



HAL
open science

Online deconvolution for industrial hyperspectral imaging systems

Yingying Song, El-Hadi Djermoune, Jie Chen, Cédric Richard, David Brie

► **To cite this version:**

Yingying Song, El-Hadi Djermoune, Jie Chen, Cédric Richard, David Brie. Online deconvolution for industrial hyperspectral imaging systems. *SIAM Journal on Imaging Sciences*, 2019, 12 (1), pp.54-86. 10.1137/18M1177640 . hal-01801272v3

HAL Id: hal-01801272

<https://hal.science/hal-01801272v3>

Submitted on 28 Nov 2018

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.

ONLINE DECONVOLUTION FOR INDUSTRIAL HYPERSPECTRAL IMAGING SYSTEMS*

YINGYING SONG[†], EL-HADI DJERMOUNE[†], JIE CHEN[‡], CÉDRIC RICHARD[§], AND DAVID BRIE[†]

Abstract. This paper proposes a hyperspectral image deconvolution algorithm for the online restoration of hyperspectral images as provided by whiskbroom and pushbroom scanning systems. We introduce a least-mean-squares (LMS)-based framework accounting for the convolution kernel non-causality and including non-quadratic (zero attracting and piece-wise constant) regularization terms. This results in the so-called sliding block regularized LMS (SBR-LMS) which maintains a linear complexity compatible with real-time processing in industrial applications. A model for the algorithm mean and mean-squares transient behavior is derived and the stability condition is studied. Experiments are conducted to assess the role of each hyper-parameter. A key feature of the proposed SBR-LMS is that it outperforms standard approaches in low SNR scenarios such as ultra-fast scanning.

Key words. hyperspectral image, online deconvolution, LMS, ZA-LMS

AMS subject classifications. 68U10

1. Introduction. Hyperspectral imaging has received considerable attention in the last decade as it combines the power of digital imaging and spectroscopy. Every pixel in a hyperspectral image provides local spectral information about a scene of interest across a large number of contiguous bands. This information can be used to characterize objects with great precision and details in a number of areas, including agricultural monitoring, industrial inspection, and defense. The core characteristics of hyperspectral images raise new data processing issues ranging from image restoration to pattern recognition [27, 10, 2, 38]. Several sensing techniques have been devised for hyperspectral imaging. They can be categorized into four main groups [46, 33]: whiskbroom (point scan), pushbroom (line scan), tunable filter (wavelength scan), and snapshot. Pushbroom systems are used in many areas such as food safety [13, 28], geo-referencing [7] and material sorting [42, 35]. The main feature of pushbroom imaging systems is that the hyperspectral data cubes are acquired slice by slice, sequentially in time (which in fact represents the so-called along track spatial dimension). This paper is a first step towards the development of advanced online (real-time) hyperspectral image processing methods required in industrial processes that aim at controlling and sorting input materials right after each line scanning. For example, this is the case for paper and chipboard plants that use recycled materials in the manufacturing process. The aim of this paper is to address the fast online (sequential) deconvolution of hyperspectral images captured by pushbroom imaging systems. However, it is worth noting that the proposed sequential deconvolution algorithm can be easily extended

*A short and preliminary version of this work appears in the conference publication [39].

Funding: This work has been supported by the FUI AAP 2015 Trispirabois Project, the Conseil Régional de Lorraine, the GDR ISIS CNRS, and the CNRS Imag'in ALOhA Project. The work of J. Chen was supported in part by National Natural Science Foundation of China (NSFC) under grant 61671382 and in part by NSF of Shenzhen under grant JCYJ2017030155315873.

[†]Centre de Recherche en Automatique de Nancy (CRAN), Université de Lorraine, CNRS, Boulevard des Aiguillettes B.P. 239 F-54506 Vandœuvre-lès-Nancy, France (firstname.lastname@univ-lorraine.fr).

[‡]School of Marine Science and Technology, Northwestern Polytechnical University, Xi'an, China (dr.jie.chen@ieee.org).

[§]Université Côte d'Azur, CNRS, OCA, Nice, France (cedric.richard@unice.fr).

38 to whiskbroom (point-by-point) systems.

39 Multichannel image restoration was carried out with Wiener methods in [29, 22].
40 Other strategies such as those in [23, 25, 27, 50] were also introduced, but only in
41 an offline setting. Let us mention the works of Jemec *et al.* [30, 31] which specifi-
42 cally address the calibration and (offline) deconvolution of pushbroom hyperspectral
43 images. The idea of restoring images in a sequential way can be traced back to the
44 end of the 70's and has susctied extensive works up to the mid 90's. Early works on
45 2D-image restoration are due to Woods and Radewan, who developed a Kalman-filter
46 (KF)-based approach allowing a pixel by pixel image restoration [48]. This approach
47 was then modified in [47] to lower the computational cost, resulting in the so-called re-
48 duced update KF which was then extended to the restoration of color images [43]. An
49 alternative KF-based approach was proposed by Biemond *et al.* [4] to perform a line-
50 by-line image restoration and which, due to Toeplitz matrix structures in the image
51 and observation models, can be efficiently implemented in the Fourier domain. This
52 approach was then extended to multichannel image restoration [21] allowing a slice
53 by slice restoration. The interest of KF-based image restoration started to decrease
54 in the mid 90's with the growth of the computational capabilities. This led to the
55 development of new iterative off-line image restoration approaches handling efficiently
56 non-quadratic terms (including edge-preserving regularization) as in [8, 3, 5]. There
57 is currently a renewed interest in extending these approaches to online estimation as
58 [40, 24, 15, 14] that comes from the need for processing tools that are able to handle
59 very large data sets. However, extending such approaches to online hyperspectral
60 image deconvolution can be tricky. For example, the online ADMM algorithm pro-
61 posed in [40] addresses the problem of estimating fixed parameters and derives online
62 algorithms that converge to the same solution as the batch ADMM. Basically they
63 consist of (sub)-gradient-based algorithms with a step size that decreases to zero as
64 the number of time samples increases. This may be irrelevant for online hyperspectral
65 image deconvolution since, in that case, the key feature is the tracking capability of
66 the algorithms.

67 The approach proposed in this paper has connections with [21] since it addresses
68 the slice-by-slice hyperspectral image restoration problem. It also uses the same ap-
69 proach to transform the non-causal blurring model into a (semi-)causal model. How-
70 ever, the KF equations need to update the state vector, the associated Kalman gain
71 and covariance matrices. This complexity may not be compatible with real-time pro-
72 cessing in industrial applications of hyperspectral imaging systems. The main motiva-
73 tion is to derive algorithms that allow to incorporate additional constraints enforcing
74 some prior properties in the restored images while maintaining a linear complexity
75 with respect to the number of unknown parameters. This work introduces an least
76 mean-squares (LMS) framework for sequential deconvolution of hyperspectral images
77 in which the subgradient is substituted by an instantaneous stochastic approximation
78 and the step size is constant. Our algorithm operates in the spirit of LMS-based
79 algorithms used, for instance, for adaptive system identification [45, 12, 9] and super-
80 resolution restoration of sequences of images [19, 16]. However, LMS-based sequential
81 deconvolution of (hyperspectral) images has never been reported in the literature. Ac-
82 counting for the specificities of the online image deconvolution (non-causal blurring,
83 low signal-to-noise ratio), we propose a sliding-block regularized least mean-squares
84 (SBR-LMS) algorithm that allows the restoration of images within a preset delay Q .

85 The paper is organized as follows. In Section 2, the image-blurring model is stated.
86 Causality of the convolution kernel and of the estimator are discussed. In Section 3,
87 the hyperspectral image deconvolution problem is introduced. The SBR-LMS algo-

88 rithm is proposed to address this problem. The criterion includes a spatial regularizer
 89 promoting the restoration of piecewise constant objects, a zero-attracting regularizer
 90 promoting the removal of the conveyor background, and a spectral regularizer pro-
 91 moting the smoothness of the spectral response. In Section 4, the transient behavior
 92 of the algorithm is analyzed in the mean and mean-squares sense. Its stability is
 93 also discussed. In Section 5, simulation results are presented to assess the accuracy of
 94 the convergence behavior model in both stationary and non-stationary scenarios. The
 95 role of each hyper-parameter is discussed. Numerical results are reported to assess the
 96 performance of the proposed approach and to compare it with state-of-the-art meth-
 97 ods. Finally, the algorithm is applied to real hyperspectral images acquired under
 98 different experimental conditions.

99 **2. Blurring and causality issues.**

100 **2.1. Scanning technologies and data structure.** Recording a 3D hyperspec-
 101 tral data cube can be performed in several ways, depending on how the 3D signal is
 102 acquired, as a sequence of either 1D or 2D signals. Each configuration results in a
 103 specific imaging system. Here, we restrict our attention to whiskbroom and push-
 104 broom scanners which are the most usual configurations in industrial applications.
 105 Figure 1 depicts schematically how a hyperspectral image is captured by whiskbroom
 106 and pushbroom imagers. In both cases, the scene is scanned by moving the imager
 107 or its field-of-view (FOV) across the scene. For example, for material analysis, the
 108 samples to be imaged are carried by a conveyor moving at constant speed, while in
 109 remote sensing applications, the imaging system is carried on an airborne platform
 110 moves over the scene. In a whiskbroom imaging system, an 1D sensor (*i.e.*, a spec-
 111 trometer) is used to form the 3D data cube. Thus, the whole spatial scene is observed
 112 point-by-point yielding a stream of spectra which are stacked into a 3D data cube.
 113 A mechanized angular movement using a scanning mirror sweeping allows to cover a
 114 spatial line from one edge to the other. Alternatively, pushbroom imaging systems
 115 make use of 2D sensors allowing to observe the scene line-by-line at each time instant.
 The stream of spatial-spectral arrays is stacked to form the 3D-data cube.

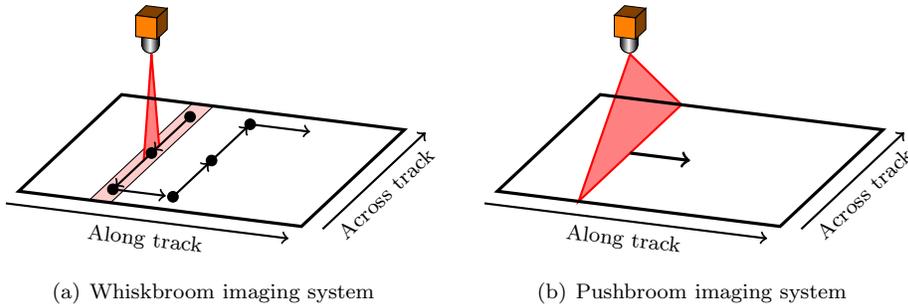


Fig. 1: Data acquisition by hyperspectral imaging systems

116 Following [36], respective pros and cons of whiskbroom and pushbroom imaging
 117 systems are as follows. Because whiskbroom systems rely on 1D detector, the cali-
 118 bration is much simpler than with pushbroom systems. The disadvantages of this
 119 design include the presence of a mechanical scanning system, the shorter integration
 120 time than in pushbroom systems, and the image forming geometry which is depen-
 121 dent on the scanning speed, the scan mirror arrangement and the conveyor movement.
 122

123 As compared to whiskbroom scan, pushbroom scanners allow for longer integration
 124 times for individual detector elements; the pixel dwell time is approximatively the
 125 inverse of the line frequency. In addition, there are distinct but fixed geometric rela-
 126 tions between the pixels within a scan line. Since 2D sensors are used as focal planes
 127 in these systems, the uniform calibration of the detector response is critical. Note
 128 that in a combined analysis of signal-to-noise ratio (SNR), uniformity, and stabil-
 129 ity, pushbroom scanners might not necessarily outperform whiskbroom systems even
 130 though they have a longer integration time. For example, [35] developed a whiskb-
 131 room hyperspectral imaging system for ultra fast material sorting application allowing
 132 to reach up to 3 meters per second scanning velocity. However, pushbroom scanners
 133 are increasingly used in industrial applications.

134 A hyperspectral image is a 3D data cube whose dimensions are generally pre-
 135 ferred to as across-track, along-track and spectral dimensions. However, in the se-
 136 quel, to avoid any confusion, they will be designated as spatial (across-track), time
 137 (along-track) and spectral dimensions. We will denote a hyperspectral image by
 138 $\mathbf{Y} \in \mathbb{R}^{N \times P \times K}$, where N , P , and K are the number of spatial, spectral and time
 139 measurements, respectively (see Figure 2). The hyperspectral image is then obtained
 140 slice by slice, each slice being denoted by $\mathbf{Y}_k \in \mathbb{R}^{N \times P}$, $k = 1, \dots, K$. The size of \mathbf{Y}
 increases with k , which can possibly grow to infinity.

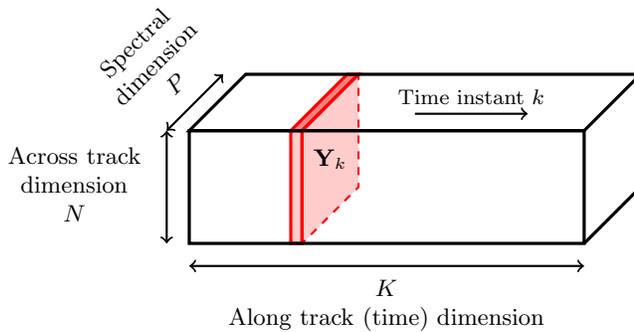


Fig. 2: Structure of a hyperspectral data cube \mathbf{Y}

141

142 **2.2. Blurring and noise.** A hyperspectral image \mathbf{Y} may suffer from distor-
 143 tions caused by the measuring devices and acquisition process such as keystone effect,
 144 smile effect, non-uniform illumination, and sensor response. These degradations may
 145 be corrected using standard calibration techniques [36, 20]. In this paper, we are pri-
 146 marily concerned with another type of distortion, namely, spatial distortions resulting
 147 in a loss of spatial resolution.

148 In a whiskbroom imaging system, the spatial resolution is mainly controlled by
 149 the size of the beam footprint on the scanned object. Additional distortion can be
 150 caused by the conveyor motion but, in this paper, we shall assume that the integra-
 151 tion time is small enough so that this distortion can be neglected. In general, the spatial
 152 sampling period ι_s is chosen equal to the beam footprint size yielding a low resolution
 153 image. To increase the spatial resolution, it is necessary to reduce ι_s , but this results
 154 in a spatial blurring of the hyperspectral image as sketched in Figure 3, where pixels
 155 are represented by the dots, and the beam footprint is represented by the circles.

156 The integration time T mainly acts on the noise level: a low value for T results in a
 157 low SNR. Consequently, assuming a constant acquisition velocity ι_s/T , any resolution
 158 improvement results in an increase of both blurring and noise level. Assuming that the
 159 scanning angle is small, the beam footprint size can be considered as invariant with
 160 respect to the pixel position. We also assume that the footprint size is independent of
 161 the wavelength. Thus, the corresponding distortion can be modeled by linear invariant
 162 convolution.

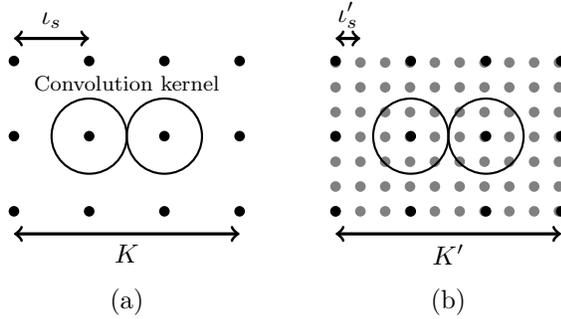


Fig. 3: Data acquisition with (a) low and (b) high spatial sampling rate. high spatial sampling rate results in a blurring of the image.

