A Multicomponent proximal algorithm for Empirical Mode Decomposition

N. Pustelnik[1], P. Borgnat[1], and P. Flandrin [1]

May 28, 2013

Abstract

The Empirical Mode Decomposition (EMD) is known to be a powerful tool adapted to the decomposition of a signal into a collection of intrinsic mode functions (IMF). A key procedure in the extraction of the IMFs is the sifting process whose main drawback is to depend on the choice of an interpolation method and to have no clear convergence guarantees. We propose a convex optimization procedure in order to replace the sifting process in the EMD. The considered method is based on proximal tools, which allow us to deal with a large class of constraints such as quasi-orthogonality or extrema-based constraints.

[1]Nelly Pustelnik, Pierre Borgnat, and Patrick Flandrin are with the Laboratoire de Physique de l'ENS Lyon, CNRS UMR 5672, F69007 Lyon, France. E-mail: nelly.pustelnik@ens-lyon.fr,pierre.borgnat@ens-lyon.fr,flandrin@ens-lyon.fr.

1 Introduction

The concept of EMD was introduced by Huang et al. [1] in order to propose an adaptive data analysis method to study trend and instantaneous frequency of non-linear and non-stationary data.

The principle of the EMD is to adaptively decompose a signal into a collection of intrinsic mode functions (IMF), which are basically a set of functions oscillating around zero but non-necessarily constant in frequency and amplitude. The IMF characterization imposes the average of the envelope defined by the local maxima and the envelope defined by the local minima to be zero. It results that a signal $x \in \mathbb{R}^N$ can be written as

$$x = \sum_{k=1}^{K} d_k + a_K, \tag{1}$$

where, for every $k \in \{1, \dots, K\}$, $d_k \in \mathbb{R}^N$ is the intrinsic mode of order k, and $a_K \in \mathbb{R}^N$ denotes the trend of order K.

At each step of the EMD, the trend and the IMF of order $k \geq 1$, respectively denoted by a_k and d_k, are extracted from the trend of order $k-1$ (note that $a_0 = x$). In other words, it is based on a particular case of trend-fluctuation decomposition [2]. In the EMD, this decomposition stage is known as the *sifting process* and it consists in:

1- initialize a temporary variable $s = a_{k-1}$,

2- identify all extrema of s,

3- interpolate between minima (resp. maxima) ending up with some envelope $e_{\min}$ (resp. $e_{\max}$),

4- compute the mean envelope $m = \frac{e_{\min}+e_{\max}}{2}$,

5- extract the residual $s = s - m$,

6- iterate Steps 2-5 until the residual s achieves a zero mean envelope,

7- let the IMF of order k be $d_k = s$ and the trend of order k be $a_k = a_{k-1} - s$.

Although this technique proved its efficiency through numerous applications (see [3] and references therein), the result of this method is highly dependent on the interpolation process in Step 3 and it has also been pointed out to be sensitive to sampling effects [4]. Most of all, this technique faces the difficulty of having no mathematical definition besides its algorithm and thus no convergence properties.

For the last ten years, many references have been focused on finding a rigorous mathematical formalism for the EMD. On one hand, Daubechies et al. [5] combined synchrosqueezing (which is a special case of reassignment methods [6]) with wavelet transform in order to model EMD. On the other hand, convex optimization tools has been explored. In [7, 8], the sifting process is

replaced by a constrained optimization procedure which looks for a trend of order k belonging to the class of spline functions. In [9], the authors decompose the signal into its local median and an IMF by solving a nonlinear optimization problem which involves a dictionary learning step. This method has been generalized in [10] in order to simultaneously extract the K modes, but it still requires a learning stage. In the image processing field, the matching of trend-fluctuation decomposition is known as geometry-texture decomposition [11]. This field of interest deals with multicomponent convex optimization techniques in order to simultaneously constrain the textural and the geometric components [12].

The proposed approach follows the idea of texture-geometry decomposition with further specific EMD features such as quasi-orthogonality and extrema-based constraints. It results that a multicomponent primal-dual algorithm will be proposed and the associated convergence property will be presented.

In Section 2, we review the recent advances in non-smooth convex optimization and we detail the most recent convex criteria designed for trend-fluctuation decomposition. In Section 3, we discuss the choice of the proposed criterion and we present a multicomponent version of the primal-dual algorithm named M+LFBF [13] as well as the associated convergence property. Experimental results and comparisons with existing methods are given in Section 4. Conclusions will be drawn in Section 5, outlying possibilities of future works.

