

A SEMI-BLIND REGULARIZATION ALGORITHM FOR INVERSE PROBLEMS WITH APPLICATION TO IMAGE DEBLURRING*

ALESSANDRO BUCCINI[†], MARCO DONATELLI[‡], AND RONNY RAMLAU[§]

Abstract. In many inverse problems the operator to be inverted depends on parameters which are not known precisely. In this article we propose a functional that involves as variables both the solution of the problem and the parameters on which the operator depends. We first prove that the functional, even if it is non-convex, admits a global minimum and that its minimization naturally leads to a regularization method. Then, using the popular Alternating Direction Multiplier Method (ADMM), we describe an algorithm to identify a stationary point of the functional. The introduction of the ADMM algorithm lets us easily introduce some constraints on the reconstructions like non-negativity and flux conservation. Since the functional is non-convex a proof of convergence of the method is given. Numerical examples prove the validity of the proposed approach.

Key words. Noisy operator, Regularization of ill-posed problems, Non-convex optimization, Semi-blind deconvolution

AMS subject classifications. 65N20, 90C26, 65J20

1. Introduction. Many applications in various scientific fields lead to inverse problems modeled as an equation of the form

$$(1) \quad B(k, f) = g,$$

where f is the desired solution, g is the measured data, and k is a variable on which the operator B depends, e.g., if $B(k, f) = k * f$ is the convolution operator then k represents the integral kernel. We consider the case in which the problem (1) is ill-posed. As happens in practical applications, we assume that is impossible to get the true data g but only a noisy version g_δ is available, such that $\|g - g_\delta\| \leq \delta$, where $\|\cdot\|$ denotes the Euclidean norm. The direct solution of the associated discrete system, usually leads to very poor reconstruction of the original f , thus arises the need for regularization methods. In many situations also k is not precisely known, but only a noisy measurement k_ϵ is available, such that $\|k - k_\epsilon\| \leq \epsilon$. We will assume that the noise that corrupts both g and k is Gaussian. Therefore, equation (1) becomes

$$B(k_\epsilon, f) = g_\delta.$$

We refer to the regularization of such a problem as *semi-blind*, since the variable k , even though is not completely unknown, has a certain degree of uncertainty on it. Blind and semi-blind deconvolution has been largely investigated [1, 2, 3, 6, 7, 8, 12, 15, 16, 18, 24, 29, 30]. The approach in [3, 16] requires that the blurring operator is diagonalized by fast transforms, which is not assumed in this paper. In [6] a double regularization approach to recover f and k was proposed, which consisted in solving

$$(2) \quad \left(k_{\alpha, \beta}^{\delta, \epsilon}, f_{\alpha, \beta}^{\delta, \epsilon} \right) = \arg \min_{k, f} J(k, f)$$

where $J(k, f) = \frac{1}{2}T(k, f) + R(k, f)$, T had the role of data-fitting term and R was the penalty term. In particular

$$\begin{aligned} T(k, f) &= \|B(k, f) - g_\delta\|^2 + \gamma \|k - k_\epsilon\|^2, \\ R(k, f) &= \alpha \|Lf\| + \beta \mathcal{R}(k), \end{aligned}$$

where $\mathcal{R}(k)$ is an appropriate penalty term and L is a regularization operator which was supposed to be bounded and continuously invertible. A numerical technique to solve (2) was proposed by the same authors in [7] using an alternating minimization over f and k .

In this work we want to improve the results in [6, 7] by introducing a more complex penalty term for f and by employing a more sophisticated numerical technique for the minimization. In particular we want to

*Submitted to the editors 2nd November 2016

Funding: The work of the first two authors is supported in part by MIUR - PRIN 2012 N. 2012MTE38N and by a grant of the group GNCS of INdAM.

[†]University of Insubria, Como, Italy (alessandro.buccini@uninsubria.it).

[‡]University of Insubria, Como, Italy (marco.donatelli@uninsubria.it).

[§]Johannes Kepler University Linz, Linz, Austria (ronny.ramlau@jku.at).

introduce for both k and f as a prior the *Total Variation* (TV). The TV norm has the ability to well preserve the edges of the signal while also reconstructing its smooth parts and has proven to well perform in various situations. Moreover, we add some constraints like the flux conservation and nonnegativity.

We would like to stress the fact that, while here we are only considering the TV norm as a prior for k , the extension to a general convex prior $\mathcal{R}(k)$ is straightforward.

The introduction of the TV and of the constraints makes the minimization of the functional we are going to consider more complicated and so the introduction of advanced numerical techniques is needed.

For the numerical solution of the minimization we are going to use the *Alternating Direction Multiplier Method* (ADMM) [9]. The ADMM algorithm was developed for solving constrained convex optimization problem, in recent years this method has been successfully applied to non-convex minimization, see, e.g., [25, 27]. However, the convergence of the algorithm is not proven in the general non-convex setting and needs to be shown for each particular problem.

We thus prove that the resulting numerical method is convergent and we give some numerical evidences of the improvement in the quality of the reconstructed images with respect to the proposal in [7], to the basic method using the measured PSF for the deconvolution, and to some methods in the literature.

To the best of our knowledge an approach like the one described here has never been proposed. In particular, most of the blind and semi-blind methods tackle the problem by considering the variables k and f separately, see, e.g., [1], while here we propose a model that consider the two unknowns jointly. Moreover, the convergence of the ADMM algorithm in this setup has never been proved. In particular, our approach differs from the other ones in the literature because the variables are not decoupled. Moreover, we do not assume that the gradients of the functional that is to be minimized are Lipschitz continuous.

This work is structured as follows: Section 2 describes the functional we want to consider and analyze some of its properties. In Section 3 we describe the algorithm for the minimization of the functional previously introduced and we prove its convergence, while in Section 4 we discuss the addition of some constraints. Section 5 is devoted to numerical examples in image deblurring. In Section 6 we draw some conclusions. Finally in Appendix A.1 we give all the detailed proofs of the results stated in the paper and in Appendix B we describe how to use the ADMM algorithm for multiple constrained minimization.

2. The regularized functional. As previously discussed, our goal is to extend the previous results from [6] using a more complex penalty term for both f and k . Let $g_\delta, k_\epsilon, f, k \in H^1$, where H^1 denotes the Sobolev space $W^{1,2}$ which is a separable Hilbert space, see, e.g., [10, Section 9.1]. We consider the minimization of the following functional

$$(3) \quad J_{\alpha,\beta}^{\delta,\epsilon}(k, f) = \|B(k, f) - g_\delta\|^2 + \alpha^{\text{TV}} \|f\|_{\text{TV}} + \alpha^{\text{E}} \|f\|^2 + \gamma \|k - k_\epsilon\|^2 + \beta \|k\|_{\text{TV}},$$

where

$$\|h\|_{\text{TV}} = \int \|\nabla h\|$$

is the total variation norm of $h \in H^1$. For simplicity of notation we denote by α the couple $(\alpha^{\text{E}}, \alpha^{\text{TV}})$.

We now show some theoretical properties of (3), in particular we want to show the existence of a global minimizer, the stability of the minima, and the fact that, if α and β are properly chosen, in relation to the noise, the minimization of $J_{\alpha,\beta}^{\delta,\epsilon}$ induces a regularization method.

ASSUMPTION 1. *The operator B is strongly continuous on its domain.*

Before proving the existence of the minimizer we first need to show some properties of $J_{\alpha,\beta}^{\delta,\epsilon}(f, k)$.

For the convenience of the reader we postpone all the proofs in Appendix A.1.

LEMMA 1. *The functional $J_{\alpha,\beta}^{\delta,\epsilon}(f, k)$ defined in (3) is positive, weakly lower semi-continuous (wlsc) and coercive with respect to the norm $\|(k, f)\|^2 := \|k\|^2 + \|f\|^2$.*

We are now able to state the existence of a global minimizer.

THEOREM 2 (Existence). *The functional $J_{\alpha,\beta}^{\delta,\epsilon}(f, k)$ defined in (3) has a global minimizer.*

We are now in the position to state the stability property

THEOREM 3 (Stability). *With the same notation as above, let $\alpha^{\text{E}}, \alpha^{\text{TV}}, \beta$, and γ be fixed. Let $(g_{\delta_j})_j$ and $(k_{\epsilon_j})_j$ be sequences such that $g_{\delta_j} \rightarrow g_\delta$ and $k_{\epsilon_j} \rightarrow k_\epsilon$, let (k_j, f_j) be minimizers obtained with data $g_{\delta_j}, k_{\epsilon_j}$. Then there exists a convergent subsequence of (k_j, f_j) and the limit of every subsequence is a minimizer of $J_{\alpha,\beta}^{\delta,\epsilon}$.*

One of the most important properties of regularization methods is the regularization property. We want that in the ideal case, i.e., when there is no noise, we can exactly recover a true solution of the problem and in particular that with minimum norm. Moreover, we want that, as the norm of the noise goes to 0, the corresponding reconstructions converge to the minimum norm solution of the problem as well. This intuitively means that, if there is not too much noise and if the parameters α^E , α^{TV} , and β are set accordingly, we can trust our method to give good approximation to the true solution.

We now define what is the *minimum norm solution* in this setup.

DEFINITION 4. *The minimum norm solution of $B(k_0, f_0) = g_0$ is*

$$f^\dagger = \arg \min_{f \in H^1} \{ \|f\|^2 + \|f\|_{TV} : B(k_0, f) = g_0 \}.$$

We would like to stress the fact that the quantity $\|f\|^2 + \|f\|_{TV}$ is strictly related to the norm of the Sobolev space H^1 which is defined as $\|f\|_{H^1}^2 = \int |f|^2 + |\nabla f|^2$.

We now state that, if α^E , α^{TV} , and β are chosen properly and dependently from the noise, the method proposed is a regularization method.

THEOREM 5 (Regularization property). *Let $(g_{\delta_j})_j$ and $(k_{\epsilon_j})_j$ be sequences such that*

$$\|g_{\delta_j} - g_0\| < \delta_j \text{ and } \|k_{\epsilon_j} - k_0\| < \epsilon_j$$

and such that $\delta_j, \epsilon_j \rightarrow 0$ as $j \rightarrow \infty$. Let α_j^E , α_j^{TV} , and β_j be sequences such that $\alpha_j^E, \alpha_j^{TV}, \beta_j \rightarrow 0$ as $j \rightarrow \infty$, moreover, assume that it holds

$$\lim_{j \rightarrow \infty} \frac{\delta_j^2 + \gamma \epsilon_j^2}{\alpha_j^E} = 0, \quad \lim_{j \rightarrow \infty} \frac{\alpha_j^{TV}}{\alpha_j^E} = 1, \quad \text{and} \quad \lim_{j \rightarrow \infty} \frac{\beta_j}{\alpha_j^E} = \eta \quad 0 < \eta < \infty.$$

Call $(k_j, f_j) := \left(k_{\alpha_j^E, \beta_j}, f_{\alpha_j^E, \beta_j} \right)$, then there exists a convergent subsequence of (k_j, f_j) such that $k_j \rightarrow k_0$ and the limit of every convergent subsequence of f_j is the minimum norm solution of $B(k_0, f) = g_0$.

3. Minimization Algorithm. Computing a minimum of (3), since it is non-convex and non-smooth. Moreover, the minimization over two variables by itself can be very challenging. The strategy proposed in [7] was an alternating minimization, where at each step one variable was fixed and the functional was minimized in respect to the other free variable. With this method it is possible to avoid the complicated minimization over two arguments, but still the minimization over only one variable can be tough. The method we are going to propose decouples the various terms inside (3) obtaining a series of simple minimization problems.

We now consider the finite dimensional case, in particular for simplicity we assume $\mathbf{f}, \mathbf{k} \in \mathbb{R}^{n \times n}$ (the extension to the case where \mathbf{f} and \mathbf{k} belong to different spaces and are not square is straightforward).

In this section we denote by \mathbf{x}_i the i th component of the vector \mathbf{x} and by $\mathbf{x}^{(j)}$ the j th iteration generated by some iterative algorithm.

Since in many application it is possible to know that the exact solution of the problem lies in some closed and convex set, it is then beneficial to constrain the computed approximation to lie in that set as well, i.e., constraining the minimization of $J_{\alpha, \beta}^{\delta, \epsilon}$. We consider the following constrained minimization problem

$$(4) \quad (\mathbf{k}^*, \mathbf{f}^*) = \arg \min_{\mathbf{k} \in \Omega_{\mathbf{k}}, \mathbf{f} \in \Omega_{\mathbf{f}}} J_{\alpha, \beta}^{\delta, \epsilon}(\mathbf{k}, \mathbf{f}),$$

where, with abuse of notation, we call $J_{\alpha, \beta}^{\delta, \epsilon}$ the discretization of the function defined in (3) and $\Omega_{\mathbf{k}}, \Omega_{\mathbf{f}} \subset \mathbb{R}^{n \times n}$ are closed and convex sets.

For computing a minimum, or at least a critical point, of $J_{\alpha, \beta}^{\delta, \epsilon}$ we use the ADMM algorithm. However, note that the classic theory of ADMM does not assure the convergence of the method in this case since $J_{\alpha, \beta}^{\delta, \epsilon}$ is non-convex, see, e.g., [9]. Thus, we are going to need a different result for its convergence. In particular, we are going to need a further assumption on $\Omega_{\mathbf{k}} \times \Omega_{\mathbf{f}}$.

Before covering the convergence property of the proposed algorithm we explicitly formulate all its ingredients.

The *isotropic Total Variation* operator in this space is defined as follows. Let $\mathbf{x} \in \mathbb{R}^{n^2}$ be the vectorization of a bidimensional array in $\mathbb{R}^{n \times n}$ obtained by stacking its columns, then

$$TV(\mathbf{x}) = \sum_{i=1}^{n^2} \|D_i \mathbf{x}\|,$$

where $D_i \mathbf{x} = ((D^{(1)} \mathbf{x})_i, (D^{(2)} \mathbf{x})_i)^t \in \mathbb{R}^2$ is

$$(5) \quad \begin{aligned} (D^{(1)} \mathbf{x})_i &= \begin{cases} \mathbf{x}_{i+n} - \mathbf{x}_i, & \text{if } 1 \leq i \leq n(n-1) \\ \mathbf{x}_{\text{mod}(i,n)} - \mathbf{x}_i, & \text{otherwise} \end{cases} \\ (D^{(2)} \mathbf{x})_i &= \begin{cases} \mathbf{x}_{i+1} - \mathbf{x}_i, & \text{if } \text{mod}(i,n) \neq 0 \\ \mathbf{x}_{i-n+1} - \mathbf{x}_i, & \text{otherwise.} \end{cases} \end{aligned}$$

With abuse of notation we will write $\|\mathbf{x}\|_{TV} = TV(\mathbf{x})$, also in the finite dimensional case.

Let us rewrite (4) in an equivalent way

$$\begin{aligned} (\mathbf{k}^*, \mathbf{f}^*) &= \arg \min_{\mathbf{k} \in \Omega_{\mathbf{k}}, \mathbf{f} \in \Omega_{\mathbf{f}}} \|B(\mathbf{k}, \mathbf{f}) - \mathbf{g}_\delta\|^2 + \alpha^E \|\mathbf{f}\|^2 + \alpha^{TV} \|\mathbf{f}\|_{TV} + \gamma \|\mathbf{k} - \mathbf{k}_\epsilon\|^2 + \beta \|\mathbf{k}\|_{TV} \\ &= \arg \min_{\substack{\tilde{\mathbf{k}} \in \Omega_{\tilde{\mathbf{k}}}, \tilde{\mathbf{f}} \in \Omega_{\tilde{\mathbf{f}}} \\ \hat{\mathbf{k}}, \hat{\mathbf{f}}, \mathbf{k}, \mathbf{f}}} \left\{ \|B(\mathbf{k}, \mathbf{f}) - \mathbf{g}_\delta\|^2 + \alpha^E \|\mathbf{f}\|^2 + \alpha^{TV} \|\hat{\mathbf{f}}\|_{TV} + \gamma \|\mathbf{k} - \mathbf{k}_\epsilon\|^2 + \beta \|\hat{\mathbf{k}}\|_{TV}, \right. \\ &\quad \left. \mathbf{k} = \tilde{\mathbf{k}}, \mathbf{f} = \tilde{\mathbf{f}}, \mathbf{k} = \hat{\mathbf{k}}, \mathbf{f} = \hat{\mathbf{f}} \right\}, \end{aligned}$$

where $\mathbf{f}, \tilde{\mathbf{f}}, \hat{\mathbf{f}}, \mathbf{k}, \tilde{\mathbf{k}}, \hat{\mathbf{k}} \in \mathbb{R}^N$ with $N = n^2$ and, with abuse of notation, $B(\mathbf{k}, \mathbf{f})$ denotes the finite dimensional version of $B(k, f)$.

