Fast Fusion of Multi-Band Images Based on Solving a Sylvester Equation

Qi Wei, *Student Member, IEEE*, Nicolas Dobigeon, *Senior Member, IEEE*,
and Jean-Yves Tourneret, *Senior Member, IEEE*

***Abstract*— This paper proposes a fast multi-band image fusion algorithm, which combines a high-spatial low-spectral resolution image and a low-spatial high-spectral resolution image. The well admitted forward model is explored to form the likelihoods of the observations. Maximizing the likelihoods leads to solving a Sylvester equation. By exploiting the properties of the circulant and downsampling matrices associated with the fusion problem, a closed-form solution for the corresponding Sylvester equation is obtained explicitly, getting rid of any iterative update step. Coupled with the alternating direction method of multipliers and the block coordinate descent method, the proposed algorithm can be easily generalized to incorporate prior information for the fusion problem, allowing a Bayesian estimator. Simulation results show that the proposed algorithm achieves the same performance as the existing algorithms with the advantage of significantly decreasing the computational complexity of these algorithms.**

***Index Terms*— Multi-band image fusion, Bayesian estimation, circulant matrix, Sylvester equation, alternating direction method of multipliers, block coordinate descent.**

I. Introduction

A. Background

IN GENERAL, a multi-band image can be represented as a 3D data cube indexed by three exploratory variables (x, y, λ), where x and y are the two spatial dimensions of the scene, and λ is the spectral dimension (covering a range of wavelengths). Typical examples of multi-band images include hyperspectral (HS) images [1], multi-spectral (MS) images [2], integral field spectrographs [3], magnetic resonance spectroscopy images etc. However, multi-band imaging generally suffers from the limited spatial resolution of the data acquisition devices, mainly due to an unsurpassable tradeoff between spatial and spectral sensitivities [4]. For example, HS images benefit from excellent spectroscopic properties with hundreds of bands but are limited by their relatively low spatial resolution compared with MS and panchromatic (PAN) images (which are acquired in much fewer bands). As a consequence, reconstructing a high-spatial and high-spectral multi-band image from two degraded and complementary observed images is a challenging but crucial issue that has been addressed in various scenarios [5]–[8]. In particular, fusing a high-spatial low-spectral resolution image and a low-spatial high-spectral image is an archetypal instance of multi-band image reconstruction, such as pansharpening (MS+PAN) [9] or hyperspectral pansharpening (HS+PAN) [10]. Generally, the linear degradations applied to the observed images with respect to (w.r.t.) the target high-spatial and high-spectral image reduce to spatial and spectral transformations. Thus, the multi-band image fusion problem can be interpreted as restoring a 3D data-cube from two degraded data-cubes. A more precise description of the problem formulation is provided in the following paragraph.

B. Problem Statement

To better distinguish spectral and spatial degradations, the pixels of the target multi-band image, which is of high-spatial and high-spectral resolution, can be rearranged to build an $m_\lambda \times n$ matrix $\mathbf{X}$, where m_λ is the number of spectral bands and $n = n_r \times n_c$ is the number of pixels in each band (n_r and n_c represents the number of rows and columns respectively). In other words, each column of the matrix $\mathbf{X}$ consists of a m_λ-valued pixel and each row gathers all the pixel values in a given spectral band. Based on this pixel ordering, any linear operation applied on the left (resp. right) side of $\mathbf{X}$ describes a spectral (resp. spatial) degradation.

In this work, we assume that two complementary images of high-spectral or high-spatial resolutions, respectively, are available to reconstruct the target high-spectral and high-spatial resolution target image. These images result from linear spectral and spatial degradations of the full resolution image $\mathbf{X}$, according to the well-admitted model

$$\begin{aligned} \mathbf{Y}_\mathrm{L} &= \mathbf{L}\mathbf{X} + \mathbf{N}_\mathrm{L} \\ \mathbf{Y}_\mathrm{R} &= \mathbf{X}\mathbf{R} + \mathbf{N}_\mathrm{R} \end{aligned} \tag{1}$$

where

- $\mathbf{X} = [\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n] \in \mathbb{R}^{m_\lambda \times n}$ is the full resolution target image,
- $\mathbf{Y}_\mathrm{L} \in \mathbb{R}^{n_\lambda \times n}$ and $\mathbf{Y}_\mathrm{R} \in \mathbb{R}^{m_\lambda \times m}$ are the observed spectrally degraded and spatially degraded images,
- $m = m_r \times m_c$ is the number of pixels of the high-spectral resolution image,

- n_λ is the number of bands of the high-spatial resolution image,
- $\mathbf{N}_\mathrm{L}$ and $\mathbf{N}_\mathrm{R}$ are additive terms that include both modeling errors and sensor noises.

The noise matrices are assumed to be distributed according to the following matrix normal distributions[1]

$$\begin{aligned}\mathbf{N}_\mathrm{L} &\sim \mathcal{MN}_{m_\lambda,m}(\mathbf{0}_{m_\lambda,m}, \mathbf{\Lambda}_\mathrm{L}, \mathbf{I}_m)\\ \mathbf{N}_\mathrm{R} &\sim \mathcal{MN}_{n_\lambda,n}(\mathbf{0}_{n_\lambda,n}, \mathbf{\Lambda}_\mathrm{R}, \mathbf{I}_n).\end{aligned}$$

Note that no particular structure is assumed for the row covariance matrices $\mathbf{\Lambda}_\mathrm{L}$ and $\mathbf{\Lambda}_\mathrm{R}$ except that they are both positive definite, which allows for considering spectrally colored noises. Conversely, the column covariance matrices are assumed to be the identity matrix to reflect the fact that the noise is pixel-independent. In practice, $\mathbf{\Lambda}_\mathrm{L}$ and $\mathbf{\Lambda}_\mathrm{R}$ depend on the sensor characteristics and can be known or learnt using cross-calibration. To simplify the problem, $\mathbf{\Lambda}_\mathrm{L}$ and $\mathbf{\Lambda}_\mathrm{R}$ are often assumed to be diagonal matrices, where the ith diagonal element is the noise variance in the ith band. Thus, the number of variables in $\mathbf{\Lambda}_\mathrm{L}$ is decreased from $\frac{n_\lambda(n_\lambda+1)}{2}$ to n_λ. Similar results hold for $\mathbf{\Lambda}_\mathrm{R}$. Furthermore, if we want to ignore the noise terms $\mathbf{N}_\mathrm{L}$ and $\mathbf{N}_\mathrm{R}$, which means the noises of $\mathbf{Y}_\mathrm{L}$ and $\mathbf{Y}_\mathrm{R}$ are both trivial for fusion, we can simply set $\mathbf{\Lambda}_\mathrm{L}$ and $\mathbf{\Lambda}_\mathrm{R}$ to identity matrices as in [10].

In most practical scenarios, the spectral degradation $\mathbf{L} \in \mathbb{R}^{n_\lambda \times m_\lambda}$ only depends on the spectral response of the sensor, which can be *a priori* known or estimated by cross-calibration [11]. The spatial degradation $\mathbf{R}$ includes warp, translation, blurring, decimation, etc. As the warp and translation can be attributed to the image co-registration problem and mitigated by precorrection, only blurring and decimation degradations, denoted $\mathbf{B}$ and $\mathbf{S}$ are considered in this work. If the spatial blurring is assumed to be space-invariant, $\mathbf{B} \in \mathbb{R}^{n\times n}$ owns the specific property of being a cyclic convolution operator acting on the bands. The matrix $\mathbf{S} \in \mathbb{R}^{n\times m}$ is a $d = d_r \times d_c$ uniform downsampling operator, which has $m = n/d$ ones on the block diagonal and zeros elsewhere, and such that $\mathbf{S}^T\mathbf{S} = \mathbf{I}_m$. Note that multiplying by $\mathbf{S}^T$ represents zero-interpolation to increase the number of pixels from m to n. Therefore, assuming $\mathbf{R}$ can be decomposed as $\mathbf{R} = \mathbf{BS} \in \mathbb{R}^{n\times m}$, the fusion model (1) can be rewritten as

$$\begin{aligned}\mathbf{Y}_\mathrm{L} &= \mathbf{LX} + \mathbf{N}_\mathrm{L}\\ \mathbf{Y}_\mathrm{R} &= \mathbf{XBS} + \mathbf{N}_\mathrm{R}\end{aligned} \tag{2}$$

where all matrix dimensions and their respective relations are summarized in Table I.

This matrix equation (1) has been widely advocated in the pansharpening and HS pansharpening problems, which consist of fusing a PAN image with an MS or an HS image [10], [12], [13]. Similarly, most of the techniques developed to fuse MS and HS images also rely on a similar linear model [14]–[20]. From an application point of view, this problem is also important as motivated by recent national programs, e.g., the Japanese next-generation space-borne HS image suite (HISUI), which fuses co-registered MS and HS images acquired over the same scene under the same conditions [21].

TABLE I
NOTATIONS

Notation	Definition	Relation
m_r	row number of spatially degraded image	$m_r = n_r/d_r$
m_c	column number of spatially degraded image	$m_c = n_c/d_c$
m	number of pixels in each band of $\mathbf{Y}_\mathrm{R}$	$m = m_r \times m_c$
n_r	row number of spectrally degraded image	$n_r = m_r \times d_r$
n_c	column number of spectrally degraded image	$n_c = m_c \times d_c$
n	number of pixels in each band of $\mathbf{Y}_\mathrm{L}$	$n = n_r \times n_c$
d_r	decimation factor in row	$d_r = n_r/m_r$
d_c	decimation factor in column	$d_c = n_c/m_c$
d	decimation factor	$d = d_r \times d_c$

To summarize, the problem of fusing high-spectral and high-spatial resolution images can be formulated as estimating the unknown matrix $\mathbf{X}$ from (2). There are two main statistical estimation methods that can be used to solve this problem. These methods are based on maximum likelihood (ML) or on Bayesian inference. ML estimation is purely data-driven while Bayesian estimation relies on prior information, which can be regarded as a regularization (or a penalization) for the fusion problem. Various priors have been already advocated to regularize the multi-band image fusion problem, such as Gaussian priors [22], [23], sparse representations [20] or total variation (TV) [24] priors. The choice of the prior usually depends on the information resulting from previous experiments or from a subjective view of constraints affecting the unknown model parameters [25], [26].

Computing the ML or the Bayesian estimators (whatever the form chosen for the prior) is a challenging task, mainly due to the large size of $\mathbf{X}$ and to the presence of the downsampling operator $\mathbf{S}$, which prevents any direct use of the Fourier transform to diagonalize the blurring operator $\mathbf{B}$. To overcome this difficulty, several computational strategies have been designed to approximate the estimators. Based on a Gaussian prior modeling, a Markov chain Monte Carlo (MCMC) algorithm has been implemented in [22] to generate a collection of samples asymptotically distributed according to the posterior distribution of $\mathbf{X}$. The Bayesian estimators of $\mathbf{X}$ can then be approximated using these samples. Despite this formal appeal, MCMC-based methods have the major drawback of being computationally expensive, which prevents their effective use when processing images of large size. Relying on exactly the same prior model, the strategy developed in [23] exploits an alternating direction method of multipliers (ADMM) embedded in a block coordinate descent method (BCD) to compute the maximum a posterior (MAP) estimator of $\mathbf{X}$. This optimization strategy allows the numerical complexity to be greatly decreased when compared to its MCMC counterpart. Based on a prior built from a sparse representation, the fusion problem is solved in [20] and [24] with the split augmented Lagrangian shrinkage algorithm (SALSA) [27], which is an instance of ADMM.

[1]The probability density function $p(\mathbf{X}|\mathbf{M}, \mathbf{\Sigma}_r, \mathbf{\Sigma}_c)$ of a matrix normal distribution $\mathcal{MN}_{r,c}(\mathbf{M}, \mathbf{\Sigma}_r, \mathbf{\Sigma}_c)$ is defined by

$$p(\mathbf{X}|\mathbf{M}, \mathbf{\Sigma}_r, \mathbf{\Sigma}_c) = \frac{\exp\left(-\frac{1}{2}\mathrm{tr}\left[\mathbf{\Sigma}_c^{-1}(\mathbf{X}-\mathbf{M})^T\mathbf{\Sigma}_r^{-1}(\mathbf{X}-\mathbf{M})\right]\right)}{(2\pi)^{rc/2}|\mathbf{\Sigma}_c|^{r/2}|\mathbf{\Sigma}_r|^{c/2}}$$

where $\mathbf{M} \in \mathbb{R}^{r\times c}$ is the mean matrix, $\mathbf{\Sigma}_r \in \mathbb{R}^{r\times r}$ is the row covariance matrix and $\mathbf{\Sigma}_c \in \mathbb{R}^{c\times c}$ is the column covariance matrix.

