A Statistical Perspective of the Empirical Mode Decomposition

Author: Marta Campi [1],

Supervisors: Gareth W. Peters [2], Matina Rassias[1], Nourddine Azzaoui [3]

Collaborators: Tomoko Matsui [4], Dorota Toczydlowska [5], Kylie-Anne Richards [5]

[1] University College London (UCL) - London, UK

[2] University of California, Santa Barbara - Santa Barbara, US

[3] Laboratoire de Mathemtiques Blaise Pascal (LBMP) - Clermont Ferrand, France

[4] The Institute of Statistical Mathematics (ISM) - Tokyo, Japan

[5] University of Technology (UTS) - Sydney, Australia

March 18, 2022 - Viva Examination

Outline

Outline

Journal Papers & International Visits

- **Published**. "Machine Learning Mitigants for Biometric Cyber Risk", Marta Campi, Gareth W. Peters, Nourddine Azzaoui, IEEE Access, Volume 9, pages:136831 - 136860.
- **In submission**
  - "Stochastic Embedding Approaches for Empirical Mode Decomposition in Parkinson's Speech Detection", Marta Campi, Dorota Toczydlowska, Gareth W. Peters, Pascal Bianchi (IEEE Transactions on Signal Processing)
- **In Preparation**:
  - "Spectral Representation Learning For Real-Valued Signals by Complex Gaussian Process', Dorota Toczydlowska, Marta Campi, Gareth W. Peters, Pascal Bianchi.
  - "Green Bond Performance and Risk Indicators", Marta Campi, Kylie-Anne Richards, Gareth W. Peters. (Annals of Actuarial Science)

Journal Papers & International Visits

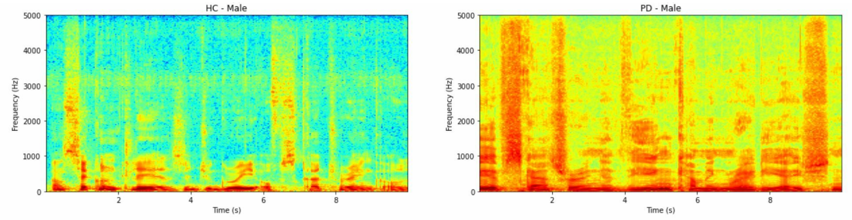
- Heriot-Watt, Business School, Visited Prof. Dimitris Christopolous, Edinburgh, Scotland, January 2022
- Heriot-Watt, Department of Actuarial Mathematics & Statistics, Visited Prof. Gareth W. Peters, Edinburgh, Scotland, June 2019
- Institute of Statistical Mathematics (ISM), Department of Statistical Modeling, Visited Prof. Tomoko Matsui, Tokyo, Japan, February 2018
- Laboratoire de Mathématiques Blaise Pascal, Université Clermont Auvergne, Visited Prof. Nourddine Azzaoui, Clermont-Ferrand, France, July 2017
- 47-th Probability Summer School Saint-Flour, Saint-Flour, France, July 2017

Outline

Objectives and Problem Statements

Stochastic Embedding Approaches for Empirical Mode Decomposition

Motivation : Perform inference and estimation procedures at different time scales or frequency components obtained from a stochastic representation of a data-driven a posteriori decomposition method.



The development of this framework requires the set up of the next components:

- **Set of Objectives**
- **Problem Statements**

Application Objectives

- Determine presence or absence of Parkinson's disease (PD) through voice recordings.
- For the sick patients, develop a surveillance tool by establishing their PD stage and the disease progression from voice recordings.

The achievement of the above objectives will require the development of the followings methodological objectives.

Methodological Objectives

- Setting up an inference procedure with these attributes:
  - → Statistical test that can be evaluated pointwise and known in closed form.
  - → Asymptotic distribution of the test statistics known in closed form.
  - → Computationally efficient procedure.
- Developing a statistical framework for the speech signal that would be amenable to the above inference procedure.

Statistical Framework for Deterministic EMD

Employ the decomposition method known as **Empirical Mode Decomposition** (EMD) [HSL$^+$98] that has the following properties: (1) **non-linear**, (2) **non-stationary**, (3) **data-driven**, (4) **not-constructive**.

Problem Statement 1 (PS1)

Identify a basis representation that captures the following characteristics for the deterministic, pathwise decomposition:

- Smooth
- Differentiable
- Continuous
- approximated by the class of polynomial basis in the L^2 space, using spline functions.
- Admits closed form of the Hilbert transform
- finite (but unknown) number of basis

A smooth spline formulation for the basis functions representations will provide the solution to PS1.

Statistical Framework for Stochastic EMD

Up until this work, the EMD has only been explored from an algorithmic deterministic, pathwise point of view. The second problem statement aims to introduce **a stochastic version** of this method, whose additional characteristics are presented below.

Problem Statement 2 (PS2)

Formulate a stochastic embedding of the basis representation obtained from EMD that is consistent at process level and observed basis levels according to the following properties

- closed on the convolution
- infinite divisibility
- closed form of finite dimensional distributions
- knowledge of sufficient statistics functions in the form of the mean and the covariance of the considered stochastic process

The solution to PS2 will be provided by embedding the EMD into a Gaussian Process framework.

Adaptive Gaussian Kernel Design through Optimal Time-Frequency EMD Partitions

Standard Gaussian process kernels often impose restrictions (as isotropy or stationarity) that will not suit the developed settings. The employed kernels need to be **adaptive** with time-varying hyperparameters or a more flexible kernel structure.

There are multiple solution to achieve such a result. In this work, the time-series kernel known as **the Fisher Kernel** is employed to provide adaptivity.

However, this kernel should be applied on partitions of time where local stationarity holds again. The following problem statement deals with how to construct a sound local partition.

Problem Statement 3 (PS3)

Identify an optimal time-frequency partition for a formulation of novel stochastic embedding. This is achieved through the following steps:

- Partition Rule Definition
- Formulation of the Optimisation Problem

The solution to PS3 will be provided by the Cross-Entropy Method.

UCL

Statistical Framework for EMD

PS1 **Deterministic EMD**

$S(t) : \Omega \to \mathbb{R}$ $s(t), t \in [0, T]$

•**Assume** $S(t)$ carrying any of these attributes

(1) non-stationary (weak or strong sense)

(2) non-linear (mean or covariance)

(3) with discontinuities

(4) not Markov

•**Assume** an optimal $\tilde{s}(t)$ $s(t) = \tilde{s}(t) + e(t) \quad t \in [0, T]$

$S(t) \overset{d}{=} \tilde{S}(t) + \epsilon(t)$

$\tilde{S}(t) : \Omega \to \mathbb{R}$ $\tilde{s}(t) = \sum_{l=1}^{L+1} \gamma_l(t)$

$\Gamma_1(t) : \Omega_1 \to \mathbb{R}$ $\gamma_1(t)$

$\Gamma_2(t) : \Omega_2 \to \mathbb{R}$ $\gamma_2(t)$

$\dots$ $\dots$

$\Gamma_{L+1}(t) : \Omega_{L+1} \to \mathbb{R}$ $\gamma_{L+1}(t)$

GOAL PS1: Find a statistical model s.t. $\gamma_l(t)$ is **smooth, differentiable, continuous** and can be **approximated** by **simple functions** or **spline polynomial functions**.

Natural Cubic Spline Formulation

$\tilde{s}(t) = a_0 + a_1 t + a_2 t^2 + a_3 (t - \tau_1)_+^3 + \cdots + a_{3+l-2} (t - \tau_{l-1})_+^3$

•Optimal repres. in closed form

• Hilbert Transform in closed form

PS2 **Stochastic Embedding of the EMD**

GOAL PS2: Characterize $\tilde{S}(t)$ through properties of $\{\Gamma_l\}_{l=1}^{L+1}$ induced by $\{\gamma_l(t)\}_{l=1}^{L+1}$.

This is equivalent to find a representation of $\tilde{S}(t)$ s.t.

(1) $\tilde{S}(t)$ is **closed under convolution**

(2) $\tilde{S}(t)$ is **infinite divisible** $\forall t$

(3) $\tilde{S}(t)$ has **closed form of the marginals**

(4) $\tilde{S}(t)$ has **sufficient statistics** given by **the mean** and **covariance functions.**

Ass. 0: Gaussian Processes satisfy (1) - (4).

Ass. 1: $\tilde{S}(t) \overset{d}{=} \sum_{l=1}^{L} \Gamma_l(t) + R(t)$

Ass. 2: $\tilde{S}(t) \sim \mathcal{GP}\left(\mu(t; \theta_\mu(t)); k(t, t'; \theta_k(t))\right)$

$\Gamma_l(t) \sim \mathcal{GP}\left(\mu(t, \theta_{\mu_l}); k_l(t, t'; \theta_l)\right)$

$R(t) \sim \mathcal{GP}\left(\mu(t; \theta_{\mu_{l+1}}); k_{l+1}(t, t'; \theta_{l+1})\right)$

A Multi-Kernel Representation of the EMD

$\tilde{S}(t) \sim \mathcal{GP}\left(\sum_{l=1}^{L+1} \mu_l(t, \theta_{\mu_l}); \sum_{l=1}^{L+1} k_l(t, t'; \theta_{k_l})\right)$

Outline

1. Journal Papers & International Visits
2. Objectives and Problem Statements
3. Deterministic EMD
4. Stochastic Embedding EMD Models
5. The Optimal Time-Frequency EMD Partition

Problem Statement 1 (PS1)

Consider the stochastic process $S(t) \in \mathbb{R}$ with $t \in \{-\infty, +\infty\}$, whose observed realisation is denoted as $s(t)$ and it is partially or discretely observed at $0 = t_1 < \cdots < t_N = T$.

