Fast Multispectral and Hyperspectral Image Fusion via Hessian Inversion

Dan Pineau [1,2,3], François Orieux [1,2] Alain Abergel [2,3],

***Abstract*—Multispectral and hyperspectral data fusion allows data restoration with increased spatial and spectral resolutions. A common approach, in particular with the presence of variant blur, is to solve an ill-posed inverse problem by minimizing a mixed criterion. This minimization usually requires an iterative gradient-based method. Still, this paper demonstrates the existence of a reachable explicit solution without the need to solve a Sylvester equation, allowing for a three order of magnitude time-saving factor wrt. iterative algorithms. We exploit this explicit resolution to accelerate the minimization by a Majorize-Minimize algorithm of a semi-quadratic convex edge-preserving criterion. We conduct experiments on realistic synthetic measurements for the James Webb Space Telescope and show that our proposed solution outperforms the state-of-the-art in computation time and quality of reconstruction.**

***Index Terms*—Inverse problems, data fusion, MM algorithm, pansharpening, multispectral, hyperspectral, variant blur, deconvolution, close form, infrared, JWST.**

I. Introduction

MANY fields, such as remote sensing, astrophysics, or Earth observation, use hyperspectral data for spectra observation. Several applications allow the use of integral field spectrographs to acquire spectra in a given field of view, but the images produced are generally poorly sampled. Parallel acquisition of high spatial resolution and well-sampled data, such as multispectral data, opens the field of data fusion to reconstruct highly spatially and spectrally resolved data [1].

Usually, multispectral (MS) and hyperspectral (HS) datasets are represented as datacubes, with two spatial dimensions and one spectral dimension, that have been spatially or spectrally degraded by the instruments' effects like transmission of optics, filters, sampling of the detector, etc. Therefore, the MS/HS fusion problem consists of exploiting the information provided by these two degraded datasets to reconstruct a hyperspectral datacube corrected from instrumental effects.

Different approaches were explored in the literature to solve the MS/HS fusion problem like matrix factorization and tensor-based approaches [2]–[6], deep neural networks approaches [7]–[10], or bayesian based approaches [11], [12].

Nonetheless, some cases of MS/HS fusion problems can require using computationally burdening algorithms, especially when fusing high-dimensional datasets. The development of computationally efficient solutions to this fusion problem was attempted in the literature. In Wei *et al.* [13], the MS/HS data fusion problem is formulated through a Sylvester equation, and an explicit solution is obtained, allowing a significant reduction of the computational cost, unlike iterative approaches. Methods based on Tucker decompositions proposed by Prévost *et. al.* [5] and Borsoi *et. al.* [6] also provide explicit solutions to the MS/HS fusion problem.

However, the problem tackled in this paper involves the fusion of datasets where instrumental effects include significant and spectrally varying spatial blurs, acquired on a relatively wide mid-infrared waveband for instance. The previously presented methods [13] or [6] did not take into account such spectrally dependent effects, making them inadequate for the problem of this paper.

In contrast, Guilloteau *et. al.* [14] proposes the solution of an inverse problem with a spectrally non-stationary spatial blur by minimizing a least squares criterion including data adequacy terms and a quadratic regularization. However, the minimization of this criterion required using an iterative algorithm instead of an explicit solution.

This paper approaches the MS/HS fusion problem with variational approaches, as done in [14]. The first contribution is the demonstration that a close form and explicit solution of a regularized least square is reachable for the same inverse problem [14], instead of the previously proposed iterative approach, and without involving the resolution of a Sylvester equation like in [13]. This proposed procedure allows for a significant time-saving factor of up to 7 000 compared to commonly used iterative approaches. A second contribution is the development of a fast convex edge-preserving solution for the MS/HS fusion problem. This method uses a semi-quadratic regularization [15] and a Majorize-Minimize algorithm [16] to improve the quality of the reconstruction of objects featuring high spatial gradients in comparison to [14] (NRMSE from 27×10^{-3} to 22×10^{-3}). These advances exploit a new efficient way to calculate, store, and apply the inverse of a Hessian matrix by exploiting its structure of diagonal blocks, later described in section III.

Section II describes the model used to represent the object observed by the instruments and the data formation processes or forward models: one for an imaging instrument, and another for an integral field spectrograph. Section III formulates the problem and details solutions for two different regularizations: quadratic and convex differentiable. Finally, section IV presents an application of our method on simulated astronomical data acquired by instruments similar to MIRI, the mid-infrared instrument of the James Webb Space Telescope (JWST). The results are presented in terms of reconstruction

[1] Université Paris-Saclay, CNRS, CentraleSupélec, Laboratoire des Signaux et Systèmes, 3 rue Joliot Curie, 91190 Gif-sur-Yvette, France
[2] Institut d'Astrophysique Spatiale (IAS), CNRS, UMR 8617, Université Paris-Saclay, 91405 Orsay, France
[3] Innovative Common Laboratory For Space Spectroscopy (INCLASS), IAS, ACRI-ST, CNRS, Université Paris-Saclay, 91405 Orsay, France

quality and time savings against iterative approaches and deep neural network methods from the literature.

II. Models description

A. Observed object model

The observed sky region is described as a discrete hyperspectral cube $\boldsymbol{x}$ with two spatial coordinates $i \in [1, I]$ and $j \in [1, J]$, and a spectral coordinate $l \in [1, L]$. As proposed in [14], [15], [17], spectral correlations are introduced within the observed object $\boldsymbol{x}$, such that the spectral content of any location $(i,\ j)$ is a linear combination of T known spectra $\boldsymbol{s}_t$. The observed object $\boldsymbol{x}$ thus writes as

$$\boldsymbol{x}[i,j,l] = \sum_{t=1}^{T} \boldsymbol{a}_t[i,j]\ \boldsymbol{s}_t[l] \tag{1}$$

where $\boldsymbol{a}_t[i,j]$ gives the abundance of the spectrum $\boldsymbol{s}_t$ at $(i,\ j)$. This assumption of a Linear Mixing Model offers several advantages. As the spectra $\boldsymbol{s}_t$ can be defined for all common wavelengths covered by the instruments, they enable efficient joint processing of all available data. Therefore, the number of unknowns describing the object becomes $P \times T$ instead of $P \times L$, where $P = I \times J$ is the number of pixels in an image. Thus, choosing $T \ll L$ significantly reduces the number of unknowns. Eq. 1 can be rewritten in a matrix form as

$$\boldsymbol{x} = \boldsymbol{T}\boldsymbol{a} \tag{2}$$

where $\boldsymbol{x} \in \mathbb{R}^{LP \times P}$ is the object with stacked monochromatic images, $\boldsymbol{T} \in \mathbb{R}^{LP \times TP}$ and $\boldsymbol{a} \in \mathbb{R}^{TP \times P}$ the stacked abundance images. The linear mixing matrix $\boldsymbol{T}$ is written as

$$\boldsymbol{T} = \begin{pmatrix} \boldsymbol{s}_1[1]\boldsymbol{I}_P & \dots & \boldsymbol{s}_T[1]\boldsymbol{I}_P \\ \vdots & & \vdots \\ \boldsymbol{s}_1[L]\boldsymbol{I}_P & \dots & \boldsymbol{s}_T[L]\boldsymbol{I}_P \end{pmatrix} \tag{3}$$

with $\boldsymbol{I}_P$ the identity matrix.

B. Imager model

Imagers contain several components, such as mirrors to focus light, filters to select the waveband to observe, and a detector to sample the light spatially. Due to the diffraction on the finite aperture of the instrument, the spectrally varying response blurs the images. Moreover, fields of view are considered relatively small, so the blurring effect can be modeled as a spatial convolution with spectrally varying impulse responses.

It is assumed that each wideband image used for the multispectral dataset can be produced from different imagers with different blurring responses. Let C be the total number of imagers. The blurred object $\boldsymbol{x}^c_{\boldsymbol{h}_\mathrm{m}}$ writes

$$\boldsymbol{x}^c_{\boldsymbol{h}_\mathrm{m}}[i,j,l] = (\boldsymbol{x} \underset{i,j}{*} \boldsymbol{h}^c_\mathrm{m})[i,j,l]$$

where $\underset{i,j}{*}$ is the 2D spatial convolution operator on the axes of i and j with $\boldsymbol{h}^c_\mathrm{m}$ the spectrally varying impulse response of imager $c \in [1, C]$.

Then, the spectral filter profiles and the detector's quantum efficiency of the imager must be taken into account. These effects write as a spectral weighting $\boldsymbol{w}^c_\mathrm{m}$ with

$$\boldsymbol{x}^c_\mathrm{w}[i,j,l] = \boldsymbol{x}^c_{\boldsymbol{h}_\mathrm{m}}[i,j,l]\ \boldsymbol{w}^c_\mathrm{m}[l]. \tag{4}$$

Finally, wideband images result from the integration of the incoming light on a given waveband

$$\boldsymbol{x}^c_\mathrm{m}[i,j] = \sum_l \boldsymbol{x}^c_\mathrm{w}[i,j,l].$$

Thus, transformations of the observed object $\boldsymbol{x}$ to obtain the collection of C wideband images $\boldsymbol{x}_\mathrm{m}$ can be written in a matrix form as

$$\boldsymbol{x}_\mathrm{m} = \boldsymbol{W}_\mathrm{m}\boldsymbol{C}_\mathrm{m}\boldsymbol{x} \tag{5}$$

where $\boldsymbol{C}_\mathrm{m} \in \mathbb{R}^{CLP \times LP}$ writes as

$$\boldsymbol{C}_\mathrm{m} = \begin{pmatrix} \begin{pmatrix} \boldsymbol{C}^1_{\mathrm{m},1} & & \\ & \ddots & \\ & & \boldsymbol{C}^1_{\mathrm{m},L} \end{pmatrix} \\ \vdots \\ \begin{pmatrix} \boldsymbol{C}^C_{\mathrm{m},1} & & \\ & \ddots & \\ & & \boldsymbol{C}^C_{\mathrm{m},L} \end{pmatrix} \end{pmatrix} \tag{6}$$

and where $\boldsymbol{C}^c_{\mathrm{m},l} \in \mathbb{R}^{P \times P}$ is a convolution matrix with the impulse response $\boldsymbol{h}^c_\mathrm{m}$ at the wavelength l.

The matrix $\boldsymbol{W}_\mathrm{m} \in \mathbb{R}^{CP \times CLP}$ operates the spectral response and spectral integration for each of the C imagers, and is written as

$$\boldsymbol{W}_\mathrm{m} = \begin{pmatrix} \boldsymbol{W}^1_\mathrm{m} & \dots & \boldsymbol{W}^C_\mathrm{m} \end{pmatrix} \tag{7}$$

where $\boldsymbol{W}^c_\mathrm{m} \in \mathbb{R}^{CP \times LP}$ writes

$$\boldsymbol{W}^c_\mathrm{m} = \begin{pmatrix} \boldsymbol{0}_P & \dots & \boldsymbol{0}_P \\ \vdots & & \vdots \\ \boldsymbol{w}^c_\mathrm{m}[1]\boldsymbol{I}_P & \dots & \boldsymbol{w}^c_\mathrm{m}[L]\boldsymbol{I}_P \\ \vdots & & \vdots \\ \boldsymbol{0}_P & \dots & \boldsymbol{0}_P \end{pmatrix} \tag{8}$$

and where the row of non-zero matrices is the row c.