Notations Throughout this paper, we denote by $\mathbb{R}^X$ the usual X-dimensional Euclidean space and by $\Gamma_0(\mathbb{R}^X)$ the class of lower semicontinuous convex functions from $\mathbb{R}^X$ to $]-\infty,+\infty]$ which are proper in the sense that $\operatorname{dom}\varphi = \left\{y \in \mathbb{R}^X \;\middle|\; \varphi(y) < +\infty\right\} \neq \varnothing$. Argmin refers to a set of minimizers while $\arg\min$ denotes a unique minimizer.

2 Signal decomposition and convex optimization

2.1 Proximal methods

During the last decades, convex optimization methods have been shown to be very effective for solving inverse problems (for instance, algorithms such as Projection Onto Convex Sets (POCS) [14] or parallel approaches such as block-iterative surrogate constraint splitting [15]). However these methods are not applicable to non-differentiable objective criteria, which become of great interest with the compressed sensing development involving ℓ_1-minimization. Consequently, since 2004, there has been a large interest for the proximal methods which are designed to deal with convex but non-necessarily differentiable functions [16]. As indicated by the name of these methods, they are built on the Moreau proximity operator.

Definition 2.1 [17] Let $\mathcal{H}$ be an Hilbert space. Let $\varphi \in \Gamma_0(\mathcal{H})$. For every $u \in \mathcal{H}$, the proximity operator of φ is

$$\operatorname{prox}_{\varphi}\colon u \mapsto \underset{v\in\mathcal{H}}{\arg\min}\ \varphi(v) + \frac{1}{2}\|u-v\|^2. \tag{2}$$

A particular example of proximity operator is the projection onto a convex set $C \subset \mathcal{H}$. Indeed, if ι_C denotes the indicator function of C (it takes on the value 0 in C and $+\infty$ in $\mathcal{H} \setminus C$) then the projection operator P_C is prox_{ι_C}. For a detailed account of the theory of proximity operators, see [16] and references therein.

The large interest for proximal tools have enabled to develop a large panel of algorithms to efficiently solve Problem 2.2.

Problem 2.2 Let $\mathcal{H}$ be an Hilbert space. For every $i \in \{1, \ldots, I\}$, let $\mathcal{G}_i$ be an Hilbert space, let $\varphi_i \in \Gamma_0(\mathcal{G}_i)$, and let $L_i : \mathcal{H} \to \mathcal{G}_i$ be a bounded linear operator. The problem consists in finding

$$\widehat{u} \in \underset{u \in \mathcal{H}}{\text{Argmin}} \sum_{i=1}^{I} \varphi_i(L_i u). \tag{3}$$

The class of proximal algorithms can be split in two groups: the primal algorithms [16] and the primal-dual algorithms [13]. To summarize, the primal algorithms generally require to compute the inverse of $\sum_i L_i^* L_i$ whereas the primal-dual iterations only involve the computation of L_i and L_i^*. However, we have to note that even if the primal-dual methods are often easier to implement than the primal methods, they generally converge slower.

A multicomponent version of Problem 2.2 can be achieved by considering the product space $\mathcal{H} = \mathcal{H}_1 \times \ldots \times \mathcal{H}_m$ equipped with the usual vector space structure and the scalar product

$$(\forall (\mathrm{u}, \mathrm{v}) \in \mathcal{H} \times \mathcal{H}), \quad (\mathrm{u}, \mathrm{v}) \mapsto \sum_{i=1}^{m} \langle u_i | v_i \rangle, \tag{4}$$

where $\mathrm{u} = (u_i)_{1 \leq i \leq m}$ denotes a generic element in $\mathcal{H}$. The readers could refer to [12] for details on the multicomponent primal proximal algorithms. This multicomponent formalism will be used in Section 3 in order to deal with the primal-dual proximal algorithm M+LFBF [13].

2.2 Signal decomposition and variational approach

The first interpretation of the EMD in term of convex optimization has been proposed by Meignen and Perrier [7]. In this former work, the authors have replaced the sifting process by convex optimization techniques. The extraction of the trend is achieved by considering, for every $k \in \{1, \ldots, K\}$,

$$a_k \in \underset{a \in \mathbb{R}^N}{\text{Argmin}} \|a\|_2 \quad \text{s.t.} \quad a \in \Pi \cap C_{a_{k-1}} \tag{5}$$

where a_k denoted the trend of order k while $d_k = a_{k-1} - a_k$ denotes the IMF of order k, Π denotes the space of spline functions, and $C_{a_{k-1}}$ denotes a constraint onto the dynamic range of the mean envelope at the location of the extrema of a_{k-1}. Note that the initialization is reduced to $a_0 = x$. To be entirely accurate, we have to specify that the variable to be optimized in (5) is not the trend but the coefficients associated to the Hermite interpolant of a. In [8], the dynamic range constraint is replaced by a constraint which imposes the symmetry of the upper and lower envelopes of $a_{k-1} - a$. The limitation stays that this approach requires a first approximation $\overline{a}$

of a_k in order to deal with a convex constraint. Note that this approach still looks for a mean envelope in the space of spline functions.

