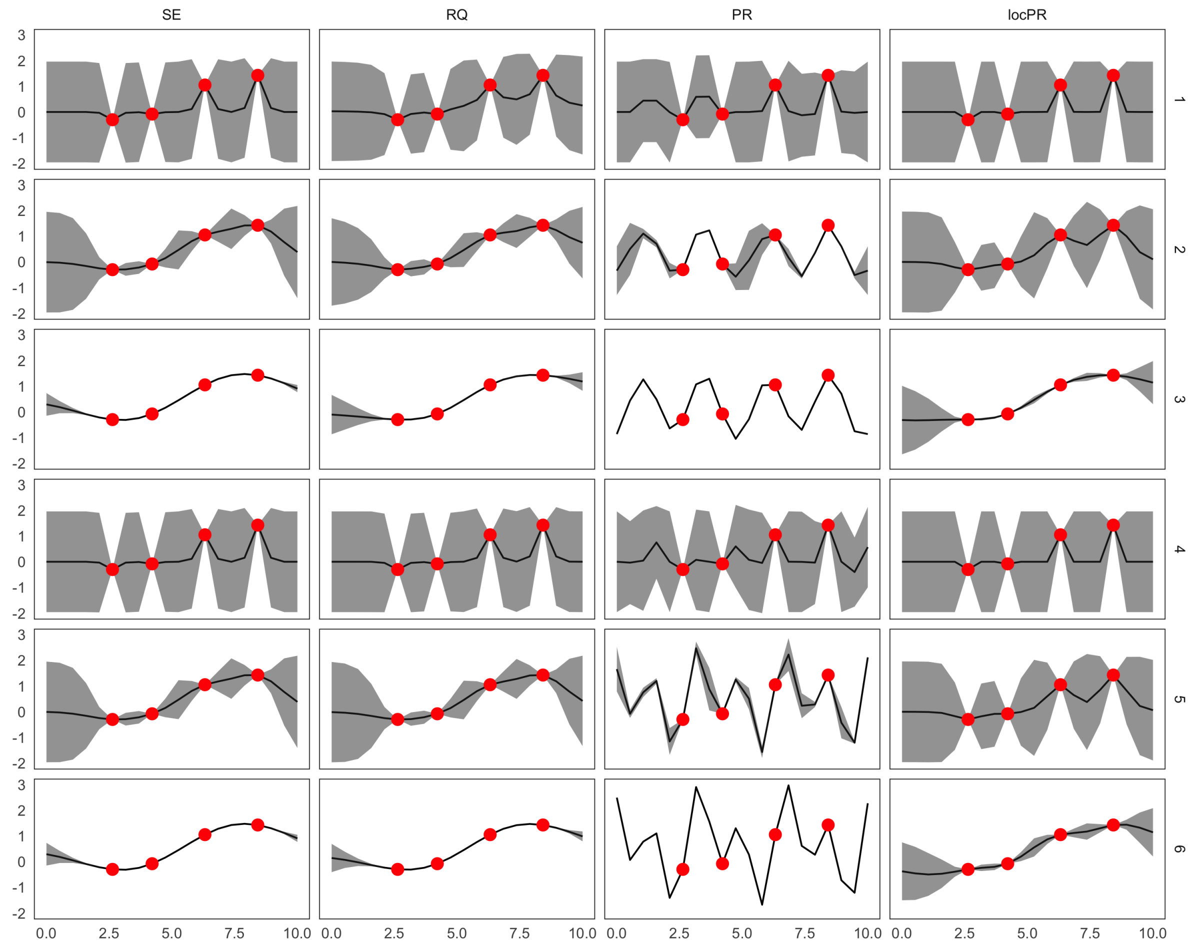

Figure 3. The predictive conditional mean of the IMF with the confidence intervals under noise-free assumption for different stationary kernel assumptions (columns wise) given 6 different sets of hyper-parameters. The red dots correspond to the observed values of the signal

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 (4), the loglikelihood of the the observation set $\{\mathbf{c}_k, \mathbf{t}\}$ is the following

$$l_k\left(\mathbf{c}_k, \mathbf{t}, \varphi_k, \Psi_k\right) = -\frac{N}{2}\log 2\pi - \frac{1}{2}\log|\mathbf{K}_k + \sigma_k^2 \mathbb{I}_N| - \frac{1}{2}\mathbf{v}_k^T\left(\mathbf{K}_k + \sigma_k^2\mathbb{I}_N\right)^{-1}\mathbf{v}_k \tag{7}$$

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

$$l_k\Big(\mathbf{c}_k, \mathbf{t}^*, \varphi_k, \Psi_k\Big) = -\frac{N}{2}\log 2\pi - \frac{M}{2}\log|K_k(\mathbf{t}^*,\mathbf{t}^*) + \sigma_k^2\mathbb{I}_{N_*}| - \frac{1}{2}\sum_{i=1}^{M}\mathbf{v}^{(i)\,T}\Big(K_k(\mathbf{t}^*,\mathbf{t}^*) + +\sigma_k^2\mathbb{I}_{N_*}\Big)^{-1}\mathbf{v}^{(i)}) \tag{8}$$

Under the formulation of the loglikelihood in Equation (7), the static parameters of the model in Equation (4) can be estimated by solving the system of equations given by

$$\nabla l_k\Big(\mathbf{c}_k, \mathbf{t}, \varphi_k, \Psi_k\Big) = \mathbf{0} \tag{9}$$

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

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

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

$$x(t) = \sum_{k=1}^{K} c_k(t) + r_K(t) + \epsilon \tag{10}$$

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

$$x(t) \sim\ GP\Big(r_K(t) + \sum_{k=1}^{K}\mu_k(t); \sum_{k=1}^{K}K_k(t,t') + \sigma^2\Big) \tag{11}$$

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

$$l(\mathbf{x}, \mathbf{t}, \sigma^2) = -\frac{N}{2}\log 2\pi - \frac{1}{2}\log|K(\mathbf{t},\mathbf{t}) + \sigma^2\mathbb{I}_N| - \frac{1}{2}(\mathbf{x} - \mu(\mathbf{t}))^T\Big(K(\mathbf{t},\mathbf{t}) + \sigma^2\mathbb{I}_N\Big)^{-1}(\mathbf{x} - \mu(\mathbf{t})) \tag{12}$$

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

$$x(t) \sim \; GP\left(r_K(t)+\sum_{k=1}^{K}m_k(t);\sum_{k=1}^{K}K_k(t,t')+2\sum_{k_1,k_2=1,k_1<k_2}^{K}K_{k1,k2}(t,t')+\sigma^2\right) \tag{13}$$

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

$$W(t) \sim \mathrm{GP}(0, K(t,t')) \tag{14}$$

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

$$B(t) = W(t) - \frac{t}{T}W(T) \tag{15}$$

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

$$\begin{cases} dB(t) = dW(t) \\ B(0) = a & ,\text{ for } 0 \le t \le T \\ B(T) = b \end{cases} \tag{16}$$

and after calculations results in the form

$$B(t) = a + (b-a)\frac{t}{T} + W(t) - \frac{t}{T}W(T) \tag{17}$$

and therefore it is a Gaussian Process

$$B(t) \sim \mathrm{GP}\Big(a + (b-a)\frac{t}{T}, K(t,t') - \frac{tt'}{T}\Big) \tag{18}$$

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