In this paper, contrary to the algorithms described above, a much more efficient method is proposed to solve explicitly an underlying Sylvester equation (SE) associated with the fusion problem derived from (2), leading to an algorithm referred to as Fast fUsion based on Sylvester Equation (FUSE). This algorithm can be implemented *per se* to compute the ML estimator in a computationally efficient manner. The proposed FUSE algorithm has also the great advantage of being easily generalizable within a Bayesian framework when considering various priors. The MAP estimators associated with a Gaussian prior similar to [22] and [23] can be directly computed thanks to the proposed strategy. When handling more complex priors such as those used in [20] and [24], the FUSE solution can be conveniently embedded within a conventional ADMM or a BCD algorithm.

C. Paper Organization

The remaining of this paper is organized as follows. Section II studies the optimization problem to be addressed in absence of any regularization, i.e., in an ML framework. The proposed fast fusion method is presented in Section III and generalized to Bayesian estimators associated with various priors in Section IV. Section V presents experimental results assessing the accuracy and the numerical efficiency of the proposed fusion method. Conclusions are finally reported in Section VI.

II. Problem Formulation

Using the statistical properties of the noise matrices $\mathbf{N}_{\mathrm{L}}$ and $\mathbf{N}_{\mathrm{R}}$, $\mathbf{Y}_{\mathrm{L}}$ and $\mathbf{Y}_{\mathrm{R}}$ have matrix Gaussian distributions, i.e.,

$$\begin{aligned} p\left(\mathbf{Y}_{\mathrm{L}}|\mathbf{X}\right) &= \mathcal{MN}_{n_\lambda,n}(\mathbf{LX}, \mathbf{\Lambda}_{\mathrm{L}}, \mathbf{I}_n) \\ p\left(\mathbf{Y}_{\mathrm{R}}|\mathbf{X}\right) &= \mathcal{MN}_{m_\lambda,m}(\mathbf{XBS}, \mathbf{\Lambda}_{\mathrm{R}}, \mathbf{I}_m). \end{aligned} \tag{3}$$

As the collected measurements $\mathbf{Y}_{\mathrm{L}}$ and $\mathbf{Y}_{\mathrm{R}}$ have been acquired by different (possibly heterogeneous) sensors, the noise matrices $\mathbf{N}_{\mathrm{L}}$ and $\mathbf{N}_{\mathrm{R}}$ are sensor-dependent and can be generally assumed to be statistically independent. Therefore, $\mathbf{Y}_{\mathrm{L}}$ and $\mathbf{Y}_{\mathrm{R}}$ are independent conditionally upon the unobserved scene $\mathbf{X} = [\boldsymbol{x}_1, \cdots, \boldsymbol{x}_n]$. As a consequence, the joint likelihood function of the observed data is

$$p\left(\mathbf{Y}_{\mathrm{L}}, \mathbf{Y}_{\mathrm{R}}|\mathbf{X}\right) = p\left(\mathbf{Y}_{\mathrm{L}}|\mathbf{X}\right) p\left(\mathbf{Y}_{\mathrm{R}}|\mathbf{X}\right). \tag{4}$$

Since adjacent HS bands are known to be highly correlated, the HS vector $\boldsymbol{x}_i$ usually lives in a subspace whose dimension is much smaller than the number of bands m_λ [28], [29], i.e., $\mathbf{X} = \mathbf{HU}$ where $\mathbf{H}$ is a full column rank matrix and $\mathbf{U} \in \mathbb{R}^{\widetilde{m}_\lambda \times n}$ is the projection of $\mathbf{X}$ onto the subspace spanned by the columns of $\mathbf{H} \in \mathbb{R}^{m_\lambda \times \widetilde{m}_\lambda}$.

Defining $\mathbf{\Psi} = \{\mathbf{Y}_{\mathrm{L}}, \mathbf{Y}_{\mathrm{R}}\}$ as the set of the observed images, the negative logarithm of the likelihood is

$$\begin{aligned} -\log p\left(\mathbf{\Psi}|\mathbf{U}\right) &= -\log p\left(\mathbf{Y}_{\mathrm{L}}|\mathbf{U}\right) - \log\ p\left(\mathbf{Y}_{\mathrm{R}}|\mathbf{U}\right) + C \\ &= \frac{1}{2}\operatorname{tr}\Big(\left(\mathbf{Y}_{\mathrm{R}} - \mathbf{HUBS}\right)^T \mathbf{\Lambda}_{\mathrm{R}}^{-1} \left(\mathbf{Y}_{\mathrm{R}} - \mathbf{HUBS}\right)\Big) \\ &\quad + \frac{1}{2}\operatorname{tr}\Big(\left(\mathbf{Y}_{\mathrm{L}} - \mathbf{LHU}\right)^T \mathbf{\Lambda}_{\mathrm{L}}^{-1} \left(\mathbf{Y}_{\mathrm{L}} - \mathbf{LHU}\right)\Big) + C \end{aligned}$$

where C is a constant. Thus, calculating the ML estimator of $\mathbf{U}$ from the observed images $\mathbf{\Psi}$, i.e., maximizing the likelihood can be achieved by solving the following problem

$$\arg\min_{\mathbf{U}}\ L(\mathbf{U}) \tag{5}$$

where

$$\begin{aligned} L(\mathbf{U}) = &\operatorname{tr}\Big(\left(\mathbf{Y}_{\mathrm{R}} - \mathbf{HUBS}\right)^T \mathbf{\Lambda}_{\mathrm{R}}^{-1} \left(\mathbf{Y}_{\mathrm{R}} - \mathbf{HUBS}\right)\Big) \\ &+ \operatorname{tr}\Big(\left(\mathbf{Y}_{\mathrm{L}} - \mathbf{LHU}\right)^T \mathbf{\Lambda}_{\mathrm{L}}^{-1} \left(\mathbf{Y}_{\mathrm{L}} - \mathbf{LHU}\right)\Big). \end{aligned}$$

Note that it is also obvious to formulate the optimization problem (5) from the linear model (2) directly in the least-squares (LS) sense [30]. However, specifying the distributions of the noises $\mathbf{N}_{\mathrm{L}}$ and $\mathbf{N}_{\mathrm{R}}$ allows us to consider the case of colored noises (band-dependent) more easily by introducing the covariance matrices $\mathbf{\Lambda}_{\mathrm{R}}$ and $\mathbf{\Lambda}_{\mathrm{L}}$, leading to the weighted LS problem (5).

In this paper, we prove that the minimization of (5) w.r.t. the target image $\mathbf{U}$ can be solved analytically, without any iterative optimization scheme or Monte Carlo based method. The resulting closed-form solution to the optimization problem is presented in Section III. Furthermore, it is shown in Section IV that the proposed method can be easily generalized to Bayesian fusion methods with appropriate prior distributions.

III. Fast Fusion Scheme

A. Sylvester Equation

Minimizing (5) w.r.t. $\mathbf{U}$ is equivalent to force the derivative of $L(\mathbf{U})$ to be zero, i.e., $\mathrm{d}L(\mathbf{U})/\mathrm{d}\mathbf{U} = 0$, leading to the following matrix equation

$$\begin{aligned} &\mathbf{H}^H \mathbf{\Lambda}_{\mathrm{R}}^{-1} \mathbf{HUBS} \left(\mathbf{BS}\right)^H + \Big(\left(\mathbf{LH}\right)^H \mathbf{\Lambda}_{\mathrm{L}}^{-1} \mathbf{LH}\Big)\mathbf{U} \\ &\quad = \mathbf{H}^H \mathbf{\Lambda}_{\mathrm{R}}^{-1} \mathbf{Y}_{\mathrm{R}} \left(\mathbf{BS}\right)^H + \left(\mathbf{LH}\right)^H \mathbf{\Lambda}_{\mathrm{L}}^{-1} \mathbf{Y}_{\mathrm{L}}. \end{aligned} \tag{6}$$

As mentioned in Section I-B, the difficulty for solving (6) results from the high dimensionality of $\mathbf{U}$ and the presence of the downsampling matrix $\mathbf{S}$. In this work, we will show that Eq. (6) can be solved analytically with two assumptions summarized below.

Assumption 1: The blurring matrix $\mathbf{B}$ *is a block circulant matrix with circulant blocks.*

The physical meaning of this assumption is that the matrix $\mathbf{B}$ stands for a convolution operator by a space-invariant blurring kernel. This assumption has been currently used in the image processing literature, e.g., [24], [31]–[33]. Moreover, the blurring matrix $\mathbf{B}$ is assumed to be known in this work. In practice, it can be learnt by cross-calibration [11] or estimated from the data directly [24]. A consequence of this assumption is that $\mathbf{B}$ can be decomposed as $\mathbf{B} = \mathbf{FDF}^H$ and $\mathbf{B}^H = \mathbf{FD}^*\mathbf{F}^H$, where $\mathbf{F} \in \mathbb{R}^{n\times n}$ is the discrete Fourier transform (DFT) matrix ($\mathbf{FF}^H = \mathbf{F}^H\mathbf{F} = \mathbf{I}_n$), $\mathbf{D} \in \mathbb{R}^{n\times n}$ is a diagonal matrix and $*$ represents the conjugate operator.

Assumption 2: The decimation matrix $\mathbf{S}$ *corresponds to downsampling the original image and its conjugate transpose* $\mathbf{S}^H$ *interpolates the decimated image with zeros.*

Again, this assumption has been widely admitted in various image processing applications, such as super-resolution [32], [34] and fusion [14], [24]. Moreover, a decimation matrix satisfies the property $\mathbf{S}^H\mathbf{S} = \mathbf{I}_m$ and the

matrix $\underline{\mathbf{S}} \triangleq \mathbf{S}\mathbf{S}^H \in \mathbb{R}^{n\times n}$ is symmetric and idempotent, i.e., $\underline{\mathbf{S}} = \underline{\mathbf{S}}^H$ and $\underline{\mathbf{S}}\underline{\mathbf{S}}^H = \underline{\mathbf{S}}^2 = \underline{\mathbf{S}}$. For a practical implementation, multiplying an image by $\underline{\mathbf{S}}$ can be achieved by doing entry-wise multiplication with an $n \times n$ mask matrix with ones in the sampled position and zeros elsewhere.

After multiplying (6) on both sides by $\left(\mathbf{H}^H\mathbf{\Lambda}_{\mathrm{R}}^{-1}\mathbf{H}\right)^{-1}$, we obtain[2]

$$\mathbf{C}_1\mathbf{U} + \mathbf{U}\mathbf{C}_2 = \mathbf{C}_3 \tag{7}$$

where

$$\begin{aligned}
\mathbf{C}_1 &= \left(\mathbf{H}^H\mathbf{\Lambda}_{\mathrm{R}}^{-1}\mathbf{H}\right)^{-1}\left((\mathbf{L}\mathbf{H})^H\mathbf{\Lambda}_{\mathrm{L}}^{-1}\mathbf{L}\mathbf{H}\right)\\
\mathbf{C}_2 &= \mathbf{B}\underline{\mathbf{S}}\mathbf{B}^H\\
\mathbf{C}_3 &= \left(\mathbf{H}^H\mathbf{\Lambda}_{\mathrm{R}}^{-1}\mathbf{H}\right)^{-1}\left(\mathbf{H}^H\mathbf{\Lambda}_{\mathrm{R}}^{-1}\mathbf{Y}_{\mathrm{R}}(\mathbf{B}\mathbf{S})^H + (\mathbf{L}\mathbf{H})^H\mathbf{\Lambda}_{\mathrm{L}}^{-1}\mathbf{Y}_{\mathrm{L}}\right).
\end{aligned}$$

Eq. (7) is a Sylvester matrix equation [35]. It is well known that an SE has a unique solution if and only if an arbitrary sum of the eigenvalues of $\mathbf{C}_1$ and $\mathbf{C}_2$ is not equal to zero [35].

B. Existence of a Solution

In this section, we study the eigenvalues of $\mathbf{C}_1$ and $\mathbf{C}_2$ to check if (7) has a unique solution. As the matrix $\mathbf{C}_2 = \mathbf{B}\underline{\mathbf{S}}\mathbf{B}^H$ is positive semi-definite, its eigenvalues include positive values and zeros [36]. In order to study the eigenvalues of $\mathbf{C}_1$, Lemma 1 is introduced below.