It is then interesting to remark that a similar problem has been looked at in image processing. This problem is known as image decomposition into texture and geometry components and, more generally, the idea is to decompose an image in elementary structures. In the context of denoising, this decomposition has been achieved in [18] with a total variation potential in order to extract a piecewise smooth component. In [11, 19], a criterion combining total variation and G-norm has been considered in order to perform geometry-texture decomposition. Note that the G-norm had been theoretically introduced few years before in [20] in order to model strong oscillations. In most recent works, the oscillating patterns have been extracted in considering ℓ_1-norm applied on frame coefficients [12, 21, 22]. The use of composite criteria have also been proposed in [23]. The general variational formulation associated to the geometry-texture decomposition is presented in Problem 2.3.

Problem 2.3

$$\text{Find} \quad (\widehat{a},\widehat{d}) \in \underset{a\in\mathbb{R}^N,\, d\in\mathbb{R}^N}{\text{Argmin}}\; h(a,d)+g(a)+f(d),$$

where $g \in \Gamma_0(\mathbb{R}^N)$ and $f \in \Gamma_0(\mathbb{R}^N)$ are potentials promoting the properties of the geometry and texture components separately, while $h \in \Gamma_0(\mathbb{R}^N \times \mathbb{R}^N)$ is a coupling term modeling their interaction.

The case of three components jointly estimated has been considered in [11].

3 Proposed approach

Based on the state-of-the-art of EMD and image decomposition, we propose to replace the sifting process by a trend-fluctuation decomposition method based on a multicomponent variational analysis. First, we have to specify the properties that we want to impose onto each component. On one hand, the IMF of order k is expected to

- have a zero-mean envelope,
- be quasi-orthogonal to the IMF of order $j < k$.

The first condition is the most difficult to impose. We propose a method derived from [8] in order to deal with it. We denote by $(t_k[\ell])_{1\leq\ell\leq L}$ the location of the local extrema of a_{k-1}, and these extrema are alternatively minima and maxima. We can approximate the first condition by considering, for every $\ell \in \{1,\ldots,L\}$,

$$\left| d\big[t_k[\ell]\big] + \frac{\alpha_\ell d\big[t_k[\ell-1]\big] + \beta_\ell d\big[t_k[\ell+1]\big]}{\alpha_\ell+\beta_\ell} \right| < \varepsilon_\ell \tag{6}$$

where $\alpha_\ell = t_k[\ell+1]-t_k[\ell]$, $\beta_\ell = t_k[\ell]-t_k[\ell-1]$, and $\varepsilon_\ell > 0$. The coefficients α_ℓ and β_ℓ are chosen so that, in Eq. (6), the extremum $d\big[t_k[\ell]\big]$ (e.g., a maximum) is approximately compared to its

mirror-point on the would-be other envelope (e.g., the minimum envelope) which would be locally defined thanks to $d\big[t_k[\ell-1]\big]$ and $d\big[t_k[\ell+1]\big]$. Note that no envelope is explicitly computed. This condition can be globally rewritten as

$$\|D_k d\|_1 \leq \varepsilon_k, \tag{7}$$

where $\varepsilon_k > 0$ and $D_k \in \mathbb{R}^{N\times N}$ denotes a linear operator which models the penalization imposed on d at each location $t_k[\ell]$. The second condition requires to impose a constraint taking the form

$$|\langle d, d_j\rangle| \leq \zeta_{k,j}, \tag{8}$$

where $\zeta_{k,j} > 0$.

On the other hand, the trend of order k has to be a smooth signal. This condition can be achieved by imposing

$$\|Aa\|_p^p \leq \eta_k, \tag{9}$$

where $\eta_k > 0$, $p \geq 1$ (typically $p = 1$ or $p = 2$), and A denotes a derivative operator (1st or 2nd order derivative). If $p = 1$, it corresponds to a total variation constraint. At last, we want to impose that the sum of the extracted trend and the IMF is close to a_{k-1}. However, we avoid to use a strict equality in order to limitate the sampling effects.

The criterion we propose to consider is summarized in Problem 3.1.