$S(t)$ carries any of the following attributes: non-stationary (weak or strong sense), non-linear (in mean or covariance), potentially with discontinuities, potentially not Markov.

Goal: Given an observed trajectory $s(t)$ of $S(t)$ and its finite, deterministic, pathwise EMD decomposition given as

$$s(t) = \sum_{l=1}^{L+1} \gamma_l(t) \tag{3.1}$$

identify a statistical model representation of $s(t)$ $\gamma_l(t)$ which satisfies the following properties:

- Smooth
- Differentiable
- Continuous
- approximated by polynomial basis in L^2 using splines functions.
- Admits closed form of the Hilbert transform
- finite (but unknown) number of bases

(1) Solution to PS1

A natural cubic spline provides a solution to PS1. The following steps are required and reviewed within this section.

- The discrete signal $s(t)$ is approximated by a natural cubic spline denoted as $\tilde{s}(t)$, which is optimal since it is the smoothest function amongst all the continuously differentiable functions (up to order 2) interpolating $s(t)$ in $[0, T]$.
- $\{\gamma_l(t)\}_{l=1}^{L}$ will also be represented by a cubic spline since recursively extracted on $\tilde{s}(t)$.
- The Hilbert transform of each $\gamma_l(t)$ represented by a cubic spline can be derived in closed form and provides the instantaneous frequency in closed form. This is highly beneficial for the proposal of the stochastic embeddings.

Stochastic process

A univariate real-valued stochastic process defined as $\{\ldots S(t_k - h), \ldots, S(t_1), S(t_2), \ldots S(t_k + h), \ldots\} = S(t)$ for $t \in \{-\infty, \infty\}$, for all $k \in \mathbb{N}, h \in \mathbb{R}$ is a sequence of random variables indexed by time t.

Which methods are available to decompose and analyse such stochastic processes?

Deterministic time	Stochastic time
Deterministic time-frequency	Stochastic time-frequency

Fourier transform

$$h(f) = \int_{-\infty}^{+\infty} s(t)e^{-j2\pi ft}dt$$

where $h(f) = \mathcal{F}[s(t)]$ is such that $h(f) : \mathbb{R} \rightarrow \mathbb{C}$ for any $f \in \mathbb{R}$.

Deterministic EMD

- **infinite number** of basis
- **parametric** structure
- **a-priori** basis
- **stationary/linear** data system

What happens if some of these assumptions do not hold?

Consider the deterministic path $s(t)$ associated with the stochastic process $S(t)$, observed at $0 = t_1 < \cdots < t_N = T$. For the EMD to exist, $s(t)$ must be converted into a continuous analog signal denoted $\tilde{s}(t)$.

The semi-parametric model known as a natural cubic spline will be used. As a consequence, the EMD bases denoted as $\{\gamma_l(t)\}_{l=1}^{L}$ will also be expressed as natural cubic splines, derived from representation $\tilde{s}(t)$.

Natural Cubic Spline

Given a set of l knots $a = \tau_1 < \tau_2 < \cdots < \tau_l = b$, a function $\tilde{s} : [a, b] \to \mathbb{R}$ is called a cubic polynomial spline if:

- $\tilde{s}(\cdot)$ is a polynomial of degree 3 on each interval (τ_j, τ_{j+1}) $(j = 1, \ldots, l - 1)$
- $\tilde{s}(\cdot)$ is twice continuously differentiable

It is then a natural cubic spline when $\tilde{s}''(a) = \tilde{s}''(b) = 0$.

Deterministic EMD

The signal representation $\tilde{s}(t)$ is expressed in the class of truncated power basis, where the knot points are placed at the sampling times ($\tau_i = t_i$)

$$\tilde{s}(t) = a_0 + a_1 t + a_2 t^2 + a_3 (t - \tau_1)_+^3 + \cdots + a_{3+l-2} (t - \tau_{l-1})_+^3 . \qquad (3.2)$$

The coefficients are estimated by standard penalised least squares

$$\sum_{i=1}^{N-1} (s(t_i) - \tilde{s}(t_i))^2 + \lambda \int_{t_i}^{t_{i+1}} \tilde{s}''(t)^2 dt \qquad (3.3)$$

with natural cubic spline constraints $\tilde{s}''(0) = \tilde{s}''(t_N) = 0$ and where $\lambda > 0$. In this case, the number of total convexity changes (oscillations) of the analog signal $\tilde{s}(t)$ within the time domain $[0, t_N]$ is denoted by to $L \in \mathbb{N}$.

Is there a finite basis decomposition admitting meaningful time and frequency domain interpretation in such a framework?
Ans: YES [HSL$^+$98]

Empirical Mode Decomposition

The Empirical Mode Decomposition of signal $\tilde{s}(t)$ is represented by the finite number of non-stationary basis functions known as **Intrinsic Mode Function** (IMFs), denoted by $\{\gamma_l(t)\}$, such that

$$\tilde{s}(t) = \sum_{l=1}^{L} \gamma_l\,(t) + r\,(t) = \left(\sum_{l=1}^{L+1} \gamma_l(t), \quad \gamma_{l+1}(t) = r(t) \right) \tag{3.4}$$

where $r\,(t)$ represents the **residual** (or final tendency) extracted, which has only a single convexity.

Define the interpolations of maxima and minima of $\tilde{s}(t)$ as **upper envelope** $\tilde{s}^U(t)$ and **lower envelope** $\tilde{s}^L(t)$ respectively and the **mean envelope** as $m_l(t) = \left(\frac{\tilde{s}^U(t)+\tilde{s}^L(t)}{2}\right)$.

The IMF basis $\{\gamma_l(t)\}_{l=1}^{L}$ are designed to satisfy the following two conditions:

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

$$\text{abs}\left(\left|\left\{\frac{d\gamma_l(t)}{dt} = 0 : \quad t \in (t_1, t_N)\right\}\right| - |\{\gamma_l(t) = 0 : \\ t \in (t_1, t_N)\}|\right) \in \{0, 1\} \tag{3.5}$$

- **Local Symmetry** The local mean value of the envelope defined by a spline through the local maxima denoted $\tilde{s}^{U_l}(t)$ and the envelope defined by a spline through the local minima denoted by $\tilde{s}^{L_l}(t)$ is equal to zero pointwise i.e.

$$m_l(t) = \left(\frac{\tilde{s}^{U_l}(t) + \tilde{s}^{L_l}(t)}{2}\right) I\left(t \in [t_1, t_N]\right) = 0 \tag{3.6}$$

Note that, The minimum requirements of the upper and lower envelopes are:

$$
\begin{aligned}
&\tilde{s}^{U_l}(t) = \gamma_l(t), \quad \textit{if} \quad \frac{d\gamma_l(t)}{dt} = 0 \quad \& \quad \frac{d^2\gamma_l(t)}{dt^2} < 0, \\
&\tilde{s}^{U_l}(t) \geq \gamma_l(t) \quad \forall t \in (t_1, t_N) \\
&\tilde{s}^{L_l}(t) = \gamma_l(t), \quad \textit{if} \quad \frac{d\gamma_l(t)}{dt} = 0 \quad \& \quad \frac{d^2\gamma_k(t)}{dt^2} > 0, \\
&\tilde{s}^{L_l}(t) \leq \gamma_l(t) \quad \forall t \in (t_1, t_N).
\end{aligned}
\tag{3.7}
$$

The residual $r(t)$ is designed to satisfy the following condition:

The residual is a curve with at most one extremum.

$$
\begin{aligned}
&osc(r(t)) \in \{0, 1\} \\
&r'(t) \geq 0 \quad \text{or} \quad r'(t) < 0 \quad \text{over} \quad [0, T]
\end{aligned}
$$

Instantaneous frequency

$$\omega_l(t) = \frac{1}{2\pi}\frac{d\theta_l(t)}{dt}$$

An analytical signal is defined as $z_l(t) = \gamma_l(t) + j\check{\gamma}_l(t)$ or $z_l(t) = a_l(t)e^{j\theta_l(t)}$ where

- $a_l(t)$ is the **amplitude** of $z_l(t)$
- $\theta_l(t) = arctan\frac{\check{\gamma}_l(t)}{\gamma_l(t)}$ is **the instantaneous phase.**

Hilbert transform (Cauchy Principal Value Integral)

$$\check{\gamma}_l(t) = \frac{1}{\pi}\lim_{\epsilon\to\infty}\int_{-\epsilon}^{+\epsilon}\frac{\gamma_l(\tau)}{t-\tau}d\tau$$

Deterministic EMD

By considering the natural cubic spline representation for $\gamma_l(t)$ per knot segmentation as a local cubic polynomial for $t \in [\tau_{i-1}, \tau_i]$, then the Hilbert transform is constructed as

$$\check{\gamma}_l(t) = \mathcal{HT}\left[\gamma_l(\tau)\right] = \frac{1}{\pi}\sum_{i=1}^{N-1}\check{\gamma}_{l_i}(t) \quad \tau_{i-1} < t \leq \tau_i \tag{3.8}$$

where $\triangle_i = \tau_i - \tau_{i-1}$ and $\check{\gamma}_{l_i}(t)$ is the Hilbert transform of the i-th polynomial:

$$\begin{aligned}\check{\gamma}_{l_i}(t) &= \left(a_{l_i}t^3 + b_{l_i}t^2 + c_{l_i}t + d_{l_i}\right)\log\left(\frac{t}{t-\triangle_i}\right) \\ &+ a_{l_i}\left(\frac{\triangle_i^2 t}{2} - \triangle_i t^2 - \frac{\triangle_i^3}{3}\right) + b_{l_i}\left(-\triangle_i t - \frac{\triangle_i^2}{2}\right) - c_{l_i}\triangle_i.\end{aligned} \tag{3.9}$$

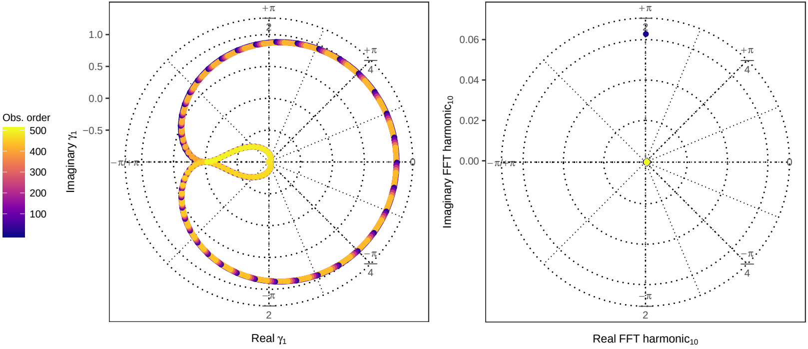
See for details [eMH20].

