# Notes: Empirical Mode Decomposition & Gaussian Processes

*

Version: May 31, 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.

## 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

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

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 $\varphi_k$ and $\Psi_k$, respectively. Therefore, the moments of $c_k(t)$ are given in the form of the following one dimensional and two dimensional mappings

$$\begin{aligned}
&\mathbb{E}_{c_k(t)|\varphi_k}\big[c_k(t)\big] = \mu_k(t),\\
&\mathbb{E}_{c_k(t)|\Psi_k}\Big[\big(c_k(t) - \mu_k(t)\big)\big(c_k(t') - \mu_k(t')\big)\Big] = K_k(t,t')
\end{aligned}$$

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 $\mathbf{t}^{(1)}, \dots, \mathbf{t}^{(M)}$, each $N_i$ dimensional for $i \in \Big\{1, \dots, M\Big\}$ 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

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

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

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 $t_j^{(i)}$.

#### 1.1.1. MLE Estimation of the Static Parameters in Gaussian Processes Models

In the following subsection we derive the MLE estimator of the vectors of parameters $\varphi_k$ and $\Psi_k$. The log likelihood of the model is obtained for two cases

1. when we assume that the observations of $c_k(t)$ are noise free;
2. when we assume that the observations of $c_k(t)$ are contaminated with a small noise;

In order for the first assumption to hold, the $M$ paths of $c_k(t)$ at the same point $t_0$ should have the same values. The second case relaxes this assumption allowing for a degree of perturbation and unequal values at the same argument.

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

We define $K_k(\cdot, \cdot)$ as a vector operator such that

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

which is used to define the $N \times N$ Gram matrix $\mathbf{K}_k$ given by

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

**Noise-Free observations**

Under the noise free assumption and given the model in Equation (2), the loglikelihood of the model of $c_k(t)$ is given by

$$l_k\left(\mathbf{c}_k, \mathbf{t}, \varphi_k, \Psi_k\right) = -\frac{N}{2}\log 2\pi - \frac{1}{2}\log|\mathbf{K_k}| - \frac{1}{2}\mathbf{v}_k^T \mathbf{K_k}^{-1}\mathbf{v}_k \tag{4}$$

and result in the following gradient with respect to the static parameters

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

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

$$l_k\left(\mathbf{c}_k, \mathbf{t}, \varphi_k, \Psi_K\right) = -\frac{N}{2}\log 2\pi - \frac{M}{2}\log|K_k(\mathbf{t}^*, \mathbf{t}^*)| - \frac{1}{2}\sum_{i=1}^{M}\mathbf{v}^{(i)\,T} K_k(\mathbf{t}^*, \mathbf{t}^*)^{-1}\mathbf{v}^{(i)}) \tag{5}$$

**Noisy Observations**

If we suspect that the realisations of $c_k(t)$ are perturbed by a zero mean Gaussian noise with the variance $\sigma_k^2$, then the loglikelihood of the model for $c_k(t)$ is adjusted to accommodate this information as following

$$l_k\Big(\mathbf{c}_k, \mathbf{t}, \varphi_k, \Psi_k\Big) = -\frac{N}{2}\log 2\pi - \frac{1}{2}\log|\mathbf{K}_k + \sigma_k^2\mathbb{I}_N| - \frac{1}{2}\mathbf{v}_k^T\Big(\mathbf{K}_k + \sigma_k^2\mathbb{I}_N\Big)^{-1}\mathbf{v}_k \tag{6}$$

what applies as well to its simplified version, which is reformulated as

$$l_k\Big(\mathbf{c}_k, \mathbf{t}^*, \varphi_k, \Psi_k\Big) = -\frac{N}{2}\log 2\pi - \frac{M}{2}\log|K_k(\mathbf{t}^*, \mathbf{t}^*) + \sigma_k^2\mathbb{I}_{N_*}| - \frac{1}{2}\sum_{i=1}^{M}\mathbf{v}^{(i)\ T}\Big(K_k(\mathbf{t}^*, \mathbf{t}^*) + +\sigma_k^2\mathbb{I}_{N_*}\Big)^{-1}\mathbf{v}^{(i)}) \tag{7}$$

The gradient of the loglikelihood with respect to the static parameters of the model is given by

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

#### 1.1.2. Predictive Distribution of IMFs

Let $\mathbf{s}$ represent a set of points where we want to predict IMFs given the models estimated in the previous subsection. The predictive distribution of $c_k$ on a new set of points denoted by $\mathbf{s}$, which is conditioned on the observed information and assuming that there is an observation error, is Gaussian with the conditional mean

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

and the conditional covariance matrix given by

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

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

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

## 1.2. Multikernel Representation of the Signal

#### 1.2.1. 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 $\sigma^2$, that is

$$x(t) = \sum_{k=1}^{K} c_k(t) + r_K(t) + \epsilon \tag{8}$$

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

$$x(t) \sim \; GP\Big( r_K(t) + \sum_{k=1}^{K} \mu_k(t); \sum_{k=1}^{K} K_k(t,t') + \sigma^2 \Big) \tag{9}$$

The scalar $\sigma^2$ can be estimated by MLE of $x(t)$, given its $M$ realization, $\mathbf{x}^{(i)}$, formed into a vector $\mathbf{x} = [\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(M)}]$. 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

$$l(\mathbf{x},\mathbf{t},\sigma^2) = -\frac{N}{2}\log 2\pi - \frac{1}{2}\log|K(\mathbf{t},\mathbf{t}) + \sigma^2\mathbb{I}_N| - \frac{1}{2}(\mathbf{x}-\mu(\mathbf{t}))^T \Big(K(\mathbf{t},\mathbf{t}) + \sigma^2\mathbb{I}_N\Big)^{-1}(\mathbf{x}-\mu(\mathbf{t})) \tag{10}$$

with corresponding gradient

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

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

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

#### 1.2.2. 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

$$x(t) \sim \; GP\Big( r_K(t) + \sum_{k=1}^{K} m_k(t); \sum_{k=1}^{K} K_k(t,t') + 2\sum_{k_1,k_2=1,k_1<k_2}^{K} K_{k1,k2}(t,t') + \sigma^2 \Big) \tag{11}$$

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

$$W(t) \sim \mathrm{GP}(0, K(t,t')) \tag{12}$$

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

$$B(t) = W(t) - \frac{t}{T}W(T) \tag{13}$$

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

$$\begin{cases} dB(t) = dW(t) \\ B(0) = a \qquad , \text{ for } 0 \le t \le T \\ B(T) = b \end{cases} \tag{14}$$

and after calculations results in the form

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

and therefore it is a Gaussian Process

$$B(t) \sim \mathrm{GP}\Big(a + (b-a)\frac{t}{T}, K(t,t') - \frac{tt'}{T}\Big) \tag{16}$$

for $0 \le t \le T$

## 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