Using the Linear Mixing Model for Eq. 5 yields

$$\boldsymbol{x}_\mathrm{m} = \boldsymbol{W}_\mathrm{m}\boldsymbol{C}_\mathrm{m}\boldsymbol{T}\boldsymbol{a} \tag{9a}$$

$$= \boldsymbol{M}\boldsymbol{a} \tag{9b}$$

where $\boldsymbol{M} \in \mathbb{R}^{CP \times TP}$ is the matrix form of the imager forward model, with each block calculated as weighted sums of convolution matrices

$$\boldsymbol{M} = \begin{pmatrix} \sum_l \boldsymbol{w}^1_\mathrm{m}[l]\boldsymbol{s}_1[l]\boldsymbol{C}^1_{\mathrm{m},l} & \dots & \sum_l \boldsymbol{w}^1_\mathrm{m}[l]\boldsymbol{s}_T[l]\boldsymbol{C}^1_{\mathrm{m},l} \\ \vdots & & \vdots \\ \sum_l \boldsymbol{w}^C_\mathrm{m}[l]\boldsymbol{s}_1[l]\boldsymbol{C}^C_{\mathrm{m},l} & \dots & \sum_l \boldsymbol{w}^C_\mathrm{m}[l]\boldsymbol{s}_T[l]\boldsymbol{C}^C_{\mathrm{m},l} \end{pmatrix}. \tag{10}$$

C. Spectrometer model

Common design choices exist between an integral field spectrograph and an imager, such as mirrors to focus light and detectors to sample the flux. Hypotheses do not differ from the imager's blur, as it varies spectrally but is spatially invariant. Therefore, the blurred observed object $\boldsymbol{x}_{\boldsymbol{h}_\mathrm{h}}$ is a convolution between the optical impulse response $\boldsymbol{h}_\mathrm{h}$ of the instrument and $\boldsymbol{x}$ as

$$\boldsymbol{x}_{\boldsymbol{h}_\mathrm{h}}[i,j,l] = (\boldsymbol{x} \underset{i,j}{*} \boldsymbol{h}_\mathrm{h})[i,j,l].$$

Then, the diffraction grating does not transmit flux for all wavelengths equally. This effect, combined with the quantum efficiency of the detector, leads to the spectral weighting of the spectrometer $\boldsymbol{w}_\mathrm{h}$, which must be taken into account as

$$\boldsymbol{x}_w[i,j,l] = \boldsymbol{x}_{\boldsymbol{h}_\mathrm{h}}[i,j,l]\boldsymbol{w}_\mathrm{h}[l].$$

The last effect is the detector spatial sampling of the incoming light with a spatial sampling step greater than the one used for the imager. It is computed as a spatial integration of every region of $d_i \times d_j$ pixels of the object $\boldsymbol{x}_w$, d_i, $d_j \in \mathbb{N}^*$. Consequently, the resulting images are spatially decimated from P to $P' = P/(d_i d_j)$ pixels per image, and each pixel value is calculated as

$$\boldsymbol{x}_\mathrm{h}[\bar{i},\bar{j},l] = \sum_{i=\bar{i}d_i}^{(\bar{i}+1)d_i} \sum_{j=\bar{j}d_j}^{(\bar{j}+1)d_j} \boldsymbol{x}_w[i,j,l] \tag{11}$$

where $\bar{i}$ and $\bar{j}$ are the new pixels coordinates in the spectrometer images.

Like the imager's model, the formation of images from the spectrometer model can be written in a matrix form as

$$\boldsymbol{x}_\mathrm{h} = \overline{\boldsymbol{S}\boldsymbol{\Sigma}}\boldsymbol{W}_\mathrm{h}\boldsymbol{C}_\mathrm{h}\boldsymbol{x} \tag{12}$$

where the matrix $\boldsymbol{C}_\mathrm{h} \in \mathbb{R}^{LP\times LP}$ contains L convolution matrices $\boldsymbol{C}_{\mathrm{h},l} \in \mathbb{R}^{P\times P}$ on its diagonal, applying the spectrally variant blur $\boldsymbol{h}_\mathrm{h}$ for each wavelength.

The matrix $\boldsymbol{W}_\mathrm{h} \in \mathbb{R}^{LP\times LP}$ operates the spectral response of the spectrometer, written as

$$\boldsymbol{W}_\mathrm{h} = \begin{pmatrix} \boldsymbol{w}_\mathrm{h}[1]\boldsymbol{I}_P & & \\ & \ddots & \\ & & \boldsymbol{w}_\mathrm{h}[L]\boldsymbol{I}_P \end{pmatrix}. \tag{13}$$

The spatial integration of the spectrometer's detector, described in Eq. 11, is modeled by combining two matrix operators: a summing operator $\overline{\boldsymbol{\Sigma}} \in \mathbb{R}^{LP\times LP}$ and a subsampling operator $\overline{\boldsymbol{S}} \in \mathbb{R}^{LP'\times LP}$. The first operator replaces the value of each pixel of every image with the sum of the region of $d_i \times d_j$ pixels surrounding them and is written as

$$\overline{\boldsymbol{\Sigma}} = \boldsymbol{I}_L \otimes \boldsymbol{\Sigma} \tag{14}$$

where $\otimes$ stands for the Kronecker product, and $\boldsymbol{\Sigma} \in \mathbb{R}^{P\times P}$ is a convolution matrix with a 2D kernel of size $d_i \times d_j$ and full of ones.

The second operator then selects one pixel for every region of $d_i \times d_j$ pixels, hence a reduction of the image size from P to P' pixels, and writes as

$$\overline{\boldsymbol{S}} = \boldsymbol{I}_L \otimes \boldsymbol{S} \tag{15}$$

where each matrix $\boldsymbol{S} \in \mathbb{R}^{P'\times P}$ subsamples one image of the input cube.

Using Eq. 3 of the Linear Mixing Model allows to write Eq. 12 as

$$\boldsymbol{x}_\mathrm{h} = \overline{\boldsymbol{S}\boldsymbol{\Sigma}}\boldsymbol{W}_\mathrm{h}\boldsymbol{C}_\mathrm{h}\boldsymbol{T}\boldsymbol{a} \tag{16a}$$

$$= \boldsymbol{H}\boldsymbol{a} \tag{16b}$$

where $\boldsymbol{H} \in \mathbb{R}^{LP'\times TP}$ is the matrix form of the spectrometer forward model and is written as

$$\boldsymbol{H} = \overline{\boldsymbol{S}\boldsymbol{\Sigma}}\boldsymbol{W}_\mathrm{h}\boldsymbol{C}_\mathrm{h}\boldsymbol{T} = \begin{pmatrix} \boldsymbol{w}_\mathrm{h}[1]\boldsymbol{s}_1[1]\boldsymbol{S}\boldsymbol{C}_{\boldsymbol{\Sigma},1} & \dots & \boldsymbol{w}_\mathrm{h}[1]\boldsymbol{s}_T[1]\boldsymbol{S}\boldsymbol{C}_{\boldsymbol{\Sigma},1} \\ \vdots & & \vdots \\ \boldsymbol{w}_\mathrm{h}[L]\boldsymbol{s}_1[L]\boldsymbol{S}\boldsymbol{C}_{\boldsymbol{\Sigma},l} & \dots & \boldsymbol{w}_\mathrm{h}[L]\boldsymbol{s}_T[L]\boldsymbol{S}\boldsymbol{C}_{\boldsymbol{\Sigma},l} \end{pmatrix} \tag{17}$$

with $\boldsymbol{C}_{\boldsymbol{\Sigma},l} = \boldsymbol{\Sigma}\boldsymbol{C}_{\mathrm{h},l}$ the joint convolution operator for the wavelength l.

III. Methodology

This section formulates the inverse problem for the reconstruction of the observed object described with abundance maps $\boldsymbol{a}$. Then, fast approaches are detailed to solve the problem for two regularization cases.

A. Formulation of the problem

The abundance maps describing the observed object are defined as the minimizer of a criterion

$$\widehat{\boldsymbol{a}} = \arg\min_{\boldsymbol{a}} \mathcal{J}(\boldsymbol{a}) \tag{18}$$

composed of two data adequacy terms and a regularization term as

$$\mathcal{J}(\boldsymbol{a}) = \mu_\mathrm{m}\|\boldsymbol{y}_\mathrm{m} - \boldsymbol{M}\boldsymbol{a}\|_2^2 + \mu_\mathrm{h}\|\boldsymbol{y}_\mathrm{h} - \boldsymbol{H}\boldsymbol{a}\|_2^2 + R(\boldsymbol{a}) \tag{19}$$

where $\mu_\mathrm{m} = 1/2\sigma_\mathrm{m}^2$, $\mu_\mathrm{h} = 1/2\sigma_\mathrm{h}^2$, and σ_m and σ_h are the standard deviations of additive white Gaussian noises. Because of the spatial convolution, a regularization $R(\boldsymbol{a})$, detailed later, is used to stabilize the solution. An explicit minimizer of this criteria with a quadratic regularization identical to [14] is given in section III-B, and a fast minimization procedure for a convex semi-quadratic regularization is described in section III-E.

B. Quadratic regularization

A formulation of Eq. 19 with a quadratic regularization (ℓ_2-norm), as proposed in Guilloteau *et al.* [14], writes as

$$\mathcal{J}_{\ell_2}(\boldsymbol{a}) = \mu_\mathrm{m}\|\boldsymbol{y}_\mathrm{m} - \boldsymbol{M}\boldsymbol{a}\|_2^2 + \mu_\mathrm{h}\|\boldsymbol{y}_\mathrm{h} - \boldsymbol{H}\boldsymbol{a}\|_2^2 + \mu_\mathrm{r}\Big(\|\overline{\boldsymbol{D}}_\mathrm{r}\boldsymbol{a}\|_2^2 + \|\overline{\boldsymbol{D}}_\mathrm{c}\boldsymbol{a}\|_2^2\Big) \tag{20}$$

where

$$\overline{\boldsymbol{D}}_\mathrm{r} = \boldsymbol{I}_T \otimes \boldsymbol{D}_\mathrm{r}, \tag{21a}$$

$$\overline{\boldsymbol{D}}_\mathrm{c} = \boldsymbol{I}_T \otimes \boldsymbol{D}_\mathrm{c}, \tag{21b}$$

and $\boldsymbol{D}_\mathrm{r}$, $\boldsymbol{D}_\mathrm{c} \in \mathbb{R}^{TP\times TP}$ are first order 2D difference matrices along the rows and columns respectively.

Minimizing $\mathcal{J}_{\ell_2}$ is equivalent to solving $\nabla\mathcal{J}_{\ell_2} = 0$, which yields

$$\widehat{\boldsymbol{a}} = \boldsymbol{Q}_{\ell_2}^{-1}\boldsymbol{q}_{\ell_2} \tag{22}$$

with

$$\boldsymbol{Q}_{\ell_2} = \mu_{\mathrm{m}}\boldsymbol{Q}_{\boldsymbol{M}} + \mu_{\mathrm{h}}\boldsymbol{Q}_{\boldsymbol{H}} + \mu_{\mathrm{r}}(\boldsymbol{Q}_{\overline{\boldsymbol{D}}_{\mathrm{r}}} + \boldsymbol{Q}_{\overline{\boldsymbol{D}}_{\mathrm{c}}}), \tag{23}$$

$$\boldsymbol{q}_{\ell_2} = \mu_{\mathrm{m}}\boldsymbol{M}^H\boldsymbol{y}_{\mathrm{m}} + \mu_{\mathrm{h}}\boldsymbol{H}^H\boldsymbol{y}_{\mathrm{h}}, \tag{24}$$

where $\cdot^H$ is the conjugate transpose operator, $\boldsymbol{Q}_{\ell_2} \in \mathbb{R}^{TP\times TP}$, $\boldsymbol{q}_{\ell_2} \in \mathbb{R}^{TP\times P}$, and

$$\boldsymbol{Q}_{\boldsymbol{M}} = \boldsymbol{M}^H\boldsymbol{M} \tag{25}$$

$$\boldsymbol{Q}_{\boldsymbol{H}} = \boldsymbol{H}^H\boldsymbol{H} \tag{26}$$

$$\boldsymbol{Q}_{\overline{\boldsymbol{D}}_{\mathrm{r}}} = \overline{\boldsymbol{D}}_{\mathrm{r}}^H\overline{\boldsymbol{D}}_{\mathrm{r}} \tag{27}$$

$$\boldsymbol{Q}_{\overline{\boldsymbol{D}}_{\mathrm{c}}} = \overline{\boldsymbol{D}}_{\mathrm{c}}^H\overline{\boldsymbol{D}}_{\mathrm{c}} \tag{28}$$

In [14], authors propose a gradient based descent algorithm to compute $\hat{\boldsymbol{a}}$ (Eq. 22). However, the minimizer $\widehat{\boldsymbol{a}}$ of $\mathcal{J}_{\ell_2}$ can be calculated explicitly if the inversion of $\boldsymbol{Q}_{\ell_2}$ is accessible. Section III-D demonstrates that the structure of $\boldsymbol{Q}_{\ell_2}$, detailed in III-C, can be exploited for its fast inversion, that $\boldsymbol{Q}_{\ell_2}^{-1}$ is easily storable and its application to $\boldsymbol{q}_{\ell_2}$ is also fast.

C. Structure of $\boldsymbol{Q}_{\ell_2}$

Lemma 1: If convolution matrices $\boldsymbol{C}_*$ and $\boldsymbol{D}_*$ are circulant, then the matrices $\boldsymbol{Q}_{\boldsymbol{M}}$, $\boldsymbol{Q}_{\boldsymbol{H}}$, and $\boldsymbol{Q}_{\overline{\boldsymbol{D}}_*}$ are square with circulant blocks of equal size. Consequently, $\boldsymbol{Q}_{\ell_2}$ is also made of circulant blocks.