Problem 3.1 Let $d_0 = 0$ and let $a_0 = x$. For every $k \in \{1, \ldots, K\}$,

$$(a_k, d_k) \in \underset{a\in\mathbb{R}^N,\, d\in\mathbb{R}^N}{\operatorname{Argmin}} \|a_{k-1} - a - d\|_2^2$$

$$\text{subject to} \quad \begin{cases} \|Aa\|_p^p \leq \eta_k, \\ \|D_k d\|_1 \leq \varepsilon_k, \\ (\forall j \in \{0, \ldots, k-1\}),\ |\langle d, d_j\rangle| \leq \zeta_{k,j}, \end{cases}$$

where $\eta_k > 0$, $\epsilon_k > 0$, $\zeta_{k,j} \geq 0$, $p \geq 1$, $A \in \mathbb{R}^{N\times N}$, and $D_k \in \mathbb{R}^{N\times N}$.

The criterion involved in Problem 3.1 is a particular case of Problem 2.2 with $I = k + 2$ and $u = (a, d)$. Consequently, Problem 3.1 can be solved with a multicomponent version of M+LFBF [13] whose iterations and convergence result are detailed in Algorithm 1 and Proposition 3.2. The choice of this specific proximal algorithm among the other ones is motivated by the presence of the linear operator D and the presence of the quadratic term for the objective function (i.e. a Lipschitz differentiable function).

Proposition 3.2 *The sequence $(a^{(n)}, d^{(n)})_{n\in\mathbb{N}}$ generated by Algorithm 1 converges to a solution (a_k, d_k) of the minimization problem involved in Problem 3.1.*

The convergence result in Proposition 3.2 is obtained by applying [13, Theorem 4.2] in $\mathcal{H} = \mathbb{R}^N \times \mathbb{R}^N$ and using Eq. (4).

In Algorithm 1, the constraints sets $(C_i^{\zeta_{k,i-2}})_{2\leq i\leq k+1}$, $C_1^{\eta_k}$, and $C_1^{\varepsilon_k}$ are involved. The definition of this convex sets subject to Problem 3.1 are specified below as well as the associated projection operators.

The projection onto the convex set $C_1^{\varepsilon_k} = \{u \in \mathbb{R}^N \mid \|u\|_1 \leq \varepsilon_k\}$ can be computed iteratively by considering [24] or by using epigraphical projection techniques as detailed in [25]. The projection onto the convex set $C_1^{\eta_k} = \{u \in \mathbb{R}^N \mid \|u\|_p^p \leq \eta_k\}$ has been previously detailed for the case where $p = 1$ and it has an explicit form when $p = 2$:

$$(\forall u \in \mathbb{R}^N), \quad P_{C_1^{\eta_k}} u = \begin{cases} u, & \text{if } \|u\|_2 \leq \eta_k \\ \frac{\eta_k u}{\|u\|_2}, & \text{otherwise.} \end{cases} \tag{10}$$

For every $i \in \{2, \ldots, k+1\}$, the projection onto the constraint set $C_i\zeta_{k,i-2} = \{u \in \mathbb{R}^N \mid |\langle u, d_{i-2}\rangle| \leq \zeta_{k,i-2}\}$ is, for every $u \in \mathbb{R}^N$,

$$P_{C_i^{\zeta_{k,i-2}}} u = \begin{cases} u, & \text{if } |\langle u, d_{i-2}\rangle| \leq \zeta_{k,i-2} \\ u + \frac{\zeta_{k,i-2} - \langle u, d_{i-2}\rangle}{\|d_{i-2}\|_2^2} d_{i-2}, & \text{otherwise.} \end{cases} \tag{11}$$

4 Experimental results

We propose to compare the proposed method (P-EMD) with the classical EMD (C-EMD) [26] and the optimization-based EMD (O-EMD) proposed in [8]. We consider two experiments for which the original signals (IMFs and trend) are represented in black, the results obtained with traditional EMD [26] are plotted in blue, the results extracted with the optimization procedure of Oberlin *et al.* [8] in green, and those by the proposed approach in red. For each figure, the signal x to be decomposed is plotted on the left column, the resuls are presented on the middle column, and a zooming in of the results is plotted on the right column.

The first experiment is presented in Figure 1. The signal to be decomposed consists in a sum of a triangular signal and an AM signal ($K = 2$) with $N = 1000$. The oscillating part is plotted on the top row while the trend is presented on the bottom row. The experimental parameters are $\eta_1 = 0.02$, $\varepsilon_1 = 0.58$, $\zeta_{1,0} = 0$, $p = 1$, and A denotes the matrix associated to the filter $(\frac{1}{4}, -\frac{1}{2}, \frac{1}{4})$. Note that the choice of $p = 1$ allows us to model the non-smooth behavior of a_1. When $p = 2$ the proposed method achieves a mean square error of 14.1×10^{-5}. For this example, it appears that all the methods give good results, however we can observe that the proposed approach allows us to avoid the spline behaviour which is not desired in the trend component.