163 For pushbroom imaging systems, the quality of the acquired images can be signif-
 164 icantly degraded by the blurring effect resulting from the miscalibration of the optical
 165 system [30, 31, 20]. According to [41], there are many reasons that can cause spa-
 166 tial degradation of the pushbroom imaging system such as optical aberrations, slit
 167 size, and detector element size. In [31], the observed image corresponds to the 3D
 168 convolution of the original image with a spatial-spectral blurring filter varying with
 169 the spatial position. Here, the dependence of the spatial blurring with respect to the
 170 spatial position is neglected. We consider situations where the spectral response is
 171 smooth enough to choose a spectral sampling ι_λ (controlled by the spectral binning)
 172 greater than the support of the spectral blurring and a spatial sampling ι_s smaller
 173 than the support of the point spread function. This results in a degradation model
 174 involving only spatial blurring (possibly depending on the wavelength) of the hyper-
 175 spectral image which is given by Equation (2.1) in the next section. The integration
 176 time is supposed to be small for fast scanning yielding a low SNR.

177 **2.3. Causality.** We shall now discuss issues related to the causality of the convo-
 178 lution kernel and associated estimates. Following [27], hyperspectral image blurring
 179 can be seen as P simultaneous spatial convolutions. For each wavelength λ_p , the
 180 blurred spatial image $\mathbf{Y}^p \in \mathbb{R}^{N \times K}$ is given by the 2D convolution:

$$181 \quad (2.1) \quad \mathbf{Y}^p = \mathbf{H}^{*p} * \mathbf{X}^p + \mathbf{Z}^p$$

182 where $\mathbf{Y}^p \in \mathbb{R}^{N \times K}$ is the blurred spatial image for each wavelength λ_p , $\mathbf{X}^p \in \mathbb{R}^{N \times K}$
 183 is the image to restore, $\mathbf{H}^{*p} \in \mathbb{R}^{M \times L}$ is a convolution kernel (filter), and $\mathbf{Z}^p \in \mathbb{R}^{N \times K}$
 184 is an additive i.i.d. noise. The symbol $*$ stands for the 2D convolution operator. First
 185 we derive a sequential causal formulation of model (2.1). Without loss of generality,
 186 we focus on the sequential model for 2D images. The image \mathbf{Y}^p , collected online, can
 187 be represented as a sequence of vectors $\mathbf{y}_k^p := [y_{1,k}^p, \dots, y_{N,k}^p]^\top$, with $k = 1, \dots, K$,

188 where \top denotes the transpose of a matrix. We use the same notation for the columns
189 of \mathbf{X}^p . We assume a finite length blurring kernel of size L along the time dimension,
190 centered around 0 which means that both past and future values of \mathbf{x}_k^p contribute to
191 the observation \mathbf{y}_k^p . In order to make the blurring kernel causal¹, it has to be shifted
192 by $(L-1)/2$. It is thus necessary to delay the observation by $(L-1)/2$ samples, that
193 is, $\tilde{\mathbf{y}}_k^p = \mathbf{y}_{k-(L-1)/2}^p$. Writing $\mathbf{H}^{\star p} = [\mathbf{h}_L^p, \dots, \mathbf{h}_1^p]$ with $\mathbf{h}_\ell^p = [h_{M,\ell}^p, \dots, h_{1,\ell}^p]^\top$, model
194 (2.1) can be expressed as:

$$195 \quad (2.2) \quad \tilde{\mathbf{y}}_k^p = \mathbf{y}_{k-(L-1)/2}^p = \sum_{\ell=1}^L \mathbf{H}_\ell^p \mathbf{x}_{k-\ell+1}^p + \mathbf{z}_{k-(L-1)/2}^p$$

196 where the measurement noise \mathbf{z}_k^p is zero mean and statistically independent of
197 the other signals. \mathbf{H}_ℓ^p is the $N \times N$ Toeplitz matrix with first column and first
198 row given by $[h_{(M+1)/2,\ell}^p, \dots, h_{M,\ell}^p, 0, \dots, 0]$ and $[h_{(M+1)/2,\ell}^p, \dots, h_{1,\ell}^p, 0, \dots, 0]$, respec-
199 tively². Relation (2.2) introduces a delay in time dimension because the filter is
200 made causal along this dimension. Another consequence of causality issues concerns
201 the estimation process of \mathbf{x}_k^p . First, \mathbf{x}_k^p is involved in past and future observations
202 ($\mathbf{y}_{k-(L-1)/2}^p, \dots, \mathbf{y}_k^p, \dots, \mathbf{y}_{k+(L-1)/2}^p$). Secondly, let us consider the problem of esti-
203 mating only \mathbf{x}_k^p from the dataset:

$$204 \quad (2.3) \quad \tilde{\mathbf{y}}^p \triangleq \text{col}\{\tilde{\mathbf{y}}_k^p\}_{k=1}^K.$$

205 where $\text{col}\{\cdot\}$ stacks its vector arguments on top of each other. The least squares
206 criterion can be written as:

$$207 \quad (2.4) \quad \|\tilde{\mathbf{y}}^p - \mathbf{F}^p \mathbf{x}^p\|^2$$

208 where \mathbf{x}^p is built similarly to $\tilde{\mathbf{y}}^p$ and \mathbf{F}^p is a Toeplitz-block-Toeplitz matrix of proper
209 dimensions. To make the dependence of the criterion on \mathbf{x}_k^p explicit, we introduce the
210 following partitions:

$$211 \quad (2.5) \quad \mathbf{x}^p = \left[\mathbf{x}_{1:k-1}^{p\top}, \mathbf{x}_k^{p\top}, \mathbf{x}_{k+1:K}^{p\top} \right]^\top,$$

$$212 \quad (2.6) \quad \mathbf{F}^p = \left[\mathbf{F}_{1:k-1}^p, \mathbf{F}_k^p, \mathbf{F}_{k+1:K}^p \right]$$

214 where $\mathbf{x}_{i:j}^p \triangleq \text{col}\{\mathbf{x}_k^p\}_{k=i}^j$ and $\mathbf{F}_{i:j}^p$ is the the submatrix formed by the columns $(i-1)N+1$
215 through jN of \mathbf{F}^p . This results in:

$$216 \quad (2.7) \quad \|\tilde{\mathbf{y}}^p - \mathbf{F}^p \mathbf{x}^p\|^2 = \|\tilde{\mathbf{y}}^p - \mathbf{F}_{1:k-1}^p \mathbf{x}_{1:k-1}^p - \mathbf{F}_{k+1:K}^p \mathbf{x}_{k+1:K}^p - \mathbf{F}_k^p \mathbf{x}_k^p\|^2.$$

217 It is now clear that optimally estimating \mathbf{x}_k^p requires all the past estimations $\hat{\mathbf{x}}_{1:k-1}^p$ and
218 future estimations $\hat{\mathbf{x}}_{k+1:K}^p$, which precludes the derivation of a sequential estimator
219 for \mathbf{x}_k^p . To address these issues, we propose to produce the estimates $\hat{\mathbf{x}}_k^p$ with a delay
220 Q . More precisely, at time instant k , we shall estimate \mathbf{x}_{k-Q+1}^p from Q observations
221 ($\tilde{\mathbf{y}}_k^p, \tilde{\mathbf{y}}_{k-1}^p, \dots, \tilde{\mathbf{y}}_{k-Q+1}^p$), coarse posterior estimates ($\hat{\mathbf{x}}_k^p, \hat{\mathbf{x}}_{k-1}^p, \dots, \hat{\mathbf{x}}_{k-Q+2}^p$) refined as k
222 increases, and anterior estimates ($\hat{\mathbf{x}}_{k-Q}^p, \hat{\mathbf{x}}_{k-Q-1}^p, \dots, \hat{\mathbf{x}}_{k-Q-L+2}^p$) which are no longer
223 updated. This is the key idea of the sliding-block LMS deconvolution algorithm
224 developed in the next section.

¹For simplicity, L is assumed to be odd.

²For simplicity, M is assumed to be odd.

225 **3. Online image deconvolution.**

226 **3.1. Block Tikhonov.** In this section, we introduce the Tikhonov-like approach
 227 proposed in [27] and a direct extension for online hyperspectral image restoration.
 228 They will serve as reference methods to assess the performance of the proposed LMS-
 229 based method. This Tikhonov-like approach, originally developed in an offline setting,
 230 consists of seeking the minimum of a criterion composed of three terms: a data fitting
 231 term, a spatial regularizer, and a spectral regularizer:

$$232 \quad (3.1) \quad \min_{\mathbf{x}} \mathcal{J}(\mathbf{x}) = \frac{1}{2} \|\mathbf{y} - \mathbf{H}\mathbf{x}\|^2 + \frac{\eta_s}{2} \|\Delta_s \mathbf{x}\|^2 + \frac{\eta_\lambda}{2} \|\Delta_\lambda \mathbf{x}\|^2.$$

233 where \mathbf{x} and \mathbf{y} denote, respectively, the original and the observed vectorized hy-
 234 perspectral images. Operator Δ_s corresponds to a Laplacian filter promoting the
 235 smoothness along both spatial and time dimensions, and Δ_λ corresponds to a first-
 236 order derivative filter along the spectral dimension. Parameters η_s and η_λ are re-
 237 spectively the spatio-temporal and spectral regularization parameters. Matrix \mathbf{H} is
 238 block-diagonal and corresponds to the convolution kernel. Because the criterion is
 239 quadratic, solution of problem (3.1) results in the following linear estimator:

$$240 \quad (3.2) \quad \hat{\mathbf{x}} = (\mathbf{H}^T \mathbf{H} + \eta_s \Delta_s^T \Delta_s + \eta_\lambda \Delta_\lambda^T \Delta_\lambda)^{-1} \mathbf{H}^T \mathbf{y}.$$

241 Due to the block diagonal structure of \mathbf{H} , the estimator (3.2) can be efficiently im-
 242 plemented in the frequency domain (see [27] for details).

243 The Block Tikhonov (BT) deconvolution approach addresses problem (3.1) in an
 244 online way by sequentially restoring spatio-spectral arrays $\mathbf{X}_k \in \mathbb{R}^{N \times P}$. Let \mathbf{x}_k and
 245 \mathbf{y}_k be the vectorized matrices $\mathbf{X}_k \in \mathbb{R}^{N \times P}$ and $\mathbf{Y}_k \in \mathbb{R}^{N \times P}$, respectively:

$$246 \quad (3.3) \quad \mathbf{x}_k \triangleq \text{col} \{ \mathbf{x}_k^p \}_{p=1}^P, \quad \mathbf{y}_k \triangleq \text{col} \{ \mathbf{y}_k^p \}_{p=1}^P.$$

247 where superscript p refers to the spectral band. This BT algorithm implements esti-
 248 mator (3.2) with sliding blocks of Q samples $[\mathbf{x}_{k-Q+1}, \dots, \mathbf{x}_k], \forall k = Q, \dots, K$ and
 249 outputs the estimated vector $\hat{\mathbf{x}}_{k-(Q-1)/2}$ as the final result³. This algorithm works
 250 fast since only quadratic regularization terms are considered. However, when other
 251 constraints are needed, such as ℓ_1 -norm-based ones, no explicit solutions are available
 252 and time consuming iterative algorithms are required.

253 In the next section, we focus on extending the zero-attracting LMS (ZA-LMS)
 254 algorithm proposed in [12] to online deconvolution, while accounting for some image
 255 specificities (non-causal blurring, presence of smooth regions separated by abrupt
 256 edges, and low SNR) resulting in the so-called Sliding-block regularized LMS (SBR-
 257 LMS).

258 **3.2. Sliding-block regularized LMS (SBR-LMS).** Consider the problem of
 259 sequentially estimating \mathbf{x}_{k-Q+1} based on the delayed observations $(\hat{\mathbf{y}}_k, \dots, \hat{\mathbf{y}}_{k-Q+1})$
 260 and past estimates $\hat{\mathbf{x}}_{k-Q}, \dots, \hat{\mathbf{x}}_{k-Q-L+2}$. Due to the non-causality of the estimation
 261 process, we propose to determine the future estimates with respect to time $k - Q + 1$,
 262 namely, $\mathbf{x}_k, \dots, \mathbf{x}_{k-Q+2}$. In what follows, to simplify notations, \mathbf{y}_k refers to the

³For simplicity, Q is chosen to be odd.

263 delayed observation \tilde{y}_k and $\hat{\cdot}$ symbol is omitted. This results in the following criterion:

$$\begin{aligned}
264 \quad \mathcal{J}(\overbrace{\mathbf{x}_k, \dots, \mathbf{x}_{k-Q+1}}^{\text{updated}} \overbrace{\mathbf{x}_{k-Q}, \dots, \mathbf{x}_{k-Q-L+2}}^{\text{past estimates}}) &= \sum_{q=1}^Q \mathbb{E} \left\| \mathbf{y}_{k-q+1} - \sum_{\ell=1}^L \mathbf{H}_\ell \mathbf{x}_{k-q-\ell+2} \right\|^2 \\
265 \quad (3.4) \quad &+ \eta_z \sum_{q=1}^Q \|\mathbf{x}_{k-q+1}\|_1 + \eta_s \sum_{q=1}^Q \|\mathbf{D}_s \mathbf{x}_{k-q+1}\|_1 + \eta_\lambda \sum_{q=1}^Q \|\mathbf{D}_\lambda \mathbf{x}_{k-q+1}\|^2 \\
266
\end{aligned}$$

267 where $\mathbb{E}\{\cdot\}$ stands for the expectation operator, $\|\cdot\|_1 = \sum_n |\{\cdot\}_n|$ denotes the ℓ_1 -
268 norm of its argument, and $\{\cdot\}_n$ stands for the n -th entry of a vector. Matrix $\mathbf{H}_\ell \triangleq$
269 $\text{blkdiag}\{\mathbf{H}_\ell^p\}_{p=1}^P$ is a block-diagonal matrix.

270 The regularizer $\|\mathbf{D}_s \mathbf{x}_{k-q+1}\|_1$ promotes the restoration of piecewise constant pat-
271 terns along the spatial dimension. The first-order filtering operator \mathbf{D}_s is defined
272 as:

$$273 \quad (3.5) \quad \mathbf{D}_s \triangleq \mathbf{I}_P \otimes \mathbf{T}_N$$

274 where \otimes stands for the Kronecker product, matrix \mathbf{I}_J denotes the $J \times J$ identity matrix
275 and \mathbf{T}_J is the Toeplitz matrix of size $(J-1) \times J$ with first column $[1, 0, \dots, 0]$ and
276 first row $[1, -1, 0, \dots, 0]$. The zero-attracting regularizer $\|\mathbf{x}_{k-q+1}\|_1$ aims at removing
277 the conveyor background. The choice of these regularization terms is thus mainly
278 motivated by the targeted application, namely, the inspection of objects put on a
279 conveyor belt. At a given wavelength, the response of the conveyor after background
280 removal is close to zero while that of the objects is supposed to be piecewise constant.
281 The strength of the first derivative regularizers along spatial dimension is controlled
282 by $\eta_s \geq 0$. The strength of the zero-attracting regularizer is controlled by $\eta_z \geq 0$.

283 The spectral regularization term $\|\mathbf{D}_\lambda \mathbf{x}_{k-q+1}\|^2$ promotes spectral smoothness
284 which corresponds to the very nature of NIR hyperspectral images of biological mater-
285 ial such as wood [44]. Matrix \mathbf{D}_λ is a first-order filtering operator along the spectral
286 dimension weighted by the coefficients $\{c_p\}_{p=1}^{P-1}$. It is defined as:

$$287 \quad (3.6) \quad \mathbf{D}_\lambda \triangleq (\text{diag}(c_1, \dots, c_{P-1}) \mathbf{T}_P) \otimes \mathbf{I}_N.$$

288 The parameter η_λ controls the strength of the spectral smoothness penalty term.
289 The spectral weights c_p were introduced to provide some additional flexibility to the
290 algorithm. Basically, the strength of the weight c_p promotes the smoothness around
291 the p -th spectral band. In our case, the spectral data are sampled on a regular grid.
292 hence, c_1, \dots, c_{P-1} are all set to 1. however, in the case where the spectral data
293 would be recorded on two non adjacent spectral bands, smoothness between these
294 bands should not be promoted and the corresponding weight should be set to a small
295 value (zero).