We now write the Augmented Lagrangian of the minimization above

$$(6) \quad \begin{aligned} \mathcal{L}_A(\tilde{\mathbf{f}}, \hat{\mathbf{f}}, \mathbf{f}, \tilde{\mathbf{k}}, \hat{\mathbf{k}}, \mathbf{k}; \boldsymbol{\lambda}, \boldsymbol{\xi}, \boldsymbol{\zeta}, \boldsymbol{\mu}) &= \|B(\mathbf{k}, \mathbf{f}) - \mathbf{g}_\delta\|^2 + \alpha^E \|\mathbf{f}\|^2 + \alpha^{TV} \|\hat{\mathbf{f}}\|_{TV} + \gamma \|\mathbf{k} - \mathbf{k}_\epsilon\|^2 + \beta \|\hat{\mathbf{k}}\|_{TV} \\ &+ \frac{\omega_1}{2} \|\tilde{\mathbf{f}} - \mathbf{f}\|^2 - \langle \boldsymbol{\lambda}, \tilde{\mathbf{f}} - \mathbf{f} \rangle + \frac{\omega_2}{2} \|\hat{\mathbf{f}} - \mathbf{f}\|^2 - \langle \boldsymbol{\xi}, \hat{\mathbf{f}} - \mathbf{f} \rangle \\ &+ \frac{\omega_3}{2} \|\tilde{\mathbf{k}} - \mathbf{k}\|^2 - \langle \boldsymbol{\zeta}, \tilde{\mathbf{k}} - \mathbf{k} \rangle + \frac{\omega_4}{2} \|\hat{\mathbf{k}} - \mathbf{k}\|^2 - \langle \boldsymbol{\mu}, \hat{\mathbf{k}} - \mathbf{k} \rangle, \end{aligned}$$

where $\boldsymbol{\lambda}, \boldsymbol{\zeta}, \boldsymbol{\xi}, \boldsymbol{\mu} \in \mathbb{R}^N$.

We apply the ADMM algorithm on (6) obtaining Algorithm 1.

We call this method SeB-A as for *Semi-blind ADMM*.

We can observe that, due to the fact that the variables inside $J_{\alpha, \beta}^{\delta, \epsilon}$ are coupled, we cannot apply the classical ADMM technique of first solving the minimization with respect to the auxiliary variables and then the minimization with respect to the primary variables, but we need to implement an hybrid approach. The variables are still divided in two blocks, but each block contains both primary and auxiliary variables. While this seems counterintuitive it well works from both the theoretical and numerical point of view.

We now formulate some assumptions that we are going to need in the following.

ASSUMPTION 2. $B(\mathbf{k}, \mathbf{f})$ is bilinear.

REMARK 6. The semi-blind deconvolution problem we are interested in satisfies Assumption 2.

Under Assumption 2 most of the minimization in Algorithm 1 are easily computed. We have that

Algorithm 1 SeB-A

Given $\mathbf{f}^{(0)}$, $\mathbf{k}^{(0)}$, $\boldsymbol{\lambda}^{(0)}$, $\boldsymbol{\xi}^{(0)}$, $\boldsymbol{\zeta}^{(0)}$, and $\boldsymbol{\mu}^{(0)}$ initial guesses for \mathbf{f} , \mathbf{k} , $\boldsymbol{\lambda}$, $\boldsymbol{\xi}$, $\boldsymbol{\zeta}$, and $\boldsymbol{\mu}$, respectively. Let $\omega_1, \omega_2, \omega_3, \omega_4 > 0$ be real fixed numbers.

```

for  $j = 0, 1, \dots$ 
   $\begin{pmatrix} \tilde{\mathbf{f}}^{(j+1)} \\ \hat{\mathbf{f}}^{(j+1)} \\ \mathbf{k}^{(j+1)} \end{pmatrix} = \arg \min_{\tilde{\mathbf{f}}, \hat{\mathbf{f}}, \mathbf{k}} \mathcal{L}_A \left( \tilde{\mathbf{f}}, \hat{\mathbf{f}}, \mathbf{k} \mid \tilde{\mathbf{k}}^{(j)}, \hat{\mathbf{k}}^{(j)}, \mathbf{f}^{(j)}; \boldsymbol{\lambda}^{(j)}, \boldsymbol{\xi}^{(j)}, \boldsymbol{\zeta}^{(j)}, \boldsymbol{\mu}^{(j)} \right)$ 
   $\begin{pmatrix} \tilde{\mathbf{k}}^{(j+1)} \\ \hat{\mathbf{k}}^{(j+1)} \\ \mathbf{f}^{(j+1)} \end{pmatrix} = \arg \min_{\tilde{\mathbf{k}}, \hat{\mathbf{k}}, \mathbf{f}} \mathcal{L}_A \left( \tilde{\mathbf{k}}, \hat{\mathbf{k}}, \mathbf{f} \mid \tilde{\mathbf{f}}^{(j+1)}, \hat{\mathbf{f}}^{(j+1)}, \mathbf{k}^{(j+1)}; \boldsymbol{\lambda}^{(j)}, \boldsymbol{\xi}^{(j)}, \boldsymbol{\zeta}^{(j)}, \boldsymbol{\mu}^{(j)} \right)$ 
   $\boldsymbol{\lambda}^{(j+1)} = \boldsymbol{\lambda}^{(j)} - \omega_1 (\tilde{\mathbf{f}}^{(j+1)} - \mathbf{f}^{(j+1)})$ 
   $\boldsymbol{\xi}^{(j+1)} = \boldsymbol{\xi}^{(j)} - \omega_2 (\hat{\mathbf{f}}^{(j+1)} - \mathbf{f}^{(j+1)})$ 
   $\boldsymbol{\zeta}^{(j+1)} = \boldsymbol{\zeta}^{(j)} - \omega_3 (\tilde{\mathbf{k}}^{(j+1)} - \mathbf{k}^{(j+1)})$ 
   $\boldsymbol{\mu}^{(j+1)} = \boldsymbol{\mu}^{(j)} - \omega_4 (\hat{\mathbf{k}}^{(j+1)} - \mathbf{k}^{(j+1)})$ 
end

```

$$\tilde{\mathbf{f}}^{(j+1)} = P_{\Omega_{\mathbf{f}}} \left(\mathbf{f}^{(j)} + \frac{\boldsymbol{\lambda}^{(j)}}{\omega_1} \right)$$

$$\mathbf{k}^{(j+1)} = (2A_{\mathbf{f}^{(j)}}^* A_{\mathbf{f}^{(j)}} + 2\gamma I + (\omega_3 + \omega_4) I)^{-1} \left(2A_{\mathbf{f}^{(j)}}^* \mathbf{g}_\delta + 2\gamma \mathbf{k}_\epsilon - \boldsymbol{\zeta}^{(j)} + \omega_3 \tilde{\mathbf{k}}^{(j)} - \boldsymbol{\mu}^{(j)} + \omega_4 \hat{\mathbf{k}}^{(j)} \right)$$

$$\tilde{\mathbf{k}}^{(j+1)} = P_{\Omega_{\mathbf{k}}} \left(\mathbf{k}^{(j+1)} + \frac{\boldsymbol{\zeta}^{(j)}}{\omega_3} \right)$$

$$\mathbf{f}^{(j+1)} = (2A_{\mathbf{k}^{(j+1)}}^* A_{\mathbf{k}^{(j+1)}} + 2\alpha^E I + (\omega_1 + \omega_2) I)^{-1} \left(2A_{\mathbf{k}^{(j+1)}}^* \mathbf{g}_\delta - \boldsymbol{\lambda}^{(j)} + \omega_1 \tilde{\mathbf{f}}^{(j+1)} - \boldsymbol{\xi}^{(j)} + \omega_2 \hat{\mathbf{f}}^{(j+1)} \right)$$

Where by A_\bullet we indicate the linear operator obtained from $B(\mathbf{k}, \mathbf{f})$ by fixing \bullet .

On the other hand the minimizations with respect of $\hat{\mathbf{f}}$ and $\hat{\mathbf{k}}$ are non trivial, in fact

$$\begin{aligned} \hat{\mathbf{f}}^{(j+1)} &= \arg \min_{\hat{\mathbf{f}}} \alpha^{\text{TV}} \left\| \hat{\mathbf{f}} \right\|_{\text{TV}} + \frac{\omega_2}{2} \left\| \hat{\mathbf{f}} - \mathbf{f}^{(j)} \right\|^2 - \left\langle \boldsymbol{\xi}^{(j)}, \hat{\mathbf{f}} - \mathbf{f}^{(j)} \right\rangle \\ &= \arg \min_{\hat{\mathbf{f}}} \left\| \hat{\mathbf{f}} \right\|_{\text{TV}} + \frac{\omega_2}{2\alpha^{\text{TV}}} \left\| \hat{\mathbf{f}} - \left(\mathbf{f}^{(j)} + \frac{\boldsymbol{\xi}^{(j)}}{\omega_2} \right) \right\|^2 \\ \hat{\mathbf{k}}^{(j+1)} &= \arg \min_{\hat{\mathbf{k}}} \beta \left\| \hat{\mathbf{k}} \right\|_{\text{TV}} + \frac{\omega_4}{2} \left\| \hat{\mathbf{k}} - \mathbf{k}^{(j+1)} \right\|^2 - \left\langle \boldsymbol{\mu}^{(j)}, \hat{\mathbf{k}} - \mathbf{k}^{(j+1)} \right\rangle \\ &= \arg \min_{\hat{\mathbf{k}}} \left\| \hat{\mathbf{k}} \right\|_{\text{TV}} + \frac{\omega_4}{2\beta} \left\| \hat{\mathbf{k}} - \left(\mathbf{k}^{(j+1)} + \frac{\boldsymbol{\mu}^{(j)}}{\omega_4} \right) \right\|^2 \end{aligned}$$

To solve this minimization problems we can use, e.g., the ADMM algorithm. Since, however, in this case the functional are proper and convex the convergence is assured by the classical ADMM theory, see, e.g., [9]. For completeness we describe here this approach.

Consider the minimization problem

$$(7) \quad \mathbf{x} = \arg \min_{\mathbf{x}} \|\mathbf{x}\|_{\text{TV}} + c \|\mathbf{x} - \mathbf{y}\|^2,$$

where $c > 0$ is a constant. We can rewrite (7) as

$$\mathbf{x} = \arg \min_{\mathbf{x}, \hat{\mathbf{x}}} \left\{ \sum_{i=1}^N \|\hat{\mathbf{x}}_i\| + c \|\mathbf{x} - \mathbf{y}\|^2, \hat{\mathbf{x}}_i = D_i \mathbf{x} \right\}.$$

The related augmented Lagrangian is

$$\mathcal{L}_A(\mathbf{x}, \hat{\mathbf{x}}; \boldsymbol{\lambda}) = \sum_{i=1}^N \|\hat{\mathbf{x}}_i\| + c \|\mathbf{x} - \mathbf{y}\|^2 + \sum_{i=1}^N \left(\frac{\omega}{2} \|\hat{\mathbf{x}}_i - D_i \mathbf{x}\|^2 - \langle \boldsymbol{\lambda}_i, \hat{\mathbf{x}}_i - D_i \mathbf{x} \rangle \right).$$

The resulting ADMM algorithm is summarized in Algorithm 2.

Algorithm 2 ADMM for TV optimization

Given $\mathbf{x}^{(0)}$ and $\boldsymbol{\lambda}^{(0)}$ initial guesses for \mathbf{x} and $\boldsymbol{\lambda}$, respectively. Let $\omega > 0$ be a real fixed number.

```

for  $j = 0, 1, \dots$ 
   $\hat{\mathbf{x}}^{(j+1)} = \arg \min_{\hat{\mathbf{x}}} \sum_{i=1}^N \left( \|\hat{\mathbf{x}}_i\| + \frac{\omega}{2} \|\hat{\mathbf{x}}_i - D_i \mathbf{x}^{(j)}\|^2 - \langle \boldsymbol{\lambda}_i^{(j)}, \hat{\mathbf{x}}_i - D_i \mathbf{x}^{(j)} \rangle \right)$ 
   $\mathbf{x}^{(j+1)} = \arg \min_{\mathbf{x}} c \|\mathbf{x} - \mathbf{y}\|^2 + \sum_{i=1}^N \left( \frac{\omega}{2} \|\hat{\mathbf{x}}_i^{(j+1)} - D_i \mathbf{x}\|^2 - \langle \boldsymbol{\lambda}_i^{(j)}, \hat{\mathbf{x}}_i^{(j+1)} - D_i \mathbf{x} \rangle \right)$ 
   $\boldsymbol{\lambda}^{(j+1)} = \boldsymbol{\lambda}^{(j)} - \omega (\hat{\mathbf{x}}^{(j+1)} - D\mathbf{x}^{(j+1)})$ 
end

```

The minimization in Algorithm 2 are easily computed. The minimization with respect to $\hat{\mathbf{x}}$ decouples in N subproblems that are easily solved using a two-dimensional shrinkage and the minimization with respect to \mathbf{x} can be achieved by solving a well-conditioned spd linear system. In particular

$$\begin{aligned} \hat{\mathbf{x}}_i^{(j+1)} &= \frac{D_i \mathbf{x}^{(j)} + \frac{\boldsymbol{\lambda}_i^{(j)}}{\omega}}{\left\| D_i \mathbf{x}^{(j)} + \frac{\boldsymbol{\lambda}_i^{(j)}}{\omega} \right\|} \circ \left(\left\| D_i \mathbf{x}^{(j)} + \frac{\boldsymbol{\lambda}_i^{(j)}}{\omega} \right\| - \frac{1}{c\omega} \right)_+ \\ \mathbf{x}^{(j+1)} &= (2cI + \omega D^* D)^{-1} \left(2c\mathbf{y} + \omega D^* \hat{\mathbf{x}}^{(j+1)} - D^* \boldsymbol{\lambda}^{(j)} \right) \end{aligned}$$

In this setup D is the linear operator which maps \mathbb{R}^N into $\mathbb{R}^{N \times 2}$ defined as $\begin{pmatrix} D^{(1)} \\ D^{(2)} \end{pmatrix}$.

We now prove the convergence of Algorithm 1 (SeB-A). This proof is very technical and inspired by [25]. The main difference between our setup and the one in [25] is that in the functional $J_{\alpha, \beta}^{\delta, \epsilon}$ that we want to minimize the variables are coupled inside the fidelity term $\|B(\mathbf{k}, \mathbf{f}) - \mathbf{g}_\delta\|^2$, while in [25] all the variables are decoupled.

