Learnable Empirical Mode Decomposition based on Mathematical Morphology

Santiago Velasco-Forero, Romain Pagès, Jesus Angulo

Learnable Empirical Mode Decomposition based on Mathematical Morphology *

Santiago Velasco-Forero[†], R. Pagès[‡], and Jesus Angulo [§]

Abstract. Empirical mode decomposition (EMD) is a fully data driven method for multiscale decomposing signals into a set of components known as intrinsic mode functions. EMD is based on lower and upper envelopes of the signal in an iterated decomposition scheme. In this paper, we put forward a simple yet effective method to learn EMD from data by means of morphological operators. We propose an end-to-end framework by incorporating morphological EMD operators into deeply learned representations, trained using standard backpropagation principle and gradient descent-based optimization algorithms. Three generalizations of morphological EMD are proposed: a) by varying the family of structuring functions, b) by varying the pair of morphological operators used to calculate the envelopes, and c) by considering a convex sum of envelopes instead of the mean point used in classical EMD. We discuss in particular the invariances that are induced by the morphological EMD representation. Experimental results on supervised classification of hyperspectral images by 1D convolutional networks demonstrate the interest of our method.

Key words. Deep Learning, Mathematical morphology, Hyperspectral image processing

AMS subject classifications. 68U10, 94A12, 68T07

1. Introduction.

Deep convolutional neural networks (DCNN) provide state-of-the-art results in many tasks for signal and image classification [4]. The DCNN architectures combine low complexity signal/image operators, like convolution with small kernels or pooling estimation, with the ability to optimize the corresponding weights of the operators in evolved and hierarchical networks. Traditional models for signal/image representation and associated feature extraction are generally not compatible with the DCNN paradigm. The main limitation is the incompatibility of the backpropagation principle used to train the parameters of the neural networks by gradient descent algorithms. In the case of traditional signal/image processing, the interpretability of the operators and features is often straightforward. We focus here in particular in the *Empirical Mode Decomposition* (EMD) [23], which is a simple and powerful technique used to represent the features of a signal (without any assumption on its frequency content) from a geometric viewpoint, basically using lower and upper envelopes of the signal in an iterated decomposition. The two main ingredients of EMD: detection of local extrema and the interpolation between them, are not naturally formulated in the neural network paradigm. Inspired by the work of Diop and co-workers [12, 11, 13], we revisit EMD using morphological operators to deal with lower/upper envelopes. Additionally, we propose three generalizations: a) by varying the family of structuring functions, b) by varying the pair of morphological operators used to calculate the envelopes, and c) by considering a convex

---

*

Funding: This work was funded by the Fondation Jacques Hadamard under PGMO-IRSDI 2019 program.

[†]CMM, MINES ParisTech, PSL Research University, France (santiago.velasco@mines-paristech.com, http://cmm.ensmp.fr/~velasco/).

[‡]École Centrale de Lyon, France (romain.pages@ecl18.ec-lyon.fr) .

[§]CMM, MINES ParisTech, PSL Research University, France (angulo@mines-paristech.com, http://cmm.ensmp.fr/~angulo/) .

sum of envelopes instead of the mean point used in classical EMD. All the parameters of our proposition can be learnt using backpropagation and gradient descent techniques and therefore the associated morphological EMD can be integrated into standard DCNN representations for end-to-end learning. The integration of morphological operators into DCNN pipelines is an active research area. First attempts were based on approximation of dilation and erosion using standard convolution [32]. More recently, straightforward approaches of dilation and erosion optimization have been explored [14, 33, 37]. However, plugging morphological operators into standard networks is far from being trivial from the optimization based on backpropagation of gradients through all layers by the chain rule. Max-plus operators are indeed differentiable only on a local and specific domain. Here we focus on standard gradient descent strategies and we provide a better understanding of how the gradient of morphological operators, in particular those associated to parametric structuring functions, is defined. Additionally, we show that our morphological EMD induces the invariance to additive shift in standard DCNN. To the best of our knowledge, these technical aspects have not been previously discussed in the field of morphological deep neural networks.

1.1. Related work. In what follows we review the state-of-the-art that is most relevant for the proposed morphological EMD.

1.1.1. Empirical Mode Decomposition. EMD is an algorithm introduced by Huang et al. [23] for analysing linear and non-stationary time series. It is a way to decompose a signal in order to obtain instantaneous frequency data. In this original version of the EMD is an iterative process which decomposes real signals f into simpler signals (modes), $f(x) = \sum_{i=1}^{M} \Phi_j(x)$, where each *mono-component* signal Φ should be written in the form $\Phi(x) = r(x)\cos(\theta x)$, where the amplitude and phase are both physically and mathematically meaningful [48]. Unlike some other common transforms like the Fourier transform for example, the EMD was built as an algorithm and lacks theoretical background then. The problem of EMD to represent a signal as a sum of amplitude modulation (AM) and frequency modulation (FM) components at multiple scales was first proposed in [31] where the problem of finding the AM-FM components and their envelopes was solved using multiscale Gabor filters and nonlinear Teager-Kaiser Energy Operators via an Energy Separation Algorithm (ESA). In the original EMD, there is no parametric family of filters used to estimate the envelopes.

From an algorithmic point of view, the EMD is obtained following the iterative process [23]:

1. Find all the local extrema of the function f.
2. Interpolate all the local maxima together to get the function $\hat{f}$ (upper envelope), and all the local minima together to get the function $\check{f}$ (lower envelope)
3. Calculate the *local mean* as the average of the both interpolations; the obtained function is called *Intrinsic Mode Function*:
$$IMF(x) = \frac{1}{2}\left(\hat{f}(x) + \check{f}(x)\right)$$
4. Iterate this process (that is called the *sifting process*) on the residual, *i.e.*,
$$r(x) = f(x) - IMF(x)$$
until a selected tolerance criterion is respected.

Thus, the original signal is decomposed as:

$$f(x) = \sum_{k=1}^{n} IMF_k(x) + r(x) \tag{1.1}$$

where IMF_k is the k-th intrinsic mode function and r is the last residual. The EMD can be efficiently applied to 1D-signals. However the selection of interpolation method for the second step gives a wide variety of possibilities, from the original formulation using cubic splines [23], passing by sparse filtering [22], filtering from wavelet based decomposition [15] and partial differential equation based formulations [10].

The EMD method can be justified only under certain very restrictive assumptions that are seldom satisfied by practical data. The EMD method is also known to be very sensitive to noisy data. Recently, a compendium of practical advice for EMD in real life examples has been presented in [50]. Some works extend EMD to 2D [12, 49, 11] and 3D images [18]. However, the main limitations of EMD for both 2D and 3D are both the choice of maxima and minima detector, and the choice of the interpolation algorithm.

An alternative characterisation of the EMD computation was introduced by Diop *et al.* in [12, 13] according to the definition of *local mean*, *i.e.*, the sifting process is fully determined by the sequence $(h_n)_{n\in\mathbb{N}}$ defined by :

$$\begin{cases} h_{n+1} = h_n - \Phi(h_n) = (\mathrm{Id} - \Phi)\, h_n \\ h_0 = f \end{cases} \tag{1.2}$$

where $\Phi(h_n) = \frac{\hat{h}_n + \check{h}_n}{2}$, and $\hat{h}_n$ (resp. $\check{h}_n$) denotes a continuous interpolation of the maxima (resp. minima) of h_n.

In the following subsection, we formulated an EMD by means of dilation and erosion operators.

1.1.2. Dilation/Erosion. We study here functions $f : E \to \overline{\mathbb{R}}$, where $\overline{\mathbb{R}}$ it allowed to be *extended-real-valued*, *i.e.*, to take values in $\overline{\mathbb{R}} = [-\infty, \infty]$. Accordingly, the set of all such functions is denoted by $\mathcal{F}(E, \overline{\mathbb{R}})$. We will use the two basic morphological operators *dilation* and *erosion*, which correspond respectively to the convolution in the $(\max, +)$ algebra and its dual.

Definition 1.1. *In mathematical morphology [47], the dilation (sup-convolution) $\delta_{SE}(f)$ of f is given by:*

$$\delta_{SE}(f)(x) := \sup_{y\in E} \{f(y) + SE(x-y)\} = \sup_{w\in E} \{f(x-w) + SE(w)\} \tag{1.3}$$

where $SE \in \mathcal{F}(E, \overline{\mathbb{R}})$ is the (additive) structuring function which determines the effect of the operator. Here the inf-addition rule $\infty - \infty = \infty$ is to be used in case of conflicting infinities. $\sup f$ *and* $\inf f$ *refer to the* supremum *(least upper bound) and* infimum *(greatest lower bound) of f. In the discrete case where the function is a finite set of points,* max *and* min *are used.*

The erosion [47] $\varepsilon_{\mathrm{SE}}(f)$, known as inf-convolution in convex analysis [35], is the adjoint operator to the dilation (1.3), and it is defined as

$$\varepsilon_{\mathrm{SE}}(f)(x) := -\delta_{\check{\mathrm{SE}}}(-f)(x) = \inf_{y\in E} \{f(y) - \mathrm{SE}(y-x)\} = \inf_{w\in E} \{f(x+w) - \mathrm{SE}(w)\} \tag{1.4}$$

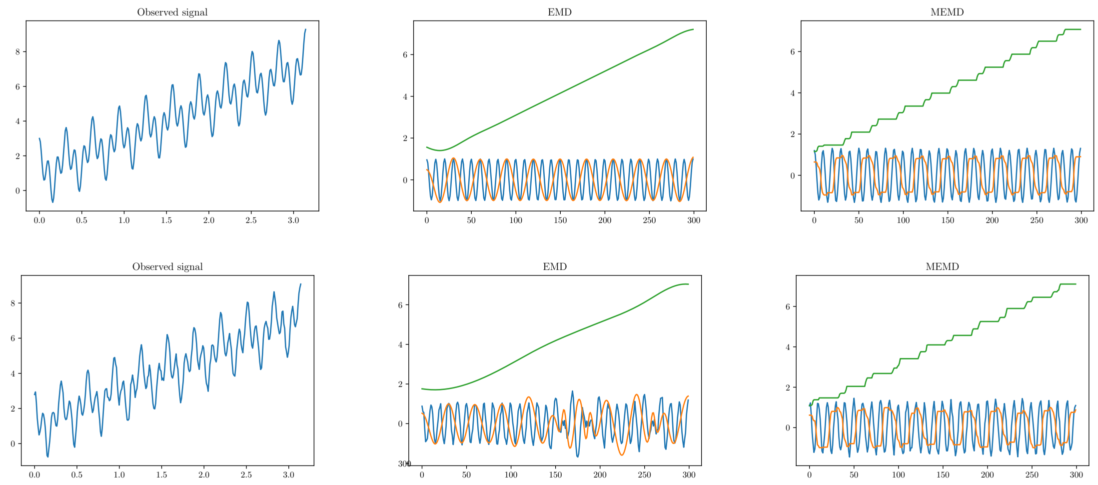


Figure 1. *First Row: Noise-free example a)* $f(x) = 2x+1+cos(20x)+cos(60x)$*, b) Classical EMD c) MEMD with flat structuring functions. Second Row: Noisy example a)* $f(x) = 2x+1+cos(20x)+cos(60x)+N(0, \frac{1}{8})$*, b) Classical EMD c) MEMD with flat structuring functions.*

where the transposed structuring function is $\check{\text{SE}}(x) = \text{SE}(-x)$.