296 The criterion (3.1) developed in [27] introduces a regularization using second order
297 derivatives along spatial and time dimensions. An inspection of criterion (3.4) may
298 lead to the assertion that no regularization term is acting along the time dimension
299 (time-regularization). However, we prove in Appendix B that, for the denoising case,
300 LMS-based algorithm implicitly introduces a time regularization and presents a time
301 delay in the estimation. A formal link between μ and the regularization strength
302 is given. A small value μ results in strong regularization strength and large delay
303 along the time dimension. Experimental results (Subsection 5.1) confirm that the
304 interpretation is valid for the deconvolution case.

305 A subgradient of (3.4) is given by:

$$306 \quad (3.7) \quad \nabla \mathcal{J}(\mathbf{x}_k, \dots, \mathbf{x}_{k-Q-L+2}) \triangleq \left(\frac{\partial \mathcal{J}}{\partial \mathbf{x}_k}, \dots, \frac{\partial \mathcal{J}}{\partial \mathbf{x}_{k-Q+1}}, \mathbf{0}_{N \times 1}, \dots, \mathbf{0}_{N \times 1} \right)$$

307 where $\mathbf{0}_{I \times J}$ denotes the $I \times J$ zero matrix. Zero terms in the subgradient indicate
308 that past estimates $\hat{\mathbf{x}}_{k-Q}, \dots, \hat{\mathbf{x}}_{k-Q-L+2}$ are no longer updated. We now derive the
309 sliding-block regularized LMS algorithm. Consider the vectorized data:

$$310 \quad (3.8) \quad \mathbf{x}'_k \triangleq \text{col}\{\mathbf{x}_{k-q+1}\}_{q=1}^{Q+L-1}, \quad \mathbf{y}'_k \triangleq \text{col}\{\mathbf{y}_{k-q+1}\}_{q=1}^{Q+L-1}.$$

311 A valid subgradient for $|x|$ is $\text{sign}(x)$ [18] where the sign function is defined as $\text{sign}(x) =$
312 0 for $x = 0$, and $\text{sign}(x) = x/|x|$ otherwise. Approximating the subgradient in (3.7)
313 by its instantaneous value yields:

$$314 \quad (3.9) \quad \nabla \mathcal{J}(\mathbf{x}'_k) = -2\Phi(\mathbf{y}'_k - \mathbf{G}\mathbf{x}'_k) + \eta_z \mathbf{\Gamma} \text{sign}(\mathbf{x}'_k) + \eta_s \mathbf{\Lambda}_s^\top \text{sign}(\mathbf{\Lambda}_s \mathbf{x}'_k) + 2\eta_\lambda \mathbf{\Lambda}_\lambda^\top \mathbf{\Lambda}_\lambda \mathbf{x}'_k$$

315 Matrices Φ , \mathbf{G} and $\mathbf{\Gamma}$ are of size $(Q+L-1)PN \times (Q+L-1)PN$. Matrix Φ is defined
316 by:

$$317 \quad (3.10) \quad \Phi \triangleq \begin{bmatrix} \Phi_{11}^\top & \mathbf{0}_{QP N \times (L-1)PN} \\ \mathbf{0}_{(L-1)PN \times (Q+L-1)PN} & \end{bmatrix}$$

318 where Φ_{11} is the matrix of size $QP N \times QP N$ defined as:

$$319 \quad (3.11) \quad \Phi_{11} \triangleq \begin{bmatrix} \mathbf{H}_1 & \cdots & \mathbf{H}_Q \\ & \ddots & \vdots \\ \mathbf{0} & & \mathbf{H}_1 \end{bmatrix}$$

320
321 in which $\mathbf{H}_\ell = \mathbf{0}_{PN \times PN}$ for $\ell > L$. It is necessary to define the matrix Φ_{11} as above
322 to properly account for the cases $Q < L$ and $Q \geq L$. Matrix \mathbf{G} has the form:

$$323 \quad (3.12) \quad \mathbf{G} \triangleq \begin{bmatrix} \mathbf{H}_1 & \cdots & \mathbf{H}_L & \mathbf{0} \\ & \ddots & & \ddots \\ \mathbf{0} & & \mathbf{H}_1 & \cdots & \mathbf{H}_L \\ \hline \mathbf{0}_{(L-1)PN \times (Q+L-1)PN} & & & & \end{bmatrix} \triangleq \begin{bmatrix} \Phi_{11} & \mathbf{G}_{12} \\ \mathbf{0}_{(L-1)PN \times (Q+L-1)PN} & \end{bmatrix},$$

324 (see Appendix A for more details). Matrix $\mathbf{\Gamma}$ is defined by:

$$325 \quad (3.13) \quad \mathbf{\Gamma} \triangleq \begin{bmatrix} \mathbf{I}_{QP N} & \mathbf{0}_{QP N \times (L-1)PN} \\ \mathbf{0}_{(L-1)PN \times (Q+L-1)PN} & \end{bmatrix}.$$

326 The first-order derivative filters for spatial and spectral dimensions are:

$$327 \quad (3.14) \quad \mathbf{\Lambda}_s \triangleq [\mathbf{I}_Q \otimes \mathbf{D}_s \quad \mathbf{0}_{QP(N-1) \times (L-1)PN}] \triangleq [\mathbf{\Lambda}_{s,1} \quad \mathbf{0}_{QP(N-1) \times (L-1)PN}],$$

328

$$329 \quad (3.15) \quad \mathbf{\Lambda}_\lambda \triangleq [\mathbf{I}_Q \otimes \mathbf{D}_\lambda \quad \mathbf{0}_{Q(P-1)N \times (L-1)PN}] \triangleq [\mathbf{\Lambda}_{\lambda,1} \quad \mathbf{0}_{Q(P-1)N \times (L-1)PN}].$$

330 Finally, the SBR-LMS algorithm for hyperspectral image deconvolution is given by:

$$331 \quad (3.16) \quad \hat{\mathbf{x}}'_{k+1} = \Omega \hat{\mathbf{x}}'_k + \mu \Phi(\mathbf{y}'_k - \mathbf{G}\hat{\mathbf{x}}'_k) - \rho_z \mathbf{\Gamma} \text{sign}(\hat{\mathbf{x}}'_k) - \rho_s \mathbf{\Lambda}_s^\top \text{sign}(\mathbf{\Lambda}_s \hat{\mathbf{x}}'_k) - \mu \eta_\lambda \mathbf{\Lambda}_\lambda^\top \mathbf{\Lambda}_\lambda \hat{\mathbf{x}}'_k$$

332 where $\rho_z = \mu\eta_z/2$, $\rho_s = \mu\eta_s/2$ and μ is a step size parameter that controls the trade
 333 off between convergence rate and algorithm stability. Matrix $\mathbf{\Omega}$ is given by:

334 (3.17)
$$\mathbf{\Omega} \triangleq \left[\begin{array}{c|c} \mathbf{I}_{QP_N} & \mathbf{0}_{QP_N \times (L-1)PN} \\ \hline \mathbf{0} & \mathbf{I}_{PN} \\ \mathbf{0} & \mathbf{0} \end{array} \middle| \begin{array}{c} \mathbf{0}_{QP_N \times (L-1)PN} \\ \mathbf{0}_{PN \times PN} \\ \mathbf{I}_{(L-2)PN} \\ \mathbf{0}_{(L-2)PN \times PN} \end{array} \right] \triangleq \left[\begin{array}{c|c} \mathbf{I}_{QP_N} & \mathbf{0}_{QP_N \times (L-1)PN} \\ \hline \mathbf{\Omega}_{21} & \mathbf{\Omega}_{22} \end{array} \right].$$

335 The upper part of matrix $\mathbf{\Omega}$ corresponds to the set of updated variables, and the
 336 lower part allows to shift the past estimates. The final result $\hat{\mathbf{x}}_{k-Q+2}$ is obtained by
 337 selecting the Q -th block of vector $\hat{\mathbf{x}}'_{k+1}$, that is,

338 (3.18)
$$\hat{\mathbf{x}}_{k-Q+2} = \mathbf{S}\hat{\mathbf{x}}'_{k+1}$$

339 where $\mathbf{S} \triangleq [\mathbf{0}_{PN \times (Q-1)PN}, \mathbf{I}_{PN}, \mathbf{0}_{PN \times (L-1)PN}]$. It is worth to mention that the pro-
 340 posed algorithm is different from the standard block-LMS algorithm for which the
 341 output \mathbf{x}_k is updated only once for every block of size Q . On the contrary, in the
 342 proposed algorithm, to account for the causality issues discussed in [Subsection 2.3](#),
 343 \mathbf{x}_k is updated Q times.

344 Finally, depending on the hyperparameter values, different LMS-like algorithms
 can be defined as presented in [Table 1](#).

Table 1: The different LMS-like algorithms and corresponding hyperparameters

| | μ | Q | η_z | η_s | η_λ |
|----------|-------|-------|----------|----------|----------------|
| ZA-LMS | > 0 | 1 | > 0 | 0 | 0 |
| SB-LMS | > 0 | > 1 | 0 | 0 | 0 |
| SBZA-LMS | > 0 | > 1 | > 0 | 0 | 0 |
| SBR-LMS | > 0 | > 1 | > 0 | > 0 | > 0 |

345

346 **3.3. Algorithm implementation and computational cost.** [Algorithm 3.1](#)
 347 details the SBR-LMS implementation. It should be noted that the storage of matrix
 348 \mathbf{H}_ℓ^p is avoided and that only 1D convolutions are required. Two different implemen-
 349 tations of the convolution are possible: direct (spatial) domain or frequency domain.
 350 We also specify the particular case of using separable convolution kernel in form of
 351 $\mathbf{H}^{*p} = \mathbf{g}^p \mathbf{f}^{p\top}$ where $\mathbf{f}^p = [f_L^p, \dots, f_1^p]^\top$ and $\mathbf{g}^p = [g_M^p, \dots, g_1^p]^\top$ which yields lower
 352 computational burden. The validity of the separable kernel assumption is mainly
 353 controlled by the calibration of the optical systems. If the optical elements are mis-
 354 aligned, then the kernel will be non separable. Otherwise, a separable Gaussian kernel
 355 is a good and widely used model (see [\[26\]](#) for details).

356 The computational complexity per iteration of SBR-LMS is given in [Table 2](#),
 357 where the costs of both direct and frequency domain implementations of the convolu-
 358 tion are given. The approximate costs in [Table 3](#) give the number of multiplications
 359 obtained by setting $Q = L = M$ and retaining the dominant complexity. In the
 360 separable kernel case, both direct and frequency domain implementations yield lower
 361 complexity than for a non-separable kernel. In the direct domain, the complexity is
 362 linear w.r.t. the number of unknowns PN . For $Q \gg \log_2 N$, we can reduce the com-
 363 plexity for the non-separable kernel case by applying the algorithm in the frequency
 364 domain while for separable kernel case, both direct domain and frequency domain
 365 yield a computational cost of the same order. As expected, when $Q \ll \log_2 N$, the
 366 frequency domain implementation provides no benefit.

Table 2: Computational cost per iteration of recursion (3.16)

| Operation | Multiplications | Additions |
|--|---|---|
| $\Omega \hat{\mathbf{x}}'_k + \mu \Phi(\mathbf{y}'_k - \mathbf{G} \hat{\mathbf{x}}'_k)$ direct domain, non-separable kernel | $PQ \left(L + \frac{Q+1}{2} \right) M \left(N - \frac{M-1}{2} \right)$ | $PQL \left[(M-1) \left(N - \frac{M-1}{2} \right) + N \right]$ $+ PQ \left[\frac{Q+1}{2} (M-1) \left(N - \frac{M-1}{2} \right) + \frac{Q-1}{2} N \right]$ |
| $\Omega \hat{\mathbf{x}}'_k + \mu \Phi(\mathbf{y}'_k - \mathbf{G} \hat{\mathbf{x}}'_k)$ direct domain, separable kernel | $PQ \left(L + \frac{Q+1}{2} \right) N + 2PQM \left(N - \frac{M-1}{2} \right)$ | $PQ \left(L + \frac{Q-1}{2} \right) N$ $+ 2PQ(M-1) \left(N - \frac{M-1}{2} \right)$ |
| $\Omega \hat{\mathbf{x}}'_k + \mu \Phi(\mathbf{y}'_k - \mathbf{G} \hat{\mathbf{x}}'_k)$ frequency domain, non-separable kernel | $4PQ \left(L + \frac{Q+1}{2} \right) N + 2PQ(L+1)N \log_2 N$ | $PQ(2L+Q-1)N + 2PQ(L+1)N \log_2 N$ |
| $\Omega \hat{\mathbf{x}}'_k + \mu \Phi(\mathbf{y}'_k - \mathbf{G} \hat{\mathbf{x}}'_k)$ frequency domain, separable kernel | $PQ(L+Q+1)N + 2PQN(2 \log_2 N + 4)$ | $PQ(L+Q)N + 4PQN \log_2 N$ |
| $\rho_z \mathbf{\Gamma} \text{sign}(\hat{\mathbf{x}}'_k)$ | 0 | 0 |
| $\rho_s \mathbf{\Lambda}_s^\top \text{sign}(\mathbf{\Lambda}_s \hat{\mathbf{x}}'_k)$ | 0 | $2(N-1)PQ$ |
| $\mu \eta_\lambda \mathbf{\Lambda}_\lambda^\top \mathbf{\Lambda}_\lambda \hat{\mathbf{x}}'_k$ | $4N(P-1)(Q+L-1)$ | $2NP(P-1)(Q+L-1)$ |

Algorithm 3.1 SBR-LMS algorithm for hyperspectral image deconvolution

Data: $\mathbf{Y} \in \mathbb{R}^{N \times K \times P}$, $\mathbf{H}^{*p} = [\mathbf{h}_L^p, \dots, \mathbf{h}_1^p]$, $\forall p = 1, \dots, P$

Result: $\mathbf{X} \in \mathbb{R}^{N \times K \times P}$

Initialization Q , μ , ρ_z , ρ_s , η_λ , c_1, \dots, c_{P-1} and \mathbf{T}_N

$$\mathbf{h}_\ell^p = \mathbf{0}_{N \times 1}, \quad \forall \ell = L + 1, \dots, Q, \quad \forall p = 1, \dots, P;$$

Selection of parameters $\mathbf{X} = \mathbf{0}_{N \times K \times P}$, $c_0 = c_P = 0$ **for** $k = L + Q - 1 : K - 1$ **do**

for $q = 1 : Q$ **do**

for $p = 1 : P$ **do**

 Non-separable convolution kernel:

$$\mathbf{err}_{k-q+1}^p = \mathbf{y}_{k-q+1}^p - \sum_{\ell=1}^L \mathbf{h}_\ell^p * \mathbf{x}_{k-q-\ell+2}^p$$

$$\mathbf{grad} = \sum_{i=1}^q \mathbf{h}_{q-i+1}^p * \mathbf{err}_{k-i+1}^p$$

 Separable convolution kernel:

$$\mathbf{err}_{k-q+1}^p = \mathbf{y}_{k-q+1}^p - \mathbf{g}^p * \sum_{\ell=1}^L f_\ell^p \mathbf{x}_{k-q-\ell+2}^p$$

$$\mathbf{grad} = \mathbf{g}^p * \sum_{i=1}^q f_{q-i+1}^p \mathbf{err}_{k-i+1}^p$$

 where $*$ is the 1D convolution operator.

$$\begin{aligned} \mathbf{x}_{k-q+2}^p = & \mathbf{x}_{k-q+1}^p + \mu \mathbf{grad} - \rho_z \text{sign}(\mathbf{x}_{k-q+1}^p) - \rho_s \mathbf{T}_N^T \text{sign}(\mathbf{T}_N \mathbf{x}_{k-q+1}^p) \\ & - \mu \eta_\lambda \left[(c_{p-1}^2 + c_p^2) \mathbf{x}_{k-q+1}^p - c_{p-1}^2 \mathbf{x}_{k-q+1}^{p-1} - c_p^2 \mathbf{x}_{k-q+1}^{p+1} \right]; \end{aligned}$$

end

end

end

Table 3: Approximate computational cost per iteration for the convolution in (3.16)

| | Direct domain | Frequency domain |
|----------------------|-----------------------|------------------------------------|
| Non-separable kernel | $\mathcal{O}(Q^3 PN)$ | $\mathcal{O}(Q^2 PN \log_2 N)$ |
| Separable kernel | $\mathcal{O}(Q^2 PN)$ | $\mathcal{O}(Q PN (Q + \log_2 N))$ |

367 **4. Transient behavior analysis.** In this section, we derive the transient behav-
 368 ior model of the deconvolution algorithm (3.16) both in the mean and mean-squares
 369 sense. The analysis provides us with an important guidance for selecting the step-
 370 size parameter, and also allows us to study the influence of the other parameters
 371 (Subsection 5.2). The stability condition of the algorithm is also studied.

