



**HAL**  
open science

# Learnable Empirical Mode Decomposition based on Mathematical Morphology

Santiago Velasco-Forero, Romain Pagès, Jesus Angulo

► **To cite this version:**

Santiago Velasco-Forero, Romain Pagès, Jesus Angulo. Learnable Empirical Mode Decomposition based on Mathematical Morphology. 2021. hal-03221652v1

**HAL Id: hal-03221652**

**<https://hal.science/hal-03221652v1>**

Preprint submitted on 9 May 2021 (v1), last revised 26 Aug 2021 (v3)

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Learnable Empirical Mode Decomposition based on Mathematical Morphology \*

Santiago Velasco-Forero<sup>†</sup>, R. Pagès<sup>‡</sup>, and Jesus Angulo<sup>§</sup>

**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) the use of a convex sum of envelopes instead of the classical 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) [24], which is a simple 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) the use of a convex sum of

\*

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

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

<sup>‡</sup>École Centrale de Lyon, France ([romain.pages@ecl18.ec-lyon.fr](mailto:romain.pages@ecl18.ec-lyon.fr)) .

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

36 envelopes instead of the classical mean point used in classical EMD. All the parameters of our  
 37 proposition can be learnt using backpropagation and gradient descent techniques and therefore  
 38 the associated morphological EMD can be integrated into standard DCNN representations for  
 39 end-to-end learning. The integration of morphological operators into DCNN pipelines is an  
 40 active research area. First attempts were based on approximation of dilation and erosion using  
 41 standard convolution [33]. More recently, straightforward approaches of dilation and erosion  
 42 optimization have been explored [14, 34, 38]. However, plugging morphological operators into  
 43 standard networks is far from being trivial from the optimization based on backpropagation  
 44 of gradients through all layers by the chain rule. Max-plus operators are indeed differentiable  
 45 only on a local and specific domain. Here we focus on standard gradient descent strategies  
 46 and we provide a better understanding of how the gradient of morphological operators, in  
 47 particular those associated to parametric structuring functions, is defined. Additionally, we  
 48 show that our morphological EMD induces the invariance to additive shift to standard DCNN.  
 49 To the best of our knowledge, these technical aspects have not been previously discussed in  
 50 the field of morphological deep neural networks.

51 **1.1. Related work.** In what follows we review the state-of-the-art that is most relevant  
 52 for the proposed morphological EMD.

53 **1.1.1. Empirical Mode Decomposition.** EMD is an algorithm introduced by Huang et  
 54 al. [24] for analysing linear and non-stationary time series. It is a way to decompose a signal  
 55 to obtain instantaneous frequency data. In this original version on the EMD is an iterative  
 56 process which decomposes real signals  $f$  into simpler signals (modes),  $f(x) = \sum_{i=1}^M \Phi_j(x)$ ,  
 57 where each *mono-component* signal  $\Phi$  should be written in the form  $\Phi(x) = r(x) \cos(\theta x)$ , where  
 58 the amplitude and phase are both physically and mathematically meaningful [49]. Unlike some  
 59 other common transforms like the Fourier transform for example, the EMD was built as an  
 60 algorithm and lacks theoretical background then. The problem of EMD to represent a signal as  
 61 a sum of amplitude modulation (AM) and frequency modulation (FM) components at multiple  
 62 scales was first proposed in [32] where the problem of finding the AM-FM components and  
 63 their envelopes was solved using multiscale Gabor filters and nonlinear Teager-Kaiser Energy  
 64 Operators via an Energy Separation Algorithm (ESA). In the case of original EMD, there is  
 65 no parametric family of filters used to estimate the envelopes.

66 From an algorithmic point of view, the EMD is obtained following the iterative process  
 67 [24]:

- 68 1. Find all the local extrema of the function  $f$ .
- 69 2. Interpolate all the local maxima together to get the function  $\hat{f}$  (upper envelope), and  
 70 all the local minima together to get the function  $\check{f}$  (lower envelope)
- 71 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.*,

$$r(x) = f(x) - IMF(x)$$

71 until a selected tolerance criterion is respected.

72 Thus, the original signal is decomposed as:

$$73 \quad (1.1) \quad f(x) = \sum_{k=1}^n IMF_k(x) + r(x)$$

74 where  $IMF_k$  is the  $k$ -th intrinsic mode function and  $r$  is the last residual. The EMD can be  
 75 efficiently applied to 1D-signals. However the selection of interpolation method for the second  
 76 step gives a wide variety of possibilities, from the original formulation using cubic splines [24],  
 77 passing by sparse filtering [23], filtering from wavelet based decomposition [15] and partial  
 78 differential equation based formulations [10].

79 The EMD method can be justified only under certain very restrictive assumptions that  
 80 are seldom satisfied by practical data. The EMD method is also known to be very sensitive  
 81 to noisy data. Recently, a compendium of practical advice for EMD in real life examples has  
 82 been presented in [51]. Some works extend EMD to 2D [12, 50], [11] and 3D images [19].  
 83 However, the main limitations of EMD for both 2D and 3D are both the choice of maxima  
 84 and minima detector, and the choice of the interpolation algorithm.

85 An alternative characterisation of the EMD computation was introduced by Diop *et al.* in  
 86 [12, 13] according to the definition of *local mean*, *i.e.*, the sifting process is fully determined  
 87 by the sequence  $(h_n)_{n \in \mathbb{N}}$  defined by :

$$88 \quad (1.2) \quad \begin{cases} h_{n+1} = h_n - \Phi(h_n) = (\text{Id} - \Phi) h_n \\ h_0 = f \end{cases}$$

89 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  
 90 (resp. minima) of  $h_n$ .

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

93 **1.1.2. Dilation/Erosion.** We study here functions  $f : E \rightarrow \overline{\mathbb{R}}$ , where  $\overline{\mathbb{R}}$  it allowed to be  
 94 *extended-real-valued*, *i.e.*, to take values in  $\overline{\mathbb{R}} = [-\infty, \infty]$ . Accordingly, the set of all such  
 95 functions is denoted by  $\mathcal{F}(E, \overline{\mathbb{R}})$ . We will use the two basic morphological operators *dilation*  
 96 and *erosion*, which correspond respectively to the convolution in the  $(\max, +)$  algebra and its  
 97 dual.

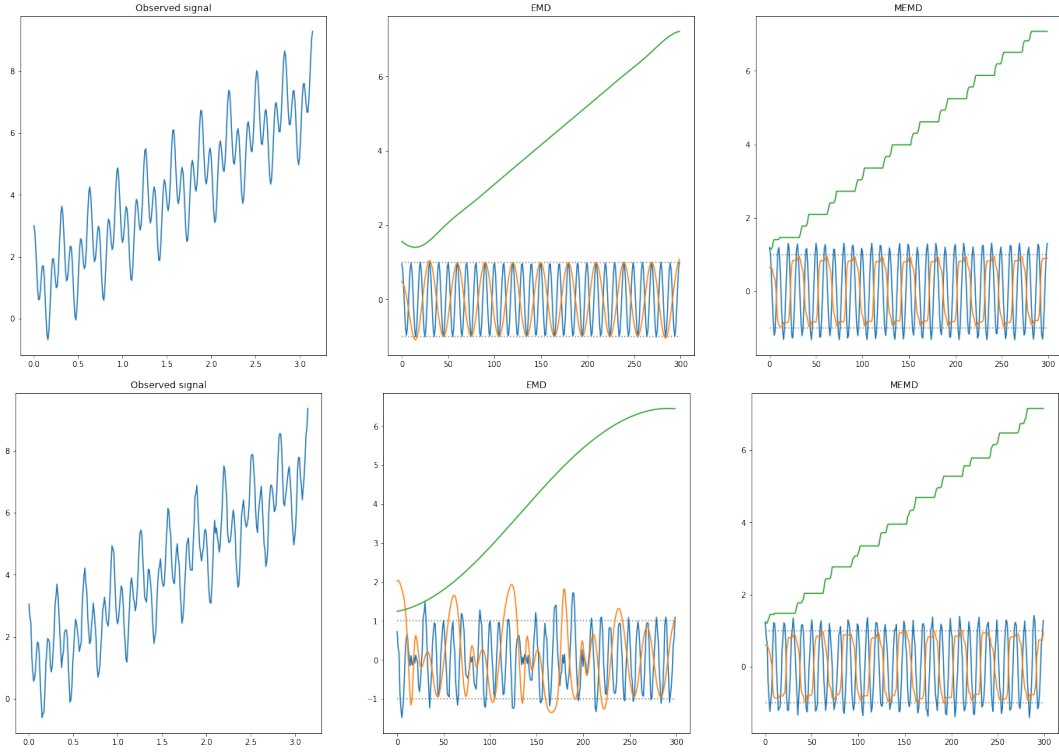
98 **Definition 1.1.** *In mathematical morphology [48], the dilation (sup-convolution)  $\delta_{SE}(f)$  of*  
 99  *$f$  is given by:*

$$100 \quad (1.3) \quad \delta_{SE}(f)(x) := \sup_y \{f(y) + SE(x - y)\} = \sup_w \{f(x - w) + SE(w)\}$$

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

105 The erosion [48]  $\varepsilon_{SE}(f)$ , known as inf-convolution in convex analysis [36], is the adjoint  
 106 operator to the dilation (1.3), and it is defined as

$$107 \quad (1.4) \quad \varepsilon_{SE}(f)(x) := -\delta_{SE}(-f)(x) = \inf_y \{f(y) - SE(y - x)\} = \inf_w \{f(w - x) - SE(w)\}$$



