Exploiting the scaling indetermination of bi-linear models in inverse problems

Samuel The, Eric Thiebaut, Loic Denis, Ferréol Soulez

Exploiting the scaling indetermination of bi-linear models in inverse problems

Samuel Thé[a], Éric Thiébaut[a], Loïc Denis[b] & Ferréol Soulez[a]

[a] Univ. Lyon, Univ. Lyon 1, ENS de Lyon, CNRS, Centre de Recherche Astrophysique de Lyon UMR5574, F-69230, Saint-Genis-Laval, France

[b] Univ. Lyon, UJM-Saint-Etienne, CNRS, Institut d Optique Graduate School, Laboratoire Hubert Curien UMR 5516, F-42023, Saint-Étienne, France

***Abstract*—Many inverse problems in imaging require estimating the parameters of a bi-linear model, e.g., the crisp image and the blur in blind deconvolution. In all these models, there is a scaling indetermination: multiplication of one term by an arbitrary factor can be compensated for by dividing the other by the same factor.**

To solve such inverse problems and identify each term of the bi-linear model, reconstruction methods rely on prior models that enforce some form of regularity. If these regularization terms verify a homogeneity property, the optimal scaling with respect to the regularization functions can be determined. This has two benefits: hyper-parameter tuning is simplified (a single parameter needs to be chosen) and the computation of the maximum *a posteriori* estimate is more efficient.

Illustrations on a blind deconvolution problem are given with an unsupervised strategy to tune the hyper-parameter.

***Index Terms*—Inverse problem, Bi-linear models, Scaling indetermination, Blind deconvolution**

I. Introduction

There are many problems in signal and image processing that can be formulated as inverse problems: the observed signals or images do not directly correspond to the parameters of interest but can be related via a model. Estimating the parameters of interest requires inverting the model. It is often the case that the model also depends on other parameters, e.g., instrumental parameters, and that these parameters need to be jointly estimated from the data. Such inverse problems are often called "*blind*" reconstruction problems to emphasize that, beyond the reconstruction of the parameters of interest, another set of parameters must be estimated. The family of bi-linear models covers a wide range of such models.

In order to identify each set of parameters in a blind reconstruction problem, regularization terms are introduced. These terms favor solutions in good agreement with our prior knowledge, for example with sufficient temporal or spatial smoothness.

There is a scaling indetermination to all bi-linear models: the model is unchanged when one set of parameters is multiplied by a factor and the other is divided by the same factor. This indetermination is solved by the regularization and the domain constraints imposed on the parameters.

We show in this paper that regularization terms that verify a homogeneity property lead to a closed form expression of the optimal scaling, for a given weighting of each regularization term. This optimal scaling can be beneficial by (i) simplifying the tuning of the relative weights of regularization terms (so-called hyper-parameters of the blind reconstruction problem); (ii) leading to a more efficient reconstruction algorithm; (iii) being applicable to an unsupervised reconstruction strategy based on the minimization of a criterion like Stein's unbiased risk estimate SURE [1].

The paper is organized as follows: Section II derives the expression of the optimal scaling and describes the proposed method. Section III illustrates this general methodology on a blind deconvolution problem, i.e., the joint estimation of a crisp image and of the blur. Section IV concludes the paper.

II. Proposed Method

A. Bi-linear models and scaling invariance

We consider a class of inverse problems for which the direct model of the data writes:

$$\boldsymbol{d} \approx \boldsymbol{x} \otimes \boldsymbol{y}\,, \tag{1}$$

where $\boldsymbol{d} \in \mathbb{D}$, $\boldsymbol{x} \in \mathbb{X}$ and $\boldsymbol{y} \in \mathbb{Y}$ are real-valued vectors: $\boldsymbol{d}$ represents the measurements while $\boldsymbol{x}$ and $\boldsymbol{y}$ represent unknowns which are combined by the bi-linear operator $\otimes$. The above symbol $\approx$ is to account for discrepancies due to the approximations of the model $\boldsymbol{x} \otimes \boldsymbol{y}$ and to measurement noise. Since the direct model $\boldsymbol{x} \otimes \boldsymbol{y}$ is bi-linear, it verifies the property:

$$\forall(\alpha, \beta, \boldsymbol{x}, \boldsymbol{y}) \in \mathbb{R}^2 \times \mathbb{X} \times \mathbb{Y},\ (\alpha\,\boldsymbol{x}) \otimes (\beta\,\boldsymbol{y}) = (\alpha\,\beta)\,(\boldsymbol{x} \otimes \boldsymbol{y}) \tag{2}$$

Instances of such problems are numerous in signal and image processing, for example:

- *Blind deconvolution* [2] with $\boldsymbol{x}$ the observed object, $\boldsymbol{y}$ the instrumental point spread function (PSF) and $\otimes$ the bi-dimensional convolution operator.
- *Blind source separation* [3] where $\boldsymbol{x}$ and $\boldsymbol{y}$ are matrices[1] while $\otimes$ denotes the usual matrix product.

