



HAL
open science

Regularization parameter estimation for non-negative hyperspectral image deconvolution

Yingying Song, David Brie, El-Hadi Djermoune, Simon Henrot

► **To cite this version:**

Yingying Song, David Brie, El-Hadi Djermoune, Simon Henrot. Regularization parameter estimation for non-negative hyperspectral image deconvolution. *IEEE Transactions on Image Processing*, 2016, 25 (11), pp.5316-5330. 10.1109/TIP.2016.2601489 . hal-01358458

HAL Id: hal-01358458

<https://hal.science/hal-01358458>

Submitted on 31 Aug 2016

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

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

Regularization parameter estimation for non-negative hyperspectral image deconvolution

Yingying Song, *Student Member, IEEE*, David Brie, *Member, IEEE*, El-Hadi Djermoune, *Member, IEEE*, and Simon Henrot

Abstract

This work aims at studying a method to automatically estimate regularization parameters of non-negative hyperspectral image deconvolution methods. The deconvolution problem is formulated as a multi-objective optimization problem and the properties of the corresponding response surface are studied. Based on these properties, the minimum distance criterion (MDC) and the maximum curvature criterion (MCC) are proposed to estimate regularization parameters especially for the non-negativity constrained deconvolution problem. MDC has good theoretical properties (convexity, uniqueness) but requires to choose a reference point. On the contrary, MCC does not need to choose any reference point but does not have interesting theoretical properties. A grid-search-based approach to minimize the computational cost of MDC and MCC is proposed. It results in fast approaches to estimate the regularization parameters. Based on simulated 2D images, the proposed approaches are compared to state-of-the-art methods, confirming the effectiveness of the MDC and MCC for the non-negativity constrained image deconvolution problem. In the case of non-negative hyperspectral image deconvolution, the fast MDC yields better performances than the fast MCC. An application to real-world hyperspectral fluorescence microscopy images is also provided; it confirms the superiority of MDC.

Keywords

Non-negative hyperspectral image deconvolution, multi-objective optimization, regularization parameter estimation

I. INTRODUCTION

A hyperspectral image can be viewed as a stack of images obtained for different wavelengths. The observed images may suffer from degradation due to the measuring device, resulting in a convolution or blurring of the images. Hyperspectral image deconvolution consists in removing the blur to restore the original images at best. This problem arises in a number of applications including microscopy [1] [2] [3], astronomy [4] [5] [6] and industrial hyperspectral imaging systems [7] [8]. Actually, hyperspectral image deconvolution is required as soon as the spatial

This work has been supported by the FUI AAP 2015 Trispirabois Project and the Conseil Régional de Lorraine.

Centre de Recherche en Automatique de Nancy (CRAN), Université de Lorraine, CNRS, Boulevard des Aiguillettes B.P. 239 F-54506 Vandœuvre-lès-Nancy, France (email: `firstname.lastname@univ-lorraine.fr` for Yingying Song, David Brie and El-Hadi Djermoune; `simon.henrot@gmail.com` for Simon Henrot).

resolution has to be increased (super-resolution) [9]. Similar problems also arise in x-ray fluorescence tomography [10] [11], even if the problem at hand is not a deconvolution problem but rather a reconstruction problem.

The deconvolution of hyperspectral images is an ill-posed inverse problem which suffers from instability. To recover accurately the original images, it is necessary to resort to some regularization techniques. This can be done by formulating the problem as the minimization of a penalized criterion incorporating prior information enforcing the spatial and spectral regularity as well as the non-negativity of the image to recover. Different hyperspectral image deconvolution methods were proposed in [12], [5], [13], [14] and [3]. They all consider separable spatial and spectral regularization terms. The effective implementation of such methods is hampered by the choice of the regularization parameters. In general, this choice is made by successive trials which can be highly time consuming. Here, we focus on Tikhonov-like hyperspectral image deconvolution with non-negativity constraint proposed in [13].

A classical approach to estimate a single regularization parameter of Tikhonov-based deconvolution is the generalized cross-validation method [15]. It was used for choosing the regularization parameter in image deconvolution in [16]. The L-curve presented in [17] and [18] is also a method for selecting a single regularization parameter of the Tikhonov-based deconvolution. Plotting in a log-log scale the data fitting term versus the penalty term yields a curve which exhibits a corner. The curvature is expected to reach a maximum value yielding an estimated regularization parameter which provides an acceptable trade-off between these two terms. However, the L-curve approach has some undesirable properties discussed in [19] and [20]. In particular, it is not convex and the maximum curvature is not unique. The L-hypersurface as a multi-objective extension of the L-curve for selecting multiple regularization parameters was introduced in [21]. However, since the curvature is not uniquely defined, the maximum curvature approach is not an appropriate choice. Thus, in [21], the minimum distance criterion (MDC), which was already introduced in [22] for the bi-objective case, is applied to the L-hypersurface for estimating the regularization parameters; this leads to a simple fixed-point iterative algorithm for computing regularization parameters in both bi-objective and multi-objective cases. But this approach can only be applied to the unconstrained Tikhonov-based deconvolution.

The goal of the present paper is to propose a general approach to estimate the regularization parameters of hyperspectral image deconvolution methods formulated as a convex multi-objective minimization problem. A key point is that it can be used indifferently for unconstrained and constrained problems. Addressing deconvolution as a multi-objective optimization problem is not very common. To the best of our knowledge, [23] is the first work mentioning the link between the L-curve and the multi-objective optimization. It is also mentioned that the use of the log-log scale results in a loss of convexity of the L-curve. Recently, [24] formulated the basis pursuit as a convex bi-objective optimization problem and proved that the corresponding Pareto front is convex and continuously differentiable over all points of interest. In fact, the Pareto front of basis pursuit is strongly connected to the regularization path for which a continuation-based approach allows fast calculation of the set of solutions when the regularization parameter is varying from 0 to $+\infty$ [25]. Two important results are proved in this paper. Firstly, the multi-objective criterion being composed of convex cost functions, its response surface is proved to be convex. This property holds for both the unconstrained and constrained cases. Secondly, as far as we know, no work is explicitly analyzing the impact of the non-negativity constraint on the regularization parameter estimation of the deconvolution algorithm. Here, we also prove that the non-negativity constraint results in a folding of the

response surface. The beneficial consequences of these two properties on the regularization parameter estimation are discussed. This will be supported by extensive simulations aiming at evaluating the Mean Squares Error (MSE) as a function of the Signal to noise Ratio (SNR).

The paper is organized as follows: in section II, we present the non-negative hyperspectral image deconvolution problem. In section III, it is formulated as a multi-objective optimization problem and the properties of the corresponding response surface are studied. In section IV, to estimate the regularization parameters, the maximum curvature criterion (MCC) and the MDC (directly applied to the response surface) are proposed and their properties are studied. To reduce the computational burden of MCC and MDC, a grid-search strategy is proposed: it is proved to be convergent for the MDC but not for the MCC. In section V and in the supplementary material [26], numerical experiments allowing to assess the performances of the proposed approaches and to compare them with state-of-the-art methods are presented. Finally, these approaches are applied to hyperspectral fluorescence microscopy data.

II. HYPERSPECTRAL IMAGE DECONVOLUTION

Hyperspectral imaging consists in observing a spatial scene at several wavelengths. Physically, such an image can be obtained as a stack of two-dimensional (2D) images equipped with optical filters or as a collection of one-dimensional (1D) spectra acquired by a spectrometer. Hyperspectral imaging is used in a wide range of applications including remote sensing [27], chemistry [28], [29], food science [30], biology [31] and medical imaging [32]. Among the different spectroscopic techniques allowing to produce hyperspectral images, we can mention Infra Red (IR), Raman [33] and fluorescence [34] microscopies. The problem at hand aims at removing the blur affecting the observed images. Such a blurring arises, for example, when we want to increase the spatial resolution of the imaging spectrometer. To do that, it is necessary to choose a spatial sampling lower than the instrument resolution.

A. Discrete representation of the blurred images

The unknown hyperspectral image is denoted by \mathbf{X} and the observed image by \mathbf{Y} . Considering that the discrete image \mathbf{X} has L wavelengths $\lambda_1, \dots, \lambda_L$, it can be seen as a stack of images $\{\mathbf{X}_l, l = 1 \dots L\}$. \mathbf{X}_l is a matrix of size $N_1 \times N_2$. By concatenating the columns of each image \mathbf{X}_l , the hyperspectral image can be reorganized into L vectors $\{\mathbf{x}_l, l = 1 \dots L, \}$ of length $N = N_1 \cdot N_2$ each, or a single vector \mathbf{x} of length NL . We use similar notations for the observed image, substituting letter y to letter x .

The blurred image corresponds to the 2D (circular¹) convolution of \mathbf{X}_l with filter \mathcal{H}_l . An equivalent formulation of the 2D convolution is obtained by defining a circulant-block-circulant convolution matrix \mathbf{H}_l of size $N \times N$. The discrete convolution can be written in matrix form as (see [35] for details):

$$\mathcal{H}_l \underset{(2D)}{*} \mathbf{X}_l = \mathbf{H}_l \mathbf{x}_l. \quad (1)$$

If we assume that the blurring affecting each spectral slice is different, then the global convolution matrix \mathbf{H} yielding the (vectorized) hyperspectral spectral image \mathbf{y} is block-diagonal; each block \mathbf{H}_l is the convolution matrix

¹In what follows, we will consider circular convolution which results in an exact discrete Fourier domain implementation of the convolution.

corresponding to the wavelength λ_l :

$$\mathbf{H} = \begin{bmatrix} \mathbf{H}_1 & 0 & \dots & 0 \\ 0 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & \mathbf{H}_L \end{bmatrix}. \quad (2)$$

Finally, the blurred and noisy spectral image is obtained by adding a noise term \mathbf{e} which results in the observation model:

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{e}. \quad (3)$$

The problem in hyperspectral image deconvolution is then the inverse problem which aims at finding the original image \mathbf{x} from the observed one.

B. Hyperspectral image deconvolution

1) *Spatial and spectral regularization*: Henrot *et al.* [36], [35] proposes to add a spectral regularization to the traditional Tikhonov [37] approach yielding a criterion composed of three terms: the data fitting, the spatial regularization and the spectral regularization:

$$\min_{\mathbf{x}} J(\mathbf{x}) = \frac{1}{2} \|\mathbf{y} - \mathbf{H}\mathbf{x}\|_2^2 + \frac{\mu_s}{2} \|\mathbf{D}_s \mathbf{x}\|_2^2 + \frac{\mu_\lambda}{2} \|\mathbf{D}_\lambda \mathbf{x}\|_2^2. \quad (4)$$

Here, μ_s and μ_λ are respectively the spatial and spectral regularization parameters. \mathbf{D}_s corresponds to a Laplacian filter and \mathbf{D}_λ corresponds to a first-order derivative filter along the spectral dimension.

The solution of problem (4) is given by

$$\mathbf{x}^* = (\mathbf{H}^T \mathbf{H} + \mu_s \mathbf{D}_s^T \mathbf{D}_s + \mu_\lambda \mathbf{D}_\lambda^T \mathbf{D}_\lambda)^{-1} \mathbf{H}^T \mathbf{y}. \quad (5)$$

Introducing the spectral regularization results in a coupling of both spatial and spectral dimensions: the slices of the hyperspectral images cannot be processed independently. Following [38], the Laplacian (second-order derivative) and first-order derivative are valid regularization differential operators. A rule of thumb to choose the differential operator is as follows: the identity operator will favor the reconstruction of null signals, first-order derivative will favor the reconstruction of constant signals, second-order derivative will favor the reconstruction of linear signals. It is worth noticing that these operators can be replaced by others. However, in practice, the choice of the differential operator is not so crucial since the criterion also includes the data fitting term and the trade-off between the data fitting and regularization terms is controlled by the regularization parameters. In fact, only the choice of the regularization parameters really matters.