Hence, with the optimal representation of $\gamma_l(t)$ provided by a cubic spline, the instantaneous frequency can be computed in closed form.

Deterministic EMD

After the EMD and the HHT of the IMFs are computed, $\tilde{s}(t)$ can be expressed as:

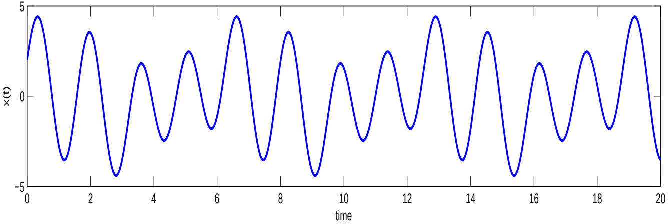
$$\tilde{s}(t) = Re\left\{\sum_{l=1}^{L+1} a_l(t)\exp\{\jmath\,\theta_l(t)\}\right\} = Re\left\{\sum_{l=1}^{L+1} a_l(t)\exp\{\jmath\int_{t_1}^{t_N} 2\pi\omega_l(t)dt\}\right\}$$



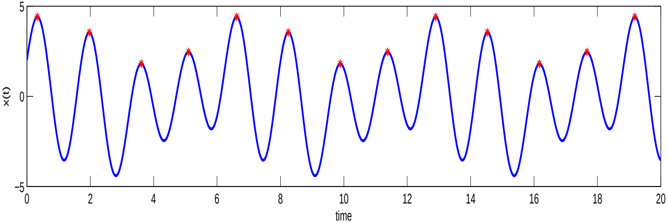
Deterministic EMD

1. **Step 1** Find local extrema of $\tilde{s}(t)$
2. **Step 2** Compute the upper envelope $\tilde{s}^U(t)$ and the lower envelope $\tilde{s}^L(t)$ by employing spline interpolations (cubic, akima, b-spline, etc.)
3. **Step 3** Update the signal $\tilde{s}(t) \leftarrow \tilde{s}(t) - \frac{\tilde{s}^U(t)+\tilde{s}^L(t))}{2}$
4. **Step 4** Repeat 1, 2 and 3 until achieving an IMF $\gamma_l(t)$
5. **Step 5** Subtract the obtained IMF from the signal $\tilde{s}(t) \leftarrow \tilde{s}(t) - \gamma_l(t)$
6. **Step 6** Repeat 1-5 until achieving a tendency

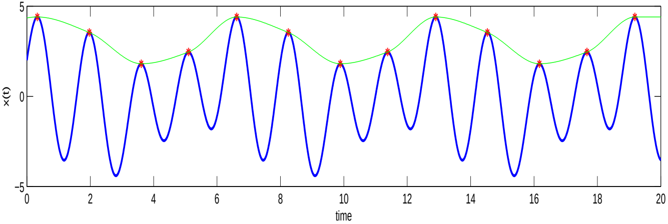
Deterministic EMD



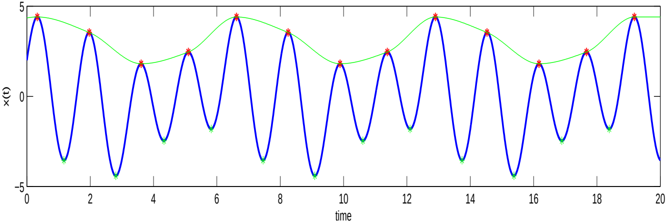
Deterministic EMD



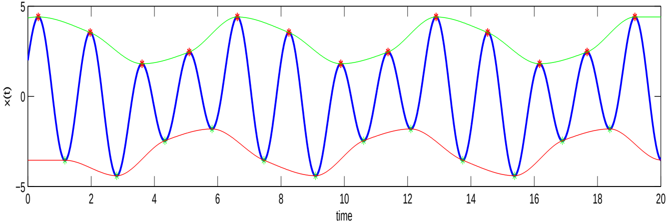
Deterministic EMD



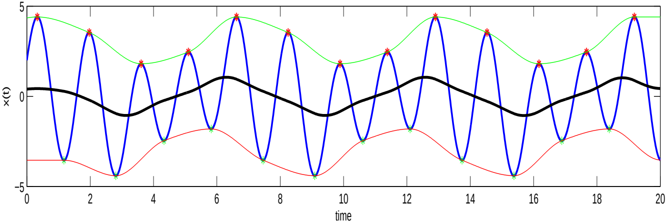
Deterministic EMD



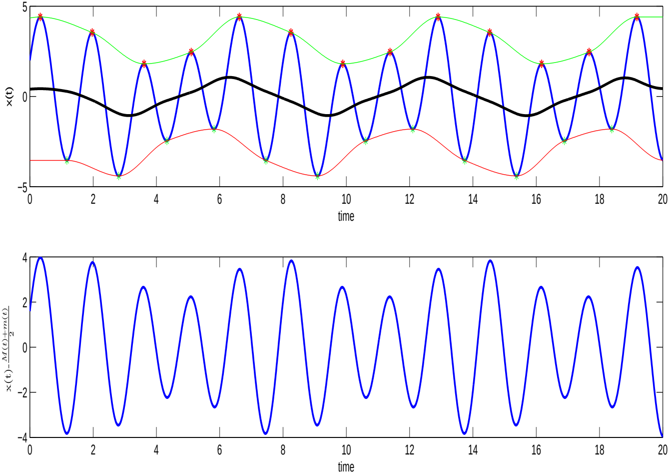
Deterministic EMD



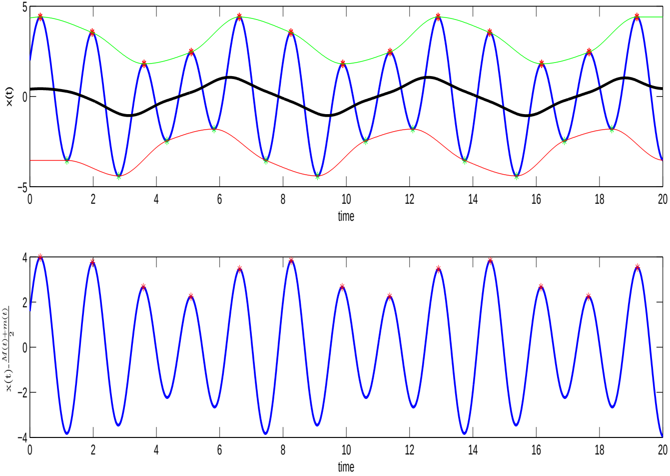
Deterministic EMD



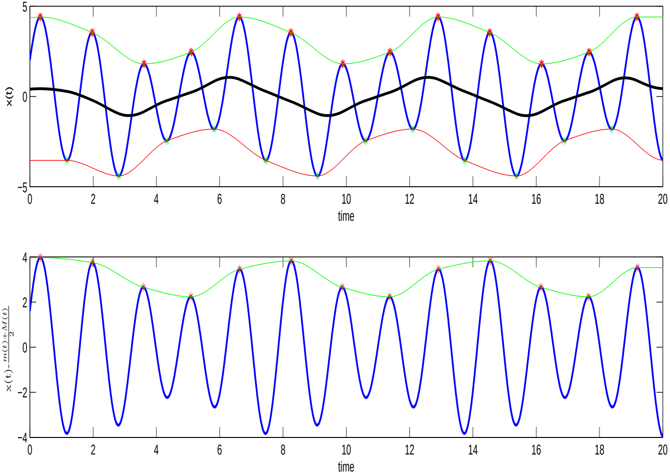
Deterministic EMD



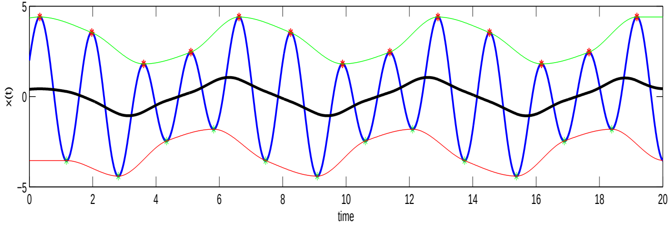
Deterministic EMD

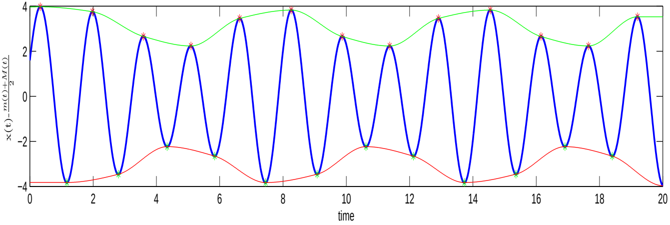


Deterministic EMD

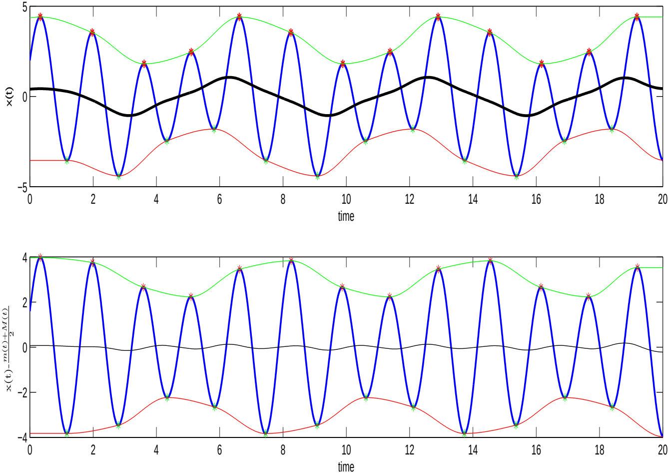


Deterministic EMD





Deterministic EMD



Outline

Problem Statement 2 (PS2)