REMARK 7. *The convergence of Algorithm 1 is proven only under Assumption 4, i.e., that the norms of the iterates $\mathbf{f}^{(j)}$ and $\mathbf{k}^{(j)}$ remain bounded. While this seems a very strong requirement, from the numerical experiments we can see that this condition is always satisfied. We provide some simple bounds φ_k and φ_f for the norms of $\mathbf{k}^{(j)}$ and $\mathbf{f}^{(j)}$, respectively. In Section 5 we show that this bounds are far from being violated in all computed examples.*

3.1. Proof of convergence. We now give the proof of convergence of Algorithm 1.

We first analyze the unconstrained case, i.e., we consider the minimization problem

$$(\mathbf{k}^*, \mathbf{f}^*) = \arg \min_{\mathbf{k}, \mathbf{f}} \|B(\mathbf{k}, \mathbf{f}) - \mathbf{g}_\delta\|^2 + \alpha^E \|\mathbf{f}\|^2 + \alpha^{\text{TV}} \|\mathbf{f}\|_{\text{TV}} + \gamma \|\mathbf{k} - \mathbf{k}_\epsilon\|^2 + \beta \|\mathbf{k}\|_{\text{TV}},$$

Let us rewrite the minimization above in an equivalent way

$$(8) \quad (\mathbf{k}^*, \mathbf{f}^*) = \arg \min_{\substack{\mathbf{k}, \mathbf{f} \\ \hat{\mathbf{k}}, \hat{\mathbf{f}}}} \left\{ \|B(\mathbf{k}, \mathbf{f}) - \mathbf{g}_\delta\|^2 + \alpha^E \|\mathbf{f}\|^2 + \alpha^{\text{TV}} \|\hat{\mathbf{f}}\|_{\text{TV}} + \gamma \|\mathbf{k} - \mathbf{k}_\epsilon\|^2 + \beta \|\hat{\mathbf{k}}\|_{\text{TV}}, \hat{\mathbf{f}} = \mathbf{f}, \hat{\mathbf{k}} = \mathbf{k} \right\}.$$

We now form the augmented Lagrangian related to the minimization problem (8) where, without loss of generality, we have chosen the same ω for all the augmentation terms

$$(9) \quad \begin{aligned} \mathcal{L}_A(\mathbf{f}, \mathbf{k}, \hat{\mathbf{f}}, \hat{\mathbf{k}}; \boldsymbol{\xi}, \boldsymbol{\mu}) &= \|B(\mathbf{k}, \mathbf{f}) - \mathbf{g}_\delta\|^2 + \alpha^E \|\mathbf{f}\|^2 + \alpha^{\text{TV}} \|\hat{\mathbf{f}}\|_{\text{TV}} + \gamma \|\mathbf{k} - \mathbf{k}_\epsilon\|^2 + \beta \|\hat{\mathbf{k}}\|_{\text{TV}} \\ &+ \frac{\omega}{2} \|\hat{\mathbf{f}} - \mathbf{f}\|^2 - \langle \boldsymbol{\xi}, \hat{\mathbf{f}} - \mathbf{f} \rangle + \frac{\omega}{2} \|\hat{\mathbf{k}} - \mathbf{k}\|^2 - \langle \boldsymbol{\mu}, \hat{\mathbf{k}} - \mathbf{k} \rangle. \end{aligned}$$

Algorithm 3

Given $\mathbf{f}^{(0)}$, $\mathbf{k}^{(0)}$, $\boldsymbol{\xi}^{(0)}$, and $\boldsymbol{\mu}^{(0)}$ initial guesses for \mathbf{f} , \mathbf{k} , $\boldsymbol{\xi}$, and $\boldsymbol{\mu}$, respectively.

```

for  $j = 0, 1, \dots$ 
   $\begin{pmatrix} \hat{\mathbf{f}}^{(j+1)} \\ \mathbf{k}^{(j+1)} \end{pmatrix} = \arg \min_{\hat{\mathbf{f}}, \mathbf{k}} \mathcal{L}_A \left( \hat{\mathbf{f}}, \mathbf{k} \mid \hat{\mathbf{k}}^{(j)}, \mathbf{f}^{(j)}; \boldsymbol{\xi}^{(j)}, \boldsymbol{\mu}^{(j)} \right)$ 
   $\begin{pmatrix} \hat{\mathbf{k}}^{(j+1)} \\ \mathbf{f}^{(j+1)} \end{pmatrix} = \arg \min_{\hat{\mathbf{k}}, \mathbf{f}} \mathcal{L}_A \left( \hat{\mathbf{k}}, \mathbf{f} \mid \hat{\mathbf{f}}^{(j+1)}, \mathbf{k}^{(j+1)}; \boldsymbol{\xi}^{(j)}, \boldsymbol{\mu}^{(j)} \right)$ 
   $\boldsymbol{\xi}^{(j+1)} = \boldsymbol{\xi}^{(j)} - \omega \left( \hat{\mathbf{f}}^{(j+1)} - \mathbf{f}^{(j+1)} \right)$ 
   $\boldsymbol{\mu}^{(j+1)} = \boldsymbol{\mu}^{(j)} - \omega \left( \hat{\mathbf{k}}^{(j+1)} - \mathbf{k}^{(j+1)} \right)$ 
end

```

The unconstrained method is described in Algorithm 3. This simplified version of our method is an algorithm to compute the solution of an unconstrained minimization problem. In our proofs we are not going to consider the constraint $(\mathbf{k}, \mathbf{f}) \in \Omega_{\mathbf{k}} \times \Omega_{\mathbf{f}}$. We are going to insert this constraint only at the very end.

To proceed we need some further assumptions

ASSUMPTION 3. *We assume that*

- (i) *If $\mathbf{k} = \mathbf{0}$ or $\mathbf{f} = \mathbf{0}$ then $B(\mathbf{k}, \mathbf{f}) = \mathbf{0}$;*
- (ii) *If for a certain set $K = \{\mathbf{k}^{(l)}, l = 1, 2, \dots\}$ it holds that $\|\mathbf{k}^{(l)}\| < C_K$, where C_K is a constant, then the operators $A_{\mathbf{k}^{(l)}} = B(\mathbf{k}^{(l)}, \cdot)$, which are linear in force of Assumption 2, have bounded norm, i.e. there exists a constant C such that $\forall \mathbf{f}$ and $\forall \mathbf{k} \in K$ it holds $\|B(\mathbf{k}, \mathbf{f})\| < C \|\mathbf{f}\|$. Similarly assume that for a certain set $F = \{\mathbf{f}^{(l)}, l = 1, 2, \dots\}$ it holds that $\|\mathbf{f}^{(l)}\| < C_F$, where C_F is a constant, then the operators $A_{\mathbf{f}^{(l)}} = B(\cdot, \mathbf{f}^{(l)})$, which are linear in force of Assumption 2, have bounded norm, i.e. there exists a constant C such that $\forall \mathbf{k}$ and $\forall \mathbf{f} \in F$ it holds $\|B(\mathbf{k}, \mathbf{f})\| < C \|\mathbf{k}\|$;*
- (iii) *The parameter ω is big enough so that*

$$\|B(\mathbf{k}, \mathbf{f}) - \mathbf{g}_\delta\|^2 + \alpha^E \|\mathbf{f}\|^2 + \frac{\omega}{2} \|\hat{\mathbf{f}} - \mathbf{f}\|^2 - \langle \boldsymbol{\xi}, \hat{\mathbf{f}} - \mathbf{f} \rangle,$$

$$\|B(\mathbf{k}, \mathbf{f}) - \mathbf{g}_\delta\|^2 + \gamma \|\mathbf{k} - \mathbf{k}_\epsilon\|^2 + \frac{\omega}{2} \|\hat{\mathbf{k}} - \mathbf{k}\|^2 - \langle \boldsymbol{\mu}, \hat{\mathbf{k}} - \mathbf{k} \rangle$$

are strongly convex with modulus ρ with respect to \mathbf{f} and \mathbf{k} , respectively.

We now assume that the norms of the iterates $\mathbf{f}^{(j)}$ and $\mathbf{k}^{(j)}$ are bounded.

ASSUMPTION 4. *The norm of the iterates $\mathbf{f}^{(j)}$ and $\mathbf{k}^{(j)}$ generated by Algorithm 3 are uniformly bounded.*

When the flux and nonnegativity constraints are imposed we can derive some bounds that the solution should satisfy. For more details about the flux constraint please refer to Section 4, here we only recall that the flux of a vector $\mathbf{x} \in \mathbb{R}^n$ is defined by

$$\text{flux}(\mathbf{x}) = \sum_{i=1}^n \mathbf{x}_i.$$

We can see that the flux bounds the norm of the iterates. Namely, if we constrain the flux of \mathbf{k} to be $\varphi_{\mathbf{k}}$ then $\|\mathbf{k}^{(j)}\| \leq \varphi_{\mathbf{k}} \forall j$, if we constrain the flux of \mathbf{f} to be $\varphi_{\mathbf{f}}$ then $\|\mathbf{f}^{(j)}\| \leq \varphi_{\mathbf{f}} \forall j$.

The bounds proposed above are derived by the following argument. We consider \mathbf{f} , but the extension to \mathbf{k} is trivial. Recalling that $\forall \mathbf{z} \|\mathbf{z}\| \leq \|\mathbf{z}\|_1$, since we are imposing that $\text{flux}(\mathbf{f}) = \varphi_{\mathbf{f}}$ and that $\mathbf{f} \geq 0$, it holds

$$\|\mathbf{f}\| \leq \|\mathbf{f}\|_1 = \text{flux}(\mathbf{f}) = \varphi_{\mathbf{f}}.$$

Assumption 4 seems indeed strong, however, as shown in Section 5, it is always satisfied in our numerical examples.

For convenience we define

$$(10) \quad \phi(\mathbf{f}, \mathbf{k}) = \|B(\mathbf{k}, \mathbf{f}) - \mathbf{g}_\delta\|^2 + \alpha^E \mathbf{f}^2 + \gamma \|\mathbf{k} - \mathbf{k}_\epsilon\|^2,$$

$$(11) \quad \psi_\alpha(\mathbf{f}) = \alpha^{\text{TV}} \|\mathbf{f}\|_{\text{TV}},$$

$$(12) \quad \psi_\beta(\mathbf{k}) = \beta \|\mathbf{k}\|_{\text{TV}}.$$

For the convenience of the reader here we only state the results and postpone the proofs to Appendix A.2.

We now state an auxiliary result that we are going to need for the following.

LEMMA 8. *Let $\boldsymbol{\xi}^{(j)}$, $\boldsymbol{\mu}^{(j)}$, $\mathbf{f}^{(j)}$, $\mathbf{k}^{(j)}$ be the iterations generated by Algorithm 3. Assume that Assumptions 1-4 hold. Then we have*

$$\begin{aligned} \|\boldsymbol{\xi}^{(j+1)} - \boldsymbol{\xi}^{(j)}\| &\leq C \|\mathbf{f}^{(j+1)} - \mathbf{f}^{(j)}\|, \\ \|\boldsymbol{\mu}^{(j+1)} - \boldsymbol{\mu}^{(j)}\| &\leq C \|\hat{\mathbf{k}}^{(j+1)} - \hat{\mathbf{k}}^{(j)}\| \end{aligned}$$

where $C > 0$ is a constant.

We now state that the value of the augmented Lagrangian decays throughout the iterations.

PROPOSITION 9. *With the same notation of Lemma 8 it holds that*

$$\begin{aligned} &\mathcal{L}_A(\mathbf{k}^{(j+1)}, \mathbf{f}^{(j+1)}, \hat{\mathbf{k}}^{(j+1)}, \hat{\mathbf{f}}^{(j+1)}; \boldsymbol{\xi}^{(j+1)}, \boldsymbol{\mu}^{(j+1)}) - \mathcal{L}_A(\mathbf{k}^{(j)}, \mathbf{f}^{(j)}, \hat{\mathbf{k}}^{(j)}, \hat{\mathbf{f}}^{(j)}; \boldsymbol{\xi}^{(j)}, \boldsymbol{\mu}^{(j)}) \\ &\leq \left(\frac{C^2}{\omega} - \frac{\rho}{2}\right) \left(\|\mathbf{f}^{(j+1)} - \mathbf{f}^{(j)}\|^2 + \|\hat{\mathbf{k}}^{(j+1)} - \hat{\mathbf{k}}^{(j)}\|^2\right) - \frac{\rho}{2} \left(\|\hat{\mathbf{f}}^{(j+1)} - \hat{\mathbf{f}}^{(j)}\|^2 + \|\mathbf{k}^{(j+1)} - \mathbf{k}^{(j)}\|^2\right). \end{aligned}$$

We are now in position to state that Algorithm 3 converges to a limit.

LEMMA 10. *Let \mathcal{L}_A be the functional defined in (9) and $\mathbf{k}^{(j)}$, $\mathbf{f}^{(j)}$, $\hat{\mathbf{k}}^{(j)}$, $\hat{\mathbf{f}}^{(j)}$, $\boldsymbol{\xi}^{(j)}$, $\boldsymbol{\mu}^{(j)}$ the iterates generated by Algorithm 3. Let Assumptions 1-4 hold. Moreover, assume that $\frac{C^2}{\omega} - \frac{\rho}{2} < 0$ we have that*

$$\lim_{j \rightarrow \infty} \mathcal{L}_A(\mathbf{k}^{(j)}, \mathbf{f}^{(j)}, \hat{\mathbf{k}}^{(j)}, \hat{\mathbf{f}}^{(j)}; \boldsymbol{\xi}^{(j)}, \boldsymbol{\mu}^{(j)}) \geq \nu,$$

where ν , defined in (22), is the global minimum of $J_{\alpha, \beta}^{\delta, \epsilon}(\mathbf{k}, \mathbf{f})$.

We can now state our primary result.

THEOREM 11. *Let $\mathbf{p}_* = (\mathbf{k}_*, \mathbf{f}_*, \hat{\mathbf{k}}_*, \hat{\mathbf{f}}_*, \boldsymbol{\xi}_*, \boldsymbol{\mu}_*)$ be the limit point of Algorithm 3. Assume that Assumptions 1-4 are satisfied. Then the followings hold*

(a) \mathbf{p}_* is a stationary point, that is

(i) $\mathbf{f}_* = \hat{\mathbf{f}}_*$ and $\mathbf{k}_* = \hat{\mathbf{k}}_*$;

(ii) $\mathbf{0} = \nabla_{\mathbf{f}} \phi(\mathbf{f}_*, \mathbf{k}_*) + \boldsymbol{\xi}_*$ and $\mathbf{0} = \nabla_{\mathbf{k}} \phi(\mathbf{f}_*, \mathbf{k}_*) + \boldsymbol{\mu}_*$;

(iii) $(\hat{\mathbf{k}}_*, \hat{\mathbf{f}}_*) \in \arg \min_{(\hat{\mathbf{k}}, \hat{\mathbf{f}})} \alpha \|\hat{\mathbf{f}}\|_{\text{TV}} + \langle \mathbf{f}_* - \hat{\mathbf{f}}, \boldsymbol{\xi}_* \rangle + \beta \|\hat{\mathbf{k}}\|_{\text{TV}} + \langle \mathbf{k}_* - \hat{\mathbf{k}}, \boldsymbol{\mu}_* \rangle$.

(b) Assume now that $\Omega_{\mathbf{f}} \times \Omega_{\mathbf{k}}$ is compact then

$$\lim_{j \rightarrow \infty} \text{dist} \left(\left(\mathbf{f}^{(j)}, \mathbf{k}^{(j)}, \hat{\mathbf{f}}^{(j)}, \hat{\mathbf{k}}^{(j)}; \boldsymbol{\xi}^{(j)}, \boldsymbol{\mu}^{(j)} \right), Z^* \right) = 0,$$

where Z^* denotes the set of stationary points and dist the Euclidean distance between sets and points.

The proof of Theorem 11 is very similar to the one of [25, Theorem 2.4]. As we mentioned above the main differences between our work and the one in [25] is that in our work the variables are coupled in the functional, while in [25] the variables are decoupled. This means that some technical details inside the proof of [25, Theorem 2.4] needs to be addressed and adapted to our scenario.

4. Flux conservation. In many cases it is possible to know if the true solution (k, f) lies in some closed and convex set $(k, f) \in \Omega_k \times \Omega_f$. Therefore, we want to restrict the domain of $J_{\alpha, \beta}^{\delta, \epsilon}$ so that it can get only values from those sets (for simplicity we suppose that $\Omega_k \times \Omega_f \subseteq \mathcal{D}(J_{\alpha, \beta}^{\delta, \epsilon})$) and our minimization becomes a constrained one. If $\Omega_k \times \Omega_f$ is closed and convex we can notice that the proof of Theorem 2 becomes a simple application of the Weierstrass Theorem.

One of the possible constraint that we can impose is the *flux conservation* that here we describe in detail. Consider, for example, the framework of convolution such that

$$(13) \quad B(k, f) = k * f.$$

Throughout this Section we will assume that k is such that

- (i) $k(\mathbf{x}) \geq 0 \forall \mathbf{x}$;
- (ii) $\int k(\mathbf{x}) d\mathbf{x} = 1$.