372 **4.1. Mean and mean-squares transient behavior model.** Using the nota-
373 tions of Subsection 3.2, Equation (2.2) can be written as: $\mathbf{y}'_k = \mathbf{G}\mathbf{x}'_k + \mathbf{z}'_k$ with \mathbf{z}'_k a
374 zero-mean measurement noise of covariance $\sigma_z^2 \mathbf{I}_{(Q+L-1)NP}$. Taking the expectation of
375 both sides of (3.16), the mean sliding block vector evolves according to the following
376 recursion:

$$377 \quad (4.1) \quad \mathbb{E}\{\hat{\mathbf{x}}'_{k+1}\} = \mathbf{A}\mathbb{E}\{\hat{\mathbf{x}}'_k\} + \mu\Phi\mathbf{G}\mathbf{x}'_k - \rho_z\Gamma\mathbb{E}\{\text{sign}(\hat{\mathbf{x}}'_k)\} - \rho_s\Lambda_s^\top\mathbb{E}\{\text{sign}(\Lambda_s\hat{\mathbf{x}}'_k)\}$$

378 where $\mathbf{A} \triangleq \Omega - \mu\Phi\mathbf{G} - \mu\eta_\lambda\Lambda_\lambda^\top\Lambda_\lambda$. The mean vector can be then obtained by using:

$$379 \quad (4.2) \quad \mathbb{E}\{\hat{\mathbf{x}}_{k-Q+2}\} = \mathbf{S}\mathbb{E}\{\hat{\mathbf{x}}'_{k+1}\},$$

380 The main difficulty in (4.1) lies in evaluating the expectation of the sign function.
381 This point will be discussed later.

382 Consider the error:

$$383 \quad (4.3) \quad \boldsymbol{\epsilon}_k = \mathbf{S}(\mathbf{y}'_k - \mathbf{G}\hat{\mathbf{x}}'_k) = \mathbf{S}(\mathbf{z}'_k - \mathbf{G}(\hat{\mathbf{x}}'_k - \mathbf{x}'_k)),$$

384 as \mathbf{z}'_k is assumed independent of other variables, the mean-squared error (MSE) can
385 be expressed as:

$$386 \quad \begin{aligned} \mathbb{E}\|\boldsymbol{\epsilon}_k\|^2 &= NP\sigma_z^2 + \text{trace}(\mathbf{G}^\top\mathbf{S}^\top\mathbf{S}\mathbf{G}\mathbb{E}\{\hat{\mathbf{x}}'_k\hat{\mathbf{x}}'^\top_k\}) \\ &\quad + \mathbf{x}'_k{}^\top\mathbf{G}^\top\mathbf{S}^\top\mathbf{S}\mathbf{G}\mathbf{x}'_k - 2\mathbf{x}'_k{}^\top\mathbf{G}^\top\mathbf{S}^\top\mathbf{S}\mathbf{G}\mathbb{E}\{\hat{\mathbf{x}}'_k\}. \end{aligned}$$

387 It follows that $\mathbb{E}\{\hat{\mathbf{x}}'_k\hat{\mathbf{x}}'^\top_k\}$ can be updated as:

$$388 \quad \begin{aligned} \mathbb{E}\{\hat{\mathbf{x}}'_{k+1}\hat{\mathbf{x}}'^\top_{k+1}\} &= \mathbf{A}\mathbb{E}\{\hat{\mathbf{x}}'_k\hat{\mathbf{x}}'^\top_k\}\mathbf{A}^\top + \mu^2\sigma_z^2\Phi\Phi^\top + \mu\Phi\mathbf{G}\mathbf{x}'_k\mathbb{E}\{\hat{\mathbf{x}}'^\top_k\}\mathbf{A}^\top \\ &\quad + \mu\mathbf{A}\mathbb{E}\{\hat{\mathbf{x}}'_k\}\mathbf{x}'_k{}^\top\mathbf{G}^\top\Phi^\top + \mu^2\Phi\mathbf{G}\mathbf{x}'_k\mathbf{x}'_k{}^\top\mathbf{G}^\top\Phi^\top \\ &\quad - \rho_z(\mathbf{U}_1 + \mathbf{U}_1^\top + \mathbf{U}_2 + \mathbf{U}_2^\top) + \rho_z^2\mathbf{U}_3 \\ &\quad - \rho_s(\mathbf{U}_4 + \mathbf{U}_4^\top + \mathbf{U}_5 + \mathbf{U}_5^\top) + \rho_s^2\mathbf{U}_6 + \rho_s\rho_z(\mathbf{U}_7 + \mathbf{U}_7^\top) \end{aligned}$$

389 where

$$\begin{aligned} 390 \quad \mathbf{U}_1 &= \mathbf{A}\mathbb{E}\{\hat{\mathbf{x}}'_k\text{sign}(\hat{\mathbf{x}}'_k)^\top\}\Gamma^\top, \\ 391 \quad \mathbf{U}_2 &= \mu\Phi\mathbf{G}\mathbf{x}'_k\mathbb{E}\{\text{sign}(\hat{\mathbf{x}}'_k)^\top\}\Gamma^\top, \\ 392 \quad \mathbf{U}_3 &= \Gamma\mathbb{E}\{\text{sign}(\hat{\mathbf{x}}'_k)\text{sign}(\hat{\mathbf{x}}'_k)^\top\}\Gamma^\top, \\ 393 \quad \mathbf{U}_4 &= \mathbf{A}\mathbb{E}\{\hat{\mathbf{x}}'_k\text{sign}(\Lambda_s\hat{\mathbf{x}}'_k)^\top\}\Lambda_s, \\ 394 \quad \mathbf{U}_5 &= \mu\Phi\mathbf{G}\mathbf{x}'_k\mathbb{E}\{\text{sign}(\Lambda_s\hat{\mathbf{x}}'_k)^\top\}\Lambda_s, \\ 395 \quad \mathbf{U}_6 &= \Lambda_s^\top\mathbb{E}\{\text{sign}(\Lambda_s\hat{\mathbf{x}}'_k)\text{sign}(\Lambda_s\hat{\mathbf{x}}'_k)^\top\}\Lambda_s, \\ 396 \quad \mathbf{U}_7 &= \Gamma\mathbb{E}\{\text{sign}(\hat{\mathbf{x}}'_k)\text{sign}(\Lambda_s\hat{\mathbf{x}}'_k)^\top\}\Lambda_s. \end{aligned}$$

398 Again, the main difficulty lies in the evaluation of the expectations in form of
399 $\mathbb{E}\{\text{sign}(u)\text{sign}(v)\}$ and $\mathbb{E}\{u\text{sign}(v)\}$. An analysis of the stochastic behavior of ZA-
400 LMS is carried out in [11] where exact expressions of these expectations are derived
401 under a Gaussian assumption on the error vectors. however, due to the size of the
402 covariance matrix (4.4), we decided to resort to use the following approximations [49]:

$$403 \quad (4.4) \quad \mathbb{E}\{\text{sign}(u)\} \approx \text{sign}(\mathbb{E}\{u\}),$$

$$404 \quad (4.5) \quad \mathbb{E}\{u\text{sign}(v)^\top\} \approx \mathbb{E}\{u\}\text{sign}(\mathbb{E}\{v\})^\top,$$

$$405 \quad (4.6) \quad \mathbb{E}\{\text{sign}(u)\text{sign}(v)^\top\} \approx \text{sign}(\mathbb{E}\{u\})\text{sign}(\mathbb{E}\{v\})^\top$$

407 which lead to sufficiently accurate models. When $|u|$ and $|v|$ are large, the approxi-
 408 mations are quite good meaning that the mean behavior is well approximated when
 409 the dynamic of the image to restore is high enough. When they are close to zero, the
 410 approximations are of lower quality. However, following [12], these approximations
 411 yield a worst case transient behavior model. In addition, this approximation does
 412 not assume any statistical model on u and v . Only their statistical independence is
 413 required.

414 The mean and mean-squares transient models will be used to study the SBR-LMS
 415 behavior in both steady-state and non-stationary cases (Subsection 5.1). In the next
 416 section, to study the stability condition of the algorithm, we focus on the steady-state
 417 case.

418 **4.2. Stability condition.** We shall now discuss the stability of the proposed
 419 algorithm in the mean and mean-squares sense. Consider a steady-state solution for
 420 which:

$$421 \quad (4.7) \quad \mathbf{x}_k^p = \bar{\mathbf{x}}^p, \quad \forall k.$$

422 The error-vector \mathbf{v}_k^p is defined as:

$$423 \quad (4.8) \quad \mathbf{v}_k^p = \bar{\mathbf{x}}^p - \hat{\mathbf{x}}_k^p$$

424 with its vectorized version

$$425 \quad (4.9) \quad \mathbf{v}_k \triangleq \text{col}\{\mathbf{v}_k^p\}_{p=1}^P$$

$$426 \quad (4.10) \quad \mathbf{v}'_k \triangleq \text{col}\{\mathbf{v}_{k-q+1}\}_{q=1}^{Q+L-1}.$$

428 By combining (3.16), (4.8) and (4.10), \mathbf{v}'_k can be updated as:

$$429 \quad (4.11) \quad \begin{aligned} \mathbf{v}'_{k+1} &= \mathbf{A}\mathbf{v}'_k - \mu\Phi\mathbf{z}'_k + (\mathbf{I} - \Omega + \mu\eta_\lambda\Lambda_\lambda^\top\Lambda_\lambda)\bar{\mathbf{x}}' + \rho_z\Gamma\text{sign}(\hat{\mathbf{x}}'_k) + \rho_s\Lambda_s^\top\text{sign}(\Lambda_s\hat{\mathbf{x}}'_k) \\ &= \mathbf{A}\mathbf{v}'_k - \mu\Phi\mathbf{z}'_k + \mu\eta_\lambda\Lambda_\lambda^\top\Lambda_\lambda\bar{\mathbf{x}}' + \rho_z\Gamma\text{sign}(\hat{\mathbf{x}}'_k) + \rho_s\Lambda_s^\top\text{sign}(\Lambda_s\hat{\mathbf{x}}'_k) \end{aligned}$$

430 where $\bar{\mathbf{x}} \triangleq \text{col}\{\bar{\mathbf{x}}^p\}_{p=1}^P$ and $\bar{\mathbf{x}}' \triangleq [\bar{\mathbf{x}}^\top, \dots, \bar{\mathbf{x}}^\top]^\top$ of size $(Q+L-1)PN \times 1$. We consider
 431 the following partition of the matrix \mathbf{A} :

$$432 \quad \mathbf{A} = \left[\begin{array}{c|c} \mathbf{I} & \mathbf{0} \\ \hline \Omega_{21} & \Omega_{22} \end{array} \right] - \mu \left[\begin{array}{c|c} \Phi_{11}^\top & \mathbf{0} \\ \hline \mathbf{0} & \mathbf{0} \end{array} \right] \left[\begin{array}{c|c} \Phi_{11} & \mathbf{G}_{12} \\ \hline \mathbf{0} & \mathbf{0} \end{array} \right] - \mu\eta_\lambda \left[\begin{array}{c|c} \Lambda_{\lambda,1}^\top\Lambda_{\lambda,1} & \mathbf{0} \\ \hline \mathbf{0} & \mathbf{0} \end{array} \right]$$

$$433 \quad (4.12) \quad \triangleq \left[\begin{array}{c|c} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \hline \Omega_{21} & \Omega_{22} \end{array} \right]$$

434

435 where Φ_{11} is defined by (3.11), Ω_{21} and Ω_{22} are defined by (3.17), $\mathbf{A}_{11} = \mathbf{I} -$
 436 $\mu(\Phi_{11}^\top\Phi_{11} + \eta_\lambda\Lambda_{\lambda,1}^\top\Lambda_{\lambda,1})$, $\mathbf{A}_{12} = -\mu\Phi_{11}^\top\mathbf{G}_{12}$.

437 *Mean stability.* We now consider the mean stability of (4.11). Taking the expecta-
 438 tion of both sides, the mean-error vector can be expressed as:

$$439 \quad (4.13) \quad \mathbb{E}\{\mathbf{v}'_{k+1}\} = \mathbf{A}\mathbb{E}\{\mathbf{v}'_k\} + \mathbf{d}_k$$

440 where

$$441 \quad (4.14) \quad \mathbf{d}_k \triangleq \mu\eta_\lambda\Lambda_\lambda^\top\Lambda_\lambda\bar{\mathbf{x}}' + \rho_z\mathbb{E}\{\text{sign}(\hat{\mathbf{x}}'_k)\} + \rho_s\Lambda_s^\top\mathbb{E}\{\text{sign}(\Lambda_s\hat{\mathbf{x}}'_k)\}.$$

442 Let $\mathbb{E}\{\mathbf{v}'_1\}$ be the initial condition. Then we obtain:

$$443 \quad (4.15) \quad \mathbb{E}\{\mathbf{v}'_{k+1}\} = (\mathbf{A})^k \mathbb{E}\{\mathbf{v}'_1\} + \sum_{n=0}^{k-1} (\mathbf{A})^n \mathbf{d}_{k-n}$$

444 The Euclidean norm of \mathbf{d}_k is bounded. Indeed, we have:

$$445 \quad (4.16) \quad \|\mathbf{d}_k\| \leq \mu\eta_\lambda \|\mathbf{\Lambda}_\lambda^\top \mathbf{\Lambda}_\lambda \bar{\mathbf{x}}'\| + \rho_z \|\mathbb{E}\{\text{sign}(\hat{\mathbf{x}}'_k)\}\| + \rho_s \|\mathbf{\Lambda}_s^\top \mathbb{E}\{\text{sign}(\mathbf{\Lambda}_s \hat{\mathbf{x}}'_k)\}\|.$$

446 The first term $\|\mathbf{\Lambda}_\lambda^\top \mathbf{\Lambda}_\lambda \bar{\mathbf{x}}'\|$ corresponds to the initial condition and is bounded if $\bar{\mathbf{x}}'$ is
447 bounded. The last two terms are also bounded:

$$448 \quad (4.17) \quad \|\mathbb{E}\{\text{sign}(\hat{\mathbf{x}}'_k)\}\| \leq \sqrt{QPN}$$

$$449 \quad (4.18) \quad \|\mathbf{\Lambda}_s^\top \mathbb{E}\{\text{sign}(\mathbf{\Lambda}_s \hat{\mathbf{x}}'_k)\}\| \leq \sqrt{QP(4N-6)}.$$

