Minimum distance criterion for non negative hyperspectral image

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

MINIMUM DISTANCE CRITERION FOR NON-NEGATIVE HYPERSPECTRAL IMAGE DECONVOLUTION

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

Université de Lorraine, Centre de Recherche en Automatique de Nancy (CRAN), CNRS
Boulevard des Aiguilletes B.P. 239 F-54506, Vandœuvre-lès-Nancy, France

ABSTRACT

This work aims at studying a method to automatically estimate regularization parameters of hyperspectral images 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) is proposed to estimate regularization parameters. It has good theoretical properties (uniqueness, robustness) from which a grid search based approach is proposed. It results in a fast approach to estimate the regularization parameters.

***Index Terms*—** Hyperspectral image deconvolution, multi-objective optimization, regularization parameter estimation

1. INTRODUCTION

Hyperspectral image deconvolution consists in removing the blur to restore the original images at best. This can be formulated 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 [1], [2] and [3]. However, the effective implementation of such methods is hampered by the choice of the regularization parameters. The goal of the present paper is to propose a general approach to estimate the regularization parameter of deconvolution methods formulated as a convex multi-objective minimization problem. Here we only consider the Tikhonov-like hyperspectral image deconvolution with non-negativity constraint of [1] but this approach can be applied to other methods.

The L-curve presented in [4], [5] is a method for selecting a single regularization parameter by plotting in a log-log scale the data fitting term versus the regularization term. This curve exhibits a corner where the curvature is expected to reach a maximum value yielding an estimated regularization parameter. However, the L-curve approach has some undesirable properties discussed in [6] and [7]. In particular, it is not convex and the maximum curvature can be not unique. L-hypersurface as a multi-objective extension of the L-curve for selecting multiple regularization parameters was introduced in [8] which also developed a minimum distance criterion (MDC) 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. The MDC was already introduced in [9] for the bi-objective case only.

More recently, [10] shows that basis pursuit corresponds to a weighted sum approach allowing to solve a convex bi-objective optimization problem. In particular, it is proved that the corresponding Pareto front is convex and continuously differentiable over all points of interest. In fact, the Pareto front of the basis pursuit is strongly connected to the regularization path for which continuation based approach allows the fast calculation of the set of all solutions when the regularization is varying from 0 to $+\infty$ [11]. It is worth mentioning that [10] and [11] do not address the problem of selecting a particular solution from the set of all solutions.

The paper is organized as follows : in section 2, we briefly present the non-negative hyperspectral image deconvolution problem. In section 3, it is formulated as a multi objective optimization problem and the properties of the corresponding response surface are studied. In section 4, the MDC directly applied to the response surface is proposed to estimate the regularization parameters. It is associated with a grid search strategy to drastically reduce the computational burden of the procedure. In section 5, we present some numerical experiments allowing to assess the effect of the non-negativity constraint on hyperspectral image deconvolution. Due to space limitation, the proofs of the theoretical results are given in [12].

2. HYPERSPECTRAL IMAGE DECONVOLUTION

We call the unknown hyperspectral image $\mathbf{x}$ and the observed image $\mathbf{y}$. The blurred image corresponds to the product of $\mathbf{x}$ by a block diagonal convolution matrix $\mathbf{H}$ (see [1] for details). The blurred and noisy hyperspectral image is obtained by adding a noise $\mathbf{e}$ which results in the observation model:

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{e}. \tag{1}$$

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

In [1], the deconvolution is stated as a minimization of a criterion composed of three terms: the data fitting, the spatial regularization and the spectral regularization. A non-negativity constraint is added to the regular Tikhonov method to guarantee the restored image to be non-negative:

$$\min_{\mathbf{x}\geq 0} J(\mathbf{x}) = \frac{1}{2}\|\mathbf{y}-\mathbf{H}\mathbf{x}\|_2^2 + \frac{\mu_s}{2}\|\mathbf{D}_s\mathbf{x}\|_2^2 + \frac{\mu_\lambda}{2}\|\mathbf{D}_\lambda\mathbf{x}\|_2^2 \quad (2)$$

Here μ_s and μ_λ are respectively the spatial and spectral regularization parameters. $\mathbf{D}_s$ corresponds to a Laplacian filter applied to each slice of the hyperspectral image and $\mathbf{D}_\lambda$ corresponds to a first order derivative filter along the spectral dimension (see [1] for more details).