(PS2) Obtain a stochastic representation of $S(t)$ that has the following properties:

- infinitely divisible at all time points
- closed under convolution
- closed forms of the marginal distributions
- knowledge of the sufficient statistics functions given by the mean and covariance functions.

(1) Solution to PS2

The class of Gaussian Processes carries the above properties and is selected for the stochastic embedding of $\tilde{s}(t)$ and $\{\gamma_l(t)\}_{l=1}^{L+1}$. The following steps are required and reviewed within this section.

- The goal is to formulate a stochastic representation that is consistent with the basis function representations provided by PS1 .
- This is established through the employment of the Gaussian process since there is an isometric isomorphism between the reproducing kernel of the smooth spline and the covariance function of the Gaussian process.
- Equality of the distribution family between the stochastic process of $\tilde{s}(t)$ and the stochastic process of $\{\gamma_l(t)\}_{l=1}^{L+1}$ will be considere, i.e. they will all be given as GPs.
- The RKHS of the stochastic processes denoted as $\mathcal{H}_{\tilde{S}}, \mathcal{H}_{\Gamma_2}, \ldots, \mathcal{H}_{\Gamma_{L+1}}$ are ordered according to the smoothness expressed by their reproducing property which is linked to the oscillations induced by the realised observed trajectories $\tilde{s}(t)$, $\{\gamma_l(t)\}_{l=1}^{L+1}$.

Stochastic Embedding EMD Models

The EMD is applied to the approximating signal $\tilde{s}(t)$, which can be given as

$$s(t) = \tilde{s}(t) + e(t) \tag{4.1}$$

where $e(t)$ represents the observed error at $t \in [0, T]$ and t is not a knot point. Hence, the stochastic embedding will be formulated for $\widetilde{S}(t)$, for which the following holds:

$$S(t) \stackrel{d}{=} \widetilde{S}(t) + \epsilon(t) \tag{4.2}$$

The two processes are equal in the interval $[0, T]$ at the knot points of the interpolated continuous signal $\tilde{s}(t)$; however, at all the the other points, this will not be the case and this will result into a residual error $\epsilon(t)$.

When the EMD is applied to $\widetilde{s}(t)$, each $\gamma_l(t)$ can be considered as the realised path of the stochastic process denoted as $\Gamma_l(t)$ and the one for the residual $r(t)$ denoted as $R(t)$. Hence, one will be given the following stochastic processes

$$\widetilde{S}(t), \Gamma_1(t), \ldots, \Gamma_L(t), R(t)$$

The aim is to construct a model given the relationship in distribution of the above sequence.

Stochastic Embedding EMD Models

The first step is to consider the following **assumption**:

$$\widetilde{S}(t) \overset{d}{=} \sum_{l=1}^{L} \Gamma_l(t) + R(t) \tag{4.3}$$

If $\widetilde{S}(t)$ was stationary then

$$\begin{gathered}\tilde{S}(t) \sim \mathcal{GP}\left(\mu(t;\boldsymbol{\theta}_\mu); k(t,t';\boldsymbol{\theta}_k)\right) \\ \mu(t;\boldsymbol{\theta}_\mu) = \sum_{l=1}^{L} \mu(t;\boldsymbol{\theta}_{\mu_l}), \quad k(t,t';\boldsymbol{\theta}_k) = \sum_{l=1}^{L} k_l(t,t';\boldsymbol{\theta}_l)\end{gathered} \tag{4.4}$$

However, $\widetilde{S}(t)$ is **assumed** non-stationary and can be decomposed into the sum of a finite number L non-stationary basis functions which are the IMFs. This is equivalent to say that $\widetilde{S}(t)$ has a formulation given as

$$\tilde{S}(t) \sim \mathcal{GP}\left(\mu(t;\boldsymbol{\theta}_\mu(t)); k(t,t';\boldsymbol{\theta}_k(t))\right) \tag{4.5}$$

where $\mu(t;\boldsymbol{\theta}_\mu(t))$, $k(t,t';\boldsymbol{\theta}_k(t))$ are both non-stationary.

Goal: is to model $\widetilde{S}(t)$ as the sum of multiple stochastic processes whose structure can be derived with the EMD.

The next **assumption** required to define the desired stochastic embedding is that each process $\Gamma_l(t)$ will also be represented as a Gaussian process such that

$$\Gamma_l(t) \sim \mathcal{GP}\left(\mu(t, \boldsymbol{\theta}_{\mu_l}); k_l(t, t'; \boldsymbol{\theta}_l)\right), \tag{4.6}$$

The **assumption** on the stochastic process of the residual is given as follows:

$$R(t) \sim \mathcal{GP}\left(\mu(t; \boldsymbol{\theta}_{\mu_{l+1}}); k_{l+1}(t, t'; \boldsymbol{\theta}_{l+1})\right) \tag{4.7}$$

A Multi-Kernel Representation of the EMD

$$\widetilde{S}(t) \sim \mathcal{GP}\left(\sum_{l=1}^{L+1} \mu_l(t, \boldsymbol{\theta}_{\mu_l}); \sum_{l=1}^{L+1} k_l(t, t'; \boldsymbol{\theta}_{k_l})\right) \quad (4.8)$$

where the mean and the kernel functions are given as the sum of the L mean and kernel functions of the stochastic processes modelling the IMFs and the residual tendency. The models for the stochastic process of the IMFs and the residual are

$$\begin{aligned} \Gamma_l(t) &\sim \mathcal{GP}\left(\mu(t, \boldsymbol{\theta}_{\mu_l}); k_l(t, t'; \boldsymbol{\theta}_l)\right) \text{ for } l = 1, \ldots, L \\ R(t) &\sim \mathcal{GP}\left(\mu(t; \boldsymbol{\theta}_{\mu_{l+1}}); k_{l+1}(t, t'; \boldsymbol{\theta}_{l+1})\right) \end{aligned} \quad (4.9)$$

Stochastic Embedding EMD Models

Three system models are now proposed based on the above formulation.

The first model assumes that the original observed signal $\tilde{s}(t)$ is obtained from a stochastic process $\tilde{S}(t)$ which is a GP given as follows:

System Model 1

$$\tilde{S}(t) \sim \mathcal{GP}\left(\mu(t;\boldsymbol{\theta}_\mu); k(t,t';\boldsymbol{\theta}_k)\right) \quad (4.10)$$

where $\mu(t;\boldsymbol{\theta}_\mu)$ and $k(t,t';\boldsymbol{\theta}_k)$ represent the mean and kernel functions respectively, $\boldsymbol{\theta}_\mu$ and $\boldsymbol{\theta}_k$ are the sets of hyperparameters of the mean and the kernel respectively.

Note that the zero mean assumption is considered and, therefore, one has $\mu(t;\boldsymbol{\theta}_\mu) = 0$.

Stochastic Embedding EMD Models

Consider $\tilde{s}(t)$ decomposed into IMFs basis functions denoted as $\{\gamma_l(t)\}_{l=1}^{L}$.
The second model assumes that each $\gamma_l(t)$ is observed from a Gaussian Process $\Gamma_l(t)$ and then reconstructs the original $\tilde{s}(t)$ by summing up the IMFs.

$$\gamma_1(t) \qquad \Gamma_1(t) \sim \mathcal{GP}(\mu_1(t;\boldsymbol{\theta}_{\mu_1}); k_1(t,t';\boldsymbol{\theta}_1))$$

$$\tilde{s}(t) \qquad \ldots$$

$$\gamma_L(t) \qquad \Gamma_L(t) \sim \mathcal{GP}(\mu_L(t;\boldsymbol{\theta}_{\mu_L}); k_L(t,t';\boldsymbol{\theta}_L))$$

System Model 2

$$\widetilde{S}(t) \stackrel{d}{=} \sum_{l=1}^{L} \Gamma_l(t) \sim \mathcal{GP}\left(\mu(t;\boldsymbol{\theta}_\mu), k(t,t';\boldsymbol{\theta}_k)\right), \tag{4.11}$$

where $\mu(t;\boldsymbol{\theta}_\mu) = \sum_{l=1}^{L} \mu_l(t;\boldsymbol{\theta}_{\mu_l})$ and $k(t,t';\boldsymbol{\theta}_k) = \sum_{l=1}^{L} k_l(t,t';\boldsymbol{\theta}_l)$.