2) *Restoration with non-negativity constraint*: The solution expressed in (5) cannot guarantee the non-negativity of the restored image. As proposed in [13], we can add a non-negativity constraint resulting in the following constrained optimization problem:

$$\begin{aligned} \min_{\mathbf{x}} J(\mathbf{x}) &= \frac{1}{2} \|\mathbf{y} - \mathbf{H}\mathbf{x}\|_2^2 + \frac{\mu_s}{2} \|\mathbf{D}_s \mathbf{x}\|_2^2 + \frac{\mu_\lambda}{2} \|\mathbf{D}_\lambda \mathbf{x}\|_2^2 \\ \text{s.t. } \mathbf{x} &\geq 0. \end{aligned} \quad (6)$$

To solve this problem, we use the quadratic penalty method proposed in [13] which consists in introducing a slack variable \mathbf{p} . The original problem is then replaced by a surrogate criterion expressed in (7)

$$\begin{aligned} \min_{\mathbf{x}, \mathbf{p}} K(\mathbf{x}, \mathbf{p}; \xi) &= J(\mathbf{x}) + \frac{\xi}{2} \|\mathbf{x} - \mathbf{p}\|_2^2 \\ \text{s.t. } \mathbf{p} &\geq 0. \end{aligned} \quad (7)$$

The solution is obtained iteratively. At each iteration, the following three steps are performed:

- unconstrained minimization of $K(\mathbf{x}, \mathbf{p}; \xi)$ with respect to \mathbf{x} ;
- constrained minimization of $K(\mathbf{x}, \mathbf{p}; \xi)$ with respect to \mathbf{p} ;
- increase of the penalty factor ξ .

These three steps are alternated until a maximum number of iterations N_{iter} is reached.

At each iteration $k = 1, \dots, N_{iter}$, when \mathbf{p}^k and ξ^k are fixed, $K(\mathbf{x}|\mathbf{p}^k, \xi^k)$ can be minimized explicitly

$$\mathbf{x}^{k+1} = (\mathbf{H}^T \mathbf{H} + \mu_s \mathbf{D}_s^T \mathbf{D}_s + \mu_\lambda \mathbf{D}_\lambda^T \mathbf{D}_\lambda + \xi^k \mathbf{I}_{NL})^{-1} (\mathbf{H}^T \mathbf{y} + \xi^k \mathbf{p}^k). \quad (8)$$

Once \mathbf{x}^{k+1} is obtained, \mathbf{p} can be updated according to

$$\mathbf{p}^{k+1} = \max(\mathbf{0}, \mathbf{x}^{k+1}). \quad (9)$$

A detailed analysis of the quadratic penalty method, including convergence, can be found in [39]. In practice, increasing the value of ξ will ensure the solution to converge to the minimum of the constrained problem. Following [39], the simplest choice is $\xi^{(k+1)} = \gamma \xi^{(k)}$ with $\gamma > 1$. The initial value of ξ should be large enough. Indeed, if it is too small, a large number of iterations may be required to reach the optimum. The choice of γ also influences the convergence rate. As the minimization of $K(\mathbf{x}|\mathbf{p}^k, \xi^k)$ yields the explicit solution (8), γ can be set to a large value. Here, the initial value of ξ is set to 1 and $\gamma = 10$.

III. HYPERSPECTRAL IMAGE DECONVOLUTION AS A MULTI-OBJECTIVE OPTIMIZATION

The starting point of our problem is the hyperspectral image deconvolution (HID) in (6). The optimal solution \mathbf{x}^* depends on both μ_s and μ_λ . If the value of μ_s increases toward infinity, the term $\|\mathbf{D}_s \mathbf{x}\|_2^2$ will be minimized. Similarly, the value of μ_λ increasing toward infinity will minimize $\|\mathbf{D}_\lambda \mathbf{x}\|_2^2$. When μ_s and μ_λ approach zero, the solution tend to minimize the data fitting term $\|\mathbf{y} - \mathbf{H}\mathbf{x}\|_2^2$. When μ_s is very small, the resulting deconvolution is generally not satisfactory because the noise is not sufficiently rejected. On the other hand, for very small μ_λ , the intensities of two adjacent spectral bands are not similar enough. But when both are large, the error between the solution and the observed image increases. This means that we cannot improve one objective without deteriorating the others. In this section, by stating the problem as a convex multiple objective optimization problem, it is possible to estimate the response surface from which the Pareto front can be deduced: this gives a characterization of the set of solutions obtained by varying the values of $\boldsymbol{\mu} = (\mu_s, \mu_\lambda)$. While the notions presented here are mainly concerned with the HID, the problem is formulated in a much more general setting which is the minimization of cost functions consisting in the weighted sum of convex objectives.

A. Multi-objective Optimization

A generic multi-objective optimization problem may be formulated as:

$$\min_{\mathbf{x}} \mathbf{J}(\mathbf{x}) = (J_1(\mathbf{x}), \dots, J_z(\mathbf{x})). \quad (10)$$

Here, \mathbf{J} is a (vector) criterion of z (which equals 3 in our case) objective functions which defines a multi-dimensional space.

1) *Ideal objective vector*: The ideal objective vector is defined as in [40]:

$$\mathbf{I} = (I_1, \dots, I_z). \quad (11)$$

The i -th component of \mathbf{I} is the minimum of the problem:

$$I_i = \min_{\mathbf{x}} J_i(\mathbf{x}), \quad i = 1, \dots, z. \quad (12)$$

The ideal objective vector \mathbf{I} corresponds to the point whose coordinates are minimum among each objective. It is termed as ideal because the value of each objective is the smallest and that is exactly the goal of the minimization problem (10). Most of the time, it cannot be reached because the objectives are conflicting: decreasing one will increase the others.

2) *Pareto front*: The notion of *domination* is defined for example in [40]. It is an important notion in multi-objective optimization since it allows to define a solution ordering, i.e. we can say that a solution is better than another one. Let $\mathbf{x}^{(1)}$ and $\mathbf{x}^{(2)}$ be two different solutions of the multi-objective minimization problem (10). The solution $\mathbf{x}^{(1)}$ is said to dominate $\mathbf{x}^{(2)}$ and we write $\mathbf{x}^{(1)} \preceq \mathbf{x}^{(2)}$, if the solution $\mathbf{x}^{(1)}$ is not worse than $\mathbf{x}^{(2)}$ in all objectives, and the solution $\mathbf{x}^{(1)}$ is strictly better than $\mathbf{x}^{(2)}$ in at least one objective:

$$\mathbf{x}^{(1)} \preceq \mathbf{x}^{(2)} \text{ iff } \begin{cases} J_i(\mathbf{x}^{(1)}) \leq J_i(\mathbf{x}^{(2)}), \forall i \in [1, \dots, z] \\ \exists j \in [1, \dots, z] \quad \text{s.t.} \quad J_j(\mathbf{x}^{(1)}) < J_j(\mathbf{x}^{(2)}). \end{cases} \quad (13)$$

Otherwise, the solution $\mathbf{x}^{(1)}$ does not dominate the solution $\mathbf{x}^{(2)}$. A solution is either dominated or non-dominated but cannot be both at the same time. The solution $\tilde{\mathbf{x}}$ is said to be non-dominated or Pareto-optimal for a multi-objective problem if all other vectors \mathbf{x} in the set of all feasible points have a higher value for at least one of the objectives J_i with $i = 1, \dots, z$. The set of all the non-dominated solutions is called Pareto front or Pareto curve or surface which means that each solution belonging to the Pareto front cannot be said better than another in the sense of domination. The shape of the Pareto surface reveals the nature of the trade-off between the different objective functions. In multi-objective optimization, the goal is to find the set of Pareto-optimal solutions rather than a single solution. Two different cases have to be distinguished:

- the case of convex criteria for which it is proved that any point of the Pareto front can be reached using the weighted sum approach;
- the case of non-convex criteria for which the weighted sum approach cannot find the non-convex part of the Pareto front [41]. A lot of attention was paid to this case (see for example [42] [40] [43] [44]).

In our case, convex criteria are considered, but we only search for a single solution being optimal according to a given criterion; the multi-objective optimization formalism is used to derive and analyze such a criterion.

3) *Weighted-sum approach*: The weighted-sum approach can solve the multi-objective problem by combining all of the objectives into a single one. With this method, the weights between objectives are assigned *a priori* before the optimization process is completed. With z objectives, the equivalent scalar objective $J(\mathbf{x}_w)$ is given by:

$$\begin{aligned} J(\mathbf{x}_w) &= \sum_{j=1}^z w_j J_j(\mathbf{x}) \\ &= \mathbf{w}^T \mathbf{J}(\mathbf{x}). \end{aligned} \quad (14)$$

Here is an example of the weighted-sum method with two objectives for which (14) simplifies to:

$$J(\mathbf{x}_{w_1, w_2}) = w_1 J_1(\mathbf{x}) + w_2 J_2(\mathbf{x}) \quad (15)$$

$$\text{and } w_1 + w_2 = 1, w_1 \geq 0, w_2 \geq 0.$$

If the weight vector is parameterized by α , so that $w_1 = 1 - \alpha$ and $w_2 = \alpha$, then the problem becomes:

$$J(\mathbf{x}_\alpha) = (1 - \alpha) J_1(\mathbf{x}) + \alpha J_2(\mathbf{x}) \quad (16)$$

$$\text{and } 0 \leq \alpha < 1.$$

In our minimization problem with the non-negativity constraint, the following two formulations are equivalent since $\mu = \alpha/(1 - \alpha)$.

$$\min_{\mathbf{x} \geq 0} (1 - \alpha) J_1(\mathbf{x}) + \alpha J_2(\mathbf{x}) \iff \min_{\mathbf{x} \geq 0} J_1(\mathbf{x}) + \mu J_2(\mathbf{x}). \quad (17)$$

In the case of three objectives, the criteria in (6) can also be written as:

$$\min_{\mathbf{x} \geq 0} J(\mathbf{x}) = J_1(\mathbf{x}) + \mu_s J_2(\mathbf{x}) + \mu_\lambda J_3(\mathbf{x}). \quad (18)$$

Each value of $\boldsymbol{\mu} = (\mu_s, \mu_\lambda)$ yields a solution:

$$\mathbf{x}_\boldsymbol{\mu} = \arg \min_{\mathbf{x} \geq 0} J(\mathbf{x}) \quad (19)$$

and gives a point in the response surface which will be denoted by $\Pi(\boldsymbol{\mu})$. Unlike the L-curve or the L-hypersurface, this response surface uses linear scales axes. For notation simplicity we will write $J_i(\mathbf{x}_\boldsymbol{\mu}) \triangleq J_i(\boldsymbol{\mu})$ and the same for $J(\mathbf{x}_\boldsymbol{\mu}) \triangleq J(\boldsymbol{\mu})$.

A necessary condition for the recovery of the Pareto front using the weighted-sum method is that all objectives are convex functions of \mathbf{x} which is the case here. In the next section, an equivalent constrained minimization formulation will be used to prove that the response surface of a convex tri-objective is convex as well.

B. Properties of the response surface

1) *Evaluating the response surface for HID problem*: To estimate the response surface of HID problem, we define (μ_s, μ_λ) on a 2D grid. Then, for each couple of parameters, the corresponding solution \mathbf{x}^* is computed using the algorithm presented in section II-B. To simplify notation, the dependence of \mathbf{x}^* on μ_s and μ_λ is omitted.