451 Following the arguments reported in [17], if the spectral norm of \mathbf{A} is less than 1,
452 the series $\sum_{n=0}^{\infty} (\mathbf{A})^n \mathbf{d}_{k-n}$ is convergent. Thus, the convergence of (4.15) is entirely
453 controlled by the spectral norm of \mathbf{A} . Because matrix \mathbf{A} is not symmetric, standard
454 stability results cannot be applied straightforwardly. Let \mathbb{D} be the closed unit disc
455 in the complex plane. We say that the matrix \mathbf{A} is discrete stable (see [6]) if $\delta(z) =$
456 $\det(\mathbf{I} - z\mathbf{A}) \neq 0, \forall z \in \mathbb{D}$. Using the Schur determinant formula, it can be shown that:

$$457 \quad (4.19) \quad \delta(z) = \det(\mathbf{I} - z\mathbf{A}_{11}) \cdot \det(\mathbf{I} - z\mathbf{\Omega}_{22} - z^2\mathbf{\Omega}_{21}(\mathbf{I} - z\mathbf{A}_{11})^{-1}\mathbf{A}_{12}).$$

458 Then, $\delta(z) \neq 0$ if and only if the following two conditions hold:

$$459 \quad (4.20) \quad \det(\mathbf{I} - z\mathbf{A}_{11}) \neq 0, \quad \forall z \in \mathbb{D}$$

$$460 \quad (4.21) \quad \det(\mathbf{I} - z\mathbf{\Omega}_{22} - z^2\mathbf{\Omega}_{21}(\mathbf{I} - z\mathbf{A}_{11})^{-1}\mathbf{A}_{12}) \neq 0, \quad \forall z \in \mathbb{D}$$

462 The condition in (4.20) is fulfilled if all the eigenvalues of the symmetric matrix \mathbf{A}_{11}
463 lie inside the unit disc, i.e., its spectral radius $r(\mathbf{A}_{11}) < 1$. The matrix $\mathbf{\Phi}_{11}^\top \mathbf{\Phi}_{11} +$
464 $\eta_\lambda \mathbf{\Lambda}_{\lambda,1}^\top \mathbf{\Lambda}_{\lambda,1}$ being the sum of a positive semi-definite matrix and a positive definite
465 matrix, it is positive definite; thus, standard stability result applies from which we
466 conclude that \mathbf{A}_{11} is stable iff:

$$467 \quad (4.22) \quad 0 < \mu < \frac{2}{r(\mathbf{\Phi}_{11}^\top \mathbf{\Phi}_{11} + \eta_\lambda \mathbf{\Lambda}_{\lambda,1}^\top \mathbf{\Lambda}_{\lambda,1})}.$$

468 In practice, it is not easy to check the second condition in (4.21) for all values of z
469 in the unit disc. However, in the case of Gaussian convolution kernels, we checked
470 experimentally that the stability of \mathbf{A} depends only on that of \mathbf{A}_{11} . For example,
471 Figure 4 displays the spectral radius of \mathbf{A} as a function of the tuning parameters
472 μ and η_λ for two Gaussian kernels. The red curve is obtained using (4.22) which
473 corresponds to the stability limit of \mathbf{A}_{11} . The region where the spectral radius of
474 \mathbf{A} is less than 1 (i.e., \mathbf{A} is stable) is highlighted in green. One can observe that for
475 different Gaussian kernels, except for some numerical errors, the stability region of
476 \mathbf{A} is the same as that of \mathbf{A}_{11} . In conclusion, we conjecture that for Gaussian filters,
477 the stability condition of \mathbf{A} is also given by (4.22). Numerical experiments seem to
478 indicate that the stability of \mathbf{A} is also controlled by the stability of \mathbf{A}_{11} for a larger
479 class of low-pass filters, including hanning, hamming and rectangular windows. We
480 also found counter examples for which the stability of \mathbf{A}_{11} does not guarantee the
481 stability of \mathbf{A} . It is worth mentioning that the stability region depends on both μ
482 and η_λ and that increasing η_λ necessitates to lower μ to guarantee the stability of the
483 algorithm.

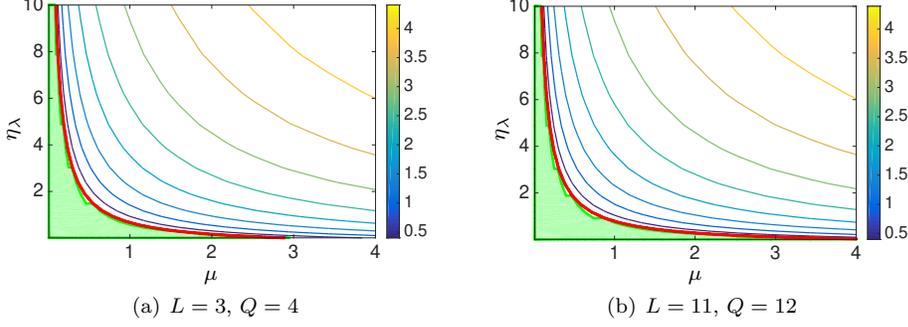


Fig. 4: Spectral radius of \mathbf{A} (in log scale) as a function of μ and η_λ . The green area corresponds to the values of μ and η_λ for which \mathbf{A} is stable. The red curve corresponds to the stability limit of \mathbf{A}_{11} .

484 *Mean-squares stability.* We now address the stability in the mean-squares sense.
 485 First, define the zero-mean misalignment vector \mathbf{w}_k as:

$$486 \quad (4.23) \quad \mathbf{w}_k = \mathbf{v}'_k - \mathbb{E}\{\mathbf{v}'_k\},$$

487 then using (4.11) and (4.13), \mathbf{w}_k can be updated as:

$$488 \quad (4.24) \quad \mathbf{w}_{k+1} = \mathbf{A}\mathbf{w}_k - \mu\Phi\mathbf{z}'_k + \rho_z\mathbf{e}_{z_k} + \rho_s\mathbf{e}_{s_k}$$

489 where

$$490 \quad (4.25) \quad \mathbf{e}_{z_k} = \text{sign}(\hat{\mathbf{x}}'_k) - \mathbb{E}\{\text{sign}(\hat{\mathbf{x}}'_k)\},$$

$$491 \quad (4.26) \quad \mathbf{e}_{s_k} = \Lambda_s^\top (\text{sign}(\Lambda_s\hat{\mathbf{x}}'_k) - \mathbb{E}\{\text{sign}(\Lambda_s\hat{\mathbf{x}}'_k)\}).$$

493 The covariance matrix of \mathbf{w}_k is defined as:

$$494 \quad (4.27) \quad \mathbf{W}_k = \mathbb{E}\{\mathbf{w}_k\mathbf{w}_k^\top\}$$

495 Since $\mathbb{E}\{\mathbf{w}_k\} = \mathbf{0}$, $\mathbb{E}\{\mathbf{e}_{z_k}\} = \mathbf{0}$ and $\mathbb{E}\{\mathbf{e}_{s_k}\} = \mathbf{0}$, \mathbf{W}_k can be updated as follows:

$$496 \quad \mathbf{W}_{k+1} = \mathbf{A}\mathbf{W}_k\mathbf{A}^\top + \mu^2\sigma_z^2\Phi\Phi^\top + \rho_z\mathbf{A}\mathbb{E}\{\mathbf{w}_k\mathbf{e}_{z_k}^\top\} + \rho_z\mathbb{E}\{\mathbf{e}_{z_k}\mathbf{w}_k^\top\}\mathbf{A}^\top + \rho_z^2\mathbb{E}\{\mathbf{e}_{z_k}\mathbf{e}_{z_k}^\top\} \\ 497 \quad + \rho_s\mathbf{A}\mathbb{E}\{\mathbf{w}_k\mathbf{e}_{s_k}^\top\} + \rho_s\mathbb{E}\{\mathbf{e}_{s_k}\mathbf{w}_k^\top\}\mathbf{A}^\top + \rho_s^2\mathbb{E}\{\mathbf{e}_{s_k}\mathbf{e}_{s_k}^\top\} \\ 498 \quad (4.28) \quad + \rho_s\rho_z\mathbb{E}\{\mathbf{e}_{s_k}\mathbf{e}_{z_k}^\top\} + \rho_s\rho_z\mathbb{E}\{\mathbf{e}_{z_k}\mathbf{e}_{s_k}^\top\}$$

500 From (4.8) and (4.23) we have:

$$501 \quad (4.29) \quad \mathbb{E}\{\mathbf{w}_k\mathbf{e}_{z_k}^\top\} = \mathbb{E}\{\hat{\mathbf{x}}'_k\}\mathbb{E}\{\mathbf{e}_{z_k}^\top\} - \mathbb{E}\{\hat{\mathbf{x}}'_k\mathbf{e}_{z_k}^\top\} = -\mathbb{E}\{\hat{\mathbf{x}}'_k\mathbf{e}_{z_k}^\top\},$$

$$502 \quad (4.30) \quad \mathbb{E}\{\mathbf{w}_k\mathbf{e}_{s_k}^\top\} = \mathbb{E}\{\hat{\mathbf{x}}'_k\}\mathbb{E}\{\mathbf{e}_{s_k}^\top\} - \mathbb{E}\{\hat{\mathbf{x}}'_k\mathbf{e}_{s_k}^\top\} = -\mathbb{E}\{\hat{\mathbf{x}}'_k\mathbf{e}_{s_k}^\top\}.$$

504 Thus, we observe:

$$505 \quad \text{trace}(\mathbf{W}_{k+1}) = \text{trace}(\mathbf{A}^\top\mathbf{A}\mathbf{W}_k) + \mu^2\sigma_z^2\text{trace}(\Phi\Phi^\top) + 2\rho_s\rho_z\mathbb{E}\{\mathbf{e}_{s_k}^\top\mathbf{e}_{z_k}\} \\ 506 \quad - 2\rho_z\text{trace}(\mathbf{A}\mathbb{E}\{\hat{\mathbf{x}}_k^\top\mathbf{e}_{z_k}\}) + \rho_z^2\mathbb{E}\{\mathbf{e}_{z_k}^\top\mathbf{e}_{z_k}\} \\ 507 \quad (4.31) \quad - 2\rho_s\text{trace}(\mathbf{A}\mathbb{E}\{\hat{\mathbf{x}}_k^\top\mathbf{e}_{s_k}\}) + \rho_s^2\mathbb{E}\{\mathbf{e}_{s_k}^\top\mathbf{e}_{s_k}\}.$$

509 It is straightforward to verify that \mathbf{e}_{z_k} and \mathbf{e}_{s_k} are bounded. With $\mathbb{E}\{\hat{\mathbf{x}}'_k\} = \bar{\mathbf{x}}' -$
510 $\mathbb{E}\{\mathbf{v}'_k\}$, the terms $\mathbb{E}\{\hat{\mathbf{x}}_k^{\top} \mathbf{e}_{z_k}\}$ and $\mathbb{E}\{\hat{\mathbf{x}}_k^{\top} \mathbf{e}_{s_k}\}$ converge if $\mathbb{E}\{\mathbf{v}'_k\}$ converges which is
511 true if \mathbf{A} is stable and $\bar{\mathbf{x}}'$ is bounded. Thus to have the stability of (4.28), only the
512 stability of its first term on the right hand side has to be studied (see for example
513 [37]). By using the same argument as in [17], the mean-squares stability of (4.28) is
514 guaranteed if the mean stability holds.

515 5. Experimental results.

516 **5.1. Validation of the transient behavior model.** The experiment described
517 below aims at validating the transient behavior model of the SBR-LMS algorithm
518 described in Subsection 4.1. We first address the steady-state case. The simulated
519 image is a 2D image whose lines are constant over time k . Its columns are set to:
520 $\mathbf{x}^{stat.} = [\mathbf{0}_{1 \times 3}, 1, 0.9, \dots, 0.1, \mathbf{0}_{1 \times 3}]^{\top}$. The convolution kernel is a Gaussian filter of
521 size 3×3 . The zero-mean i.i.d. Gaussian noise \mathbf{z}_k is added to reach a 16 dB SNR.
522 The SNR is defined as follows:

$$523 \quad (5.1) \quad \text{SNR} = 10 \log_{10} \frac{\|\mathbf{H}^* * \mathbf{X}\|_F^2}{\|\mathbf{Z}\|_F^2}.$$

524 The block size Q is set to 5, the initial value \mathbf{x}_0 to zero and the step size μ to
525 0.01. Experimental results were obtained by averaging the estimations over 50 runs.
526 The convergence behavior in the mean and mean-squares sense of SBR-LMS with
527 different regularization parameter values are presented in Figure 5. The experimental
528 curves (in blue) and the theoretical curves (4.1)–(4.4) (in red) are superimposed thus
529 confirming the accuracy of the model. It appears that the zero-attracting property
530 results in a faster convergence to zero (Figure 5(c)) than that of the algorithm without
531 any regularization (Figure 5(a)). The first order derivative regularizer (Figure 5(e)) is
532 favoring the reconstruction of piecewise constant objects along the spatial dimension
533 by decreasing the difference between two adjacent rows. However, both zero-attracting
534 and the first order derivative properties introduce a bias on the amplitudes explaining
535 why the MSEs shown in Figure 5(d) and Figure 5(f) are larger than that shown in
536 Figure 5(b).

537 The next experiment intends to show that the proposed model is still valid in the
538 non-stationary case. The results are reported in Figure 6. The n -th pixel at time k
539 of the 2D simulated image is defined as:

$$540 \quad (5.2) \quad x_{nk}^{(nonstat.)} = x_n^{(stat.)} + \frac{|x_n^{(stat.)}|}{10} \sin\left(\frac{2\pi}{T_o}k + 2\pi\frac{n-1}{K}\right)$$

541 with $T_o = 100$. The convolution filter and noise level are the same as before. Q is
542 set to 5, ρ_z and ρ_s are set to $0.2 \cdot 10^{-3}$. The simulated images (in black dots) have
543 to be compared with the superimposed experimental (blue) and theoretical (red)
544 curves. Figure 6(a), Figure 6(b) and Figure 6(c), Figure 6(d) confirm that the step
545 size influences the convergence speed and the estimation variance. A large value for
546 μ results in an increase of the convergence speed and estimation variance. Also, the
547 delay mentioned in Subsection 3.2 and Appendix B can be observed on these figures.
548 Increasing the step size (Figure 6(c)) results in a decreasing of the delay.