Stochastic Embedding EMD Models

Consider the extracted instantaneous frequencies $\{\omega_l\}_{l=1}^{L}$. A construction of System Model is given in the following diagram:

$$
\begin{array}{ccccccc}
 & \gamma_1(t) \rightarrow & \omega_1(t) & & & & \Gamma_1^{(BL)}(t)|\Pi = \hat{\Pi} \sim \mathcal{GP}(\mu_1^{BL}(t;\boldsymbol{\theta}_{\mu_1^{BL}}), k_1^{BL}(t,t';\boldsymbol{\theta}_{k_1^{BL}})) \\
 \nearrow & & & & & & \\
\tilde{s}(t) & \dots & & \longrightarrow \quad \hat{\Pi} \longrightarrow & \dots & & \dots \\
 \searrow & & & & & & \\
 & \gamma_L(t) \rightarrow & \omega_L(t) & & & & \Gamma_M^{(BL)}(t)|\Pi = \hat{\Pi} \sim \mathcal{GP}(\mu_M^{BL}(t;\boldsymbol{\theta}_{\mu_M^{BL}}), k_M^{BL}(t,t';\boldsymbol{\theta}_{k_M^{BL}}))
\end{array}
$$

System Model 3 can be constructed once the partition $\hat{\Pi}$ is obtained. The model is given by the IF being in a certain interval frequency band. The partition $\hat{\Pi}$ is estimated by a realisation of the stochastic process $\widetilde{S}(t)$ and, therefore, is conditioned upon it. This is stated in the distributional assumption of the different processes $\Gamma_m^{(BL)}$ which are indeed given as $\Gamma_m^{(BL)}|\Pi = \hat{\Pi}$.

Formally, by considering $\{\gamma_l(t)\}_{l=1}^{L}$, $\{\omega_l(t)\}_{l=1}^{L}$ and the partition Π, the following set of aggregated IMFs are obtained

System Model 3

$$\begin{cases} \gamma_1^{(BL)}(t) = \gamma_1(t)\mathbb{1}_{\left\{\omega_1(t)\in\bigcup_{d=1}^{D}\Pi_{1,d}\right\}} + \ldots + \gamma_K(t)\mathbb{1}_{\left\{\omega_L(t)\in\bigcup_{d=1}^{D}\Pi_{1,d}\right\}} \\ \gamma_2^{(BL)}(t) = \gamma_1(t)\mathbb{1}_{\left\{\omega_1(t)\in\bigcup_{d=1}^{D}\Pi_{2,d}\right\}} + \ldots + \gamma_K(t)\mathbb{1}_{\left\{\omega_L(t)\in\bigcup_{d=1}^{D}\Pi_{2,d}\right\}} \\ \vdots \\ \gamma_M^{(BL)}(t) = \gamma_1(t)\mathbb{1}_{\left\{\omega_1(t)\in\bigcup_{d=1}^{D}\Pi_{M,d}\right\}} + \ldots + \gamma_K(t)\mathbb{1}_{\left\{\omega_L(t)\in\bigcup_{d=1}^{D}\Pi_{M,d}\right\}} \end{cases}$$

which results in construction of $\gamma_m^{(BL)}(t)$ for $m = 1, \ldots, M$ such that

$$\forall_m \; \gamma_m^{(BL)}(t) = \sum_{l=1}^{L} \gamma_l(t)\mathbb{1}_{\left\{\omega_l(t)\in\bigcup_{d=1}^{D}\Pi_{m,d}\right\}} \tag{4.12}$$

Each $\gamma_m^{(BL)}(t)$ is embedded within a Gaussian Process, Γ_m^{BL}. Therefore, the original signal will then correspond to

System Model 3

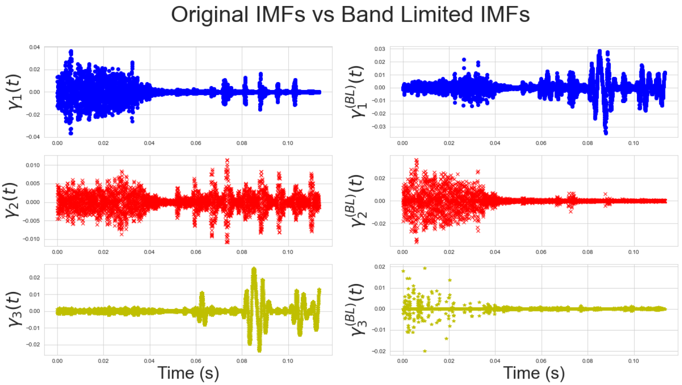
$$\tilde{s}(t) = \sum_{m=1}^{M-1} \gamma_m^{(BL)}(t) = \sum_{l=1}^{L} \gamma_l(t) \tag{4.13}$$

and the stochastic process $\widetilde{S}(t)$ is represented via multi-kernel representation exploiting $\Gamma_m^{BL}(t)$, that is

$$\widetilde{S}(t) \stackrel{d}{=} \sum_{m=1}^{M} \Gamma_m^{BL}(t) \sim \mathcal{GP}\left(\mu_s(t;\boldsymbol{\theta}_{\mu_s}), k_s(t,t';\boldsymbol{\theta}_{k_s})\right), \tag{4.14}$$

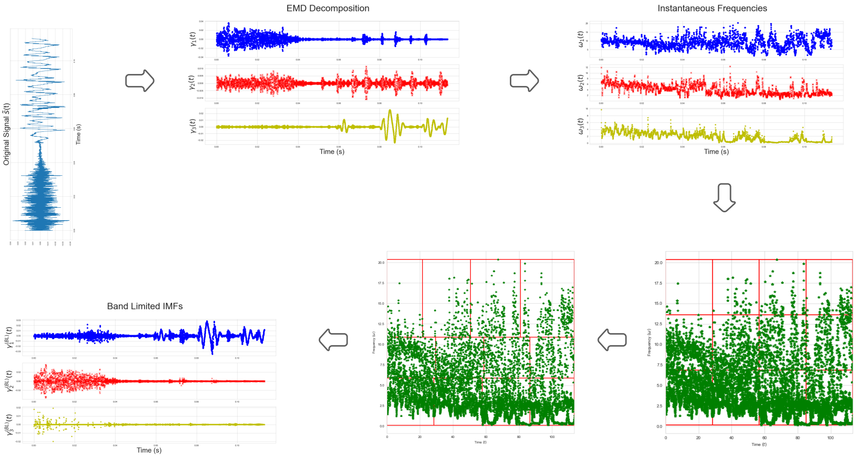
where $\mu_s(t;\boldsymbol{\theta}_{\mu_s}) = \sum_{m=1}^{M} \mu_m^{BL}(t)$ and $k_s(t,t';\boldsymbol{\theta}_{k_s}) = \sum_{m=1}^{M} k_m^{BL}(t,t';\boldsymbol{\theta}_{k_M^{BL}})$.

The following figure compares the original IMFs extracted on a given speech signal and the obtained IMFs Band Limited.



Stochastic Embedding EMD Models

SYSTEM MODEL 3 - CONSTRUCTION PROCEDURE



Outline

(1) Solution to PS3

PS3 aims to identify an optimal time-frequency partition to formulate a more powerful stochastic embedding which relies on SM3 and models specific frequency bandwidths. The achievement of such a partition rule Π^* is established through the following steps.

- Procedure for partition rule definition.
- Notion of optimality definition.
- Formulation of the Optimisation Problem.
- The Cross-Entropy Solution for the optimisation problem.
- The discrete case using the Multinomial distribution.

The Optimal Time-Frequency EMD Partition

1) Procedure for Partition Rule Definition:

- Consider the instantaneous frequencies denoted as $\omega_1(t), \omega_2(t), \dots, \omega_L(t)$. These continuous functions are discretely evaluated on the mesh $0 = t_0 < t_1 < \cdots < t_N = T$, the sampling points of the $\{\gamma_l(t)\}_{l=1}^{L}$.
- Consider these as a set of NL points on the time-frequency plane given as $\mathbf{p}_{l,n} = (t_n, \omega_l(t_n))$, for $l = 1, \dots, L$ and $n = 1, \dots, N$.
- Denote $\mathcal{T} = [t_0, t_N]$ the time interval and $\mathcal{I} = [\omega_0, \omega_M]$ the frequency interval, where $\omega_0 = \min_{n,l} \omega_l(t_n)$ and $\omega_M = \max_{n,l} \omega_l(t_n)$.
- Define the two-dimensional rectangle $\Pi = \mathcal{I} \times \mathcal{T}$, which total area is given as follows:

$$|\Pi| := |\mathcal{I}| \times |\mathcal{T}| = (\omega_M - \omega_0)(t_N - t_0). \tag{5.1}$$

Goal

Representing the area $|\Pi|$ via an optimal partition Π^* defined through a discretised representation over a grid of $M \times D$ smaller rectangles.

The Optimal Time-Frequency EMD Partition

Assumptions

- **Ass.1** Partition the frequency domain into M subintervals $\mathcal{I}_m := [\omega_{m-1}, \omega_m]$ s. t.

$$\mathcal{I} = \bigcup_{m=1}^{M} \mathcal{I}_m, \text{ s.t. } \bigcap_{m=1}^{M} \mathcal{I}_m = \emptyset \text{ and } |\mathcal{I}| = \sum_{m=1}^{M} |\mathcal{I}_m| \text{ for } m = 1, \ldots, M. \quad (5.2)$$

- **Ass.2** Further divide the rectangle $\mathcal{I}_m \times [t_0, t_N]$ into D smaller rectangles with equal width, by partitioning $\mathcal{T} = [t_0, t_N]$ into D intervals $\mathcal{T}_{m,d} = [s_{m,d-1}, s_{m,d}]$ s. t. $t_0 = s_0 < s_{m,d-1} < s_{m,d} \leq s_{m,D} = t_N$. This yields to

$$\mathcal{T} = \bigcup_{d=1}^{D} \mathcal{T}_{m,d}, \text{ s.t. } \bigcap_{d=1}^{D} \mathcal{T}_{m,d} = \emptyset \text{ and } |\mathcal{T}| = \sum_{d=1}^{D} |\mathcal{T}_{m,d}| \text{ for } d = 1, \ldots, D \quad (5.3)$$

- **Remark 1**: it is not necessary that $|\mathcal{T}_{m,d}| = |\mathcal{T}_{m',d}|$ for $\forall\ m \neq m'$.
- **Remark 2**: for a fixed m and $d = 1, \ldots, D$, $\Pi_{m,d}$ share the same subinterval of $\mathcal{I}$ on the frequency axis, $\mathcal{I}_m$.