Figure 1 shows three different empirical response surfaces estimated from the same simulated example (see section V) for $N_{iter} = 1, 5, 10$. For each response surface, the hyperparameters (μ_s, μ_λ) are sampled on a 20×20 regular logarithmic scale varying from 0.1 to 1000. The case $N_{iter} = 1$ corresponds to the response surface obtained with the unconstrained Tikhonov solution with spatial and spectral regularizations (section II-B1). The two others correspond to the response surface obtained with the non-negative constrained Tikhonov solution of section II-B2. For all cases, the penalty factor ξ is evolving similarly.

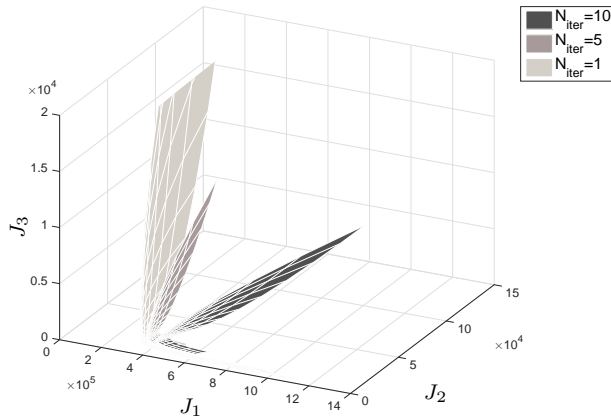


Fig. 1: Estimated response surface for different values of N_{iter}

2) *Convexity of the response surface*: Because $J(\mathbf{x})$ is the sum of three convex objectives and the non-negative orthant is convex, problem (18) remains convex. We follow an approach similar to that of [24] to prove the convexity of the response surface of the multi-objective minimization problem (18).

Theorem 1. *If $J(\mathbf{x})$ is convex, then the response surface of problem (18) is convex.*

Proof: First, the tri-objective optimization problem (18) can be written in the following equivalent form:

$$\min_{\mathbf{x} \geq 0} J_1(\mathbf{x}) \quad \text{subject to} \quad \begin{cases} J_2(\mathbf{x}) \leq \tau_s, \\ J_3(\mathbf{x}) \leq \tau_\lambda. \end{cases} \quad (20)$$

Let $\mathbf{x}_{\tau_s, \tau_\lambda}$ be the optimal solution of (20) and $\Pi(\tau_s, \tau_\lambda)$ be the response surface, then for each $\tau_s \geq 0, \tau_\lambda \geq 0$:

$$\Pi(\tau_s, \tau_\lambda) = J_1(\mathbf{x}_{\tau_s, \tau_\lambda}). \quad (21)$$

The equivalence of problems (18) and (20) implies that there exists a unique hyperparameter (μ_s, μ_λ) yielding the same solution as (τ_s, τ_λ) . In other words, there is a one-to-one correspondence between (μ_s, μ_λ) and (τ_s, τ_λ) .

Equation (21) can be restated as:

$$\Pi(\tau_s, \tau_\lambda) = \inf_{\mathbf{x} \geq 0} f(\mathbf{x}, \tau_s, \tau_\lambda) \quad (22)$$

where

$$f(\mathbf{x}, \tau_s, \tau_\lambda) = J_1(\mathbf{x}) + \varphi_{\tau_s}(\mathbf{x}) + \varphi_{\tau_\lambda}(\mathbf{x}) \quad (23)$$

$$\varphi_{\tau_s}(\mathbf{x}) = \begin{cases} 0 & \text{if } J_2(\mathbf{x}) \leq \tau_s \\ \infty & \text{otherwise} \end{cases} \quad (24)$$

$$\varphi_{\tau_\lambda}(\mathbf{x}) = \begin{cases} 0 & \text{if } J_3(\mathbf{x}) \leq \tau_\lambda \\ \infty & \text{otherwise.} \end{cases} \quad (25)$$

Note that $\varphi_{\tau_s}(\mathbf{x})$ is convex in (\mathbf{x}, τ_s) [45] and the same for $\varphi_{\tau_\lambda}(\mathbf{x})$ which is convex in $(\mathbf{x}, \tau_\lambda)$. Since the objective $J_1(\mathbf{x})$ is convex in \mathbf{x} , f is then convex in $(\mathbf{x}, \boldsymbol{\tau})$, where $\boldsymbol{\tau} = (\tau_s, \tau_\lambda)$. Let $\boldsymbol{\tau}_1$ and $\boldsymbol{\tau}_2$ be non-negative, \mathbf{x}_1 and \mathbf{x}_2 be the corresponding minimizers of (22) and $\alpha \in [0, 1]$. We then have:

$$\begin{aligned} \Pi((1-\alpha)\boldsymbol{\tau}_1 + \alpha\boldsymbol{\tau}_2) &= \inf_{\mathbf{x} \geq 0} f(\mathbf{x}, (1-\alpha)\boldsymbol{\tau}_1 + \alpha\boldsymbol{\tau}_2) \\ &\leq f((1-\alpha)\mathbf{x}_1 + \alpha\mathbf{x}_2, (1-\alpha)\boldsymbol{\tau}_1 + \alpha\boldsymbol{\tau}_2) \\ &\leq (1-\alpha)f(\mathbf{x}_1, \boldsymbol{\tau}_1) + \alpha f(\mathbf{x}_2, \boldsymbol{\tau}_2) \\ &= (1-\alpha)\Pi(\boldsymbol{\tau}_1) + \alpha\Pi(\boldsymbol{\tau}_2). \end{aligned} \quad (26)$$

Hence, the 2D response surface $\Pi(\boldsymbol{\tau})$ is convex. ■

Let us mention that the bi-objective unconstrained case can be written as follows:

$$\min J_1(\mathbf{x}) \quad \text{subject to} \quad J_2(\mathbf{x}) \leq \tau. \quad (27)$$

In this case, [24] proved that Π is a nonincreasing² function of τ . This implies that the Pareto front exactly coincides with the response curve³. This is no longer true for the constrained problem at hand.

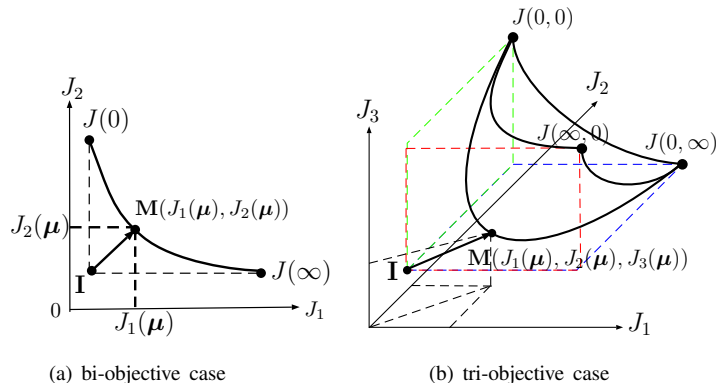


Fig. 2: Representation of the response surface for the unconstrained bi-objective and tri-objective cases: it corresponds to the Pareto front. The ideal point is denoted by \mathbf{I} .

3) *Shape of the response surface with non-negativity constraint:* Following [24], in the unconstrained bi-objective case, the response curve is convex and monotonically decreasing, as represented in Figure 2(a). This can be extended to the response surface corresponding to the unconstrained tri-objective case (Figure 2(b)). It is convex; its intersection with a plane parallel to either (J_1, J_2) or (J_1, J_3) or (J_2, J_3) also defines a monotonically decreasing function. In this case, the Pareto front coincides with the response surface since no point of the response surface is dominated by another one. This behavior is experimentally observed when we use the unconstrained deconvolution (case $N_{iter} = 1$ in Figure 1).

²In the bi-objective case, the regularization parameter is a scalar.

³In the bi-objective case, this is a curve.

On the contrary, when a non-negativity constraint is enforced, the estimated response surface is no longer as in Figure 2(b). A folding of the response surface is observed ($N_{iter} = 5$ and $N_{iter} = 10$ in Figure 1). This results from the constrained data fitting term J_1 which is decreasing and then increasing as μ_s (or μ_λ) increases. In this case, only the set of non-dominated points of the response surface is corresponding to the Pareto front.

When comparing the estimated response surfaces obtained by using $N_{iter} = 5$ and $N_{iter} = 10$, it appears that the case $N_{iter} = 5$ gives an intermediate response between the unconstrained case ($N_{iter} = 1$) and the case $N_{iter} = 10$ for which the convergence of the algorithm is experimentally verified. Actually, this shows how the quadratic penalty approach is progressively modifying the response surface until it converges to the response surface with non-negativity constraint. The folding of the response surface is proved in the following theorem.

Theorem 2. *Let us consider the following constrained optimization problem:*

$$\min_{\mathbf{x} \geq 0} (1 - \alpha)J_1(\mathbf{x}) + \alpha J_2(\mathbf{x}) \quad s.t. \quad J_3(\mathbf{x}) = \tau_\lambda. \quad (28)$$

Then the data fitting J_1 has a unique minimum as α varies from 0 to 1.

Proof: When the non-negativity constraint is enforced, the data fitting can be written as in equation (29)

$$\|\mathbf{y} - \mathbf{H} \max(0, \mathbf{x}_\alpha)\|_2^2 = \sum_{i \in \Omega_\alpha^+} (y_i - [\mathbf{H}\mathbf{x}_\alpha]_i)^2 + \sum_{i \in \Omega_\alpha^-} y_i^2 \quad (29)$$

with $\Omega_\alpha^+ = \{i \mid x_{\alpha,i} > 0\}$ and $\Omega_\alpha^- = \{i \mid x_{\alpha,i} \leq 0\}$. Ω_α^+ corresponds to the set of points where the constraint is not active while Ω_α^- is the active constraint set. We have $\Omega_\alpha = \Omega_\alpha^+ \cup \Omega_\alpha^-$ and $\emptyset = \Omega_\alpha^+ \cap \Omega_\alpha^-$.

Let $n = |\Omega_\alpha|$ and $n^+ = |\Omega_\alpha^+|$, $n^- = |\Omega_\alpha^-|$. By considering $n \rightarrow +\infty$, we can introduce the following probabilities:

$$\begin{aligned} \nu &= P_\alpha(x_i \in \Omega_\alpha^+) = \lim_{n \rightarrow +\infty} \frac{n^+}{n} \\ 1 - \nu &= P_\alpha(x_i \in \Omega_\alpha^-) = \lim_{n \rightarrow +\infty} \frac{n^-}{n}. \end{aligned} \quad (30)$$

By taking the expectation, (29) can be rewritten as:

$$\mathbb{E} [\|\mathbf{y} - \mathbf{H} \max(0, \mathbf{x}_\alpha)\|_2^2] = \nu \mathbb{E} [\|\mathbf{y} - \mathbf{H}\mathbf{x}_\alpha\|_2^2] + (1 - \nu) \mathbb{E} [\|\mathbf{y}\|_2^2]. \quad (31)$$

When $\alpha \rightarrow 1$, \mathbf{x}_α is highly regularized which means that it is very smooth and ν should be close to 1. On the contrary, when $\alpha \rightarrow 0$, \mathbf{x}_α is less regularized and ν is decreasing to a value $\nu_{\min} > 0$. In other words, $\nu \in (0, 1]$. At this point it is necessary to assume that there is a one-to-one correspondence between α and ν .