Lemma 1: If the matrix $\mathbf{A}_1 \in \mathbb{R}^{n\times n}$ is symmetric (resp. Hermitian) positive definite and the matrix $\mathbf{A}_2 \in \mathbb{R}^{n\times n}$ is symmetric (resp. Hermitian) positive semi-definite, the product $\mathbf{A}_1\mathbf{A}_2$ is diagonalizable and all the eigenvalues of $\mathbf{A}_1\mathbf{A}_2$ are non-negative.

Proof: See Appendix A. □

According to Lemma 1, since the matrix $\mathbf{C}_1$ is the product of a symmetric positive definite matrix $\left(\mathbf{H}^H\mathbf{\Lambda}_{\mathrm{R}}^{-1}\mathbf{H}\right)^{-1}$ and a symmetric semi-definite matrix $(\mathbf{L}\mathbf{H})^H\mathbf{\Lambda}_{\mathrm{L}}^{-1}\mathbf{L}\mathbf{H}$, it is diagonalizable and all its eigenvalues are non-negative. As a consequence, the eigen-decomposition of $\mathbf{C}_1$ can be expressed as $\mathbf{C}_1 = \mathbf{Q}\Lambda_C\mathbf{Q}^{-1}$, where $\Lambda_C = \mathrm{diag}\left(\lambda_C^1, \cdots, \lambda_C^{\widetilde{m}_\lambda}\right)$ ($\mathrm{diag}\left(\lambda_C^1, \cdots, \lambda_C^{\widetilde{m}_\lambda}\right)$ is a diagonal matrix whose elements are $\lambda_C^1, \cdots, \lambda_C^{\widetilde{m}_\lambda}$) and $\lambda_C^i \geq 0$, $\forall i$. Therefore, as long as zero is not an eigenvalue of $\mathbf{C}_1$ (or equivalently $\mathbf{C}_1$ is invertible), any sum of eigenvalues of $\mathbf{C}_1$ and $\mathbf{C}_2$ is different from zero (more accurately, this sum is greater than 0), leading to the existence of a unique solution of (7).

However, the invertibility of $\mathbf{C}_1$ is not always guaranteed depending on the forms and dimensions of $\mathbf{H}$ and $\mathbf{L}$. For example, if $n_\lambda < \widetilde{m}_\lambda$, meaning that the number of MS bands is smaller than the subspace dimension, the matrix $(\mathbf{L}\mathbf{H})^H\mathbf{\Lambda}_{\mathrm{L}}^{-1}\mathbf{L}\mathbf{H}$ is rank deficient and thus (7) has no unique solution. In cases where $\mathbf{C}_1$ is singular, a regularization or prior information is necessary to be introduced to ensure (7) has a unique solution. In this section, we focus on the case when $\mathbf{C}_1$ is non-singular. The generalization to Bayesian estimators based on specific priors already considered in the literature will be elaborated in Section IV.

[2]The invertibility of the matrix $\mathbf{H}^H\mathbf{\Lambda}_{\mathrm{R}}^{-1}\mathbf{H}$ is guaranteed since $\mathbf{H}$ has full column rank and $\mathbf{\Lambda}_{\mathrm{R}}$ is positive definite.

C. A Classical Algorithm for the Sylvester Matrix Equation

A classical algorithm for obtaining a solution of the SE is the Bartels-Stewart algorithm [35]. This algorithm decomposes $\mathbf{C}_1$ and $\mathbf{C}_2$ into Schur forms using a QR algorithm and solves the resulting triangular system via back-substitution. However, as the matrix $\mathbf{C}_2 = \mathbf{B}\underline{\mathbf{S}}\mathbf{B}^H$ is very large for our application ($n \times n$, where n is the number of image pixels), it is unfeasible to construct the matrix $\mathbf{C}_2$, let alone use the QR algorithm to compute its Schur form (which has the computational cost $\mathcal{O}(n^3)$ arithmetical operations). The next section proposes an innovative strategy to obtain an analytical expression of the SE (7) by exploiting the specific properties of the matrices $\mathbf{C}_1$ and $\mathbf{C}_2$ associated with the fusion problem.

D. Proposed Closed-Form Solution

Using the decomposition $\mathbf{C}_1 = \mathbf{Q}\Lambda_C\mathbf{Q}^{-1}$ and multiplying both sides of (7) by $\mathbf{Q}^{-1}$ leads to

$$\Lambda_C\mathbf{Q}^{-1}\mathbf{U} + \mathbf{Q}^{-1}\mathbf{U}\mathbf{C}_2 = \mathbf{Q}^{-1}\mathbf{C}_3. \tag{8}$$

Right multiplying (8) by $\mathbf{F}\underline{\mathbf{D}}$ on both sides and using the definitions of matrices $\mathbf{C}_2$ and $\mathbf{B}$ yields

$$\Lambda_C\mathbf{Q}^{-1}\mathbf{U}\mathbf{F}\underline{\mathbf{D}} + \mathbf{Q}^{-1}\mathbf{U}\mathbf{F}\underline{\mathbf{D}}\left(\mathbf{F}^H\underline{\mathbf{S}}\mathbf{F}\underline{\mathbf{D}}\right) = \mathbf{Q}^{-1}\mathbf{C}_3\mathbf{F}\underline{\mathbf{D}} \tag{9}$$

where $\underline{\mathbf{D}} = (\mathbf{D}^*)\,\mathbf{D}$ is a real diagonal matrix. Note that $\mathbf{U}\mathbf{F}\underline{\mathbf{D}} = \mathbf{U}\mathbf{B}\mathbf{F}\underline{\mathbf{D}} \in \mathbb{R}^{\widetilde{m}_\lambda\times n}$ can be interpreted as the Fourier transform of the blurred target image, which is a complex matrix. Eq. (9) can be regarded as an SE w.r.t. $\mathbf{Q}^{-1}\mathbf{U}\mathbf{F}\underline{\mathbf{D}}$, which has a simpler form compared to (7) as Λ_C is a diagonal matrix. The next step in our analysis is to simplify the matrix $\mathbf{F}^H\underline{\mathbf{S}}\mathbf{F}\underline{\mathbf{D}}$ appearing on the left hand side of (9). First, we introduce the following lemma.

Lemma 2: The following equality holds

$$\mathbf{F}^H\underline{\mathbf{S}}\mathbf{F} = \frac{1}{d}\mathbf{J}_d \otimes \boldsymbol{I}_m \tag{10}$$

where $\mathbf{F}$ and $\underline{\mathbf{S}}$ are defined as in Section III-A, $\mathbf{J}_d$ is the $d \times d$ matrix of ones and $\boldsymbol{I}_m$ is the $m \times m$ identity matrix.

Proof: See Appendix B. □

This lemma shows that the spectral aliasing resulting from a downsampling operator applied to a multi-band image in the spatial domain can be easily formulated as a Kronecker product in the frequency domain.

Then, let introduce the following $md \times md$ matrix

$$\mathbf{P} = \underbrace{\begin{bmatrix} \mathbf{I}_m & \mathbf{0} & \cdots & \mathbf{0} \\ -\mathbf{I}_m & \mathbf{I}_m & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ -\mathbf{I}_m & \mathbf{0} & \cdots & \mathbf{I}_m \end{bmatrix}}_{d} \tag{11}$$

whose inverse[3] can be easily computed

$$\mathbf{P}^{-1} = \begin{bmatrix} \mathbf{I}_m & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{I}_m & \mathbf{I}_m & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{I}_m & \mathbf{0} & \cdots & \mathbf{I}_m \end{bmatrix}.$$

[3]Note that left multiplying a matrix by $\mathbf{P}$ corresponds to subtracting the first row blocks from all the other row blocks. Conversely, right multiplying by the matrix $\mathbf{P}^{-1}$ means replacing the first (block) column by the sum of all the other (block) columns.

Right multiplying both sides of (9) by $\mathbf{P}^{-1}$ leads to

$$\Lambda_C \bar{\mathbf{U}} + \bar{\mathbf{U}}\mathbf{M} = \bar{\mathbf{C}}_3 \tag{12}$$

where $\bar{\mathbf{U}} = \mathbf{Q}^{-1}\mathbf{U}\mathbf{F}\mathbf{D}\mathbf{P}^{-1}$, $\mathbf{M} = \mathbf{P}\left(\mathbf{F}^H\underline{\mathbf{S}}\mathbf{F}\underline{\mathbf{D}}\right)\mathbf{P}^{-1}$ and $\bar{\mathbf{C}}_3 = \mathbf{Q}^{-1}\mathbf{C}_3\mathbf{F}\mathbf{D}\mathbf{P}^{-1}$. Eq. (12) is a Sylvester matrix equation w.r.t. $\bar{\mathbf{U}}$ whose solution is significantly easier than for (8), thanks to the simple structure of the matrix $\mathbf{M}$ outlined in the following lemma.

Lemma 3: The following equality holds

$$\mathbf{M} = \frac{1}{d}\begin{bmatrix} \sum_{i=1}^{d}\underline{\mathbf{D}}_i & \underline{\mathbf{D}}_2 & \cdots & \underline{\mathbf{D}}_d \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} \end{bmatrix} \tag{13}$$

where the matrix $\underline{\mathbf{D}}$ has been partitioned as follows

$$\underline{\mathbf{D}} = \begin{bmatrix} \underline{\mathbf{D}}_1 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \underline{\mathbf{D}}_2 & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \underline{\mathbf{D}}_d \end{bmatrix}$$

with $\underline{\mathbf{D}}_i$ $m \times m$ real diagonal matrices.

Proof: See Appendix C. □

This lemma, which exploits the equality (10) and the resulting specific structure of the matrix $\mathbf{F}^H\underline{\mathbf{S}}\mathbf{F}\underline{\mathbf{D}}$, allows the matrix $\mathbf{M}$ to be written block-by-block, with nonzero blocks only located in its first (block) row (see (13)). Finally, using this simple form of $\mathbf{M}$, the solution $\bar{\mathbf{U}}$ of the SE (12) can be computed block-by-block as stated in the following theorem.

Theorem 1: Let $(\bar{\mathbf{C}}_3)_{l,j}$ denotes the jth block of the lth band of $\bar{\mathbf{C}}_3$ for any $l = 1, \cdots, \widetilde{m}_\lambda$. Then, the solution $\bar{\mathbf{U}}$ of the SE (12) *can be decomposed as*

$$\bar{\mathbf{U}} = \begin{bmatrix} \bar{\mathbf{u}}_{1,1} & \bar{\mathbf{u}}_{1,2} & \cdots & \bar{\mathbf{u}}_{1,d} \\ \bar{\mathbf{u}}_{2,1} & \bar{\mathbf{u}}_{2,2} & \cdots & \bar{\mathbf{u}}_{2,d} \\ \vdots & \vdots & \ddots & \vdots \\ \bar{\mathbf{u}}_{\widetilde{m}_\lambda,1} & \bar{\mathbf{u}}_{\widetilde{m}_\lambda,2} & \cdots & \bar{\mathbf{u}}_{\widetilde{m}_\lambda,d} \end{bmatrix} \tag{14}$$

with

$$\bar{\mathbf{u}}_{l,j} = \begin{cases} (\bar{\mathbf{C}}_3)_{l,j}\left(\frac{1}{d}\sum_{i=1}^{d}\underline{\mathbf{D}}_i + \lambda_C^l \boldsymbol{I}_m\right)^{-1}, & j = 1, \\ \frac{1}{\lambda_C^l}\left[(\bar{\mathbf{C}}_3)_{l,j} - \frac{1}{d}\bar{\mathbf{u}}_{l,1}\underline{\mathbf{D}}_j\right], & j = 2, \cdots, d. \end{cases} \tag{15}$$

Proof: See Appendix D. □

Note that $\mathbf{u}_{l,j} \in \mathbb{R}^{1\times m}$ denotes the jth block of the lth band. Note also that the matrix $\frac{1}{d}\sum_{i=1}^{d}\underline{\mathbf{D}}_i + \lambda_C^l\mathbf{I}_n$ appearing in the expression of $\bar{\mathbf{u}}_{l,1}$ is an $n \times n$ real diagonal matrix whose inversion is trivial. The final estimator of $\mathbf{X}$ is obtained as follows[4]

$$\hat{\mathbf{X}} = \mathbf{H}\mathbf{Q}\bar{\mathbf{U}}\mathbf{P}\mathbf{D}^{-1}\mathbf{F}^H. \tag{16}$$

[4]It may happen that the diagonal matrix $\mathbf{D}$ does not have full rank (containing zeros in diagonal) or is ill-conditioned (having very small numbers in diagonal), resulting from the property of blurring kernel. In this case, $\mathbf{D}^{-1}$ can be replaced by $(\mathbf{D} + \tau\mathbf{I}_m)^{-1}$ for regularization purpose, where τ is a small penalty parameter [31].