**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.

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

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

- 110 1. The operators in (1.3) and (1.4) are translation invariant.
- 111 2. (1.3) and (1.4) correspond to one another through the duality relation  $\delta_{SE}(f)(x) \leq$   
112  $g(x) \iff f(x) \leq \varepsilon_{SE}(g)(x)$ , called *adjunction*[16].
- 113 3. An operator  $\xi$  is called *extensive* (resp. *antiextensive*) if  $\xi(f)(x) \geq f(x)$  (resp.  
114  $\xi(f)(x) \leq f(x)$ )  $\forall x$ . The dilation (1.3) (resp. erosion (1.4)) is extensive (resp. antiex-  
115 tensive) if and only if  $SE(0) \geq 0$ , i.e., the structuring function evaluated at the origin  
116 is non-negative.
- 117 4. An operator  $\xi$  is called *increasing* if  $f(x) \geq g(x) \Rightarrow \xi(f)(x) \geq \xi(g)(x) \forall x$ . The dilation  
118 (1.3) and erosion (1.4) are increasing for all SE.
- 119 5.  $\varepsilon_{SE}(f)(x) \leq f(x) \leq \delta_{SE}(f)(x)$  if and only if  $SE(0) \geq 0$ .
- 120 6.  $\delta_{SE}$  (resp.  $\varepsilon_{SE}$ ) does not introduce any local maxima (resp. local minima) if  $SE \leq 0$   
121 and  $SE(0) = 0$ . In this case, we say that SE is *centered*.

122 *Proof.* (1) and (2) are classical results from [48]. As explained in [21] and [31], the *ad-*  
123 *junction* is related to a well-known concept in group and lattice theory, the *Galois connection*.  
124 (3)  $\forall f, \delta_{SE}(x) \geq f(x) \Rightarrow \forall f, \sup (f(x-w) + SE(w) - f(x))(x) \geq 0 \Rightarrow SE(0) \geq 0$  Now,

125  $\sup f(x - w) + SE(w) \geq f(x) + SE(0)$ , if  $SE(0) \geq 0 \Rightarrow \sup f(x - w) + SE(w) \geq f(x)$ . (4) and  
 126 (6) are easy to prove directly from the definition of the operators. It has been also proved  
 127 in the original paper of inf-convolution (Proposition 6.d) in [36]. From (3) and (4) is easy to  
 128 prove (5). ■

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

$$132 \quad (1.5) \quad B(y) = \begin{cases} 0 & \text{if } y \in B \\ -\infty & \text{if } y \in B^c \end{cases}$$

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

$$136 \quad (1.6) \quad B_\lambda(w) = \begin{cases} 0 & \text{if } \|w\| \leq \lambda \\ -\infty & \text{if } \|w\| > \lambda \end{cases}$$

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

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

$$140 \quad (1.7) \quad \Phi_{\varepsilon, \delta, B_\lambda}(f)(x) := \frac{\delta_{B_\lambda}(f)(x) + \varepsilon_{B_\lambda}(f)(x)}{2}$$

141 The operator (1.7) was proposed to generate a EMD based on solving a morphological PDE  
 142 [13]. As a manner of example, classical and morphological MEMD are shown for a mono-  
 143 component signal in the first row of Figure 1. In the second row of Figure 1, we illustrated  
 144 how the addition of noisy perturbed more the results of classical EMD than the proposed  
 145 morphological one.

146 **Remark 1.4.** Note that using (1.7) twice, the first residual (1.2) is  $2(f - \Phi_\lambda(f)) = (f -$   
 147  $\delta_{B_\lambda}(f)) + (f - \varepsilon_{B_\lambda}(f)) = 2f - \delta_{B_\lambda}(f) - \varepsilon_{B_\lambda}(f)$ . This expression, up to a minus sign,  
 148 corresponds just to the so-called *morphological Laplace operator* [54], and therefore provides  
 149 an interpretation of the EMD as an iterated second-order derivative decomposition of the  
 150 function  $f$ .

151 **1.2. Our proposal.** The main motivation of this paper is to define EMD learnable in the  
 152 sense of neural networks approaches. Note that last property in Remark 1.2 together with  
 153 the extensivity/antiextensivity (*i.e.*, upper/lower envelopes) imply that the pair of operators  
 154  $(\varepsilon_{SE}, \delta_{SE})$  are candidate functions for  $(\hat{h}, \check{h})$  in (1.2). Accordingly, we proposed a simple  
 155 generalization by considering non-flat structuring functions.

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

$$157 \quad (1.8) \quad \Phi_{\varepsilon, \delta, SE}(f) = \frac{\delta_{SE}(f)(x) + \varepsilon_{SE}(f)(x)}{2}$$

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

160 **1.3. Contributions of the paper.** In what follows we study,

- 161 • A formulation of EMD based on pairs of morphological operators in a general case.
- 162 • The proposition of a parametric morphological empirical mode whose sifting process
- 163 is invariant to additive intensity shifts.
- 164 • A approach to learn the structuring functions of a morphological operator in a deep
- 165 learning framework.
- 166 • A convex sum of envelopes instead of mean point to learn morphological EMD.
- 167 • A number of numerical experiments for hyperspectral signal classification to illustrate
- 168 the relevance of our proposal.

169 **1.4. Organization of the paper.** The rest of the paper is organised as follows. In [sec-](#)  
 170 [tion 2](#), we review the general definition of Empirical Mode Decomposition approach to decom-  
 171 pose signals and we introduce how morphological extensive/antiextensive filters are naturally  
 172 adapted to implement a Morphological Empirical Mode computation. We consider different  
 173 possibilities in the choice of structuring functions and the pair of lower and upper envelopes.  
 174 Additionally, an  $\alpha$ -MEM is proposed as a generalization of mean of envelopes. [Section 3](#) is  
 175 devoted to the implementation of morphological EMD operators as layers in a neural network  
 176 pipelines. [Section 4](#) presents the experimental results of hyperspectral image classification  
 177 using DCNNs which integrates morphological EMD layers. Conclusions and perspectives are  
 178 discussed in [section 5](#).

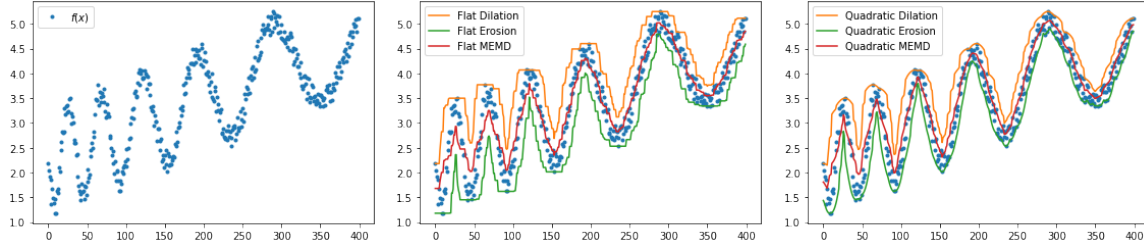
179 **2. Morphological Empirical Mode and its variants.** In this section, three kind of gen-  
 180 eralization will be explored: a) different type of structuring functions, b) different pairs of  
 181 functions to compute the lower and upper envelopes, and c) a convex sum of lower and upper  
 182 envelopes.

183 **2.1. Varying the structuring function.** In this subsection, firstly we will study a paramet-  
 184 ric family of symmetric quadratic shape structuring functions. Secondly, similarly to classical  
 185 CNNs, the structuring function plays a similar role to the kernel. Accordingly a structuring  
 186 function without any parametric constraint is also considered.

187 **2.1.1. Quadratic MEM.** From the theory of morphological scale-spaces, the most useful  
 188 nonflat structuring functions are those which depend on a scale parameter [[22](#), [47](#)]. The only  
 189 separable and rotationally invariant structuring functions is the called *quadratic structuring*  
 190 *function*[[52](#)]:

$$191 \quad (2.1) \quad q_\lambda(w) = -\frac{\|w\|^2}{2\lambda},$$

192 such that the corresponding dilation and erosion are equal to the Lax–Oleinik operators or  
 193 viscosity solutions of the standard Hamilton–Jacobi PDE, also known as morphological PDE:  
 194  $u_t(t, x) \mp \|u_x(t, x)\|^2 = 0$ ,  $(t, x) \in (0, +\infty) \times E$ ;  $u(0, x) = f(x)$ ,  $x \in E$ . It plays also a canonical  
 195 role in the definition of dilation and erosion on Riemannian manifolds [[2](#)] and their behaviour  
 196 with respect to the maxima/minima is well understood [[26](#)]. The morphological PDE was  
 197 proposed and analyzed using 2D boundary propagation in [[53](#)] and further analyzed using the  
 198 morphological slope transform in [[20](#)].



**Figure 2.** a) Original signal, b) Flat dilation/erosion based Morphological Empirical Mode (1.7) with a disk of  $\lambda = 5$ , c) Quadratic dilation/erosion based Morphological Empirical Mode (2.3) with  $\lambda = 3$ .