1) Structure of $\boldsymbol{Q}_{\boldsymbol{M}}$: As shown in Eq. 10, the matrix $\boldsymbol{M}$ contains $C\times T$ blocks $\boldsymbol{M}_{c,t}$ that are sums of circulant matrices

$$\boldsymbol{M}_{c,t} = \sum_l \boldsymbol{w}_{\mathrm{m}}^c[l]\boldsymbol{s}_t[l]\boldsymbol{C}_{\mathrm{m},l}. \tag{29}$$

As a result, the blocks $\boldsymbol{M}_{c,t}$ are circulant themselves. Then, the matrix $\boldsymbol{Q}_{\boldsymbol{M}}$ is a square matrix of $T\times T$ blocks that writes

$$(\boldsymbol{Q}_{\boldsymbol{M}})_{t,t'} = \sum_c \boldsymbol{M}_{c,t}^H\boldsymbol{M}_{c,t'} \tag{30}$$

that are also circulant [18].

2) Structure of $\boldsymbol{Q}_{\boldsymbol{H}}$: From Eq. 17, the $L\times T$ blocks of $\boldsymbol{H}$ writes

$$\boldsymbol{H}_{l,t} = \boldsymbol{w}_{\mathrm{h}}[l]\boldsymbol{s}_t[l]\boldsymbol{S}\boldsymbol{C}_{\boldsymbol{\Sigma},l} \tag{31}$$

Therefore, $\boldsymbol{Q}_{\boldsymbol{H}}$ is a square matrix of $T\times T$ blocks that writes as

$$(\boldsymbol{Q}_{\boldsymbol{H}})_{t,t'} = \sum_l \boldsymbol{w}_{\mathrm{h}}[l]^2\boldsymbol{s}_t[l]\boldsymbol{s}_{t'}[l]\boldsymbol{C}_{\boldsymbol{\Sigma},l}^H\boldsymbol{S}^H\boldsymbol{S}\boldsymbol{C}_{\boldsymbol{\Sigma},l}. \tag{32}$$

However, the structure of these blocks in the Fourier space leads to exploitable matrix structures. Indeed, writing $\boldsymbol{C}_{\boldsymbol{\Sigma},l} = \boldsymbol{F}^H\boldsymbol{\Lambda}_l\boldsymbol{F}$ gives

$$\boldsymbol{C}_{\boldsymbol{\Sigma},l}^H\boldsymbol{S}^H\boldsymbol{S}\boldsymbol{C}_{\boldsymbol{\Sigma},l} = \boldsymbol{F}^H\boldsymbol{\Lambda}_l^H\boldsymbol{F}\boldsymbol{S}^H\boldsymbol{S}\boldsymbol{F}^H\boldsymbol{\Lambda}_l\boldsymbol{F}. \tag{33}$$

As $\boldsymbol{S}$ carries a subsampling while $\boldsymbol{S}^H$ fills images with zeros, the resulting operation of $\boldsymbol{S}^H\boldsymbol{S}$ is a spatial weighing. More precisely, the matrix $\boldsymbol{S}^H\boldsymbol{S}$ is diagonal with 1 every d_id_j coefficients and zeros elsewhere, i.e. the diagonal is a Dirac comb of step d_id_j. Therefore, $\boldsymbol{F}\boldsymbol{S}^H\boldsymbol{S}\boldsymbol{F}^H$ is a circulant convolution matrix with a Dirac comb as kernel, and as presented in Wei *et al.* [13], writes

$$\boldsymbol{F}\boldsymbol{S}^H\boldsymbol{S}\boldsymbol{F}^H = \frac{1}{d_id_j}\boldsymbol{J}_{d_id_j}\otimes\boldsymbol{I}_{P'} \tag{34}$$

where $\boldsymbol{J}_{d_id_j} \in \mathbb{R}^{d_id_j\times d_id_j}$ is full of "1".

Multiplying Eq. 34 by $\boldsymbol{\Lambda}_l$ on the right gives

$$\begin{aligned}
&\boldsymbol{F}\boldsymbol{S}^H\boldsymbol{S}\boldsymbol{F}^H\boldsymbol{\Lambda}_l \\
&= \frac{1}{d_id_j}\begin{pmatrix}\boldsymbol{I}_{P'} & \dots & \boldsymbol{I}_{P'}\\ \vdots & & \vdots \\ \boldsymbol{I}_{P'} & \dots & \boldsymbol{I}_{P'}\end{pmatrix}\begin{pmatrix}\boldsymbol{\Lambda}_{l,1} & & \\ & \ddots & \\ & & \boldsymbol{\Lambda}_{l,d_id_j}\end{pmatrix} \\
&= \frac{1}{d_id_j}\begin{pmatrix}\boldsymbol{\Lambda}_{l,1} & \dots & \boldsymbol{\Lambda}_{l,d_id_j}\\ \vdots & & \vdots \\ \boldsymbol{\Lambda}_{l,1} & \dots & \boldsymbol{\Lambda}_{l,d_id_j}\end{pmatrix}
\end{aligned} \tag{35}$$

where $\boldsymbol{\Lambda}_{l,d'}$, $d' \in [1, d_id_j]$, are subdivisions of the diagonal of the initial matrix $\boldsymbol{\Lambda}_l$.

Multiplying Eq. 35 by $\boldsymbol{\Lambda}_l^H$ on the left gives

$$\begin{aligned}
&\boldsymbol{\Lambda}_l^H\boldsymbol{F}\boldsymbol{S}^H\boldsymbol{S}\boldsymbol{F}^H\boldsymbol{\Lambda}_l = \\
&\qquad\frac{1}{d_id_j}\begin{pmatrix}\boldsymbol{\Lambda}_{l,1}^H\boldsymbol{\Lambda}_{l,1} & \dots & \boldsymbol{\Lambda}_{l,1}^H\boldsymbol{\Lambda}_{l,d_id_j}\\ \vdots & & \vdots \\ \boldsymbol{\Lambda}_{l,d_id_j}^H\boldsymbol{\Lambda}_{l,1} & \dots & \boldsymbol{\Lambda}_{l,d_id_j}^H\boldsymbol{\Lambda}_{l,d_id_j}\end{pmatrix}.
\end{aligned} \tag{36}$$

As each block of $\boldsymbol{\Lambda}_l^H\boldsymbol{F}\boldsymbol{S}^H\boldsymbol{S}\boldsymbol{F}^H\boldsymbol{\Lambda}_l$ is diagonal, the matrices $\boldsymbol{C}_{\boldsymbol{\Sigma},l}^H\boldsymbol{S}^H\boldsymbol{S}\boldsymbol{C}_{\boldsymbol{\Sigma},l}$ defined at Eq. 33 are thus made circulant blocks, which turns also true for the blocks of $\boldsymbol{Q}_{\boldsymbol{H}}$ (Eq. 32).

3) Structures of $\boldsymbol{Q}_{\overline{\boldsymbol{D}}_*}$: Using Eqs. 21a and 21b gives expressions for $\boldsymbol{Q}_{\overline{\boldsymbol{D}}_{\mathrm{r}}}$ and $\boldsymbol{Q}_{\overline{\boldsymbol{D}}_{\mathrm{c}}}$ as

$$\boldsymbol{Q}_{\overline{\boldsymbol{D}}_{\mathrm{r}}} = \boldsymbol{I}_T\otimes\boldsymbol{D}_{\mathrm{r}}^H\boldsymbol{D}_{\mathrm{r}} \text{ and} \tag{37}$$

$$\boldsymbol{Q}_{\overline{\boldsymbol{D}}_{\mathrm{c}}} = \boldsymbol{I}_T\otimes\boldsymbol{D}_{\mathrm{c}}^H\boldsymbol{D}_{\mathrm{c}}. \tag{38}$$

Since $\boldsymbol{D}_*$ are circulant, $\boldsymbol{D}_*^H\boldsymbol{D}_*$ are circulant and $\boldsymbol{Q}_{\overline{\boldsymbol{D}}_*}$ are block matrices with circulant blocks.

D. Inversion of $\boldsymbol{Q}_{\ell_2}$

Proposition 1: If

$$\ker\boldsymbol{Q}_{\boldsymbol{M}} \cap \ker\boldsymbol{Q}_{\boldsymbol{H}} \cap \ker\boldsymbol{Q}_{\overline{\boldsymbol{D}}_{\mathrm{r}}} \cap \ker\boldsymbol{Q}_{\overline{\boldsymbol{D}}_{\mathrm{c}}} = \varnothing$$

then $\boldsymbol{Q}_{\ell_2}^{-1}$ exists and is a block matrix with circulant blocks. Moreover, thanks to Fourier transform, $\boldsymbol{Q}_{\ell_2}^{-1}$ is tractable, storable in Fourier space and applicable. The proof is described Appendix A.

E. Semi-quadratic regularization

The quadratic penalization of the previous regularization may turn inadequate when restoring high spatial gradients in images and can induce ringing artifacts. To solve this issue, non-quadratic regularizations can be convenient alternatives [19], [20]. A common choice is to use a strictly differentiable convex function φ within the regularization function R, such that φ is [16]

1) φ is $\mathcal{C}^1$, even, coercive,
2) $\forall u \in \mathbb{R}^+$, $\varphi(\sqrt{u})$ is concave,
3) $\forall u \in \mathbb{R}$, $0 < \dot{\varphi}(u)/u < +\infty$.

Therefore, the new convex criterion to minimize writes as

$$\mathcal{J}_{\ell_{2,1}}(\boldsymbol{a}) = \mu_{\mathrm{m}}\|\boldsymbol{y}_{\mathrm{m}} - \boldsymbol{M}\boldsymbol{a}\|_2^2 + \mu_{\mathrm{h}}\|\boldsymbol{y}_{\mathrm{h}} - \boldsymbol{H}\boldsymbol{a}\|_2^2 + \mu_{\mathrm{r}} \sum_{t,i,j} \Big(\varphi(\boldsymbol{D}_{\mathrm{r}}\boldsymbol{a}_t)[i,j] + \varphi(\boldsymbol{D}_{\mathrm{c}}\boldsymbol{a}_t)[i,j] \Big) \quad (39)$$

As proposed by Geman and Yang (GY) [20] and used in [17], the half-quadratic strategy aims to minimize $\mathcal{J}_{\ell_{2,1}}(\boldsymbol{a})$ by defining an augmented criterion $\mathcal{K}(\boldsymbol{a}, \boldsymbol{b}_{\mathrm{r}}, \boldsymbol{b}_{\mathrm{c}})$ such that

$$\min_{\boldsymbol{b}_{\mathrm{r}}, \boldsymbol{b}_{\mathrm{c}}} \mathcal{K}(\boldsymbol{a}, \boldsymbol{b}_{\mathrm{r}}, \boldsymbol{b}_{\mathrm{c}}) = \mathcal{J}_{\ell_{2,1}}(\boldsymbol{a}) \quad (40)$$

where $\boldsymbol{b}_{\mathrm{r}}$ and $\boldsymbol{b}_{\mathrm{c}}$ are auxiliary variables. Therefore, under light hypothesis [19] $\mathcal{K}$ is convex and a relaxed version of $\mathcal{J}$ with common minimizer, i.e.

$$\hat{\boldsymbol{a}}, \hat{\boldsymbol{b}}_{\mathrm{r}}, \hat{\boldsymbol{b}}_{\mathrm{c}} = \arg\min_{\boldsymbol{a}, \boldsymbol{b}_{\mathrm{r}}, \boldsymbol{b}_{\mathrm{c}}} \mathcal{K} \quad \text{and} \quad \hat{\boldsymbol{a}} = \arg\min_{\boldsymbol{a}} \mathcal{J}_{\ell_{2,1}}. \quad (41)$$

This convex [19] surrogate criterion writes as

$$\mathcal{K}(\boldsymbol{a}, \boldsymbol{b}_{\mathrm{r}}, \boldsymbol{b}_{\mathrm{c}}) = \mu_{\mathrm{m}}\|\boldsymbol{y}_{\mathrm{m}} - \boldsymbol{M}\boldsymbol{a}\|_2^2 + \mu_{\mathrm{h}}\|\boldsymbol{y}_{\mathrm{h}} - \boldsymbol{H}\boldsymbol{a}\|_2^2 + \mu_{\mathrm{r}}\Big(\|\overline{\boldsymbol{D}_{\mathrm{r}}}\boldsymbol{a} - \boldsymbol{b}_{\mathrm{r}}\|_2^2 + \overline{\zeta}(\boldsymbol{b}_{\mathrm{r}}) + \|\overline{\boldsymbol{D}_{\mathrm{c}}}\boldsymbol{a} - \boldsymbol{b}_{\mathrm{c}}\|_2^2 + \overline{\zeta}(\boldsymbol{b}_{\mathrm{c}}) \Big) \quad (42)$$

where $\overline{\zeta}(\boldsymbol{u}) = \sum_i \zeta(\boldsymbol{u}_i)$, with ζ linked to φ through convex duality [19].