The second experiment is presented in Figure 2. The signal consists in a sum of a three components ($K = 3$) of size $N = 1000$: the first mode and the trend have a constant amplitude and frequency while the second mode denotes a AM signal. The experimental parameters are $\eta_1 = 37.2$, $\eta_2 = 0.05$, $\varepsilon_1 = 13.6$, $\varepsilon_2 = 69.2$, $\zeta_{1,0} = 0$, $\zeta_{2,0} = 0$, $\zeta_{2,1} = 0.1$, $p = 1$, and A denotes the matrix associated to the filter $(\frac{1}{4}, -\frac{1}{2}, \frac{1}{4})$. The results obtained with the proposed approach are very close to the classical EMD results for a_2 and d_1 and are better than the convex optimization procedure developed in [8] for a_2. However, we have to remark that in the low amplitude part of d_2 our approach introduces some undesired residual oscillations.

The computational time is about 20 secondes for the first experiment and 100 secondes for the second one.

5 Conclusions and perspectives

We propose an efficient method in order to replace the sifting process in EMD. This work follows the one in [8] but it requires neither the use of splines nor a first estimate of the trend of order k. The proposed approach is based on proximal tools and its good behaviour is evaluated and compared to the state-of-the-art methods.

The proposed approach allows us to be robust to sampling effects and to avoid a spline interpolation procedure which is known to lead to some artefacts. Moreover, the proposed approach has convergence guarantees that classical EMD does not have. The first results are encouraging and they pave the way to many questions such as: how could we easily select the parameters ε and η which have a key role in the efficiency of the proposed approach or do we need to add other constraints? How is the sensitivity of these parameters regarding the decomposition results? How can we insure the robustness to noise?

Algorithm 1 – Multicomponent M+LFBF.

Initialization

Let $\tau = 4 + \sqrt{\max(\|A\|^2, \|D_k\|^2) + (k-1)}$

Let $\sigma \in]0, 1/(\tau+1)[, \gamma \in [\sigma, (1-\sigma)/\tau)], a^{(0)} \in \mathbb{R}^N, d^{(0)} \in \mathbb{R}^N$

For every $i \in \{1, \ldots, k+1\}$, $v_i^{(0)} \in \mathbb{R}^N$ and $\widetilde{v}_i^{(0)} \in \mathbb{R}^N$

For $n = 0, 1, \ldots$

– Steps involving gradient and adjoint operators–

$y^{(n)} = a^{(n)} - \gamma\Big(2(a^{(n)} + d^{(n)} - a_{k-1}) + A^\top v_1^{(n)} + \sum_{i=2}^{k+1} v_i^{(n)}\Big)$

$\widetilde{y}^{(n)} = d^{(n)} - \gamma\Big(2(a^{(n)} + d^{(n)} - a_{k-1}) + D_k^\top \widetilde{v}_1^{(n)} + \sum_{i=2}^{k+1} \widetilde{v}_i^{(n)}\Big)$

– Steps involving linear operators–

$z_1^{(n)} = v_1^{(n)} + \gamma A a^{(n)}$

$\widetilde{z}_1^{(n)} = \widetilde{v}_1^{(n)} + \gamma D_k d^{(n)}$

– Steps involving proximity operator computation–

$p_1^{(n)} = z_1^{(n)} - \gamma P_{C_1^{\eta_k}}(z_1^{(n)}/\gamma)$

$\widetilde{p}_1^{(n)} = \widetilde{z}_1^{(n)} - \gamma P_{C_1^{\varepsilon_k}}(\widetilde{z}_1^{(n)}/\gamma)$

– Steps involving linear operators–

$q_1^{(n)} = p_1^{(n)} + \gamma A y^{(n)}$

$\widetilde{q}_1^{(n)} = \widetilde{p}_1^{(n)} + \gamma D_k \widetilde{y}^{(n)}$

– Updating steps–

$v_1^{(n+1)} = v_1^{(n)} - z_1^{(n)} + q_1^{(n)}$

$\widetilde{v}_1^{(n+1)} = \widetilde{v}_1^{(n)} - \widetilde{z}_1^{(n)} + \widetilde{q}_1^{(n)}$

For $i = 2, \ldots, k+1$

– Updating steps–

$z_i^{(n)} = v_i^{(n)} + \gamma a^{(n)}$

$\widetilde{z}_i^{(n)} = \widetilde{v}_i^{(n)} + \gamma d^{(n)}$

– Step involving proximity operator computation–

$\widetilde{p}_i^{(n)} = \widetilde{z}_i^{(n)} - \gamma P_{C_i^{\zeta_{k,i-2}}}(\widetilde{z}_i^{(n)}/\gamma)$