199 *Remark 2.1.* The erosion by a quadratic structuring function with parameter  $\lambda$  is defined  
 200 by  
 (2.2)

$$201 \quad \varepsilon_{q_\lambda}(f)(x) := \inf_y \{f(y) - q_\lambda(y - x)\} = \inf_w \{f(w - x) - q_\lambda(w)\} = \inf_w \left\{ f(w - x) + \frac{\|w\|^2}{2\lambda} \right\}$$

202 The erosion of a function  $f$  by a quadratic structuring function with parameter  $\lambda$  is  
 203 known as the *Moreau envelope* or *Moreau-Yosida approximation* [36, 44, 41], which offers many  
 204 benefits specially for optimization purposes [35]. Additionally, (2.2) induces an additive scale-  
 205 space [20, 25], i.e.,  $\varepsilon_{q_{\lambda_1}}(\varepsilon_{q_{\lambda_2}}(f)) = \varepsilon_{q_{\lambda_1 + \lambda_2}}(f)$ .

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

$$208 \quad (2.3) \quad \Phi_{\varepsilon, \delta, q_\lambda}(f) = \frac{\varepsilon_{q_\lambda}(f) + \delta_{q_\lambda}(f)}{2}.$$

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

210 **2.1.2. Nonflat Morphological MEM.** The most general case of *nonflat structuring func-*  
 211 *tion* involves different additive weights  $W_y(x)$  at each position  $x$  of the local neighborhood  $B$   
 212 centered at pixel  $y$ , i.e., a nonflat structuring function  $SE_W$  of support shape  $B$  at  $y$  is defined  
 213 as

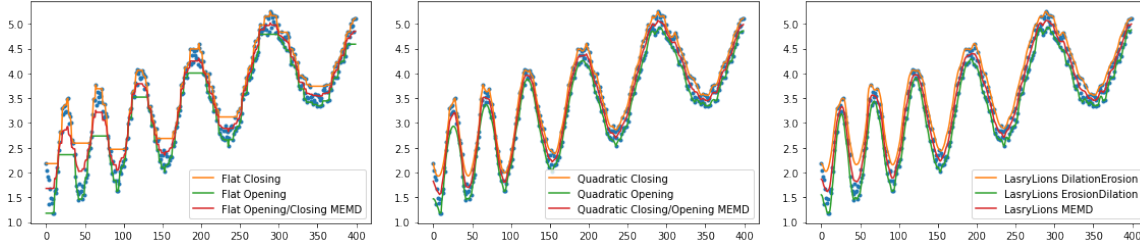
$$214 \quad (2.4) \quad SE_{W_y}(x) = \begin{cases} W_y(x) & \text{if } x \in B(y) \\ -\infty & \text{otherwise} \end{cases}$$

215 The case (2.4) includes flat, nonflat, either local or nonlocal structuring functions [55]. In the  
 216 translation invariant case, the weighting function  $W_y(x)$  is equal for all  $y \in E$ .

217 **2.2. Varying the Envelope.** We have explored above several possible structuring functions  
 218 that produce multiple pairs of  $(\varepsilon_{SE}, \delta_{SE})$  as basic ingredient for the Morphological Empirical  
 219 Mode (1.8). At this point, we can consider the use of the composition of erosion and dilation  
 220 to obtain other upper/lower envelopes, typically of the form  $(\delta_{SE}(\varepsilon_{SE}), \varepsilon_{SE}(\delta_{SE}))$ .

221 **2.2.1. Opening/Closing MEM.** The theory of morphological filtering is based on the  
 222 opening  $\gamma_{SE}(f)(x)$  and closing  $\varphi_{SE}(f)(x)$  operators, obtained respectively by the composi-  
 223 tion product of erosion-dilation and dilation-erosion, i.e.,  $\gamma_{SE}(f)(x) = \delta_{SE}(\varepsilon_{SE}(f))(x)$  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  $c = .9$

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

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

$$233 \quad (2.5) \quad \Phi_{\gamma, \varphi, SE}(f) = \frac{\gamma_{SE}(f) + \varphi_{SE}(f)}{2}.$$

234 For the case of flat disks  $B_\lambda$ , the operator (2.5) was called morphological locally monotonic  
 235 (LOMO) filter in [5]. A signal is monotonic over an interval if it is either non-increasing or  
 236 non-decreasing over that interval. A 1-D signal is *locally monotonic* of degree  $n$  (LOMO- $n$ )  
 237 if and only if the signal is monotonic within every internal of length  $n$ . In the general case, a  
 238 LOMO filter of  $f$  is defined as the fixed point of iterating  $\Phi_{\gamma, \varphi, B_\lambda}(f)$ , which is simultaneously  
 239 idempotent to both the opening and closing by a flat disk as structuring function. Two  
 240 examples of (2.5) for both flat and quadratic structuring function for the 1D signal with noise  
 241 presented in Figure 2 is shown in Figure 3.

242 **2.2.2. Lasry–Lions MEM.** Besides their feature extraction properties, morphological di-  
 243 lation and erosion using quadratic structuring functions are a powerful tool for Lipschitz  
 244 regularization. For the nonconvex case, the Lasry–Lions double envelope  $h_{\mu, \lambda}$  is defined as  
 245 the composition of two different Moreau envelopes, or using the morphological vocabulary,  
 246 the composition of an erosion followed by a dilation with a quadratic structuring function.  
 247 For all  $0 < c < 1$  and  $0 < \lambda$ , the so-called Lasry–Lions regularizers [28] are defined as

$$248 \quad \gamma_\lambda^c(f)(x) := \delta_{q_{c\lambda}}(\varepsilon_{q_\lambda}(f))(x),$$

$$249 \quad \varphi_\lambda^c(f)(x) := \varepsilon_{q_{c\lambda}}(\delta_{q_\lambda}(f))(x),$$

250 such that if  $f$  is bounded, the functions  $\gamma_\lambda^c$  and  $\varphi_\lambda^c$  are bounded and one has the ordering  
 251 properties for the following envelopes:

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

$$253 \quad \gamma_{\lambda_1}^c(f)(x) \leq \gamma_{\lambda_2}^c(f)(x) \leq f \leq \varphi_{\lambda_2}^c(f)(x) \leq \varphi_{\lambda_1}^c(f)(x);$$

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

$$255 \quad \gamma_{\lambda}^{c_2}(f)(x) \leq \gamma_{\lambda}^{c_1}(f)(x) \leq f \leq \varphi_{\lambda}^{c_1}(f)(x) \leq \varphi_{\lambda}^{c_2}(f)(x).$$

256 For any bounded function  $f$ , Lasry–Lions regularizers provided a function with a Lipschitz  
257 continuous gradient, *i.e.*,

$$258 \quad |\nabla \gamma_{\lambda}^c(f)(x) - \nabla \gamma_{\lambda}^c(f)(y)| \leq M_{\lambda,c} \|x - y\|, \quad |\nabla \varphi_{\lambda}^c(f)(x) - \nabla \varphi_{\lambda}^c(f)(y)| \leq M_{\lambda,c} \|x - y\|.$$

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

$$261 \quad \text{Lip}(\gamma_{\lambda}^c(f)) \leq \text{Lip}(f) \quad \text{and} \quad \text{Lip}(\varphi_{\lambda}^c(f)) \leq \text{Lip}(f),$$

262 with

$$263 \quad \text{Lip}(g) = \sup \left\{ \frac{|g(x) - g(y)|}{\|x - y\|}; \quad x, y \in \mathbb{R}^n, \quad x \neq y \right\}.$$

264 For more details on the properties of Lasry–Lions regularizers in the context of mathe-  
265 matical morphology, see [1].

266 *Remark 2.4.* The following statements are interesting about the composition of quadratic  
267 morphological operators [44, 9]. Let  $0 < \mu < \lambda$

- 268 1.  $\varepsilon_{q_{\lambda}}(\gamma_{q_{\lambda}}(f)) = \varepsilon_{q_{\lambda}}(f)$
- 269 2.  $\gamma_{q_{\mu}}(\varepsilon_{q_{\lambda-\mu}}(f)) = \varepsilon_{q_{\lambda-\mu}}(\gamma_{q_{\lambda}}(f))$
- 270 3.  $\gamma_{q_{\lambda-c\lambda}} \varphi_{\lambda}^c(f) = \varphi_{\lambda}^c(f)$

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

$$273 \quad (2.6) \quad \Phi_{\gamma, \varphi, c, \lambda}(f) := \frac{\gamma_{\lambda}^c(f) + \varphi_{\lambda}^c(f)}{2}.$$

274 An example of (2.6) for the 1D signal with noise presented in Figure 2 is shown in Fig-  
275 ure 3(c).

276 **2.3. Parametric family of morphological empirical mode operator.** The choices of the  
277 structuring function and the class of lower and upper envelopes give extra possibilities for the  
278 formulation of an EMD approach. Besides, a third degree of freedom is considered now by  
279 including a parameter to weight the contribution of the two envelopes. We have been inspired  
280 by the recent work on proximal average [9] to propose a convex generalization of MEMs.

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

$$283 \quad (2.7) \quad \Phi_{\hat{h}, \check{h}}^{\alpha}(f) = \alpha \hat{h}(f) + (1 - \alpha) \check{h}(f).$$

284 **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  
 285 the abstract group  $g \in G$ . We say a function  $\phi$  is invariant to  $g$  if for all transformations  $T_g$ ,  
 286 and for all  $f \in \mathcal{F}(E, \overline{\mathbb{R}})$  one has

$$287 \quad (2.8) \quad \phi(T_g(f)) = \phi(f)$$

288 This says that the feature extracted by  $\phi$  does not change as the transformation is applied to  
 289 the input.