The benefit of this strategy is to transform a high-dimensional convex minimization with correlation problem into two separate minimization problems: a quadratic one with correlation (minimizing $\mathcal{K}$ with $\boldsymbol{b}_{\mathrm{r}}$ and $\boldsymbol{b}_{\mathrm{c}}$ fixed), and a separable one wrt. to the auxiliary variables (minimizing $\mathcal{K}$ with $\boldsymbol{a}$ fixed), as

$$\begin{cases} \widehat{\boldsymbol{a}}^{(k)} &= \arg\min_{\boldsymbol{a}} \mathcal{K}\left(\boldsymbol{a}, \widehat{\boldsymbol{b}}_{\mathrm{r}}^{(k)}, \widehat{\boldsymbol{b}}_{\mathrm{c}}^{(k)}\right), & (43) \\ \widehat{\boldsymbol{b}}_{\mathrm{r}}^{(k+1)}, \widehat{\boldsymbol{b}}_{\mathrm{c}}^{(k+1)} &= \arg\min_{\boldsymbol{b}_{\mathrm{r}}, \boldsymbol{b}_{\mathrm{c}}} \mathcal{K}\left(\widehat{\boldsymbol{a}}^{(k)}, \boldsymbol{b}_{\mathrm{r}}, \boldsymbol{b}_{\mathrm{c}}\right). & (44) \end{cases}$$

The alternative resolution of these minimization problems leads to finding the triplet ($\widehat{\boldsymbol{a}}$, $\widehat{\boldsymbol{b}}_{\mathrm{r}}$, $\widehat{\boldsymbol{b}}_{\mathrm{c}}$) that minimizes $\mathcal{K}$, and according to Eq. 41, $\widehat{\boldsymbol{a}}$ will minimize the original criterion $\mathcal{J}_{\ell_{2,1}}$.

Solving the first problem (Eq. 43) yields

$$\widehat{\boldsymbol{a}} = \boldsymbol{Q}_{\ell_{2,1}}^{-1}\boldsymbol{q}_{\ell_{2,1}} \quad (45)$$

where

$$\boldsymbol{Q}_{\ell_{2,1}} = \mu_{\mathrm{m}}\boldsymbol{Q}_{\boldsymbol{M}} + \mu_{\mathrm{h}}\boldsymbol{Q}_{\boldsymbol{H}} + \mu_{\mathrm{r}}\Big(\boldsymbol{Q}_{\overline{\boldsymbol{D}}_{\mathrm{r}}} + \boldsymbol{Q}_{\overline{\boldsymbol{D}}_{\mathrm{c}}} \Big) \quad (46)$$

$$\boldsymbol{q}_{\ell_{2,1}} = \mu_{\mathrm{m}}\boldsymbol{M}^H\boldsymbol{y}_{\mathrm{m}} + \mu_{\mathrm{h}}\boldsymbol{H}^H\boldsymbol{y}_{\mathrm{h}} + \mu_{\mathrm{r}}\Big(\overline{\boldsymbol{D}}_{\mathrm{r}}^H\boldsymbol{b}_{\mathrm{r}} + \overline{\boldsymbol{D}}_{\mathrm{c}}^H\boldsymbol{b}_{\mathrm{c}} \Big) \quad (47)$$

and where expressions of $\boldsymbol{Q}_{\boldsymbol{M}}$, $\boldsymbol{Q}_{\boldsymbol{H}}$, $\boldsymbol{Q}_{\overline{\boldsymbol{D}}_{\mathrm{r}}}$ and $\boldsymbol{Q}_{\overline{\boldsymbol{D}}_{\mathrm{c}}}$ are found from Eq. 25 to 28.

Noticing that $\boldsymbol{Q}_{\ell_{2,1}} = \boldsymbol{Q}_{\ell_2}$, where the latter matrix was defined in section III-B for the resolution of the quadratic criterion $\mathcal{J}_{\ell_2}$, leads to invert $\boldsymbol{Q}_{\ell_{2,1}}$ as it was done for $\boldsymbol{Q}_{\ell_2}$ in section III-D. As $\boldsymbol{Q}_{\ell_{2,1}}^{-1}$ does not depend on $\boldsymbol{a}$, $\boldsymbol{b}_{\mathrm{r}}$ or $\boldsymbol{b}_{\mathrm{c}}$, it can only be calculated once for a given problem, at the start of the alternative minimization procedure. As a consequence, the solution to Eq. 43 is explicit and fast.

The second minimization problem Eq. 44, used to update the auxiliary variables $\boldsymbol{b}_{\mathrm{r}}$ and $\boldsymbol{b}_{\mathrm{c}}$, can be reformulated in two distinct problems with

$$\begin{cases} \widehat{\boldsymbol{b}}_{\mathrm{r}} &= \arg\min_{\boldsymbol{b}_{\mathrm{r}}} \frac{1}{2}\|\overline{\boldsymbol{D}_{\mathrm{r}}}\boldsymbol{a} - \boldsymbol{b}_{\mathrm{r}}\|_2^2 + \overline{\zeta}(\boldsymbol{b}_{\mathrm{r}}), & (48) \\ \widehat{\boldsymbol{b}}_{\mathrm{c}} &= \arg\min_{\boldsymbol{b}_{\mathrm{c}}} \frac{1}{2}\|\overline{\boldsymbol{D}_{\mathrm{c}}}\boldsymbol{a} - \boldsymbol{b}_{\mathrm{c}}\|_2^2 + \overline{\zeta}(\boldsymbol{b}_{\mathrm{c}}). & (49) \end{cases}$$

Cancelling the gradients of both of these criteria gives formulations for the minimizers [17] as

$$\widehat{\boldsymbol{b}}_{\mathrm{r}} = \overline{\boldsymbol{D}}_{\mathrm{r}}\boldsymbol{a} - \frac{1}{2}\overline{\varphi}'\left(\overline{\boldsymbol{D}}_{\mathrm{r}}\boldsymbol{a}\right), \quad (50)$$

$$\widehat{\boldsymbol{b}}_{\mathrm{c}} = \overline{\boldsymbol{D}}_{\mathrm{c}}\boldsymbol{a} - \frac{1}{2}\overline{\varphi}'\left(\overline{\boldsymbol{D}}_{\mathrm{c}}\boldsymbol{a}\right) \quad (51)$$

where $\overline{\varphi}'$ independently applies φ', the first derivative of φ, to every element of the input vector.

As illustrated, both minimization problems of Eqs. 43 and 44 are solved explicitly. Consequently, this procedure is expected to converge quickly in many test cases. This algorithm is described in Alg. 1, where the stopping criterion can be chosen to exploit the least square distance between two iterates, the value of the criterion $\mathcal{J}_{\ell_{2,1}}$ of Eq. 39, or the number of iterations.

Algorithm 1: Convex $\ell_{2,1}$ HS-MS fusion

Input : μ_{m}, $\boldsymbol{y}_{\mathrm{m}}$, $\boldsymbol{M}$, μ_{h}, $\boldsymbol{y}_{\mathrm{h}}$, $\boldsymbol{H}$, μ_{r}, $\overline{\boldsymbol{D}}_{\mathrm{r}}$, $\overline{\boldsymbol{D}}_{\mathrm{c}}$, $\overline{\varphi}'$
$\boldsymbol{P} \leftarrow \boldsymbol{Q}_{\ell_{2,1}}^{-1}$; `# See Eq. (46) and Appendix A`
$\bar{\boldsymbol{y}} \leftarrow \mu_{\mathrm{m}}\boldsymbol{M}^H\boldsymbol{y}_{\mathrm{m}} + \mu_{\mathrm{h}}\boldsymbol{H}^H\boldsymbol{y}_{\mathrm{h}}$;
while *Stopping criterion is not met* **do**
  $\boldsymbol{b}_{\mathrm{r}} \leftarrow \overline{\boldsymbol{D}}_{\mathrm{r}}\boldsymbol{a} - \frac{1}{2}\overline{\varphi}'\left(\overline{\boldsymbol{D}}_{\mathrm{r}}\boldsymbol{a}\right)$
  $\boldsymbol{b}_{\mathrm{c}} \leftarrow \overline{\boldsymbol{D}}_{\mathrm{c}}\boldsymbol{a} - \frac{1}{2}\overline{\varphi}'\left(\overline{\boldsymbol{D}}_{\mathrm{c}}\boldsymbol{a}\right)$
  $\boldsymbol{a} \leftarrow \boldsymbol{P}\left(\bar{\boldsymbol{y}} + \mu_{\mathrm{r}}\left(\overline{\boldsymbol{D}}_{\mathrm{r}}^H\boldsymbol{b}_{\mathrm{r}} + \overline{\boldsymbol{D}}_{\mathrm{c}}^H\boldsymbol{b}_{\mathrm{c}}\right)\right)$
end
Output: $\boldsymbol{a}$

IV. Experimental Results

The proposed method for the fusion of multispectral and hyperspectral data can be applied to a relatively wide variety of cases. Indeed, the impulse response can differ between the imager and the spectrometer model. Moreover, the wideband images obtained in the multispectral dataset can originate from different imagers (with different impulse responses and spectral responses) as long as their fields of view and spatial sampling steps are the same. Note that the fusion with imagery data with different sampling spatial steps or the fusion with more than one hyperspectral datacubes is possible by adding more data attachment terms in the originally formulated criterion of Eq. 19, but this was not explored in this paper.

In this section, the proposed algorithms' performances were evaluated by simulating the fusion of MS and HS data from

the mid-infrared instrument MIRI of the James Webb Space Telescope [21]. Instrument models described in sections II-B and II-C have been adapted for the imager MIRIM [22] and the Mid Resolution Spectrometer (MRS) [23] belonging to MIRI.

The instrument models chosen for this study are described in section IV-A and the synthetic observation datasets in section IV-B. The performances of data fusion approaches are compared with reconstruction quality metrics, explained in section IV-C. Results obtained with the quadratic and semi-quadratic regularizations, introduced in sections III-B and III-E, are presented in terms of time benefits and quality of reconstruction in sections IV-D and IV-E. Finally, inverse problem approaches are compared with a neural network method of the literature in section IV-F.

A. Study case instrument models

The PSF of the JWST, obtained with *webbpsf* [24], was used to simulate the impulse response used in the imager and spectrometer models of sections II-B and II-C. The FWHM of this PSF is proportionally linear to the wavelength, as illustrated in Fig. 1.

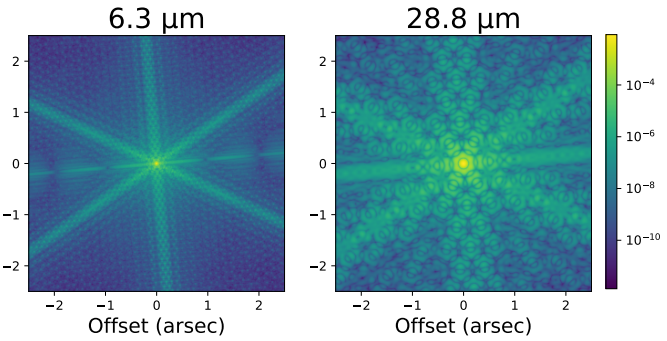


Fig. 1: PSF of MIRI/JWST.

The spectral response of the imager uses the real Photon Conversion Efficiency (PCE) curves of MIRIM [21], where $C = 9$ different spectral channels are considered [22]. The spectral response induced by the MRS is composed of 4 channels with 3 sub-channels each [23]. However, a simplified spectrometer model is considered with one unique channel and no sub-channel, with a spectral response taken to be the maximum of the concatenated PCE curves of the MRS [21].

This model applies a spatial decimation with $d_i = d_j = 4$, i.e. integrates square region of 4×4 pixels into one, which induces a spatial subsampling for wavelength shorter than 20 µm [21].