[1] In the case of blind source separation, $\boldsymbol{x}$ may represent the *end-members* while $\boldsymbol{y}$ represent the mixture coefficients.

- *Low-rank approximations* where a matrix $\boldsymbol{d}$ is approximated as the matrix product of a tall matrix and a wide matrix.

To account for the approximate relation in Eq. (1), the inverse problem of recovering the unknowns $\boldsymbol{x}$ and $\boldsymbol{y}$ given the data $\boldsymbol{d}$ involves fitting the model $\boldsymbol{x} \otimes \boldsymbol{y}$ to the data. Formally, this fitting can be expressed by:

$$\min_{\boldsymbol{x}\in\mathbb{X},\boldsymbol{y}\in\mathbb{Y}} \mathcal{G}(\boldsymbol{x} \otimes \boldsymbol{y}) \tag{3}$$

where $\mathcal{G}\colon \mathbb{D} \to \mathbb{R}$ implements a distance between the model $\boldsymbol{x} \otimes \boldsymbol{y}$ and the data $\boldsymbol{d}$.

As a consequence of the Property (2), the relation

$$\mathcal{G}((\alpha\,\boldsymbol{x}) \otimes (\boldsymbol{y}/\alpha)) = \mathcal{G}(\boldsymbol{x} \otimes \boldsymbol{y}) \tag{4}$$

holds for any $\alpha \in \mathbb{R}^\star \equiv \mathbb{R}\backslash\{0\}$. This means that there is a multiplicative degeneracy in the inverse problem which cannot be disentangled by the data alone whatever the *metric* implemented by $\mathcal{G}$. In the general case, the Problem (3) is extremely ill-posed due to its numerous degeneracies (such as swap, shift or reducibility degeneracies in the blind deconvolution case [4]).

Hence, it is necessary to add some *a priori* knowledge on both unknowns $\boldsymbol{x}$ and $\boldsymbol{y}$ by the means of regularization penalties and/or by tightening the feasible sets $\mathbb{X}$ and/or $\mathbb{Y}$ [5]. The inverse problem then amounts to solving the following optimization problem:

$$\min_{\boldsymbol{x}\in\mathbb{X},\boldsymbol{y}\in\mathbb{Y}} \big\{\mathcal{F}(\boldsymbol{x},\mu,\boldsymbol{y},\nu) = \mathcal{G}(\boldsymbol{x}\otimes\boldsymbol{y}) + \mu\,\mathcal{J}(\boldsymbol{x}) + \nu\,\mathcal{K}(\boldsymbol{y})\big\} \tag{5}$$

where $\mathcal{J}\colon \mathbb{X} \to \mathbb{R}_+$ and $\mathcal{K}\colon \mathbb{Y} \to \mathbb{R}_+$ are regularization penalties to enforce the prior constraints, $\mu \in \mathbb{R}_+$ and $\nu \in \mathbb{R}_+$ are hyper-parameters to tune the relative weights of the different terms of the objective function $\mathcal{F}$. Many *a priori* models have been proposed in the literature, such as quadratic regularization [5], *total variation* (TV) [6] or *edge-preserving smoothness* [7].

B. Exploiting the scaling indetermination

In order to exploit the scaling indetermination of the direct model, we require that the regularization penalties $\mathcal{J}(\boldsymbol{x})$ and $\mathcal{K}(\boldsymbol{y})$ be homogeneous functions of degree $q > 0$ and $r > 0$ respectively. That is:

$$\mathcal{J} \in \Omega(\mathbb{X}, q) \quad \text{and} \quad \mathcal{K} \in \Omega(\mathbb{Y}, r)\,, \tag{6}$$

where $\Omega(\mathbb{U}, p)$ denotes the set of homogeneous functions from the convex cone[2] $\mathbb{U}$ to $\mathbb{R}_+$ of degree p. In other words, we assume that $\mathbb{X}$ and $\mathbb{Y}$ are convex cones and that:

$$\mathcal{J}(\alpha\,\boldsymbol{x}) = \alpha^q\,\mathcal{J}(\boldsymbol{x}) \geq 0\,, \tag{7a}$$

$$\mathcal{K}(\alpha\,\boldsymbol{y}) = \alpha^r\,\mathcal{K}(\boldsymbol{y}) \geq 0\,, \tag{7b}$$

hold for any $\boldsymbol{x} \in \mathbb{X}$, $\boldsymbol{y} \in \mathbb{Y}$ and $\alpha \in \mathbb{R}_+$.