Remark 1.2. $\forall f, g \in \mathcal{F}(E, \overline{\mathbb{R}})$

1. The operators (1.3) and (1.4) are translation invariant.
2. (1.3) and (1.4) correspond to one another through the duality relation $\delta_{\text{SE}}(f)(x) \leq g(x) \iff f(x) \leq \varepsilon_{\text{SE}}(g)(x)$, called *adjunction* [16].
3. An operator ξ is called *increasing* if $f(x) \geq g(x) \Rightarrow \xi(f)(x) \geq \xi(g)(x)\ \forall x$. The dilation (1.3) and erosion (1.4) are increasing for all SE.
4. An operator ξ is called *extensive* (resp. *antiextensive*) if $\xi(f)(x) \geq f(x)$ (resp. $\xi(f)(x) \leq f(x)$), $\forall x$. The dilation (1.3) (resp. erosion (1.4)) is extensive (resp. antiextensive) if and only if $\text{SE}(0) \geq 0$, *i.e.*, the structuring function evaluated at the origin is non-negative.
5. $\varepsilon_{\text{SE}}(f)(x) \leq f(x) \leq \delta_{\text{SE}}(f)(x)$ if and only if $\text{SE}(0) \geq 0$.
6. δ_{SE} (resp. ε_{SE}) does not introduce any local maxima (resp. local minima) if $\text{SE} \leq 0$ and $\text{SE}(0) = 0$. In this case, we say that SE is *centered*.

Proof. (1) and (2) are classical results from [47]. As explained in [20] and [30], the *adjunction* is related to a well-known concept in group and lattice theory, the *Galois connection*. (3) and (6) are easy to prove directly from the definition of the operators. It has been also proved in the original paper of inf-convolution (Proposition 6.d) in [35]. (4) $\forall f, \delta_{\text{SE}}(x) \geq f(x) \Rightarrow \forall f, \sup\left(f(x-w) + \text{SE}(w) - f(x)\right)(x) \geq 0 \Rightarrow \text{SE}(0) \geq 0$. Now, $\sup f(x-w) + \text{SE}(w) \geq f(x) + SE(0)$, if $\text{SE}(0) \geq 0 \Rightarrow \sup f(x-w) + \text{SE}(w) \geq f(x)$. From (3) and (4) is easy to prove (5). ■

The most commonly studied framework for dilation/erosion of functions is based on *flat structuring functions*, where structuring elements are viewed as *shapes*. More precisely, given the structuring element $B \subseteq E$, its associated structuring function is

$$B(y) = \begin{cases} 0 & \text{if } y \in B \\ -\infty & \text{if } y \in B^c \end{cases} \tag{1.5}$$

Hence, the flat dilation $\delta_B(f)$ and flat erosion $\varepsilon_B(f)$ can be computed respectively by the moving local maxima and minima filters. The shape of B is often a disk of radius λ, denoted by B_λ.

$$B_\lambda(w) = \begin{cases} 0 & \text{if } \|w\| \leq \lambda \\ -\infty & \text{if } \|w\| > \lambda \end{cases} \tag{1.6}$$

A Morphological Empirical Mode Decomposition (MEMD) where the pair $(\hat{h}, \check{h})$ correspond to $(\varepsilon_{B_\lambda}, \delta_{B_\lambda})$ has been proposed in [13].

Definition 1.3. *The Flat Morphological Empirical Mode [13] is defined as*

$$\Phi_{\varepsilon,\delta,B_\lambda}(f)(x) := \frac{\delta_{B_\lambda}(f)(x) + \varepsilon_{B_\lambda}(f)(x)}{2} \tag{1.7}$$

The operator (1.7) was proposed to generate an EMD based on solving a morphological PDE [13]. As a manner of example, EMD and MEMD are shown for a mono-component signal in the first row of Figure 1. In the second row of Figure 1, we illustrated how the addition of noisy perturbed more the results of classical EMD than the proposed morphological one.

Remark 1.4. Note that using (1.7) twice, the first residual (1.2) is $2(f - \Phi_\lambda(f)) = (f - \delta_{B_\lambda}(f)) + (f - \varepsilon_{B_\lambda}(f)) = 2f - \delta_{B_\lambda}(f) - \varepsilon_{B_\lambda}(f)$. This expression, up to a minus sign, corresponds just to the so-called *morphological Laplace operator [53]*, and therefore provides an interpretation of the EMD as an iterated second-order derivative decomposition of the function f.

1.2. Our proposal. The main motivation of this paper is to define EMD learnable in the sense of neural networks approaches. Note that last property in Remark 1.2 together with the extensivity/antiextensivity (*i.e.*, upper/lower envelopes) imply that the pair of operators $(\varepsilon_{\mathrm{SE}}, \delta_{\mathrm{SE}})$ are candidate functions for $(\hat{h}, \check{h})$ in (1.2). Accordingly, we proposed a simple generalization by considering non-flat structuring functions.

Definition 1.5. *The Morphological Empirical Mode (MEM) is defined as*

$$\Phi_{\varepsilon,\delta,SE}(f) = \frac{\delta_{SE}(f)(x) + \varepsilon_{SE}(f)(x)}{2} \tag{1.8}$$

This operator can be formulated in any dimension (from 1D to nD signals) and avoid using an interpolation method which is the bottleneck of the original definition of EMD.

1.3. Contributions of the paper. In what follows we study,

- A formulation of EMD based on pairs of morphological operators in a general case.
- The proposition of a parametric morphological empirical mode whose sifting process is invariant to additive intensity shifts.

- A approach to learn the structuring functions of a morphological operator in a deep learning framework.
- A convex sum of envelopes instead of mean points to learn morphological EMD.
- A number of numerical experiments for hyperspectral signal classification to illustrate the relevance of our proposal.

1.4. Organization of the paper. The rest of the paper is organised as follows. In section 2, we review the general definition of Empirical Mode Decomposition approach to decompose signals and we introduce how morphological extensive/antiextensive filters are naturally adapted to implement a MEMD computation. We consider different possibilities in the choice of structuring functions and the pair of lower and upper envelopes. Additionally, an α-MEM is proposed as a generalization of the mean of envelopes. Section 3 is devoted to the implementation of morphological EMD operators as layers in a neural network pipeline. Section 4 presents the experimental results of hyperspectral image classification using DCNNs which integrate morphological EMD layers. Conclusions and perspectives are discussed in section 5.

2. Morphological Empirical Mode and its variants. In this section, three kinds of generalization will be explored: a) different types of structuring functions, b) different pairs of functions to compute the lower and upper envelopes, and c) a convex sum of lower and upper envelopes.

2.1. Varying the structuring function. In this subsection, firstly we will study a parametric family of symmetric quadratic shape structuring functions. Secondly, similarly to classical CNNs, the structuring function plays a similar role to the kernel in standard convolution. Accordingly a structuring function without any parametric constraint is also considered.

2.1.1. Quadratic MEM. From the theory of morphological scale-spaces, the most useful nonflat structuring functions are those which depend on a scale parameter [21, 46]. The only separable and rotationally invariant structuring functions is the called *quadratic structuring function*[51]:

$$q_\lambda(z) = -\frac{\|z\|^2}{2\lambda}, \tag{2.1}$$

such that the corresponding dilation and erosion are equal to the Lax–Oleinik operators or viscosity solutions of the standard Hamilton–Jacobi PDE, also known as morphological PDE: $u_t(t,x) \mp \|u_x(t,x)\|^2 = 0$, $(t,x) \in (0,+\infty) \times E$; $u(0,x) = f(x)$, $x \in E$. It plays also a canonical role in the definition of dilation and erosion on Riemannian manifolds [2] and their behaviour with respect to the maxima/minima is well understood [25]. The morphological PDE was proposed and analyzed using 2D boundary propagation in [52] and further analyzed using the morphological slope transform in [19].

Remark 2.1. The erosion by a quadratic structuring function with parameter λ is defined by

$$\varepsilon_{q_\lambda}(f)(x) := \inf_{y\in E}\{f(y) - q_\lambda(y-x)\} = \inf_{z\in E}\{f(z-x) - q_\lambda(z)\} = \inf_{z\in E}\left\{f(z-x) + \frac{\|z\|^2}{2\lambda}\right\}. \tag{2.2}$$

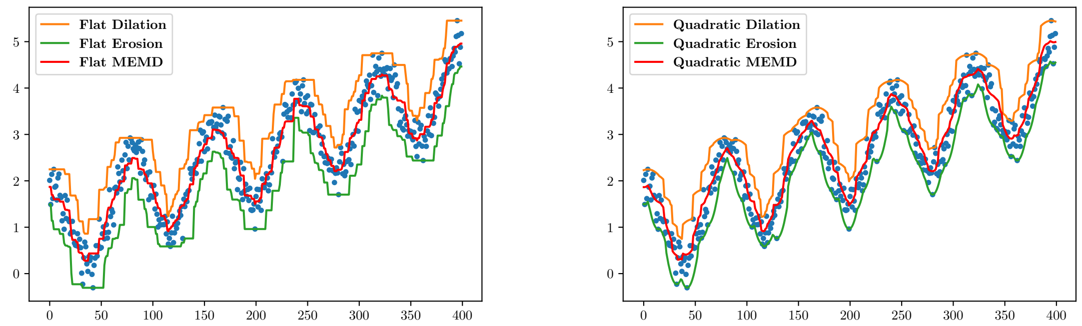


Figure 2. *The blue points corresponds to the observed signal, a) Flat dilation/erosion based Morphological Empirical Mode (1.7) with a disk of $\lambda = 5$, b) Quadratic dilation/erosion based Morphological Empirical Mode (2.3) with $\lambda = 3$.*

The erosion of a function f by a quadratic structuring function with parameter λ is known as the *Moreau envelope* or *Moreau-Yosida approximation* [35, 43, 40], which offers many benefits specially for optimization purposes [34]. Additionally, (2.2) induces an additive scale-space [19, 24], *i.e.*, $\varepsilon_{q_{\lambda_1}}(\varepsilon_{q_{\lambda_2}}(f)) = \varepsilon_{q_{\lambda_1+\lambda_2}}(f)$.