The term $\mathbb{E} [\|\mathbf{y}\|_2^2]$ is the norm of the data \mathbf{y} which is constant with respect to ν . Thus, $(1 - \nu)\mathbb{E} [\|\mathbf{y}\|_2^2]$ is a linearly decreasing function of ν . The term $\mathbb{E} [\|\mathbf{y} - \mathbf{H}\mathbf{x}_\alpha\|_2^2]$ is the estimation error with no non-negativity constraint. It is an increasing function of ν (or α). More precisely, as $\mathbb{E} [\|\mathbf{y} - \mathbf{H}\mathbf{x}_\alpha\|_2^2]$ is an increasing function of ν , $\nu \mathbb{E} [\|\mathbf{y} - \mathbf{H}\mathbf{x}_\alpha\|_2^2] = O(\nu^a)$ with $a > 1$. Thus, there exists a value of ν for which the data fitting term is minimum. ■

Remark 1. *Instead of considering problem (28) which amounts to looking at a slice of the response surface parallel to (J_1, J_2) , we could have considered*

$$\min_{\mathbf{x} \geq 0} (1 - \alpha)J_1(\mathbf{x}) + \alpha J_3(\mathbf{x}) \quad s.t. \quad J_2(\mathbf{x}) = \tau_s$$

which corresponds to a slice of the response surface parallel to (J_1, J_2) .

Remark 2. Due to the folding in the constrained convex multi-objective case, the Pareto front does not coincide with the response surface: actually, the Pareto front is the set of non-dominated points belonging to the response surface. This property has an important practical consequence: while the ideal point can be determined easily in the unconstrained case by setting the hyperparameters to particular values, this is no longer true in the constrained case. But this folding also has a very positive consequence since it will make the regularization parameter estimation easier (see section V-A)

IV. CHOOSING THE REGULARIZATION PARAMETERS

The response surface gives the set of all solutions of the convex multi-objective problem. The goal of this section is to choose among this set a particular solution which in turn consists in estimating the regularization parameters; here we will pay special attention to the non-negativity constrained multi-objective problem.

In the case of unconstrained bi-objective optimization problems, [18] proposes to find the point with maximum curvature on the L-curve which is a log-log plot of the norm of a regularized solution versus the norm of the corresponding fitting error. Actually, this is nothing but the response surface (which is also the Pareto front, see section III-B3) plot in a log-log scale. However, this change of scale leads to a loss of convexity of the L-curve. In the case of multi-objective optimization problems, [21] extended the notion of L-curve to the L-hypersurface. Also, rather than using the maximum curvature, they proposed a minimum distance criterion to choose the optimal hyperparameters.

To preserve the convexity property, we will work directly on the response surface without resorting to a logarithmic scale. We propose the maximum curvature criterion (MCC) and the minimum distance criterion (MDC). An efficient algorithm to evaluate the MDC solution is also proposed.

A. Maximum curvature criterion

Since we have the regularization parameter $\boldsymbol{\mu} = (\mu_s, \mu_\lambda)$ for the tri-objective problem (18), the response surface $\Pi(\boldsymbol{\mu})$ is a two-dimensional manifold (surface) in \mathbb{R}_+^3 . To calculate the curvature, we need to estimate the first and second derivatives of each objective $J_1 = \|\mathbf{y} - \mathbf{H}\mathbf{x}\|_2^2$, $J_2 = \|\mathbf{D}_s\mathbf{x}\|_2^2$ and $J_3 = \|\mathbf{D}_\lambda\mathbf{x}\|_2^2$ with respect to μ_s and μ_λ . If f is a function of both variables $\{\mu_s, \mu_\lambda\}$, we can, for example, estimate the first partial derivative of f with respect to μ_s by:

$$f'_{\mu_s}(i, j) \approx \frac{f(\mu_{s,i}, \mu_{\lambda,j}) - f(\mu_{s,i-1}, \mu_{\lambda,j})}{\mu_{s,i} - \mu_{s,i-1}} \quad (32)$$

where $(\mu_{s,i}, \mu_{\lambda,j})$ is a discrete grid over which f is computed. Similarly, the second partial derivative of f is estimated by:

$$f''_{\mu_s}(i, j) \approx \frac{f(\mu_{s,i+1}, \mu_{\lambda,j}) + f(\mu_{s,i-1}, \mu_{\lambda,j}) - 2f(\mu_{s,i}, \mu_{\lambda,j})}{\mu_{s,i+1} - \mu_{s,i-1}}. \quad (33)$$

Now we define the curvature as follows.

Definition 1 (Curvature). Let $x = J_1(\boldsymbol{\mu})$, $y = J_2(\boldsymbol{\mu})$ and $z = J_3(\boldsymbol{\mu})$, κ_{μ_s} and κ_{μ_λ} are the curvatures along μ_s

and μ_λ respectively.

$$\begin{aligned}\kappa_{\mu_s} &= \frac{\sqrt{a_{\mu_s}^2 + b_{\mu_s}^2 + c_{\mu_s}^2}}{(x_{\mu_s}'^2 + y_{\mu_s}'^2 + z_{\mu_s}'^2)^{3/2}} \\ \kappa_{\mu_\lambda} &= \frac{\sqrt{a_{\mu_\lambda}^2 + b_{\mu_\lambda}^2 + c_{\mu_\lambda}^2}}{(x_{\mu_\lambda}'^2 + y_{\mu_\lambda}'^2 + z_{\mu_\lambda}'^2)^{3/2}}\end{aligned}\quad (34)$$

where $a = z''y' - y''z'$, $b = x''z' - z''x'$, $c = y''x' - x''y'$. The curvature of the surface is defined by

$$\kappa = \kappa_{\mu_s} \cdot \kappa_{\mu_\lambda}.\quad (35)$$

Then we can formulate the maximum curvature criterion as follows:

Definition 2 (Maximum Curvature Criterion).

$$\boldsymbol{\mu}^* = \arg \max_{\boldsymbol{\mu}} \kappa(\boldsymbol{\mu}).\quad (36)$$

Note that the proposed definition of the curvature is a simplified version which does not correspond to the mean curvature as defined in differential geometry. Indeed, in our case, it is estimated as the average of two curvatures along two predefined directions while, in differential geometry, the mean curvature corresponds to the average of the principal curvatures which require the estimation of the curvatures along all possible directions.

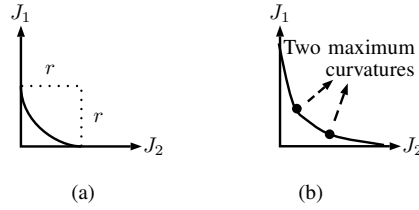


Fig. 3: Examples of response surfaces with non-unique maximum curvature

The maximum curvature criterion suffers from two main shortcomings. One of them is related to the discrete derivative evaluation which is highly sensitive to noise. This noise comes from the use of an iterative solver which provides only approximate solutions thus yielding a noisy response surface. Another important issue is the non-uniqueness of the maximum curvature criterion. To illustrate this point, let us consider the two following bi-objective examples. The first one corresponds to a response curve having a convex quarter circle shape, as shown in Figure 3(a): the curvature is a constant and, thus, not unique. The second case corresponds to a convex response surface front whose curvature has two maxima highlighted by the two dots in Figure 3(b). Despite the fact that the non-negativity constraint increases the curvature of the response surface and makes the MCC more efficient, it cannot fully overcome these shortcomings. Instead, we propose in the next section the MDC.

B. Minimum distance criterion

As mentioned before, the MDC is applied directly to the response surface whose convexity is central in establishing the properties of the criterion. The ideal point as defined in (11) corresponds to the minimum of all objective

functions. Even if it is a non-reachable solution, it can be considered as a reference point and the optimal point of the response surface will be the one having the minimum distance to this ideal point. However, this choice of the reference point is somewhat arbitrary; this will be discussed in remark 4 and in the experimental section V-A. Let us introduce the MDC by first defining the distance to the ideal point.

Definition 3 (Distance to Ideal Point). *Let $\mathbf{I} = (I_1, \dots, I_z)$ denotes the coordinates of the ideal point. The function $D(\boldsymbol{\mu})$ is the squared distance from the ideal point \mathbf{I} to the point $\mathbf{M}(\boldsymbol{\mu}) = (J_1(\boldsymbol{\mu}), \dots, J_z(\boldsymbol{\mu}))$ on the response surface.*

$$D(\boldsymbol{\mu}) = \sum_{i=1}^z (J_i(\boldsymbol{\mu}) - I_i)^2. \quad (37)$$

The MDC is defined as follows.

Definition 4 (Minimum Distance Criterion).

$$\boldsymbol{\mu}^* = \arg \min_{\boldsymbol{\mu}} D(\boldsymbol{\mu}). \quad (38)$$

The key property of the MDC is that it admits a unique minimum. Its proof relies on a geometrical interpretation of the MDC.

Theorem 3. *If the response surface is convex, the MDC admits a unique minimum.*

Proof: To find the minimum of $D(\boldsymbol{\mu})$, we have to find $\boldsymbol{\mu}$ such that:

$$\begin{aligned} \frac{\partial D(\boldsymbol{\mu})}{\partial \boldsymbol{\mu}} &= \frac{\partial ((J_1(\boldsymbol{\mu}) - I_1)^2 + \dots + (J_z(\boldsymbol{\mu}) - I_z)^2)}{\partial \boldsymbol{\mu}} \\ &= 2 \left(\frac{\partial J_1(\boldsymbol{\mu})}{\partial \boldsymbol{\mu}} (J_1(\boldsymbol{\mu}) - I_1) + \dots \right. \\ &\quad \left. + \frac{\partial J_z(\boldsymbol{\mu})}{\partial \boldsymbol{\mu}} (J_z(\boldsymbol{\mu}) - I_z) \right) \\ &= 2 \begin{bmatrix} \frac{\partial J_1}{\partial \boldsymbol{\mu}}, \dots, \frac{\partial J_z}{\partial \boldsymbol{\mu}} \end{bmatrix} \begin{bmatrix} J_1(\boldsymbol{\mu}) - I_1 \\ \vdots \\ J_z(\boldsymbol{\mu}) - I_z \end{bmatrix} \\ &= 0. \end{aligned} \quad (39)$$

In (39), $\mathbf{T} = \left[\frac{\partial J_1}{\partial \boldsymbol{\mu}}, \dots, \frac{\partial J_z}{\partial \boldsymbol{\mu}} \right]^T$ is the matrix whose columns span the tangent plane to the response surface at the point \mathbf{M} and $\vec{\mathbf{IM}} = [J_1(\boldsymbol{\mu}) - I_1, \dots, J_z(\boldsymbol{\mu}) - I_z]^T$. As (39) equals zero, we have :

$$\vec{\mathbf{IM}} \perp \mathbf{T} \quad (40)$$

which means that each column of \mathbf{T} is orthogonal to $\vec{\mathbf{IM}}$. Any point satisfying the orthogonality condition is thus a critical point of $D(\boldsymbol{\mu})$. As the response surface is convex, the critical point is necessarily the unique minimum of $D(\boldsymbol{\mu})$ [45]. ■

This theorem is illustrated in Figure 4 showing the tangent plane orthogonal to $\vec{\mathbf{IM}}$.

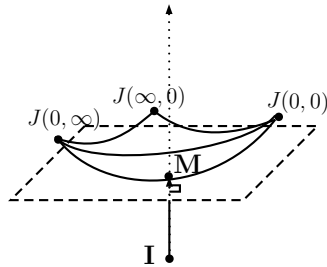


Fig. 4: Minimum distance criterion : the solution corresponds to $\mathbf{T} \perp \vec{\mathbf{IM}}$.