It is then straightforward to show, using the convolution theorem, the following

LEMMA 12. *Let k be an integral kernel with compact support and let $g = k * f$, then*

$$\int f = \int g.$$

After a discretization procedure by a collocation method, replacing (13) in (1), the latter becomes

$$\mathbf{g} = A_k \mathbf{x}.$$

Lemma 12 implies that

- A_k has no negative entries;
- If the periodic boundary conditions are imposed, then
 - the row-sum and column-sum of A is 1, i.e., the entries of the vector which discretize k sum up to 1;
 - If $\mathbf{y} = A_k \mathbf{z}$, then $\mathbf{1}^t \mathbf{y} = \mathbf{1}^t \mathbf{z}$, where $\mathbf{1} = (1, 1, \dots, 1)^t$.

DEFINITION 13. *Let $\mathbf{x} \in \mathbb{R}^n$, we call*

$$\text{flux}(\mathbf{x}) = \mathbf{1}^t \mathbf{x},$$

where $\mathbf{1} = (1, 1, \dots, 1)^t$.

REMARK 14. *In the noise free case and when periodic boundary conditions are employed it holds that*

$$(14) \quad \text{flux}(\mathbf{g}) = \text{flux}(\mathbf{x}),$$

where \mathbf{x}, \mathbf{g} are the discretizations of the true signal f and the noise free g , respectively.

In the noisy case (14) does not hold in general. Let us call \mathbf{g}_δ the discretization of g_δ so that $\mathbf{g}_\delta = \mathbf{g} + \boldsymbol{\eta}$, where $\boldsymbol{\eta}$ represents the discretized noise. It holds $\text{flux}(\mathbf{g}_\delta) = \text{flux}(\mathbf{g}) + \text{flux}(\boldsymbol{\eta})$, however, assuming that $\boldsymbol{\eta}$ is white Gaussian noise, we have that $\text{flux}(\boldsymbol{\eta}) \approx 0$. Therefore, in this case we have that $\text{flux}(\mathbf{g}_\delta) \approx \text{flux}(\mathbf{x})$.

From this considerations we can see why we would like to constrain the reconstructed image to lie also in

$$(15) \quad \Omega_F = \{\mathbf{x} \in \mathbb{R}^n \mid \text{flux}(\mathbf{x}) = \text{flux}(\mathbf{g}_\delta)\}.$$

REMARK 15. *The set Ω_F in (15) is a closed and convex set.*

We now construct P_{Ω_F} , the metric projection over Ω_F . By definition of metric projection we have

$$P_{\Omega_F}(\mathbf{x}) = \arg \min_{\mathbf{y} \in \Omega_F} \frac{1}{2} \|\mathbf{x} - \mathbf{y}\|^2.$$

Consider the Fourier matrix F . The Fourier matrix $F \in \mathbb{C}^{n \times n}$ is defined as

$$(F)_{j,k} = e^{i(j-1)(k-1)2\pi/n} \quad j, k = 1, \dots, n, \quad i^2 = -1.$$

When normalized by $\frac{1}{\sqrt{n}}$ F is a unitary matrix and so $\|F\mathbf{z}\| = \|\mathbf{z}\|$ for all \mathbf{z} . Note that the first row of the matrix F is the constant vector $\frac{\mathbf{1}^t}{\sqrt{n}}$. Hence the first entry of $F\mathbf{z}$, for some $\mathbf{z} \in \mathbb{R}^n$, is $\frac{\mathbf{1}^t \mathbf{z}}{\sqrt{n}}$ which is the flux divided by \sqrt{n} of \mathbf{z} by Definition 13. Namely, $\hat{\mathbf{z}}_1 = \frac{\text{flux}(\mathbf{z})}{\sqrt{n}}$ for $\hat{\mathbf{z}} = F\mathbf{z}$. This implies that

$$\hat{\Omega}_F = \{\hat{\mathbf{x}} = F\mathbf{x} \mid \mathbf{x} \in \Omega_F\} = \left\{ \hat{\mathbf{x}} \in \mathbb{R}^n \mid \hat{\mathbf{x}}_1 = \frac{\text{flux}(\mathbf{g}_\delta)}{\sqrt{n}} \right\}.$$

Consider now

$$P_{\Omega_F}(\mathbf{x}) = \arg \min_{\mathbf{y} \in \Omega_F} \frac{1}{2} \|\mathbf{x} - \mathbf{y}\|^2 = \arg \min_{\mathbf{y} \in \Omega_F} \frac{1}{2} \|F\mathbf{x} - F\mathbf{y}\|^2 = F^* \arg \min_{\hat{\mathbf{y}} \in \hat{\Omega}_F} \frac{1}{2} \|\hat{\mathbf{x}} - \hat{\mathbf{y}}\|^2,$$

where we have called $\hat{\mathbf{x}} = F\mathbf{x}$, $\hat{\mathbf{y}} = F\mathbf{y}$. The solution of the last minimization problem follows straightforward from the definition of $\hat{\Omega}_F$. Defining

$$\hat{\mathbf{z}} = \arg \min_{\hat{\mathbf{y}} \in \hat{\Omega}_F} \frac{1}{2} \|\hat{\mathbf{x}} - \hat{\mathbf{y}}\|^2,$$

the j -th entry of $\hat{\mathbf{z}}$ is

$$(16) \quad \hat{z}_j = \begin{cases} \frac{\text{flux}(\mathbf{g}_\delta)}{\sqrt{n}} & \text{if } j = 1 \\ \hat{\mathbf{x}}_j & \text{otherwise.} \end{cases}$$

Finally, we have that

$$(17) \quad P_{\Omega_F}(\mathbf{x}) = F^* \hat{\mathbf{z}},$$

with $\hat{\mathbf{z}}$ defined in (16).

In practice the computation of P_{Ω_F} does not need any FFT and can be done in $O(n)$ arithmetic operations. Let us call \mathbf{v}_j , for $j = 1, \dots, n$, the normalized vectors of the Fourier basis. The expansion of the vector \mathbf{x} in this base is

$$\mathbf{x} = \sum_{j=1}^n \hat{\mathbf{x}}_j \mathbf{v}_j = \frac{\hat{\mathbf{x}}_1}{\sqrt{n}} \mathbf{1} + \sum_{j=2}^n \hat{\mathbf{x}}_j \mathbf{v}_j.$$

According to (17) and (16), it holds

$$(18) \quad \begin{aligned} P_{\Omega_F}(\mathbf{x}) &= \frac{\text{flux}(\mathbf{g}_\delta)}{\sqrt{n}} \frac{\mathbf{1}}{\sqrt{n}} + \sum_{j=2}^n \hat{\mathbf{x}}_j \mathbf{v}_j \\ &= \frac{\text{flux}(\mathbf{g}_\delta)}{n} \mathbf{1} + \sum_{j=2}^n \hat{\mathbf{x}}_j \mathbf{v}_j \\ &= \frac{\text{flux}(\mathbf{g}_\delta)}{n} \mathbf{1} - \frac{\hat{\mathbf{x}}_1}{\sqrt{n}} \mathbf{1} + \frac{\hat{\mathbf{x}}_1}{\sqrt{n}} \mathbf{1} + \sum_{j=2}^n \hat{\mathbf{x}}_j \mathbf{v}_j \\ &= \frac{\text{flux}(\mathbf{g}_\delta) - \hat{\mathbf{x}}_1 \sqrt{n}}{n} \mathbf{1} + \mathbf{x} \\ &= \frac{\text{flux}(\mathbf{g}_\delta) - \text{flux}(\mathbf{x})}{n} \mathbf{1} + \mathbf{x}, \end{aligned}$$

where the last equation holds recalling that $\hat{\mathbf{x}}_1 = \frac{\text{flux}(\mathbf{x})}{\sqrt{n}}$. Note that the computation of $P_{\Omega_F}(\mathbf{x})$ by (18) requires only $O(n)$ operations.

5. Numerical examples. We now give some numerical examples. Firstly we reformulate our algorithm to be more efficient, then we compare our approach with the one in [7] on the same example proposed in their work. Finally we compare our approach with other blind methods in the literature. For each example we compare the result obtained by our proposal with the blind method that we considered from the literature that provided the best result.

We consider the framework of image deblurring with spatially invariant blur, in this setting \mathbf{k} will be the PSF. The blurring phenomenon can be modeled as a Fredholm integral of the first kind

$$(19) \quad B(k, f) = \int k(s-x, t-y) f(s, t) ds dt = k * f,$$

which, due to the spatial invariance property, reduces to a convolution. If we fix one of the two variables and discretize (19) by a collocation method, we get a linear system which is severely ill-conditioned, meaning

that the singular values of its matrix decreases to zero exponentially with no significant spectral gap, see, e.g., [22, 23].

We call $A_{\mathbf{k}}$ the discretization of $B(k, \cdot)$ and $A_{\mathbf{f}}$ the discretized version of $B(\cdot, f)$. For simplicity we assume that both the image and the PSF are periodic and so $A_{\mathbf{f}}$ and $A_{\mathbf{k}}$ are BCCB matrices. Since we imposed periodic boundary condition on D defined in (5) we have that D is a BCCB matrix as well. Thanks to this choice all the matrices involved are diagonalizable using the Fourier transform and so all the linear systems involved can be solved exactly using the FFT in $O(n \log n)$ flops. While this assumption is not needed for the theoretical results it greatly simplifies the computations and lowers the computational effort. If other boundary conditions are employed then the solution of the linear systems involved requires the usage of some iterative algorithm, in this setting the Conjugate Gradient method would perform extremely well since all the system matrices are symmetric positive definite and very well conditioned [21, 5].

Both the image and the PSF should not have negative values, thus we want to constrain them inside the nonnegative cone. Since the assumptions of Section 4 hold, we can use the flux conservation on \mathbf{f} . Moreover, since we know that \mathbf{k} should sum up to 1, we want to constrain the flux of \mathbf{k} to be 1.

Summarizing we set $\Omega_{\mathbf{f}} = \Omega_0 \cap \Omega_F$ and $\Omega_{\mathbf{k}} = \Omega_0 \cap \Omega_1$, where Ω_0 is the nonnegative cone, Ω_F is defined in (15), and $\Omega_1 = \{\mathbf{k} \in \mathbb{R}^{n \times n} \mid \text{flux}(\mathbf{k}) = 1\}$.

Projecting on both $\Omega_{\mathbf{f}}$ and $\Omega_{\mathbf{k}}$ is not trivial then we have to resort to the technique described in Algorithm 7 for decoupling the projection on the two components of $\Omega_{\mathbf{f}}$ and $\Omega_{\mathbf{k}}$. By doing so we are able to perform the projections into the nonnegative cone and on Ω_F in $O(n)$ flops. We are then able to introduce both the nonnegativity and the flux conservation constraints in our method.

For the evaluation of the performances of the method we use the Signal to Noise Ratio (SNR), which is computed as

$$\text{SNR}(\mathbf{f}) = 20 \log_{10} \left(\frac{\|\mathbf{f}^\dagger\|}{\|\mathbf{f} - \mathbf{f}^\dagger\|} \right).$$

The discussion on how to choose the appropriate α^E , α^{TV} , and β is out of the scope of this paper and thus we choose the optimal one, i.e., the one which gives the highest value of SNR among some tested ones. For a discussion on how to select the regularization parameter for the TV regularization see, e.g., [34, 28, 31, 14, 17, 19, 26, 13, 32, 33]. Moreover, for a discussion on multiparameter selection see, e.g., [20].

In Lemma 10 we derive a theoretical condition on ω_j to ensure the convergence of the algorithm. In particular, we state that ω_j should be big enough so that $\frac{C^2}{\omega} - \frac{\rho}{2} < 0$, where $\omega = \min_j \omega_j$. However, from a computational point of view, the choice of a too big ω is undesirable since it is trivial to see that the bigger the ω the slower the method. Thus, we heuristically set $\omega_j \equiv 1$ for all j .

For stopping both methods we use the relative distance between two consecutive iterations, i.e., we stop as soon as

$$\frac{\|\mathbf{f}^{(j-1)} - \mathbf{f}^{(j)}\|}{\|\mathbf{f}^{(j-1)}\|} < 10^{-4}.$$

Since in the considered examples we use the flux conservation (14) and nonnegativity constraints, we have that the bounds in Assumption 4 are

$$(20) \quad \varphi_k = 1 \quad \text{and} \quad \varphi_f = \text{flux}(\mathbf{g}_\delta).$$

In the computed examples we are going to show that, at least experimentally, the norm of the iterates are bounded by the quantities (20) and so that Assumption 4 holds.

For simulating real data we add white Gaussian noise. We will refer to the ratio $\sigma = \frac{\delta}{\|\mathbf{g}\|}$ as noise level.

We want to show that the introduction of the knowledge of the presence of the noise in the PSF and the flux constraint helps in getting better reconstructions. We compare SeB-A against the deconvolution obtained using directly the noisy PSF, i.e., the reconstruction obtained by minimizing

$$(21) \quad \mathbf{f} = \arg \min_{\mathbf{f} \in \Omega_0} \|B(\mathbf{k}_\epsilon, \mathbf{f}) - \mathbf{g}_\delta\|^2 + \alpha^{TV} \|\mathbf{f}\|_{TV} + \alpha^E \|\mathbf{f}\|^2.$$

This algorithm is very similar to the one proposed in [11], the only difference is the presence of $\|\mathbf{f}\|^2$ in the regularization term. To minimize (21) we proceed as in [11], i.e. decoupling the variables, forming the Augmented Lagrangian and then using ADMM.

The decoupled problem in finite dimension becomes

$$\mathbf{f} = \arg \min_{\tilde{\mathbf{f}} \in \Omega_0, \hat{\mathbf{f}}, \mathbf{f}} \left\{ \|B(\mathbf{k}_\epsilon, \mathbf{f}) - \mathbf{g}_\delta\|^2 + \alpha^{\text{TV}} \sum_{i=1}^N \|\hat{\mathbf{f}}_i\| + \alpha^{\text{E}} \|\mathbf{f}\|^2, \right. \\ \left. \tilde{\mathbf{f}} = \mathbf{f}, \hat{\mathbf{f}}_i = D_i \mathbf{f}, \forall i = 1, \dots, N \right\}.$$

The related Augmented Lagrangian is

$$\mathcal{L}_A(\tilde{\mathbf{f}}, \hat{\mathbf{f}}, \mathbf{f}, \boldsymbol{\lambda}, \boldsymbol{\xi}) = \|B(\mathbf{k}_\epsilon, \mathbf{f}) - \mathbf{g}_\delta\|^2 + \alpha^{\text{TV}} \sum_{i=1}^N \|\hat{\mathbf{f}}_i\| + \alpha^{\text{E}} \|\mathbf{f}\|^2 \\ + \frac{\omega_1}{2} \|\tilde{\mathbf{f}} - \mathbf{f}\|^2 - \lambda^t (\tilde{\mathbf{f}} - \mathbf{f}) + \sum_{i=1}^N \left[\frac{\omega_2}{2} \|\hat{\mathbf{f}}_i - D_i \mathbf{f}\|^2 - \boldsymbol{\xi}_i^t (\hat{\mathbf{f}}_i - D_i \mathbf{f}) \right].$$

The ADMM method is described in Algorithm 4.

Algorithm 4 Tikhonov-TV

Let $\mathbf{f}^{(0)}$, $\boldsymbol{\lambda}^{(0)}$, and $\boldsymbol{\xi}^{(0)}$ be initial guesses for \mathbf{f} , $\boldsymbol{\lambda}$, and $\boldsymbol{\xi}$, respectively. Let $\omega_1, \omega_2 > 0$ be real fixed positive numbers.

```

for  $j = 0, 1, \dots$ 
   $\begin{pmatrix} \tilde{\mathbf{f}}^{(j+1)} \\ \hat{\mathbf{f}}^{(j+1)} \end{pmatrix} = \arg \min_{\tilde{\mathbf{f}}, \hat{\mathbf{f}}} \mathcal{L}_A(\tilde{\mathbf{f}}, \hat{\mathbf{f}} | \mathbf{f}^{(j)}; \boldsymbol{\lambda}^{(j)}, \boldsymbol{\xi}^{(j)})$ 
   $\mathbf{f}^{(j+1)} = \arg \min_{\mathbf{f}} \mathcal{L}_A(\mathbf{f} | \tilde{\mathbf{f}}^{(j+1)}, \hat{\mathbf{f}}^{(j+1)}; \boldsymbol{\lambda}^{(j)}, \boldsymbol{\xi}^{(j)})$ 
   $\boldsymbol{\lambda}^{(j+1)} = \boldsymbol{\lambda}^{(j)} - \omega_1 (\tilde{\mathbf{f}}^{(j+1)} - \mathbf{f}^{(j+1)})$ 
   $\boldsymbol{\xi}^{(j+1)} = \boldsymbol{\xi}^{(j)} - \omega_2 (\hat{\mathbf{f}}^{(j+1)} - D \mathbf{f}^{(j+1)})$ 
end

```

The various minimizations in Algorithm 1 are exactly performed, see [11] and Algorithm 5.