Algorithm 1 Fast Fusion of Multi-Band Images Based on Solving a Sylvester Equation (FUSE)

Input: $\mathbf{Y}_\mathrm{L}$, $\mathbf{Y}_\mathrm{R}$, $\mathbf{\Lambda}_\mathrm{L}$, $\mathbf{\Lambda}_\mathrm{R}$, $\mathbf{L}$, $\mathbf{B}$, $\mathbf{S}$, $\mathbf{H}$

// *Circulant matrix decomposition:* $\mathbf{B} = \mathbf{F}\mathbf{D}\mathbf{F}^H$

$\mathbf{D} \leftarrow \mathrm{Dec}(\mathbf{B})$;

$\underline{\mathbf{D}} = \mathbf{D}^*\mathbf{D}$;

// *Calculate* $\mathbf{C}_1$

$\mathbf{C}_1 \leftarrow \left(\mathbf{H}^H\mathbf{\Lambda}_\mathrm{R}^{-1}\mathbf{H}\right)^{-1}\left((\mathbf{L}\mathbf{H})^H\mathbf{\Lambda}_\mathrm{L}^{-1}\mathbf{L}\mathbf{H}\right)$;

// *Eigen-decomposition of* $\mathbf{C}_1$*:* $\mathbf{C}_1 = \mathbf{Q}\mathbf{\Lambda}_C\mathbf{Q}^{-1}$

$(\mathbf{Q}, \mathbf{\Lambda}_C) \leftarrow \mathrm{EigDec}(\mathbf{C}_1)$;

// *Calculate* $\bar{\mathbf{C}}_3$

$\bar{\mathbf{C}}_3 \leftarrow \mathbf{Q}^{-1}\left(\mathbf{H}^H\mathbf{\Lambda}_\mathrm{R}^{-1}\mathbf{H}\right)^{-1}(\mathbf{H}^H\mathbf{\Lambda}_\mathrm{R}^{-1}\mathbf{Y}_\mathrm{R}(\mathbf{B}\mathbf{S})^H + (\mathbf{L}\mathbf{H})^H\mathbf{\Lambda}_\mathrm{L}^{-1}\mathbf{Y}_\mathrm{L})\mathbf{B}\mathbf{F}\mathbf{P}^{-1}$;

// *Calculate* $\bar{\mathbf{U}}$ *block by block (*d *blocks) and band by band (*$\widetilde{m}_\lambda$ *bands)*

for $l = 1$ **to** $\widetilde{m}_\lambda$ **do**

  // *Calculate the* 1*st block in* l*th band*

  $\bar{\mathbf{u}}_{l,1} = (\bar{\mathbf{C}}_3)_{l,1}\left(\frac{1}{d}\sum_{i=1}^{d}\underline{\mathbf{D}}_i + \lambda_C^l\mathbf{I}_n\right)^{-1}$;

  // *Calculate other blocks in* l*th band*

  for $j = 2$ **to** d **do**

    $\bar{\mathbf{u}}_{l,j} = \frac{1}{\lambda_C^l}\left((\bar{\mathbf{C}}_3)_{l,j} - \frac{1}{d}\bar{\mathbf{u}}_{l,1}\underline{\mathbf{D}}_j\right)$;

  end

end

Set $\mathbf{X} = \mathbf{H}\mathbf{Q}\bar{\mathbf{U}}\mathbf{P}\mathbf{D}^{-1}\mathbf{F}^H$;

Output: $\mathbf{X}$

Algorithm 1 summarizes the derived FUSE steps required to calculate the estimated image $\hat{\mathbf{X}}$.

E. Complexity Analysis

The most computationally expensive part of the proposed algorithm is the computation of matrices $\mathbf{D}$ and $\bar{\mathbf{C}}_3$ because of the FFT and iFFT operations. Using the notation $\mathbf{C}_4 = \mathbf{Q}^{-1}\left(\mathbf{H}^H\mathbf{\Lambda}_\mathrm{R}^{-1}\mathbf{H}\right)^{-1}$, the matrix $\bar{\mathbf{C}}_3$ can be rewritten

$$\begin{aligned}\bar{\mathbf{C}}_3 &= \mathbf{C}_4\left(\mathbf{H}^H\mathbf{\Lambda}_\mathrm{R}^{-1}\mathbf{Y}_\mathrm{R}(\mathbf{B}\mathbf{S})^\mathbf{H} + (\mathbf{L}\mathbf{H})^\mathbf{H}\mathbf{\Lambda}_\mathrm{L}^{-1}\mathbf{Y}_\mathrm{L}\right)\mathbf{B}\mathbf{F}\mathbf{P}^{-1} \\ &= \mathbf{C}_4\left(\mathbf{H}^H\mathbf{\Lambda}_\mathrm{R}^{-1}\mathbf{Y}_\mathrm{R}\mathbf{S}^H\mathbf{F}\mathbf{D}^* + (\mathbf{L}\mathbf{H})^H\mathbf{\Lambda}_\mathrm{L}^{-1}\mathbf{Y}_\mathrm{L}\mathbf{F}\right)\mathbf{D}\mathbf{P}^{-1}.\end{aligned} \tag{17}$$

The most heavy step in computing (17) is the decomposition $\mathbf{B} = \mathbf{F}\mathbf{D}\mathbf{F}^H$ (or equivalently the FFT of the blurring kernel), which has a complexity of order $\mathcal{O}(n\log n)$. The calculations of $\mathbf{H}^H\mathbf{\Lambda}_\mathrm{R}^{-1}\mathbf{Y}_\mathrm{R}\mathbf{S}^H\mathbf{F}\mathbf{D}^*$ and $(\mathbf{L}\mathbf{H})^H\mathbf{\Lambda}_\mathrm{L}^{-1}\mathbf{Y}_\mathrm{L}\mathbf{F}$ require one FFT operation each. All the other computations are made in the frequency domain. Note that the multiplication by $\mathbf{D}\mathbf{P}^{-1}$ has a cost of $\mathcal{O}(n)$ operations as $\mathbf{D}$ is diagonal, and $\mathbf{P}^{-1}$ reduces to block shifting and addition. The left multiplication with $\mathbf{Q}^{-1}\left(\mathbf{H}^H\mathbf{\Lambda}_\mathrm{R}^{-1}\mathbf{H}\right)^{-1}$ is of order $\mathcal{O}(\widetilde{m}_\lambda^2 n)$. Thus, the calculation of $\mathbf{C}_3\mathbf{B}\mathbf{F}\mathbf{P}^{-1}$ has a total complexity of order $\mathcal{O}(n\cdot\max\left\{\log n, \widetilde{m}_\lambda^2\right\})$.

IV. Generalization to Bayesian Estimators

As mentioned in Section III-B, if the matrix $(\mathbf{LH})^H \mathbf{\Lambda}_{\mathrm{L}}^{-1} \mathbf{LH}$ is singular or ill-conditioned (e.g., when the number of MS bands is smaller than the dimension of the subspace spanned by the pixel vectors, i.e., $n_\lambda < \widetilde{m}_\lambda$), a regularization or prior information $p(\mathbf{U})$ has to be introduced to ensure the Sylvester matrix equation (12) has a unique solution. The resulting estimator $\mathbf{U}$ can then be interpreted as a Bayesian estimator. Combining the likelihood (4) and the prior $p(\mathbf{U})$, the posterior distribution of $\mathbf{U}$ can be written as

$$\begin{aligned} p(\mathbf{U}|\mathbf{\Psi}) &\propto p(\mathbf{\Psi}|\mathbf{U})\, p(\mathbf{U}) \\ &\propto p(\mathbf{Y}_{\mathrm{L}}|\mathbf{U})\, p(\mathbf{Y}_{\mathrm{R}}|\mathbf{U})\, p(\mathbf{U}) \end{aligned}$$

where $\propto$ means "proportional to" and where we have used the independence between the observation vectors $\mathbf{Y}_{\mathrm{L}}$ and $\mathbf{Y}_{\mathrm{R}}$.

The mode of the posterior distribution $p(\mathbf{U}|\mathbf{\Psi})$ is the so-called MAP estimator, which can be obtained by solving the following optimization problem

$$\arg\min_{\mathbf{U}} L(\mathbf{U}) \tag{18}$$

where

$$\begin{aligned} L(\mathbf{U}) = &\frac{1}{2} \operatorname{tr}\left(\left(\mathbf{Y}_{\mathrm{R}} - \mathbf{HUBS}\right)^{\mathbf{T}} \mathbf{\Lambda}_{\mathrm{R}}^{-\mathbf{1}} \left(\mathbf{Y}_{\mathrm{R}} - \mathbf{HUBS}\right)\right) \\ &+ \frac{1}{2} \operatorname{tr}\left(\left(\mathbf{Y}_{\mathrm{L}} - \mathbf{LHU}\right)^{T} \mathbf{\Lambda}_{\mathrm{L}}^{-1} \left(\mathbf{Y}_{\mathrm{L}} - \mathbf{LHU}\right)\right) - \log p(\mathbf{U}). \end{aligned} \tag{19}$$

Different Bayesian estimators corresponding to different choices of $p(\mathbf{U})$ have been considered in the literature. These estimators are first recalled in the next sections. We will then show that the explicit solution of the SE derived in Section III can be used to compute the MAP estimator of $\mathbf{U}$ for these prior distributions.

A. Gaussian Prior

Gaussian priors have been used widely in image processing [37]–[39], and can be interpreted as a Tikhonov regularization [40]. Assume that a matrix normal distribution is assigned *a priori* to the projected target image $\mathbf{U}$

$$p(\mathbf{U}) = \mathcal{MN}_{\widetilde{m}_\lambda, n}(\boldsymbol{\mu}, \mathbf{\Sigma}, \mathbf{I}_n) \tag{20}$$

where $\boldsymbol{\mu}$ and $\mathbf{\Sigma}$ are the mean and covariance matrix of the matrix normal distribution. Note that the covariance matrix $\mathbf{\Sigma}$ explores the correlations between HS band and controls the distance between $\mathbf{U}$ and its mean $\boldsymbol{\mu}$. Forcing the derivative of $L(\mathbf{U})$ in (18) to be zero leads to the following SE

$$\mathbf{C}_1 \mathbf{U} + \mathbf{U} \mathbf{C}_2 = \mathbf{C}_3 \tag{21}$$

where

$$\begin{aligned} \mathbf{C}_1 &= \left(\mathbf{H}^H \mathbf{\Lambda}_{\mathrm{R}}^{-1} \mathbf{H}\right)^{-1} \left((\mathbf{LH})^H \mathbf{\Lambda}_{\mathrm{L}}^{-1} \mathbf{LH} + \mathbf{\Sigma}^{-1}\right) \\ \mathbf{C}_2 &= \mathbf{B}\underline{\mathbf{S}}\mathbf{B}^H \\ \mathbf{C}_3 &= \left(\mathbf{H}^H \mathbf{\Lambda}_{\mathrm{R}}^{-1} \mathbf{H}\right)^{-1} (\mathbf{H}^H \mathbf{\Lambda}_{\mathrm{R}}^{-1} \mathbf{Y}_{\mathrm{R}} (\mathbf{BS})^H \\ &\quad + (\mathbf{LH})^H \mathbf{\Lambda}_{\mathrm{L}}^{-1} \mathbf{Y}_{\mathrm{L}} + \mathbf{\Sigma}^{-1} \boldsymbol{\mu}). \end{aligned} \tag{22}$$

The matrix $\mathbf{C}_1$ is positive definite as long as the covariance matrix $\mathbf{\Sigma}^{-1}$ is positive definite. Algorithm 1 can thus be adapted to a matrix normal prior case by simply replacing $\mathbf{C}_1$ and $\mathbf{C}_3$ by their new expressions defined in (22).