Under these assumptions, the identity:

$$\boxed{\mathcal{F}(\alpha\,\boldsymbol{x},\mu,\boldsymbol{y}/\alpha,\nu) = \mathcal{F}\big(\boldsymbol{x},\mu\,\alpha^q,\boldsymbol{y},\nu\,\alpha^{-r}\big)}\,, \tag{8}$$

[2]A convex cone $\mathbb{U}$ is such that: $\boldsymbol{u} \in \mathbb{U} \Rightarrow \alpha\,\boldsymbol{u} \in \mathbb{U}$ for any $\alpha \geq 0$. Examples of convex cones are $\mathbb{R}^n$ or $\mathbb{R}^n_+$.

holds $\forall\alpha \in \mathbb{R}^\star_+$. This identity shows that scaling the unknown parameters in such a way that the model is left unchanged amounts to changing the regularization terms of the objective function.

We now introduce $\mathcal{E}(\alpha)$ the value of the objective function in Eq. (8) as a function of the scaling factor α:

$$\begin{aligned}\mathcal{E}(\alpha) &= \mathcal{F}(\alpha\,\boldsymbol{x},\mu,\boldsymbol{y}/\alpha,\nu)\\ &= \mathcal{G}(\boldsymbol{x}\otimes\boldsymbol{y}) + \mu\,\alpha^q\,\mathcal{J}(\boldsymbol{x}) + \nu\,\alpha^{-r}\,\mathcal{K}(\boldsymbol{y})\,,\end{aligned} \tag{9}$$

for any given $\boldsymbol{x} \in \mathbb{X}$, $\boldsymbol{y} \in \mathbb{Y}$, $\mu > 0$, $\nu > 0$ and $\alpha \neq 0$ such that $\mathcal{J}(\boldsymbol{x}) > 0$ and $\mathcal{K}(\boldsymbol{y}) > 0$. The derivative of $\mathcal{E}(\alpha)$ writes:

$$\mathcal{E}'(\alpha) = q\,\mu\,\alpha^{q-1}\,\mathcal{J}(\boldsymbol{x}) - r\,\nu\,\alpha^{-r-1}\,\mathcal{K}(\boldsymbol{y})\,.$$

Since $\mathcal{J}(\boldsymbol{x}) > 0$ and $\mathcal{K}(\boldsymbol{y}) > 0$, $\mathcal{E}(\alpha)$ is strictly convex on $\mathbb{R}^\star_+$. The optimal scaling factor which minimizes $\mathcal{E}(\alpha)$ is $\widehat{\alpha}$ the only positive root of $\mathcal{E}'(\alpha)$:

$$\boxed{\widehat{\alpha}(\boldsymbol{x},\mu,\boldsymbol{y},\nu) = \left(\frac{r\,\nu\,\mathcal{K}(\boldsymbol{y})}{q\,\mu\,\mathcal{J}(\boldsymbol{x})}\right)^{\frac{1}{q+r}}}\,. \tag{10}$$

Replacing the optimal scaling parameter $\widehat{\alpha}$ in the objective function yields:

$$\begin{aligned}\min_{\alpha>0} \mathcal{F}(\alpha\,\boldsymbol{x},\mu,\boldsymbol{y}/\alpha,\nu) &= \mathcal{F}(\alpha\,\boldsymbol{x},\mu,\boldsymbol{y}/\alpha,\nu)\big|_{\alpha=\widehat{\alpha}(\boldsymbol{x},\mu,\boldsymbol{y},\nu)}\\ &= \mathcal{H}\big(\boldsymbol{x},\boldsymbol{y},\widehat{\eta}(\mu,\nu)\big)\end{aligned} \tag{11}$$

where:

$$\boxed{\mathcal{H}(\boldsymbol{x},\boldsymbol{y},\eta) = \mathcal{G}(\boldsymbol{x}\otimes\boldsymbol{y}) + \eta\,\mathcal{J}(\boldsymbol{x})^{\frac{r}{q+r}}\,\mathcal{K}(\boldsymbol{y})^{\frac{q}{q+r}}}\,, \tag{12}$$

and:

$$\boxed{\widehat{\eta}(\mu,\nu) = \Big((r/q)^{\frac{q}{q+r}} + (q/r)^{\frac{r}{q+r}}\Big)\,(\mu^r\,\nu^q)^{\frac{1}{q+r}}}\,. \tag{13}$$

The criterion defined in Eq. (12) shows that when considering the shape of the solution (that is $\boldsymbol{x}$ and $\boldsymbol{y}$ up to some scaling factor), the regularity of the two variables is mutually dependent, indeed only the value of $\mu^r\,\nu^q$ matters to set the importance of the regularization. As a consequence a single hyper-parameter, η, instead of two, μ and ν, is sufficient to tune the regularization level. This property can be exploited to simplify the tuning of the regularization. Minimization problems of the form of equation (5) generally have several local minima. The solution reached by minimization algorithms, therefore, depends on the optimization path followed during the iterations. Exploiting relation (10) to select, at each iteration, the optimal scaling can also help avoiding some poor local minima.