– Updating steps–

$q_i^{(n)} = \gamma y^{(n)}$

$\widetilde{q}_i^{(n)} = \widetilde{p}_i^{(n)} + \gamma \widetilde{y}^{(n)}$

$v_i^{(n+1)} = v_i^{(n)} - z_i^{(n)} + q_i^{(n)}$

$\widetilde{v}_i^{(n+1)} = \widetilde{v}_i^{(n)} - \widetilde{z}_i^{(n)} + \widetilde{q}_i^{(n)}$

– Steps involving gradient and adjoint operators–

$u^{(n)} = y^{(n)} - \gamma\Big(2(y^{(n)} + \widetilde{y}^{(n)} - a_{k-1}) + A^\top p_1^{(n)}\Big)$

$\widetilde{u}^{(n)} = \widetilde{y}^{(n)} - \gamma\Big(2(y^{(n)} + \widetilde{y}^{(n)} - a_{k-1}) + D_k^\top \widetilde{p}_1^{(n)} + \sum_{i=2}^{k+1} \widetilde{p}_i^{(n)}\Big)$

– Updating steps–

$a^{(n+1)} = a^{(n)} - y^{(n)} + u^{(n)}$

$d^{(n+1)} = d^{(n)} - \widetilde{y}^{(n)} + \widetilde{u}^{(n)}$

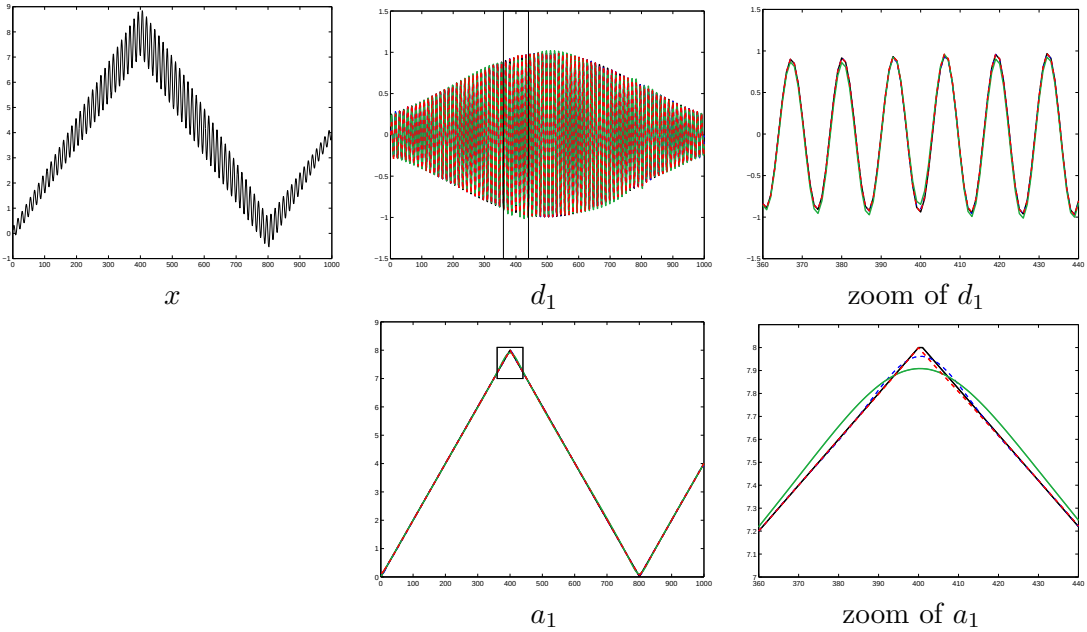

Figure 1: Sum of two signals. The signal x is on the left, the results are on the middle, and a zoom of the results is on the right. Signal is in black, traditional EMD is in blue, the optimization procedure of [8] is in magenta, and the proposed approach is in red. The mean square errors for the different methods are similar for a_1 and d_1: 8.3×10^{-5} (C-EMD), 49×10^{-5} (O-EMD), 7.5×10^{-5} (P-EMD).

References

[1] N. E. Huang, Z. Shen, S. R. Long, M. C. Wu, H. Shih, Q. Zheng, N.-C. Yen, C. C. Tung, and H. H. Liu, "The empirical mode decomposition and the Hilbert spectrum for nonlinear and nonstationary time series analysis," *Proc. R. Soc. Lond. A*, vol. 454, pp. 903–995, 1998.

[2] T. Alexandrov, S. Bianconcini, E. Bee Dagum, P. Maass, and T. McElroy, "A review of some modern approaches to the problem of trend extraction," *US Census Bureau TechReport RRS2008/03*, 2008.

[3] N. E. Huang and S. S. Shen, *Hilbert-Huang transform and its applications*, vol. 5, World Scientific, Singapore, 2005.