B. Non-Gaussian Prior

When the projected image $\mathbf{U}$ is assigned a non-Gaussian prior, the objective function $L(\mathbf{U})$ in (18) can be split into a data term $f(\mathbf{U})$ corresponding to the likelihood and a regularization term $\phi(\mathbf{U})$ corresponding to the prior in a Bayesian framework as

$$L(\mathbf{U}) = f(\mathbf{U}) + \phi(\mathbf{U}) \tag{23}$$

where

$$\begin{aligned} f(\mathbf{U}) = &\frac{1}{2} \operatorname{tr}\left(\left(\mathbf{Y}_{\mathrm{R}} - \mathbf{HUBS}\right)^{T} \mathbf{\Lambda}_{\mathrm{R}}^{-1} \left(\mathbf{Y}_{\mathrm{R}} - \mathbf{HUBS}\right)\right) \\ &+ \frac{1}{2} \operatorname{tr}\left(\left(\mathbf{Y}_{\mathrm{L}} - \mathbf{LHU}\right)^{T} \mathbf{\Lambda}_{\mathrm{L}}^{-1} \left(\mathbf{Y}_{\mathrm{L}} - \mathbf{LHU}\right)\right) \end{aligned}$$

and

$$\phi(\mathbf{U}) = -\log p(\mathbf{U}).$$

The optimization of (23) w.r.t. $\mathbf{U}$ can be solved efficiently by using an ADMM that consists of two steps: 1) solving a surrogate optimization problem associated with a Gaussian prior and 2) applying a proximity operator [41]. This strategy can be implemented in the image domain or in the frequency domain. The resulting algorithms, referred to as FUSE-within-ADMM (FUSE-ADMM) are described below.

1) Solution in Image Domain: Eq. (23) can be rewritten as

$$L(\mathbf{U}, \mathbf{V}) = f(\mathbf{U}) + \phi(\mathbf{V}) \quad \text{s.t. } \mathbf{U} = \mathbf{V}.$$

The augmented Lagrangian associated with this problem is

$$L_\mu(\mathbf{U}, \mathbf{V}, \boldsymbol{\lambda}) = f(\mathbf{U}) + \phi(\mathbf{V}) + \boldsymbol{\lambda}^T (\mathbf{U} - \mathbf{V}) + \frac{\mu}{2} \|\mathbf{U} - \mathbf{V}\|_F^2 \tag{24}$$

or equivalently

$$L_\mu(\mathbf{U}, \mathbf{V}, \mathbf{W}) = f(\mathbf{U}) + \phi(\mathbf{V}) + \frac{\mu}{2} \|\mathbf{U} - \mathbf{V} - \mathbf{W}\|_F^2 \tag{25}$$

where $\mathbf{W}$ is the scaled dual variable. This optimization problem can be solved by an ADMM as follows

$$\begin{aligned} (\mathbf{U}^{k+1}, \mathbf{V}^{k+1}) &= \arg\min_{\mathbf{U},\mathbf{V}} f(\mathbf{U}) + \phi(\mathbf{V}) + \frac{\mu}{2} \|\mathbf{U} - \mathbf{V} - \mathbf{W}^k\|_F^2 \\ \mathbf{W}^{k+1} &= \mathbf{W}^k - (\mathbf{U}^{k+1} - \mathbf{V}^{k+1}). \end{aligned}$$

The updates of the derived ADMM algorithm are

$$\begin{aligned} \mathbf{U}^{k+1} &= \arg\min_{\mathbf{U}} f(\mathbf{U}) + \frac{\mu}{2} \|\mathbf{U} - \mathbf{V}^k - \mathbf{W}^k\|_F^2 \\ \mathbf{V}^{k+1} &= \operatorname{prox}_{\phi,\mu}(\mathbf{U}^{k+1} - \mathbf{W}^k) \\ \mathbf{W}^{k+1} &= \mathbf{W}^k - (\mathbf{U}^{k+1} - \mathbf{V}^{k+1}). \end{aligned} \tag{26}$$

- **Update U**: Instead of using any iterative update method, the optimization w.r.t. $\mathbf{U}$ can be solved analytically by using Algorithm 1 as for the Gaussian prior investigated in Section IV-A. For this, we can set $\boldsymbol{\mu} = \mathbf{V}^k + \mathbf{W}^k$ and $\mathbf{\Sigma}^{-1} = \mu \mathbf{I}_{\widetilde{m}_\lambda}$ in (22). However, the computational complexity of updating $\mathbf{U}$ in each iteration is $\mathcal{O}(n \log n)$

because of the FFT and iFFT steps required for computing $\bar{\mathbf{C}}_3$ and $\mathbf{U}$ from $\bar{\mathbf{U}}$.

- **Update V**: The update of $\mathbf{V}$ requires computing a proximity operator, which depends on the form of $\phi(\mathbf{V})$. When the regularizer $\phi(\mathbf{V})$ is simple enough, the proximity operator can be evaluated analytically. For example, if $\phi(\mathbf{V}) \equiv \|\mathbf{V}\|_1$, then

$$\text{prox}_{\phi,\mu}(\mathbf{U}^{k+1} - \mathbf{W}^k) = \text{soft}\left(\mathbf{U}^{k+1} - \mathbf{W}^k, \frac{1}{\mu}\right)$$

where soft is the soft-thresholding function defined as

$$\text{soft}(g, \tau) = \text{sign}(g)\max(|g| - \tau, 0).$$

More examples of proximity computations can be found in [41].

- **Update W**: The update of $\mathbf{W}$ is simply a matrix addition whose implementation has a small computational cost.

2) Solution in Frequency Domain: Recalling that $\mathbf{B} = \mathbf{F}\mathbf{D}\mathbf{F}^H$, a less computationally expensive solution is obtained by rewriting $L(\mathbf{U})$ in (23) as

$$L(\mathcal{U}, \mathcal{V}) = f(\mathcal{U}) + \phi(\mathcal{V}) \quad \text{s.t. } \mathcal{U} = \mathcal{V}$$

where $\mathcal{U} = \mathbf{U}\mathbf{F}$ is the Fourier transform of $\mathbf{U}$, $\mathcal{V} = \mathbf{V}\mathbf{F}$ is the Fourier transform of $\mathbf{V}$, and

$$\begin{aligned} f(\mathcal{U}) = &\frac{1}{2}\text{tr}\left((\mathbf{Y}_\text{R} - \mathbf{H}\mathcal{U}\mathbf{D}\mathbf{F}^H\mathbf{S})^T\mathbf{\Lambda}_\text{R}^{-1}(\mathbf{Y}_\text{R} - \mathbf{H}\mathcal{U}\mathbf{D}\mathbf{F}^H\mathbf{S})\right) \\ &+ \frac{1}{2}\text{tr}\left(\left(\mathbf{Y}_\text{L} - \mathbf{L}\mathbf{H}\mathcal{U}\mathbf{F}^H\right)^T\mathbf{\Lambda}_\text{L}^{-1}\left(\mathbf{Y}_\text{L} - \mathbf{L}\mathbf{H}\mathcal{U}\mathbf{F}^H\right)\right) \end{aligned}$$

and

$$\phi(\mathcal{V}) = -\log p(\mathcal{V}).$$

Thus, the ADMM updates, defined in the image domain by (26), can be rewritten in the frequency domain as

$$\begin{aligned} \mathcal{U}^{k+1} &= \arg\min_{\mathcal{U}} f(\mathcal{U}) + \frac{\mu}{2}\|\mathcal{U} - \mathcal{V}^k - \mathcal{W}^k\|_F^2 \\ \mathcal{V}^{k+1} &= \text{prox}_{\phi,\mu}(\mathcal{U}^{k+1} - \mathcal{W}^k) \\ \mathcal{W}^{k+1} &= \mathcal{W}^k - (\mathcal{U}^{k+1} - \mathcal{V}^{k+1}). \end{aligned} \tag{27}$$

where $\mathcal{W}$ is the dual variable in frequency domain. At the $(k+1)$th ADMM iteration, updating $\mathcal{U}$ can be efficiently conducted thanks to an SE solver similar to Algorithm 1, where the matrix $\bar{\mathbf{C}}_3$ is defined by

$$\bar{\mathbf{C}}_3 = \mathbf{C}_s + \mathbf{C}_c\left(\mathcal{V}^k + \mathcal{W}^k\right)\mathbf{D}\mathbf{P}^{-1} \tag{28}$$

with

$$\begin{aligned} \mathbf{C}_s = &\mathbf{Q}^{-1}\left(\mathbf{H}^H\mathbf{\Lambda}_\text{R}^{-1}\mathbf{H}\right)^{-1} \\ &\left(\mathbf{H}^H\mathbf{\Lambda}_\text{R}^{-1}\mathbf{Y}_\text{R}\mathbf{S}^H\mathbf{F}\mathbf{D}^H + (\mathbf{L}\mathbf{H})^H\mathbf{\Lambda}_\text{L}^{-1}\mathbf{Y}_\text{L}\mathbf{F}\right)\mathbf{D}\mathbf{P}^{-1} \\ \mathbf{C}_c = &\mathbf{Q}^{-1}\left(\mathbf{H}^H\mathbf{\Lambda}_\text{R}^{-1}\mathbf{H}\right)^{-1}\mathbf{\Sigma}^{-1}. \end{aligned}$$

Note that the update of $\bar{\mathbf{C}}_3$ does not require any FFT computation since $\mathbf{C}_s$ and $\mathbf{C}_c$ can be calculated once and are not updated in the ADMM iterations.

C. Hierarchical Bayesian Framework

When using a Gaussian prior, a hierarchical Bayesian framework can be constructed by introducing a hyperprior to the hyperparameter vector $\mathbf{\Phi} = \{\boldsymbol{\mu}, \mathbf{\Sigma}\}$. Several priors have been investigated in the literature based on generalized Gaussian distributions, sparsity-promoted ℓ_1 or ℓ_0 regularizations, ℓ_2 smooth regularization, or TV regularization. Denoting as $p(\mathbf{\Phi})$ the prior of $\mathbf{\Phi}$, the optimization w.r.t. $\mathbf{U}$ can be replaced by an optimization w.r.t. $(\mathbf{U}, \mathbf{\Phi})$ as follows

$$\begin{aligned} (\mathbf{U}, \mathbf{\Phi}) &= \arg\max_{\mathbf{U},\mathbf{\Phi}} p(\mathbf{U}, \mathbf{\Phi}|\mathbf{\Psi}) \\ &= \arg\max_{\mathbf{U},\mathbf{\Phi}} p(\mathbf{Y}_\text{L}|\mathbf{U})\, p(\mathbf{Y}_\text{R}|\mathbf{U})\, p(\mathbf{U}|\mathbf{\Phi})\, p(\mathbf{\Phi}). \end{aligned}$$

A standard way of solving this problem is to optimize alternatively between $\mathbf{U}$ and $\mathbf{\Phi}$ using the following updates

$$\begin{aligned} \mathbf{U}^{k+1} &= \arg\max_{\mathbf{U}} p(\mathbf{Y}_\text{L}|\mathbf{U})\, p(\mathbf{Y}_\text{R}|\mathbf{U})\, p\left(\mathbf{U}|\mathbf{\Phi}^k\right) \\ \mathbf{\Phi}^{k+1} &= \arg\max_{\mathbf{\Phi}} p\left(\mathbf{U}^{k+1}|\mathbf{\Phi}\right) p(\mathbf{\Phi}). \end{aligned}$$

The update of $\mathbf{U}^{k+1}$ can be solved using Algorithm 1 whereas the update of $\mathbf{\Phi}$ depends on the form of the hyperprior $p(\mathbf{\Phi})$. The derived optimization method is referred to as FUSE-within-BCD (FUSE-BCD).

It is interesting to note that the strategy of Section IV-B proposed to handle the case of a non-Gaussian prior can be interpreted as a special case of a hierarchical updating. Indeed, if we interpret $\mathbf{V} + \mathbf{d}$ and $\frac{1}{\mu}\mathbf{I}_{\widetilde{m}_\lambda}$ in (25) as the mean $\boldsymbol{\mu}$ and covariance matrix $\mathbf{\Sigma}$, the ADMM update (26) can be considered as the iterative updates of $\mathbf{U}$ and $\boldsymbol{\mu} = \mathbf{V} + \mathbf{d}$ with fixed $\mathbf{\Sigma} = \frac{1}{\mu}\mathbf{I}_{\widetilde{m}_\lambda}$.