Since the minimized functional in (21) is convex it admits a unique minimizer and thus Algorithm 4 converges to it in force of the classical ADMM theory.

The implementation of SeB-A Algorithm 1 can be quite expensive. In fact, since there is no closed form for the minimization with respect to the auxiliary variables $\hat{\mathbf{k}}$ and $\hat{\mathbf{f}}$, the usage of some iterative method is required. This implies that the computational cost of SeB-A can be fairly high. In order to damp this cost we now present a different implementation of the algorithm. We will refer to this method as Computational SeB-A (CSeB-A). This algorithm does not require any inner cycle for the solution of the intermediate problems, however, we are not able to present a rigorous result of convergence. In the first numerical example we are going to show that SeB-A and CSeB-A give equivalent results and thus in the following experiments we are going to use only the latter for computational convenience.

Let us rewrite (4) explicitly and reformulate as in [11]

$$(\mathbf{k}^*, \mathbf{f}^*) = \arg \min_{\mathbf{k} \in \Omega_{\mathbf{k}}, \mathbf{f} \in \Omega_{\mathbf{f}}} \|B(\mathbf{k}, \mathbf{f}) - \mathbf{g}_\delta\|^2 + \alpha^{\text{E}} \|\mathbf{f}\|^2 + \alpha^{\text{TV}} \sum_{i=1}^N \|D_i \mathbf{f}\| + \gamma \|\mathbf{k} - \mathbf{k}_\epsilon\|^2 + \beta \sum_{i=1}^N \|D_i \mathbf{k}\| \\ = \arg \min_{\substack{\mathbf{k} \in \Omega_{\mathbf{k}}, \mathbf{f} \in \Omega_{\mathbf{f}} \\ \tilde{\mathbf{k}}, \tilde{\mathbf{f}}, \mathbf{k}, \mathbf{f}}} \left\{ \|B(\mathbf{k}, \mathbf{f}) - \mathbf{g}_\delta\|^2 + \alpha^{\text{E}} \|\mathbf{f}\|^2 + \alpha^{\text{TV}} \sum_{i=1}^N \|\hat{\mathbf{f}}_i\| + \gamma \|\mathbf{k} - \mathbf{k}_\epsilon\|^2 + \beta \sum_{i=1}^N \|\hat{\mathbf{k}}_i\|, \right. \\ \left. \mathbf{k} = \tilde{\mathbf{k}}, \mathbf{f} = \tilde{\mathbf{f}}, D_i \mathbf{k} = \hat{\mathbf{k}}_i, D_i \mathbf{f} = \hat{\mathbf{f}}_i, i = 1, \dots, N \right\},$$

where, $\mathbf{f}, \tilde{\mathbf{f}}, \mathbf{k}, \tilde{\mathbf{k}} \in \mathbb{R}^N$, $\hat{\mathbf{f}}, \hat{\mathbf{k}} \in \mathbb{R}^{N \times 2}$ and we have $\hat{\mathbf{f}}_i = \begin{pmatrix} \hat{f}_{i,1} \\ \hat{f}_{i,2} \end{pmatrix}$ and similarly for $\hat{\mathbf{k}}_i$.

We now write the Augmented Lagrangian of the minimization above

$$\begin{aligned}
 \mathcal{L}_A & \left(\tilde{\mathbf{f}}, \hat{\mathbf{f}}, \mathbf{f}, \tilde{\mathbf{k}}, \hat{\mathbf{k}}, \mathbf{k}; \boldsymbol{\lambda}, \boldsymbol{\xi}, \boldsymbol{\zeta}, \boldsymbol{\mu} \right) \\
 & = \|B(\mathbf{k}, \mathbf{f}) - \mathbf{g}_\delta\|^2 + \alpha^E \|\mathbf{f}\|^2 + \alpha^{\text{TV}} \sum_{i=1}^N \|\hat{\mathbf{f}}_i\| + \gamma \|\mathbf{k} - \mathbf{k}_\epsilon\|^2 + \beta \sum_{i=1}^N \|\hat{\mathbf{k}}_i\| \\
 & + \frac{\omega_1}{2} \|\tilde{\mathbf{f}} - \mathbf{f}\|^2 - \langle \boldsymbol{\lambda}, (\tilde{\mathbf{f}} - \mathbf{f}) \rangle + \sum_{i=1}^N \left[\frac{\omega_2}{2} \|\hat{\mathbf{f}}_i - D_i \mathbf{f}\|^2 - \langle \boldsymbol{\xi}_i, \hat{\mathbf{f}}_i - D_i \mathbf{f} \rangle \right] \\
 & + \frac{\omega_3}{2} \|\tilde{\mathbf{k}} - \mathbf{k}\|^2 - \langle \boldsymbol{\zeta}, (\tilde{\mathbf{k}} - \mathbf{k}) \rangle + \sum_{i=1}^N \left[\frac{\omega_4}{2} \|\hat{\mathbf{k}}_i - D_i \mathbf{k}\|^2 - \langle \boldsymbol{\mu}_i, \hat{\mathbf{k}}_i - D_i \mathbf{k} \rangle \right],
 \end{aligned}$$

where $\boldsymbol{\lambda}, \boldsymbol{\zeta} \in \mathbb{R}^N$ and $\boldsymbol{\xi}, \boldsymbol{\mu} \in \mathbb{R}^{N \times 2}$.

We can apply the ADMM method obtaining Algorithm 5.

Algorithm 5 CSeB-A

Given $\mathbf{f}^{(0)}, \mathbf{k}^{(0)}, \boldsymbol{\lambda}^{(0)}, \boldsymbol{\xi}^{(0)}, \boldsymbol{\zeta}^{(0)}$ and $\boldsymbol{\mu}_0$ initial guesses for $\mathbf{f}, \mathbf{k}, \boldsymbol{\lambda}, \boldsymbol{\xi}, \boldsymbol{\zeta}$, and $\boldsymbol{\mu}$, respectively. Let $\omega_1, \omega_2, \omega_3, \omega_4 > 0$ be real fixed numbers.

```

for  $j = 0, 1, \dots$ 
     $\begin{pmatrix} \tilde{\mathbf{f}}^{(j+1)} \\ \hat{\mathbf{f}}^{(j+1)} \\ \mathbf{k}^{(j+1)} \end{pmatrix} = \arg \min_{\tilde{\mathbf{f}}, \hat{\mathbf{f}}, \mathbf{k}} \mathcal{L}_A \left( \tilde{\mathbf{f}}, \hat{\mathbf{f}}, \mathbf{k} \mid \tilde{\mathbf{k}}^{(j)}, \hat{\mathbf{k}}^{(j)}, \mathbf{f}^{(j)}; \boldsymbol{\lambda}^{(j)}, \boldsymbol{\xi}^{(j)}, \boldsymbol{\zeta}^{(j)}, \boldsymbol{\mu}^{(j)} \right)$ 
     $\begin{pmatrix} \tilde{\mathbf{k}}^{(j+1)} \\ \hat{\mathbf{k}}^{(j+1)} \\ \mathbf{f}^{(j+1)} \end{pmatrix} = \arg \min_{\tilde{\mathbf{k}}, \hat{\mathbf{k}}, \mathbf{f}} \mathcal{L}_A \left( \tilde{\mathbf{k}}, \hat{\mathbf{k}}, \mathbf{f} \mid \tilde{\mathbf{f}}^{(j+1)}, \hat{\mathbf{f}}^{(j+1)}, \mathbf{k}^{(j+1)}; \boldsymbol{\lambda}^{(j)}, \boldsymbol{\xi}^{(j)}, \boldsymbol{\zeta}^{(j)}, \boldsymbol{\mu}^{(j)} \right)$ 
     $\boldsymbol{\lambda}^{(j+1)} = \boldsymbol{\lambda}^{(j)} - \omega_1 (\tilde{\mathbf{f}}^{(j+1)} - \mathbf{f}^{(j+1)})$ 
     $\boldsymbol{\xi}^{(j+1)} = \boldsymbol{\xi}^{(j)} - \omega_2 (\hat{\mathbf{f}}^{(j+1)} - D\mathbf{f}^{(j+1)})$ 
     $\boldsymbol{\zeta}^{(j+1)} = \boldsymbol{\zeta}^{(j)} - \omega_3 (\tilde{\mathbf{k}}^{(j+1)} - \mathbf{k}^{(j+1)})$ 
     $\boldsymbol{\mu}^{(j+1)} = \boldsymbol{\mu}^{(j)} - \omega_4 (\hat{\mathbf{k}}^{(j+1)} - D\mathbf{k}^{(j+1)})$ 
end
    
```

As we stated above, thanks to Assumption 2 and to the fact that all the matrices involved are BCCB matrices, the minimization above are easily computed and all have a closed form.

$$\begin{aligned}
 \tilde{\mathbf{f}}^{(j+1)} & = P_{\Omega_f} \left(\mathbf{f}^{(j)} + \frac{\boldsymbol{\lambda}^{(j)}}{\omega_1} \right) \\
 \hat{\mathbf{f}}_i^{(j+1)} & = \frac{\left(D_i \mathbf{f}^{(j)} + \frac{1}{\omega_2} \boldsymbol{\xi}_i^{(j)} \right)}{\left\| D_i \mathbf{f}^{(j)} + \frac{1}{\omega_2} \boldsymbol{\xi}_i^{(j)} \right\|} \circ \left(\left\| D_i \mathbf{f}^{(j)} + \frac{1}{\omega_2} \boldsymbol{\xi}_i^{(j)} \right\| - \frac{\alpha^{\text{TV}}}{\omega_2} \right)_+ \\
 \mathbf{k}^{(j+1)} & = \left(2A_{\mathbf{f}^{(j)}}^* A_{\mathbf{f}^{(j)}} + 2\gamma I + \omega_3 I + \omega_4 D^* D \right)^{-1} \left(2A_{\mathbf{f}^{(j)}}^* \mathbf{g}_\delta + 2\gamma \mathbf{k}_\epsilon - \boldsymbol{\zeta}^{(j)} + \omega_3 \tilde{\mathbf{k}}^{(j)} - D^* \boldsymbol{\mu}^{(j)} + \omega_4 D^* \hat{\mathbf{k}}^{(j)} \right) \\
 \tilde{\mathbf{k}}^{(j+1)} & = P_{\Omega_k} \left(\mathbf{k}^{(j+1)} + \frac{\boldsymbol{\zeta}^{(j)}}{\omega_3} \right) \\
 \hat{\mathbf{k}}_i^{(j+1)} & = \frac{\left(D_i \mathbf{k}^{(j+1)} + \frac{1}{\omega_4} \boldsymbol{\mu}_i^{(j)} \right)}{\left\| D_i \mathbf{k}^{(j+1)} + \frac{1}{\omega_4} \boldsymbol{\mu}_i^{(j)} \right\|} \circ \left(\left\| D_i \mathbf{k}^{(j+1)} + \frac{1}{\omega_4} \boldsymbol{\mu}_i^{(j)} \right\| - \frac{\beta}{\omega_4} \right)_+ \\
 \mathbf{f}^{(j+1)} & = \left(2A_{\mathbf{k}^{(j+1)}}^* A_{\mathbf{k}^{(j+1)}} + 2\alpha^E I + \omega_1 I + \omega_2 D^* D \right)^{-1} \left(2A_{\mathbf{k}^{(j+1)}}^* \mathbf{g}_\delta - \boldsymbol{\lambda} + \omega_1 \tilde{\mathbf{f}}^{(j+1)} - D^* \boldsymbol{\xi}^{(j)} + \omega_2 D^* \hat{\mathbf{f}}^{(j+1)} \right)
 \end{aligned}$$

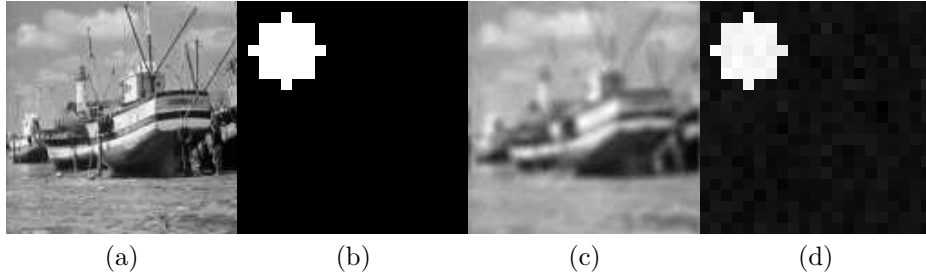


FIGURE 1. Boat test problem: (a) Test image, (b) Noise-free PSF, (c) Blurred and noisy image, (d) Noisy PSF.

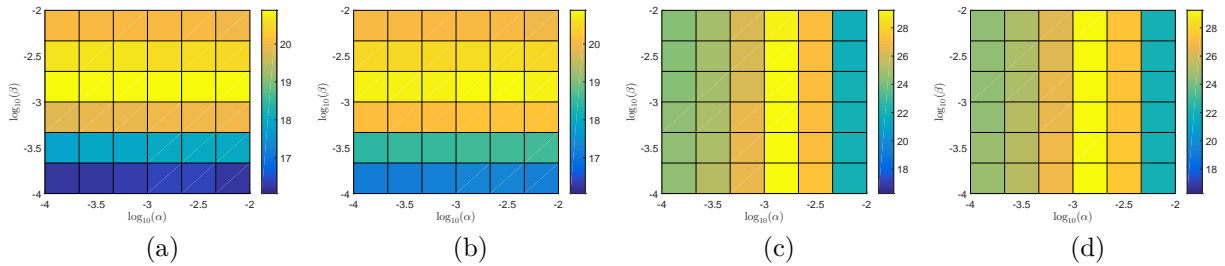


FIGURE 2. Boat test problem SNRs comparison: SNRs comparison between SeB-A and CSeB-A against α and β . (a) SNR of the image for SeB-A, (b) SNR of the image for CSeB-A, (c) SNR of the PSF for SeB-A, (d) SNR of the image for CSeB-A.

In all the example we do not consider the boundary conditions and we assume that all the images are periodic

We can now proceed with the numerical tests. The computations were performed on a laptop pc with an Intel Core i7 6700HQ with 16GB of RAM running Matlab 2016a 64-bit. The results for dbl-RTSL in Subsection 5.2 were taken directly from [7].

5.1. Equivalence between SeB-A and CSeB-A. We would like now to show that SeB-A and CSeB-A give equivalent results. For this purpose we use a relatively small example, since the images involved are of 128×128 pixels. We consider the image deblurring problem in Figure 1. We take the image in Figure 1(a) and blur it with an out of focus PSF then add white Gaussian noise, such that $\sigma = 0.02$. We then add 70% of Gaussian noise to the PSF \mathbf{k} to obtain \mathbf{k}_ϵ .

For simplicity we fix $\alpha^E = \alpha^{TV} = \alpha$ and we run both SeB-A and CSeB-A on several choices of α and β . In particular we test the method on a 7×7 grid of parameters logarithmically spaced between 10^{-4} and 10^{-2} . In Figure 2 we show the SNRs obtained with different choices of α and β with the two methods. We can see that the errors are almost the same. Moreover, we can see that the SNRs obtained with CSeB-A are slightly better than the one obtained with SeB-A. This can be due to the fact that in SeB-A we approximate the minimization with respect to $\hat{\mathbf{f}}$ and $\hat{\mathbf{k}}$ whereas in CSeB-A all the minimization are exact (up to machine precision).

We can also observe that changing α does not affect too much the quality of the reconstruction of \mathbf{k} and, similarly, changing β does not affect too much the quality of the reconstruction of \mathbf{f} .