The Optimal Time-Frequency EMD Partition

Partition Rule Definition

Given the above procedure, the rectangle Π can now be partitioned by defining *MD* rectangles $\Pi_{m,d} = \mathcal{I}_m \times \mathcal{T}_{m,d}$ for $m = 1, \ldots, M$ and $d = 1, \ldots, D$. The rectangles that are defined by this partition are assumed to not overlap and as a result they satisfy

$$\Pi = \bigcup_{m,d}^{D} \Pi_{m,d}, \text{ s.t. } \bigcap_{m,d} \Pi_{m,d} = \emptyset \text{ and } |\Pi| = \sum_{m,d} |\Pi_{m,d}| \qquad (5.4)$$

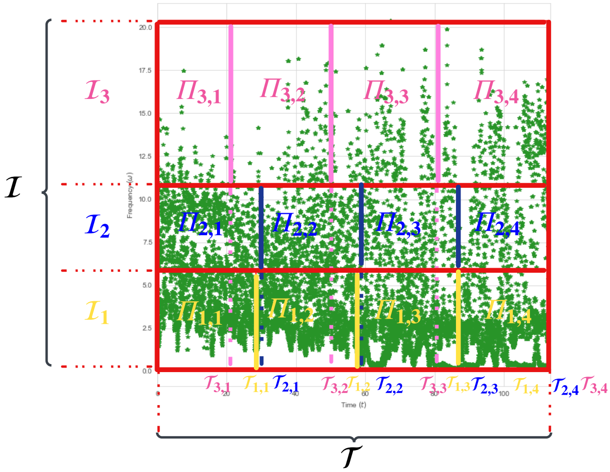
Remark 3: the rectangles $\Pi_{m,d}$ and $\Pi_{m',d}$, when $m \neq m'$, $m, m' = 1, \ldots, M$, that have the same index d do not share the same subintervals on the time axis since $\mathcal{T}_{m,d} \neq \mathcal{T}_{m',d}$.

Partition Rule Definition

$$|\Pi| := |\mathcal{I}| \times |\mathcal{T}|$$

$M = 3$

$D = 4$



The Optimal Time-Frequency EMD Partition

The next step corresponds to define the following sets of points:

$$\mathcal{P} = \bigcup \mathcal{P}_{m,d} \text{ where } \mathcal{P}_{m,d} = \left\{ \mathbf{p}_{l,n} : \mathbf{p}_{l,n} \in \Pi_{m,d} \right\} \tag{5.5}$$

The cardinality of the above set is denoted by $|\mathcal{P}_{m,d}|$.

(1) Notion of Optimal Partition Π^*

- The notion of optimality will reflect a concept of equi-energy partition in the time-frequency plane.
- The IFs samples produce a 2-d histogram. Therefore, to achieve an equi-energy partition rule, the problem becomes equivalent to solving a density optimisation problem producing an empirical ECDF as close as possible to a uniform distribution in 2d (for a given number of partitions in time and frequency).
- This can be reposed in a discrete quantised framework as a combinatorial search problem.

2) Formulation of the Optimisation Problem

To formulate an optimisation problem identifying the optimal Π^*, the parameters of the given Π are identified through the following considerations. Learning Π^* will correspond to identify the optimal set of parameters.

Parametrisation of Π

- The optimal partition Π^* is specified for the available sample set $\mathbf{p}_{l,n}$.
- Assume one partitions the given Π according to $M-1$ frequency horizontal partitions and D time vertical partitions.
- **Remark 1**: by construction, the horizontal frequency partitions are assumed constant over time. The vertical time partitions might vary within each frequency band.
- Therefore, the partition Π is parametrised according to a sequence of frequency parameters $\omega_1, \dots, \omega_{M-1}$, defining subintervals of $\mathcal{I}$, and D sequences of time parameters $s_{m,1}, \dots, s_{m,D-1}$, defining subintervals of $\mathcal{T}$ for different m. The set of parameters that are to be estimated is denoted by:

$$\boldsymbol{\psi} = \left[\omega_1, \dots, \omega_{M-1}, s_{1,1}, \dots, s_{1,D-1}, \dots, s_{m,1}, \dots, s_{m,D-1}, \dots, s_{M,1}, \dots, s_{M,D-1}\right]. \quad (5.6)$$

The Optimal Time-Frequency EMD Partition

Goal

Learn an optimal partition Π^* (hence the optimal set ψ^*) producing an empirical distribution function of the IFs in each sub-rectangle $\Pi_{m,d}$ which is as close to uniform distribution across the domain area Π as possible.

The solution this problem is provided by the Cross-Entropy Method!

The Cross-Entropy method relies on Importance Sampling and requires the definition of an **Importance distribution**.

Importance distribution Characterisation

- Consider a discrete random variable X that would be representative of the boundaries defining the optimal partition Π^*, hence the tuples (m, d). This random variable X is characterised by the Importance distribution.
- X is defined on the the indexes of the sub-rectangles $\Pi_{m,d}$ with corresponding probabilities that sum to 1.
- As a result, the set of possible values taken by X consists of DM tuples (m, d), for $m = 1, \ldots, M$ and $d = 1, \ldots, D$. Therefore, X controls assignments of a points to sub-rectangles $\Pi_{m,d}$.

The Optimal Time-Frequency EMD Partition

Target and Empirical Distributions of X

Ideally, the target distribution of X will be uniform with p.d.f $\pi(x)$

$$\text{Target} : \pi(x) = \prod_{m,d} \pi_{m,d}^{\mathbf{1}_{\{x=(m,d)\}}} \text{ for } \pi_{m,d} = \mathbb{P}\big(X = (m,d)\big) = \frac{|\Pi_{m,d}|}{|\Pi|}. \tag{5.7}$$

Therefore, it associates the probability of drawing tuple (m, d) to the proportion of the area of rectangle $\Pi_{m,d}$ to the overall area of Π. However, one has an empirical distribution for X, depending on $p_{l,n}$ and denoted as $\hat{\pi}(x)$:

$$\text{Empirical} : \hat{\pi}(x) = \prod_{m,d} \hat{\pi}_{m,d}^{\mathbf{1}_{\{x=(m,d)\}}} \text{ for } \hat{\pi}_{m,d} = \hat{\mathbb{P}}\big(X = (m,d)\big) = \frac{|\mathcal{P}_{m,d}|}{LN}, \tag{5.8}$$

Therefore, the probability of drawing tuple (m, d) reflects the proportion of the number of points $p_{n,l}$ that lay within the rectangle $\Pi_{m,d}$ to the overall sample size.

Remark that both $\pi_{m,d}$ and $\hat{\pi}_{m,d}$ satisfy

$$\sum_{m,d} \pi_{m,d} = 1 \text{ and } \sum_{m,d} \hat{\pi}_{m,d} = 1. \tag{5.9}$$

Goal

Select the support of X in such a way that the Kullback-Leibler divergence, measuring similarity between the two proposed distributions, is minimised.

Remark: The Kullback-Leibler divergence defined as

$$KL(\pi, \hat{\pi}) = \int_{x \in \mathcal{X}} \pi(x) \log \left(\frac{\pi(x)}{\hat{\pi}(x)} \right) dx. \tag{5.10}$$

Since X is a discrete random variable that ranges of values is countable and takes values in the set of tuples $\mathcal{X} = \{(m, d)\}_{m,d}$, the integration problem in 5.10 can be rewritten as a sum over the elements of the set $\mathcal{X}$, that is

$$KL(\pi, \hat{\pi}; \psi) = \log KN - \log |\Pi| + \frac{1}{|\Pi|} \sum_{m=1}^{M} \sum_{d=1}^{d} \left\{ |\Pi_{m,d}| \Big(\log |\Pi_{m,d}| - \log |\mathcal{P}_{m,d}| \Big) \right\} \tag{5.11}$$

The Optimal Time-Frequency EMD Partition

The vector of parameters belong to the multidimensional parameters space Ψ defined by the following constraints on its elements

$$\Psi = \begin{cases} \omega_1, \ldots, \omega_{M-1} \in (\omega_0, \omega_M) \text{ such that } \omega_0 < \omega_1 < \ldots < \omega_{M-1} < \omega_M, \\ s_{1,1}, \ldots, s_{1,N_1-1} \in (t_0, t_N) \text{ such that } t_0 < s_{1,1} < \ldots < s_{1,D-1} < t_N, \\ \vdots \\ s_{m,1}, \ldots, s_{m,N_m-1} \in (t_0, t_N) \text{ such that } t_0 < s_{m,1} < \ldots < s_{m,D-1} < t_N, \\ \vdots \\ s_{M,1}, \ldots, s_{M,N_M-1} \in (t_0, t_N) \text{ such that } t_0 < s_{M,1} < \ldots < s_{M,D-1} < t_N. \end{cases} \tag{5.12}$$

The Objective Function

The objective function of the constrained optimisation problem that finds optimal partitioning of Π that minimizes distance between the empirical and target distributions is specified by

$$\psi^* = \underset{\psi \in \Psi}{\text{argmin}}\, KL(\pi, \hat{\pi}; \psi) = \underset{\psi \in \Psi}{\text{argmax}} -KL(\pi, \hat{\pi}; \psi) \tag{5.13}$$

3)The Cross-Entropy Method

The objective function is optimised with respect to the vector of parameters ψ that belongs to the parameter space Ψ and is the domain of the objective function.