B. Synthetic data

Typical MS and HS data that MIRI/JWST could acquire were created from a set of $T = 5$ abundance maps $\boldsymbol{a}$ derived from real observations [25], [26], shown Fig. 3. These maps are associated with 5 spectra $\boldsymbol{s}$ [27] of length $L = 300$ covering the MIRI spectral range (from 4.6 to 28.6 µm), shown Fig. 2. The maps $\boldsymbol{a}$ and spectra $\boldsymbol{s}$ have been used as first inputs for the imager and spectrometer model to simulate multispectral and hyperspectral data.

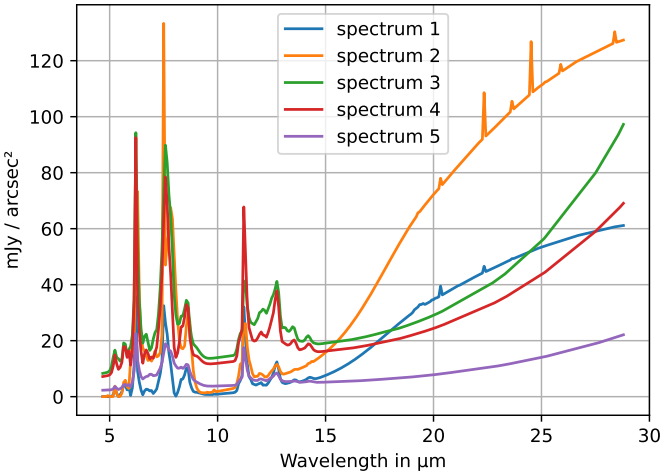


Fig. 2: Spectra for the Linear Mixing Model.

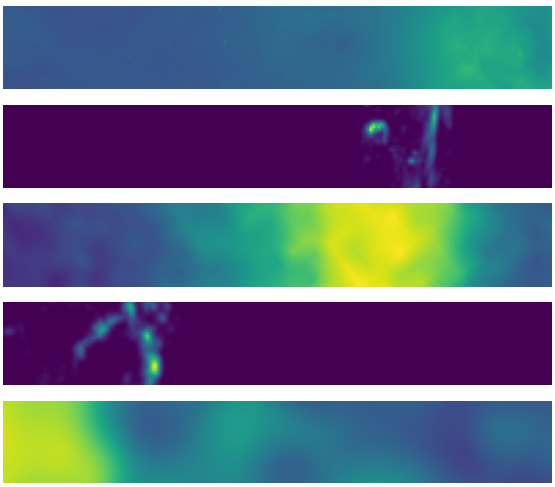


Fig. 3: Original abundance maps $\boldsymbol{a}$

The readout and shot noise present for real observations were also simulated by corrupting images with an additive white Gaussian noise so that the signal-to-noise ratio $\mathrm{SNR}_{\mathrm{dB}}$ in both the MS and HS datasets would be the same, where

$$\mathrm{SNR}_{\mathrm{dB}}(\boldsymbol{y}) = 10 \log_{10} \left(\frac{\|\boldsymbol{y}\|_2^2}{N\sigma^2} \right) \tag{52}$$

with σ the standard deviation of the Gaussian noise, and N the number of values in $\boldsymbol{y}$, i.e. $N = P \times C$ for the MS data, and $N = P' \times L$ for HS data.

C. Reconstruction quality metrics

The quality of the reconstructed cubes was evaluated using four different criteria : the Normalized Root Mean Squared Error (NRMSE), the Structural Dissimilarity (aDSSIM), the Spectral Angle Mapper (aSAM), and the Peak Signal-to-Noise Ratio (PSNR). For all criteria except the PSNR, the lower the value, the better.

The Normalized Root Mean Squared Error (NRMSE) [28] measures the cumulative squared error between each pixel of

the reconstructed cube $\widehat{\boldsymbol{x}}$ and the true cube $\boldsymbol{x}$, and writes as

$$\text{NRMSE}(\boldsymbol{x}, \widehat{\boldsymbol{x}}) = \frac{\|\boldsymbol{x} - \widehat{\boldsymbol{x}}\|_2}{\|\boldsymbol{x}\|_2} \tag{53}$$

with a value between 0 and 1, the lower the better the reconstruction.

The Structural SIMilarity (SSIM) [29] evaluates the quality of the perceived spatial quality of a reconstructed image w.r.t. the original one. To evaluate the spatial quality of the full reconstructed cube, the average SSIM between all images of both cubes $\boldsymbol{x}$ and $\widehat{\boldsymbol{x}}$ is calculated and rescaled between 0 to 1, where 0 means perfect spatial similarity and 1 perfect anti-correlation. This value is called the average Structural Dissimilarity (aDSSIM) and is written as

$$\text{aDSSIM}(\boldsymbol{x}, \widehat{\boldsymbol{x}}) = \frac{1}{2}\left(1 - \frac{1}{L}\sum_{l=1}^{L} \text{SSIM}(\boldsymbol{x}_l, \widehat{\boldsymbol{x}}_l)\right). \tag{54}$$

The average Spectral Angle Mapper (aSAM) [30] measures the spectral distorsion between $\boldsymbol{x}$ and $\widehat{\boldsymbol{x}}$. It calculates the average angular distance between the spectra of both cubes at all pixels. It is formulated as

$$\text{aSAM}(\boldsymbol{x}, \widehat{\boldsymbol{x}}) = \frac{1}{P}\sum_{p=1}^{P} \arccos\left(\frac{\langle \boldsymbol{x}_p, \widehat{\boldsymbol{x}}_p \rangle}{||\boldsymbol{x}_p||_2 ||\widehat{\boldsymbol{x}}_p||_2}\right) \tag{55}$$

and returns a value between 0 and π. The lower this value is, the better the spectral content of the original cube $\boldsymbol{x}$ was recovered in $\widehat{\boldsymbol{x}}$.

The Peak Signal-to-Noise Ratio [31] evaluates the quality of the reconstruction by calculating the maximum ratio of a pixel value to the noise value. It is calculated as

$$\text{PSNR}(\boldsymbol{x}, \widehat{\boldsymbol{x}}) = 20\log_{10}\left(\frac{\max(\boldsymbol{x})}{\sqrt{\text{MSE}(\boldsymbol{x}, \widehat{\boldsymbol{x}})}}\right) \tag{56}$$

and the higher it is, the higher the quality of reconstruction is.

D. Reconstruction with ℓ_2 approach

The explicit procedure for the minimization of the quadratically regularized criterion $\mathcal{J}_{\ell_2}$ proposed in section III-B, now called the ℓ_2 approach, was used to reconstruct abundance maps $\boldsymbol{a}_{\ell_2}$ from both MS and HS datasets.

Its performance was compared with a gradient-based iterative approach, as proposed in [14], regarding the quality of reconstruction and computational load. Although the exact algorithm used in [14] is not mentioned, the Linear Conjugate Gradient (LCG) [32] optimization algorithm was specifically used for the comparison with the proposed ℓ_2 explicit approach of this paper. The value for the only hyperparameter μ_r was chosen for both methods to maximize the quality of the reconstructed cube $\widehat{\boldsymbol{x}}_{\ell_2}$ wrt. the original cube $\boldsymbol{x}$, according to the NRMSE, defined at Eq. 53.

Inverse problem approaches were also compared with more "naive" reconstruction methods, such as the coaddition, which consists of manually inverting the spectral response for both MS and HS datasets and reformating them to allow the calculation of the average cube. To do so, images of the MS dataset were duplicated for each wavelength on their corresponding spectral band, and images of the HS dataset were upsampled to match the dimensions of MS images.

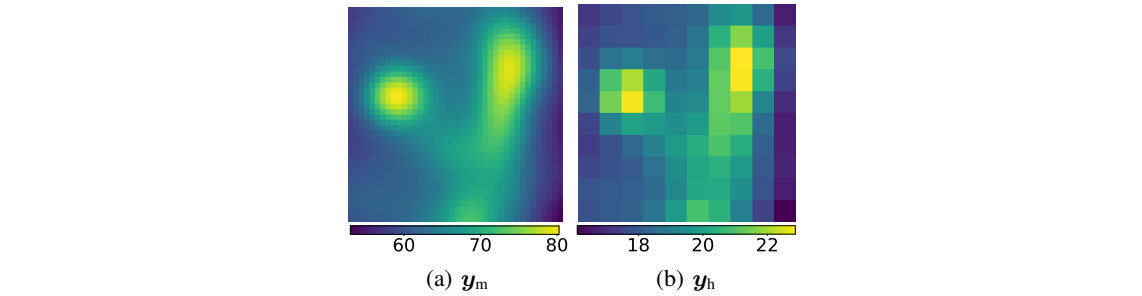
Fig. 4 compares the reconstructions of the coaddition method, the iterative inverse problem approach (LCG) [14], and the exact solution found with the proposed procedure at a relatively long wavelength (25.5 µm) where instruments data are blurry. Two images are shown for the iterative approach: one after 0.05 seconds of calculations (the computation time of the explicit approach, which corresponds to one iteration), and one after convergence of the algorithm (approximately 6 minutes, 7 000 iterations), with a convergence defined as reaching 0.1% of $\min(\mathcal{J}_{\ell_2})$[1]. Note that precalculations, such as the calculation of $\boldsymbol{Q}$ (Eq. 23), $\boldsymbol{Q}^{-1}$ (section III-D) and $\boldsymbol{q}_{\ell_2}$ (Eq. 24) in the Fourier space were not included in these time measurements. For information, in that case, precalculations required 15 seconds for LCG and 17 seconds for the proposed method (longer time due to the inversion of $\boldsymbol{Q}$, not required for LCG). Tab. I compares the reconstruction quality measurements between the different reconstruction methods.

As expected, results for the coaddition method are relatively poor as the method does not deconvolve images. The intensities of pixels are slightly closer to the original than those given in the instrument's data, but much more precisely recovered with inverse problem approaches. Efficient deconvolution results of the latter approaches are explained by their ability to take into account instrument models (sections II-B and II-C) and exploit the high spatial resolution of short wavelengths to recover high-frequency features at long wavelengths through the spectral correlations introduced with the Linear Mixing Model (section II-A).

At equal computation time, the proposed method performs a significantly finer restoration of the signal than the iterative approach. When allowed to converge, the iterative approach performs visually equivalently to the proposed method, as expected. However, the reconstruction with the explicit approach is 7 000 times faster while demonstrating slightly superior reconstruction quality, as shown in Tab. I. Indeed, for the reconstruction of $T = 5$ abundance maps of size 124×248 from an MS dataset of size 124×248×9 and an HS dataset of size 31×62×300, the proposed explicit ℓ_2 approach required only 0.05 seconds for $\widehat{\boldsymbol{a}}_{\ell_2} = \boldsymbol{Q}_{\ell_2}^{-1}\boldsymbol{q}_{\ell_2}$ on a processor i9-10885H, 2.40 GHz (x 16), 64 Go RAM. In contrast, the iterative approach requires a computational time exponentially proportional to the signal-to-noise ratio SNR_dB, as shown in Fig. 5. Therefore, for a low noise case (SNR_dB = 100 dB), approximately 6 minutes were necessary with the LCG algorithm to reach 0.1% of the minimum of the criteria $\min(\mathcal{J}_{\ell_2})$. Indeed, the less noisy the observation data, i.e. the higher the SNR_dB, the more high frequencies need to be restored, thus requiring a more significant number of iterations. This issue is avoided with the proposed explicit ℓ_2 approach.

It is worth noting that the time-saving factor relative to the LCG algorithm could be higher by setting a convergence threshold stricter than 0.1% of $\min(\mathcal{J}_{\ell_2})$. In any case, such a time-saving factor could save hours of computational time for the joint processing of heavier MS and HS datasets, e.g. real

[1]The value of $\min(\mathcal{J}_{\ell_2})$ comes from the proposed explicit approach.



(a) $\boldsymbol{y}_{\mathrm{m}}$ (b) $\boldsymbol{y}_{\mathrm{h}}$

(c) Coaddition (d) LCG [14] in 0.05 seconds (e) LCG [14] in 6 minutes (f) Exact solution in 0.05 seconds (g) Original

Fig. 4: Imagery data $\boldsymbol{y}_{\mathrm{m}}$ and spectrometry data $\boldsymbol{y}_{\mathrm{h}}$ with $\mathrm{SNR}_{\mathrm{dB}} = 100$ dB on top, and reconstruction with the coaddition method Fig. 4c. The next figures show results of the minimization of $\mathcal{J}_{\ell_2}$ with the LCG algorithm in 0.05 second and 6 minutes (value of criterion at 0.1% of $\min(\mathcal{J}_{\ell_2})$), and with the demonstrated explicit method (section III-B) in 0.05 seconds. The last figure is the original image. Central wavelength for all images is 25.5 µm. Unit is mJy arcsec^{-2}. The dynamical scales for the three last figures are identical.