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

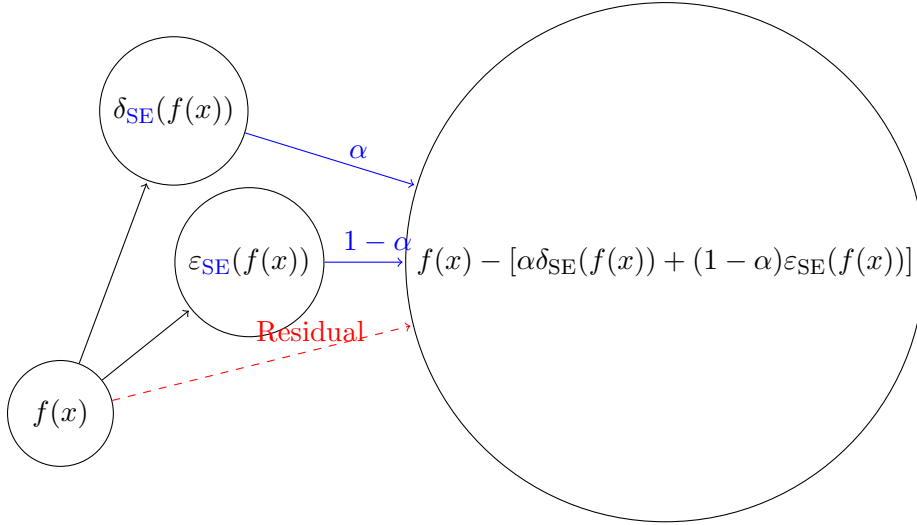
291 **Remark 2.8.** For any SE,  $\forall 0 \leq \alpha \leq 1$ , and all the pairs  $(\check{h}, \hat{h})$  previously considered, the  
 292 operator (2.7) is increasing, invariant to translation, and the sifting process  $f - \Phi_{\check{h}, \hat{h}}^\alpha(f)$  is  
 293 invariant to additive intensity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

$$294 \quad (f(x) + c) - \Phi_{\check{h}, \hat{h}}^\alpha(f(x) + c) = f(x) - \Phi_{\check{h}, \hat{h}}^\alpha(f(x)).$$

295 **3. Learnable Morphological Empirical Mode Decomposition.** One of the main advan-  
 296 tages of EMD is that it can be considered as a parameter-free decomposition [51] and, for  
 297 this reason, the inclusion of the structuring function and the parameter  $\alpha$  can be seen as  
 298 inconvenient. However, in the following, we consider EMD, in the context of learning from  
 299 data [30], where one would be interested in using EMD decomposition as a preprocessing of  
 300 an input signal before using a machine learning or deep learning method [43, 3, 27].

301 **3.1. Introduction.** The simplest form of a neural network is the called *multilayer archite-*  
 302 *cture*, which is a simply stack by composition of modules, each module implements a function  
 303  $X_n = F_n(\theta_n, X_{n-1})$ , where  $X_n$  is a vector representing the output of module,  $\theta_n$  is the vector  
 304 of learnable parameters in the module, and  $X_{n-1}$  is the module input vector (as well as the  
 305 output of the previous module). The input of the first module  $X_0$  is an input pattern  $Z_0$ , the  
 306 output of the whole system is the one of the last module which denoted  $Z_l$ , where  $l$  is the  
 307 *number of layers*. In *gradient-based learning methods*, given a cost function  $\mathcal{L}^p(\cdot, \cdot)$  measuring  
 308 the discrepancy between the output of the system  $Z_l^p$  and  $D^p$  the “correct” or desired output  
 309 for the  $p$ -th input pattern. One is interested on minimizing the average discrepancy over a  
 310 set of input/output pairs called the *training set*,  $\{(Z_0^0, D^0), (Z_0^1, D^1), \dots, (Z_0^n, D^n)\}$ . The net-  
 311 work is initialized with randomly chosen weights  $\theta^0$ . The gradient of the error function with  
 312 respect to each parameter is computed and gradient descent is used to update the weights  
 313 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  $\eta$  is a learning rate, and  
 314 the computation of  $\frac{\partial \mathcal{L}(\theta)}{\partial \theta^i}$ , is performed by *backpropagation algorithm* through the layers [45].  
 315 Additionally, for structured data as images, *convolutional neural networks* (CNN) are nowa-  
 316 days the recommended solution. In CNNs, the same operator is computed in each pixel of  
 317 the image. This mechanism is called, *weight sharing*, and it has several advantages such as  
 318 it can reduce the model complexity and make the network easier to train [39]. Including any  
 319 new layer, like EMD, requires therefore the computation of the corresponding gradient of the  
 320 layer with respect to the parameters to be learnt.

321 **3.2. Derivatives of Morphological EMD.**



**Figure 4.** The Morphological Empirical Mode layer corresponds to a residual layer [18] where the processing block is the average between upper and lower envelopes  $(\hat{h}, \check{h})$ , in this case the pair  $(\varepsilon, \delta)$  is used as examples.

322 **3.2.1. Dilation and erosion.** Our approach involves dilation and erosion operators as  
 323 defined in (1.3) and (1.4). In general for rank operators, their derivative is zero in every  
 324 coordinate, except for that of the value attending the desired rank [42][37]. Accordingly, for  
 325 dilation operator,  $\delta_\lambda(x) = \sup_w \{f(x-w) + \text{SE}_\lambda(w)\} = \sup_w [u(w)]$  derivative with respect of  
 326 a parameter in the additive structuring function is given by

$$327 \quad (3.1) \quad \frac{\partial \delta_\lambda(x)}{\partial \lambda} = \frac{\partial \delta_\lambda(x)}{\partial u(w)} \frac{\partial u(w)}{\partial \lambda} = \begin{cases} \frac{\partial \text{SE}(w, \lambda)}{\partial \lambda} & \text{if } w \in \arg \max_w \delta_\lambda(x) \\ 0 & \text{otherwise} \end{cases}$$

328 where the operator  $\arg \max_x f(x) := \{x \mid \forall y : f(y) \leq f(x)\}$ . In other words,  $\arg \max$  is the  
 329 set of points  $x$ , for which  $f(x)$  attains the largest value of the function. Note that we do not  
 330 regard maximum as being attained at any  $x$  when  $f(x) = \infty$ , nor do we regard the minimum  
 331 as being attained at any  $x$  when  $f(x) = -\infty$ . Thus (3.1) means that there is no gradient with  
 332 respect to non-maximum values. Similarly for the erosion,  $\varepsilon_\lambda(x) = \inf_w [f(x+w) - \text{SE}_\lambda(w)] =$   
 333  $\inf_{w \in \text{SE}_\lambda} [u(w)]$

$$334 \quad (3.2) \quad \frac{\partial \varepsilon_\lambda(x)}{\partial \lambda} = \frac{\partial \varepsilon_\lambda(x)}{\partial u(w)} \frac{\partial u(w)}{\partial \lambda} = \begin{cases} -\frac{\partial \text{SE}(w, \lambda)}{\partial \lambda} & \text{if } w \in \arg \min \varepsilon_\lambda(x) \\ 0 & \text{otherwise} \end{cases}$$

335 there is only gradient with respect to minimum values.

336 Thus, we can compute the derivative for the different cases of morphological EMD (See  
 337 Table 1). For Quadratic EMD, the evolution of the parameter  $\lambda$  depends on the difference of  
 338 the norm to the value where the morphological operator attends their value, normalised by  
 339 the square of the current value of  $\lambda$ . However, curiously for general nonparametric MEMs  $\Phi_W$   
 340 their derivative does not depend on the scale of the parameter. The derivative for composition  
 341 operators, as opening or closing, can be easily compute by the chain rule.

342 **3.3. Implementation.** Different methods for learning morphological operators in neural  
 343 networks have been proposed in the literature:

Name	Expression	Partial Derivative
Quadratic MEM (2.3)	$\Phi_\lambda(x)$	$\frac{\partial \Phi_\lambda(x)}{\partial \lambda} = \frac{ y_\epsilon^\lambda(x) ^2 -  y_\delta^\lambda(x) ^2}{4\lambda^2}$
Nonflat MEM	$\Phi_W(x)$	$\frac{\partial \Phi_W}{\partial w_i} = \begin{cases} -\frac{1}{2} & \text{if } i \in y_\epsilon^W \\ \frac{1}{2} & \text{if } i \in y_\delta^W \\ 0 & \text{otherwise} \end{cases}$

Table 1

Considered  $\Phi$  and its correspondent derivative.  $y_\delta^W(x) \in \arg \max_y [f(x - y) - B(y, W)]$  and  $y_\epsilon^\lambda(x) \in \arg \min_y [f(x + y) + B(y, W)]$ .

- 344 1. Replace maximum and minimum operator by smooth differentiable approximations,  
345 making possible the use of conventional gradient descent learning approach via back-  
346 propagation, for instance using an approximation by counter-harmonic mean[33] or  
347 other generalizations [29].  
348 2. Morphological operations can be computed by combinations of depthwise and point-  
349 wise convolution with depthwise pooling [38] allowing the use of classical optimization  
350 procedures.  
351 3. Use original definition of morphological operator, and in the backpropagation step  
352 follows the approach used in classical max-pooling layers [6, 14, 34].