549 **5.2. Effects of the parameters.** To assess the influence of the parameters Q ,
550 μ , ρ_s and ρ_z , numerical simulations were conducted on a 2D image. The original
551 image shown in Figure 7(a) was blurred by a Gaussian kernel of size 15×15 with

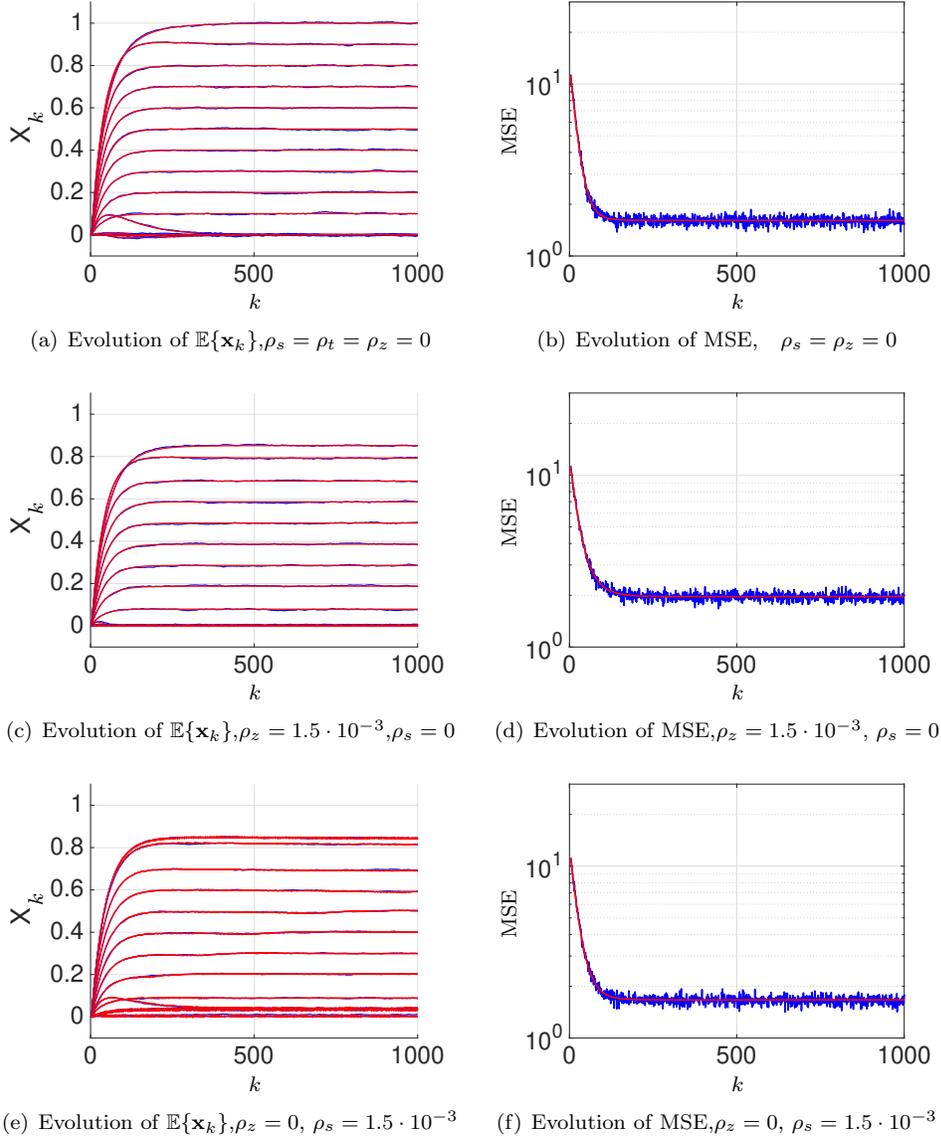


Fig. 5: Transient behavior model in stationary environment

552 full width at half-maximum (FWHM) set to 7 pixels. A Gaussian noise was added to
 553 reach a 5 dB SNR. This blurred and noisy image is shown in [Figure 7\(b\)](#). [Figure 7\(c\)](#)
 554 to [Figure 7\(f\)](#) show respectively the results corresponding to the ZA-LMS algorithm
 555 obtained by setting $Q = 1$, $\mu = 0.06$ and $\rho_z = 0.02$, the SB-LMS algorithm with
 556 hyperparameters $Q = 15$, $\mu = 0.006$, the SBZA-LMS (sliding block LMS with zero-
 557 attracting term only) with hyperparameters $Q = 15$, $\mu = 0.006$ and $\rho_z = 0.05$ and
 558 the SBR-LMS with hyperparameters $Q = 15$, $\mu = 0.006$, $\rho_z = 0.05$ and $\rho_s = 0.01$.
 559 The image restored with ZA-LMS in [Figure 7\(c\)](#) exhibits lower noise level than the

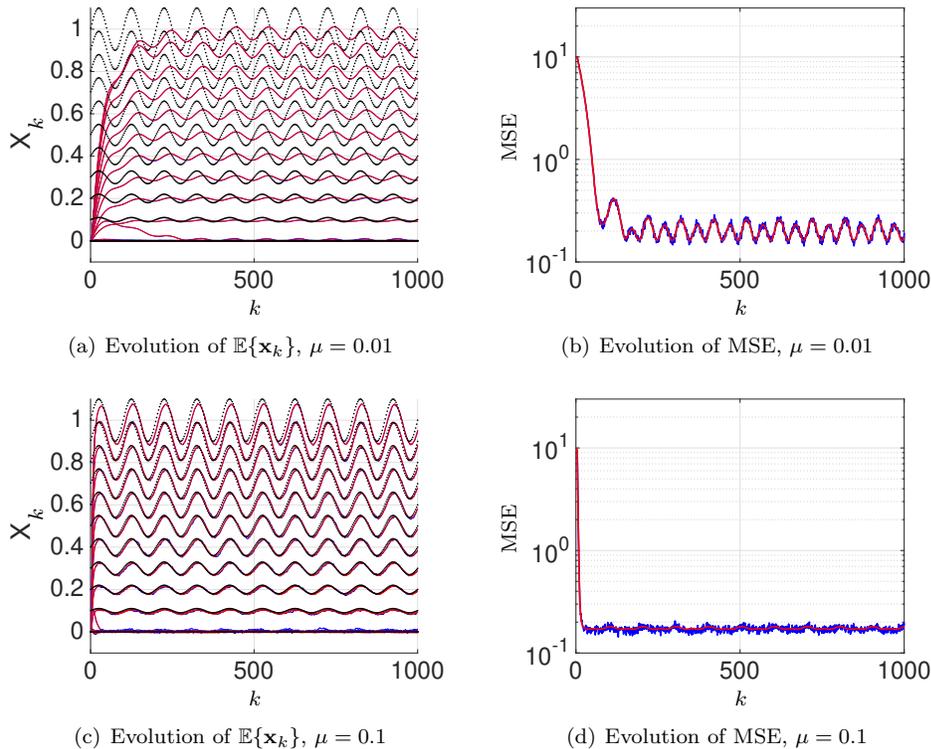


Fig. 6: Transient behavior model validation in non-stationary environment

560 original image but deblurring is very limited. Better results are obtained when the
 561 block size increases as shown in Figure 7(d). The image restored by SBZA-LMS has
 562 a better resolution and a lower noise level as illustrated in Figure 7(e). As mentioned
 563 in Subsection 3.2 and Appendix B, the SB-LMS and SBZA-LMS induce a time (and
 564 no spatial) regularization resulting in horizontal structures visible on the restored
 565 images. This effect is attenuated by introducing a spatial regularization as shown in
 566 Figure 7(f).

567 Figure 8 shows the influence of the spectral regularization parameter η_λ on a
 568 simulated hyperspectral image. Original and blurred noisy images corresponding to
 569 three different wavelengths are shown in Figure 8(a) and Figure 8(b), respectively.
 570 The blurring kernel and the noise level are the same as those in Figure 7(b). By comparing
 571 the estimated results shown in Figure 8(c) (without spectral regularization) and
 572 Figure 8(d) (with spectral regularization), it appears that spectral regularization
 573 can help to recover objects whose amplitude is very small (at a given wavelength)
 574 as pattern 1 in the second band. However, the spectral regularization may also induce
 575 artifacts: for example, pattern 2 appears in the second band when the spectral
 576 regularization is used while it is not present in the original image.

577 We can assess the effect of the parameter Q on the convergence of the SBR-LMS
 578 by examining the transient behavior of the MSE. The simulated image is the same
 579 as the one in Subsection 5.1 (2D image whose lines are constant over time k). It was
 580 blurred by Gaussian filter of size 5×5 with FWHM equal to 3 and the SNR of the

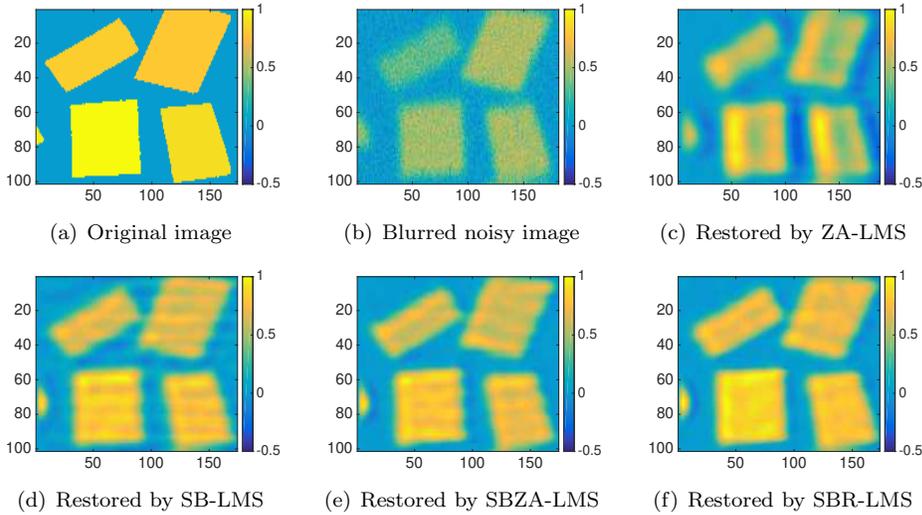


Fig. 7: Influence of different hyperparameters

581 image was set to 40 dB. The initial value \mathbf{x}_0 was set to zero. Experimental MSE were
 582 obtained by averaging over 100 runs. The result comparing the transient behavior
 583 of MSE for different values of Q is shown in Figure 9. Convergence speed using the
 584 SBR-LMS increases when Q increases. Extensive simulations shows that Q should be
 585 chosen slightly larger than L (convolution kernel length along the time dimension).
 586 No improvements are observed when Q is much larger than L .

587 The next experiment aims at showing that the evolution of each parameter (the
 588 others being fixed) w.r.t. the SNR is complies with intuition. Figure 10 shows the
 589 evolution of the MSE as a function of μ , ρ_z , ρ_s and η_λ respectively for different values
 590 of the SNR. For a given SNR, the MSE curve (horizontal lines on Figure 10) decreases
 591 and then increases: the MSE minimum value (red point) corresponds to the best
 592 tradeoff between bias and variance. The values of ρ_z , ρ_s and η_λ minimizing the MSE
 593 increase as the noise level increases. The value of μ minimizing the MSE decreases
 594 as the noise level increases: this is in accordance with the time regularization implicitly
 595 enforced by the LMS (presented in Appendix B) whose strength is controlled by $1/\mu$.
 596

597 Let us now concentrate on the parameter ρ_z which controls the strength of the ZA
 598 terms. In Subsection 5.1, it was mentioned that ZA was favoring a faster convergence
 599 to zero. At the same time it can also slow down the convergence to non-zero values.
 600 To illustrate this point, we consider the following experiment. A simulated image of
 601 size $27 \times 500 \times 4$, shown in Figure 11(a) includes many zero values and some non-zero
 602 values. It was blurred by a Gaussian filter of size 3×3 . Q was set to 5, $\mu = 0.01$,
 603 $\sigma_z^2 = .001$, $\rho_s = \rho_t = \eta_\lambda = 0$, $c_p = 1$. The initial value \mathbf{x}_0 was set to one. The practical
 604 MSE (blue) was obtained by averaging over 10 runs while the theoretical MSE (red)
 605 was evaluated according to (4.4). They are shown in Figure 11(b) for three different
 606 values of ρ_z . The result contains three parts. At first, the algorithm converges from
 607 one to smaller values, and larger ρ_z results in a faster convergence. However, as the
 608 first part of the image contains some non-zero values, the large value of ρ_z introduces

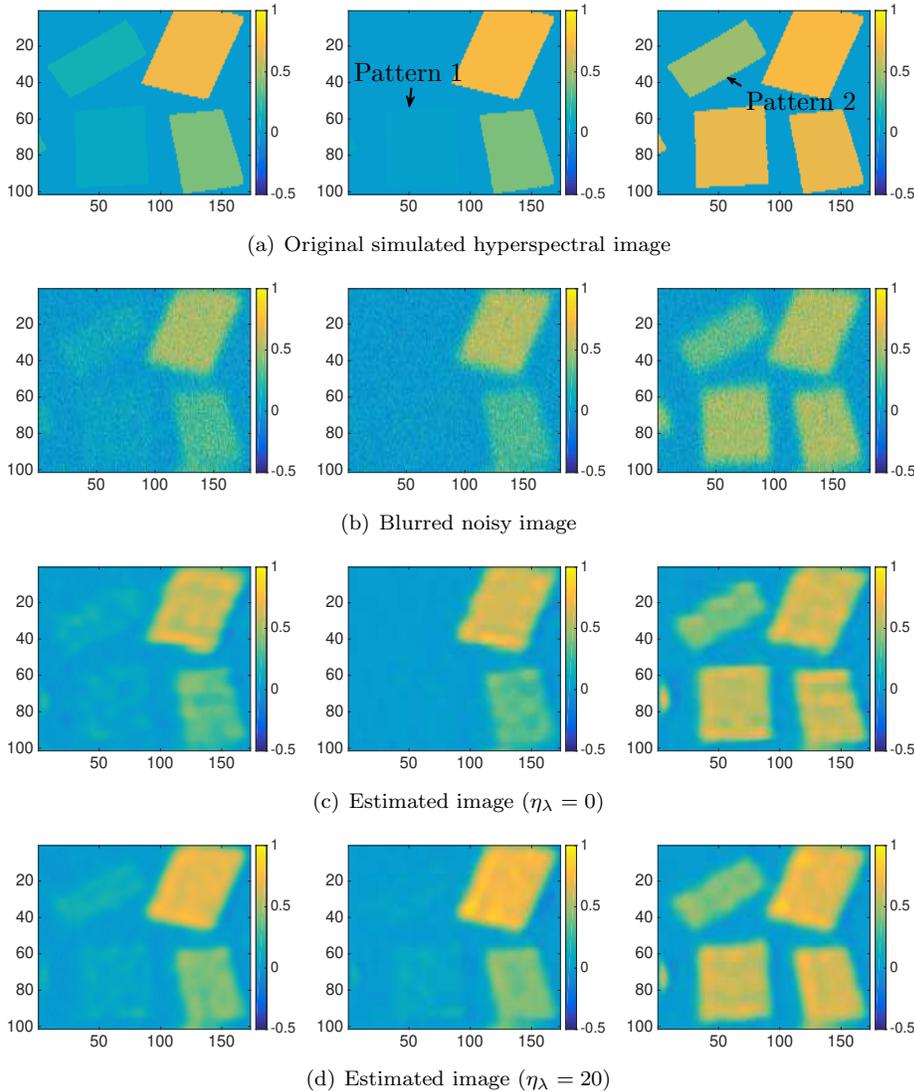


Fig. 8: Influence of spectral regularization parameter

609 a large bias resulting in a large MSE after convergence (around $k = 500$). In the
 610 second part of the image, pixels were all set to zero. The convergence rate of the
 611 algorithm increases with ρ_z . In that case, the zero-attracting does not introduce any
 612 bias and the MSEs after convergence (around $k = 1000$) are the same for all values
 613 of ρ_z . In the third part of the image, the algorithm starts from zero values and has
 614 to converge to non-zero values. Since the zero-attracting algorithm is pulling down
 615 the values to zero, large ρ_z will slow down the convergence rate. We note that the
 616 MSE after convergence (around $k = 1700$) are the same as that of the first part
 617 of the image. This experiment shows that the ZA term introduces an asymmetric
 618 transient behavior: it results in a faster convergence when starting from a non-zero

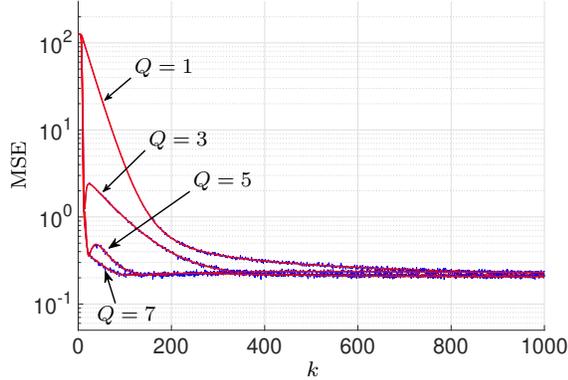


Fig. 9: Evolution to MSE with different values of Q

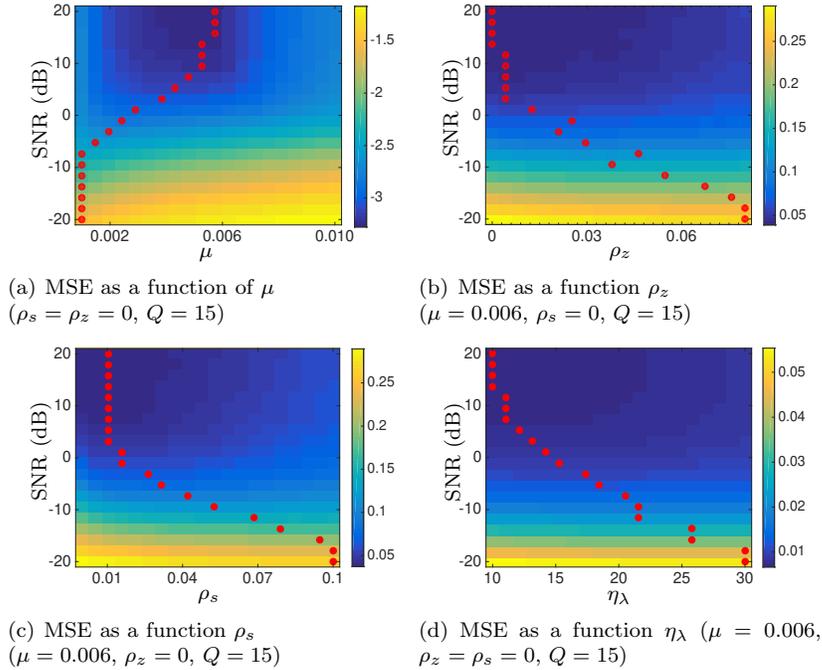


Fig. 10: Evaluation of MSE as a function of parameters for different values of the SNR

619 value toward a zero value and a slower convergence when starting from a zero value
620 toward a non-zero value.