Remark 3. *The evaluation of the MDC requires the determination of the ideal point which is difficult to obtain when the non-negativity constraint is enforced. This is due to the folding discussed in section III-B3. We propose to define it by determining the points of the response surface corresponding to the unconstrained Tikhonov solution for three values of μ equal to $(0, 0)$, $(0, \infty)$, $(\infty, 0)$ ⁴. The ideal point coordinates are then obtained by finding the minimum coordinate of each of the three points (see Figure 2(b)). It is important to notice that determining the ideal point is not time consuming since the 3 points of the unconstrained response surface are computed using the fast frequency domain implementation of the unconstrained Tikhonov estimator.*

Remark 4. *The convexity of the MDC is stated for a reference point chosen as the ideal point. The MDC remains convex for any other choice of the reference point, but the optimal point depends on this choice. More precisely, if the curvature of the response surface is low in the vicinity of the optimal point, then the estimated point will vary a lot with the chosen reference point while it does not if the curvature is large. The folding of the response surface resulting from the non-negativity constraint yields an increase of the response surface curvature, thus stabilizing the estimated point using MDC.*

C. The case of unimodal MDC

Looking carefully at the shape of the estimated response surface in Figure 1 reveals that it may be slightly non-convex. In fact, the loss of convexity was experimentally observed when the deconvolution is possibly affected by numerical errors. We made a large number of experiments to identify the possible causes of this loss of convexity and we found that this phenomenon (sometimes but not always) arises in situations where the bandwidth of the signal to recover was much greater than the bandwidth of the convolution kernel. For the deconvolution problem at hand, it also depends on the conditioning of the convolution matrix. This phenomenon occurs mainly for small values of μ_s and μ_λ . A detailed analysis of this phenomenon is very complicated and is out of the scope of this paper. In all the considered cases, the loss of convexity resulted in a unimodal MDC. This motivates the study of the uniqueness of unimodal MDC. For simplicity reasons, only the bi-objective case is considered.

The minimum distance criterion consists in finding the vector $\vec{\mathbf{IM}}$ orthogonal to the tangent vector \mathbf{T} at the point \mathbf{M} . Let α (resp. β) be the angle of $\vec{\mathbf{IM}}$ (resp. \mathbf{T}) with the horizontal axis, as shown in Figure 5(a). If the response surface Π is convex (as the black curves in Figures 5(a) and 5(b)), when $\vec{\mathbf{IM}}$ is moving from the vertical direction

⁴In practice, the hyperparameter values cannot be set to ∞ but are fixed to large values

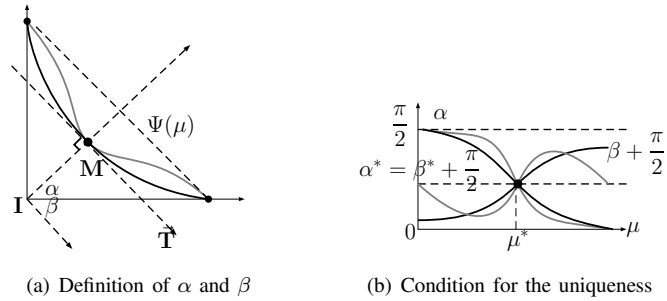


Fig. 5: Uniqueness of the MDC.

to the horizontal direction, the angle α is monotonically decreasing from $\frac{\pi}{2}$ to 0 and the angle $\beta \in [-\frac{\pi}{2}, 0]$ is monotonically increasing, as shown in Figure 5(b). There is only one point where $\alpha^* = \beta^* + \frac{\pi}{2}$. So there is only one point satisfying the orthogonality condition.

Let us examine what happens when the response surface Π is no longer convex, as highlighted by the gray curves in Figure 5(a) and 5(b). We assume that $D(\mu)$ is unimodal and admits the same minimum as in the convex case. This means that $D(\mu)$ is decreasing $\forall \mu < \mu^*$ and is increasing $\forall \mu > \mu^*$. There exists a line $\Psi(\mu)$ whose slope equals that of the tangent vector at \mathbf{M} , such that $\Pi(\mu) \leq \Psi(\mu)$. The slope of $\Psi(\mu)$ equals β^* . If $D(\mu)$ is unimodal, $\alpha(\mu)$ is monotonically decreasing from $\frac{\pi}{2}$ to 0 and we have the following inequalities:

$$\begin{cases} \beta(\mu) - \beta^* < \alpha(\mu) - \alpha^* & \text{if } \mu < \mu^* \\ \beta(\mu) - \beta^* > \alpha(\mu) - \alpha^* & \text{if } \mu > \mu^*. \end{cases} \quad (41)$$

So for the unimodal case, the unique intersection of α and $\beta(\mu) + \frac{\pi}{2}$ is in μ^* .

This proves that if the distance criterion is unimodal, it always admits a unique minimum even if the response surface is no longer convex. In that respect, the MDC can be said to be robust to the loss of convexity observed on the empirical response surface. We now turn our attention to the proposal of a fast method to estimate the MDC.

D. A grid-search strategy for MDC

To begin, let us give one remark about the computational cost of the response surface estimation on a 2D grid. It can be very high if the grid size is high. To give some figures, the evaluation of the response surface of the non-negative Tikhonov solution on a 20×20 grid (black surface in Figure 1) takes more than 2 days with an Intel Core i7 2.2 Ghz processor. However, by properly exploiting the property of the response surface it is possible to design a fast approach aiming at finding a particular point on the response surface. We propose to use a grid-search method which is proved to be convergent for unimodal criteria [46].

Figure 6 illustrates the grid refinement. For the bi-objective case with one single regularization parameter, at the first level $r = 1$, we have only four points $\mu_i^{(1)} (i = 1, \dots, 4)$ on which the response surface is estimated. Then the grid is refined by defining a new search segment on which four new points $\mu_i^{(2)} (i = 1, \dots, 4)$ are defined. The procedure is repeated until a maximum number of levels is reached. In the tri-objective case with two regularization

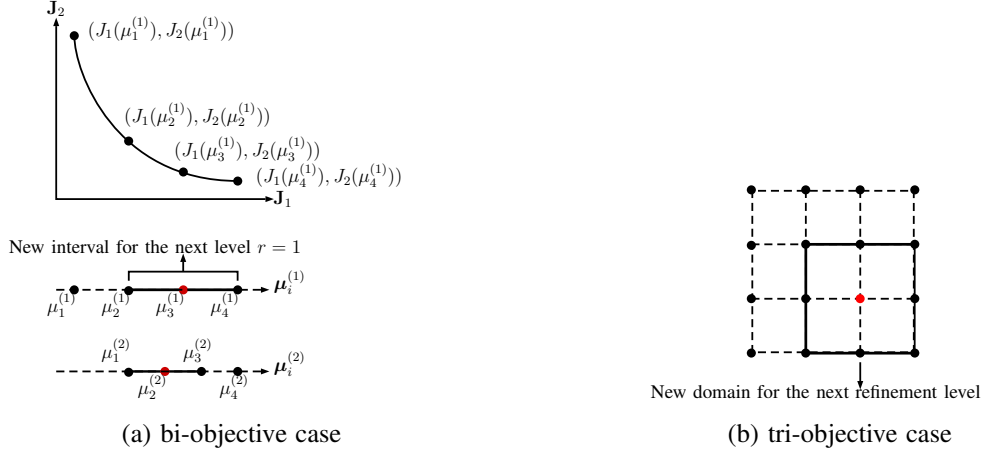


Fig. 6: Grid refinement method

parameters, we define 4×4 points for the r -th level and choose an optimum point among them. Note that the response surface should be evaluated on the four central points of the grid. Then we select the points around it as the new domain. This new domain is refined to find a new optimum point. The procedure is repeated iteratively. The whole procedure is summarized in Algorithm 1.

Algorithm 1: Grid minimization of MDC

Data: Parameters of the initialized level $\boldsymbol{\mu} = (\mu_{1,1}^{(1)}, \dots, \mu_{4,4}^{(1)})$; number of refinement levels R ; image \mathbf{y}_l , matrix \mathcal{H}_l , matrix \mathcal{D}_l for $l = 1, \dots, L$; Δ_λ

Find the ideal point \mathbf{I} ;

for $r = 1 : R$ **do**

$\Pi =$ **Procedure 2** ($\mathbf{y}_l, \mathcal{H}_l, \mathcal{D}_l$ for $l = 1, \dots, L$; $\Delta_\lambda; \boldsymbol{\mu}$);

Calculate the distance \mathbf{D}_r for each value of Π ;

Find the minimum distance d_{min} in \mathbf{D}_r and the corresponding indexes m^*, n^* and $\boldsymbol{\mu}^*$;

New domain for the grid refinement

$$\mu_{1,1}^{(r+1)} = \mu_{m^*-1, n^*-1}^{(r)};$$

$$\mu_{4,1}^{(r+1)} = \mu_{m^*+1, n^*-1}^{(r)};$$

$$\mu_{1,4}^{(r+1)} = \mu_{m^*-1, n^*+1}^{(r)};$$

$$\mu_{4,4}^{(r+1)} = \mu_{m^*+1, n^*+1}^{(r)};$$

Calculate the grid points into the new domain $\boldsymbol{\mu} = (\mu_{1,1}^{(r+1)}, \dots, \mu_{4,4}^{(r+1)})$;

end

Result: $\boldsymbol{\mu}^*$

In what follows, the number of levels is fixed to 6 which gives approximately the same resolution as the 20×20

Procedure 2: Evaluation of the points on the response surface

Data: image y_l , matrix \mathcal{H}_l , matrix \mathcal{D}_l for $l = 1, \dots, L$; matrix Δ_λ ; $\boldsymbol{\mu}$
for $n = 2 : 3$ **do**
for $m = 2 : 3$ **do**
 $\mathbf{x} = \text{Deconvolution}(y_l, \mathcal{H}_l, \mathcal{D}_l \text{ for } l = 1, \dots, L; \Delta_\lambda; (\mu_s, \mu_\lambda) \text{ corresponds to } \mu_{m,n}^{(r)});$

 Calculate the response surface $\Pi(m, n) = (J_1(m, n), J_2(m, n), J_3(m, n));$
end
end
Result: Π

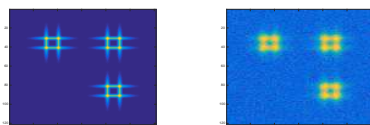
grid of section III-B1 but requires only 24 evaluations of the response surface instead of 400. Note that such a grid strategy has also been used to maximize the MCC with a reasonable computation time. However, as the MCC is not necessarily unimodal, the procedure cannot be guaranteed to converge.

V. EXAMPLES AND EXPERIMENTS

In this section, some numerical and experimental results will be presented to illustrate the effectiveness of proposed MCC and MDC for estimating the regularization parameters. First, a simple bi-objective image deconvolution problem is used to assess the performances of the proposed approaches and to compare them to state-of-the-art regularization parameter estimation methods. Then we address the tri-objective hyperspectral images deconvolution problem. We begin by giving an illustrative example and then we compare the performances of the two proposed approaches (MCC and MDC). The performance assessment is conducted by evaluating the mean square error (MSE) as a function of the signal-to-noise ratio (SNR): the lower the MSE, the better the performances.

A. Performances of MCC and MDC for 2D image deconvolution

In this section, we consider a bi-objective 2D image deconvolution problem. We use the simulated image as shown in Figure 7 which corresponds to a single slice from the simulated hyperspectral cube in section V-B.