Definition 2.2. *The quadratic morphological empirical mode (QMEM) is defined as a MEM where the pair $(\hat{h}, \check{h})$ corresponds to erosion/dilation with a quadratic structuring functions,*

$$\Phi_{\varepsilon,\delta,q_\lambda}(f) = \frac{\varepsilon_{q_\lambda}(f) + \delta_{q_\lambda}(f)}{2}. \tag{2.3}$$

An example of (2.3) for a 1D signal with noise is shown in Figure 2.

2.1.2. Nonflat Morphological MEM.

The most general case of *nonflat structuring function* involves different additive weights $W_y(x)$ at each position x of the local neighborhood B centered at pixel y, *i.e.*, a nonflat structuring function SE_W of support shape B at y is defined as

$$\mathrm{SE}_{W_y}(x) = \begin{cases} W_y(x) & \text{if } x \in B(y) \\ -\infty & \text{otherwise} \end{cases} \tag{2.4}$$

The case (2.4) includes flat, nonflat, either local or nonlocal structuring functions [54]. In the translation invariant case, the weighting function $W_y(x)$ is equal for all $y \in E$.

2.2. Varying the Envelope.

We have explored above several possible structuring functions that produce multiple pairs of $(\varepsilon_{\mathrm{SE}}, \delta_{\mathrm{SE}})$ as basic ingredient for the Morphological Empirical Mode (1.8). At this point, we can consider the use of the composition of erosion and dilation to obtain other upper/lower envelopes, typically of the form $(\delta_{\mathrm{SE}} \circ \varepsilon_{\mathrm{SE}}, \varepsilon_{\mathrm{SE}} \circ \delta_{\mathrm{SE}})$.

2.2.1. Opening/Closing MEM.

The theory of morphological filtering is based on the opening $\gamma_{\mathrm{SE}}(f)(x)$ and closing $\varphi_{\mathrm{SE}}(f)(x)$ operators, obtained respectively by the composition product of erosion-dilation and dilation-erosion, *i.e.*, $\gamma_{\mathrm{SE}}(f)(x) = \delta_{\mathrm{SE}}(\varepsilon_{\mathrm{SE}}(f))(x)$ and

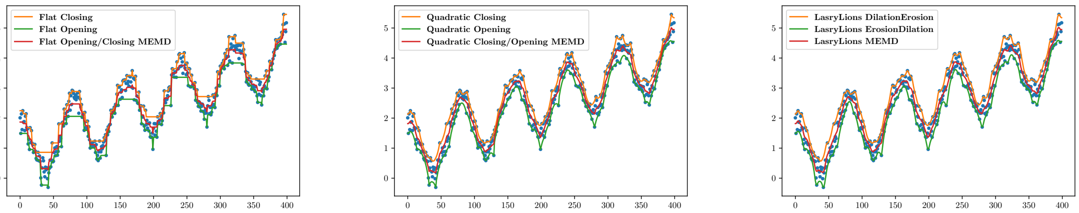


Figure 3. *a) Flat OCMEM with a disk of $\lambda = 5$, b) Quadratic OCMEM with $\lambda = 3$ and c) Lasry-Lions MEM with $\lambda = 3$ and $c = .9$*

$\varphi_{\mathrm{SE}}(f)(x) = \varepsilon_{\mathrm{SE}}\left(\delta_{\mathrm{SE}}(f)\right)(x)$. Opening (resp. closing) is increasing, idempotent and anti-extensive (resp. extensive), independently of the properties of the structuring function. The opening can be seen as the supremum of the invariant parts of f under-swept by SE and it can be again rewritten as a maximal lower envelope of structuring functions (resp. minimal upper envelope of negative symmetric structuring functions). We highlight that the *quadratic envelope* also called as *proximal hull* [7] is an opening with a quadratic structuring function, *i.e.*, a quadratic erosion followed by a quadratic dilation.

Definition 2.3. *The opening/closing morphological empirical mode (OCMEM) is defined as a MEM where the pair $(\hat{h}, \check{h})$ corresponds to $(\gamma_{SE}, \varphi_{SE})$,* i.e.,

$$\Phi_{\gamma,\varphi,SE}(f) = \frac{\gamma_{SE}(f) + \varphi_{SE}(f)}{2}. \tag{2.5}$$

For the case of flat disks B_λ, the operator (2.5) was called a morphological locally monotonic (LOMO) filter in [5]. A signal is monotonic over an interval if it is either non-increasing or non-decreasing over that interval. A 1-D signal is *locally monotonic* of degree n (LOMO-n) if and only if the signal is monotonic within every interval of length n. In the general case, a LOMO filter of f is defined as the fixed point of iterating $\Phi_{\gamma,\varphi,B_\lambda}(f)$, which is simultaneously idempotent to both the opening and closing by a flat disk as structuring function. Two examples of (2.5) for both flat and quadratic structuring function for the 1D signal with noise are shown in Figure 3.

2.2.2. Lasry–Lions MEM. Besides their feature extraction properties, morphological dilation and erosion using quadratic structuring functions are a powerful tool for Lipschitz regularization. For the nonconvex case, the Lasry–Lions double envelope is defined as the composition of two different Moreau envelopes, or using the morphological vocabulary, the composition of an erosion followed by a dilation with quadratic structuring functions. For all $0 < c < 1$ and $0 < \lambda$, the so-called Lasry–Lions regularizers [27] are defined as

$$\gamma_\lambda^c(f)(x) := \delta_{q_{c\lambda}}\left(\varepsilon_{q_\lambda}(f)\right)(x),$$
$$\varphi_\lambda^c(f)(x) := \varepsilon_{q_{c\lambda}}\left(\delta_{q_\lambda}(f)\right)(x),$$

such that if f is bounded, the functions γ_λ^c and φ_c^λ are bounded and one has the ordering properties for the following envelopes:

- if $\lambda_1 \geq \lambda_2 > 0$, for any $0 < c < 1$ then

$$\gamma^c_{\lambda_1}(f)(x) \leq \gamma^c_{\lambda_2}(f)(x) \leq f \leq \varphi^c_{\lambda_2}(f)(x) \leq \varphi^c_{\lambda_1}(f)(x);$$

- if $0 < c_2 < c_1 < 1$, for any $\lambda > 0$ then

$$\gamma^{c_2}_{\lambda}(f)(x) \leq \gamma^{c_1}_{\lambda}(f)(x) \leq f \leq \varphi^{c_1}_{\lambda}(f)(x) \leq \varphi^{c_2}_{\lambda}(f)(x).$$

For any bounded function f, Lasry–Lions regularizers provide a function with a Lipschitz continuous gradient, *i.e.*,

$$|\nabla\gamma^c_\lambda(f)(x) - \nabla\gamma^c_\lambda(f)(y)| \leq M_{\lambda,c}\|x-y\|, \quad |\nabla\varphi^c_\lambda(f)(x) - \nabla\varphi^c_\lambda(f)(y)| \leq M_{\lambda,c}\|x-y\|.$$

where the Lipschitz constant is $M_{\lambda,c} = \max\left((c\lambda)^{-1}, ((1-c)\lambda)^{-1}\right)$. If f is bounded and Lipschitz continuous, one has

$$\mathrm{Lip}(\gamma^c_\lambda(f)) \leq \mathrm{Lip}(f) \text{ and } \mathrm{Lip}(\varphi^c_\lambda(f)) \leq \mathrm{Lip}(f),$$

with

$$\mathrm{Lip}(g) = \sup\left\{\frac{|g(x)-g(y)|}{\|x-y\|};\ x, y \in \mathbb{R}^n,\ x \neq y\right\}.$$

For more details on the properties of Lasry–Lions regularizers in the context of mathematical morphology, see [1].

Remark 2.4. The following statements are interesting about the composition of quadratic morphological operators [43, 9]. Let $0 < \mu < \lambda$,

1. $\varepsilon_{q_\lambda}(\gamma_{q_\lambda}(f)) = \varepsilon_{q_\lambda}(f)$;
2. $\gamma_{q_\mu}(\varepsilon_{q_{\lambda-\mu}}(f)) = \varepsilon_{q_{\lambda-\mu}}(\gamma_{q_\lambda}(f))$;
3. $\gamma_{q_{\lambda-c\lambda}}\varphi^c_\lambda(f) = \varphi^c_\lambda(f)$.

Definition 2.5. *The Lasry-Lions morphological empirical mode (LLMEM) is defined as a MEM where the pair* $(\hat{h}, \check{h})$ *corresponds to* $(\gamma^c_\lambda, \varphi^c_\lambda)$, i.e.,

$$\Phi_{\gamma,\varphi,c,\lambda}(f) := \frac{\gamma^c_\lambda(f) + \varphi^c_\lambda(f)}{2}. \tag{2.6}$$

An example of (2.6) for a 1D signal is shown in Figure 3(c).

2.3. Parametric family of morphological empirical mode operators. The choices of the structuring function and the class of lower and upper envelopes give extra possibilities for the formulation of an EMD approach. Besides, a third degree of freedom is considered now by including a parameter to weight the contribution of the two envelopes. We have been inspired by the recent work on proximal average [9] to propose a convex generalization of MEMs.