To minimize (2), we use the quadratic penalty method proposed in [1] which consists in introducing a slack variable $\mathbf{p}$ into a surrogate criterion expressed in (3)

$$\min_{\mathbf{x},\mathbf{p}} K(\mathbf{x},\mathbf{p};\xi) = J(\mathbf{x}) + \frac{\xi}{2}\|\mathbf{x}-\mathbf{p}\|_2^2 \quad \text{s.t.} \quad \mathbf{p}\geq 0 \quad (3)$$

The solution is obtained iteratively. At each iteration, the following three steps are performed: unconstrained minimization of $\mathbf{x}^{k+1}$ when $\mathbf{p}^k$ and ξ^k are fixed; constrained minimization of $\mathbf{p}^{k+1} = \max(\mathbf{0}, \mathbf{x}^{k+1})$; increase of the penalty factor ξ according to $\xi^{(k+1)} = \gamma\xi^{(k)}, \gamma > 1$. These three steps are alternated until a maximum number of iterations N_{iter} is reached. In the used implementation of the algorithm, the case $N_{iter} = 1$ corresponds to the unconstrained solution which serves as initial solution in the iterative procedure.

3. HYPERSPECTRAL IMAGE DECONVOLUTION AS A MULTI-OBJECTIVE OPTIMIZATION

3.1. Multi-objective Optimization

Problem (2) can be stated as a multiple objective optimization problem as in (4):

$$\min_{\mathbf{x}\geq 0} J(\mathbf{x}) = J_1(\mathbf{x}) + \mu_s J_2(\mathbf{x}) + \mu_\lambda J_3(\mathbf{x}) \quad (4)$$

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

$$\mathbf{x}_{\boldsymbol{\mu}} = \arg\min_{\mathbf{x}\geq 0} J(\mathbf{x}) \quad (5)$$

and gives a point in the response surface. Unlike the L-curve or the L-hypersurface, this response surface uses linear scales axes. For notation simplicity we will write $J_k(\mathbf{x}_{\boldsymbol{\mu}}) \triangleq J_k(\boldsymbol{\mu})$ and the same for $J(\mathbf{x}_{\boldsymbol{\mu}}) \triangleq J(\boldsymbol{\mu})$. The ideal objective vector is defined as in [13] :

$$\mathbf{I} = (I_1, \cdots, I_z) \quad (6)$$

where z is the number of objectives which equals 3 in our case. The k-th component of the ideal objective vector $\mathbf{I}$ is the constrained minimum of problem (7):

$$I_k = \min_{\mathbf{x}\geq 0} J_k(\mathbf{x}). \quad (7)$$

In multi-objective literature, the ideal objective vector is said to be a non-existent solution which means that it can never be reached since it does not belong to the response surface.

3.2. Properties of the response surface

Because $J(\mathbf{x})$ consists in the sum of three convex objectives and the non-negative orthant is convex, this problem remains convex.

Theorem 1. *If $J(\mathbf{x})$ is convex, the response surface of problem (4) is convex.*

The demonstration of this theorem follows the same lines of the approach [10] but is extended to the tri-objective case.

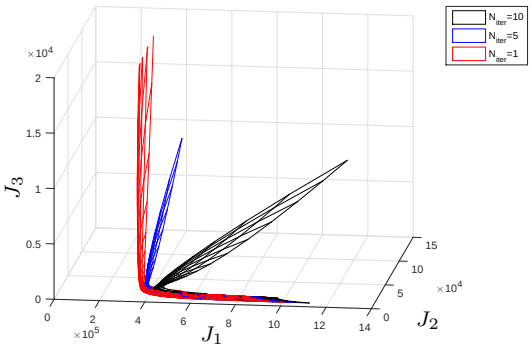
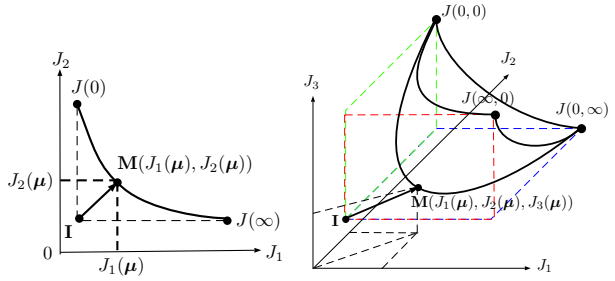