(a) Original image (b) Blurred and noisy image (SNR= 10 dB)

Fig. 7: An example of simulated image

The first experiment aims at evaluating the performances of the proposed approaches using the unconstrained Tikhonov. They are compared to two state-of-the-art methods: the L-curve approach [18] and the generalized cross-

validation (GCV) method [15]. The MSEs as a function of the SNR obtained for different methods are shown on Figure 8(a). The SNR is defined as follows: $\text{SNR} = 10 \log_{10} \|\mathbf{H}\mathbf{x}\|_2^2 / \|\mathbf{e}\|_2^2$. The second experiment aims at evaluating the performances of the criteria when using the non-negative Tikhonov approach. Note that in that case, GCV cannot be used since this algorithm cannot account for a non-negativity constraint. Thus, only the MSEs of MCC, MDC and L-curve are evaluated. The results are shown in Figure 8(b). To overcome the multiple maximum curvature problem which may occur in MCC and L-curve criteria, we follow the recommendation in [18]. Starting with a low value of the regularization parameter, the first maximum curvature is chosen to estimate the regularization parameter.

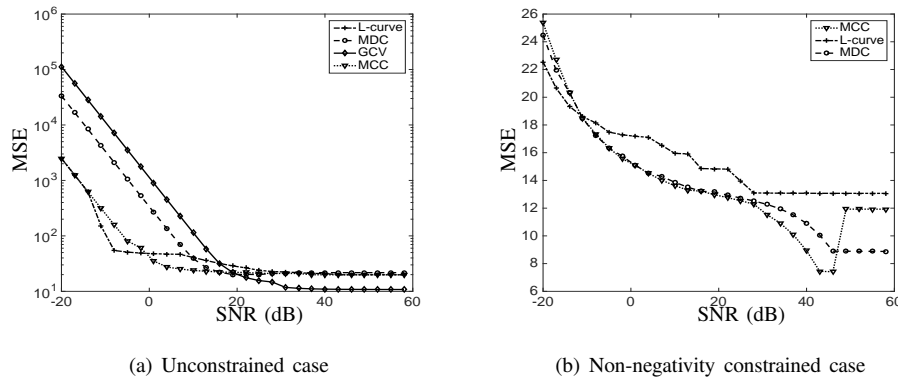


Fig. 8: Performances of the 2D deconvolution with optimal parameter μ_s selected by different approaches

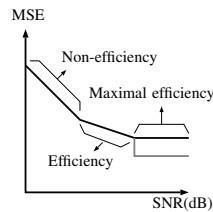


Fig. 9: Typical shape of the MSE

The MSE curve includes three main parts which are sketched on Figure 9. The non-efficiency zone corresponds to the part of the curve for which the MSE increases as fast as the noise level. The efficiency zone corresponds to the part of the curve for which the MSE increases at a lower rate than the noise: this is the zone where deconvolution is effective. Finally, the third horizontal part corresponds to the best performance of the regularized deconvolution method. The minimum value of the MSE in this third part is also depending on both \mathbf{H} and \mathbf{x} . When the bandwidth of the filter \mathbf{H} is lower than the bandwidth of \mathbf{x} , even in noise-free situations, deconvolution cannot restore the signal \mathbf{x} outside the frequency range (bandwidth) covered by the filter. In fact, this minimum MSE reflects the ill-conditioning of the matrix \mathbf{H} . It decreases as the condition number decreases. For example, the gray curve in Figure 9 corresponds to a case where the conditioning is better than that of the black curve.

In the unconstrained case, no approach performs uniformly better than the others. GCV reaches the lowest minimum value of the MSE which is about 10^1 . This shows that GCV works better than the other approaches for high SNR values. Both L-curve and MCC perform better than the other approaches when the SNR is low. For SNR ranging between $[0, 20]$ dB, the best performances are achieved by the MCC. Note that for SNR smaller than -14 dB (three first points), the maximum curvature of the L-curve is negative which is somewhat incoherent with the L-curve approach since the response curve no longer has the L-corner. The performances of the MDC (Figure 8(a)) are not very satisfying. As mentioned in remark 4, a strong folding of the response curve will decrease the sensitivity of the MDC to the choice of the reference point. In the unconstrained case, there is no folding of the response curve which results in the high sensitivity of the MDC; this explains the poor performances of the MDC. On the contrary, the MCC appears to be less sensitive to the folding of the response curve.

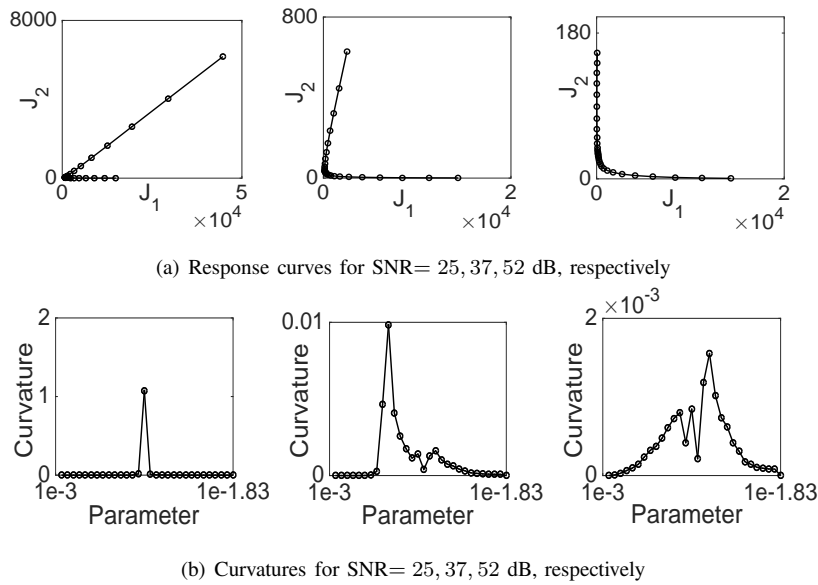


Fig. 10: Response curves and curvatures for different SNRs

When the non-negativity constraint is enforced, the L-curve does not yield satisfying results anymore. This can be attributed to the complex shape of the L-curve (which is reinforced by the non-negativity constraint) associated to the curvature maxima. At high SNR, the MSE reaches a minimum value similar to the unconstrained case. For SNR smaller than 28 dB, MCC and MDC behave similarly. MCC has the best performance for SNR in $[28, 46]$ dB. Indeed, the folding of the response curve decreases as the SNR increases since the constraint is active on fewer image points, see Figure 10(a). This explains why the MCC which is less sensitive to the folding, performs better than the MDC. However, the MCC may suffer from the non-uniqueness of the maximum curvature, see Figure 10(b). For example, after 46 dB, the first local maximum of the curvature no longer corresponds to the correct optimal point. This explains the step observed on the MSE of the MCC in Figure 8(b). Finally, it is worth mentioning that, in the non-negativity constrained case, both MCC and MDC reach a horizontal asymptote lower than the 10^1 reached by GCV which is the best performing method at high SNR values in the unconstrained case. This illustrates

the regularizing effect of the non-negativity constraint.

B. An illustrative example of the non-negativity constrained hyperspectral image deconvolution

To simulate the blurred hyperspectral images, we first generate the unblurred image according to the instantaneous mixture model:

$$\mathbf{X} = \sum_k \mathbf{A}_k \circ \mathbf{s}_k. \quad (42)$$

Here, \mathbf{A}_k represents the k -th abundance (spatial source) which is a function of the spatial variables, \mathbf{s}_k represents the k -th endmember (spectral source) and \circ is the outer (tensor) product. In the example of Figure 11, an instantaneous mixture of 3 sources is considered. The abundance maps of size (120×120) are shown on the upper row while the endmembers, which include 32 spectral bands, are on the lower row. These endmembers correspond to NIR spectra of wood (raw, varnished and painted) samples. They were chosen because of their relative smoothness making the spectral smoothness penalty effective. However, in the supplementary material [26], we added other examples corresponding to different types of endmembers and abundance profiles. Eight slices of the resulting unblurred hyperspectral image are shown in Figure 12(a).

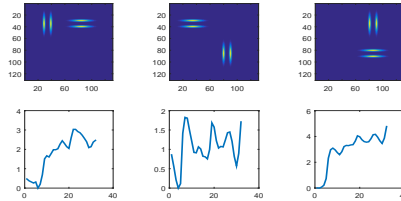
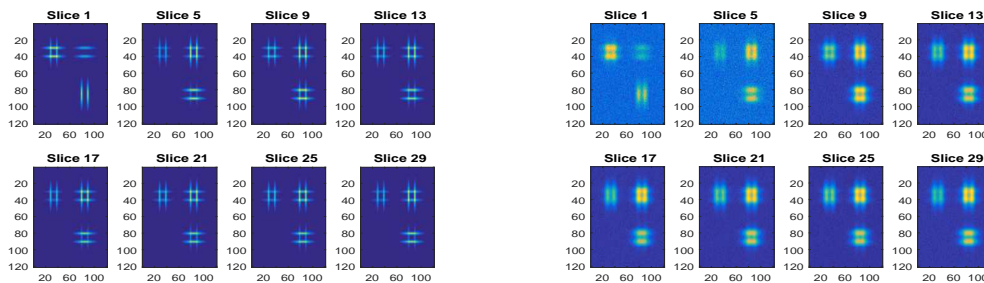


Fig. 11: Abundance maps and endmembers used to simulate the unblurred hyperspectral image.



(a) Unblurred hyperspectral image

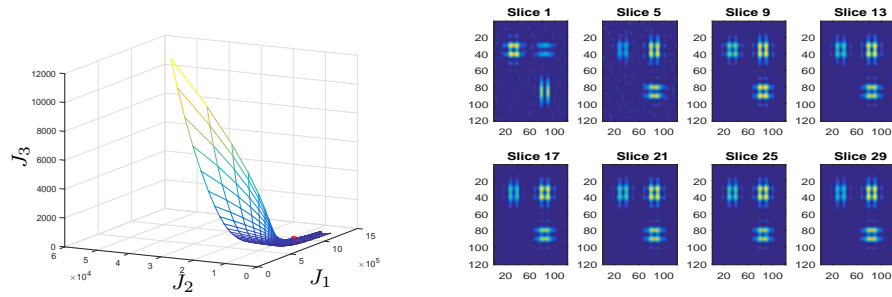
(b) Blurred noisy hyperspectral image \mathbf{y} (SNR= 20 dB)

Fig. 12: Simulation of the hyperspectral image

The convolution filter \mathcal{H}_l is assumed to be a low-pass Gaussian filter of size (11×11) and its full width at half maximum is set to 5 points in both dimensions. It is invariant with respect to l . The blurring is implemented in

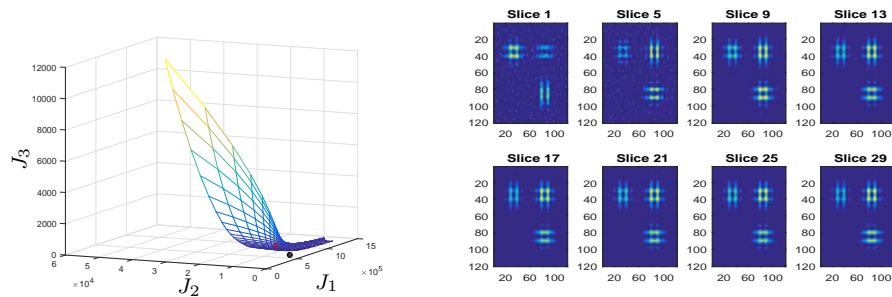
the Fourier domain (circular convolution). Note that the filter invariance assumption adopted here for simplicity, is reasonable for applications such as Raman hyperspectral imaging systems, fluorescence confocal microscopy and industrial NIR spectro-imaging system. However, in some applications such as NIR microscopy, the variation of the filter with respect to l has to be taken into account (see [35] for details).

A Gaussian noise is then added to the blurred image yielding the hyperspectral image of Figure 12(b). The noise level is the same for all bands. The simulated blurred hyperspectral image results in a difficult problem since the bandwidth of the unblurred image \mathbf{x} is much larger than that of the filter \mathbf{H} . We also have to mention that the simulated hyperspectral image was chosen to favor non-negative deconvolution. This is because the simulated unblurred image includes a large amount of zero values.