Definition 2.6. *Let α be a real value with $0 \leq \alpha \leq 1$, the α-Morphological Empirical Mode based on the pair $(\check{h}, \hat{h})$ is defined as:*

$$\Phi^\alpha_{\check{h},\hat{h}}(f) = \alpha\hat{h}(f) + (1-\alpha)\check{h}(f). \tag{2.7}$$

Definition 2.7. *Let $T_g : \mathcal{F}(E, \overline{\mathbb{R}}) \mapsto \mathcal{F}(E, \overline{\mathbb{R}})$ be a set of transformations on the space E for the abstract group $g \in G$. We say a function ϕ is* invariant *to g if for all transformations T_g, and for all $f \in \mathcal{F}(E, \overline{\mathbb{R}})$ one has*

$$\phi(T_g(f)) = \phi(f) \tag{2.8}$$

This says that the feature extracted by ϕ does not change as the transformation is applied to the input.

In this context, an important fact to consider are the invariances of the operator (2.7).

Remark 2.8. For any SE, $\forall 0 \leq \alpha \leq 1$, and all the pairs $(\check{h}, \hat{h})$ previously considered, the operator (2.7) is increasing, invariant to translation, and the sifting process $f - \Phi^{\alpha}_{\check{h},\hat{h}}(f)$ is invariant to additive intensity shifts, *i.e.*, $\forall c \in \mathbb{R}$ and $\forall f \in \mathcal{F}(E, \overline{\mathbb{R}})$,

$$(f(x) + c) - \Phi^{\alpha}_{\check{h},\hat{h}}(f(x) + c) = f(x) - \Phi^{\alpha}_{\check{h},\hat{h}}(f(x)).$$

3. Learnable Morphological Empirical Mode Decomposition.

One of the main advantages of EMD is that it can be considered as a parameter-free decomposition [50] and, for this reason, the inclusion of the structuring function and the parameter α can be seen as inconvenient. However, in the following, we consider EMD in the context of learning from data [29], where one would be interested in using EMD decomposition as a preprocessing of an input signal before using a machine learning or deep learning methods [42, 3, 26].

3.1. Neural network-based learning of parameters.

The simplest form of a neural network is the called *multilayer architecture*, which is a stack by composition of modules, each module implements a function $X_n = F_n(\theta_n, X_{n-1})$, where X_n is a vector representing the output of module, θ_n is the vector of learnable parameters in the module, and X_{n-1} is the module input vector (as well as the output of the previous module). The input of the first module X_0 is an input pattern Z_0, the output of the whole system is the one of the last module which denoted Z_l, where l is the *number of layers*. In *gradient-based learning methods*, given a cost function $\mathcal{L}^p(\cdot, \cdot)$ measuring the discrepancy between the output of the system Z_l^p and D^p the "correct" or desired output for the p-th input pattern. One is interested on minimizing the average discrepancy over a set of input/output pairs called the *training set*, $\{(Z_0^0, D^0), (Z_0^1, D^1), \ldots, (Z_0^n, D^n)\}$. The network is initialized with randomly chosen weights θ^0. The gradient of the error function with respect to each parameter is computed and gradient descent is used to update the weights in each layer, *i.e.*, for the i-th iteration, $\theta^{i+1} = \theta^i - \eta \frac{\partial \mathcal{L}(\theta)}{\partial \theta^i}$ where η is a learning rate, and the computation of $\frac{\partial \mathcal{L}(\theta)}{\partial \theta^i}$, is performed by *backpropagation algorithm* through the layers [44]. Additionally, for data structured like images, *convolutional neural networks* (CNN) are nowadays the recommended solution. In CNNs, the same operator is computed in each pixel of the image. This mechanism is called *weight sharing*, and it has several advantages such as it can reduce the model complexity and make the network easier to train [38]. Including any new layer, like EMD, requires therefore the computation of the corresponding gradient of the layer with respect to the parameters to be learnt.

3.2. Derivatives of Morphological EMD in discrete domains.

3.2.1. Derivative of dilation and erosion. Our approach involves dilation and erosion operators as defined in (1.3) and (1.4). However, in the discrete domain as it is the case of nD images, the sup operator is computed via max. Consequently, for dilation operator (1.3), is computed by $\delta_\lambda(x) = \max_w \{f(x-w) + \mathrm{SE}_\lambda(w)\}$. To understand how to compute the derivative of $\delta_\lambda(x)$ with respect to λ, we rewrite $\delta_\lambda(x) = \max_{w \in \mathrm{SE}_\lambda} u(w)$. The max operator has no gradient with respect to non-maximum values, since changing them slightly does not affect the output. In general for rank operators, their derivative is zero in every coordinate, except for that of the value attending the desired rank [41, 36]. Accordingly, the derivative with respect to a parameter in the additive structuring function is given by

$$\frac{\partial \delta_\lambda(x)}{\partial \lambda} = \frac{\partial \delta_\lambda(x)}{\partial u(w)} \frac{\partial u(w)}{\partial \lambda} = \begin{cases} \frac{\partial \mathrm{SE}_\lambda(w)}{\partial \lambda} & \text{if } w \in \arg\max_x \delta_\lambda(x) \\ 0 & \text{otherwise} \end{cases} \tag{3.1}$$

where the operator $\arg\max_x f(x) := \{x \mid \forall y : f(y) \leq f(x)\}$. In other words, arg max is the set of points x, for which $f(x)$ attains the largest value of the function. Note that we do not regard maximum as being attained at any x when $f(x) = \infty$, nor do we regard the minimum as being attained at any x when $f(x) = -\infty$. Similarly for the erosion, $\varepsilon_\lambda(x) = \min_w [f(x+w) - \mathrm{SE}_\lambda(w)] = \min_{w \in \mathrm{SE}_\lambda} u(w)$

$$\frac{\partial \varepsilon_\lambda(x)}{\partial \lambda} = \frac{\partial \varepsilon_\lambda(x)}{\partial u(w)} \frac{\partial u(w)}{\partial \lambda} = \begin{cases} -\frac{\partial \mathrm{SE}_\lambda(w)}{\partial \lambda} & \text{if } w \in \arg\min_x \varepsilon_\lambda(x) \\ 0 & \text{otherwise} \end{cases} \tag{3.2}$$

there is only gradient with respect to minimum values. As a manner of example, for the dilation by quadratic structuring element (2.1), one has

$$\frac{\partial q_\lambda(z)}{\partial \lambda} = (2\lambda^2)^{-1} \|z\|^2 \implies \frac{\partial \delta_\lambda(x)}{\partial \lambda} = \begin{cases} \frac{\|w\|^2}{2\lambda^2} & \text{if } w \in \arg\max_x \delta_\lambda(x) \\ 0 & \text{otherwise} \end{cases}$$

Therefore, for Quadratic EMD (2.3) the derivative with respect of λ,

$$\frac{\partial \Phi_\lambda(x)}{\partial \lambda} = \frac{\|w_\delta\|^2 - \|w_\varepsilon\|^2}{4\lambda^2},$$

where $w_\delta \in \arg\max_x \delta_\lambda(x)$ and $w_\delta \in \arg\min_x \varepsilon_\lambda(x)$. Thus, the evolution of the parameter λ depends on the difference of the norm to the value where the morphological operator attains their value, normalised by the square of the current value of λ. Curiously the nonflat translation invariant MEM (2.4) has a derivative that does not depend on the scale of the parameters, *i.e*, for $\mathrm{SE}_W = [w_0, \ldots, w_k]$,

$$\frac{\partial \Phi_{\mathrm{SE}_W}(x)}{\partial w_i} = \begin{cases} 1/2 \text{ if } & w_i \in \arg\max_x \delta_{SE_W}(x) \\ -1/2 \text{ if } & w_i \in \arg\min_x \varepsilon_{SE_W}(x) \\ & 0 \text{ otherwise} \end{cases} \tag{3.3}$$

Finally, the derivative for composition operators, as opening or closing, can be easily computed by the chain rule.

3.3. Implementation. Different methods for learning morphological operators in neural networks have been proposed in the literature:

1. Replace maximum and minimum operator by smooth differentiable approximations, making possible the use of conventional gradient descent learning approach via back-propagation, for instance using an approximation by counter-harmonic mean [32] or other generalizations [28].
2. Morphological operations can be computed by combinations of depthwise and point-wise convolution with depthwise pooling [37] allowing the use of classical optimization procedures.
3. Use original definition of morphological operator, and in the backpropagation step follows the approach used in max-pooling layers [6, 14, 33].

We follow the last approach. That means that the gradient in (3.1) and (3.2) will have values different from zero *only* for the first element equal to the arg max or arg min instead of the complete equivalence class. This is the implementation used in deep learning modules based on Tensorflow or Pytorch. An implementation of our approach is available in http://www.cmm.mines-paristech.fr/∼velasco/morpholayers/

3.3.1. Example of learning parameters in morphological operators. We present a dummy example of supervised classification in two classes for 1D signals of dimension p. Both classes have been generated by the function $f(x) = \sin(\frac{2\pi}{c}(x+\epsilon))$), for $x = 0, \ldots, 10$, with spatial step of 0.02 and where ϵ is a random realisation of a normalized Gaussian distribution. For the first class, we have used a period $c = 2$ and for the second class a period $c = 1.75$. Some examples are illustrated in Figure 4(a). We explore the training process by using a simple architecture: $\hat{z} := model(x) = \frac{1}{1+\exp(-\frac{1}{p}\sum_{i=1}^{p}\delta_\lambda(x_i))}$, *i.e*, a morphological dilation followed by a global average pooling with a sigmoid activation function, also called the logistic function. Now, we want to show the computation of the partial derivative with respect to a given loss function. As a manner of example, we use the mean squared error as a loss function, *i.e.*, $loss(z,\hat{z}) = (z-\hat{z})^2$. One can compute the gradient $\frac{\partial loss(z,\hat{z})}{\partial \lambda}$ by using the chain rule of derivative

$$\frac{\partial loss(z,\hat{z})}{\partial \lambda} = \frac{\partial loss(z,\hat{z})}{\partial \hat{z}} \frac{\partial \hat{z}}{\partial \sigma} \frac{\partial \sigma}{\partial \sum \delta_\lambda/p} \frac{\partial \sum \delta_\lambda/p}{\partial \lambda},$$

where $\sigma(x) := \frac{1}{1+\exp(-x)}$ is the sigmoid function. Remember that the derivative of the sigmoid function is $\sigma(x)\sigma(1-x)$. By defining $m = \sum_{i=1}^{p}\delta_\lambda(x_i)/p$, the mean value of the dilation, which is used as decision function, the derivative of the parameter of the dilation with respect to the loss function can be written by

$$\frac{\partial loss(z,\hat{z})}{\partial \lambda} = \frac{(2m)(m(1-m))}{p}\sum_{i=1}^{p}\frac{\partial \delta_\lambda(x)}{\partial \lambda}.$$

The first term is computed in the forward pass and it is the same for every parameter. We decided to train a nonflat structuring function, so from (3.3), one can interpret the second term as a counts the number of number of times that the spatial position in the structuring function attains the maximal value, which is illustrated in Figure 4(c) for the last epoch of the training. Additionally, the evolution of structuring function weights is given in Figure 4(d). As a manner of example, two signals and its corresponding learned dilation are shown for the initialization (as a flat structuring function) in Figure 4(e) and after convergence in Figure 4(f). Finally, the decision function (mean value of the learned dilation) is shown for all the training examples at initialisation Figure 4(g) and after convergence Figure 4(h). We highlight that the learned structuring function seems to be an asymmetric quadratic with a negative shift in the maximum value, which can be considerer as an additive bias. That is the reason why the learned dilation is less than the original function in Figure 4(f).