V. Experimental Results

This section applies the proposed fusion method to three kinds of priors that have been investigated in [20], [23], and [24] for the fusion of multi-band images. Note that these three methods require to solve a minimization problem similar to (18). All the algorithms have been implemented using MATLAB R2013A on a computer with Intel(R) Core(TM) i7-2600 CPU@3.40GHz and 8GB RAM. The MATLAB codes and all the simulation results are available in the first author's homepage.[5]

A. Fusion Quality Metrics

To evaluate the quality of the proposed fusion strategy, five image quality measures have been investigated. Referring to [20] and [42], we propose to use the restored signal to noise ratio (RSNR), the averaged spectral angle mapper (SAM), the universal image quality index (UIQI), the relative dimensionless global error in synthesis (ERGAS) and the degree of distortion (DD) as quantitative measures. The RSNR is defined by the negative logarithm of the distance between the estimated and reference images. The larger RSNR, the better the fusion. The definition of SAM, UIQI, ERGAS and DD can be found in [20]. The smaller SAM, ERGAS and DD,

[5] http://wei.perso.enseeiht.fr/

TABLE II

PERFORMANCE OF HS+MS FUSION METHODS: RSNR (IN dB), UIQI, SAM (IN DEGREE), ERGAS, DD (IN 10^{-3}) AND TIME (IN SECOND)

Regularization	Methods	RSNR	UIQI	SAM	ERGAS	DD	Time
supervised naive Gaussian	ADMM [23]	29.321	0.9906	1.555	0.888	7.115	126.83
	FUSE	29.372	0.9908	1.551	0.879	7.092	0.38
unsupervised naive Gaussian	ADMM-BCD [23]	29.084	0.9902	1.615	0.913	7.341	99.55
	FUSE-BCD	29.077	0.9902	1.623	0.913	7.368	1.09
sparse representation	ADMM-BCD [20]	29.582	0.9911	1.423	0.872	6.678	162.88
	FUSE-BCD	29.688	0.9913	1.431	0.856	6.672	73.66
TV	ADMM [24]	29.473	0.9912	1.503	0.861	6.922	134.21
	FUSE-ADMM	29.631	0.9915	1.477	0.845	6.788	90.99

the better the fusion. The larger UIQI, the better the fusion. The maps of the residual errors, computed in terms of root mean square errors averaged over the bands, are also available in the associated technical report [43].

B. Fusion of HS and MS Images

The reference image considered here as the high-spatial and high-spectral image is a $512 \times 256 \times 93$ HS image acquired over Pavia, Italy, by the reflective optics system imaging spectrometer (ROSIS). This image was initially composed of 115 bands that have been reduced to 93 bands after removing the water vapor absorption bands. A composite color image of the scene of interest is shown in Fig. 1 (right).

Our objective is to reconstruct the high-spatial high-spectral image $\mathbf{X}$ from a low-spatial high-spectral HS image $\mathbf{Y}_{\mathrm{R}}$ and a high-spatial low-spectral MS image $\mathbf{Y}_{\mathrm{L}}$. First, $\mathbf{Y}_{\mathrm{R}}$ has been generated by applying a 5×5 Gaussian filter and by down-sampling every $d_r = d_c = 4$ pixels in both vertical and horizontal directions for each band of the reference image. Second, a 4-band MS image $\mathbf{Y}_{\mathrm{L}}$ has been obtained by filtering $\mathbf{X}$ with the LANDSAT-like reflectance spectral responses [44]. The HS and MS images are both contaminated by zero-mean additive Gaussian noises. Our simulations have been conducted with $\mathrm{SNR}_{\mathrm{H},i} = 35$dB for the first 43 bands of the HS image and $\mathrm{SNR}_{\mathrm{H},i} = 30$dB for the remaining 50 bands, with

$$\mathrm{SNR}_{\mathrm{H},i} = 10 \log \left(\frac{\| (\mathbf{XBS})_i \|_F^2}{s_{\mathrm{H},i}^2} \right).$$

For the MS image

$$\mathrm{SNR}_{\mathrm{M},j} = 10 \log \left(\frac{\| (\mathbf{LX})_j \|_F^2}{s_{\mathrm{M},j}^2} \right) = 30\mathrm{dB}$$

for all spectral bands.

The observed HS and MS images are shown in Fig. 1 (left and middle). Note that the HS image has been scaled for better visualization (i.e., the HS image contains $d = 16$ times fewer pixels than the MS image) and that the MS image has been displayed using an arbitrary color composition. The subspace transformation matrix $\mathbf{H}$ has been defined as the PCA following the strategy of [20].

Fig. 1. Pavia dataset: HS image (left), MS image (middle) and reference image (right).

1) Example 1 (HS+MS Fusion With a Naive Gaussian Prior): We first consider the Bayesian fusion model initially proposed in [22]. This method assumed a naive Gaussian prior for the target image, leading to an ℓ_2-regularization of the fusion problem. The mean of this Gaussian prior was fixed to an interpolated HS image. The covariance matrix of the Gaussian prior can be fixed *a priori* (supervised fusion) or estimated jointly with the unknown image within a hierarchical Bayesian method (unsupervised fusion). Recall that the estimator studied in [22] was based on a hybrid Gibbs sampler generating samples distributed according to the posterior of interest. An ADMM step embedded in a BCD method (ADMM-BCD) was also proposed in [23] providing a significant computational cost reduction. This section compares the performance of this ADMM-BCD algorithm with the performances of the proposed FUSE-based methods for these fusion problems.

For the supervised case, the explicit solution of the SE can be constructed directly following the Gaussian prior-based generalization in Section IV-A. Conversely, for the unsupervised case, the generalized version denoted FUSE-BCD and described in Section IV-C is exploited, which requires embedding the closed-form solution into a BCD algorithm (refer [23] for more details). The estimated images obtained with the different algorithms are depicted in Fig. 2 and are visually very similar. More quantitative results are reported in the first four lines of Table II and confirm the similar

Fig. 2. HS+MS fusion results. Row 1 and 2: state-of-the-art-methods and corresponding proposed fast fusion methods (FUSE), respectively, with various regularizations: supervised naive Gaussian prior (1st column), unsupervised naive Gaussian prior (2nd column), sparse representation (3rd column) and TV (4th column).

performance of these methods in terms of the various fusion quality measures (RSNR, UIQI, SAM, ERGAS and DD). However, the computational time of the proposed algorithm is reduced by a factor larger than 200 (supervised) and 90 (unsupervised) due to the existence of a closed-form solution for the Sylvester matrix equation.

2) Example 2 (HS+MS Fusion With a Sparse Representation): This section investigates a Bayesian fusion model based on the Gaussian prior associated with a sparse representation introduced in [20]. The basic idea of this approach was to design a prior that results from the sparse decomposition of the target image on a set of dictionaries learned empirically. Some parameters needed to be adjusted by the operator (regularization parameter, dictionaries and supports) whereas the other parameters (sparse codes) were jointly estimated with the target image. In [20], the MAP estimator associated with this model was reached using an optimization algorithm that consists of an ADMM step embedded in a BCD method (ADMM-BCD). Using the strategy proposed in Section IV-C, this ADMM step can be avoid by exploiting the FUSE solution. Thus, the performance of the ADMM-BCD algorithm in [20] is compared with the performance of the FUSE-BCD scheme as described in Section IV-C. As shown in Fig. 2 and the 5th and 6th lines of Table II, the performances of both algorithms are quite similar. However, the proposed solution exhibits a significant complexity reduction.

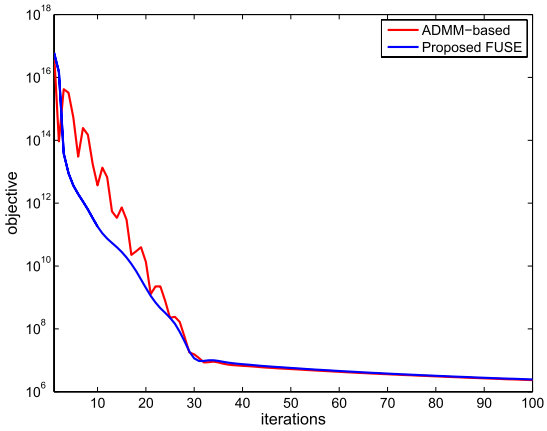

Fig. 3. Convergence speeds of the ADMM [24] and the proposed FUSE-ADMM with the TV-regularization.

3) Example 3 (HS+MS Fusion With TV Regularization): The third experiment is based on a TV regularization (can be interpreted as a specific instance of a non-Gaussian prior) studied in [24]. The regularization parameter of this model needs to be fixed by the user. The ADMM-based method investigated in [24] requires to compute a TV-based proximity operator (which increases the computational cost when compared to the previous algorithms). To solve this optimization problem, the frequency domain SE solution derived in Section IV-B2 can be embedded in an ADMM algorithm. The fusion results

Fig. 4. Hyperspectral pansharpening results. 1st column: HS image. 2nd column: PAN image. 3rd column: Reference image. 4th column: ADMM [23]. 5th column: Proposed method.

obtained with the ADMM method of [24] and the proposed FUSE-ADMM method are shown in Fig. 2 and are quite similar. The last two lines of Table II confirms this similarity more quantitatively by using the quality measures introduced in Section V-A. Note that the computational time obtained with the proposed explicit fusion solution is reduced when compared to the ADMM method. In order to complement this analysis, the convergence speeds of the FUSE-ADMM algorithm and the ADMM method of [24] are studied by analyzing the evolution of the objective function for the two fusion solutions. Fig. 3 shows that the FUSE-ADMM algorithm converges faster at the starting phase and gives smoother convergence result.

C. Hyperspectral Pansharpening

When $n_\lambda = 1$, the fusion of HS and MS images reduces to the HS pansharpening (HS+PAN) problem, which is the extension of conventional pansharpening (MS+PAN) and has become an important and popular application in the area of remote sensing [10]. In order to show that the proposed method is also applicable to this problem, we consider the fusion of HS and PAN images using another HS dataset. The reference image, considered here as the high-spatial and high-spectral image, is an HS image of size $396 \times 184 \times 176$ acquired over Moffett field, CA, in 1994 by the JPL/NASA airborne visible/infrared imaging spectrometer (AVIRIS) [45]. This image was initially composed of 224 bands that have been reduced to 176 bands after removing the water vapor absorption bands. The HS image has been generated by applying a 5×5 Gaussian filter on each band of the reference image. Besides, a PAN image is obtained by successively averaging the adjacent bands in visible bands ($1 \sim 41$ bands) according to realistic spectral responses. In addition, the HS and PAN images have been both contaminated by zero-mean additive Gaussian noises. The SNR of the HS image is 35dB for the first 126 bands and 30dB for the last remaining bands. The SNR of the PAN image is 30dB.

The FUSE based method is compared with the ADMM method[6] of [23] to solve the supervised pansharpening problem (i.e., with fixed hyperparameters). The results are displayed in Fig. 4 whereas more quantitative results are reported in Table III. Again, the proposed FUSE-based method provides similar qualitative and quantitative fusion results with a significant computational cost reduction. More results for the HS pansharpening are also available in a recently published review paper, where the authors compare eleven fusion algorithms, including the proposed FUSE-based strategies, on three different datasets [10].

[6]Due to space limitation, only the Gaussian prior of [23] is considered in this experiment. However, additional simulation results for other priors are available in the technical report [43].

TABLE III
PERFORMANCE OF THE PANSHARPENING METHODS: RSNR (IN dB), UIQI, SAM (IN DEGREE), ERGAS, DD (IN 10^{-2}) AND TIME (IN SECOND)

Methods	RSNR	UIQI	SAM	ERGAS	DD	Time
ADMM [23]	18.630	0.9800	4.952	3.564	1.306	94.36
FUSE	18.695	0.9803	4.904	3.541	1.292	0.39

VI. CONCLUSION

This paper developed a fast multi-band image fusion method based on an explicit solution of a Sylvester equation. This method was applied to both the fusion of multispectral and hyperspectral images and to the fusion of panchromatic and hyperspectral images. Coupled with the alternating direction method of multipliers and the block coordinate descent, the proposed algorithm can be easily generalized to compute Bayesian estimators for different fusion problems. Besides, the analytical solution of the Sylvester equation can be embedded in a block coordinate descent algorithm to compute the solution of a fusion model based on hierarchical Bayesian inference. Numerical experiments showed that the proposed fast fusion method compares competitively with the ADMM based methods, with the advantage of reducing the computational complexity significantly. Future work will consist of incorporating learning of the subspace transform matrix $\mathbf{H}$ into the fusion scheme. Implementing the proposed fusion scheme in real datasets will also be interesting.

Appendix A
Proof of Lemma 1

As $\mathbf{A}_1$ is symmetric (resp. Hermitian) positive definite, $\mathbf{A}_1$ can be decomposed as $\mathbf{A}_1 = \mathbf{A}_1^{\frac{1}{2}}\mathbf{A}_1^{\frac{1}{2}}$, where $\mathbf{A}_1^{\frac{1}{2}}$ is also symmetric (resp. Hermitian) positive definite thus invertible. Therefore, we have

$$\mathbf{A}_1\mathbf{A}_2 = \mathbf{A}_1^{\frac{1}{2}}\left(\mathbf{A}_1^{\frac{1}{2}}\mathbf{A}_2\mathbf{A}_1^{\frac{1}{2}}\right)\mathbf{A}_1^{-\frac{1}{2}}. \tag{29}$$

As $\mathbf{A}_1^{\frac{1}{2}}$ and $\mathbf{A}_2$ are both symmetric (resp. Hermitian) matrices, $\mathbf{A}_1^{\frac{1}{2}}\mathbf{A}_2\mathbf{A}_1^{\frac{1}{2}}$ is also a symmetric (resp. Hermitian) matrix that can be diagonalized. As a consequence, $\mathbf{A}_1\mathbf{A}_2$ is similar to a diagonalizable matrix, and thus it is diagonalizable.