There is no closed form solution for the general Problem (5), so most practical algorithms to solve this problem are iterative: starting from some initial guess $\big(\boldsymbol{x}^{(0)},\boldsymbol{y}^{(0)}\big) \in \mathbb{X}\times\mathbb{Y}$, the objective function is progressively minimized to yield improved estimates $\big(\boldsymbol{x}^{(m)},\boldsymbol{y}^{(m)}\big) \in \mathbb{X}\times\mathbb{Y}$ after m iterations.

To take advantage of the scaling invariance there are several possible strategies, the most obvious ones are:

1) Choose a regularization level $\eta > 0$ and minimize the objective function $\mathcal{H}(\boldsymbol{x},\boldsymbol{y},\eta)$ defined in Eq. (12).

Algorithm 1: Rescaled alternating strategy with two hyper-parameters (one is arbitrary).

Input: $\mu > 0$, $\nu > 0$, $\alpha_0 > 0$ and $\boldsymbol{y}^{(0)} \in \mathbb{Y}$ s.t. $\mathcal{K}\big(\boldsymbol{y}^{(0)}\big) > 0$.

Output: A local optimum $(\boldsymbol{x}^\star, \boldsymbol{y}^\star) \in \mathbb{X} \times \mathbb{Y}$.

for $k = 0, 1, 2, \ldots, m-1$ **do**

  while *true* **do**

    $\boldsymbol{x}^{(k+1)} \approx \widehat{\boldsymbol{x}}\big(\boldsymbol{y}^{(k)}, \alpha_k^q\, \mu\big)$ see Eq. (14a)

    $\alpha_{k+1/2} = \widehat{\alpha}\big(\boldsymbol{x}^{(k+1)}, \mu, \boldsymbol{y}^{(k)}, \nu\big)$ see Eq. (10)

    if $k > 0$ *or* $\alpha_{k+1/2} \approx \alpha_k$ **then break**;

    $\alpha_k \leftarrow \alpha_{k+1/2}$

  $\boldsymbol{y}^{(k+1)} \approx \widehat{\boldsymbol{y}}\big(\boldsymbol{x}^{(k+1)}, \alpha_{k+1/2}^{-r}\, \nu\big)$ see Eq. (14b)

  $\alpha_{k+1} = \widehat{\alpha}\big(\boldsymbol{x}^{(k+1)}, \mu, \boldsymbol{y}^{(k+1)}, \nu\big)$ see Eq. (10)

Result: $(\boldsymbol{x}^\star, \boldsymbol{y}^\star) = \big(\alpha_m\, \boldsymbol{x}^{(m)}, \boldsymbol{y}^{(m)}/\alpha_m\big)$

2) Choose the regularization levels $\mu > 0$ and $\nu > 0$ (one of which can have an arbitrary fixed value) and minimize the criterion $\mathcal{F}(\boldsymbol{x}, \mu\, \widehat{\alpha}^q, \boldsymbol{y}, \nu\, \widehat{\alpha}^{-r})$ with $\widehat{\alpha}$ the best scaling factor given by Eq. (10) and either set before the first iteration, *i.e.* according to $\boldsymbol{x}^{(0)}$ and $\boldsymbol{y}^{(0)}$, or at every stages of the optimization, *i.e.* according to $\boldsymbol{x}^{(k)}$ and $\boldsymbol{y}^{(k)}$. Optimally scaling the variables at every iterations (or the hyper-parameters) has been shown to speed-up the convergence [8] so the former variant should be more efficient.

Remarks: (i) Whatever the strategy, it is necessary that $\mathcal{J}\big(\boldsymbol{x}^{(k)}\big) > 0$ and $\mathcal{K}\big(\boldsymbol{y}^{(k)}\big) > 0$ for all iterates k. (ii) Owing to the identity in (8), the scaling of the hyper-parameters in the 2nd method is equivalent to a scaling of the variables.

Directly minimizing the objective function $\mathcal{H}(\boldsymbol{x}, \boldsymbol{y}, \eta)$ defined in Eq. (12) is not very practical as it requires to solve a non-linear problem whether the variables $\boldsymbol{x}$ and $\boldsymbol{y}$ are jointly, hierarchically or alternately updated. The 2nd strategy is easier to implement if $\boldsymbol{x}$ and $\boldsymbol{y}$ are updated alternately as it can re-use existing methods to solve the two half-problems of estimating $\boldsymbol{x}$ given $\boldsymbol{y}$ and μ and estimating $\boldsymbol{y}$ given $\boldsymbol{x}$ and ν:

$$\begin{aligned}\widehat{\boldsymbol{x}}(\boldsymbol{y}, \mu) &= \arg\min_{\boldsymbol{x} \in \mathbb{X}} \mathcal{F}(\boldsymbol{x}, \mu, \boldsymbol{y}, \nu) \\ &= \arg\min_{\boldsymbol{x} \in \mathbb{X}} \{\mathcal{G}(\boldsymbol{x} \otimes \boldsymbol{y}) + \mu\, \mathcal{J}(\boldsymbol{x})\}\,, \qquad (14a) \\ \widehat{\boldsymbol{y}}(\boldsymbol{x}, \nu) &= \arg\min_{\boldsymbol{y} \in \mathbb{Y}} \mathcal{F}(\boldsymbol{x}, \mu, \boldsymbol{y}, \nu) \\ &= \arg\min_{\boldsymbol{y} \in \mathbb{Y}} \{\mathcal{G}(\boldsymbol{x} \otimes \boldsymbol{y}) + \nu\, \mathcal{K}(\boldsymbol{y})\}\,. \qquad (14b)\end{aligned}$$

As previously discussed, these updates should be performed with rescaled hyper-parameters to apply the optimal scaling factor given in Eq. (10) before each update. These considerations suggest the alternating method implemented by our Algorithm 1 whose output (up to a scaling of the unknowns) depends on a single tuning hyper-parameter, the other being fixed at an arbitrary value.

An advantage of the proposed algorithm is that only $\boldsymbol{y}^{(0)}$ has to be provided. However, since $\boldsymbol{x}^{(0)}$ is unknown, the optimal initial scaling α_0 cannot be computed by the closed form expression (10). To automatically determine a good initial scaling, our algorithm implements a simple alternated method to jointly estimate $\boldsymbol{x}^{(1)}$ and α_0. This simple procedure ensures that the other iterations (and hence the output of the algorithm) do not depend on the scaling of $\boldsymbol{y}^{(0)}$. In other words, taking $\boldsymbol{y}^{(0)} = \beta\, \boldsymbol{y}_{\text{inp}}$ for a given $\boldsymbol{y}_{\text{inp}}$ gives the same results whatever the value of $\beta > 0$. As shown by Fig. 1, this feature is important to avoid that the solution and the convergence speed be driven by the scaling of the initial variables.

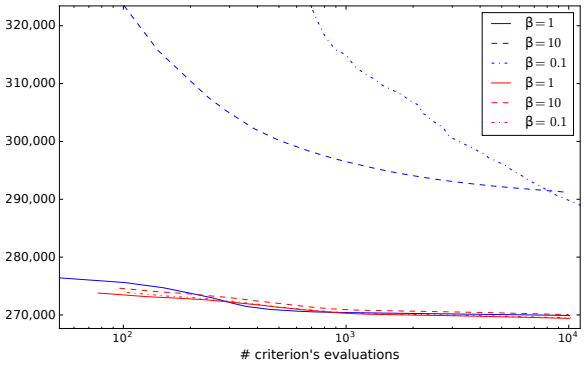

Fig. 1. Convergence curves for two versions of a blind deconvolution algorithm initialized with $\boldsymbol{y}^{(0)} = \beta\, \boldsymbol{y}_{\text{inp}}$ with different values of β but same input PSF $\boldsymbol{y}_{\text{inp}}$. In blue, the unscaled algorithm, that is alternated minimization of $\mathcal{F}(\boldsymbol{x}, \mu, \boldsymbol{y}, \nu)$ as defined in (5). In red, the proposed algorithm.

The updates of the variables $\boldsymbol{x}$ and $\boldsymbol{y}$ according to Eq. (14a) and (14b) do not have to be very precise, it is enough to seek for an approximate solution ensuring a sufficient reduction of the objective function. Starting at the previous estimate, a few iterations of the optimization algorithms designed to solve Eq. (14a) and (14b) will do the job. Thanks to the symmetry of the problem, the roles of $(\boldsymbol{x}, \mathbb{X}, \mu)$ and $(\boldsymbol{y}, \mathbb{Y}, \nu)$ can be exchanged and the algorithm started with an initial $\boldsymbol{x}^{(0)}$.

III. Application to Blind Deconvolution

Blind deconvolution is typical of the class of problems considered in this paper and we choose it to demonstrate the interest of exploiting the scale invariance property. The objective of image blind deconvolution [2], [8], [9] is to recover the object of interest, say $\boldsymbol{x}$, and the instrumental *point spread function* (PSF), say $\boldsymbol{y}$, given a single observed image, the data $\boldsymbol{d}$, of the object, this image being degraded by the blurring due to the PSF and by the noise. Such a problem is challenging but has a great interest in situations where the PSF is unknown or expensive to calibrate (*e.g.*, astronomy, medical imaging, *etc.*).