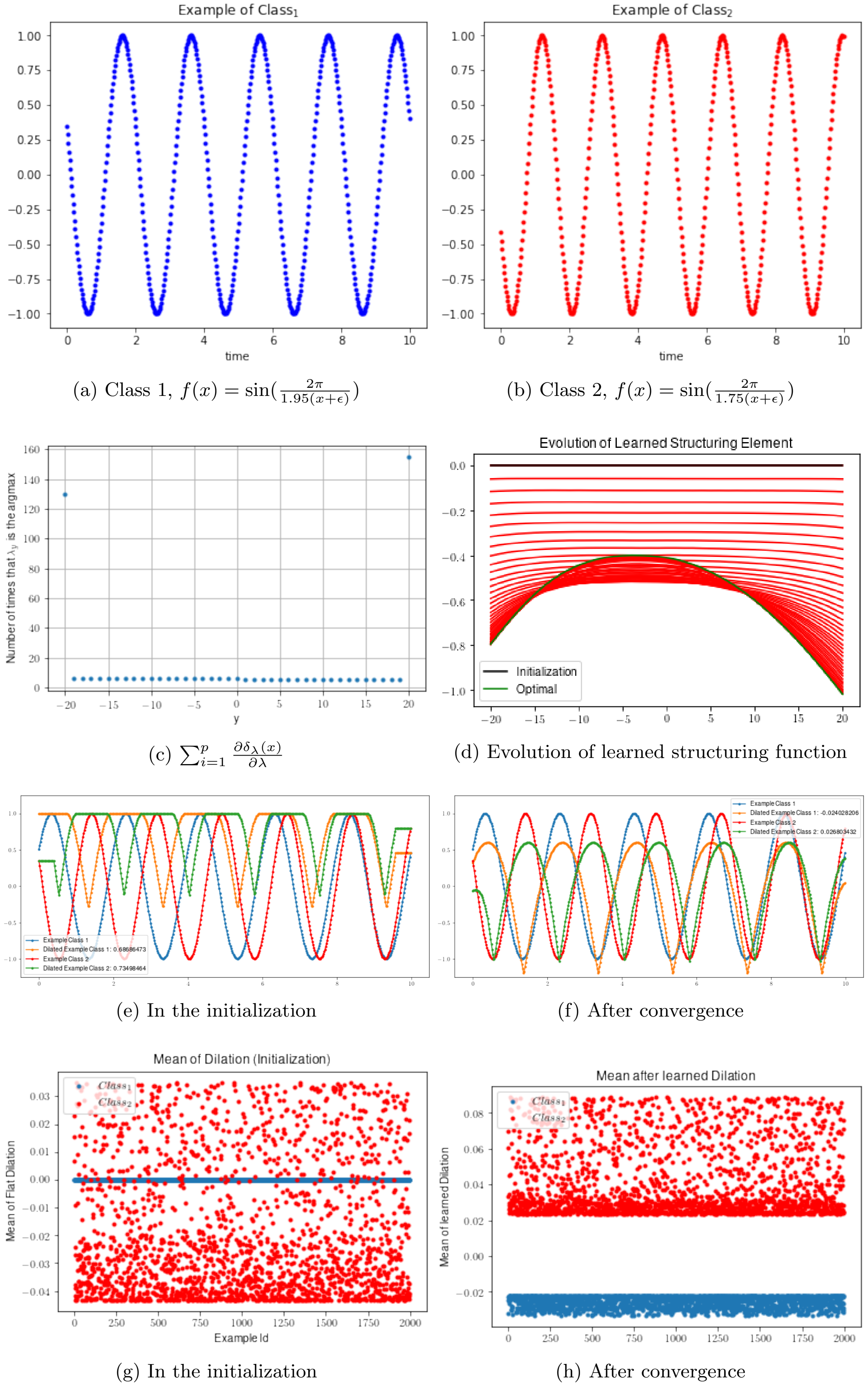
4. Experimental results on hyperspectral classification. In this section, we investigate the application of the proposed morphological empirical mode layer to the problem of signal classification. In particular, we will focus in the case of supervised classification of high-dimensional 1D signals in hyperspectral images. The architecture chosen as baseline is the one recommended in [39] and illustrated in Figure 5. More specifically, the network is composed of convolution layers, RELU, max-pooling. Each stage consists of twenty convolution layers with a kernel size of 24 channels followed by ReLU activation, and a dense layer with batch normalization. In the experimenal section, the proposed morphological empirical mode will be used as the first layer of an architecture of the baseline neural network.

4.1. Considered datasets. The aim of this section is to compare the results obtained by different proposed EMD for 1D supervised classification problems. Accordingly, we used as benchmark two classical hyperspectral images (HIS):

- *Pavia University* hyperspectral is a scene acquired by the ROSIS sensor in the north of Italy. The dataset contains nine different classes including multiple solid structures, natural objects and shadows (Figure 6(a-c)). After discarding the noisy bands, the considered scene contains 103 spectral bands, with a size of 610×340 pixels with spatial resolution of 1.3 mpp and covering the spectral range from 0.43 to 0.86 μm.
- *Indian Pines* dataset is a hyperspectral image captured over an agricultural area characterized by its crops of regular geometry and also irregular forest regions. The scene consists of 145×145 pixels and with 224 spectral bands, which have been collected in the wavelength range from 0.4 to 2.5 μm. There are 16 different classes for training/testing set with a highly unbalanced distribution (Figure 6(d-f)).

4.1.1. Protocol. HSI scenes generally suffer from high intraclass variability and interclass similarity, resulting from uncontrolled phenomena such as variations in illumination, presence of areas shaded and/or covered by clouds, among others. Accordingly, the selection of training samples must be carried out very carefully. Deep learning models for HSI have been traditionally trained by extracting random samples from available ground-truth. However, some works emphasize that the random sampling strategy has a great influence on the reliability and the quality of the solution obtained in HSI [39]. In order to avoid this important issue, we have followed the recommendation of using *spatial-disjoint samples*, *i.e*, we have used a strict spatial separation between training and testing sets, allowing us to compare our models in a difficult and realistic case. The selected training and testing samples have been illustrated in Figure 6(b-c) for Pavia University and (e-f) for Indian Pines datasets.

In gradient descent approaches the selection of random initialization of the parameter



(a) Class 1, $f(x) = \sin(\frac{2\pi}{1.95(x+\epsilon)})$ (b) Class 2, $f(x) = \sin(\frac{2\pi}{1.75(x+\epsilon)})$

(c) $\sum_{i=1}^{p} \frac{\partial \delta_\lambda(x)}{\partial \lambda}$ (d) Evolution of learned structuring function

(e) In the initialization (f) After convergence

(g) In the initialization (h) After convergence

Figure 4. *Evolution in the case of Nonflat structuring function learning in a classification problem based on dilation and average pooling.*

Layer (type)	Output Shape	Param #
InputLayer	(None, 103, 1, 1)	0
conv2d (Conv2D)	(None, 80, 1, 20)	500
max_pooling2d	(None, 16, 1, 20)	0
flatten	(None, 320)	0
dense	(None, 100)	32100
batch_normalization	(None, 100)	400
activation	(None, 100)	0
dense_1 (Dense)	(None, 9)	909
Total params: 33,909		

Layer (type)	Output Shape	Param #
input (InputLayer)	[(None, 103, 1, 10)]	0
conv2d (Conv2D)	(None, 80, 1, 20)	4820
max_pooling2d	(None, 16, 1, 20)	0
flatten (Flatten)	(None, 320)	0
dense (Dense)	(None, 100)	32100
batch_normalization	(None, 100)	400
activation (Activation)	(None, 100)	0
dense (Dense)	(None, 9)	909
Total params: 38,229		

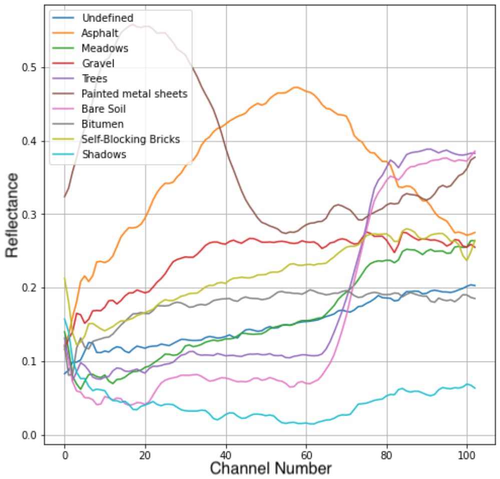
Figure 5. *Baseline architecture vs Baseline architecture applied to EMD. The baseline uses a 20 convolutions 2D with a kernel size of* $(24, 1)$ *followed by a max-pooling reduction of size* $(5, 1)$ *and a RELU activation. For the case presented in the experimental section the same baseline architecture is used. In (b) is the same baseline architecture adapted for ten empirical modes.*

value is critical. The aim of this initialization is to prevent layer activation outputs from exploding or vanishing during the course of a forward pass [17]. While the source of difficulty is well-understood, there is no universal remedy. For our MEM layers, we have used the following initialization:

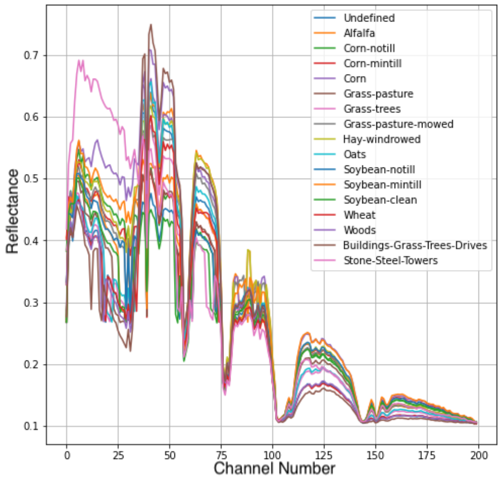
1. For non-flat structuring functions, a flat structuring element, *i.e.*, SE_W is a zero matrix.
2. For quadratic structuring functions, λ is a random realization of a uniform distribution between one and four, and for the parameter c, a uniform distribution between .5 and .95.
3. For the parameter α in (2.7), the value .5 is used.