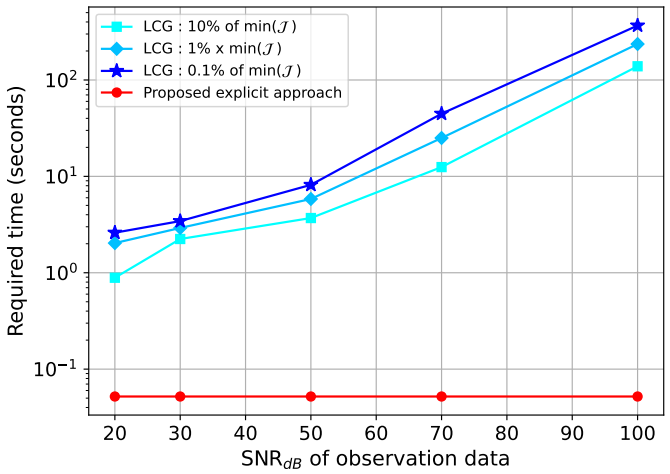


Fig. 5: Time required for the Linear Conjugate Gradient (LCG) algorithm [14] to reach 10%, 1% and 0.1% of $\min(\mathcal{J}_{\ell_2})$ (Eq. 20), compared to the time required for the proposed explicit solution as a function of the signal-to-noise ratio $\mathrm{SNR}_{\mathrm{dB}}$ of the MS and HS datasets. Required time for the proposed explicit approach: 0.05 seconds. Precalculations are not taken into account.

observation datasets.

Nonetheless, our explicit ℓ_2 approach was tested for the fusion of datasets with dimensions equal to the ones used in [14], i.e. 11 images of size 90×900 pixels for the MS dataset, and 4974 images of 30×300 pixels for the HS dataset, with only $T = 4$ spectra templates $\boldsymbol{s}$ used for the Linear Mixing Model. In that case, the fastest approach proposed by [14], the "vectorized approach", allowed a reconstruction after approximately 36 minutes of preprocessing and 20 seconds of processing. The number of iterations or time per iteration was not mentioned in the paper. In comparison, our proposed explicit approach provided the exact minimizer of $\mathcal{J}_{\ell_2}$ with 4 minutes of preprocessing and again only 0.05 seconds for the calculation of $\widehat{\boldsymbol{a}} = \boldsymbol{Q}^{-1}\boldsymbol{q}_{\ell_{2,1}}$. Indeed, the application of the Hessian $\boldsymbol{Q}^{-1}$ is not influenced by the number of wavelengths of the input data thanks to the subspace approximation.

E. Reconstruction with $\ell_{2,1}$ approach

The second proposed algorithm aiming to minimize the semi-quadratically regularized criterion $\mathcal{J}_{\ell_{2,1}}$ of Eq.39, now called the $\ell_{2,1}$ approach, has also been tested for the reconstruction of abundance maps and its performance compared with the ℓ_2 approach.

The chosen convex non quadratic function φ for the penalization term in $\mathcal{J}_{\ell_{2,1}}$ is the Huber function, defined as

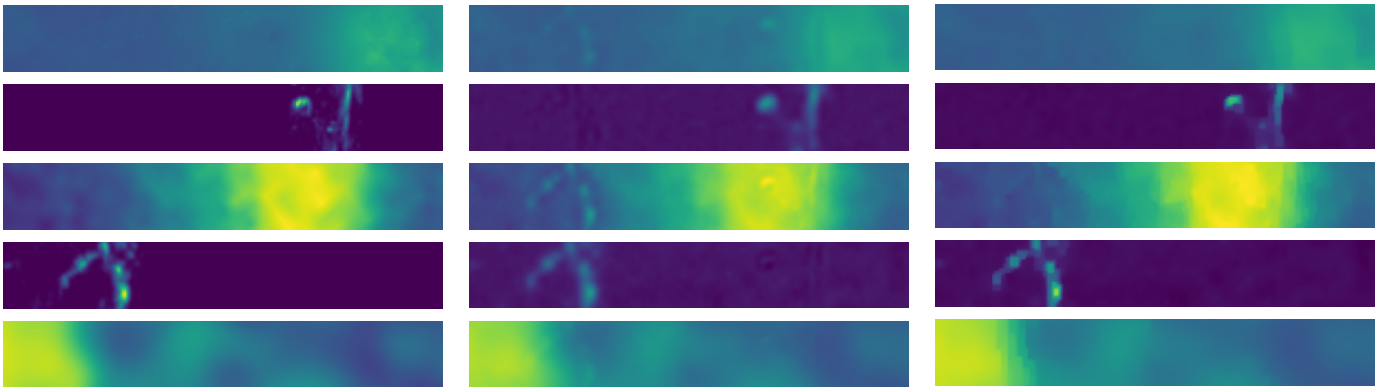
$$\varphi(\delta) = \begin{cases} \delta^2 & \text{if } |\delta| < \theta, \\ 2\theta|\delta| - \theta^2 & \text{otherwise,} \end{cases} \tag{57}$$

where $\theta \in \mathbb{R}_+$ is the threshold at which the Huber function switches from quadratic to linear. Thus, higher gradients are less penalized than with a quadratic regularization (ℓ_2 approach), which allows for a finer reconstruction of edges.

The $\ell_{2,1}$ approach requires setting two hyperparameters : the regularization parameter μ_{r}, and the Huber threshold θ. Again, hyperparameters were chosen to minimize $\mathrm{NRMSE}(\boldsymbol{x}, \widehat{\boldsymbol{x}}_{\ell_{2,1}})$,

Methods	NRMSE $\times 10^{-3}$	dSSIM $\times 10^{-5}$	SAM $\times 10^{-3}$	PSNR [dB]	Time [s]
Coaddition	130	1158	113	37	0.6
LCG [14] (equal time)	43	270	30	47	**0.05**
LCG [14] (after convergence)	3.2	**5.1**	**1.1**	**69**	367 (6 min)
Proposed explicit solution	**3.1**	**5.1**	**1.1**	**69**	**0.05**

TABLE I: Comparison of reconstruction quality measurements between the coaddition, LCG at an equal time with the proposed method and when it converged, and the proposed explicit method for the minimization of the quadratic criterion $\mathcal{J}_{\ell_2}$ defined at Eq. 20.



(a) Original abundance maps $\boldsymbol{a}$ (b) Exact minimum of $\mathcal{J}_{\ell_2}$ [14] (c) Minimum of $\mathcal{J}_{\ell_{2,1}}$

Fig. 6: Comparison between original abundance maps $\boldsymbol{a}$, on the left, and the ones reconstructed with the explicit minimization of $\mathcal{J}_{\ell_2}$ and with the fast minimization of $\mathcal{J}_{\ell_{2,1}}$, where $\text{SNR}_{\text{dB}} = 30$ dB for both MS and HS input datasets. Each of these abundance maps is associated in order with the spectra of Fig. 2. All maps share the same dynamic range.

where $\widehat{\boldsymbol{x}}_{\ell_{2,1}}$ is the hyperspectral cube reconstructed from the abundance maps $\widehat{\boldsymbol{a}}_{\ell_{2,1}}$.

Ideally, the reconstructed abundance maps should perfectly separate ("unmix") features with different spectra originally superposed in the instrument data. To verify the unmixing for both proposed approaches, the reconstructed maps from both ℓ_2 and $\ell_{2,1}$ approaches are shown in figure 6 for MS and HS datasets with a signal-to-noise ratio of 30 dB. Even though low-frequency features are well recovered with both approaches, the unmixing of high-frequency features (or edges) with the ℓ_2 approach is limited. In fact, a higher noise forces a higher quadratic regularization to stabilize the solution, with the side effect of strongly penalizing any high-frequency feature. Thus, the solution minimizing $\mathcal{J}_{\ell_2}$ in a high noise case are maps sharing these higher frequencies to avoid the presence of sharp edges. In contrast, the $\ell_{2,1}$ approach performs a visibly superior unmixing, and the recovery of high-frequency features is also enhanced.

Images from the cubes reconstructed from the abundance maps of figure 6 are shown on top of Fig. 8. Even if the ℓ_2 approach shows clear deblurring and denoising capabilities despite its inaccurate spectral unmixing, the $\ell_{2,1}$ approach performs a superior edge reconstruction, as expected. On top of an improved spatial reconstruction, the $\ell_{2,1}$ approach performed a significantly finer reconstruction of spectra at longer wavelengths than the ℓ_2 approach, especially in high gradient regions as shown figure 7. Errors on the reconstructed spectra with the coaddition method also tend to be inaccurate

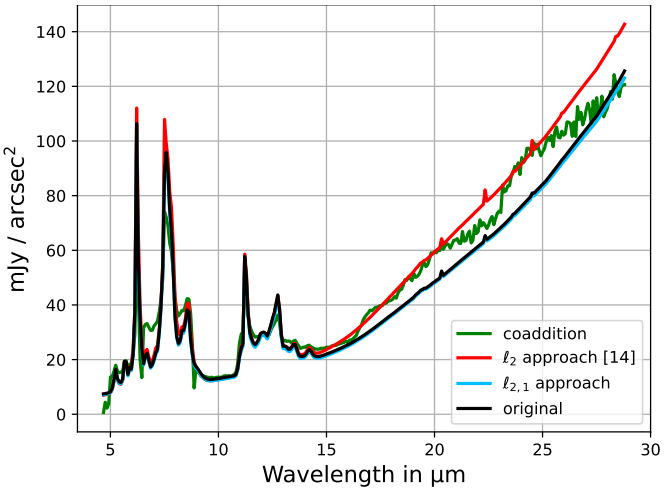
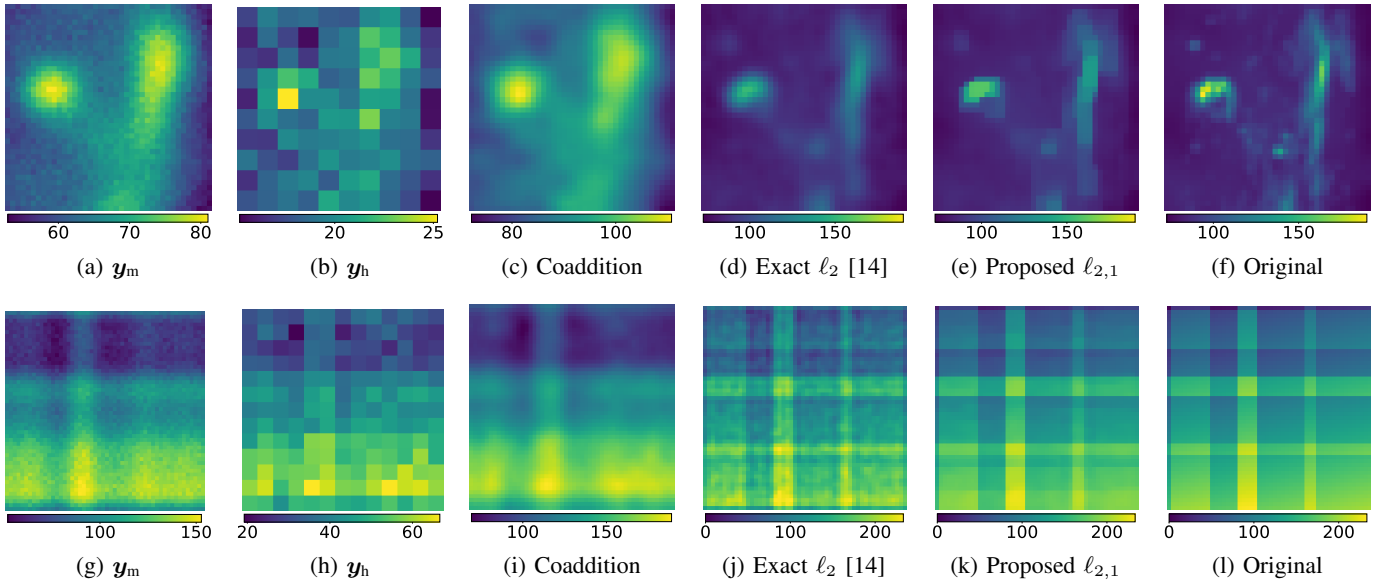


Fig. 7: Example of spectra reconstructed with the coaddition, the ℓ_2 [14] and the $\ell_{2,1}$ approaches for the same spatial position, taken in a high gradient region. SNR of input dataset: 30 dB.

at longer wavelengths, firstly due to the missing deconvolution, and secondly to the inversion of the low spectral response at long wavelengths, causing a noticeable noise effect.