We conclude by showing, in Figure 3, the best restorations of \mathbf{k} and \mathbf{f} for both methods and in Table 1 the SNRs of the best restorations. From both visual inspection and the comparison of the resulting SNRs we can see that the difference between the two methods in term of accuracy is very small.

5.2. Comparison with dbl-RTLS. We now compare our approach to the one in [7] on the exact same example proposed in that work. In Figure 4 we show the true image, the PSF, and the noise-free blurred image, we add different levels of noise to both the image and the PSF and analyze the behavior in each situation. From the visual inspection of the blurred image we can see that the periodic boundary conditions were employed in its construction and that the center of the convolution is not in the peak of the PSF. Here we keep this behavior so to be able to directly compare our results with the ones in [7]. In Table 2 we show the comparison of the SNR with the different levels of noise. Finally, in Figure 5 we show the reconstructions.

From these comparisons we can see that the proposed approach is able to get a better restoration of the image and in particular of the PSF. The gap between the two approaches gets more and more evident as the quantity of noise increases.

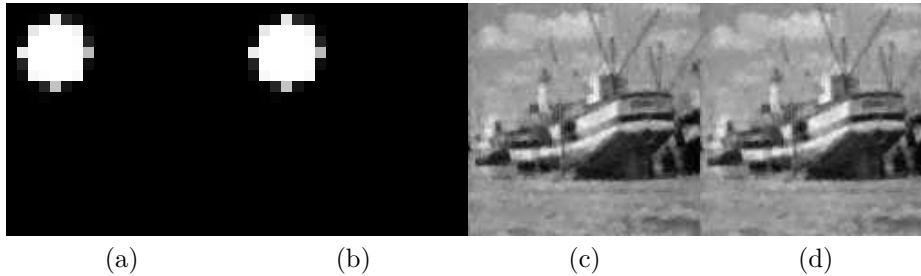


FIGURE 3. Boat test problem optimal reconstructions: (a) PSF computed with SeB-A (SNR=29.129), (b) PSF computed with CSeB-A (SNR=29.211), (c) Image computed with SeB-A (SNR=20.886), (d) Image computed with CSeB-A (SNR=20.859).

SNR \mathbf{f}		SNR \mathbf{k}	
SeB-A	CSeB-A	SeB-A	CSeB-A
20.886	20.859	29.129	29.211

TABLE 1

Boat test problem: Comparison of the SNR obtained with SeB-A and CSeB-A for the optimal choice of parameters.

5.3. Comparison with deconvblind. We now test our method on the Satellite image blurred with an atmospheric PSF and compare the result with the one obtained with the Matlab function `deconvblind`. We consider the image and the PSF provided in the `AtmosphericBlur30` dataset in the toolbox `Restore Tools` [4]. We compute the blurred image by convolving the PSF provided by the toolbox with periodic boundary conditions and we add white Gaussian noise such that $\delta = 0.05 \|g\|$, i.e., with $\sigma = 0.05$. Moreover, we add white Gaussian noise to the PSF, so that $\epsilon = 0.7 \|k\|$. All the corresponding images are shown in Figure 6.

In this example we use the Matlab function `deconvblind`. This function implements a blind Richardson-Lucy algorithm for the solution of blind deconvolution. For a fair comparison we select as initial guess for the PSF the noisy PSF. Moreover we compare, as above, with Tikhonov-TV. For both CSeB-A and Tikhonov-TV we select the best regularization parameter and for `deconvblind` the optimal number of iterations. In Figure 7 we can see the reconstructed images for the three methods and the corresponding SNRs are displayed in Table 2.

From the SNR comparison we can see that the proposed algorithm is able to outperform both Tikhonov-TV and `deconvblind`. This is confirmed also by the visual inspection of the reconstructions.

We note that, since `deconvblind` implements the Richardson-Lucy algorithm and that the PSF is scaled to sum up to 1 then the flux and nonnegativity constraints are satisfied by this method. However, this is not enough for providing a good restoration of the image.

5.4. Comparison with [1]. We now compare our method with the one described in [1] denoted by BID-ADMM. We use the Grain image blurred and we blur it, using periodic boundary conditions, with a non symmetric PSF. We add white Gaussian noise such that $\sigma = 0.01$ and we also add white Gaussian noise to the PSF, so that $\epsilon = 0.8 \|k\|$. All the corresponding images are shown in Figure 8.

Similarly to the previous example the reconstruction obtained with CSeB-A is much better than which obtained with Tikhonov-TV and BID-ADMM.

From the comparison of the SNRs in Table 2 we can see that BID-ADMM is not able to well reconstruct the PSF. However, from a visual inspection of the reconstructed PSF in Figure 9 it does not seem too far from the original one. The main disadvantage of this method is that there is no projection in the nonnegative cone and thus the algorithm is not able to well reconstruct the black area in the low-right corner. We have to stress that the BID-ADMM algorithm does not need the estimation of any parameter while our method relies on a good choice of a set of parameters. However, in [1] the authors were not able to provide a complete theoretical analysis of their algorithm while our method is proven to converge under Assumption 4.

5.5. Bounds. In this last subsection we show that the bounds proposed in Assumption 4 are far from being violated in all the computed examples. In Figure 10 we compare the norm of the iterates against the proposed bounds. The norms $\|\mathbf{f}^{(j)}\|$ are normalized so that the bound is 1, i.e., we plot $\|\mathbf{f}^{(j)}\| / \text{flux}(\mathbf{g}_\delta)$.

In Figure 10(a) we plot the norm of $\mathbf{f}^{(j)}$ against the iteration j and in Figure 10(b) we visualize the norm of $\mathbf{k}^{(j)}$. From this plots we can see that the norm of the iterates increases for the first few iterations and then stabilizes on the final value. From this analysis we can determine that, at least in the computed examples, the

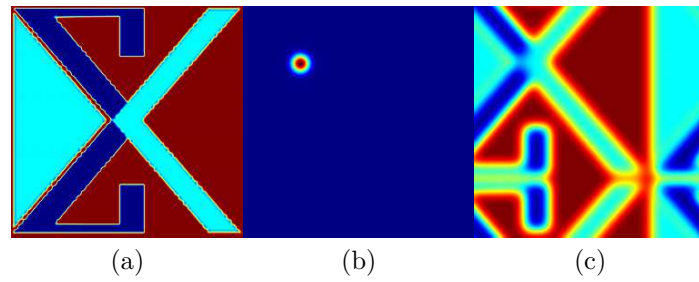


FIGURE 4. Example from [7]: (a) Test image (256×256 pixels), (b) PSF (256×256 pixels), (c) Blurred image (without noise).

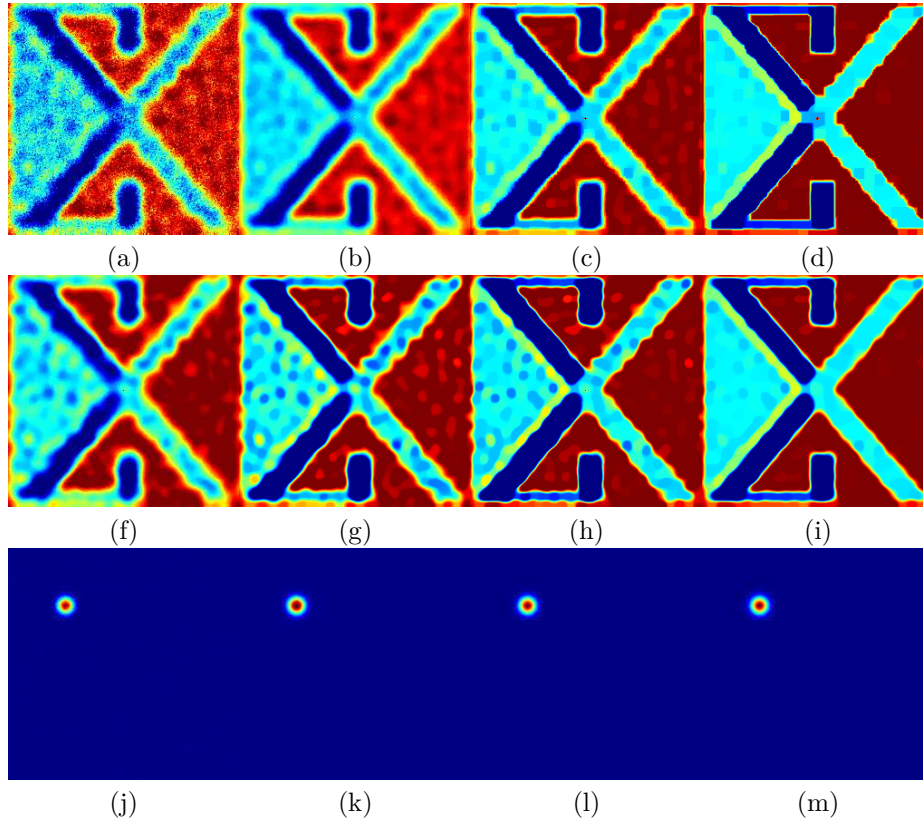


FIGURE 5. Example from [7] reconstructions with different noise levels: on the first row the reconstructed images with Tikhonov-TV, on the second and third row the reconstructed images and PSF with CS&B-A, respectively. From left to right with 8%, 4%, 2%, and 1% of noise.

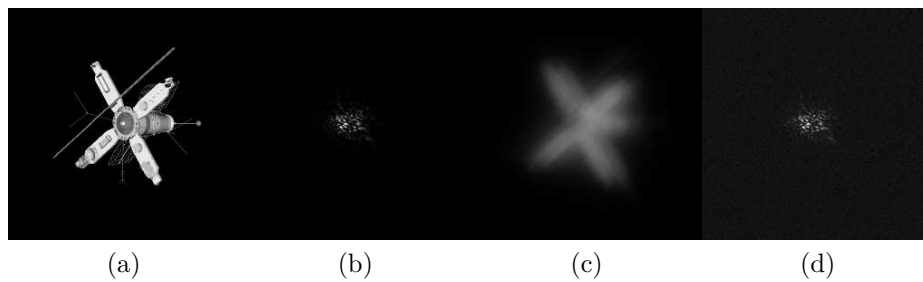


FIGURE 6. Atmospheric blur test problem: (a) Test image (256×256 pixels), (b) Noise-free PSF (256×256 pixels), (c) Blurred and noisy image, (d) Noisy PSF.

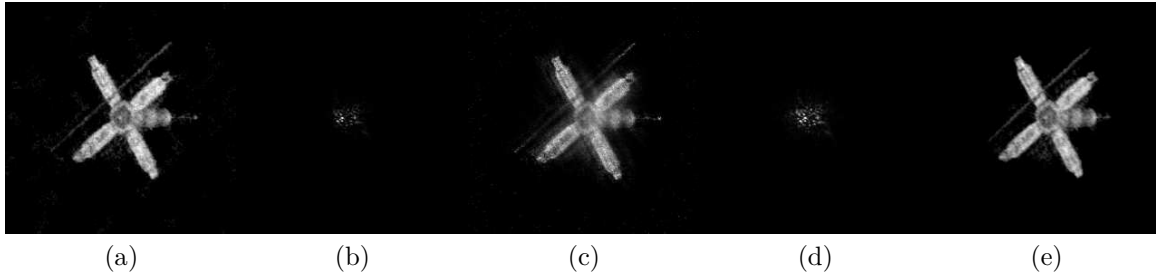


FIGURE 7. Atmospheric blur test problem reconstructions: (a) Image computed with Tikhonov-TV, (b) PSF computed with *deconvblind*, (c) Image computed with *deconvblind*, (d) Image computed with CSeB-A, (e) PSF computed with CSeB-A.

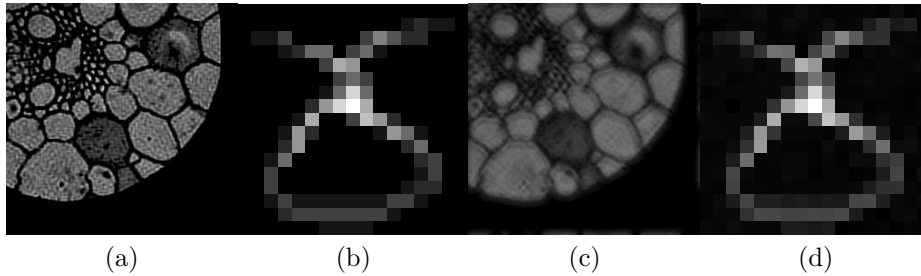


FIGURE 8. Grain test problem: (a) Test image (256×256 pixels), (b) Noise-free PSF (17×17 pixels), (c) Blurred and noisy image, (d) Noisy PSF.

bounds proposed in Assumption 4 hold and thus the hypothesis of Theorem 11 are satisfied.

6. Conclusions. In this work we have proposed a regularization method for solving semi-blind inverse problems. We have developed a functional which incorporates the informations available on both the operator and the solution. The developed functional is non-convex and non-smooth, however, we were able to prove the existence of a global minimum, the stability of this minima, and that the minimization of the functional induces a regularization method.

We proposed an algorithm for computing stationary points of the functional based on the ADMM algorithm and we proved its convergence under some mild assumptions.

We provided some numerical examples that proved the good performances of the proposed algorithm under several situations and when compared with other methods from the literature.

Acknowledgments. The authors would like to thank the referees for their insightful comments that greatly improved the readability and the general quality of this paper.

REFERENCES

- [1] M. S. ALMEIDA AND M. A. FIGUEIREDO, *Blind image deblurring with unknown boundaries using the alternating direction method of multipliers*, in Image Processing (ICIP), 2013 20th IEEE International Conference on, IEEE, 2013, pp. 586–590.
- [2] M. S. C. ALMEIDA AND L. B. ALMEIDA, *Blind and Semi-Blind deblurring of natural images*, IEEE Transactions on Image Processing, 19 (2010), pp. 36–52.
- [3] A. BECK, A. BEN-TAL, AND C. KANZOW, *A fast method for finding the global solution of the regularized structured total least squares problem for image deblurring*, SIAM J. Matrix Anal. Appl., 30 (2008), pp. 419–443.
- [4] S. BERISHA AND J. G. NAGY, *Iterative methods for image restoration*, tech. report, Department of Mathematics and Computer Science, Emory University, 2012. <http://www.mathcs.emory.edu/~nagy/RestoreTools/IR.pdf>.
- [5] Å. BJÖRCK, *Numerical methods in matrix computations*, Springer, 2015.
- [6] I. R. BLEYER AND R. RAMLAU, *A double regularization approach for inverse problems with noisy data and inexact operator*, Inverse Problems, 29 (2013), p. 025004.
- [7] I. R. BLEYER AND R. RAMLAU, *An alternating iterative minimization algorithm for the double-regularised total least square functional*, Inverse Problems, 31 (2015), p. 075004.
- [8] S. BONETTINI, A. CORNELIO, AND M. PRATO, *A new semiblind deconvolution approach for fourier-based image restoration: An application in astronomy*, SIAM Journal on Imaging Sciences, 6 (2013), pp. 1736–1757.
- [9] S. BOYD, N. PARIKH, E. CHU, B. PELEATO, AND J. ECKSTEIN, *Distributed optimization and statistical learning via the alternating direction method of multipliers*, Found. Trends Mach. Learn., 3 (2011), pp. 1–122.
- [10] H. BREZIS, *Functional analysis, Sobolev spaces and partial differential equations*, Springer Science & Business Media, 2010.

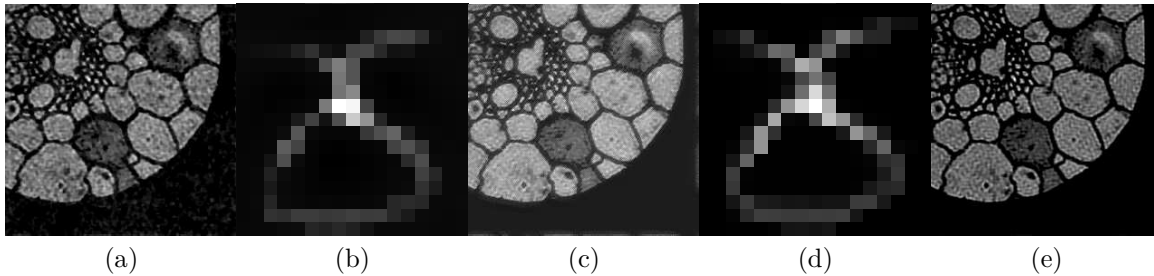


FIGURE 9. Grain test problem reconstructions: (a) Image computed with Tikhonov-TV, (b) PSF computed with BID-ADMM, (c) Image computed with BID-ADMM, (d) Image computed with CSeB-A, (e) PSF computed with CSeB-A.