353 We follows the last approach. That means that the gradient in (1.3) and (1.4) will have  
354 values different from zero only for the first element equal to the arg max or arg min instead  
355 of the complete equivalence class. This is the implementation used in deep learning mod-  
356 ules based on as Tensorflow or Pytorch. An implementation of our approach is available in  
357 <http://www.cmm.mines-paristech.fr/~velasco/morpholayers/>

358 **3.3.1. Example of learning parameters in morphological operators.** We present a dummy  
359 example of supervised classification in two classes for a 1D signal in dimension  $p$ . Both classes  
360 have been generated by  $f(x) = \sin(2\pi/(c(x + \epsilon)))$  for  $x = 0, \dots, 10$  with spatial step of 0.02,  
361 where  $\epsilon$  is a random realisation of a normalized Gaussian distribution. For first class, we  
362 have used a period  $c = 2$  and for the second class a period  $c = 1.75$ . Some examples are  
363 illustrated in Figure 5(a). We explore the training process by using a simple architecture:  
364  $\hat{z} := model(x) = \frac{1}{1 + \exp(-\frac{1}{p} \sum_{i=1}^p \delta_\lambda(x_i))}$ , *i.e.*, a morphological dilation followed by a global av-  
365 erage pooling with a sigmoid activation function, also called the logistic function. Now, we  
366 want to show the computation of the partial derivative with respect to a loss function. As a  
367 manner of example, we use the mean squared error as loss function, *i.e.*,  $loss(z, \hat{z}) = (z - \hat{z})^2$ .

368 One can compute the gradient  $\frac{\partial loss(z, \hat{z})}{\partial \lambda}$  by using the chaining rule of derivative

$$369 \frac{\partial loss(z, \hat{z})}{\partial \lambda} = \frac{\partial loss(z, \hat{z})}{\partial \hat{z}} \frac{\partial \hat{z}}{\partial \sigma} \frac{\partial \sigma}{\partial \sum \delta_\lambda/p} \frac{\partial \sum \delta_\lambda/p}{\partial \lambda},$$

370 where  $\sigma$  is the sigmoid function. Remember that the derivative of the sigmoid function  $\sigma(x)$ ,  
371 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  
372 as decision function, the derivative of the parameter of the dilation with respect to the loss

373 function can be written by

$$374 \quad \frac{\partial \text{loss}(z, \hat{z})}{\partial \lambda} = \frac{(2m)(m(1-m))}{p} \sum_{i=1}^p \frac{\partial \delta_\lambda(x)}{\partial \lambda}.$$

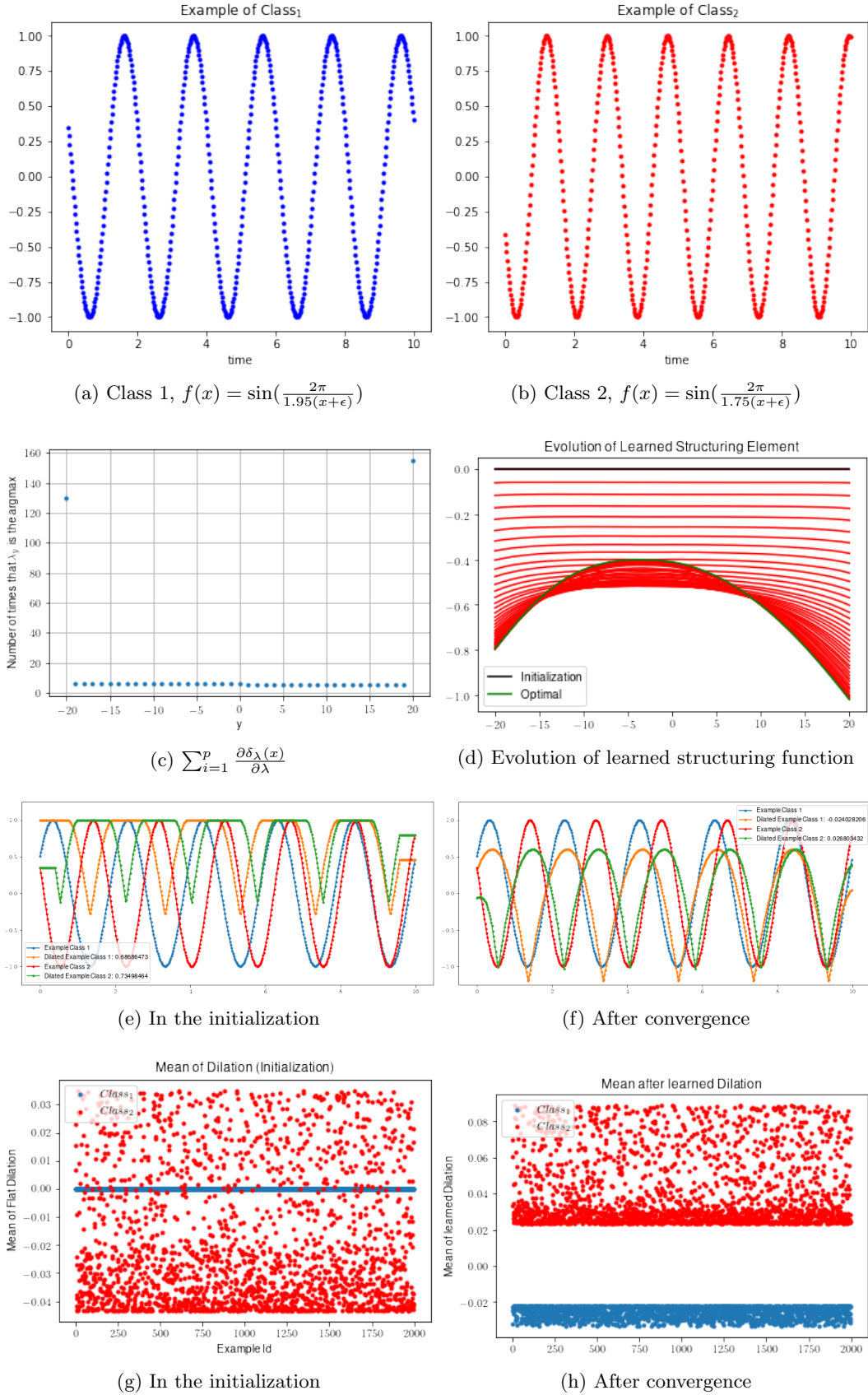
375 The first term is computed in the forward pass and it is the same for every parameter. We  
 376 decided to train a nonflat structuring function, so from (3.1), one can interpret the second  
 377 term as a counts the number of number of times that the spatial position in the structuring  
 378 function attains the maximal value, which is illustrated in Figure 5(c) for the last epoch of the  
 379 training. Additionally, the evolution of structuring function weights is given in Figure 5(d).  
 380 As a manner of example, two signals and its corresponding learned dilation are shown for the  
 381 initialization (as a flat structuring function) in Figure 5(e) and after convergence in Figure 5(f).  
 382 Finally, the decision function (mean value of the learned dilation) is shown for all the training  
 383 examples at initialisation Figure 5(g) and after convergence Figure 5(h). We highlight that  
 384 the learned structuring function seems to be an asymmetric quadratic with an additive bias.

385 **4. Experimental results on hyperspectral classification.** In this section, we investigate  
 386 the application of the proposed morphological empirical mode (Figure 4) to the problem of  
 387 signal classification. In particular, we will focus in the case of supervised classification of high-  
 388 dimensional 1D signals in hyperspectral images. The architecture chosen as baseline is the  
 389 one recommended in [40] and illustrated Figure 6. More specifically, the network is composed  
 390 of convolution layer, RELU, max-pooling. Each stage consists of twenty convolution layers  
 391 with a kernel size of 24 channels followed by ReLU activation, and a dense layer with batch  
 392 normalization. In the experimental section, the proposed morphological empirical mode will  
 393 be used as the first layer of an architecture of the baseline neural network.

394 **4.1. Considered datasets.** The aim of this section is to compare the results obtained by  
 395 different proposed EMD for 1D supervised classification problem. Accordingly, we used as  
 396 benchmark two classical hyperspectral images:

- 397 • *Pavia University* hyperspectral is a scene acquired by the ROSIS sensor in the north  
 398 of Italy. The dataset contains nine different classes including multiple solid structures,  
 399 natural objects and shadows (Figure 7(a-c)). After discarding the noisy bands, the  
 400 considered scene contains 103 spectral bands, with a size of  $610 \times 340$  pixels with  
 401 spatial resolution of 1.3 mpp and covering the spectral range from 0.43 to 0.86  $\mu\text{m}$ .
- 402 • *Indian Pines* dataset is a hyperspectral image captured over an agricultural area char-  
 403 acterized by its crops of regular geometry and also irregular forest regions. The scene  
 404 consists of  $145 \times 145$  pixels and with 224 spectral bands, which have been collected  
 405 in the wavelength range from 0.4 to 2.5  $\mu\text{m}$ . There are 16 different classes for train-  
 406 ing/testing set with a highly unbalanced distribution (Figure 7(d-f)).

407 **4.1.1. Protocol.** HSI scenes generally suffer from high intraclass variability and interclass  
 408 similarity, resulting from uncontrolled phenomena such as variations in illumination, presence  
 409 of areas shaded and/or covered by clouds, among others. Accordingly, the selection of training  
 410 samples must be carried out very carefully. We use spatial-disjoint samples which is the most  
 411 difficult and realist case according to previous works [40].