4.1.2. Quantitative results. We explore the use of proposed EMDs as preprocessing layers, that means instead of learning the classification task from the original spectral signals, we will use the residual of a single step of the decomposition by MEMD. The parameters of the MEMD are learned in a gradient-based learning method. As a manner of comparison, we report in Figure 8 and Figure 9 the accuracy over testing samples for different proposed envelopes by varying both the number of MEM from 10 to 40 and the type of structuring function. Each point is the performance for the best model trained from different random initialization and an early stopping parameter of ten, i.e, we have stopped the training process if it is not improving during ten successive epochs. As it is common in supervised classification problems, we have used categorical cross-entropy as loss function. Additionally, for quantitative comparisons, we have reported best, mean and standard deviation after ten repetitions on both Indian Pines HSI (Table 1) and Pavia University HSI (Table 2). In general, the following results can be highlighted:

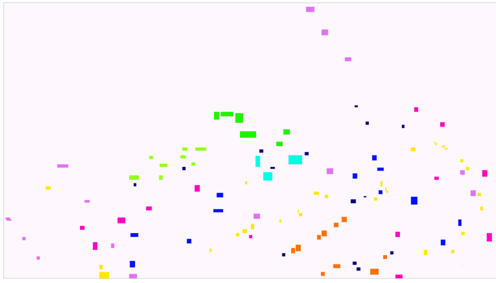
- Learning the parameter in the α-Morphological Empirical Mode (2.7) improves the performance. This can be observed in Table 1) and Pavia University HSI (Table 2) by comparing the performance of models trained with $\alpha = 0.5$ and models where this parameter is learned. Additionally, in Figure 8 and Figure 9 this fact has been highlighted by using different colors in the representation.
- Quadratic MEMDs perform significantly worse than non-flat ones. However, we would like to highlight that the number of parameters is less in the first case.



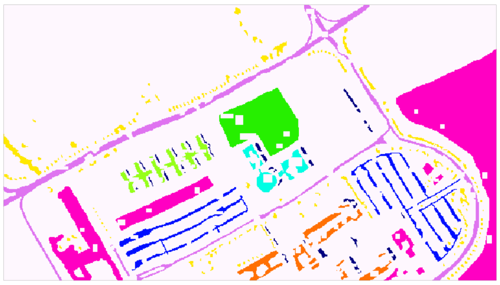
(a) In Pavia University HSI, each data point corresponding to a vector is 103 dimensions. An example per class is shown from the training set in (c).



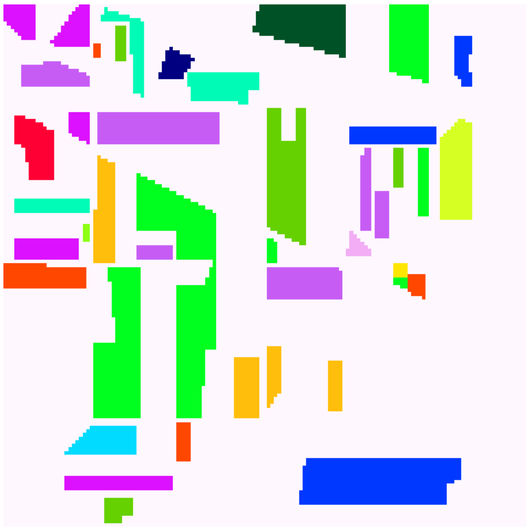
(b) In Indian Pines HSI, each data point corresponds to a vector in 224 dimensions. An example per class is shown from the training set in (e).



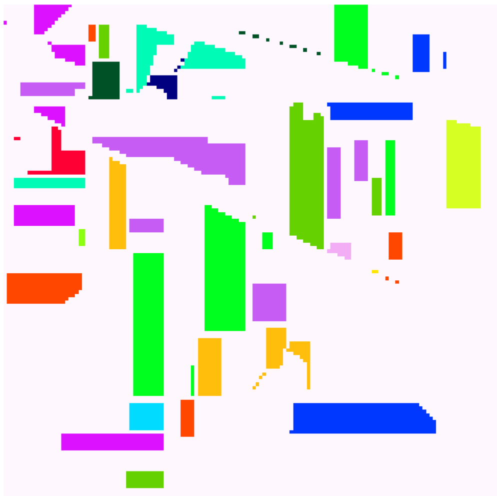
(c) Spatial position of training set in nine classes for Pavia University HSI.



(d) Spatial position of testing set for Pavia University HSI.



(e) Spatial position of training set in 16 classes for Indian Pines HSI.



(f) Spatial position of testing set for Indian Pines HSI.

Figure 6. *For considered HSI dataset, (a) an example per class in Pavia University and (b) Indian Pines. Spatial disjoint distribution of training and testing sets: for Pavia University in (c-d) and for Indian Pines in (e-f). In both cases, white pixels are not considered in the evaluation.*

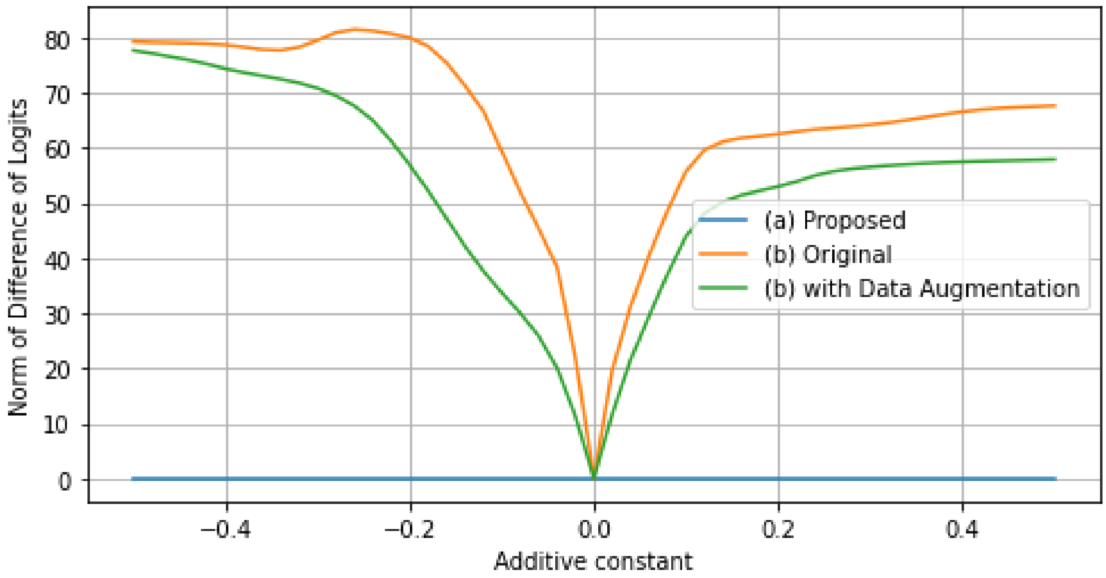


Figure 7. *Analysis of invariance against additive shift for the training sample of Indian Pines. Norm of the Difference in the predictions with and without additive shift,* i.e., $||pred(x) - pred(x + c)||_2^2$ *for different values of c is given for three models: a) MEMD by* (ε, δ)*, b) baseline model, c) baseline model with a data augmentation by random additive constant. We highlight that by* Remark 2.8 *all the MEMD based models are invariant to additive shifts.*

- In the considered HSI supervised classification problems, the best of the proposed approaches have a performance equivalent to our baseline, which is the state-of-the-art for the considered problems (Table 3). However, we remark that the inclusion of morphological EMDs induces an invariant to additive intensity shifts in the classification model. To illustrate this fact, we have trained a classical model Figure 5 with and without a random data augmentation by using an additive shift as transformation. That is the usual approach to include some invariance in deep learning models. This gives an improvement in the invariance measure in Figure 7. We highlight that by Remark 2.8 all the MEMD based models are intrinsically invariant to additive shifts, which is illustrated in Figure 7.