[4] G. Rilling and P. Flandrin, "Sampling effects on the empirical mode decomposition," *Adv. Adapt. Data Anal.*, vol. 1, no. 1, 2009.

[5] I. Daubechies, J. Lu, and H.-T. Wu, "Synchrosqueezed wavelet transforms: An empirical mode decomposition-like tool," *Appl. Comp. Harm. Analysis*, vol. 30, no. 2, pp. 243–261, Mar. 2011.

[6] P. Flandrin, F. Auger, and E. Chassande-Mottin, "Time-frequency re-assignment — From principles to algorithms," in *Applications in Time- Frequency Signal Processing*, A. Papandreou-Suppappola, Ed., chapter 5, pp. 179–203. CRC Press, Boca Raton, FL, 2003.

[7] S. Meignen and V. Perrier, "A new formulation for Empirical Mode Decomposition based on constrained optimization," *IEEE Signal Process. Lett.*, vol. 14, no. 12, pp. 932–935, Dec. 2007.

[8] T. Oberlin, S. Meignen, and V. Perrier, "An alternative formulation for the Empirical Mode Decomposition," 2011, preprint, http://hal.archives-ouvertes.fr/hal-00553107.

[9] T. Y. Hou and Z. Shi, "Adaptive data analysis via sparse time-frequency representation," *Adv. Adapt. Data Anal.*, vol. 3, no. 1-2, pp. 1–28, Mar. 2011.

[10] T. Y. Hou and Z. Shi, "Data-driven time-frequency analysis," 2012, preprint, http://arxiv.org/abs/1202.5621.

[11] J.-F. Aujol, G. Gilboa, T. Chan, and S. Osher, "Structure-texture image decomposition - modeling, algorithms, and parameter selection," *Int. J. Comp. Vis.*, vol. 67, no. 1, pp. 111–136, Apr. 2006.