A. Statement of the problem

Assuming that the noise statistics is well approximated by a multi-variate Gaussian distribution, an excellent way to measure the distance between the model $\boldsymbol{x} \otimes \boldsymbol{y}$ and the data $\boldsymbol{d}$ is provided by the co-log-likelihood of the data:

$$\mathcal{G}(\boldsymbol{x} \otimes \boldsymbol{y}) = \|\boldsymbol{d} - \boldsymbol{x} \otimes \boldsymbol{y}\|_{\mathbf{W}}^2\,, \qquad (15)$$

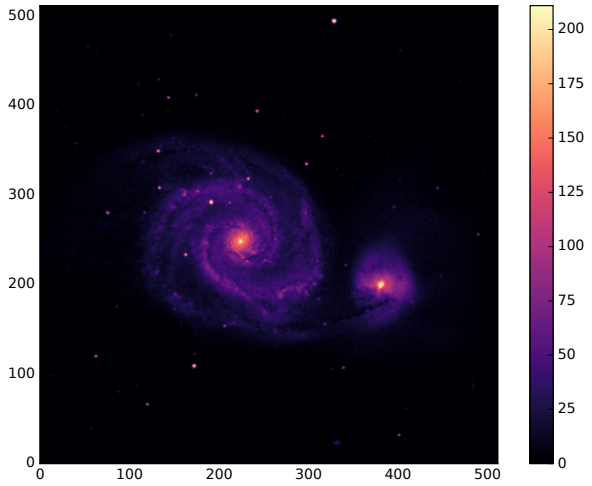

Fig. 2. The true object (the galaxy M51).

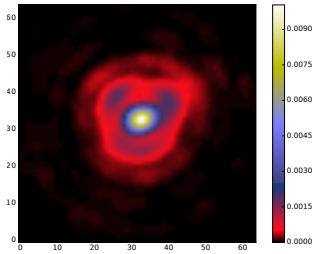

Fig. 3. The instrumental PSF (courtesy François Roddier).

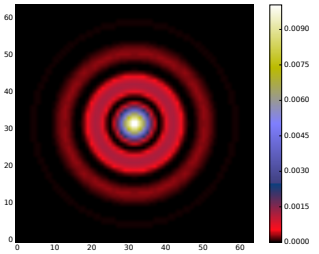

Fig. 4. Initial PSF $\boldsymbol{y}^{(0)}$ assumed in Algorithm 1.

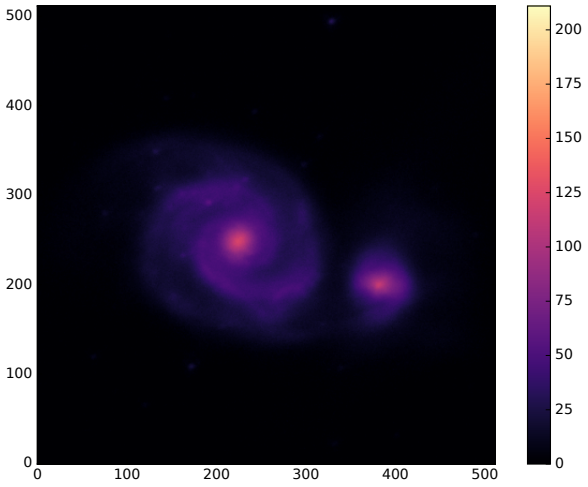

Fig. 5. The data.

where $\otimes$ denotes discrete convolution product, implemented by means of the *fast Fourier transform* (FFT), $\mathbf{W}$ is the precision matrix of the data and $\|\boldsymbol{u}\|^2_{\mathbf{W}} \overset{\text{def}}{=} \boldsymbol{u}^{\mathsf{T}} \cdot \mathbf{W} \cdot \boldsymbol{u}$ is a Mahalanobis squared norm. In practice, we considered independent noise so that the precision matrix $\mathbf{W}$ is diagonal with diagonal terms given by:

$$W_{j,j} = \begin{cases} 1/\mathrm{Var}(d_j) & \text{if } j^{\text{th}} \text{ data pixel is valid,} \\ 0 & \text{otherwise.} \end{cases} \tag{16}$$

This definition provides simple means to consistently account for defective pixels.

As strict constraints such as positivity are beneficial to solve the blind deconvolution problem of images [9], we impose that both the object and the PSF be nonnegative everywhere (for physical reasons, this is always true for the PSF in optics). This means that $\mathbb{X}$ and $\mathbb{Y}$ are equal to $\mathbb{R}^n_+$ with n the number of pixels.