Fig. 1: Estimated response surface for different values of N_{iter}

Figure 1 shows three different empirical response surfaces estimated from the same simulated example (see section 5) for $N_{iter} = 1, 5, 10$. For each response surface, the hyperparameters are sampled on a 20×20 regular logarithmic scale varying from 0.1 to 1000. The red one ($N_{iter} = 1$) corresponds to the Tikhonov solution without the non-negativity constraint. The two others correspond to the response surface obtained with non-negative Tikhonov solution of section 2. For both cases the penalty factor ξ is evolving similarly. Following [10], 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 (red response surface in figure 1). On the contrary, when a non-negativity constraint is enforced, a folding of the response surface is observed (figure 1, $N_{iter} = 5, 10$). This results from the constrained estimator data fitting J_1 which is proved to be decreasing and then increasing as μ_s

(or μ_λ) increases. In this case, only the set of non-dominated point of the response surface is corresponding to the Pareto front.



(a) bi-objective case (b) tri-objective case

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

4. REGULARIZATION PARAMETERS ESTIMATION

4.1. Minimum distance criterion (MDC)

To estimate the regularization parameters, we propose the MDC. Such a criterion was already proposed in [8] where it is applied to the L-hypersurface. Instead, we use this criterion directly on the response surface. The main reason for this is that the L-hypersurface is no longer guaranteed to be convex while the response surface is. The convexity of the response surface is central in establishing the properties of the MDC. The ideal point as defined in section 3.1 corresponds to the minimum of all objective functions. Even it is a non-existent 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. Let us introduce the MDC by defining first, the distance to the ideal point and then the MDC.

Definition 1 (Distance to Ideal Point)**.** *Let* $\mathbf{I} = (I_1, \cdots, I_z)$ *denotes the coordinates of the ideal point. The minimum distance function* $D(\boldsymbol{\mu})$ *is the distance from the ideal Point* $\mathbf{I}$ *to the point* $\mathbf{M}(\boldsymbol{\mu}) = (J_1(\boldsymbol{\mu}), \cdots, J_z(\boldsymbol{\mu}))$ *on the response surface.*

$$D(\boldsymbol{\mu}) = \sum_{i=1}^{z} (J_i(\boldsymbol{\mu}) - I_i)^2 \tag{8}$$

Definition 2 (Minimum Distance Criterion)**.**

$$\min_{\boldsymbol{\mu}} D(\boldsymbol{\mu}) \tag{9}$$

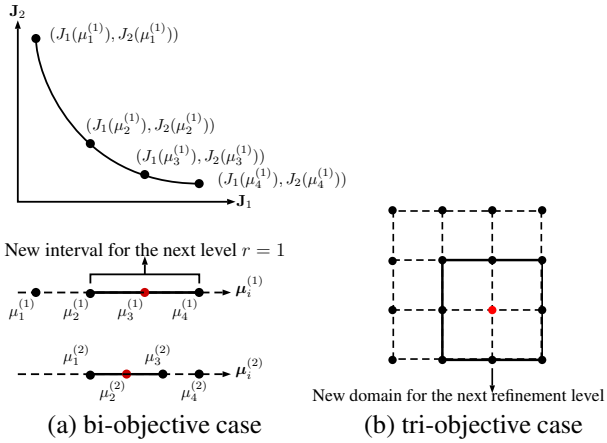
Theorem 2. *If the response surface is convex, the MDC admits a unique minimum.*

The proof of this theorem relies on a geometrical interpretation of the MDC.

Let us mention that the empirical response surface may be slightly non-convex, something which can be attributed to numerical errors. We also proved that the MDC still admits a unique minimum if it is unimodal (hence not necessarily convex). In that respect, the MDC can be said to be robust to the loss of convexity (see [12] for details).

The evaluation of the MDC requires the determination of the ideal point which is difficult to obtain when the non-negative constraint is enforced. This is due to the folding discussed in section 3.2. We propose to define it by determining the points of the response surface corresponding to the unconstrained Tikhonov solution for three values of $\boldsymbol{\mu}$ equal to $(0,0), (0,\infty), (\infty,0)$[1]. The ideal point coordinates are determined by finding the minimum coordinate of each of the three points (see figure 2(b)).