Further experiments were conducted to compare the edge reconstruction capabilities of both ℓ_2 and $\ell_{2,1}$ approaches with the same imager and spectrometer model. Simulated noised MS and HS datasets made of crosses with sharp



(a) $\boldsymbol{y}_{\mathrm{m}}$ (b) $\boldsymbol{y}_{\mathrm{h}}$ (c) Coaddition (d) Exact ℓ_2 [14] (e) Proposed $\ell_{2,1}$ (f) Original

(g) $\boldsymbol{y}_{\mathrm{m}}$ (h) $\boldsymbol{y}_{\mathrm{h}}$ (i) Coaddition (j) Exact ℓ_2 [14] (k) Proposed $\ell_{2,1}$ (l) Original

Fig. 8: Imagery data $\boldsymbol{y}_{\mathrm{m}}$ and spectrometry data $\boldsymbol{y}_{\mathrm{h}}$ with $\mathrm{SNR}_{\mathrm{dB}}$ = 30 dB, images from cubes reconstructed with the explicit minimization of $\arg\min \mathcal{J}_{\ell_2}$ and fast minimization of $\arg\min \mathcal{J}_{\ell_{2,1}}$, and the original image, for simulated astronomical images (top) and images made of crosses and gradients (bottom). The central wavelength for all images is 25.5 µm. Unit is mJy arcsec^{-2}. Same dynamical scales for figures (d), (e), and (f), and for (j), (k), and (l).

Methods	NRMSE $\times 10^{-3}$	dSSIM $\times 10^{-5}$	SAM $\times 10^{-3}$	PSNR [dB]	Time [s]
Coaddition	133	1476	119	37	0.6
Exact solution of ℓ_2 [14]	27	241	5.8	50	**0.05**
Proposed $\ell_{2,1}$ approach	**22**	**179**	**4.0**	**52**	19 (300 iter.)

TABLE II: Comparison of reconstruction quality measurements between the coaddition, the ℓ_2 and the $\ell_{2,1}$ approaches for input MS and HS datasets with $\mathrm{SNR}_{\mathrm{dB}}$ = 30 dB.

edges and smooth gradients were created, and the resulting reconstructions are shown at the bottom of Fig. 8. The optimal reconstruction with the ℓ_2 approach is noisy, and the edges are poorly reconstructed. Conversely, the $\ell_{2,1}$ approach still demonstrates high denoising and edge-preserving capabilities (see Tab. II for a comparison of reconstruction quality measurements).

Compared to the iterative gradient-based ℓ_2 approach from the literature [14], the proposed $\ell_{2,1}$ approach is superior in both quality of reconstruction *and* computational time. Indeed, for every case tested in this work, the output of the proposed approach at any given computational time had a lower NRMSE to the true object than the output obtained with the literature.

However, the $\ell_{2,1}$ approach can present some limitations compared to the ℓ_2 approach. Indeed, it is iterative, thus necessarily slower, and requires setting convergence thresholds. Moreover, it involves setting 2 hyperparameters (μ_{r} and θ) instead of 1 for the ℓ_2 approach (μ_{r}), which complicates the search for an optimal solution.

As demonstrated in this study, the ℓ_2 approach could be suited for the fusion of low-noise datasets or datasets mainly composed of low-frequency features. Other situations would require using the $\ell_{2,1}$ approach for an improved reconstruction of high-frequency features, superior spectral unmixing, and resilience to noise.

F. Comparison with neural network approaches

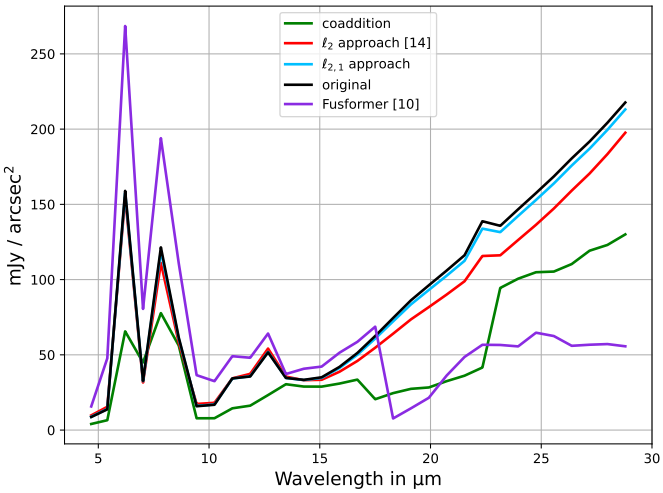
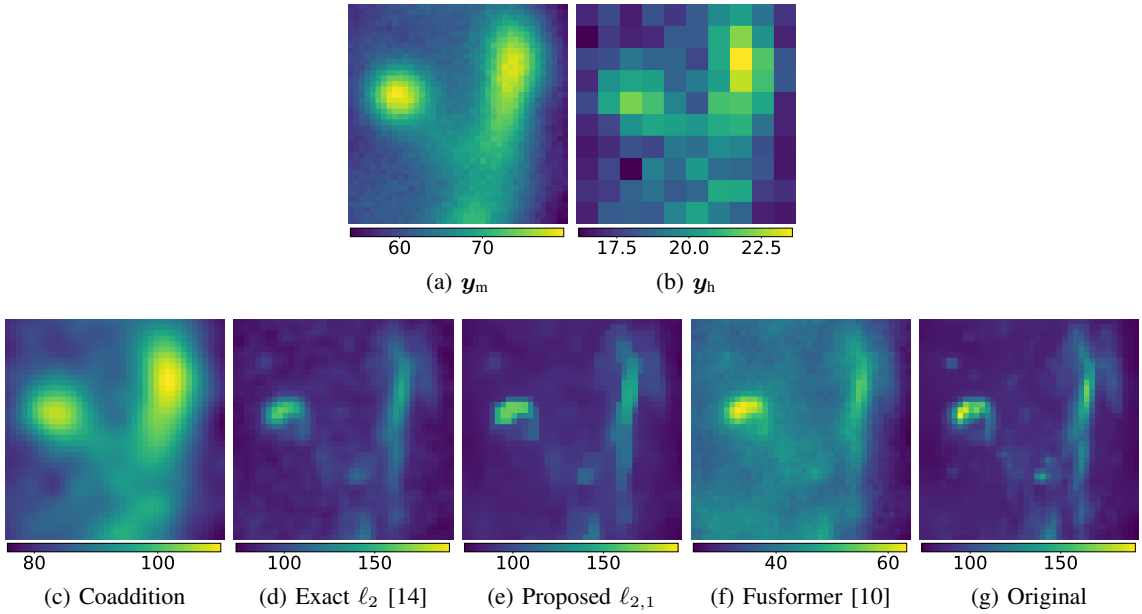


Fig. 9: Example of spectra reconstructed with the coaddition, the ℓ_2 [14] and the $\ell_{2,1}$ approaches, and Fusformer [10] for the same spatial position, taken in a high gradient region. SNR of input datasets: 40 dB.



(a) $\boldsymbol{y}_{\mathrm{m}}$ (b) $\boldsymbol{y}_{\mathrm{h}}$ (c) Coaddition (d) Exact ℓ_2 [14] (e) Proposed $\ell_{2,1}$ (f) Fusformer [10] (g) Original

Fig. 10: Imagery data $\boldsymbol{y}_{\mathrm{m}}$ and spectrometry data $\boldsymbol{y}_{\mathrm{h}}$ with $\mathrm{SNR}_{\mathrm{dB}} = 40$ dB on top, and below are reconstructions with the coaddition method, the ℓ_2 and $\ell_{2,1}$ approaches, Fusformer, and the original image. All results were produced from the detailed reduced MS and HS datasets. The central wavelength for all images is 25.5 µm. Unit is mJy arcsec^{-2}. Same dynamical scale for figures (d), (e), and (g).

The study was extended with a comparison with neural network approaches as they have been increasingly used for MS/HS fusion. As demonstrated in [10], the self-attention feature of a transformer embedded in a greater neural network architecture allowed it to outperform other known methods of the literature in terms of quality of reconstruction. Indeed, the transformer enabled to inspect the global relationships between features, whereas convolution kernels, for example in CNN-based methods, could only account for local relationships of features [10]. This architecture, called Fusformer, was compared with the proposed inverse problem approaches for the fusion of astronomical data.

A trained model of Fusformer provided by its authors was used for this purpose. New astronomical datasets were created to satisfy the model requirements on the input dimensions : 3 $\times$ 48 $\times$ 48 pixels for the MS dataset, and 31 $\times$ 12 $\times$ 12 for the HS dataset. These reduced datasets were noised so their SNR is 40 dB, and were the ones used as inputs for all compared methods. Moreover, a normalization of these input datasets was required to set their values between 0 and 1. Then, the output cube of the network was denormalized such that its boundary values were the same as for the input datasets.

Images of the reconstructed cubes with the coaddition, ℓ_2 and $\ell_{2,1}$ approaches and Fusformer are shown Fig. 10. It can be seen that Fusformer successfully recovered some high-frequency features, which is assumed related to the training of the network in the high-pass domain [10]. However, low frequencies were not well recovered, i.e. background values are too intense relative to high-frequency features compared to the original image. Furthermore, a substantial drawback of this network (and most deep neural networks) is that output values are intrinsically bounded between 0 and 1 before denormalization, yet such inverse or deconvolution problems usually lead to restoring unknown values of flux potentially falling outside of a predefined interval. Consequently, naive denormalization cannot restore correct values of flux, as seen in Fig. 10f. Therefore, even though spatial information can be well recovered with this method, it turns inadequate for spectra reconstruction, as shown with reconstructed spectra of Fig. 9.

V. Conclusion

Typical imager and spectrometer models have been described considering a wavelength-dependent spatial blur, different spectral responses, spectral integrations for the imager, and spatial integrations and subsampling for the spectrometer. Efforts were made to extend the applicability of these models such that the impulse response can be different between the imager and spectrometer model, and such that the wideband images themselves originate from different imager models.

The data fusion problem was first posed through the formulation of a regularized least squares criterion, for which a calculable explicit solution has been demonstrated by exploiting an efficient way to calculate the Hessian matrix. Then, the problem was formulated with an edge-preserving semi-quadratic regularization. This latter problem involved an alternating minimization problem, which has been accelerated with the same Hessian inversion procedure.

These data fusion methods have been applied to simulated astronomical images from the James Webb Space Telescope. The benefits of the explicit solution for the exact minimization of the quadratic criterion against an iterative approach have been proven with a time-saving factor of 7 000 for the fusion of low noise data ($\mathrm{SNR}_{\mathrm{dB}}$ = 100 dB).

A comparison of the approaches with quadratic ℓ_2 and semi-quadratic $\ell_{2,1}$ regularizations has also been carried out and highlighted the improved ability of the latter approach to unmix spectral information, to better reconstruct spatial information (especially in high gradient areas) and spectra, which also led to superior denoising capabilities.

Finally, a performance comparison with a deep neural network architecture with an embedded transformer trained for the fusion of MS and HS datasets has been conducted and led to the conclusion that the proposed inverse problem approaches outperformed such a method in this context of data fusion.

Further work could include estimations of optimal values of hyperparameters to develop a non-supervised reconstruction procedure. This would result in drastic time savings as iterative searches for optimal hyperparameter values would be avoided. Other works could be conducted to improve the quality of the reconstruction, for example by imposing a positivity constraint, or by using wavelet or deep neural network-based regularizations.

REFERENCES

[1] N. Yokoya *et al.*, "Hyperspectral and multispectral data fusion: A comparative review of the recent literature," *IEEE Geoscience and Remote Sensing Magazine*, vol. 5, no. 2, pp. 29–56, 2017.

[2] N. Yokoya, T. Yairi, and A. Iwasaki, "Coupled nonnegative matrix factorization unmixing for hyperspectral and multispectral data fusion," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 50, no. 2, pp. 528–537, 2012.

[3] R. Dian and S. Li, "Hyperspectral image super-resolution via subspace-based low tensor multi-rank regularization," *IEEE Transactions on Image Processing*, vol. 28, no. 10, pp. 5135–5146, 2019.