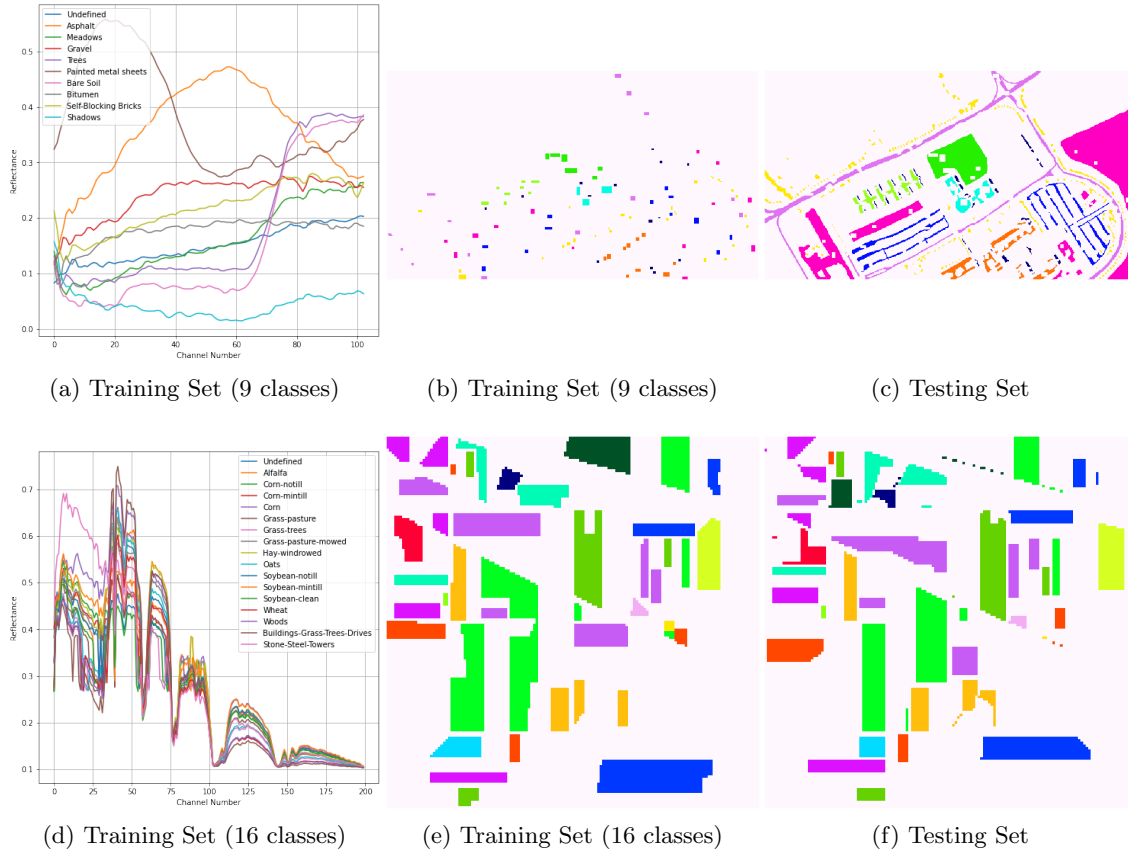


**Figure 5.** Evolution in the case of Nonflat structuring function

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 6.** 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.



**Figure 7.** (a) An example per class of the spectral information for Pavia University in spatial-disjoint (b) training and (c) testing examples. (d) An example per class of the spectral information for Indian Pines HSI using spatial disjoint (e) training and (f) testing examples.



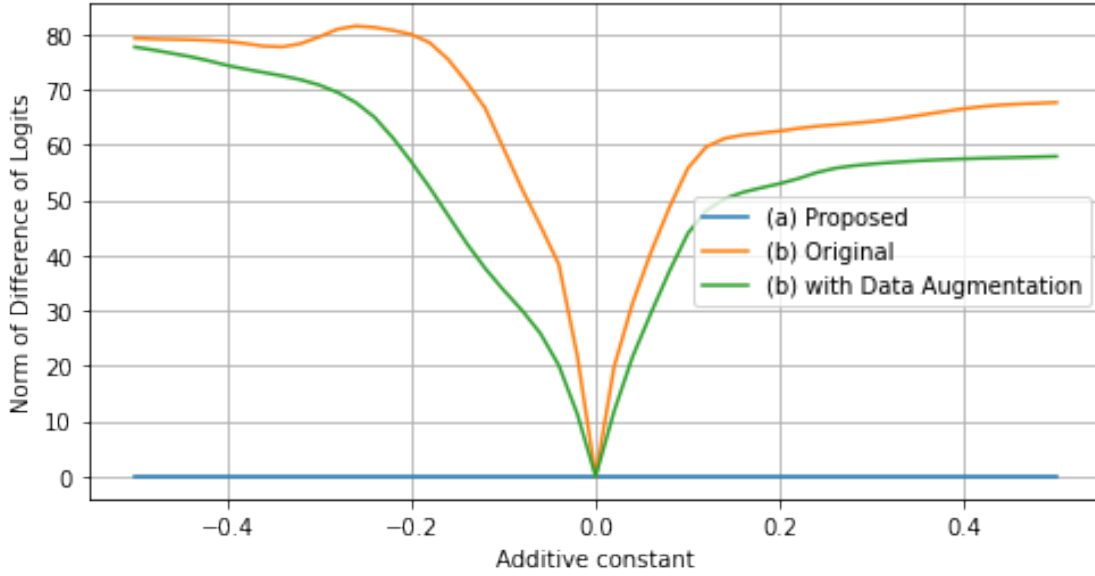
412 In gradient descent approaches the selection of random initialization of the parameter  
 413 value is critical. The aim of this initialization is to prevent layer activation outputs from  
 414 exploding or vanishing during the course of a forward pass [17]. While the source of difficulty  
 415 is well-understood, there is no universal remedy. For our MEM layers, we have used the  
 416 following initialization:

- 417 1. For non-flat structuring functions, a flat structuring element, *i.e.*,  $W$  a zero.
- 418 2. For quadratic structuring functions,  $\lambda$  is a random realization of a uniform distribution  
 419 between one and four, and for the parameter  $c$ , a uniform distribution between .5 and  
 420 .95.
- 421 3. For the parameter in the convex sum,  $\alpha$ , the value 1/2 is used.

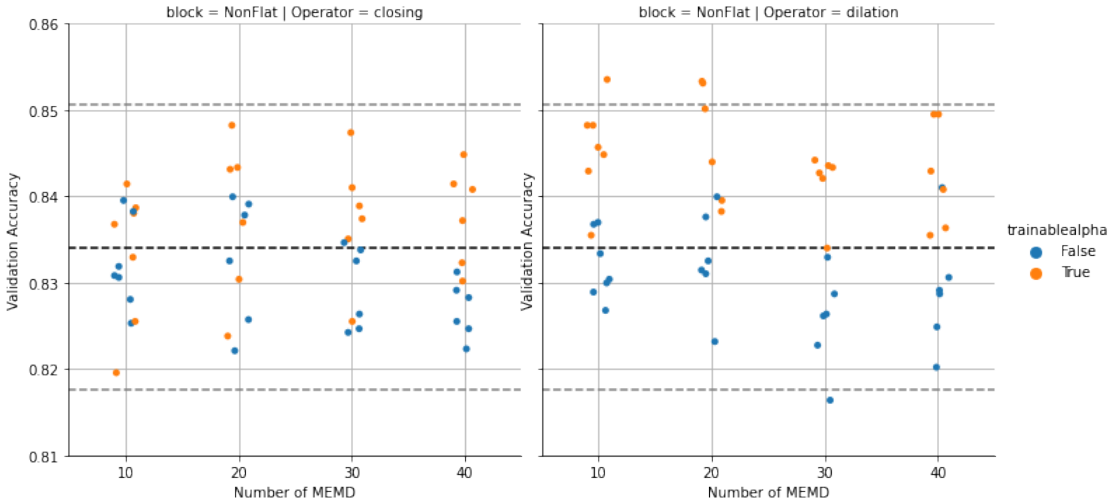
422 **4.1.2. Quantitative results.** We explore the use of proposed EMDs as feature extraction  
 423 layers, that means instead of learning from original spectral information, we will use the  
 424 residual of a single sifting process by MEMD which parameters are learned in a gradient-  
 425 based learning method by using the categorical cross-entropy as loss function.

426 The accuracy over testing set for ten different random initialization, for different proposed  
 427 envelopes by varying both the number of MEM from 10 to 40, and the type of structuring  
 428 function are shown in Figure 9 and Figure 10. In general, as expected, learning the param-  
 429 eter  $\alpha$  improves performance. We can observe that Quadratic MEMDs perform significantly  
 430 worse than other approaches. For a quantitative comparison, we have reported best, mean  
 431 and standard deviation after ten repetitions on both Indian Pines HSI (Table 2) and Pavia  
 432 University HSI (Table 3). From them, one can see that Nonflat MEMD performs much better  
 433 than the others structuring functions. In both cases, the best of proposed approaches have a  
 434 performance equivalent to our baseline, which is the state-of-the-art for the considered prob-  
 435 lems (Table 4). However, we remark that the inclusion of morphological EMDs induces an  
 436 invariant to additive intensity shifts in the classification model. This fact is shown in Figure 8,  
 437 for three models trained on Indian Pines.