The goal is to identify the optimal partition Π^* treated as a parameter that have to be learnt in the optimal Importance Distribution whose set of parameters is denoted by ψ.

The optimisation problem is solved by considering the level sets of the objective function $\{\psi : KL(\psi) \geq \gamma\}$ for $\gamma \in \mathbb{R}$.

When $\gamma = \widehat{KL} = \text{argmax}_{\psi \in \Psi} KL(\psi)$, then $\{\psi : KL(\psi) \geq \gamma\} = \{\psi^\star\}$.

Next, define a family of probability measure $\{\mathbb{P}_{\varphi'} : \varphi' \in \Phi\}$ on Ψ with densities $\{f_{\varphi'} : \varphi' \in \Phi\}$ that are parametrised by $\varphi' \in \Phi$.

Let $\mathbb{E}_{\varphi'}$ denote the expectation taken with respect to $\mathbb{P}_{\varphi'}$. Fix φ' and γ and define a rare event probability problem:

$$\mathbb{P}_{\varphi'}\left[KL(\psi) \geq \gamma\right] = \mathbb{E}_{\varphi'}\left[\mathbb{I}_{\{KL(\psi) \leq \gamma\}}\right] = \int_{\Psi} \mathbb{I}_{\{KL(\psi) \leq \gamma\}} f_{\varphi'}(\psi) d\psi \qquad (5.14)$$

Instead of approximating this probability naively by sampling from $f_{\varphi'}$, the importance sampling method is used. Let $g_{\varphi''}$ denote the importance sampler, where $\varphi'' \in \Phi$. Importance sampling approximates the rare event probability by

$$\begin{aligned}\mathbb{P}_{\varphi'}\left[KL(\psi) \geq \gamma\right] &= \int_{\Psi} \mathbb{I}_{\{KL(\psi)\leq\gamma\}} f_{\varphi'}(\psi) d\psi = \int_{\Psi} \mathbb{I}_{\{KL(\psi)\leq\gamma\}} \frac{f_{\varphi'}(\psi)}{g_{\varphi''}(\psi)} g_{\varphi''}(\psi) d\psi \\ &= \mathbb{E}_{\varphi''}\left[\mathbb{I}_{\{KL(\psi)\leq\gamma\}} \frac{f_{\varphi'}(\psi)}{g_{\varphi''}(\psi)}\right] \approx \frac{1}{S}\sum_{i=1}^{S}\left\{\mathbb{I}_{\{KL(\psi^i)\leq\gamma\}} \frac{f_{\varphi'}(\psi^i)}{g_{\varphi''}(\psi^i)}\right\}\end{aligned}$$

where vectors ψ^i for $i = 1, \ldots, S$ are iid generated from $g_{\varphi''}(\psi)$. The optimal importance sampler $g_{\varphi''}$ is selected through the cross-entropy criterion:

$$\begin{aligned}\varphi^\star &= \underset{\varphi''\in\Phi}{\operatorname{argmax}} \int_{\Psi} \mathbb{I}_{\{KL(\psi)\leq\gamma\}} f_{\varphi'}(\psi) \log \frac{f_{\varphi'}(\psi)}{g_{\varphi''}(\psi)} d\psi \\ &\approx \underset{\varphi''\in\Phi}{\operatorname{argmax}} \frac{1}{S}\sum_{i=1}^{S} \mathbb{I}_{\{KL(\psi)\leq\gamma\}} \log g_{\varphi''}(\psi^i)\end{aligned} \tag{5.15}$$

where vectors ψ^i for $i = 1, \ldots, S$ are iid samples generated from $f_{\varphi'(\psi)}$. The CEM starts from an initial sampling distribution $g_{\varphi_0^\star}$ and iteratively updates $\hat{\gamma}$ and $g_{\varphi''}$.

4)Utilising Kernel Density Estimator in Kullback-Leibler Divergence of the Partitioning Problem

During the estimation process specifying the optimal partition Π, certain sub-rectangles $\Pi_{m,d}$ do not contain any of the sample points $\mathbf{p}_{l,n} = (t_n, \omega_l(t_n)) \in \Pi$ for $\Pi = \mathcal{I} \times \mathcal{T}$ for $n = 1, \dots N$ and $l = 1, \dots, L$.

Hence, the corresponding set $\mathcal{P}_{m,d}$ is empty, i.e. $\mathcal{P}_{m,d} = \emptyset$. Consequently, the probabilities $\pi^e_{m,d}(x) = \frac{|\mathcal{P}_{m,d}|}{LN}$ equal zero and their logarithms tend to infinity. To avoid these numerical difficulties $\pi^e_{m,d}(x)$ is approximated by a kernel density estimator $\hat{\pi}^e_{m,d}(x; k, h)$ parametrised by kernel $k : \Pi \times \Pi \to \mathbb{R}$ and bandwidth $h > 0$ such that

$$\hat{\pi}^e_{m,d}(x; k, h) = \int_{\Pi_{m,d}} \hat{\pi}(\mathbf{p}; k; h) d\mathbf{p} = \int_{\omega_{m-1}}^{\omega_m} \int_{s_{m,d-1}}^{s_{m,d}} \hat{\pi}(\mathbf{p}; k; h) d\mathbf{p},$$

where $\hat{\pi}(\mathbf{p}; k; h) : \Pi \to [0, 1]$ is a kernel density estimator of points $\mathbf{p} = (t, \omega(t)) \in \Pi$ specified on a sample set $\mathbf{p}_{l,n}$

$$\hat{\pi}(\mathbf{p}; k; h) = \frac{1}{Nh} \prod_{n=1}^{N} \prod_{k=1}^{K} k\left(\frac{\mathbf{p} - \mathbf{p}_{n,k}}{h}\right) \ \text{such that} \ \int_{\Pi} \hat{\pi}(\mathbf{p}; k, h) d\mathbf{p} = 1.$$

The Optimal Time-Frequency EMD Partition

The objective function of the partitioning problem is reformulated to be the Kullback-Leibler divergence between $\pi(x)$ and

$$\text{Empirical}: \hat{\pi}^e(x;k,h) = \prod_{m,d} \left(\hat{\pi}^e_{m,d}(x;k,h)\right)^{\mathbf{1}_{\{x=(m,d)\}}}, \tag{5.16}$$

that is

$$KL(\pi,\hat{\pi}^e;\psi) = -\log|\Pi| + \frac{1}{|\Pi|}\sum_{m=1}^{M}\sum_{d=1}^{d}|\Pi_{m,d}|\Big(\log|\Pi_{m,d}| - \log\hat{\pi}^e(x=(m,d);k,h)\Big) \tag{5.17}$$

and with a numerical trick, for $C > 0$ being a very small number, i.e. $C = 10^{-100}$

$$\begin{aligned} KL(\hat{\pi}^e,\pi;\psi) &= \int_{x\in\mathcal{X}} \pi(x)\log\left(\frac{\pi(x)}{\hat{\pi}^e(x;k,h)}\right)dx \\ &= -\log|\Pi| - \log C + \frac{1}{|\Pi|}\sum_{m=1}^{M}\sum_{d=1}^{d}|\Pi_{m,d}|\Big(\log|\Pi_{m,d}| - \log\frac{\hat{\pi}^e(x=(m,d);k,h)}{C}\Big) \end{aligned} \tag{5.18}$$

5)Cross-Entropy Method Selection of Importance Distribution: Discrete Case via Multinomial Distribution

The optimisation problem in 5.15 is solved through a discretisation of the intervals $\mathcal{I}$ and $\mathcal{T}$. In such a way, a CEM method with an IS distribution reflecting the distribution of discrete random variables that determine the partitioning of the rectangle Π is taken into account.

Regular Dense Grids

Consider regular dense grids of $\mathcal{I}$ and $\mathcal{T}$ constructed as follows:

1. Partition $\mathcal{I}$ into small N_ω intervals of size $\Delta_\omega = \frac{\omega_M - \omega_0}{N_\omega}$, and define $\mathcal{I}^{grid}_{n_\omega} = \omega_0 + [n_\omega - 1, n_\omega]\Delta_\omega$ for $n_\omega = 1, \ldots, N_\omega$, therefore $|\mathcal{I}^{grid}_{a}| = \Delta_\omega$;
2. Partition $\mathcal{T}$ into small N_τ intervals of size $\Delta_\tau = \frac{t_N - t_0}{N_\tau}$, and define $\mathcal{T}^{grid}_{n_\tau} = \omega_0 + [n_\tau - 1, n_\tau]\Delta_\tau$ for $n_\tau = 1, \ldots, N_\tau$, therefore, $|\mathcal{T}^{grid}_{\tau}| = \Delta_\tau$.

Probabilistic Model of $\mathcal{I}$

Define a probabilistic model to partition $\mathcal{I}$ into M subintervals, $\mathcal{I}_m$ for $m = 1, \ldots, M$. Define (M)-dimensional multinomial random vector $\mathbf{X}$ whose entries X_m on the support of $\{0, \ldots, N_\omega\}$ indicates how many subsequent grids $\mathcal{I}^{grid}_{n_\omega}$ are connected to construct partitions $\mathcal{I}_m$ and corresponding break points $\omega_{m-1}, \omega_m \in \mathcal{I}$.