For this paper, we consider the following two regularizations:

$$\mathcal{H}_2(\boldsymbol{u}) = \sum\nolimits_j \|\mathbf{D}_j \cdot \boldsymbol{u}\|^2 , \tag{17}$$

$$\mathcal{H}_1(\boldsymbol{u}) = \sum\nolimits_j \sqrt{\|\mathbf{D}_j \cdot \boldsymbol{u}\|^2 + \varepsilon \, \|\boldsymbol{u}\|^2} , \tag{18}$$

where $\varepsilon \geq 0$ is another tuning parameter and $\mathbf{D}_j \colon \mathbb{U} \to \mathbb{R}^p$ yields a discrete approximation of the p-dimensional gradient of its argument ($p = 2$ for images) at j-th position. These regularizations are homogeneous functions of respective degrees 2 and 1, they both favor the smoothness of the variables $\boldsymbol{u}$. The second regularization, $\mathcal{H}_1(\boldsymbol{u})$, implements *total variation* (TV) regularization [6] for $\varepsilon = 0$ and a homogeneous version of *edge-preserving smoothness* [7] for $\varepsilon > 0$.

B. Results

We generated synthetic data given an image of the galaxy M51 (Fig. 2) and the PSF of the 3.60 m Canada-France Hawaii telescope equipped with one of the first adaptive optics systems (Fig. 3). The resulting data, shown in Fig. 5, is the true object blurred by the PSF and with additional i.i.d. random Gaussian noise of standard deviation $\sigma = 1.5\,\mathrm{ADU}$ which corresponds to a PSNR of $\simeq 80$.

For the initial guess of the PSF $\boldsymbol{y}^{(0)}$, we took the instrumental response of a perfect 3.60 m telescope, that is an Airy function (Fig. 4).

For the regularizations, we first considered the same quadratic regularization, given in Eq. (17), for the object $\boldsymbol{x}$ and the PSF $\boldsymbol{y}$. As the PSF is rather smooth (as expected from the physics) while the object has sharp features (stars), we also considered a homogeneous edge-preserving regularization for the object while keeping a quadratic smoothness for the PSF: $\mathcal{J}(\boldsymbol{x}) = \mathcal{H}_1(\boldsymbol{x})$ and $\mathcal{J}(\boldsymbol{y}) = \mathcal{H}_2(\boldsymbol{y})$.

As the considered objective functions are differentiable (for $\varepsilon > 0$), we used VMLM-B [8], a limited memory quasi-Newton method implementing bound constraints, to solve the sub-problems (14a) and (14b) on the nonnegative orthant. To ensure convergence, the alternating strategy implemented by Algorithm 1 is applied for $m = 200$ iterations, while each of the sub-problems is solved by 10 iterations of VMLM-B.

We used the SURE criterion [1], an unbiased estimation of the mean squared error, to determine the regularization level. We fixed ν and run Algorithm 1 for a wide range of values of μ to produce the curves shown by Fig. 6. A global minimum of SURE criterion is obtained for $\mu \simeq 0.05$ when $\boldsymbol{x}$ and $\boldsymbol{y}$ have the same quadratic regularizations and for $\mu \simeq 0.02$ when edge-preserving regularization is applied to $\boldsymbol{x}$. Figures 9 and 7 show the recovered instrumental PSF and object both with

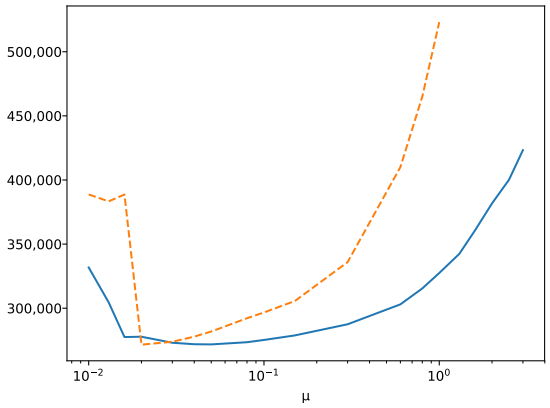

Fig. 6. The SURE criterion $\mathcal{R}(\mu)$ as a function of μ only (*i.e.*, for ν optimally set), for quadratic regularization for both the object and the PSF (blue line), and for edge-preserving and quadratic regularizations for the object and the PSF respectively and a given value of ε (orange dashed line).

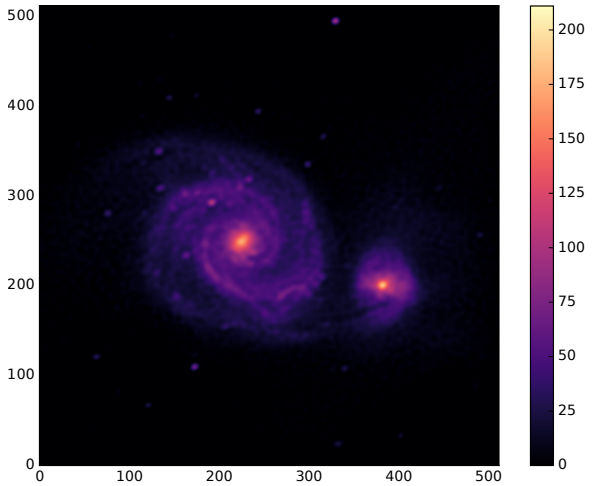

Fig. 7. The recovered object with a quadratic smoothness regularization.

quadratic smoothness priors and for the optimal value $\mu = 0.05$. Figures 10 and 8 show the instrumental PSF and object recovered when imposing quadratic smoothness to the PSF and edge-preserving smoothness to the object for the optimal value $\mu = 0.02$ in that case. Clearly, the value of the hyper-parameter μ found by the proposed procedure is correct in the two considered cases. Besides, imposing more suitable edge-preserving smoothness is beneficial for the object in particular for the recovering of sharp features such as the foreground stars and the cores of the two galaxies.