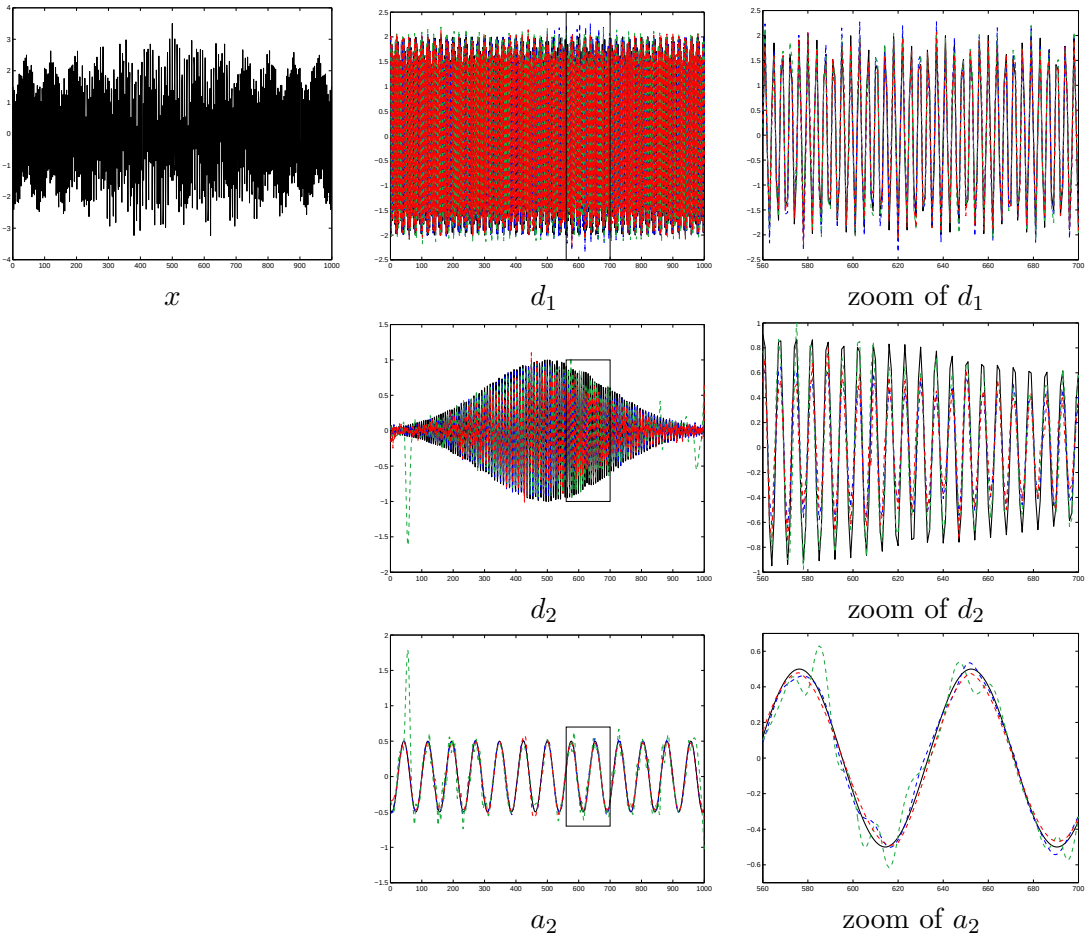

Figure 2: Sum of three signals. The signal x is on the left, the results are on the middle, and a zoom of the results is on the right. Signal is in black, traditional EMD is in blue, the optimization procedure of [8] is in magenta, and the proposed approach is in red. The mean square errors are for d_1: 0.8×10^{-2} (C-EMD), 1.5×10^{-2} (O-EMD), 1.6×10^{-2} (P-EMD), for d_2: 0.8×10^{-2} (C-EMD), 4.1×10^{-2} (O-EMD), 3.9×10^{-2} (P-EMD), and for a_2: 0.4×10^{-3} (C-EMD), 37×10^{-3} (O-EMD), 1.5×10^{-3} (P-EMD).

[12] L. M. Bricenõ-Arias, P. L. Combettes, J.-C. Pesquet, and N. Pustelnik, "Proximal algorithms for multicomponent image processing," *J. Math. Imag. Vis.*, vol. 41, no. 1, pp. 3–22, Sep. 2011.

[13] P. L. Combettes and J.-C. Pesquet, "Primal-dual splitting algorithm for solving inclusions with mixtures of composite, Lipschitzian, and parallel-sum type monotone operators," *Set-Valued Var. Anal.*, 2011.

[14] H. J. Trussell and M. R. Civanlar, "The feasible solution in signal restoration," *IEEE Trans. Acous., Speech Signal Process.*, vol. 32, no. 2, pp. 201–212, Apr. 1984.

[15] P. L. Combettes, "A block-iterative surrogate constraint splitting method for quadratic signal recovery," *IEEE Trans. Signal Process.*, vol. 51, no. 7, pp. 1771–1782, July 2003.

[16] P. L. Combettes and J.-C. Pesquet, "Proximal splitting methods in signal processing," in *Fixed-Point Algorithms for Inverse Problems in Science and Engineering*, H. H. Bauschke, R. Burachik, P. L. Combettes, V. Elser, D. R. Luke, and H. Wolkowicz, Eds., pp. 185–212. Springer-Verlag, New York, 2010.

[17] J. J. Moreau, "Proximité et dualité dans un espace hilbertien," *Bull. Soc. Math. France*, vol. 93, pp. 273–299, 1965.

[18] L. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Physica D*, vol. 60, no. 1-4, pp. 259–268, Nov. 1992.

[19] J.-F. Aujol, G. Aubert, L. Blanc-Féraud, and A. Chambolle, "Image decomposition into a bounded variation component and an oscillating component," *Int. J. Comp. Vis.*, vol. 22, pp. 71–88, 2005.

[20] Y. Meyer, *Oscillating patterns in image processing and in some nonlinear evolution equations*, AMS, Providence, RI, 2001.

[21] J.-L. Starck, M. Elad, and D. Donoho, "Image decomposition via the combination of sparse representations and a variational approach," *IEEE Trans. Image Process.*, vol. 14, pp. 1570–1582, 2005.

[22] S. Anthoine, E. Pierpaoli, and I. Daubechies, "Deux méthodes de déconvolution et séparation simultanées; application à la reconstruction des amas de galaxies," *Trait. Signal*, vol. 23, no. 5-6, pp. 439–447, 2006.

[23] Ph Ciuciu, J. Idier, and J.-F. Giovannelli, "Regularized estimation of mixed spectra using a circular Gibbs-Markov model," *IEEE Trans. Signal Process.*, vol. 49, no. 10, pp. 2202–2213, Oct. 2001.

[24] E Van Den Berg and M. P. Friedlander, "Probing the Pareto frontier for basis pursuit solutions," *SIAM J. Sci. Comput.*, vol. 31, no. 2, pp. 890–912, Nov. 2008.

[25] G. Cherchia, N. Pustelnik, J.-C. Pesquet, and B. Pesquet-Popescu, "A proximal approach approach for constrained cosparse modelling," in *Proc. Int. Conf. Acoust., Speech Signal Process.*, Kyoto, Japan, March 25-30 2012, 4p.

[26] G. Rilling, P. Flandrin, and P. Gonçalvès, "On empirical mode decomposition and its algorithms," in *IEEE-EURASIP Workshop on Nonlinear Signal and Image Processing (NSIP-03)*, 2003.