438 **5. Discussion.** The paper investigated the formulation of EMD based on morphological  
 439 operators and its integration into deep learning architectures. The training of the layers  
 440 realizing the EMD process allows them to adapt the morphological models to the signals to  
 441 be classified. The assessments have been done with 1D signals from hyperspectral images  
 442 (*i.e.*, pixelwise spectra), but the proposed approaches are applicable to CNN architectures for  
 443  $n$ D images, without conceptual or algorithmic problem. 1D signals have been used for the  
 444 only reason that the effects of the process on such signals are easier to interpret in a research  
 445 perspective. Several variants of the morphological layers have been used. However, we think  
 446 that for a better understanding of some of the elements of the approach: behaviour of the  
 447 gradient of the layers during the optimization, contribution of the different parts of the signals  
 448 to the optimization, effect of the initialization, etc. a deeper theoretical and empirical study is  
 449 required. Additionally, we have illustrated the use of only one decomposition but the presented  
 450 framework allows us to go further. There are some interesting approaches to propose more  
 451 adapted optimization schemes [8], which reveals remarkable properties of network pruning by  
 452 these operators [56]. Finally, we will explore the use of: a) another structuring functions as  
 453 Poweroid or Anisotropic Quadratic proposed in [47] for MEMD, b) MEMD to produce Scale  
 454 Equivariant Neural Networks as in [46].



**Figure 8.** 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  $(\varepsilon, \delta)$ , 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.



**Figure 9.** Test accuracy for spatial-disjoint samples in Indian Pines Hyperspectral image. MEMD produces by envelopes for opening/closing (left figure) and erosion/dilation (right figure) by flat structuring function. 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 10 epochs). The horizontal lines indicates the maximum/average/minimum performance of baseline architecture [40] on original data. Blue points corresponds to  $\alpha = .5$ , where this parameter was not learned.

Type	Operator	$\alpha$	Overall Val. Acc.		Overall Training Acc.	
			Best	$\mu \pm \sigma$	Best	$\mu \pm \sigma$
Baseline	—	—	85.035	83.929±0.654	93.443	91.413±1.696
NonFlat	$(\gamma, \varphi)$	.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
	$(\varepsilon, \delta)$	.5	83.252	82.764 ± 0.576	97.451	95.226 ± 2.065
		True	<b>85.311</b>	84.052 ± 1.227	95.922	94.015 ± 2.717
	$(\varepsilon, \delta)$ 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	0.94874 ± 1.120
Quadratic	$(\gamma, \varphi)$	.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
	$(\varepsilon, \delta)$	.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	<del>82.424</del>	81.299 ± 0.983	96.941	95.674 ± 0.927

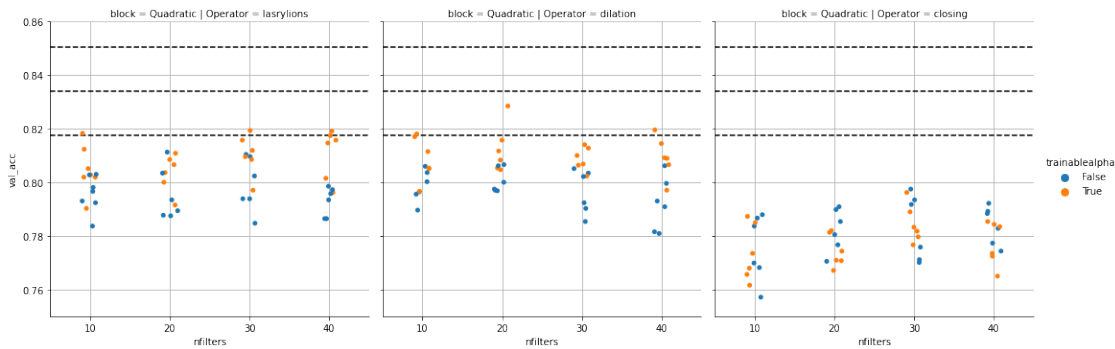
Table 2

Experiment on hyperspectral Indian Pines Disjoint classification problem. Each experiment have been repeat ten times varying the initialization of base architecture. 20 filter 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)

Type	Operator	$\alpha$	Overall Val. Acc.		Overall Training Acc.	
			Best	$\mu \pm \sigma$	Best	$\mu \pm \sigma$
Baseline	—	—	85.468	83.396± 2.42	92.527	86.447 ± 8.960
NonFlat	$(\gamma, \varphi)$	.5	79.543	78.189 ± .726	95.715	92.219 ± 2.408
		True	82.353	79.293 ± 1.767	96.353	91.525 ± 4.335
	$(\varepsilon, \delta)$	.5	84.261	82.681 ± .798	93.726	88.794 ± 4.998
		True	84.133	82.529 ± 1.131	93.879	89.735 ± 2.118
	$(\varepsilon, \delta)$ , SE(0) ≥ 0	.5	83.908	81.740 ± 1.295	93.216	84.575 ± 7.3338
		True	<b>85.483</b>	83.994 ± 1.238	94.389	89.617 ± 3.289
Quadratic	$(\gamma, \varphi)$	.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
	$(\varepsilon, \delta)$	.5	77.411	75.432 ± 1.193	95.052	86.470 ± 5.9395
		True	<i>81.196</i>	77.923 ± 1.700	92.476	86.593 ± 6.585
	Lasry-Lions	.5	77.461	76.396 ± .614	97.067	90.826 ± 6.223
		True	80.971	78.501 ± 1.332	96.123	87.082 ± 8.221

Table 3

Experiment on hyperspectral Pavia University for a disjoint training sample. Nine different classes. Each experiment have been repeat ten times varying the initialization of base architecture. 20 filter 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)



**Figure 10.** Test accuracy for spatial-disjoint samples in Indian Pines Hyperspectral image. MEMD produces by envelopes for Lasry-Lions operator (left figure), erosion/dilation (central figure), and opening/closing (right figure) by quadratic structuring functions. 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 indicates the maximum/average/minimum performance of baseline architecture [40] on original data. Blue points corresponds to  $\alpha = .5$ , where this parameter was not learned.

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
$\Phi_{\varepsilon, \delta}^{\alpha} + \text{CNN1D}$	85.31	85.48

**Table 4**

Comparison (in terms of OA) between different HSI classification models trained on spatial-disjoint samples.

455 **Acknowledgments.** This work has been supported by *Fondation Mathématique Jacques*  
 456 *Hadamard* (FMJH) under the PGMO-IRSDI 2019 program.

457

## REFERENCES

- 458 [1] J. ANGULO, *Lipschitz Regularization of Images supported on Surfaces using Riemannian Morphological*  
 459 *Operators*. working paper or preprint, Nov. 2014, <https://hal-mines-paristech.archives-ouvertes.fr/hal-01108130>.  
 460  
 461 [2] J. ANGULO AND S. VELASCO-FORERO, *Riemannian mathematical morphology*, *Pattern Recognition Letters*, 47 (2014), pp. 93–101.  
 462  
 463 [3] J. BEDI AND D. TOSHNIWAL, *Empirical mode decomposition based deep learning for electricity demand*  
 464 *forecasting*, *Ieee Access*, 6 (2018), pp. 49144–49156.  
 465 [4] Y. BENGIO, Y. LECUN, AND H. HINTON, *Deep learning*, *Nature*, 521 (2015), pp. 436–444.  
 466 [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,  
 467 pp. 157–161.  
 468 [6] Y.-L. BOUREAU, J. PONCE, AND Y. LECUN, *A theoretical analysis of feature pooling in visual recognition*,  
 469 in *ICML 2010*, 2010, pp. 111–118.  
 470 [7] M. CARLSSON, *On convex envelopes and regularization of non-convex functionals without moving global*  
 471