621 **5.3. Performances.** To illustrate the performance of our algorithm, we gen-
622 erated an unblurred hyperspectral image of 261 slices with spatial size 171 and 16
623 spectral bands. The chosen spatial image corresponds to 5 piecewise constant pat-
624 terns put on the background whose response is zero. The spectra of the different

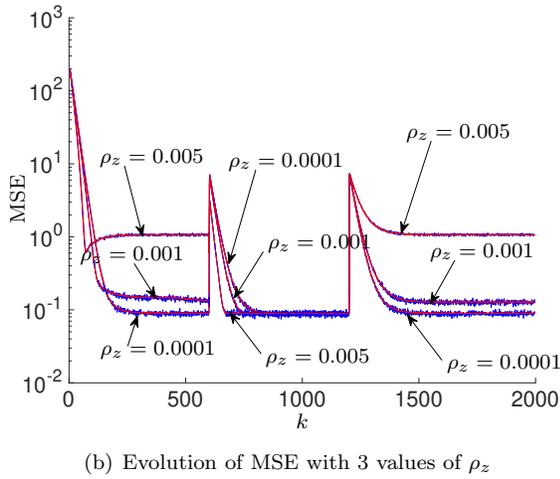
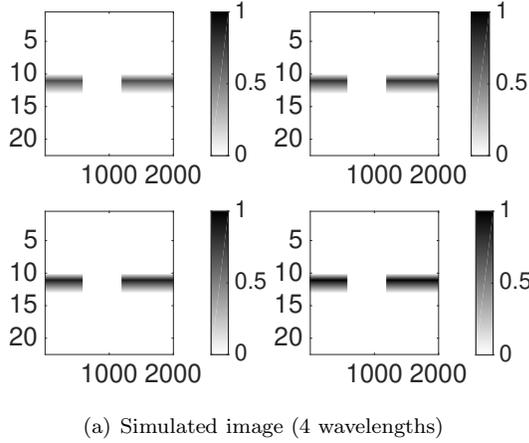


Fig. 11: MSE as function of k for different values of ρ_z

625 objects were chosen to be smooth making the spectral smoothness penalty effective.
 626 The convolution filter \mathbf{H}^* was assumed to be a low-pass Gaussian filter of size (9×9) .
 627 It is invariant with respect to spectral band p . A Gaussian noise was then added to
 628 the blurred image. The noise level was the same for all bands.

629 The SBR-LMS results are compared to an off-line deconvolution algorithm pro-
 630 posed in [27] which corresponds to the Tikhonov approach with ℓ_2 spatial and spec-
 631 tral regularizers. It is also compared to the BT approach presented in [Subsection 3.1](#)
 632 whose block size is the same as the block size of the SBR-LMS. Note that better per-
 633 formance would certainly be achieved by using much more refined methods enforcing
 634 additional constraints (non-negativity, edge preserving) at the price of an increased
 635 computational burden. That is why we decided to restrict the comparison to decon-
 636 volution approaches having a computational burden similar to that of the SBR-LMS.
 637 Three versions of SBR-LMS algorithms are considered for performance evaluation:
 638 SBR-LMS without spatial and zero-attracting regularizations, SBR-LMS with spatial

639 regularization but without zero-attracting regularization, SBR-LMS with spatial and
 640 zero-attracting regularizations.

641 As a performance measure, we consider the MSE as a function of the SNR and
 642 the goal is to assess the performance of the proposed SBR-LMS and to compare them
 643 to those of alternative approaches. As shown in the previous section, the optimal
 644 hyperparameter values are depending on the noise level. Thus, comparing the perfor-
 645 mances of different methods would require to determine for each noise level the set of
 646 optimal parameters by an exhaustive grid search which, due to the large number of
 647 hyperparameters (2 for Tikhonov, 3 for Block Tikonov and up to 5 for SBR-LMS), is
 648 out of reach in reasonable time. To handle this problem, while keeping a fair compar-
 649 ison between the different methods, we propose the following performance evaluation:
 650 the strength of the spectral regularization is fixed to a small value (the same for all
 651 methods) and the block size Q is fixed to 11 (BT, SBR-LMS). All the other param-
 652 eters are fixed by a grid search yielding the minimum MSE for a fixed SNR= 12 dB.
 653 These parameter values are then used for all other noise levels. Figure 12 shows that
 654 SBR-LMS without spatial and zero-attracting regularization terms yields the highest
 655 MSE for all noise level. This can be attributed to the fact that, in that case, only a
 656 time regularization is used. For high SNR, Tikhonov and BT give the lowest MSEs.
 657 however, when SNR is lower than 12 dB, SBR-LMS with spatial regularization has
 658 almost the same performance as the Tikhonov approach and better performance than
 659 BT algorithm. Adding a zero-attracting regularization improves the performance at
 660 low SNR even more. In conclusion, if we focus on sequential deconvolution, BT has
 to be chosen if the SNR is high while SBR-LMS should be preferred for low SNR.

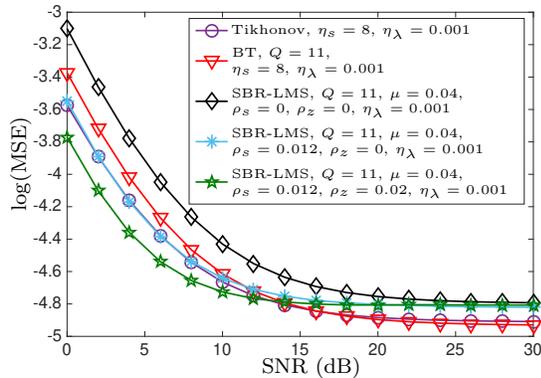


Fig. 12: Performances of the hyperspectral image deconvolution algorithms

661

662 **5.4. Real hyperspectral image deblurring.** The last experiment aims at
 663 illustrating the performance of the SBR-LMS algorithm on real blurred hyperspectral
 664 images of size $481 \times 656 \times 28$ (spatial, time and spectral sizes respectively) with
 665 wavelengths varying from 947.71 nm to 1707.7 nm with an increment of 27.68 nm.
 666 The spectral response of the conveyor (background) was estimated from data in an
 667 area of size $140 \times 140 \times 28$. It was then subtracted from each pixel of the hyperspectral
 668 image. The imaged objects put on the conveyor are pieces of woods (wood wastes).
 669 The convolution filter was estimated from data (using a calibration target) to be
 670 a Gaussian kernel of size 7×7 with FWHM of 3 points. No dependence of the

671 convolution kernel w.r.t. the wavelength was observed.

672 First, we compare different algorithms in the case of an image acquired using a
673 large integration time (2.146 ms) resulting in a high SNR. This corresponds to the
674 situation where Tikhonov and BT are expected to yield the best results. **Figure 13**
675 presents the deconvolution result obtained on the whole real hyperspectral image
676 (only 3 slices corresponding to wavelengths 1058.64 nm, 1198.12 nm and 1479.78 nm
677 are shown). **Figure 14** is a part selected from the whole image. The original images
678 are shown in **Figure 13(a)** and **Figure 14(a)**. The images restored with SBR-LMS
679 ($\mu = 2.3$, $Q = 9$, $\rho_z = 0.005$, $\rho_s = 0.01$, $\eta_\lambda = 0.001$) are shown in **Figure 13(b)** and
680 **Figure 14(b)**. **Figure 13(c)** and **Figure 14(c)** correspond to the Tikhonov approach
681 with ℓ_2 spatial and spectral regularizers. **Figure 13(d)** and **Figure 14(d)** correspond to
682 the BT approach. The hyperparameters of the Tikhonov approaches were estimated
683 by the minimum distance criterion (MDC) proposed in [38]. The FWHB of the kernel
684 being equal to 3 points, the blurring of the real images is quite limited. However, an
685 increased resolution allowing to recover the fine texture of the objects can be observed
686 on the zoomed view restored images in **Figure 14(b)**-**Figure 14(d)**. More importantly,
687 **Figure 13** and **Figure 14** show that at high SNR, the SBR-LMS yields similar
688 performances as those of the reference methods (Tikhonov and BT).

689 The second image in **Figure 15(a)** corresponds to the same observed scene with
690 an integration time of 0.013 ms which results in a low SNR. The image restored with
691 SBR-LMS ($\mu = 0.5$, $Q = 9$, $\rho_z = 0.1$, $\rho_s = 0.05$, $\eta_\lambda = 0.001$) is shown in **Figure 15(b)**.
692 Results obtained by Tikhonov and BT methods, whose hyperparameters are estimated
693 by the MDC approach, are shown in **Figure 15(b)** and **Figure 15(c)**, respectively. The
694 bias introduced by the regularization terms can be observed in **Figure 15(b)** since
695 the dynamic range is lower than that of Tikhonov methods. It clearly appears that
696 the noise level on the background estimated by SBR-LMS is much lower than that
697 of the Tikhonov approaches. Due to the noise level, all approaches cannot reveal fine
698 structures accurately, but the edges are well preserved.

699 **6. Conclusions.** In this work, the online deconvolution problem of hyperspec-
700 tral images collected by industrial pushbroom imaging systems was addressed. The
701 contribution of this work is the derivation of the SBR-LMS algorithm that allows the
702 fast slice-by-slice hyperspectral image restoration, accounting for convolution kernel
703 non-causality and low SNR issues. The key feature of the proposed approach relies on
704 the possibility to include non-quadratic regularization terms while maintaining a com-
705 putational complexity linear w.r.t. the number of unknowns. The transient behavior
706 model of the algorithm was analyzed; it allows to assess the influence of each regu-
707 larization parameter. Experimental results on both simulated and real hyperspectral
708 images showed that the SBR-LMS algorithm has good noise removal and deblurring
709 capabilities, especially at low SNR which is the relevant case for industrial imaging
710 systems.

711 The effective implementation of such method is hampered by the choice of the
712 regularization parameters. In general, this choice is made by successive trials which
713 can be highly time consuming. Future works will focus on the automatic learning
714 of hyperparameters. A first idea could be to learn the optimal parameters in a way
715 similar to what is done in [38] on a typical sample and to use these fixed parameters
716 for the online processing. A more promising approach would be to perform the online
717 learning of these parameters: for example, [32], a variable-parameter algorithm is
718 proposed to adjust optimally the algorithm parameters of ZA-LMS. This will avoid
719 the necessity of having representative image samples and will meet the requirements

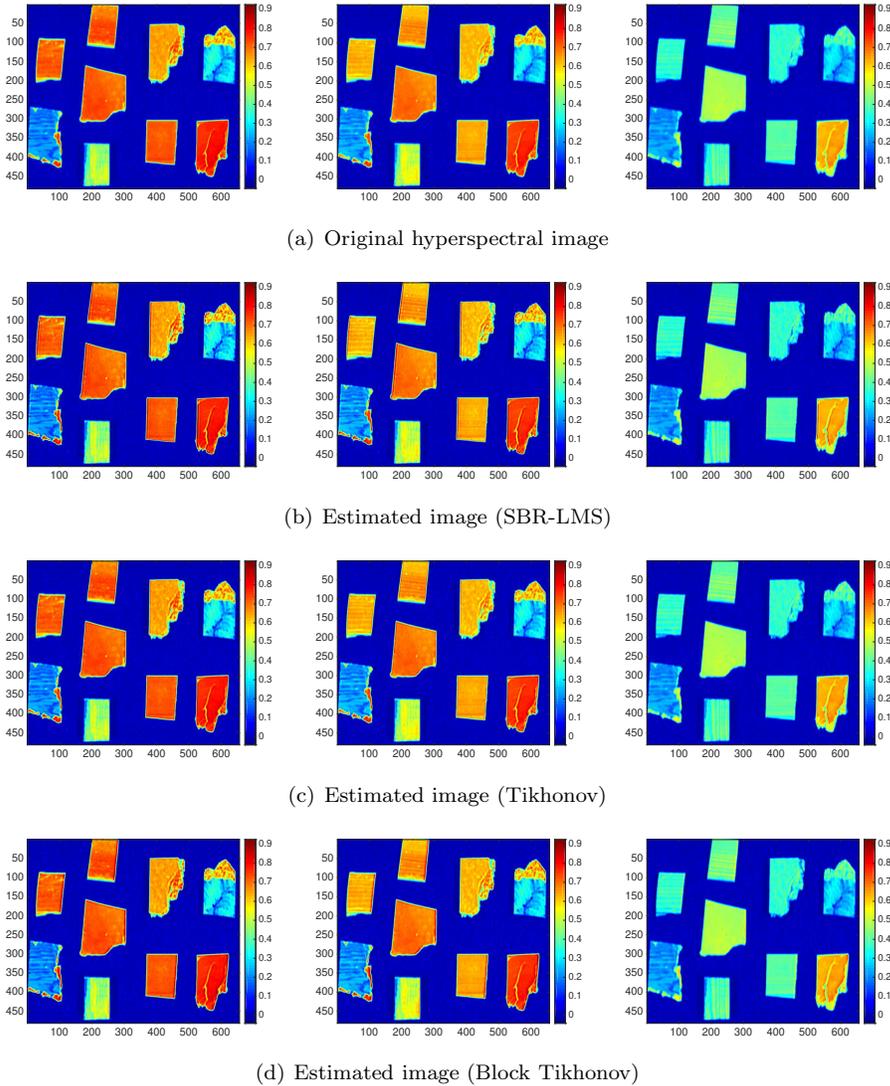


Fig. 13: hyperspectral image restoration at 3 wavelengths.

720 of a full online processing. To circumvent the asymmetric transient behavior induced
 721 by the ℓ_1 norm zero-attracting regularization, an ℓ_0 -like regularizer such as introduced
 722 in [17] could be considered. For different types of materials and/or spectroscopies,
 723 other spectral regularizations could be used such as TV-like spectral regularization
 724 proposed in [1] for promoting piecewise constant spectra reconstruction. Finally,
 725 accounting for the low-rank structure of the data to restore (which results from a
 726 non-negative linear mixing model of the data to restore), a joint online deconvolution
 727 and unmixing algorithm is worth being studied. This is expected to yield a very low
 728 computational burden and accurate image restoration approach.