[4] J. Liu, Z. Wu, L. Xiao, J. Sun, and H. Yan, "A truncated matrix decomposition for hyperspectral image super-resolution," *IEEE Transactions on Image Processing*, vol. 29, pp. 8028–8042, 2020.

[5] C. Prévost, K. Usevich, P. Comon, and D. Brie, "Hyperspectral super-resolution with coupled tucker approximation: Recoverability and svd-based algorithms," *IEEE Transactions on Signal Processing*, vol. 68, pp. 931–946, 2020.

[6] R. A. Borsoi, C. Prévost, K. Usevich, D. Brie, J. C. Bermudez, and C. Richard, "Coupled tensor decomposition for hyperspectral and multispectral image fusion with inter-image variability," *IEEE Journal of Selected Topics in Signal Processing*, vol. 15, no. 3, pp. 702–717, 2021.

[7] F. Palsson, J. R. Sveinsson, and M. O. Ulfarsson, "Multispectral and hyperspectral image fusion using a 3-d-convolutional neural network," *IEEE Geoscience and Remote Sensing Letters*, vol. 14, no. 5, pp. 639–643, 2017.

[8] Q. Xie, M. Zhou, Q. Zhao, D. Meng, W. Zuo, and Z. Xu, "Multispectral and hyperspectral image fusion by ms/hs fusion net," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019, pp. 1585–1594.

[9] J.-F. Hu, T.-Z. Huang, L.-J. Deng, T.-X. Jiang, G. Vivone, and J. Chanussot, "Hyperspectral image super-resolution via deep spatiospectral attention convolutional neural networks," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 33, no. 12, pp. 7251–7265, 2021.

[10] J.-F. Hu, T.-Z. Huang, L.-J. Deng, H.-X. Dou, D. Hong, and G. Vivone, "Fusformer: A transformer-based fusion network for hyperspectral image super-resolution," *IEEE Geoscience and Remote Sensing Letters*, vol. 19, pp. 1–5, 2022.

[11] Y. Zhang, S. De Backer, and P. Scheunders, "Noise-resistant wavelet-based bayesian fusion of multispectral and hyperspectral images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 47, no. 11, pp. 3834–3843, 2009.

[12] M. Simoes, J. Bioucas-Dias, L. B. Almeida, and J. Chanussot, "A convex formulation for hyperspectral image superresolution via subspace-based regularization," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 53, no. 6, pp. 3373–3388, 2014.

[13] Q. Wei *et al.*, "Fast fusion of multi-band images based on solving a sylvester equation," *IEEE Transactions on Image Processing*, vol. 24, no. 11, pp. 4109–4121, 2015.

[14] C. Guilloteau *et al.*, "Hyperspectral and multispectral image fusion under spectrally varying spatial blurs–application to high dimensional infrared astronomical imaging," *IEEE Transactions on Computational Imaging*, vol. 6, pp. 1362–1374, 2020.

[15] R. Abi-Rizk *et al.*, "Super-resolution hyperspectral reconstruction with majorization-minimization algorithm and low-rank approximation," *IEEE Transactions on Computational Imaging*, vol. 8, pp. 260–272, 2022.

[16] E. Chouzenoux, J. Idier, and S. Moussaoui, "A majorize–minimize strategy for subspace optimization applied to image restoration," *IEEE Transactions on Image Processing*, vol. 20, no. 6, pp. 1517–1528, 2011.

[17] M. E. Hadj-Youcef, "Spatio spectral reconstruction from low resolution multispectral data: application to the mid-infrared instrument of the james webb space telescope," Ph.D. dissertation, Université Paris-Saclay (ComUE), 2018.

[18] B. Hunt, "A matrix theory proof of the discrete convolution theorem," *IEEE Transactions on Audio and Electroacoustics*, vol. 19, no. 4, pp. 285–288, 1971.

[19] J. Idier, "Convex half-quadratic criteria and interacting auxiliary variables for image restoration," *IEEE transactions on image processing*, vol. 10, no. 7, pp. 1001–1009, 2001.

[20] D. Geman and C. Yang, "Nonlinear image recovery with half-quadratic regularization," *IEEE transactions on Image Processing*, vol. 4, no. 7, pp. 932–946, 1995.

[21] STScI, "JWST documentation website," https://jwst-sdocs.stsci.edu, accessed: 2023-03-06.

[22] P. Bouchet *et al.*, "The mid-infrared instrument for the james webb space telescope, iii: Mirim, the miri imager," *Publications of the Astronomical Society of the Pacific*, vol. 127, no. 953, p. 612, 2015.

[23] M. Wells *et al.*, "The mid-infrared instrument for the james webb space telescope, vi: The medium resolution spectrometer," *Publications of the Astronomical Society of the Pacific*, vol. 127, no. 953, p. 646, 2015.

[24] M. D. Perrin *et al.*, "Simulating point spread functions for the james webb space telescope with webbpsf," in *Space Telescopes and Instrumentation 2012: Optical, Infrared, and Millimeter Wave*, vol. 8442. SPIE, 2012, pp. 1193–1203.

[25] C. Guilloteau *et al.*, "Simulated jwst data sets for multispectral and hyperspectral image fusion," *The Astronomical Journal*, vol. 160, no. 1, p. 28, 2020.

[26] E. Habart *et al.*, "High angular resolution near-ir view of the Orion bar revealed by keck/nirc2," *arXiv preprint arXiv:2206.08245*, 2022.

[27] P. Team, O. Berné, *et al.*, "Pdrs4all: A jwst early release science program on radiative feedback from massive stars," *Publications of the Astronomical Society of the Pacific*, vol. 134, no. 1035, p. 054301, 2022.

[28] J. R. Fienup, "Invariant error metrics for image reconstruction," *Appl. Opt.*, vol. 36, no. 32, pp. 8352–8357, Nov 1997. [Online]. Available: https://opg.optica.org/ao/abstract.cfm?URI=ao-s36-s32-s8352

[29] Z. Wang *et al.*, "Image quality assessment: from error visibility to structural similarity," *IEEE transactions on image processing*, vol. 13, no. 4, pp. 600–612, 2004.

[30] R. H. Yuhas *et al.*, "Discrimination among semi-arid landscape endmembers using the spectral angle mapper (sam) algorithm," in *JPL, Summaries of the Third Annual JPL Airborne Geoscience Workshop. Volume 1: AVIRIS Workshop*, 1992.

[31] N. Instruments, "Peak signal-to-noise ratio as an image quality metric," 2013.

[32] J. Nocedal and S. J. Wright, *Numerical Optimization*, 2nd ed., ser. Springer Series in Operations Research. Springer, 2006.

[33] M. A. Hadj-Youcef *et al.*, "Fast joint multiband reconstruction from wideband images based on low-rank approximation," *IEEE Transactions on Computational Imaging*, vol. 6, pp. 922–933, 2020.

ACKNOWLEDGMENTS

This work is funded by the Centre National d'Études Spatiales (CNES) and the Agence Nationale de la Recherche (ANR), and was supported by a collaboration with ACRI-ST as part of the INCLASS joint laboratory.

APPENDIX A
PROOF OF $\boldsymbol{Q}_{\ell_2}$ INVERSION

We consider a square matrix made of diagonal blocks $\boldsymbol{\Delta}_{t,t'}$

$$\begin{pmatrix} \boldsymbol{\Delta}_{1,1} & \dots & \boldsymbol{\Delta}_{1,T} \\ \vdots & & \vdots \\ \boldsymbol{\Delta}_{T,1} & \dots & \boldsymbol{\Delta}_{T,T} \end{pmatrix}. \tag{58}$$

Let $\boldsymbol{P}$ be a permutation matrix such that

$$\begin{pmatrix} \boldsymbol{\Delta}_{1,1} & \dots & \boldsymbol{\Delta}_{1,T} \\ \vdots & & \vdots \\ \boldsymbol{\Delta}_{T,1} & \dots & \boldsymbol{\Delta}_{T,T} \end{pmatrix} = \boldsymbol{P} \begin{pmatrix} \boldsymbol{\Delta}'_1 & & \\ & \ddots & \\ & & \boldsymbol{\Delta}'_P \end{pmatrix} \boldsymbol{P} \tag{59}$$

where the blocks $\boldsymbol{\Delta}_{t,t'}$ are diagonal, and the blocks $\boldsymbol{\Delta}'_p$, $p \in [1, P]$, are matrices of size $T \times T$, built so that the coefficient of the row t and column t' is the p^{th} coefficient of the diagonal of $\boldsymbol{\Delta}_{t,t'}$

$$\left(\boldsymbol{\Delta}'_p\right)_{t,t'} = \left(\boldsymbol{\Delta}_{t,t'}\right)_{p,p}.$$

Such permutations allow for the creation of a new block diagonal matrix where all Fourier coefficients have been sorted by frequency (one block per frequency). As done in [33], this matrix is directly invertible. Using the orthogonality of permutations matrices, i.e. $\boldsymbol{P}^{-1} = \boldsymbol{P}^T$, the inversion of Eq. 59 writes as

$$\begin{pmatrix} \boldsymbol{\Delta}_{1,1} & \dots & \boldsymbol{\Delta}_{1,T} \\ \vdots & & \vdots \\ \boldsymbol{\Delta}_{T,1} & \dots & \boldsymbol{\Delta}_{T,T} \end{pmatrix}^{-1} \tag{60a}$$

$$= \boldsymbol{P}^{-1} \begin{pmatrix} \boldsymbol{\Delta}'_1 & & \\ & \ddots & \\ & & \boldsymbol{\Delta}'_P \end{pmatrix}^{-1} \boldsymbol{P}^{-1} \tag{60b}$$

$$= \boldsymbol{P}^{-1} \begin{pmatrix} (\boldsymbol{\Delta}'_1)^{-1} & & \\ & \ddots & \\ & & (\boldsymbol{\Delta}'_P)^{-1} \end{pmatrix} \boldsymbol{P}^{-1} \tag{60c}$$

$$= \boldsymbol{P}^T \begin{pmatrix} \boldsymbol{\delta}'_1 & & \\ & \ddots & \\ & & \boldsymbol{\delta}'_P \end{pmatrix} \boldsymbol{P}^T \tag{60d}$$

$$= \begin{pmatrix} \boldsymbol{\delta}_{1,1} & \dots & \boldsymbol{\delta}_{1,T} \\ \vdots & & \vdots \\ \boldsymbol{\delta}_{T,1} & \dots & \boldsymbol{\delta}_{T,T} \end{pmatrix} \tag{60e}$$

where $\boldsymbol{\delta}_{t,t'}$ have the same size and structure than $\boldsymbol{\Delta}_{t,t'}$, i.e. diagonal matrices of size $P \times P$.

The inversion of $\boldsymbol{Q}_{\ell_2}$ is then calculated as

$$\boldsymbol{Q}_{\ell_2}^{-1} = \left(\overline{\boldsymbol{F}}^H \begin{pmatrix} \boldsymbol{\Delta}_{1,1} & \dots & \boldsymbol{\Delta}_{1,T} \\ \vdots & & \vdots \\ \boldsymbol{\Delta}_{T,1} & \dots & \boldsymbol{\Delta}_{T,T} \end{pmatrix} \overline{\boldsymbol{F}} \right)^{-1} \tag{61a}$$

$$= \overline{\boldsymbol{F}}^H \begin{pmatrix} \boldsymbol{\delta}_{1,1} & \dots & \boldsymbol{\delta}_{1,T} \\ \vdots & & \vdots \\ \boldsymbol{\delta}_{T,1} & \dots & \boldsymbol{\delta}_{T,T} \end{pmatrix} \overline{\boldsymbol{F}} \tag{61b}$$

which proves that blocks of $\boldsymbol{Q}_{\ell_2}^{-1}$ indeed follow a diagonal structure in the Fourier space, as for $\boldsymbol{Q}_{\ell_2}$. This property allows for fast computation and application of $\boldsymbol{Q}_{\ell_2}^{-1}$ on $\boldsymbol{q}_{\ell_2}$ in the Fourier space to explicitly obtain the minimizer of $\mathcal{J}_{\ell_2}$ with Eq. 22.

Therefore, the inversion of the initial matrix $\boldsymbol{Q}_{\ell_2}$ require T^2 `fft` to diagonalize the blocks, and P inversion of matrix of size $T \times T$. In the case of this paper, P is the number of pixels, and T is the number of spectral templates. Such configuration can be solved efficiently by an actual personal computer and linear algebra libraries.