- 472 *minima*, Journal of Optimization Theory and Applications, 183 (2019), pp. 66–84.
- 473 [8] V. CHARISOPOULOS AND P. MARAGOS, *Morphological perceptrons: Geometry and training algorithms*, in
- 474 Mathematical Morphology and Its Applications to Signal and Image Processing, Springer Interna-
- 475 tional Publishing, 2017, pp. 3–15.
- 476 [9] J. CHEN, X. WANG, AND C. PLANIDEN, *A proximal average for prox-bounded functions*, SIAM Journal
- 477 on Optimization, 30 (2020), pp. 1366–1390.
- 478 [10] E. DELÉCHELLE, J. LEMOINE, AND O. NIANG, *Empirical mode decomposition: an analytical approach for*
- 479 *sifting process*, IEEE Signal Processing Letters, 12 (2005), pp. 764–767.
- 480 [11] E.-H. S. DIOP AND R. ALEXANDRE, *Analysis of intrinsic mode functions based on curvature motion-like*
- 481 *pdes*, Springer, (2015).
- 482 [12] R. A. E.-H. S. DIOP AND L. MOISAN, *Intrinsic nonlinear multiscale image decomposition : a 2d empirical*
- 483 *mode decomposition-like tool*, ELSEVIER, (2011).
- 484 [13] R. A. E.-H. S. DIOP AND V. PERRIER, *A pde model for 2d intrinsic mode functions*, IEEE ICIP, (2009).
- 485 [14] G. FRANCHI, A. FEHRI, AND A. YAO, *Deep morphological networks*, Pattern Recognition, 102 (2020),
- 486 p. 107246.
- 487 [15] J. GILLES, *Empirical wavelet transform*, IEEE transactions on signal processing, 61 (2013), pp. 3999–4010.
- 488 [16] J. GOUTSIAS AND H. HEIJMANS, *Mathematical Morphology*, IOS Press, 2000.
- 489 [17] B. HANIN AND D. ROLNICK, *How to start training: The effect of initialization and architecture*, in
- 490 Advances in Neural Information Processing Systems, 2018, pp. 571–581.
- 491 [18] K. HE, X. ZHANG, S. REN, AND J. SUN, *Deep residual learning for image recognition*, in Proceedings of
- 492 the IEEE conference on computer vision and pattern recognition, 2016, pp. 770–778.
- 493 [19] Z. HE, J. LI, L. LIU, AND Y. SHEN, *Three-dimensional empirical mode decomposition (temd): A fast*
- 494 *approach motivated by separable filters*, Signal Processing, 131 (2017), pp. 307–319.
- 495 [20] H. J. HEIJMANS AND P. MARAGOS, *Lattice calculus of the morphological slope transform*, Signal Process-
- 496 ing, 59 (1997), pp. 17–42.
- 497 [21] H. J. HEIJMANS AND C. RONSE, *The algebraic basis of mathematical morphology i. dilations and erosions*,
- 498 Computer Vision, Graphics, and Image Processing, 50 (1990), pp. 245–295.
- 499 [22] H. J. HEIJMANS AND R. VAN DEN BOOMGAARD, *Algebraic framework for linear and morphological scale-*
- 500 *spaces*, Journal of Visual Communication and Image Representation, 13 (2002), pp. 269–301.
- 501 [23] T. Y. HOU AND Z. SHI, *Adaptive data analysis via sparse time-frequency representation*, Advances in
- 502 Adaptive Data Analysis, 3 (2011), pp. 1–28.
- 503 [24] N. HUANG, S. ZHENG, S. LONG, M. WU, H. SHIH, Q. ZHENG, N.-C. YEN, C. TUNG, AND H. LIU, *The*
- 504 *empirical mode decomposition and the hilbert spectrum for nonlinear and non-stationary time series*
- 505 *analysis*, The Royal Society, 454 (1998), pp. 903–995.
- 506 [25] P. T. JACKWAY, *Morphological scale-spaces*, in Advances in Imaging and Electron Physics, P. W. Hawkes,
- 507 ed., vol. 99, Elsevier, 1997, pp. 1–64, [https://doi.org/https://doi.org/10.1016/S1076-5670\(08\)](https://doi.org/https://doi.org/10.1016/S1076-5670(08)70240-4)
- 508 [70240-4](https://www.sciencedirect.com/science/article/pii/S1076567008702404), <https://www.sciencedirect.com/science/article/pii/S1076567008702404>.
- 509 [26] P. T. JACKWAY AND M. DERICHE, *Scale-space properties of the multiscale morphological dilation-erosion*,
- 510 IEEE TPAMI, 18 (1996), pp. 38–51, <https://doi.org/10.1109/34.476009>.
- 511 [27] X.-B. JIN, N.-X. YANG, X.-Y. WANG, Y.-T. BAI, T.-L. SU, AND J.-L. KONG, *Deep hybrid model*
- 512 *based on emd with classification by frequency characteristics for long-term air quality prediction*,
- 513 Mathematics, 8 (2020), p. 214.
- 514 [28] P.-L. L. J.M. LASRY, *A remark on regularization in hilbert spaces*, Israel Journal of Mathematics, (1986).
- 515 [29] A. KIRSZENBERG, G. TOCHON, E. PUYBAREAU, AND J. ANGULO, *Going beyond p-convolutions to learn*
- 516 *grayscale morphological operators*, arXiv preprint arXiv:2102.10038, (2021).
- 517 [30] D. LOONEY AND D. P. MANDIC, *A machine learning enhanced empirical mode decomposition*, in IEEE
- 518 ICASSP, IEEE, 2008, pp. 1897–1900.
- 519 [31] P. MARAGOS, *Representations for morphological image operators and analogies with linear operators*,
- 520 Advances in imaging and electron physics, 177 (2013), pp. 45–187.
- 521 [32] P. MARAGOS, J. F. KAISER, AND T. F. QUATIERI, *Energy separation in signal modulations with appli-*
- 522 *cation to speech analysis*, IEEE transactions on signal processing, 41 (1993), pp. 3024–3051.
- 523 [33] J. MASCI, J. ANGULO, AND J. SCHMIDHUBER, *A learning framework for morphological operators using*
- 524 *counter-harmonic mean*, in Mathematical Morphology and Its Applications to Signal and Image
- 525 Processing, Springer, 2013, pp. 329–340.

- 526 [34] R. MONDAL, M. S. DEY, AND B. CHANDA, *Image restoration by learning morphological opening-closing*  
527 *network*, Mathematical Morphology - Theory and Applications, 4 (01 Jan. 2020), pp. 87 – 107.
- 528 [35] J.-J. MOREAU, *Proximité et dualité dans un espace hilbertien*, Bulletin de la Société mathématique de  
529 France, 93 (1965), pp. 273–299.
- 530 [36] J. J. MOREAU, *Inf-convolution, sous-additivité, convexité des fonctions numériques*, Journal de  
531 Mathématiques Pures et Appliquées, (1970).
- 532 [37] M. NAKASHIZUKA, *Image regularization with higher-order morphological gradients*, in 2015 23rd European  
533 Signal Processing Conference (EUSIPCO), IEEE, 2015, pp. 1820–1824.
- 534 [38] K. NOGUEIRA, J. CHANUSSOT, M. D. MURA, W. R. SCHWARTZ, AND J. A. DOS SANTOS, *An introduction*  
535 *to deep morphological networks*, 2019, <https://arxiv.org/abs/1906.01751>.
- 536 [39] S. J. NOWLAN AND G. E. HINTON, *Simplifying neural networks by soft weight-sharing*, Neural computa-  
537 tion, 4 (1992), pp. 473–493.
- 538 [40] M. PAOLETTI, J. HAUT, J. PLAZA, AND A. PLAZA, *Deep learning classifiers for hyperspectral imaging:*  
539 *A review*, ISPRS Journal of Photogrammetry and Remote Sensing, 158 (2019), pp. 279–317.
- 540 [41] N. PARIKH AND S. BOYD, *Proximal algorithms*, Foundations and Trends in optimization, 1 (2014),  
541 pp. 127–239.
- 542 [42] L. F. PESSOA AND P. MARAGOS, *Mrl-filters: A general class of nonlinear systems and their optimal*  
543 *design for image processing*, IEEE TIP, 7 (1998), pp. 966–978.
- 544 [43] X. QIU, Y. REN, P. N. SUGANTHAN, AND G. A. AMARATUNGA, *Empirical mode decomposition based*  
545 *ensemble deep learning for load demand time series forecasting*, Applied Soft Computing, 54 (2017),  
546 pp. 246–255.
- 547 [44] R. T. ROCKAFELLAR AND R. J.-B. WETS, *Variational analysis*, vol. 317, Springer Science & Business  
548 Media, 2009.
- 549 [45] R. ROJAS, *The backpropagation algorithm*, in Neural networks, Springer, 1996, pp. 149–182.
- 550 [46] M. SANGALLI, S. BLUSSEAU, S. VELASCO-FORERO, AND J. ANGULO, *Scale equivariant neural networks*  
551 *with morphological scale-spaces*, in Discrete Geometry and Mathematical Morphology, Springer, 2021,  
552 pp. 1–13.
- 553 [47] M. SCHMIDT AND J. WEICKERT, *Morphological counterparts of linear shift-invariant scale-spaces*, Journal  
554 of Mathematical Imaging and Vision, 56 (2016), pp. 352–366.
- 555 [48] J. SERRA, *Image Analysis and Mathematical Morphology*, Academic Press, Inc., Orlando, FL, USA, 1983.
- 556 [49] R. C. SHARPLEY AND V. VATCHEV, *Analysis of the intrinsic mode functions*, Constructive Approximation,  
557 24 (2006), pp. 17–47.
- 558 [50] S. SINCLAIR AND G. PEGRAM, *Empirical mode decomposition in 2-d space and time: a tool for space-time*  
559 *rainfall analysis and nowcasting*, Hydrology and Earth System Sciences, 9 (2005), pp. 127–137.
- 560 [51] A. STALLONE, A. CICONE, AND M. MATERASSI, *New insights and best practices for the successful use of*  
561 *empirical mode decomposition, iterative filtering and derived algorithms*, Scientific reports, 10 (2020),  
562 pp. 1–15.
- 563 [52] R. VAN DEN BOOMGAARD, L. DORST, S. MAKRAM-EBEID, AND J. SCHAVEMAKER, *Quadratic structuring*  
564 *functions in mathematical morphology*, in Mathematical morphology and its applications to image and  
565 signal processing, Springer, 1996, pp. 147–154.
- 566 [53] R. VAN DEN BOOMGAARD AND A. SMEULDERS, *The morphological structure of images: The differential*  
567 *equations of morphological scale-space*, IEEE transactions on pattern analysis and machine intelli-  
568 gence, 16 (1994), pp. 1101–1113.
- 569 [54] L. J. VAN VLIET, I. T. YOUNG, AND G. L. BECKERS, *A nonlinear laplace operator as edge detector in*  
570 *noisy images*, Computer Vision, Graphics, and Image Processing, 45 (1989), pp. 167 – 195.
- 571 [55] S. VELASCO-FORERO AND J. ANGULO, *On nonlocal mathematical morphology*, in International Sympo-  
572 sium on Mathematical Morphology and Its Applications to Signal and Image Processing, Springer,  
573 2013, pp. 219–230.
- 574 [56] Y. ZHANG, S. BLUSSEAU, S. VELASCO-FORERO, I. BLOCH, AND J. ANGULO, *Max-plus operators ap-*  
575 *plied to filter selection and model pruning in neural networks*, in Mathematical Morphology and Its  
576 Applications to Signal and Image Processing, Springer International Publishing, 2019, pp. 310–322.