(a) Response surface with the point of maximum curvature (red point) (b) Deconvolution with parameters found by the MCC ($\mu_s = 88.5867$, $\mu_\lambda = 0.1624$)

Fig. 13: Result of the non-negative deconvolution problem by using MCC



(a) Response surface with the point at minimum distance (red one) from the ideal point 2.9764, $\mu_\lambda = 33.5982$ (black one) (b) Deconvolution with parameters found by the MDC ($\mu_s =$

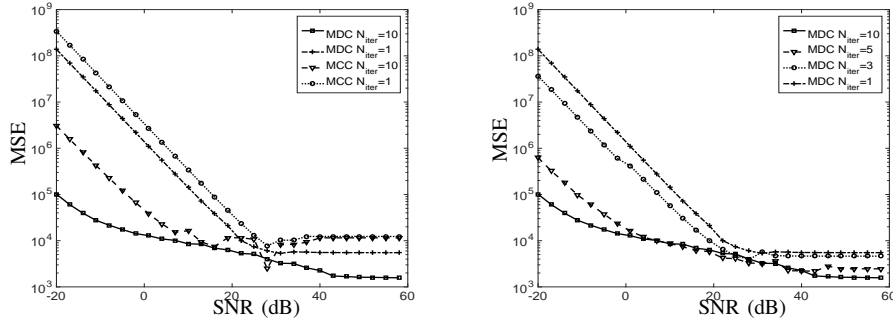
Fig. 14: Result of the non-negative deconvolution problem by using MDC

We applied MCC and MDC to the non-negativity constrained deconvolution problem of the hyperspectral image shown in Figure 11. The response surface and the point corresponding to the MCC are shown in Figure 13(a).

The result of the deconvolution with the parameters ($\mu_s = 88.5867$ and $\mu_\lambda = 0.1624$) found by this method is shown in Figure 13(b). The response surface and the point corresponding to MDC are shown in Figure 14(a). The corresponding parameters are $\mu_s = 2.9764$, $\mu_\lambda = 33.5982$. The corresponding restored images are shown in Figure 14(b). We can observe that for this deconvolution problem MDC works better than MCC. The poor performance of the MCC results from the multiple maximum curvatures. In fact, the point having the maximum curvature does not yield the best result.

The grid-search strategy with a number of levels fixed to 6 was applied to MDC. The estimated point ($\mu_s = 2.1274$, $\mu_\lambda = 31.3741$) is close to ($\mu_s = 2.9764$, $\mu_\lambda = 33.5982$), the point found on the whole response surface. The two points do not coincide exactly because the grids do not. The application of the grid search to MCC is highly sensitive to the choice of the initial grid. This is still a consequence of the already mentioned curvature multiple maxima. In the particular example considered here, using the same initial grid as for MCC yields a point ($\mu_s = 119.7262$, $\mu_\lambda = 7.2457$) which is completely different from ($\mu_s = 88.5867$, $\mu_\lambda = 0.1624$), the point found before on the whole response surface; the corresponding restored image (not shown) is not satisfying as well.

C. Performances of MCC and MDC for non-negative hyperspectral image deconvolution



(a) Performances of the hyperspectral image deconvolution with optimal parameters (μ_s, μ_λ) selected by image deconvolution with optimal parameters (μ_s, μ_λ) MCC and MDC

(b) Performances of the non-negative hyperspectral image deconvolution selected by MDC

Fig. 15: Performances of the hyperspectral image deconvolution with optimal parameters (μ_s, μ_λ) selected by MCC and MDC

In the case of non-negative hyperspectral image deconvolution, as far as we know, no other approach than MCC and MDC can be used. The MSE of MCC and MDC corresponding to the unconstrained ($N_{iter} = 1$) and constrained ($N_{iter} = 10$) are shown in Figure 15(a). For MCC, the curves obtained with $N_{iter} = 10$ and $N_{iter} = 1$ tend to the same horizontal asymptote which is about 10^4 while it is about 10^3 for MDC. There is almost a factor 10 between the MSEs obtained with MCC and MDC. In fact, the poor behavior of the MCC associated to the grid search is only reflecting the already mentioned multiple maximum curvature problem.

Let us now examine the behavior of MDC. When $N_{iter} = 1$ (unconstrained Tikhonov approach), the efficiency zone is in the interval $[20, 30]$ dB and MSE reaches its minimum when the SNR is greater than 30 dB. This

is because MDC does not give good results in the unconstrained case. However, when $N_{iter} = 10$ (constrained Tikhonov approach), not only the minimum MSE is decreased but the efficiency zone (which is between -10 dB and 40 dB) increases significantly. This effect shows that MDC works better in the non-negativity constrained case and the non-negativity constraint improves the effectiveness of the deconvolution. This highlights the stabilizing property of the non-negativity constraint as proved in [47]. Once again, recall that the considered example favors non-negative deconvolution. This is because it does include large regions where the original hyperspectral image is null (or close to 0).

The Tikhonov approach with a non-negativity constraint is an iterative algorithm which converges to the optimal solution as the number of iterations increases. Figure 15(b) shows how the MSE obtained by MDC for different values of $N_{iter} = 1, 3, 5, 10$ is gradually changing from the unconstrained to the constrained case. In fact, the study of the non-negative deconvolution performance as a function of N_{iter} aims at evaluating how the convergence of the algorithm is affecting the performances of the MDC. Increasing the number of iteration allows to gradually increase the efficiency zone until the algorithm convergence is reached ($N_{iter} = 10$ for this example).

Extensive simulations investigating the behavior of MDC and MCC for different types of hyperspectral images can be found in [26]. The analysis of the results shows that MDC always performs better than MCC. Also, the corresponding MSEs are more stable (smooth) than those of MCC. This is due to the multiple maximum problem of MCC which renders the MSE behavior a bit erratic. The non-negativity constraint really matters when the image includes many zeros. Increasing the number of points on which the positivity constraint is active, will also increase the folding of the response surface resulting in an accurate regularization parameter estimation. When the number of zeros is low, the non-negativity constraint is no longer relevant and both MDC and MCC are not very efficient. It may even happen that, for high SNR, the unconstrained deconvolution and associated MDC and MCC yields better solutions. See example 5 in [26] (which is a kind of worst-case scenario) for $\text{SNR} > 30$ dB. Finally, the estimated regularization parameters with MDC (associated to non-negative deconvolution) is linked to the nature of the image to recover. Spatially (resp. spectrally) smooth images yield large values of μ_s (resp. μ_λ). Conversely, spatially (resp. spectrally) peaky images yield low values of μ_s (resp. μ_λ). It corresponds to what intuition suggests. This is another evidence of the interest of MDC.

D. Application to hyperspectral fluorescence microscopy

A real-world example is included. It corresponds to an image of bacterial biosensors using hyperspectral fluorescence microscopy. A bacterial biosensor is a genetically modified bacteria which reacts to a stressing element (here iron, Fe) by producing a fluorescent protein (GFP). The hyperspectral fluorescence images will give indications of the Fe spatial concentration. This hyperspectral image size is $(512 \times 512 \times 16)$ and the pixel size is $0.117 \mu\text{m}$ along each dimension. The 16 wavelengths are ranging from 455nm to 605nm . It was obtained by Carl Zeiss Bio-Rad confocal microscope. The PSF of the microscope is evaluated according to [48] as a function of the imaging parameters (excitation wavelength, emission wavelength, numerical aperture and pixel size). This results in a 7×7 Gaussian approximation of the PSF.

Figure 16 is a part selected from the whole image. It shows raw data (upper row), restored data with the regularization parameters estimated by MCC (middle row) and restored data with the regularization parameters

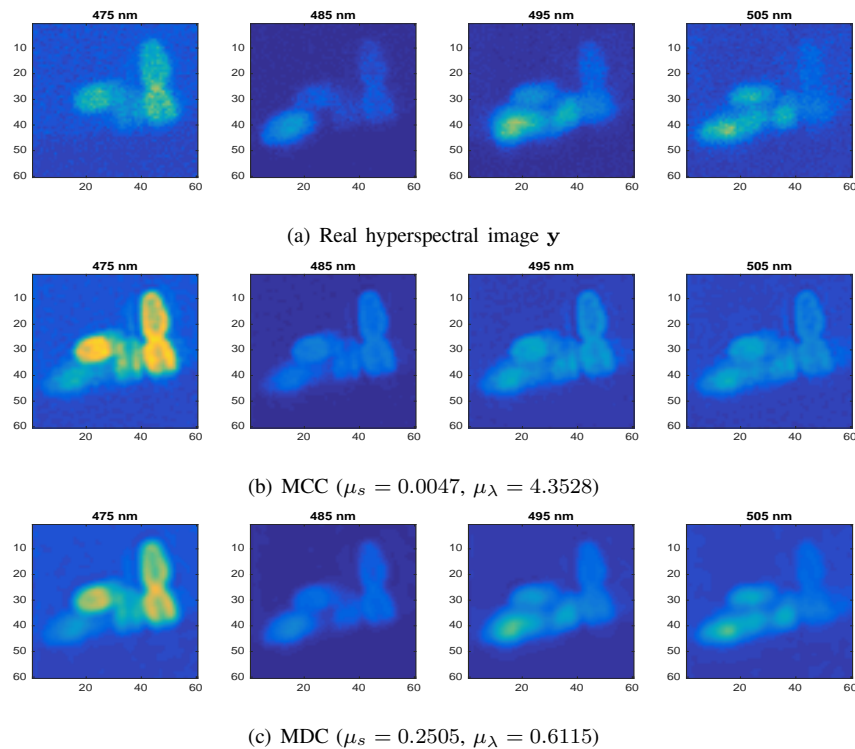


Fig. 16: Result of the non-negative deconvolution problem by using MCC and MDC

estimated by MDC (lower row). It should be noted that this image includes both peaky and smooth parts along the spectral dimension; this makes the choice of a global spectral regularization parameter not obvious. A large regularization parameter will over-smooth the peaky part while a low regularization parameter will under-regularize the smooth part.

Both results show an improved resolution. However, a closer look at the results of MCC reveals that the spectral regularization parameter is over-estimated. This results in a spectral over-smoothing which makes some high intensity patterns of bacteria remaining on adjacent spectral bands (see for example 16(b)). This is less visible for MDC.

In fluorescence microscopy, the noise is typically modeled by a Poisson distribution due to photon counting in optical devices. Following [49], it includes two main contributions: the shot noise and the dark noise. This noise model is also well adapted to other types of hyperspectral images involving photon counting. The Poisson distribution is a non-negative support probability density function and, in that respect, it is well suited to the non-negative nature of hyperspectral images. When the SNR is high enough (large integration time), the Poisson noise can be well approximated by an additive Gaussian noise whose variance depends on the signal amplitude. For low SNR, the approximation is no longer valid. However, the Gaussian assumption is adopted in a large majority of works dealing with hyperspectral images. The application of the proposed methods to real hyperspectral data illustrates their relative insensitivity to the noise model.

VI. CONCLUSION

In this work, the estimation of the regularization parameters of non-negativity constrained hyperspectral image deconvolution algorithms is stated as a multi-objective optimization problem whose response surface is proved to be convex. A first contribution of this work is to show that the non-negativity constraint results in a folding of the response surface. A consequence is that, unlike the unconstrained case, the response surface does not coincide with the Pareto front. But this folding results in an increase of the curvature of the response surface which is making the regularization parameter estimation easier.