4.2. A grid search strategy to minimize the MDC



(a) bi-objective case (b) tri-objective case

Fig. 3: Grid refinement method

To reduce the computational burden of the MDC, we use a grid search approach which is proved to be convergent for unimodal criterion [14]. Figure 3 illustrate the grid refinement. For the bi-objective case with one single regularization parameter, at the first level $r = 1$, we have only four points $\boldsymbol{\mu}_i^{(1)} (i = 1, \cdots, 4)$ on which the response surface is estimated. Then the grid is refined by defining a new search segment on which four new points $\boldsymbol{\mu}_i^{(2)} (i = 1, \cdots, 4)$ are defined. The procedure is repeated until a maximum number of levels is reached. In the tri-objective case with two regularization parameters, we define 4×4 points for r-th level and choose an optimum point among them. Then we select the points around it as the new domain. We refine this new domain to choose a new optimum point. The procedure is repeated iteratively. The algorithm corresponding to this strategy is proposed in [12]. In what follows, the number of levels

[1] In practice, the value of the hyperparameters cannot be fixed to ∞ but are fixed to large values

is fixed to 6 which gives approximately the same resolution as the 20×20 grid of section 3. At each level, only 4 new points of the response surface need to be estimated thus requiring 24 evaluations for the 6 levels instead of 400 required for the 20×20 grid.

5. EXPERIMENTS

To simulate the blurred hyperspectral images, we first simulate the unblurred image according to the instantaneous mixture model[2]. The convolution filter $\mathbf{H}$ is assumed to be a low pass gaussian filter whose bandwidth is lower than the bandwidth of $\mathbf{x}$ resulting in a difficult deconvolution problem. A gaussian noise is then added to the blurred image. The size of the resulting hyperspectral image is $(120 \times 120 \times 32)$. It should also be mentioned that the simulation example was designed to favor the non-negative deconvolution.

Here we use the MDC to answer the following question : does the non-negativity constraint improve the deconvolution results? The MDC yields a different solution for each value of the SNR. Thus it is possible to evaluate the performance of the unconstrained and constrained deconvolution algorithms (coupled with an automatic regularization parameter estimation) by evaluating the mean square error (MSE) as a function of the SNR.

The MSE curve includes three main parts which are represented on figure 4. 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 with a lower rate than the noise. This is the zone where the deconvolution is effective. Finally the third horizontal part corresponds to the maximal performance of the deconvolution. As the bandwidth of the filter $\mathbf{H}$ is lower than the bandwidth of $\mathbf{x}$, even in noise free situations, the deconvolution can not restore $\mathbf{x}$ outside the frequency range (bandwidth) covered by $\mathbf{H}$. The minimal MSE value reflects the ill-conditioning of the matrix $\mathbf{H}$. It decreases as the condition number of $\mathbf{H}$ decreases.

The Tikhonov approach with non-negativity constraint is an iterative algorithm which converges to the optimal solution as the number of iteration increases. In fact, analyzing the non-negative deconvolution performance evolution as a function of N_{iter} aims at evaluating how the convergence of the algorithm is affecting the quality of the deconvolution results.

The MSEs obtained for different values of N_{iter} are shown in figure 5. Clearly, the non-negativity constraint improves the effectiveness of the regularization parameter estimation. This highlights the stabilizing property of the non negativity constraint as proved in [15]. For example, when $N_{iter} = 1$ (unconstrained Tikhonov approach), the efficiency zone is in the interval $[20, 30]$ dB and the MSE reaches the minimal MSE value when the SNR is greater than 30 dB.

[2] An instantaneous mixture of 3 sources having 32 spectral bands is considered.

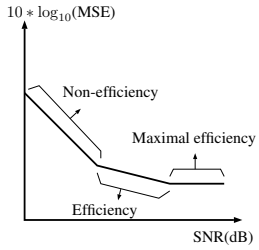


Fig. 4: Typical shape of the MSE

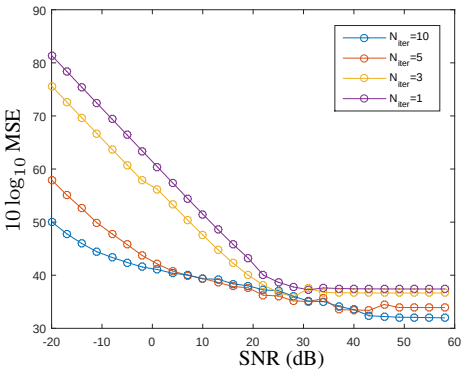


Fig. 5: Mean squared error of the deconvolution algorithms as a function of the SNR

However when $N_{iter} = 10$ (constrained Tikhonov approach), not only the minimum MSE is decreased but the efficiency zone (which is approximatively between -10 dB and 40 dB) increases significantly. The MSE obtained for the intermediate values of N_{iter} (3 and 5) shows how it is gradually changing from the unconstrained to the constrained case.