Similarly, $\mathbf{A}_2$ can be written as $\mathbf{A}_2 = \mathbf{A}_2^{\frac{1}{2}}\mathbf{A}_2^{\frac{1}{2}}$, where $\mathbf{A}_2^{\frac{1}{2}}$ is positive semi-definite. Thus, $\mathbf{A}_1^{\frac{1}{2}}\mathbf{A}_2\mathbf{A}_1^{\frac{1}{2}} = \mathbf{A}_1^{\frac{1}{2}}\mathbf{A}_2^{\frac{1}{2}}\mathbf{A}_2^{\frac{1}{2}}\mathbf{A}_1^{\frac{1}{2}}$ is positive semi-definite showing that all its eigenvalues are non-negative. As similar matrices share equal similar eigenvalues, the eigenvalues of $\mathbf{A}_1\mathbf{A}_2$ are non-negative.

Appendix B
Proof of Lemma 2

The n dimensional DFT matrix $\mathbf{F}$ can be written explicitly as follows

$$\mathbf{F} = \frac{1}{\sqrt{n}}\begin{bmatrix} 1 & 1 & 1 & 1 & \cdots & 1 \\ 1 & \omega & \omega^2 & \omega^3 & \cdots & \omega^{n-1} \\ 1 & \omega^2 & \omega^4 & \omega^6 & \cdots & \omega^{2(n-1)} \\ 1 & \omega^3 & \omega^6 & \omega^9 & \cdots & \omega^{3(n-1)} \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & \omega^{n-1} & \omega^{2(n-1)} & \omega^{3(n-1)} & \cdots & \omega^{(n-1)(n-1)} \end{bmatrix}$$

where $\omega = e^{-\frac{2\pi i}{n}}$ is a primitive nth root of unity in which $i = \sqrt{-1}$. The matrix $\underline{\mathbf{S}}$ can also be written as follows

$$\underline{\mathbf{S}} = \mathbf{E}_1 + \mathbf{E}_{1+d} + \cdots + \mathbf{E}_{1+(m-1)d}$$

where $\mathbf{E}_i \in \mathbb{R}^{n\times n}$ is a matrix containing only one non-zero element equal to 1 located at the ith row and ith column as follows

$$\mathbf{E}_i = \begin{bmatrix} 0 & \cdots & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ 0 & \cdots & 1 & \cdots & 0 \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & \cdots & 0 \end{bmatrix}.$$

It is obvious that $\mathbf{E}_i$ is an idempotent matrix, i.e., $\mathbf{E}_i = \mathbf{E}_i^2$. Thus, we have

$$\mathbf{F}^H\mathbf{E}_i\mathbf{F} = (\mathbf{E}_i\mathbf{F})^H\mathbf{E}_i\mathbf{F} = \begin{bmatrix}\mathbf{0}^T \cdots \mathbf{f}_i^H \cdots \mathbf{0}^T\end{bmatrix}\begin{bmatrix}\mathbf{0} \\ \vdots \\ \mathbf{f}_i \\ \vdots \\ \mathbf{0}\end{bmatrix} = \mathbf{f}_i^H\mathbf{f}_i$$

where $\mathbf{f}_i = \frac{1}{\sqrt{n}}\begin{bmatrix}1 & \omega^{i-1} & \omega^{2(i-1)} & \omega^{3(i-1)} \cdots \omega^{(n-1)(i-1)}\end{bmatrix}$ is the ith row of the matrix $\mathbf{F}$ and $\mathbf{0} \in \mathbb{R}^{1\times n}$ is the zero vector of dimension $1 \times n$. Straightforward computations lead to

$$\mathbf{f}_i^H\mathbf{f}_i = \frac{1}{n}\begin{bmatrix} 1 & \omega^{i-1} & \cdots & \omega^{(i-1)(n-1)} \\ \omega^{-(i-1)} & 1 & \cdots & \omega^{(i-1)(n-2)} \\ \vdots & \vdots & \ddots & \vdots \\ \omega^{-(i-1)(n-1)} & \omega^{-(i-1)(n-2)} & \cdots & 1 \end{bmatrix}.$$

Using the ω's property $\sum_{i=1}^{n}\omega^i = 0$ and $n = md$ leads to

$$\begin{aligned}&\mathbf{f}_1^H\mathbf{f}_1 + \mathbf{f}_{1+d}^H\mathbf{f}_{1+d} + \cdots \mathbf{f}_{1+(m-1)d}^H\mathbf{f}_{1+(m-1)d} \\ &= \frac{1}{n}\begin{bmatrix} \begin{bmatrix} m & 0 & \cdots & 0 \\ 0 & m & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & m \end{bmatrix} & \cdots & \begin{bmatrix} m & 0 & \cdots & 0 \\ 0 & m & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & m \end{bmatrix} \\ \vdots & \ddots & \vdots \\ \begin{bmatrix} m & 0 & \cdots & 0 \\ 0 & m & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & m \end{bmatrix} & \cdots & \begin{bmatrix} m & 0 & \cdots & 0 \\ 0 & m & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & m \end{bmatrix} \end{bmatrix} \\ &= \frac{1}{d}\begin{bmatrix} \mathbf{I}_m & \cdots & \mathbf{I}_m \\ \vdots & \ddots & \vdots \\ \mathbf{I}_m & \cdots & \mathbf{I}_m \end{bmatrix} = \frac{1}{d}\mathbf{J}_d \otimes \mathbf{I}_m.\end{aligned}$$

Appendix C
Proof of Lemma 3

According to Lemma 2, we have

$$\mathbf{F}^H\underline{\mathbf{S}}\mathbf{F}\underline{\mathbf{D}} = \frac{1}{d}\left(\mathbf{J}_d \otimes \mathbf{I}_m\right)\underline{\mathbf{D}} = \frac{1}{d}\begin{bmatrix} \underline{\mathbf{D}}_1 & \underline{\mathbf{D}}_2 & \cdots & \underline{\mathbf{D}}_d \\ \vdots & \vdots & \ddots & \vdots \\ \underline{\mathbf{D}}_1 & \underline{\mathbf{D}}_2 & \cdots & \underline{\mathbf{D}}_d \end{bmatrix} \tag{30}$$

Thus, multiplying (30) by $\mathbf{P}$ on the left side and by $\mathbf{P}^{-1}$ on the right side leads to

$$\begin{aligned}\mathbf{M} &= \mathbf{P}\left(\mathbf{F}^H\underline{\mathbf{S}}\mathbf{F}\underline{\mathbf{D}}\right)\mathbf{P}^{-1} \\ &= \frac{1}{d}\begin{bmatrix} \underline{\mathbf{D}}_i & \underline{\mathbf{D}}_2 & \cdots & \underline{\mathbf{D}}_d \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} \end{bmatrix}\mathbf{P}^{-1} \\ &= \frac{1}{d}\begin{bmatrix} \sum_{i=1}^{d}\underline{\mathbf{D}}_i & \underline{\mathbf{D}}_2 & \cdots & \underline{\mathbf{D}}_d \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} \end{bmatrix}\end{aligned}$$

Appendix D
Proof of Theorem 1

Substituting (13) and (14) into (12) leads to (32), as shown at the top of the next page, where

$$\bar{\mathbf{C}}_3 = \begin{bmatrix} (\bar{\mathbf{C}}_3)_{1,1} & (\bar{\mathbf{C}}_3)_{1,2} & \cdots & (\bar{\mathbf{C}}_3)_{1,d} \\ (\bar{\mathbf{C}}_3)_{2,1} & (\bar{\mathbf{C}}_3)_{2,2} & \cdots & (\bar{\mathbf{C}}_3)_{2,d} \\ \vdots & \vdots & \ddots & \vdots \\ (\bar{\mathbf{C}}_3)_{d,1} & (\bar{\mathbf{C}}_3)_{d,2} & \cdots & (\bar{\mathbf{C}}_3)_{d,d} \end{bmatrix}. \tag{31}$$

$$\begin{bmatrix} \bar{\mathbf{u}}_{1,1}\left(\frac{1}{d}\sum_{i=1}^{d}\underline{\mathbf{D}}_i+\lambda_C^1\mathbf{I}_n\right) & \lambda_C^1\bar{\mathbf{u}}_{1,2}+\frac{1}{d}\bar{\mathbf{u}}_{1,1}\underline{\mathbf{D}}_2 & \cdots & \lambda_C^1\bar{\mathbf{u}}_{1,d}+\frac{1}{d}\bar{\mathbf{u}}_{1,1}\underline{\mathbf{D}}_d \\ \bar{\mathbf{u}}_{2,1}\left(\frac{1}{d}\sum_{i=1}^{d}\underline{\mathbf{D}}_i+\lambda_C^2\mathbf{I}_n\right) & \lambda_C^2\bar{\mathbf{u}}_{2,2}+\frac{1}{d}\bar{\mathbf{u}}_{2,1}\underline{\mathbf{D}}_2 & \cdots & \lambda_C^2\bar{\mathbf{u}}_{2,d}+\frac{1}{d}\bar{\mathbf{u}}_{2,1}\underline{\mathbf{D}}_d \\ \vdots & \vdots & \ddots & \vdots \\ \bar{\mathbf{u}}_{\widetilde{m}_\lambda,1}\left(\frac{1}{d}\sum_{i=1}^{d}\underline{\mathbf{D}}_i+\lambda_C^{\widetilde{m}_\lambda}\mathbf{I}_n\right) & \lambda_C^{\widetilde{m}_\lambda}\bar{\mathbf{u}}_{\widetilde{m}_\lambda,2}+\frac{1}{d}\bar{\mathbf{u}}_{\widetilde{m}_\lambda,1}\underline{\mathbf{D}}_2 & \cdots & \lambda_C^{\widetilde{m}_\lambda}\bar{\mathbf{u}}_{\widetilde{m}_\lambda,d}+\frac{1}{d}\bar{\mathbf{u}}_{\widetilde{m}_\lambda,1}\underline{\mathbf{D}}_d \end{bmatrix} = \bar{\mathbf{C}}_3 \tag{32}$$

Identifying the first (block) columns of (32) allows us to compute the element $\bar{\mathbf{u}}_{1,1}$ for $l = 1, ..., d$ as follows

$$\bar{\mathbf{u}}_{l,1} = (\bar{\mathbf{C}}_3)_{l,1}\left(\frac{1}{d}\sum_{i=1}^{d}\underline{\mathbf{D}}_i+\lambda_C^l\mathbf{I}_n\right)^{-1}$$

for $l = 1, \cdots, \widetilde{m}_\lambda$. Using the values of $\bar{\mathbf{u}}_{l,1}$ determined above, it is easy to obtain $\bar{\mathbf{u}}_{l,2}, \cdots, \bar{\mathbf{u}}_{l,d}$ as

$$\bar{\mathbf{u}}_{l,j} = \frac{1}{\lambda_C^l}\left[(\bar{\mathbf{C}}_3)_{l,j} - \frac{1}{d}\bar{\mathbf{u}}_{l,1}\underline{\mathbf{D}}_j\right]$$

for $l = 1, \cdots, \widetilde{m}_\lambda$ and $j = 2, \cdots, d$.

REFERENCES

[1] D. Landgrebe, "Hyperspectral image data analysis," *IEEE Signal Process. Mag.*, vol. 19, no. 1, pp. 17–28, Jan. 2002.

[2] K. Navulur, *Multispectral Image Analysis Using the Object-Oriented Paradigm* (Remote Sensing Applications Series). Boca Raton, FL, USA: CRC Press, 2006.

[3] R. Bacon *et al.*, "The SAURON project—I. The panoramic integral-field spectrograph," *Monthly Notices Roy. Astron. Soc.*, vol. 326, no. 1, pp. 23–35, 2001.

[4] C.-I. Chang, Ed., *Hyperspectral Data Exploitation: Theory and Applications*. New York, NY, USA: Wiley, 2007.

[5] H. Aanaes, J. R. Sveinsson, A. A. Nielsen, T. Bovith, and J. A. Benediktsson, "Model-based satellite image fusion," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 5, pp. 1336–1346, May 2008.

[6] T. Stathaki, *Image Fusion: Algorithms and Applications*. San Diego, CA, USA: Academic, 2011.