Example		SNR f			SNR k	
		Compared Method	Tikhonov-TV	CSeB-A	Compared Method	CSeB-A
Example from [7]	1%	15.190	16.379	17.409	15.997	36.071
	2%	13.099	14.344	16.251	15.041	28.496
	4%	12.116	12.192	14.309	13.638	24.284
	8%	8.627	10.006	12.586	11.679	16.255
Satellite		7.8018	11.630	11.856	6.0673	16.655
Grain		11.924	12.844	23.268	1.597	22.925

TABLE 2

Comparison of the SNR for all the examples. For each method we compare our proposal with Tikhonov-TV (Algorithm 2) and another method from the literature, the latter differs from example to example. For the example from [7] we compare with the method in the same paper, for the satellite example we compare with the `deconvblind` Matlab command, and for the grain example we compare with BID-ADMM from [1]. Note that Tikhonov-TV does not take into account the noise in the PSF.

- [11] R. H. CHAN, M. TAO, AND X. YUAN, *Constrained total variation deblurring models and fast algorithms based on alternating direction method of multipliers*, SIAM Journal on imaging Sciences, 6 (2013), pp. 680–697.
- [12] T. F. CHAN AND C.-K. WONG, *Total Variation blind deconvolution*, IEEE Transactions on Image Processing, 7 (1998), pp. 370–375.
- [13] K. CHEN, E. L. PICCOLOMINI, AND F. ZAMA, *An automatic regularization parameter selection algorithm in the total variation model for image deblurring*, Numerical Algorithms, 67 (2014), pp. 73–92.
- [14] C. CLASON, B. JIN, AND K. KUNISCH, *A duality-based splitting method for ℓ_1 - ℓ_1 -tv image restoration with automatic regularization parameter choice*, SIAM Journal on Scientific Computing, 32 (2010), pp. 1484–1505.
- [15] A. CORNELIO, F. PORTA, AND M. PRATO, *A convergent least-squares regularized blind deconvolution approach*, Applied Mathematics and Computation, 259 (2015), pp. 173–186.
- [16] M. DONATELLI AND N. MASTRONARDI, *Fast deconvolution with approximated PSF by RSTLS with antireflective boundary conditions*, J. Comput. Appl. Math., 236 (2012), pp. 3992–4005.
- [17] Y. DONG, M. HINTERMÜLLER, AND M. M. RINCON-CAMACHO, *Automated regularization parameter selection in multi-scale total variation models for image restoration*, Journal of Mathematical Imaging and Vision, 40 (2011), pp. 82–104.
- [18] D. A. FISH, J. G. WALKER, A. M. BRINICOMBE, AND E. R. PIKE, *Blind deconvolution by means of the Richardson–Lucy algorithm*, J. Opt. Soc. Am. A, 12 (1995), pp. 58–65.
- [19] K. FRICK, P. MARNITZ, A. MUNK, ET AL., *Statistical multiresolution dantzig estimation in imaging: fundamental concepts and algorithmic framework*, Electronic Journal of Statistics, 6 (2012), pp. 231–268.
- [20] S. GAZZOLA AND L. REICHEL, *A new framework for multi-parameter regularization*, BIT Numerical Mathematics, 56 (2016), pp. 919–949.
- [21] M. HANKE, *Conjugate gradient type methods for ill-posed problems*, vol. 327, CRC Press, 1995.
- [22] M. HANKE AND P. C. HANSEN, *Regularization methods for large-scale problems*, Surv. Math. Ind., 3 (1993), pp. 253–315.
- [23] P. C. HANSEN, *Rank Deficient and Discrete Ill-Posed Problems: Numerical Aspects of Linear Inversion*, SIAM, 1998.
- [24] L. HE, A. MARQUINA, AND S. J. OSHER, *Blind deconvolution using TV regularization and Bregman iteration*, International Journal of Imaging Systems and Technology, 15 (2005), pp. 74–83.
- [25] M. HONG, Z.-Q. LUO, AND M. RAZAVIYAYN, *Convergence analysis of alternating direction method of multipliers for a family of nonconvex problems*, SIAM Journal on Optimization, 26 (2016), pp. 337–364.
- [26] S. KINDERMANN, L. D. MUTIMBU, AND E. RESMERITA, *A numerical study of heuristic parameter choice rules for total variation regularization*, Journal of Inverse and Ill-Posed Problems, 22 (2014), pp. 63–94.
- [27] A. LANZA, S. MORIGI, AND F. SGALLARI, *A nonsmooth nonconvex sparsity-promoting variational approach for deblurring images corrupted by impulse noise*, in Computational Vision and Medical Image Processing (VipIMAGE 2015, Tenerife, Spain, October 19–21, 2015), CRC Press, 2015, p. 87.
- [28] K. NIINIMÄKI, M. LASSAS, K. HÄMÄLÄINEN, A. KALLONEN, V. KOLEHMAINEN, E. NIEMI, AND S. SILTANEN, *Multiresolution parameter choice method for total variation regularized tomography*, SIAM Journal on Imaging Sciences, 9 (2016), pp. 938–974.

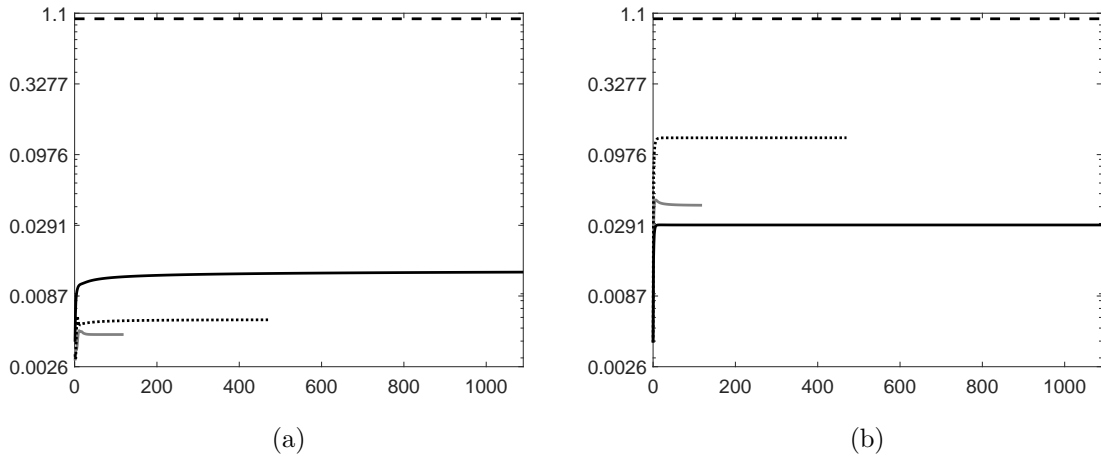


FIGURE 10. Comparison of the norm of the iterates (normalized by the bound) generated by CSeB-A with the proposed bounds for all the examples: (a) $\|\mathbf{f}^{(j)}\|$ against j , (b) $\|\mathbf{k}^{(j)}\|$ against j . The gray lines are related to the example from [7] (with 8% of noise), the black solid lines are related to the atmospheric blur example, the black dotted line are related to the grain example, and the dashed line are the bounds.

- [29] J. P. OLIVEIRA, M. A. FIGUEIREDO, AND J. M. BIOCAS-DIAS, *Blind estimation of motion blur parameters for image deconvolution*, in Iberian Conference on Pattern Recognition and Image Analysis, Springer, 2007, pp. 604–611.
- [30] M. PRATO, A. LA CAMERA, S. BONETTINI, S. REBEGOLDI, M. BERTERO, AND P. BOCCACCI, *A blind deconvolution method for ground based telescopes and fizeau interferometers*, *New Astronomy*, 40 (2015), pp. 1–13.
- [31] H. RULLGÅRD, *A new principle for choosing regularization parameter in certain inverse problems*, arXiv preprint arXiv:0803.3713, (2008).
- [32] B. SIXOU, A. TOMA, L. DENIS, AND F. PEYRIN, *Iterative choice of the optimal regularization parameter in tv image deconvolution*, in *Journal of Physics: Conference Series*, vol. 464, IOP Publishing, 2013, p. 012005.
- [33] A. TOMA, B. SIXOU, AND F. PEYRIN, *Iterative choice of the optimal regularization parameter in tv image restoration.*, *Inverse Problems & Imaging*, 9 (2015).
- [34] Y.-W. WEN AND R. H. CHAN, *Parameter selection for total-variation-based image restoration using discrepancy principle*, *IEEE Transactions on Image Processing*, 21 (2012), pp. 1770–1781.

Appendix A. Proofs of Section 2 and Subsection 3.1.

A.1. Proofs of Section 2.

We now give all the proofs of the results of Section 2. We start by giving the proof of Lemma 1 which shows some basic properties of the functional $J_{\alpha,\beta}^{\delta,\epsilon}$.

Proof (Lemma 1). It is obvious that $J_{\alpha,\beta}^{\delta,\epsilon}$ is positive and that is wslc since it is a sum of wslc functions. We have to prove that it is coercive. The coercivity is trivially showed, in fact, it holds $J_{\alpha,\beta}^{\delta,\epsilon}(k, f) \geq \gamma \|k - k_\epsilon\|^2 + \alpha^E \|f\|^2 \rightarrow \infty$ as $\|(k, f)\| \rightarrow \infty$. \square

We now give the proof of Theorem 2, i.e., we show that the functional $J_{\alpha,\beta}^{\delta,\epsilon}$ admits a global minimum.

Proof (Theorem 2). From Lemma 1 we know that $J_{\alpha,\beta}^{\delta,\epsilon}$ is positive, proper, and coercive, thus $\exists (k, f) \in \mathcal{D}(J_{\alpha,\beta}^{\delta,\epsilon})$, where $\mathcal{D}(J_{\alpha,\beta}^{\delta,\epsilon})$ denotes the domain of $J_{\alpha,\beta}^{\delta,\epsilon}$, such that $J_{\alpha,\beta}^{\delta,\epsilon}(k, f) < \infty$. Let us call

$$(22) \quad \nu := \inf \left\{ J_{\alpha,\beta}^{\delta,\epsilon}(k, f) : (k, f) \in \mathcal{D}(J_{\alpha,\beta}^{\delta,\epsilon}) \right\},$$

we want to see that ν is attained, meaning that the infimum is actually a minimum.

By definition of ν there exist $M > 0$ and $(k_j, f_j) \in \mathcal{D}(J_{\alpha,\beta}^{\delta,\epsilon})$ such that

$$(23) \quad J_{\alpha,\beta}^{\delta,\epsilon}(k_j, f_j) \rightarrow \nu \quad \text{and} \quad J_{\alpha,\beta}^{\delta,\epsilon}(k_j, f_j) \leq M \quad \forall j.$$

From (23) we get that $\alpha^E \|f_j\|^2 \leq M$ and $\gamma \|k_j - k_\epsilon\|^2 \leq M$, moreover, $\|k_j\| - \|k_\epsilon\| \leq \|k_j - k_\epsilon\| \leq \left(\frac{M}{\gamma}\right)^{\frac{1}{2}}$. Thus the following bounds hold

$$\|k_j\| \leq \left(\frac{M}{\gamma}\right)^{\frac{1}{2}} + \|k_\epsilon\| \quad \text{and} \quad \|f_j\| < \left(\frac{M}{\alpha^E}\right)^{\frac{1}{2}},$$

i.e., the sequence $(k_j, f_j)_j$ is uniformly bounded, so there exists a subsequence (k_j, f_j) (with abuse of notation we use the same indexes) such that $k_j \rightharpoonup \bar{k}$ and $f_j \rightharpoonup \bar{f}$, i.e., $(k_j, f_j) \rightharpoonup (\bar{k}, \bar{f})$.

We now prove that ν is the minimum of the functional $J_{\alpha, \beta}^{\delta, \epsilon}$ and is attained in (\bar{k}, \bar{f}) , i.e., (\bar{k}, \bar{f}) is a global minimizer. By wslc of $J_{\alpha, \beta}^{\delta, \epsilon}$ we have

$$\nu \leq J_{\alpha, \beta}^{\delta, \epsilon}(\bar{k}, \bar{f}) \leq \liminf_{j \rightarrow \infty} J_{\alpha, \beta}^{\delta, \epsilon}(k_j, f_j) = \lim_{j \rightarrow \infty} J_{\alpha, \beta}^{\delta, \epsilon}(k_j, f_j) = \nu.$$

So $\nu = J_{\alpha, \beta}^{\delta, \epsilon}(\bar{k}, \bar{f})$ is the minimum of the functional and (\bar{k}, \bar{f}) is a global minimizer. \square

We now prove Theorem 3 which shows that the minima of $J_{\alpha, \beta}^{\delta, \epsilon}$ are stable.

Proof (Theorem 3). Because (k_j, f_j) are minimizers it holds that

$$(24) \quad J_{\alpha, \beta}^{\delta_j, \epsilon_j}(k_j, f_j) \leq J_{\alpha, \beta}^{\delta_j, \epsilon_j}(k, f) \quad \forall (k, f) \in \mathcal{D}(J_{\alpha, \beta}^{\delta_j, \epsilon_j}).$$

Let us indicate with (\tilde{k}, \tilde{f}) the minimizers of $J_{\alpha, \beta}^{\delta, \epsilon}$, namely $(\tilde{k}, \tilde{f}) := (k_{\alpha, \beta}^{\delta, \epsilon}, f_{\alpha, \beta}^{\delta, \epsilon})$.

Since $J_{\alpha, \beta}^{\delta_j, \epsilon_j}(\tilde{k}, \tilde{f}) \rightarrow J_{\alpha, \beta}^{\delta, \epsilon}(\tilde{k}, \tilde{f})$ there exists $\tilde{c} > 0$ such that $J_{\alpha, \beta}^{\delta_j, \epsilon_j}(\tilde{k}, \tilde{f}) \leq \tilde{c}$ for j large enough. The latter implies that $(\|k_j - k_\epsilon\|)_j$ and $(\|f_j\|)_j$ are uniformly bounded and so, like in Theorem 2, it holds that $(k_j, f_j)_j$ is uniformly bounded.

With abuse of notation there exists a subsequence $(k_j, f_j)_j$ such that $k_j \rightharpoonup \bar{k}$ and $f_j \rightharpoonup \bar{f}$. By wslc of B and of $\|\cdot\|$ we have

$$\|B(\bar{k}, \bar{f}) - g_\delta\| \leq \liminf_{j \rightarrow \infty} \|B(k_j, f_j) - g_{\delta_j}\| \quad \text{and} \quad \|k_j - k_\epsilon\| \leq \liminf_{j \rightarrow \infty} \|k_j - k_{\epsilon_j}\|.$$

From (24) it derives

$$\begin{aligned} J_{\alpha, \beta}^{\delta, \epsilon}(\bar{k}, \bar{f}) &\leq \liminf_{j \rightarrow \infty} J_{\alpha, \beta}^{\delta_j, \epsilon_j}(k_j, f_j) \leq \limsup_{j \rightarrow \infty} J_{\alpha, \beta}^{\delta_j, \epsilon_j}(k, f) \\ &= \lim_{j \rightarrow \infty} J_{\alpha, \beta}^{\delta_j, \epsilon_j}(k, f) = J_{\alpha, \beta}^{\delta, \epsilon}(k, f), \quad \forall (k, f) \in \mathcal{D}(J_{\alpha, \beta}^{\delta, \epsilon}). \end{aligned}$$

In particular $J_{\alpha, \beta}^{\delta, \epsilon}(\bar{k}, \bar{f}) \leq J_{\alpha, \beta}^{\delta, \epsilon}(\tilde{k}, \tilde{f})$, but (\tilde{k}, \tilde{f}) is a minimizer and so $J_{\alpha, \beta}^{\delta, \epsilon}(\bar{k}, \bar{f}) = J_{\alpha, \beta}^{\delta, \epsilon}(\tilde{k}, \tilde{f})$, implying that

$$\lim_{j \rightarrow \infty} J_{\alpha, \beta}^{\delta_j, \epsilon_j}(k_j, f_j) = J_{\alpha, \beta}^{\delta, \epsilon}(\bar{k}, \bar{f}).$$

We have proven the weak convergence of k_j and f_j to \bar{k} and \bar{f} , respectively. We now have to show that the convergence is also strong. It is enough to prove that $\|\bar{k}\| \geq \limsup_{j \rightarrow \infty} \|k_j\|$ and $\|\bar{f}\| \geq \limsup_{j \rightarrow \infty} \|f_j\|$.