IV. Conclusions

We have demonstrated that the scaling indetermination of bi-linear models in inverse problems can, theoretically and for homogeneous regularizations, be exploited to simplify the tuning of the regularization (as one less tuning parameter is necessary) and to speed-up convergence. We have proposed a simple alternating strategy to solve the inverse problem in this context and whatever the scaling of the initialization. We have applied our algorithm to the blind deconvolution of astronomical images under various regularizations whose importance is automatically tuned by minimizing Stein's

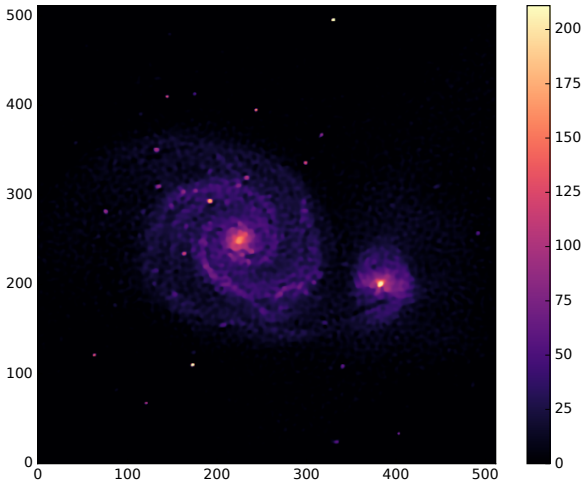

Fig. 8. The recovered object with an homogeneous edge-preserving smoothness regularization.

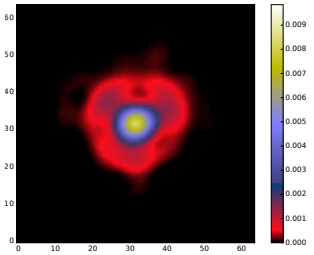

Fig. 9. The recovered PSF with quadratic smoothness regularizations on the PSF and on the object.

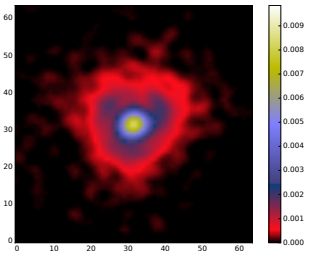

Fig. 10. The recovered PSF with a quadratic smoothness regularization while imposing edge preserving smoothness for the object.

unbiased estimator of the mean weighted error. On the basis of simulated images, we have shown that the proposed algorithm was able to recover good estimates of the object and the instrumental point spread function (PSF) given a single blurred and noisy image.

References

[1] C. M. Stein, "Estimation of the mean of a multivariate normal distribution," *The Annals of Statistics*, vol. 9, no. 6, p. 1135–1151, Nov 1981.

[2] G. R. Ayers and J. C. Dainty, "Iterative blind deconvolution and its applications," *Opt. Lett.*, vol. 13, no. 7, pp. 547–549, July 1988.

[3] P. Comon and C. Jutten, *Handbook of Blind Source Separation: Independent component analysis and applications*. Academic press, 2010.

[4] F. Soulez and M. Unser, "Superresolution with optically-motivated blind deconvolution," in *Laser Applications to Chemical, Security and Environmental Analysis*. Optical Society of America, 2016, pp. JT3A–38.

[5] A. Tarantola, *Inverse Problem Theory and Methods for Model Parameter Estimation*. SIAM, 2005.

[6] L. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Physica D*, vol. 60, pp. 259–268, 1992.

[7] P. Charbonnier, L. Blanc-Féraud, G. Aubert, and M. Barlaud, "Deterministic edge-preserving regularization in computed imaging," vol. 6, no. 2, pp. 298–311, feb 1997.

[8] É. Thiébaut, "Optimization issues in blind deconvolution algorithms," in *Astronomical Data Analysis II*, J.-L. Starck and F. D. Murtagh, Eds., vol. 4847. Bellingham, Washington: SPIE, 2002, pp. 174–183.

[9] É. Thiébaut and J.-M. Conan, "Strict a priori constraints for maximum likelihood blind deconvolution," vol. 12, no. 3, pp. 485–492, March 1995.