[7] M. Gong, Z. Zhou, and J. Ma, "Change detection in synthetic aperture radar images based on image fusion and fuzzy clustering," *IEEE Trans. Image Process.*, vol. 21, no. 4, pp. 2141–2151, Apr. 2012.

[8] A. P. James and B. V. Dasarathy, "Medical image fusion: A survey of the state of the art," *Inf. Fusion*, vol. 19, pp. 4–19, Sep. 2014.

[9] B. Aiazzi, L. Alparone, S. Baronti, A. Garzelli, and M. Selva, "25 years of pansharpening: A critical review and new developments," in *Signal and Image Processing for Remote Sensing*, C. H. Chen, Ed., 2nd ed. Boca Raton, FL, USA: CRC Press, 2011, ch. 28, pp. 533–548.

[10] L. Loncan *et al.*, "Hyperspectral pansharpening: A review," *IEEE Geosci. Remote Sens. Mag.*, to appear.

[11] N. Yokoya, N. Mayumi, and A. Iwasaki, "Cross-calibration for data fusion of EO-1/hyperion and terra/ASTER," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 6, no. 2, pp. 419–426, Apr. 2013.

[12] I. Amro, J. Mateos, M. Vega, R. Molina, and A. K. Katsaggelos, "A survey of classical methods and new trends in pansharpening of multispectral images," *EURASIP J. Adv. Signal Process.*, vol. 2011, no. 79, pp. 1–22, 2011.

[13] M. González-Audícana, J. L. Saleta, R. G. Catalán, and R. García, "Fusion of multispectral and panchromatic images using improved IHS and PCA mergers based on wavelet decomposition," *IEEE Trans. Geosci. Remote Sens.*, vol. 42, no. 6, pp. 1291–1299, Jun. 2004.

[14] R. C. Hardie, M. T. Eismann, and G. L. Wilson, "MAP estimation for hyperspectral image resolution enhancement using an auxiliary sensor," *IEEE Trans. Image Process.*, vol. 13, no. 9, pp. 1174–1184, Sep. 2004.

[15] R. Molina, A. K. Katsaggelos, and J. Mateos, "Bayesian and regularization methods for hyperparameter estimation in image restoration," *IEEE Trans. Image Process.*, vol. 8, no. 2, pp. 231–246, Feb. 1999.

[16] R. Molina, M. Vega, J. Mateos, and A. K. Katsaggelos, "Variational posterior distribution approximation in Bayesian super resolution reconstruction of multispectral images," *Appl. Comput. Harmon. Anal.*, vol. 24, no. 2, pp. 251–267, 2008.

[17] Y. Zhang, A. Duijster, and P. Scheunders, "A Bayesian restoration approach for hyperspectral images," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 9, pp. 3453–3462, Sep. 2012.

[18] N. Yokoya, T. Yairi, and A. Iwasaki, "Coupled nonnegative matrix factorization unmixing for hyperspectral and multispectral data fusion," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 2, pp. 528–537, Feb. 2012.

[19] Q. Wei, N. Dobigeon, and J.-Y. Tourneret, "Bayesian fusion of hyperspectral and multispectral images," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, Florence, Italy, May 2014, pp. 3176–3180.

[20] Q. Wei, J. Bioucas-Dias, N. Dobigeon, and J.-Y. Tourneret, "Hyperspectral and multispectral image fusion based on a sparse representation," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 7, pp. 3658–3668, Jul. 2015.

[21] N. Yokoya and A. Iwasaki, "Hyperspectral and multispectral data fusion mission on hyperspectral imager suite (HISUI)," in *Proc. IEEE Int. Conf. Geosci. Remote Sens. Symp. (IGARSS)*, Melbourne, VIC, Australia, Jul. 2013, pp. 4086–4089.

[22] Q. Wei, N. Dobigeon, and J.-Y. Tourneret, "Bayesian fusion of multi-band images," *IEEE J. Sel. Topics Signal Process.*, to appear.

[23] Q. Wei, N. Dobigeon, and J.-Y. Tourneret, "Bayesian fusion of multispectral and hyperspectral images using a block coordinate descent method," in *Proc. IEEE GRSS Workshop Hyperspectral Image Signal Process., Evol. Remote Sens. (WHISPERS)*, Tokyo, Japan, Jun. 2015, pp. 1–5.

[24] M. Simoes, J. Bioucas-Dias, L. B. Almeida, and J. Chanussot, "A convex formulation for hyperspectral image superresolution via subspace-based regularization," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 6, pp. 3373–3388, Jun. 2015.

[25] C. P. Robert, *The Bayesian Choice: From Decision-Theoretic Foundations to Computational Implementation* (Springer Texts in Statistics), 2nd ed. New York, NY, USA: Springer-Verlag, 2007.

[26] A. Gelman, J. B. Carlin, H. S. Stern, D. B. Dunson, A. Vehtari, and D. B. Rubin, *Bayesian Data Analysis*, 3rd ed. Boca Raton, FL, USA: CRC Press, 2013.

[27] M. V. Afonso, J. M. Bioucas-Dias, and M. A. T. Figueiredo, "An augmented Lagrangian approach to the constrained optimization formulation of imaging inverse problems," *IEEE Trans. Image Process.*, vol. 20, no. 3, pp. 681–695, Mar. 2011.

[28] C.-I. Chang, X.-L. Zhao, M. L. G. Althouse, and J. J. Pan, "Least squares subspace projection approach to mixed pixel classification for hyperspectral images," *IEEE Trans. Geosci. Remote Sens.*, vol. 36, no. 3, pp. 898–912, May 1998.

[29] J. M. Bioucas-Dias and J. M. P. Nascimento, "Hyperspectral subspace identification," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 8, pp. 2435–2445, Aug. 2008.

[30] C. L. Lawson and R. J. Hanson, *Solving Least Squares Problems*, vol. 161. Englewood Cliffs, NJ, USA: Prentice-Hall, 1974.

[31] R. L. Lagendijk and J. Biemond, *Iterative Identification and Restoration of Images*, vol. 118. New York, NY, USA: Springer-Verlag, 1990.

[32] M. Elad and A. Feuer, "Restoration of a single superresolution image from several blurred, noisy, and undersampled measured images," *IEEE Trans. Image Process.*, vol. 6, no. 12, pp. 1646–1658, Dec. 1997.

[33] M. Elad and Y. Hel-Or, "A fast super-resolution reconstruction algorithm for pure translational motion and common space-invariant blur," *IEEE Trans. Image Process.*, vol. 10, no. 8, pp. 1187–1193, Aug. 2001.

[34] S. C. Park, M. K. Park, and M. G. Kang, "Super-resolution image reconstruction: A technical overview," *IEEE Signal Process. Mag.*, vol. 20, no. 3, pp. 21–36, May 2003.
[35] R. H. Bartels and G. W. Stewart, "Solution of the matrix equation AX + XB = C [F4]," *Commun. ACM*, vol. 15, no. 9, pp. 820–826, 1972.
[36] R. A. Horn and C. R. Johnson, *Matrix Analysis*. Cambridge, U.K: Cambridge Univ. Press, 2012.
[37] R. C. Hardie, K. J. Barnard, and E. E. Armstrong, "Joint MAP registration and high-resolution image estimation using a sequence of undersampled images," *IEEE Trans. Image Process.*, vol. 6, no. 12, pp. 1621–1633, Dec. 1997.
[38] M. T. Eismann and R. C. Hardie, "Application of the stochastic mixing model to hyperspectral resolution enhancement," *IEEE Trans. Geosci. Remote Sens.*, vol. 42, no. 9, pp. 1924–1933, Sep. 2004.
[39] N. A. Woods, N. P. Galatsanos, and A. K. Katsaggelos, "Stochastic methods for joint registration, restoration, and interpolation of multiple undersampled images," *IEEE Trans. Image Process.*, vol. 15, no. 1, pp. 201–213, Jan. 2006.
[40] A. N. Tikhonov and V. Y. Arsenin, *Solutions of Ill-Posed Problems* (Scripta Series in Mathematics). Great Falls, MT, USA: Winston, 1977.
[41] P. L. Combettes and J.-C. Pesquet, "Proximal splitting methods in signal processing," in *Fixed-Point Algorithms for Inverse Problems in Science and Engineering* (Springer Optimization and Its Applications), H. H. Bauschke, R. S. Burachik, P. L. Combettes, V. Elser, D. R. Luke, and H. Wolkowicz, Eds. New York, NY, USA: Springer-Verlag, 2011, pp. 185–212.
[42] Y. Zhang, S. De Backer, and P. Scheunders, "Noise-resistant wavelet-based Bayesian fusion of multispectral and hyperspectral images," *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 11, pp. 3834–3843, Nov. 2009.
[43] Q. Wei, N. Dobigeon, and J.-Y. Tourneret. (Jun. 2015). "Fast fusion of multi-band images based on solving a Sylvester equation—Complementary results and supporting materials," Dept. IRIT/INP-ENSEEIHT, Univ. Toulouse, Toulouse, France, Tech. Rep. IRIT-ENSEEIHT. [Online]. Available: http://wei.perso.enseeiht.fr/papers/2015-WEI-TR-Sylvester.pdf2015
[44] D. J. Fleming, "Effect of relative spectral response on multi-spectral measurements and NDVI from different remote sensing systems," Ph.D. dissertation, Dept. Geogr., Univ. Maryland, College Park, MD, USA, 2006.
[45] R. O. Green *et al.*, "Imaging spectroscopy and the airborne visible/infrared imaging spectrometer (AVIRIS)," *Remote Sens. Environ.*, vol. 65, no. 3, pp. 227–248, 1998.

Qi Wei (S'13) was born in Shanxi, China, in 1989. He received the B.Sc. degree in electrical engineering from Beihang University, Beijing, China, in 2010. He is currently pursuing the Ph.D. degree with INP–ENSEEIHT, National Polytechnic Institute of Toulouse, University of Toulouse. In 2012, he was an Exchange Master Student with the Signal Processing and Communications Group, Department of Signal Theory and Communications, Universitat Politècnica de Catalunya. He is also with the Signal and Communications Group, IRIT Laboratory. His research has been focused on Bayesian estimation in statistical signal processing, in particular, inverse problems in image processing. He has served as a reviewer of the IEEE JOURNAL OF SELECTED TOPICS IN SIGNAL PROCESSING, the IEEE TRANSACTIONS ON IMAGE PROCESSING, the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING, and several conferences.

Nicolas Dobigeon (S'05–M'08–SM'13) was born in Angoulême, France, in 1981. He received the Engineering degree in electrical engineering from ENSEEIHT, Toulouse, France, in 2004, and the M.Sc., Ph.D., and Habilitation Diriger des Recherches degrees in signal processing from INP Toulouse, in 2004, 2007, and 2012, respectively. From 2007 to 2008, he was a Post-Doctoral Research Associate with the Department of Electrical Engineering and Computer Science, University of Michigan, Ann Arbor.

He has been with INP Toulouse, University of Toulouse, since 2008, where he is currently an Associate Professor. He conducts his research within the Signal and Communications Group, IRIT Laboratory, and is also an Affiliated Faculty Member with the TeSA Laboratory. His recent research activities have been focused on statistical signal and image processing, with a particular interest in Bayesian inverse problems with applications to remote sensing, biomedical imaging, and genomics.

Jean-Yves Tourneret (SM'08) received the Ingénieur degree in electrical engineering from ENSEEIHT, Toulouse, in 1989, and the Ph.D. degree from INP Toulouse, in 1992. He is currently a Professor with the University of Toulouse (ENSEEIHT) and a member of the IRIT Laboratory (UMR 5505 of the CNRS). His research activities are centered around statistical signal and image processing with a particular interest to Bayesian and Markov chain Monte Carlo methods.

He has been involved in the organization of several conferences, including the European Conference on Signal Processing in 2002 (Program Chair), ICASSP'06 (plenaries), the Statistical Signal Processing Workshop (SSP) in 2012 (international liaisons), the International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP) in 2013 (local arrangements), SSP (special sessions) in 2014, and the Workshop on Machine Learning for Signal Processing in 2014 (special sessions). He has been the General Chair of the CIMI Workshop on Optimization and Statistics in Image Processing in Toulouse in 2013 (with F. Malgouyres and D. Kouamé) and CAMSAP in 2015 (with P. Djuric). He has been a member of different technical committees, including the Signal Processing Theory and Methods Committee of the IEEE Signal Processing Society (2001–2007, 2010–present). He has served as an Associate Editor of the IEEE TRANSACTIONS ON SIGNAL PROCESSING (2008–2011) and the *EURASIP Journal on Signal Processing* (since 2013).