5. Discussion. The paper investigated the formulation of EMD based on morphological operators and its integration into deep learning architectures. The training of the layers realizing the EMD process allows them to adapt the morphological models to the signals to be classified. The assessments have been done for supervised classification problem in 1D signals from hyperspectral images (*i.e.*, pixelwise spectra), but the proposed approaches are applicable to CNN architectures for nD images, without conceptual or algorithmic problem. 1D signals have been used for the only reason that the effects of the process on such signals are easier to interpret in a research perspective. Several variants of the morphological layers have been used. However, we think that for a better understanding of some of the elements of the approach: behaviour of the gradient of the layers during the optimization, contribution of

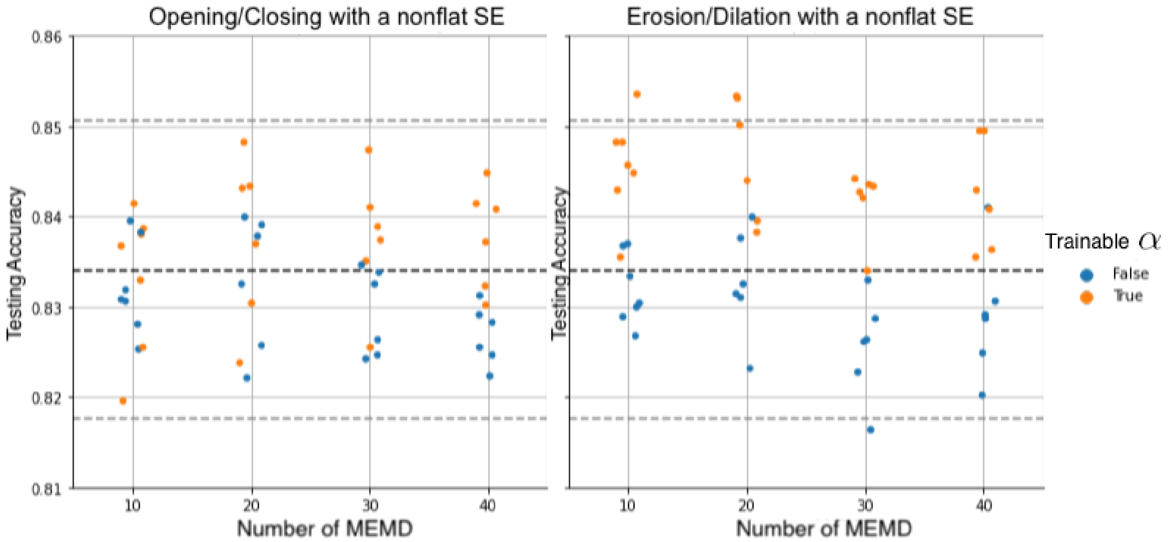


Figure 8. *Test accuracy for spatial-disjoint samples in Indian Pines Hyperspectral image. Envelopes produces by opening/closing MEMD (left figure) and erosion/dilation MEMD(right figure) by flat structuring function are used in the architecture presented inFigure 5. In both the number of MEMD varies from 10 to 40. Each point is the performance for the best model trained from different random initialization and same early stopping parameter (patience of ten epochs). The horizontal lines indicate the maximum/average/minimum performance of baseline architecture [39] on original data. Blue points correspond to $\alpha = .5$,* i.e., *when this parameter was not learned.*

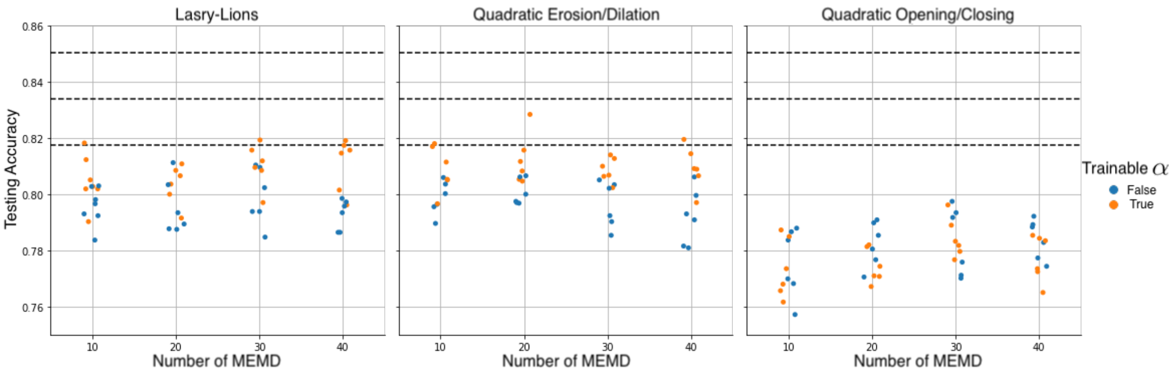


Figure 9. *Test accuracy for spatial-disjoint samples in Indian Pines Hyperspectral image. Envelopes produces by Lasry-Lions operator (left figure), erosion/dilation (central figure), and opening/closing (right figure) by quadratic structuring functions are used in the architecture presented inFigure 5. The number of MEMD varies from 10 to 40. Each point is the performance for the best model trained from different random initialization and same early stopping parameter (patience of 10 epochs). The horizontal lines indicate the maximum/average/minimum performance of baseline architecture [39] on original data. Blue points correspond to $\alpha = .5$, where this parameter was not learned.*

			Overall Val. Acc.		Overall Training Acc.	
Type	Operator	α	Best	$\mu \pm \sigma$	Best	$\mu \pm \sigma$
Baseline	—	—	85.035	83.929±0.654	93.443	91.413±1.696
NonFlat	(γ, φ)	.5	84.080	83.239 ± 0.512	97.012	95.495±1.184
		True	84.420	83.490 ± 0.656	97.223	96.012 ± 0.847
	(ε, δ)	.5	83.252	82.764 ± 0.576	97.451	95.226 ± 2.065
		True	**85.311**	84.052 ± 1.227	95.922	94.015 ± 2.717
	(ε, δ) SE(0) ≥ 0	.5	83.379	82.870 ± 0.261	96.889	95.621 ± 1.043
		True	85.247	83.821 ± 0.787	96.168	94.874 ± 1.120
Quadratic	(γ, φ)	.5	79.495	78.024 ± 0.754	96.080	93.580 ± 2.625
		True	80.959	77.971 ± 1.563	97.645	95.043 ± 1.565
	(ε, δ)	.5	81.363	79.798 ± 1.006	96.484	94.964± 1.111
		True	81.596	80.847 ± 0.537	97.223	95.066± 1.191
	Lasry-Lions	.5	81.384	79.909 ± 0.876	96.924	95.273 ± 1.336
		True	*82.424*	81.299 ± 0.983	96.941	95.674 ± 0.927

Table 1

Experiment on hyperspectral Indian Pines Disjoint *classification problem. Each experiment has been repeated ten times varying the initialization of base architecture. Twenty filters of MEMD in a single level of simplification. The training was performed without any data augmentation technique. The constraint* $SE(0) \geq 0$ *is used to assure the order relation among envelopes (See* Remark 1.2*)*

			Overall Val. Acc.		Overall Training Acc.	
Type	Operator	α	Best	$\mu \pm \sigma$	Best	$\mu \pm \sigma$
Baseline	—	—	85.468	83.396± 2.420	92.527	86.447 ± 8.960
NonFlat	(γ, φ)	.5	79.543	78.189 ± 0.726	95.715	92.219 ± 2.408
		True	82.353	79.293 ± 1.767	96.353	91.525 ± 4.335
	(ε, δ)	.5	84.261	82.681 ± 0.798	93.726	88.794 ± 4.998
		True	84.133	82.529 ± 1.131	93.879	89.735 ± 2.118
	(ε, δ), SE(0) ≥ 0	.5	83.908	81.740 ± 1.295	93.216	84.575 ± 7.334
		True	**85.483**	83.994 ± 1.238	94.389	89.617 ± 3.289
Quadratic	(γ, φ)	.5	74.516	70.821 ± 2.023	91.201	80.951 ± 6.432
		True	73.539	69.399 ± 2.339	93.828	87.360 ± 6.443
	(ε, δ)	.5	77.411	75.432 ± 1.193	95.052	86.470 ± 5.939
		True	*81.196*	77.923 ± 1.700	92.476	86.593 ± 6.585
	Lasry-Lions	.5	77.461	76.396 ± 0.614	97.067	90.826 ± 6.223
		True	80.971	78.501 ± 1.332	96.123	87.082 ± 8.221

Table 2

Experiment on hyperspectral Pavia University *for a* disjoint training sample. *Nine different classes. Each experiment has been repeated ten times varying the initialization of base architecture. Twenty filters of MEMD in a single level of simplification. The training was performed without any data augmentation technique. The constraint* $SE(0) \geq 0$ *is used to assure the order relation among envelopes (See* Remark 1.2*)*

Method	Indian Pines	Pavia University
Random Forest	65.79	69.64
Multinomial Logistic regression	83.81	72.23
Support Vector Machines	85.08	77.80
MLP	83.81	81.96
CNN1D	85.03	85.47
$\Phi^{\alpha}_{\varepsilon,\delta}$+ CNN1D	85.31	85.48

Table 3

Comparison (in terms of OA) between different HSI classification models trained on spatial-disjoint samples. The performance for first four models are included for comparison from [39].

the different parts of the signals to the optimization, effect of the initialization, etc. a deeper theoretical and empirical study is required. Additionally, we have illustrated the use of only one decomposition but the presented framework allows us to go further. In the future work, we are planning to use some interesting approaches to propose more adapted optimization schemes [8] for max-plus based layers, which reveals remarkable properties of network pruning by these operators [55]. Additionally, we will explore: a) the use for the MEMD of other structuring functions as Poweroids or Anisotropic Quadratic functions as proposed in [46], b) to consider the interest of MEMD to produce Scale Equivariant Neural Networks as in [45]. Finally, we highlight that the study of theoretical properties of morphological networks in the sense of their *expressiveness and universality* [56] is fundamental to have a full understanding of the limits of these types of layers when they are integrated in DL architectures.

Acknowledgments. This work has been supported by *Fondation Mathématique Jacques Hadamard* (FMJH) under the PGMO-IRSDI 2019 program. This work was granted access to the Jean Zay supercomputer under the allocation 2020-AD011012212.

REFERENCES

[1] J. Angulo, *Lipschitz Regularization of Images supported on Surfaces using Riemannian Morphological Operators*. Preprint, Nov. 2014, https://hal-mines-paristech.archives-ouvertes.fr/hal-01108130.

[2] J. Angulo and S. Velasco-Forero, *Riemannian mathematical morphology*, Pattern Recognition Letters, 47 (2014), pp. 93–101.

[3] J. Bedi and D. Toshniwal, *Empirical mode decomposition based deep learning for electricity demand forecasting*, Ieee Access, 6 (2018), pp. 49144–49156.

[4] Y. Bengio, Y. LeCun, and H. Hinton, *Deep learning*, Nature, 521 (2015), pp. 436–444.

[5] J. Bosworth and S. T. Acton, *The morphological lomo filter for multiscale image processing*, in Proceedings 1999 International Conference on Image Processing (Cat. 99CH36348), vol. 4, IEEE, 1999, pp. 157–161.