6. CONCLUSION

A first contribution of this work is to set the problem of hyperspectral images deconvolution as a multi-objective optimization problem whose response surface is proved to be convex. But, contrarily to the unconstrained case, the response surface does not coincide with the Pareto front. A second contribution is the proposal of the MDC to estimate the optimal values of the regularization parameters μ_s and μ_λ. We proved that this criterion admits a unique minimum even if the MDC is non convex but only unimodal. A fast grid search algorithm is proposed to estimate the point of the response surface minimizing the MDC. The MDC performances are assessed with respect to the SNR for two deconvolution algorithms (with and without non-negativity constraint). It appears that the non-negativity constraint results in a significant improvement of the estimation accuracy, thus confirming the regularization property of the non-negativity constraint. Illustrative examples and a comparison of the MDC with state of the art methods can be found in [12].

7. REFERENCES

[1] Simon Henrot, Charles Soussen, and David Brie, "Fast positive deconvolution of hyperspectral images," *IEEE Transactions on Image Processing*, vol. 22, no. 2, pp. 828–833, 2013.

[2] Simon Henrot, Saïd Moussaoui, Charles Soussen, and David Brie, "Edge-preserving nonnegative hyperspectral image restoration," in *38th International Conference on Acoustics, Speech, and Signal Processing, ICASSP 2013*, 2013.

[3] Simon Henrot, Charles Soussen, Saïd Moussaoui, and David Brie, "Restauration positive d'images hyperspectrales avec préservation des contours," in *XXIVe Colloque GRETSI Traitement du Signal & des Images, GRETSI 2013*, 2013.

[4] Per Christian Hansen, "Analysis of discrete ill-posed problems by means of the L-curve," *SIAM review*, vol. 34, no. 4, pp. 561–580, 1992.

[5] Per Christian Hansen, *The L-curve and its use in the numerical treatment of inverse problems*, IMM, Department of Mathematical Modelling, Technical University of Denmark, 1999.

[6] Curtis R Vogel, "Non-convergence of the L-curve regularization parameter selection method," *Inverse problems*, vol. 12, no. 4, pp. 535, 1996.

[7] Martin Hanke, "Limitations of the L-curve method in ill-posed problems," *BIT Numerical Mathematics*, vol. 36, no. 2, pp. 287–301, 1996.

[8] Murat Belge, Misha E Kilmer, and Eric L Miller, "Efficient determination of multiple regularization parameters in a generalized L-curve framework," *Inverse Problems*, vol. 18, no. 4, pp. 1161, 2002.

[9] Teresa Reginska, "A regularization parameter in discrete ill-posed problems," *SIAM Journal on Scientific Computing*, vol. 17, no. 3, pp. 740–749, 1996.

[10] Ewout Van Den Berg and Michael P Friedlander, "Probing the Pareto frontier for basis pursuit solutions," *SIAM Journal on Scientific Computing*, vol. 31, no. 2, pp. 890–912, 2008.

[11] David L Donoho and Yaakov Tsaig, "Fast solution of-norm minimization problems when the solution may be sparse," *Information Theory, IEEE Transactions on*, vol. 54, no. 11, pp. 4789–4812, 2008.

[12] Yingying Song, David Brie, El-Hadi Djermoune, and Simon Henrot, "Regularization parameter estimation for non-negative hyperspectral image deconvolution," Research report, CRAN <hal-01204500>, Sept. 2015.

[13] Kalyanmoy Deb, *Multi-objective optimization using evolutionary algorithms*, John Wiley & Sons, 2001.

[14] Jinhyo Kim, *Iterated grid search algorithm on unimodal criteria*, Ph.D. thesis, Virginia Polytechnic Institute and State University, 1997.

[15] Johnathan M Bardsley, Jorma K Merikoski, and Roberto Vio, "The stabilizing properties of nonnegativity constraints in least-squares image reconstruction," *International Journal of Pure and Applied Mathematics*, vol. 43, no. 1, pp. 95, 2008.