A second contribution of this work is the proposal of the MCC and MDC to estimate the optimal values of the regularization parameters μ_s and μ_λ for the non-negativity constrained tri-objective optimization problem. MCC aims at finding the point of the response surface with maximum curvature while MDC aims at finding the point of the response surface having the minimum distance from the ideal point. We also proved that this criterion admits a unique minimum even if the distance criterion is unimodal (and hence non-convex). A fast grid-search algorithm is proposed to estimate the point of the response surface maximizing MCC or minimizing MDC. Another very positive consequence of the response surface folding is that it decreases the sensitivity of both MDC and MCC. Finally, simulations were used to assess the performances of the proposed MCC and MDC. In addition, an application to a hyperspectral fluorescence microscopy is provided. In fact, MDC results is an efficient method to estimate the regularization parameters of non-negative hyperspectral image deconvolution.

Future works will focus on the extension of the proposed approaches to solve edge-preserving image deconvolution problems. We also intend to develop new approaches aiming at jointly performing the deconvolution and unmixing of hyperspectral images. The application of the MDC to real hyperspectral fluorescence data raises an interesting problem in image restoration when the convolution kernel is poorly known.

REFERENCES

- [1] P. Lasch and D. Naumann, "Spatial resolution in infrared microspectroscopic imaging of tissues," *Biochimica et Biophysica Acta (BBA)-Biomembranes*, vol. 1758, no. 7, pp. 814–829, 2006.
- [2] S. Henrot, C. Soussen, M. Dossot, and D. Brie, "Does deblurring improve geometrical hyperspectral unmixing?" *Image Processing, IEEE Transactions on*, vol. 23, no. 3, pp. 1169–1180, 2014.
- [3] S. Henrot, C. Soussen, S. Moussaoui, and D. Brie, "Edge-preserving nonnegative deconvolution of hyperspectral fluorescence microscopy images," CRAN-IRCCyN, Research Report, Jul. 2015. [Online]. Available: <https://hal.archives-ouvertes.fr/hal-01171524>
- [4] J.-F. Giovannelli and A. Coulais, "Positive deconvolution for superimposed extended source and point sources," *Astronomy & Astrophysics*, vol. 439, no. 1, pp. 401–412, 2005.
- [5] S. Bongard, F. Soulez, É. Thiébaud, and É. Pecontal, "3D deconvolution of hyper-spectral astronomical data," *Monthly Notices of the Royal Astronomical Society*, vol. 418, no. 1, pp. 258–270, 2011.
- [6] X.-L. Zhao, F. Wang, T.-Z. Huang, M. K. Ng, and R. J. Plemmons, "Deblurring and sparse unmixing for hyperspectral images," *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 51, no. 7, pp. 4045–4058, 2013.
- [7] K. Lawrence, B. Park, W. R. Windham, and C. Mao, "Calibration of a pushbroom hyperspectral imaging system for agricultural inspection," *Transactions of the ASAE*, vol. 46, no. 2, p. 513, 2003.
- [8] Pellenc Selective Technology. Mistral Product. [Online]. Available: <http://www.pellencst.com/products/>
- [9] T. Akgun, Y. Altunbasak, and R. M. Mersereau, "Super-resolution reconstruction of hyperspectral images," *Image Processing, IEEE Transactions on*, vol. 14, no. 11, pp. 1860–1875, 2005.

- [10] P. J. La Rivière and P. A. Vargas, “Monotonic penalized-likelihood image reconstruction for x-ray fluorescence computed tomography,” *Medical Imaging, IEEE Transactions on*, vol. 25, no. 9, pp. 1117–1129, 2006.
- [11] D. Gürsoy, T. Biçer, A. Lanzirrotti, M. G. Newville, and F. De Carlo, “Hyperspectral image reconstruction for x-ray fluorescence tomography,” *Optics express*, vol. 23, no. 7, pp. 9014–9023, 2015.
- [12] N. P. Galatsanos, A. K. Katsaggelos, R. T. Chin, and A. D. Hillery, “Least squares restoration of multichannel images,” *IEEE Transactions on Signal Processing*, vol. 39, no. 10, pp. 2222–2236, 1991.
- [13] S. Henrot, C. Soussen, and D. Brie, “Fast positive deconvolution of hyperspectral images,” *IEEE Transactions on Image Processing*, vol. 22, no. 2, pp. 828–833, 2013.
- [14] S. Henrot, S. Moussaoui, C. Soussen, and D. Brie, “Edge-preserving nonnegative hyperspectral image restoration,” in *38th International Conference on Acoustics, Speech, and Signal Processing, ICASSP 2013*, 2013.
- [15] G. H. Golub, M. Heath, and G. Wahba, “Generalized cross-validation as a method for choosing a good ridge parameter,” *Technometrics*, vol. 21, no. 2, pp. 215–223, 1979.
- [16] N. P. Galatsanos and A. K. Katsaggelos, “Methods for choosing the regularization parameter and estimating the noise variance in image restoration and their relation,” *IEEE Transactions on Image Processing*, vol. 1, no. 3, pp. 322–336, 1992.
- [17] P. C. Hansen, “Analysis of discrete ill-posed problems by means of the L-curve,” *SIAM review*, vol. 34, no. 4, pp. 561–580, 1992.
- [18] P. C. Hansen, *The L-curve and its use in the numerical treatment of inverse problems*. Department of Mathematical Modelling, Technical University of Denmark, 1999.
- [19] C. R. Vogel, “Non-convergence of the L-curve regularization parameter selection method,” *Inverse problems*, vol. 12, no. 4, p. 535, 1996.
- [20] M. Hanke, “Limitations of the L-curve method in ill-posed problems,” *BIT Numerical Mathematics*, vol. 36, no. 2, pp. 287–301, 1996.
- [21] M. Belge, M. E. Kilmer, and E. L. Miller, “Efficient determination of multiple regularization parameters in a generalized L-curve framework,” *Inverse Problems*, vol. 18, no. 4, p. 1161, 2002.
- [22] T. Reginska, “A regularization parameter in discrete ill-posed problems,” *SIAM Journal on Scientific Computing*, vol. 17, no. 3, pp. 740–749, 1996.
- [23] L. Kaufman and A. Neumaier, “Regularization of ill-posed problems by envelope guided conjugate gradients,” *Journal of Computational and Graphical Statistics*, vol. 6, no. 4, pp. 451–463, 1997.
- [24] E. Van Den Berg and M. P. Friedlander, “Probing the Pareto frontier for basis pursuit solutions,” *SIAM Journal on Scientific Computing*, vol. 31, no. 2, pp. 890–912, 2008.
- [25] D. L. Donoho and Y. Tsaig, “Fast solution of l_1 -norm minimization problems when the solution may be sparse,” *IEEE Transactions on Information Theory*, vol. 54, no. 11, pp. 4789–4812, 2008.
- [26] Y. Song, D. Brie, E.-H. Djermoune, and S. Henrot, “Regularization parameter estimation for non-negative hyperspectral image deconvolution: supplementary material,” CRAN, Research Report, May 2016.
- [27] A. F. Goetz, “Three decades of hyperspectral remote sensing of the earth: A personal view,” *Remote Sensing of Environment*, vol. 113, pp. S5–S16, 2009.
- [28] L. Duponchel, W. Elmi-Rayaleh, C. Ruckebusch, and J.-P. Huvenne, “Multivariate curve resolution methods in imaging spectroscopy: influence of extraction methods and instrumental perturbations,” *Journal of Chemical information and computer sciences*, vol. 43, no. 6, pp. 2057–2067, 2003.
- [29] S. Piqueras, L. Duponchel, R. Tauler, and A. de Juan, “Monitoring polymorphic transformations by using in situ Raman hyperspectral imaging and image multiset analysis,” *Analytica Chimica Acta*, vol. 819, pp. 15–25, 2014.
- [30] A. Gowen, C. O’Donnell, P. Cullen, G. Downey, and J. Frias, “Hyperspectral imaging—an emerging process analytical tool for food quality and safety control,” *Trends in Food Science & Technology*, vol. 18, no. 12, pp. 590–598, 2007.
- [31] T. Zimmermann, J. Rietdorf, and R. Pepperkok, “Spectral imaging and its applications in live cell microscopy,” *FEBS letters*, vol. 546, no. 1, pp. 87–92, 2003.
- [32] G. Lu and B. Fei, “Medical hyperspectral imaging: a review,” *Journal of Biomedical Optics*, vol. 19, no. 1, pp. 010901–010901, 2014.

- [33] R. Salzer and H. W. Siesler, *Infrared and Raman spectroscopic imaging*. John Wiley & Sons, 2009.
- [34] Y. Hiraoka, T. Shimi, and T. Haraguchi, "Multispectral imaging fluorescence microscopy for living cells." *Cell Structure and Function*, vol. 27, no. 5, pp. 367–374, 2002.
- [35] S. Henrot, "Déconvolution et séparation d'images hyperspectrales en microscopie," Ph.D. dissertation, Université de Lorraine, 2013.
- [36] S. Henrot, C. Soussen, and D. Brie, "Fast deconvolution of large fluorescence hyperspectral images," in *3rd Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing, WHISPERS 2011*, Lisbonne, Portugal, Jun. 2011, p. CDR0M, 4 pages. [Online]. Available: <https://hal.archives-ouvertes.fr/hal-00593546>
- [37] A. Tikhonov and V. Arsenin, *Solutions of ill-posed problems*, ser. Scripta Series in Mathematics. Winston Washington, DC, 1977.
- [38] J. Idier, Ed., *Bayesian Approach to Inverse Problems*. ISTE Ltd and John Wiley & Sons Inc, avr. 2008.
- [39] J. Nocedal and S. Wright, "Numerical optimization, series in operations research and financial engineering," *Springer, New York, USA*, 2006.
- [40] K. Deb, *Multi-objective optimization using evolutionary algorithms*. John Wiley & Sons, 2001.
- [41] I. Das and J. E. Dennis, "A closer look at drawbacks of minimizing weighted sums of objectives for pareto set generation in multicriteria optimization problems," *Structural optimization*, vol. 14, no. 1, pp. 63–69, 1997.
- [42] I. Y. Kim and O. De Weck, "Adaptive weighted-sum method for bi-objective optimization: Pareto front generation," *Structural and multidisciplinary optimization*, vol. 29, no. 2, pp. 149–158, 2005.
- [43] K. Deb and H. Jain, "An evolutionary many-objective optimization algorithm using reference-point-based nondominated sorting approach, part I: Solving problems with box constraints," *IEEE Transactions on Evolutionary Computation*, vol. 18, no. 4, pp. 577–601, 2014.
- [44] H. Jain and K. Deb, "An evolutionary many-objective optimization algorithm using reference-point based nondominated sorting approach, part II: handling constraints and extending to an adaptive approach," *IEEE Transactions on Evolutionary Computation*, vol. 18, no. 4, pp. 602–622, 2014.
- [45] S. Boyd and L. Vandenberghe, *Convex optimization*. Cambridge university press, 2004.
- [46] J. Kim, "Iterated grid search algorithm on unimodal criteria," Ph.D. dissertation, Virginia Polytechnic Institute and State University, 1997.
- [47] J. M. Bardsley, J. K. Merikoski, and R. Vio, "The stabilizing properties of nonnegativity constraints in least-squares image reconstruction," *International Journal of Pure and Applied Mathematics*, vol. 43, no. 1, p. 95, 2008.
- [48] B. Zhang, J. Zerubia, and J.-C. Olivo-Marin, "Gaussian approximations of fluorescence microscope point-spread function models," *Applied Optics*, vol. 46, no. 10, pp. 1819–1829, 2007.
- [49] P. Pankajakshan, "Blind deconvolution for confocal laser scanning microscopy," Ph.D. dissertation, Universit Nice Sophia Antipolis, 2009.