[6] Y.-L. Boureau, J. Ponce, and Y. LeCun, *A theoretical analysis of feature pooling in visual recognition*, in ICML 2010, 2010, pp. 111–118.

[7] M. Carlsson, *On convex envelopes and regularization of non-convex functionals without moving global minima*, Journal of Optimization Theory and Applications, 183 (2019), pp. 66–84.

[8] V. Charisopoulos and P. Maragos, *Morphological perceptrons: Geometry and training algorithms*, in Mathematical Morphology and Its Applications to Signal and Image Processing, Springer International Publishing, 2017, pp. 3–15.

[9] J. Chen, X. Wang, and C. Planiden, *A proximal average for prox-bounded functions*, SIAM Journal

on Optimization, 30 (2020), pp. 1366–1390.

[10] E. Deléchelle, J. Lemoine, and O. Niang, *Empirical mode decomposition: an analytical approach for sifting process*, IEEE Signal Processing Letters, 12 (2005), pp. 764–767.

[11] E. H. S. Diop and R. Alexandre, *Analysis of intrinsic mode functions based on curvature motion-like pdes*, in Curves and Surfaces, J.-D. Boissonnat, A. Cohen, O. Gibaru, C. Gout, T. Lyche, M.-L. Mazure, and L. L. Schumaker, eds., Cham, 2015, Springer International Publishing, pp. 202–209.

[12] E. H. S. Diop, R. Alexandre, and L. Moisan, *Intrinsic nonlinear multiscale image decomposition: A 2D empirical mode decomposition-like tool*, Computer Vision and Image Understanding, 116 (2012), pp. 102–119. Virtual Representations and Modeling of Large-scale Environments (VRML).

[13] E.-H. S. Diop, R. Alexandre, and V. Perrier, *A PDE model for 2d intrinsic mode functions*, IEEE ICIP, (2009).

[14] G. Franchi, A. Fehri, and A. Yao, *Deep morphological networks*, Pattern Recognition, 102 (2020), p. 107246.

[15] J. Gilles, *Empirical wavelet transform*, IEEE Transactions on Signal Processing, 61 (2013), pp. 3999–4010.

[16] J. Goutsias and H. Heijmans, *Mathematical Morphology*, IOS Press, 2000.

[17] B. Hanin and D. Rolnick, *How to start training: The effect of initialization and architecture*, in Advances in Neural Information Processing Systems, 2018, pp. 571–581.

[18] Z. He, J. Li, L. Liu, and Y. Shen, *Three-dimensional empirical mode decomposition (TEMD): A fast approach motivated by separable filters*, Signal Processing, 131 (2017), pp. 307–319.

[19] H. J. Heijmans and P. Maragos, *Lattice calculus of the morphological slope transform*, Signal Processing, 59 (1997), pp. 17–42.

[20] H. J. Heijmans and C. Ronse, *The algebraic basis of mathematical morphology i. dilations and erosions*, Computer Vision, Graphics, and Image Processing, 50 (1990), pp. 245–295.

[21] H. J. Heijmans and R. van den Boomgaard, *Algebraic framework for linear and morphological scale-spaces*, Journal of Visual Communication and Image Representation, 13 (2002), pp. 269–301.

[22] T. Y. Hou and Z. Shi, *Adaptive data analysis via sparse time-frequency representation*, Advances in Adaptive Data Analysis, 3 (2011), pp. 1–28.

[23] N. Huang, S. Zheng, S. Long, M. Wu, H. Shih, Q. Zheng, N.-C. Yen, C. Tung, and H. Liu, *The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis*, The Royal Society, 454 (1998), pp. 903–995.

[24] P. T. Jackway, *Morphological scale-spaces*, in Advances in Imaging and Electron Physics, P. W. Hawkes, ed., vol. 99, Elsevier, 1997, pp. 1–64.

[25] P. T. Jackway and M. Deriche, *Scale-space properties of the multiscale morphological dilation-erosion*, IEEE Transactions on Pattern Analysis and Machine Intelligence, 18 (1996), pp. 38–51.

[26] X.-B. Jin, N.-X. Yang, X.-Y. Wang, Y.-T. Bai, T.-L. Su, and J.-L. Kong, *Deep hybrid model based on emd with classification by frequency characteristics for long-term air quality prediction*, Mathematics, 8 (2020), p. 214.

[27] P.-L. L. J.M. Lasry, *A remark on regularization in Hilbert spaces*, Israel Journal of Mathematics, 55 (1986), pp. 257–266.

[28] A. Kirszenberg, G. Tochon, É. Puybareau, and J. Angulo, *Going beyond p-convolutions to learn grayscale morphological operators*, in Discrete Geometry and Mathematical Morphology, J. Lindblad, F. Malmberg, and N. Sladoje, eds., Cham, 2021, Springer International Publishing, pp. 470–482.

[29] D. Looney and D. P. Mandic, *A machine learning enhanced empirical mode decomposition*, in IEEE ICASSP, IEEE, 2008, pp. 1897–1900.

[30] P. Maragos, *Representations for morphological image operators and analogies with linear operators*, Advances in imaging and electron physics, 177 (2013), pp. 45–187.

[31] P. Maragos, J. F. Kaiser, and T. F. Quatieri, *Energy separation in signal modulations with application to speech analysis*, IEEE Transactions on Signal Processing, 41 (1993), pp. 3024–3051.

[32] J. Masci, J. Angulo, and J. Schmidhuber, *A learning framework for morphological operators using counter–harmonic mean*, in Mathematical Morphology and Its Applications to Signal and Image Processing, Springer, 2013, pp. 329–340.

[33] R. Mondal, M. S. Dey, and B. Chanda, *Image restoration by learning morphological opening-closing network*, Mathematical Morphology - Theory and Applications, 4 (01 Jan. 2020), pp. 87 – 107.

[34] J.-J. Moreau, *Proximité et dualité dans un espace hilbertien*, Bulletin de la Société mathématique de France, 93 (1965), pp. 273–299.
[35] J. J. Moreau, *Inf-convolution, sous-additivité, convexité des fonctions numériques*, Journal de Mathématiques Pures et Appliquées, (1970).
[36] M. Nakashizuka, *Image regularization with higher-order morphological gradients*, in 2015 23rd European Signal Processing Conference (EUSIPCO), IEEE, 2015, pp. 1820–1824.
[37] K. Nogueira, J. Chanussot, M. D. Mura, W. R. Schwartz, and J. A. dos Santos, *An introduction to deep morphological networks*, 2019, https://arxiv.org/abs/1906.01751.
[38] S. J. Nowlan and G. E. Hinton, *Simplifying neural networks by soft weight-sharing*, Neural computation, 4 (1992), pp. 473–493.
[39] M. Paoletti, J. Haut, J. Plaza, and A. Plaza, *Deep learning classifiers for hyperspectral imaging: A review*, ISPRS Journal of Photogrammetry and Remote Sensing, 158 (2019), pp. 279–317.
[40] N. Parikh and S. Boyd, *Proximal algorithms*, Foundations and Trends in optimization, 1 (2014), pp. 127–239.
[41] L. F. Pessoa and P. Maragos, *MRL-filters: A general class of nonlinear systems and their optimal design for image processing*, IEEE Transactions on Image Processing, 7 (1998), pp. 966–978.
[42] X. Qiu, Y. Ren, P. N. Suganthan, and G. A. Amaratunga, *Empirical mode decomposition based ensemble deep learning for load demand time series forecasting*, Applied Soft Computing, 54 (2017), pp. 246–255.
[43] R. T. Rockafellar and R. J.-B. Wets, *Variational analysis*, vol. 317, Springer Science & Business Media, 2009.
[44] R. Rojas, *The backpropagation algorithm*, in Neural networks, Springer, 1996, pp. 149–182.
[45] M. Sangalli, S. Blusseau, S. Velasco-Forero, and J. Angulo, *Scale equivariant neural networks with morphological scale-spaces*, in Discrete Geometry and Mathematical Morphology, J. Lindblad, F. Malmberg, and N. Sladoje, eds., Cham, 2021, Springer International Publishing, pp. 483–495.
[46] M. Schmidt and J. Weickert, *Morphological counterparts of linear shift-invariant scale-spaces*, Journal of Mathematical Imaging and Vision, 56 (2016), pp. 352–366.
[47] J. Serra, *Image Analysis and Mathematical Morphology*, Academic Press, Inc., Orlando, FL, USA, 1983.
[48] R. C. Sharpley and V. Vatchev, *Analysis of the intrinsic mode functions*, Constructive Approximation, 24 (2006), pp. 17–47.
[49] S. Sinclair and G. Pegram, *Empirical mode decomposition in 2-d space and time: a tool for space-time rainfall analysis and nowcasting*, Hydrology and Earth System Sciences, 9 (2005), pp. 127–137.
[50] A. Stallone, A. Cicone, and M. Materassi, *New insights and best practices for the successful use of empirical mode decomposition, iterative filtering and derived algorithms*, Scientific reports, 10 (2020), pp. 1–15.
[51] R. Van Den Boomgaard, L. Dorst, S. Makram-Ebeid, and J. Schavemaker, *Quadratic structuring functions in mathematical morphology*, in Mathematical morphology and its applications to image and signal processing, Springer, 1996, pp. 147–154.
[52] R. Van den Boomgaard and A. Smeulders, *The morphological structure of images: The differential equations of morphological scale-space*, IEEE Transactions on Pattern Analysis and Machine Intelligence, 16 (1994), pp. 1101–1113.
[53] L. J. van Vliet, I. T. Young, and G. L. Beckers, *A nonlinear Laplace operator as edge detector in noisy images*, Computer Vision, Graphics, and Image Processing, 45 (1989), pp. 167 – 195.
[54] S. Velasco-Forero and J. Angulo, *On nonlocal mathematical morphology*, in International Symposium on Mathematical Morphology and Its Applications to Signal and Image Processing, Springer, 2013, pp. 219–230.
[55] Y. Zhang, S. Blusseau, S. Velasco-Forero, I. Bloch, and J. Angulo, *Max-plus operators applied to filter selection and model pruning in neural networks*, in Mathematical Morphology and Its Applications to Signal and Image Processing, Springer International Publishing, 2019, pp. 310–322.
[56] D.-X. Zhou, *Universality of deep convolutional neural networks*, Applied and computational harmonic analysis, 48 (2020), pp. 787–794.