729 **Appendix A. Structure of matrix \mathbf{G} for both cases $Q \geq L$ and $Q < L$.**

735 written as:

$$736 \quad (\text{A.3}) \quad \mathbf{G} \triangleq \left[\begin{array}{c|c} \Phi_{11} & \mathbf{G}_{12} \\ \hline \mathbf{0}_{(L-1)PN \times (Q+L-1)PN} & \end{array} \right].$$

737 **Appendix B. SB-LMS interpreted as a time-regularized Tikhonov-like**
 738 **estimator.** This appendix reveals that there is a hidden time-regularization intro-
 739 duced by the SB-LMS (that is the SBR-LMS where $\rho_s = \rho_z = \eta_\lambda = 0$) whose strength
 740 is controlled by the step size μ . Setting the spatial and spectral regularization terms
 741 to zero in (3.16), the SB-LMS is defined as:

$$742 \quad (\text{B.1}) \quad \hat{\mathbf{x}}'_{k+1} = \Omega \hat{\mathbf{x}}'_k + \mu \Phi (\mathbf{y}'_k - \mathbf{G} \hat{\mathbf{x}}'_k)$$

744 Denoting $\hat{\mathbf{X}}'(z)$ as the z -transform of $\hat{\mathbf{x}}'_k$, Equation (B.1) can be expressed as:

$$745 \quad (\text{B.2}) \quad \hat{\mathbf{X}}'(z) = \mu (z\mathbf{I} - \Omega + \mu \Phi \mathbf{G})^{-1} \Phi \mathbf{Y}'(z)$$

746 In another word, $\hat{\mathbf{X}}'(z)$ is the output of a linear filter:

$$747 \quad (\text{B.3}) \quad \mathbf{F}(z) = \mu (z\mathbf{I} - \Omega + \mu \Phi \mathbf{G})^{-1} \Phi$$

748 with input $\mathbf{Y}'(z)$. It is well known that Tikhonov estimators result in a linear filtering
 749 of the data. Thus, the question we are addressing is to link these two linear filters.
 750 To simplify the analysis, we consider the denoising problem for which the convolution
 751 kernel is set to be identity. In that case, the block length Q is set to 1 (which results
 752 in $\hat{\mathbf{x}}'_k = \hat{\mathbf{x}}_k$) and (B.2) can be simplified as:

$$753 \quad (\text{B.4}) \quad \hat{\mathbf{X}}(z) = \frac{\mu}{z - 1 + \mu} \mathbf{Y}(z)$$

754 Assuming a unit sampling step size, the forward transform is given by $s = z - 1$ where
 755 s is the Laplace parameter. The Fourier transform on the filter $F(z) = \frac{\mu}{z - 1 + \mu}$ is
 756 obtained by setting $s = j\omega$ where ω is the angular velocity:

$$757 \quad (\text{B.5}) \quad F(\omega) = \frac{1}{1 + j\omega/\mu} = \frac{1}{(1 + \omega^2/\mu^2)^{0.5}} \exp^{-j \arctan(\omega/\mu)}$$

758 Combining the first-order approximation of $\arctan(\omega/\mu) \approx \omega/\mu$ together with the
 759 series expansion of $\sqrt{1 + x^2}$ can be used to give a low-pass approximation of the filter:

$$760 \quad (\text{B.6}) \quad F(\omega) \approx \frac{1}{1 + \sum_{i=1}^{\infty} \beta_i(\mu) \omega^{2i}} \exp^{-j \omega/\mu}$$

761 where $\beta_i(\mu) = \frac{(2i)!}{(2\mu)^{2i} (i!)^2}$. Following [34], any filter of the form (B.6) results from to
 762 the minimization of a criterion:

$$763 \quad (\text{B.7}) \quad \mathcal{J}(\mathbf{x}_{k+k_0}, \mathbf{y}_k) = \|\mathbf{y}_k - \mathbf{x}_{k+k_0}\|^2 + \sum_{i=1}^{\infty} \beta_i(\mu) \|\mathbf{D}_i \mathbf{x}_{k+k_0}\|^2$$

765 where \mathbf{D}_i is the i th-order (discrete) derivative matrix along the time dimension and
 766 $k_0 = 1/\mu$. This means the LMS-based algorithms can be interpreted as delayed

767 Tikhonov-like algorithms. The first point to mention is that LMS-based estimators
 768 present a time delay in the estimation which is (approximately) proportional to $1/\mu$.
 769 The smaller μ is, the larger delay is. This has to be opposed to Tikhonov estimators
 770 which are null-phase filters and do not introduce any delay. The second point is
 771 related to the regularization parameters $\beta_i(\mu)$ which is proportional to $1/(2\mu)^{2i}$. A
 772 small value μ results in strong regularization strength along the time dimension.

773

REFERENCES

- 774 [1] H. K. AGGARWAL AND A. MAJUMDAR, *Hyperspectral image denoising using spatio-spectral total*
 775 *variation*, IEEE Geoscience and Remote Sensing Letters, 13 (2016), pp. 442–446.
- 776 [2] R. AMMANOUIL, A. FERRARI, C. RICHARD, AND D. MARY, *Blind and fully constrained unmixing*
 777 *of hyperspectral images*, IEEE Transactions on Image Processing, 23 (2014), pp. 5510–5518.
- 778 [3] A. BECK AND M. TEBoulLE, *A fast iterative shrinkage-thresholding algorithm for linear inverse*
 779 *problems*, SIAM journal on imaging sciences, 2 (2009), pp. 183–202.
- 780 [4] J. BIEMOND, J. RIESKE, AND J. GERBRANDS, *A fast Kalman filter for images degraded by both*
 781 *blur and noise*, IEEE Transactions on Acoustics, Speech, and Signal Processing, 31 (1983),
 782 pp. 1248–1256.
- 783 [5] S. BOYD, N. PARIKH, E. CHU, B. PELEATO, J. ECKSTEIN, ET AL., *Distributed optimization and*
 784 *statistical learning via the alternating direction method of multipliers*, Foundations and
 785 Trends® in Machine learning, 3 (2011), pp. 1–122.
- 786 [6] Z. M. BUCHEVATS AND L. T. GRUYITCH, *Linear Discrete-Time Systems*, CRC Press, 2017.
- 787 [7] C. CARIOU AND K. CHEHDI, *Automatic georeferencing of airborne pushbroom scanner images*
 788 *with missing ancillary data using mutual information*, IEEE Transactions on Geoscience
 789 and Remote Sensing, 46 (2008), pp. 1290–1300.
- 790 [8] P. CHARBONNIER, L. BLANC-FÉRAUD, G. AUBERT, AND M. BARLAUD, *Deterministic edge-*
 791 *preserving regularization in computed imaging*, IEEE Transactions on image processing, 6
 792 (1997), pp. 298–311.
- 793 [9] J. CHEN, C. RICHARD, J.-C. M. BERMUDEZ, AND P. HONEINE, *Nonnegative least-mean-square*
 794 *algorithm*, IEEE Transactions on Signal Processing, 59 (2011), pp. 5225–5235.
- 795 [10] J. CHEN, C. RICHARD, AND P. HONEINE, *Nonlinear unmixing of hyperspectral data based on*
 796 *a linear-mixture/nonlinear-fluctuation model*, IEEE Transactions on Signal Processing, 61
 797 (2013), pp. 480–492.
- 798 [11] J. CHEN, C. RICHARD, Y. SONG, AND D. BRIE, *Transient performance analysis of zero-*
 799 *attracting lms*, IEEE Signal Processing Letters, 23 (2016), pp. 1786–1790.
- 800 [12] Y. CHEN, Y. GU, AND A. O. HERO, *Regularized least-mean-square algorithms*, arXiv preprint
 801 arXiv:1012.5066, (2010).
- 802 [13] Y. R. CHEN, R. W. HUFFMAN, B. PARK, AND M. NGUYEN, *Transportable spectrophotometer*
 803 *system for on-line classification of poultry carcasses*, Applied Spectroscopy, 50 (1996),
 804 pp. 910–916.
- 805 [14] E. CHOUZENOUX AND J.-C. PESQUET, *A stochastic majorize-minimize subspace algorithm for*
 806 *online penalized least squares estimation*, IEEE Transactions on Signal Processing, 65
 807 (2017), pp. 4770–4783.
- 808 [15] P. L. COMBETTES AND J.-C. PESQUET, *Stochastic forward-backward and primal-dual approxi-*
 809 *mation algorithms with application to online image restoration*, in 24th European Signal
 810 Processing Conference (EUSIPCO), 2016, pp. 1813–1817.
- 811 [16] G. H. COSTA AND J.-C. M. BERMUDEZ, *Statistical analysis of the LMS algorithm applied to*
 812 *super-resolution image reconstruction*, IEEE Transactions on Signal Processing, 55 (2007),
 813 pp. 2084–2095.
- 814 [17] P. DI LORENZO AND A. H. SAYED, *Sparse distributed learning based on diffusion adaptation*,
 815 IEEE Transactions on signal processing, 61 (2013), pp. 1419–1433.
- 816 [18] E. M. EKSIÖGLU, *Sparsity regularised recursive least squares adaptive filtering*, IET signal
 817 processing, 5 (2011), pp. 480–487.
- 818 [19] M. ELAD AND A. FEUER, *Superresolution restoration of an image sequence: adaptive filtering*
 819 *approach*, IEEE Transactions on Image Processing, 8 (1999), pp. 387–395.
- 820 [20] F. W. ESMONDE-WHITE, K. A. ESMONDE-WHITE, AND M. D. MORRIS, *Minor distortions with*
 821 *major consequences: correcting distortions in imaging spectrographs*, Applied spectroscopy,
 822 65 (2011), pp. 85–98.
- 823 [21] N. P. GALATSANOS AND R. T. CHIN, *Restoration of color images by multichannel Kalman*
 824 *filtering*, IEEE Transactions on signal processing, 39 (1991), pp. 2237–2252.

- 825 [22] N. P. GALATSANOS, A. K. KATSAGGELOS, R. T. CHIN, AND A. D. HILLERY, *Digital restoration*
826 *of multichannel images*, IEEE Transactions on Acoustics, Speech, and Signal Processing,
827 37 (1989), pp. 415–421.
- 828 [23] N. P. GALATSANOS, A. K. KATSAGGELOS, R. T. CHIN, AND A. D. HILLERY, *Least squares*
829 *restoration of multichannel images*, IEEE Transactions on Signal Processing, 39 (1991),
830 pp. 2222–2236.
- 831 [24] W. GAO, J. CHEN, C. RICHARD, AND J. HUANG, *Online dictionary learning for kernel LMS*,
832 IEEE Transactions on Signal Processing, 62 (2014), pp. 2765–2777.
- 833 [25] J. F. GIOVANNELLI AND A. COULAIS, *Positive deconvolution for superimposed extended source*
834 *and point sources*, Astronomy & Astrophysics, 439 (2005), pp. 401–412.
- 835 [26] J. GOODMAN, *Introduction to Fourier optics*, McGraw-hill, 2008.
- 836 [27] S. HENROT, C. SOUSSEN, AND D. BRIE, *Fast positive deconvolution of hyperspectral images*,
837 IEEE Transactions on Image Processing, 22 (2013), pp. 828–833.
- 838 [28] H. HUANG, L. LIU, AND M. O. NGADI, *Recent developments in hyperspectral imaging for as-*
839 *essment of food quality and safety*, Sensors, 14 (2014), pp. 7248–7276.
- 840 [29] B. R. HUNT AND O. KUBLER, *Karhunen-Loeve multispectral image restoration, part I: Theory*,
841 IEEE Transactions on Acoustics, Speech, and Signal Processing, 32 (1984), pp. 592–600.
- 842 [30] J. JEMEC, F. PERNUŠ, B. LIKAR, AND M. BÜRMEIN, *Push-broom hyperspectral image calibration*
843 *and enhancement by 2D deconvolution with a variant response function estimate*, Optics
844 Express, 22 (2014), pp. 27655–27668.
- 845 [31] J. JEMEC, F. PERNUŠ, B. LIKAR, AND M. BÜRMEIN, *Deconvolution-based restoration of SWIR*
846 *pushbroom imaging spectrometer images*, Optics Express, 24 (2016), pp. 24704–24718.
- 847 [32] D. JIN, J. CHEN, C. RICHARD, AND J. CHEN, *Adaptive parameters adjustment for group*
848 *reweighted zero-attracting*, in 43th International Conference on Acoustics, Speech, and
849 Signal Processing, (ICASSP), 2018.
- 850 [33] Q. LI, X. HE, Y. WANG, H. LIU, D. XU, AND F. GUO, *Review of spectral imaging technology*
851 *in biomedical engineering: achievements and challenges*, Journal of biomedical optics, 18
852 (2013), pp. 100901–100901.
- 853 [34] S. MOUSSAOUI, D. BRIE, AND A. RICHARD, *Regularization aspects in continuous-time model*
854 *identification*, Automatica, 41 (2005), pp. 197–208.
- 855 [35] PELLENC SELECTIVE TECHNOLOGY, *Mistral product*. <http://www.pellencst.com/products>. Ac-
856 cessed: 2016-09-01.
- 857 [36] M. E. SCHAEPMAN, *Imaging spectrometers*, The SAGE handbook of remote sensing, (2009),
858 pp. 166–178.
- 859 [37] K. SHI AND P. SHI, *Convergence analysis of sparse lms algorithms with l1-norm penalty based*
860 *on white input signal*, Signal Processing, 90 (2010), pp. 3289–3293.
- 861 [38] Y. SONG, D. BRIE, E.-H. DJERMOUNE, AND S. HENROT, *Regularization parameter estimation*
862 *for non-negative hyperspectral image deconvolution*, IEEE Transactions on Image Process-
863 ing, 25 (2016), pp. 5316–5330.
- 864 [39] Y. SONG, E.-H. DJERMOUNE, J. CHEN, C. RICHARD, AND D. BRIE, *Online deconvolution for*
865 *pushbroom hyperspectral imaging systems*, in 7th IEEE International Workshop on Com-
866 putational Advances in Multi-Sensor Adaptive Processing, CAMSAP 2017, 2017.
- 867 [40] T. SUZUKI, *Dual averaging and proximal gradient descent for online alternating direction mul-*
868 *tiplier method*, in International Conference on Machine Learning, 2013, pp. 392–400.
- 869 [41] D. TAO, G. JIA, Y. YUAN, AND H. ZHAO, *A digital sensor simulator of the pushbroom offner*
870 *hyperspectral imaging spectrometer*, Sensors, 14 (2014), pp. 23822–23842.
- 871 [42] P. TATZER, M. WOLF, AND T. PANNER, *Industrial application for inline material sorting using*
872 *hyperspectral imaging in the NIR range*, Real-Time Imaging, 11 (2005), pp. 99–107.
- 873 [43] A. M. TEKALP AND G. PAVLOVIĆ, *Multichannel image modeling and kalman filtering for mul-*
874 *tispectral image restoration*, Signal Processing, 19 (1990), pp. 221–232.
- 875 [44] S. TSUCHIKAWA AND H. KOBORI, *A review of recent application of near infrared spectroscopy*
876 *to wood science and technology*, Journal of Wood Science, 61 (2015), pp. 213–220.
- 877 [45] B. WIDROW AND S. D. STEARNS, *Adaptive signal processing*, Englewood Cliffs, NJ, Prentice-
878 Hall, Inc., 1985.
- 879 [46] R. M. WILLET, M. F. DUARTE, M. A. DAVENPORT, AND R. G. BARANIUK, *Sparsity and*
880 *structure in hyperspectral imaging: sensing, reconstruction, and target detection*, IEEE
881 Signal Processing Magazine, 31 (2014), pp. 116–126.
- 882 [47] J. WOODS AND V. INGLE, *Kalman filtering in two dimensions: Further results*, IEEE Transac-
883 tions on Acoustics, Speech, and Signal Processing, 29 (1981), pp. 188–197.
- 884 [48] J. WOODS AND C. RADEWAN, *Kalman filtering in two dimensions*, IEEE Transactions on In-
885 formation Theory, 23 (1977), pp. 473–482.
- 886 [49] S. ZHANG AND J. ZHANG, *Transient analysis of zero attracting NLMS algorithm without gaus-*

- 887 *sian inputs assumption*, Signal processing, 97 (2014), pp. 100–109.
888 [50] X. ZHAO, F. WANG, T. HUANG, M. K. NG, AND R. J. PLEMMONS, *Deblurring and sparse*
889 *unmixing for hyperspectral images*, IEEE Transactions on Geoscience and Remote Sensing,
890 51 (2013), pp. 4045–4058.