The Optimal Time-Frequency EMD Partition

Therefore, the multinomial random vector **X** models the number of grid points out of N_ω that belong to each of M intervals with probabilities of being in an interval being $0 \leq p_1, \ldots, p_M \leq 1$ for $\sum_{m=1}^{M} p_m = 1$.

Distribution of X

The distribution function of **X** is formulated as

$$\pi(\mathbf{x}; \mathbf{p}) = \pi(x_1, \ldots, x_M; p_1, \ldots, p_M) = \frac{N_\omega!}{\prod_{m=1}^{M} x_m!} \prod_{m=1}^{M} p_m^{x_m}. \tag{5.19}$$

for $\mathbf{p} = [p_1, \ldots, p_M]$. Recall that $\sum_{m=1}^{M} X_m = N_\omega$ since **X** divides N_ω points into M subsets.

For instance, for realisations of $X_1,\ X_2$ such that $x_1 = 2$ and $x_2 = 5$, the partitions $\mathcal{I}_1 = [\omega_0, \omega_1]$ and $\mathcal{I}_2 = [\omega_1, \omega_2]$ are given by

$$\omega_1 = \omega_0 + \Delta_\omega x_1 \text{ and } \omega_1 = \omega_1 + \Delta_\omega x_2 = \omega_0 + \Delta_\omega(x_1 + x_2)$$

This example gives an intuition for the general rule

$$\omega_m = \omega_0 + \Delta_\omega \sum_{m'=1}^{m} x_{m'} \text{ for } m = 1, \dots M - 1.$$

and defines the approach to sample $W_1, \dots, W_{M-1}$ via change of variables such that $W_m = \omega_0 + \Delta_\omega \sum_{m'=1}^{m} X_{m'}$ for $m = 1, \dots M - 1$. The realisation of $W_1, \dots, W_{M-1}$, denoted by $\omega_1, \dots, \omega_{M-1}$, represent the break points defining partitions $\mathcal{I}_1, \dots, \mathcal{I}_M$. Also, recall that ω_0 and $W_M = \omega_M$ are fixed.

Probabilistic Model of $\mathcal{T}$

Model M independent not identical partitions of the time-domain interval $\mathcal{T}$ into D subintervals by following the same steps as before. Define M independent multinomial random variables that are D-dimensional, each, denoted by $\mathbf{X}'_m$ for $m = 1, \dots, M$, whose entries $X'_{m,d}$ on the support of $\{0, \dots, N_\tau\}$, for $d = 1, \dots, D$, specify how many subsequent grids $\mathcal{T}^{grid}_{n_\tau}$ are connected to construct partitions $\mathcal{T}_{m,d}$ of $\mathcal{T}$ and determine break points $s_{m,d-1}, s_{m,d} \in \mathcal{T}$.

Denote their distributions by $\pi(\mathbf{x}'_m; \mathbf{p}'_m)$ for $\mathbf{p}'_m = [p'_{m,1}, \ldots, p'_{m,D}]$ such that $\sum_{d=1}^{D} p'_{m,d} = 1$. For every $m = 1, \ldots, M$ this construction satisfies $\sum_{d=1}^{D} X'_{m,d} = N_\tau$ and

$$s_{m,d} = t_0 + \Delta_\tau \sum_{d'=1}^{d} x'_{m,d'} \text{ for } d = 1, \ldots, D-1, \; m = 1, \ldots M.$$

where $x'_{m,d}$ is a realisation of $X'_{m,d}$.

Therefore, the random variables $S_{m,1}, \ldots, S_{m,D-1}$ for $m = 1, \ldots, M$ are defined via change of variables such that $S_{m,d} = t_0 + \Delta_\tau \sum_{d'=1}^{d} X'_{m,d'}$ for $d = 1, \ldots D-1$ with realisations $s_{m,1}, \ldots, s_{m,D-1}$ representing the break points of the partitions $\mathcal{T}_{m,1}, \ldots, \mathcal{T}_{m,D}$. Again, recall that t_0 and $S_{m,D} = t_N$ are fixed for every $m = 1, \ldots, M$.

The Joint Distribution

Given this model, the joint distribution of $\Psi = [W_1, \ldots, W_{M-1}, S_{1,1}, \ldots, S_{M,D-1}$ can we written as

$$g(\psi; \varphi) = C\, \pi(\mathbf{x}_m; \mathbf{p}) \prod_{m=1}^{M} \pi(\mathbf{x}'_m; \mathbf{p}'_m), \tag{5.20}$$

$$\log g(\psi;\varphi) = \log C + \log(N_\omega!) + \sum_{m=1}^{M} \{\log(x_m!) + x_m \log(p_m)\} + M\log(N_\omega!) + \sum_{m=1}^{M}\sum_{d=1}^{D}\left\{\log(x'_{m,d}!) + x'_{m,d}\log(p'_{m,d})\right\}.$$

The objective function of the estimation problem with constraint imposed on $\mathbf{P} = [\mathbf{p}, \mathbf{p}'_1, \dots, \mathbf{p}'_M] \in [0,1]$ is then formulated as

$$\Lambda(\mathbf{P},\lambda) = \sum_{s=1}^{S}\left\{\mathbf{1}_{\{KL(\hat{\pi},\pi;\psi^{(s)})\leq\gamma\}}\Bigg(\log C + \log(N_\omega!) + \sum_{m=1}^{M}\left\{\log(x_m^{(s)}!) + x_m^{(s)}\log(p_m)\right\} + M\log(N_\omega!) + \sum_{m=1}^{M}\sum_{d=1}^{D}\left\{\log(x_{m,d}^{'(s)}!) + x_{m,d}^{'(s)}\log(p'_{m,d})\right\}.\Bigg)\right\} + \lambda\left(1 - \sum_{m=1}^{M} p_m\right) + \sum_{m=1}^{M}\lambda_m\left(1 - \sum_{d=1}^{D} p'_{m,d}\right). \quad (5.21)$$

The Optimal Time-Frequency EMD Partition

Consequently

$$\begin{cases} \frac{\partial \Lambda(\mathbf{P},\lambda)}{\partial p_1} = \sum_{s=1}^{S} \left\{ \mathbf{1}_{\left\{KL(\hat{\pi},\pi;\psi^{(s)})\leq\gamma\right\}} \frac{x_1^{(s)}}{p_1} \right\} - \lambda = 0 \\ \vdots \\ \frac{\partial \Lambda(\mathbf{P},\lambda)}{\partial p_M} = \sum_{s=1}^{S} \left\{ \mathbf{1}_{\left\{KL(\hat{\pi},\pi;\psi^{(s)})\leq\gamma\right\}} \frac{x_M^{(s)}}{p_M} \right\} - \lambda = 0 \\ 1 - \sum_{m=1}^{M} p_m = 0 \end{cases}$$

$$\Rightarrow \begin{cases} p_1^* = \frac{1}{\lambda} \sum_{s=1}^{S} \left\{ \mathbf{1}_{\left\{KL(\hat{\pi},\pi;\psi^{(s)})\leq\gamma\right\}} x_1^{(s)} \right\} \\ \vdots \\ p_M^* = \frac{1}{\lambda} \sum_{s=1}^{S} \left\{ \mathbf{1}_{\left\{KL(\hat{\pi},\pi;\psi^{(s)})\leq\gamma\right\}} x_M^{(s)} \right\} \\ \sum_{m=1}^{M} p_m = 1. \end{cases}$$

Since $\sum_{m=1}^{M} p_m = 1$ and $\sum_{m=1}^{M} x_m^{(s)} = N_\omega$

$$\frac{1}{\lambda} \sum_{s=1}^{S} \left\{ \mathbf{1}_{\left\{KL(\hat{\pi},\pi;\psi^{(s)})\leq\gamma\right\}} \sum_{m=1}^{M} x_m^{(s)} \right\} = 1 \Rightarrow \lambda = N_\omega \sum_{s=1}^{S} \mathbf{1}_{\left\{KL(\hat{\pi},\pi;\psi^{(s)})\leq\gamma\right\}}$$

The Optimal Time-Frequency EMD Partition

and finally

$$\hat{p}_m = \frac{\sum_{s=1}^{S} \left\{ \mathbf{1}_{\left\{KL(\hat{\pi},\pi;\psi^{(s)})\leq\gamma\right\}} \frac{x_m^{(s)}}{N_\omega} \right\}}{\sum_{s=1}^{S} \mathbf{1}_{\left\{KL(\hat{\pi},\pi;\psi^{(s)})\leq\gamma\right\}}} \tag{5.22}$$

Following the same steps, then

$$\hat{p}'_{m,d} = \frac{\sum_{s=1}^{S} \left\{ \mathbf{1}_{\left\{KL(\hat{\pi},\pi;\psi^{(s)})\leq\gamma\right\}} \frac{x_{m,d}^{'(s)}}{N_\tau} \right\}}{\sum_{s=1}^{S} \mathbf{1}_{\left\{KL(\hat{\pi},\pi;\psi^{(s)})\leq\gamma\right\}}} \tag{5.23}$$

References

- Mina B. Abd el Malek and Samer S. Hanna, *The hilbert transform of cubic splines*, Communications in Nonlinear Science and Numerical Simulation **80** (2020), 104983.
- Norden E Huang, Zheng Shen, Steven R Long, Manli C Wu, Hsing H Shih, Quanan Zheng, Nai-Chyuan Yen, Chi Chao Tung, and Henry H Liu, *The empirical mode decomposition and the hilbert spectrum for nonlinear and non-stationary time series analysis*, Proceedings of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences **454** (1998), no. 1971, 903–995.