Let us suppose that $\exists \tau$ such that $\tau = \limsup_{j \rightarrow \infty} \|f_j\| > \|\bar{f}\|$. So there exists a subsequence $(f_n)_n$ of $(f_j)_j$ such that $f_n \rightharpoonup \bar{f}$ and $\|f_n\| \rightarrow \tau$.

$$\begin{aligned} &\lim_{n \rightarrow \infty} \left(\|B(k_n, f_n) - g_{\delta_n}\|^2 + \alpha \|f_n\|_{TV} + \gamma \|k_n - k_{\epsilon_n}\|^2 + \beta \|k_n\|_{TV} \right) \\ &= \|B(\bar{k}, \bar{f}) - g_\delta\|^2 + \alpha^{TV} \|\bar{f}\|_{TV} + \gamma \|\bar{k} - k_\epsilon\|^2 + \beta \|\bar{k}\|_{TV} + \alpha^E \left(\|\bar{f}\|^2 - \lim_{n \rightarrow \infty} \|f_n\|^2 \right) \\ &= \|B(\bar{k}, \bar{f}) - g_\delta\|^2 + \alpha^{TV} \|\bar{f}\|_{TV} + \gamma \|\bar{k} - k_\epsilon\|^2 + \beta \|\bar{k}\|_{TV} + \alpha^E \left(\|\bar{f}\|^2 - \tau^2 \right) \\ &< \|B(\bar{k}, \bar{f}) - g_\delta\|^2 + \alpha^{TV} \|\bar{f}\|_{TV} + \gamma \|\bar{k} - k_\epsilon\|^2 + \beta \|\bar{k}\|_{TV}, \end{aligned}$$

which contradicts the wslc of B and the norms, so $f_j \rightarrow \bar{f}$.

Similarly we prove that $k_j \rightarrow \bar{k}$. Let us suppose that $\exists \tau$ such that $\tau = \limsup_j \|k_j - k_\epsilon\| > \|\bar{k} - k_\epsilon\|$. So there exists a subsequence $(k_n)_n$ of $(k_j)_j$ such that $k_n - k_\epsilon \rightharpoonup \bar{k} - k_\epsilon$ and $\|k_n - k_\epsilon\| \rightarrow \tau$. By triangular inequality

$$\|k_n - k_\epsilon\| - \|k_\epsilon - k_{\epsilon_n}\| \leq \|k_n - k_{\epsilon_n}\| \leq \|k_n - k_\epsilon\| + \|k_\epsilon - k_{\epsilon_n}\|,$$

so $\lim_{n \rightarrow \infty} \|k_n - k_{\epsilon_n}\| = \lim_{n \rightarrow \infty} \|k_n - k_\epsilon\|$. Thus

$$\begin{aligned} & \lim_{n \rightarrow \infty} \left(\|B(k_n, f_n) - g_{\delta_n}\|^2 + \beta \|k_n\|_{TV} \right) \\ &= \|B(\bar{k}, \bar{f}) - g_\delta\|^2 + \beta \|\bar{k}\|_{TV} + \beta \left(\gamma \|\bar{k} - k_\epsilon\|^2 - \lim_{n \rightarrow \infty} \|k_n - k_\epsilon\|^2 \right) \\ &= \|B(\bar{k}, \bar{f}) - g_\delta\|^2 + \beta \|\bar{k}\|_{TV} + \beta \left(\gamma \|\bar{k} - k_\epsilon\|^2 - \tau^2 \right) \\ &< \|B(\bar{k}, \bar{f}) - g_\delta\|^2 + \beta \|\bar{k}\|_{TV}, \end{aligned}$$

which contradicts the wslc of B and the norms, so $k_j \rightarrow \bar{k}$. \square

Finally we give the proof of Theorem 5 which ensures that the minimization of the functional $J_{\alpha, \beta}^{\delta, \epsilon}$ induces a regularization method.

Proof (Theorem 5). Since (k_j, f_j) is a minimizer we have $J_{\alpha_j, \beta_j}^{\delta_j, \epsilon_j}(k_j, f_j) \leq J_{\alpha_j, \beta_j}^{\delta_j, \epsilon_j}(k, f)$ for all $(k, f) \in \mathcal{D}(J_{\alpha_j, \beta_j}^{\delta_j, \epsilon_j})$, in particular

$$0 \leq J_{\alpha_j, \beta_j}^{\delta_j, \epsilon_j}(k_j, f_j) \leq J_{\alpha_j, \beta_j}^{\delta_j, \epsilon_j}(k_0, f^\dagger) \leq \delta_j^2 + \gamma \epsilon_j^2 + \alpha_j^{\text{TV}} \|f^\dagger\|_{TV} + \alpha^{\text{E}} \|f^\dagger\|^2 + \beta_j \|k_0\|_{TV},$$

so $\|B(k_j, f_j) - g_{\delta_j}\|^2$, $\|k_j - k_{\epsilon_j}\|^2$, $\|f_j\|^2$, $\|f_j\|_{TV}$, and $\|k_j\|_{TV}$ are uniformly bounded.

There exists a subsequence $(k_n, f_n)_n$ of $(k_j, f_j)_j$ such that $(k_n, f_n) \rightharpoonup (\bar{k}, \bar{f})$. We want to show that $\bar{k} = k_0$ and that \bar{f} is the minimum norm solution. Moreover, we want to prove that the convergence is strong.

Let us firstly show that $\bar{k} = k_0$. Indeed, it holds

$$\begin{aligned} 0 &\leq \|B(\bar{k}, \bar{f}) - g_0\|^2 + \gamma \|\bar{k} - k_0\|^2 \\ &\leq \liminf_{n \rightarrow \infty} \|B(k_n, f_n) - g_{\delta_n}\|^2 + \gamma \|k_n - k_{\epsilon_n}\|^2 \\ &\leq \liminf_{n \rightarrow \infty} \delta_n^2 + \gamma \epsilon_n^2 + \alpha_n^{\text{E}} \|f^\dagger\|^2 + \alpha_n^{\text{TV}} \|f^\dagger\|_{TV} + \beta_n \|k_0\|_{TV} \\ &= 0, \end{aligned}$$

thus $\bar{k} = k_0$ and $B(\bar{k}, \bar{f}) = g_0$.

We now show that \bar{f} is the minimum norm solution. We have that

$$\|f_n\|^2 + \frac{\alpha_n^{\text{TV}}}{\alpha_n^{\text{E}}} \|f_n\|_{TV} + \frac{\beta_n}{\alpha_n^{\text{E}}} \|k_n\|_{TV} \leq \frac{\delta_n^2 + \gamma}{\alpha_n^{\text{E}}} + \|f^\dagger\|^2 + \frac{\alpha_n^{\text{TV}}}{\alpha_n^{\text{E}}} \|f^\dagger\|_{TV} + \frac{\beta_n}{\alpha_n^{\text{E}}} \|k_0\|_{TV}.$$

We get

$$\begin{aligned} \|\bar{f}\|^2 + \|\bar{f}\|_{TV} + \eta \|\bar{k}\|_{TV} &\leq \liminf_{n \rightarrow \infty} \left(\|f_n\|^2 + \|f_n\|_{TV} + \eta \|k_n\|_{TV} \right) \\ &= \liminf_{n \rightarrow \infty} \left(\|f_n\|^2 + \frac{\alpha_n^{\text{TV}}}{\alpha_n^{\text{E}}} \|f_n\|_{TV} + \frac{\beta_n}{\alpha_n^{\text{E}}} \|k_n\|_{TV} \right) \\ &\leq \liminf_{n \rightarrow \infty} \left(\frac{\delta_n^2 + \gamma}{\alpha_n^{\text{E}}} + \|f^\dagger\|^2 + \frac{\alpha_n^{\text{TV}}}{\alpha_n^{\text{E}}} \|f^\dagger\|_{TV} + \frac{\beta_n}{\alpha_n^{\text{E}}} \|k_0\|_{TV} \right) \\ &= \|f^\dagger\|^2 + \|f^\dagger\|_{TV} + \eta \|k_0\|_{TV}, \end{aligned}$$

but $\bar{k} = k_0$ and so \bar{f} is the minimum norm solution.

We finally prove that $f_n \rightarrow f^\dagger$ and $k_n \rightarrow k_0$.

We start with f_n , it is sufficient to show that $\|f_n\| \rightarrow \|f^\dagger\|$ or equivalently (by wslc of the norm) that $\limsup_{n \rightarrow \infty} \|f_n\| \leq \|f^\dagger\|$. Let us suppose that there exists τ such that $\tau = \limsup_{n \rightarrow \infty} \|f_n\|^2 > \|f^\dagger\|^2$ and so there is a subsequence $(f_l)_l$ of $(f_n)_n$ such that $f_l \rightharpoonup \bar{f}$ and $\|f_l\|^2 \rightarrow \tau$, so

$$\limsup_{l \rightarrow \infty} \frac{\beta_l}{\alpha_l} \|k_l\|_{TV} = \eta \|k_0\|_{TV} + \left(\|\bar{f}\|^2 - \limsup_{l \rightarrow \infty} \|f_l\|^2 \right) < \eta \|k_0\|_{TV},$$

which is a contradiction to the wsc of the norm. So we have that $f_n \rightarrow \bar{f}$.

As for k_n we have

$$\|k_n - k_0\| \leq \|k_n - k_{\epsilon_n}\| + \|k_{\epsilon_n} - k_0\| \leq \|k_n - k_{\epsilon_n}\| + \epsilon_n \rightarrow 0,$$

which leads to the thesis. \square

A.2. Proofs of Subsection 3.1. We give now the proof of Lemma 8.

Proof (Lemma 8). We prove the first inequality. Consider the optimality condition for $\mathbf{f}^{(j+1)}$ obtained differentiating (9), thus $\mathbf{0} = \nabla_{\mathbf{f}}\phi(\mathbf{f}^{(j+1)}, \mathbf{k}^{(j+1)}) + \boldsymbol{\xi}^{(j)} - \omega(\hat{\mathbf{f}}^{(j+1)} - \mathbf{f}^{(j+1)})$, where ϕ is defined in (10). Using the update rule for $\boldsymbol{\xi}^{(j+1)}$ we get

$$(25) \quad -\boldsymbol{\xi}^{(j+1)} = \nabla_{\mathbf{f}}\phi(\mathbf{f}^{(j+1)}, \mathbf{k}^{(j+1)}).$$

Combining Assumption 4 with Assumption 3(ii), we get that the linear operators $\{\nabla_{\mathbf{f}}\phi(\cdot, \mathbf{k}^{(j)})\}_j$ have uniformly bounded norm, i.e., there exists a constant $C_{\mathbf{f}} > 0$ such that $\|\nabla_{\mathbf{f}}\phi(\mathbf{x}, \mathbf{k}^{(j+1)}) - \nabla_{\mathbf{f}}\phi(\mathbf{y}, \mathbf{k}^{(j+1)})\| \leq C_{\mathbf{f}}\|\mathbf{x} - \mathbf{y}\|$. Hence we have

$$\|\boldsymbol{\xi}^{(j+1)} - \boldsymbol{\xi}^{(j)}\| = \|\nabla_{\mathbf{f}}\phi(\mathbf{f}^{(j+1)}, \mathbf{k}^{(j+1)}) - \nabla_{\mathbf{f}}\phi(\mathbf{f}^{(j)}, \mathbf{k}^{(j+1)})\| \leq C_{\mathbf{f}}\|\mathbf{f}^{(j+1)} - \mathbf{f}^{(j)}\|.$$

We now move to the second inequality. Considering the optimality condition of (9) for $\hat{\mathbf{k}}^{(j+1)}$ and denoting with $\partial\psi_{\beta}$ the subdifferential of ψ_{β} defined in (12), we get

$$\mathbf{0} \in \partial\psi_{\beta}(\hat{\mathbf{k}}^{(j+1)}) - \boldsymbol{\mu}^{(j)} + \omega(\hat{\mathbf{k}}^{(j+1)} - \mathbf{k}^{(j+1)}) = \partial\psi_{\beta}(\hat{\mathbf{k}}^{(j+1)}) - \boldsymbol{\mu}^{(j+1)}.$$

Thus it holds $\boldsymbol{\mu}^{(j+1)} - \boldsymbol{\mu}^{(j)} \in \partial\psi_{\beta}(\hat{\mathbf{k}}^{(j+1)}) - \partial\psi_{\beta}(\hat{\mathbf{k}}^{(j)})$, hence

$$\|\boldsymbol{\mu}^{(j+1)} - \boldsymbol{\mu}^{(j)}\| \leq \sup\|\partial\psi_{\beta}(\hat{\mathbf{k}}^{(j+1)}) - \partial\psi_{\beta}(\hat{\mathbf{k}}^{(j)})\|.$$

By Assumption 4 we have that $\|\hat{\mathbf{k}}^{(j)}\|$ is uniformly bounded and thus there exists $C_{\hat{\mathbf{k}}} > 0$ such that

$$\|\boldsymbol{\mu}^{(j+1)} - \boldsymbol{\mu}^{(j)}\| \leq C_{\hat{\mathbf{k}}}\|\hat{\mathbf{k}}^{(j+1)} - \hat{\mathbf{k}}^{(j)}\|.$$

Calling $C = \max\{C_{\mathbf{f}}, C_{\hat{\mathbf{k}}}\}$ we have the thesis. \square

We show the proof of Proposition 9.

Proof (Proposition 9). We split the difference above as

$$(26) \quad \begin{aligned} & \mathcal{L}_A(\mathbf{k}^{(j+1)}, \mathbf{f}^{(j+1)}, \hat{\mathbf{k}}^{(j+1)}, \hat{\mathbf{f}}^{(j+1)}; \boldsymbol{\xi}^{(j+1)}, \boldsymbol{\mu}^{(j+1)}) - \mathcal{L}_A(\mathbf{k}^{(j)}, \mathbf{f}^{(j)}, \hat{\mathbf{k}}^{(j)}, \hat{\mathbf{f}}^{(j)}; \boldsymbol{\xi}^{(j)}, \boldsymbol{\mu}^{(j)}) \\ &= \mathcal{L}_A(\mathbf{k}^{(j+1)}, \mathbf{f}^{(j+1)}, \hat{\mathbf{k}}^{(j+1)}, \hat{\mathbf{f}}^{(j+1)}; \boldsymbol{\xi}^{(j+1)}, \boldsymbol{\mu}^{(j+1)}) - \mathcal{L}_A(\mathbf{k}^{(j+1)}, \mathbf{f}^{(j+1)}, \hat{\mathbf{k}}^{(j+1)}, \hat{\mathbf{f}}^{(j+1)}; \boldsymbol{\xi}^{(j)}, \boldsymbol{\mu}^{(j)}) \\ &+ \mathcal{L}_A(\mathbf{k}^{(j+1)}, \mathbf{f}^{(j+1)}, \hat{\mathbf{k}}^{(j+1)}, \hat{\mathbf{f}}^{(j+1)}; \boldsymbol{\xi}^{(j)}, \boldsymbol{\mu}^{(j)}) - \mathcal{L}_A(\mathbf{k}^{(j)}, \mathbf{f}^{(j)}, \hat{\mathbf{k}}^{(j)}, \hat{\mathbf{f}}^{(j)}; \boldsymbol{\xi}^{(j)}, \boldsymbol{\mu}^{(j)}). \end{aligned}$$

Consider the first part

$$(27) \quad \begin{aligned} & \mathcal{L}_A(\mathbf{k}^{(j+1)}, \mathbf{f}^{(j+1)}, \hat{\mathbf{k}}^{(j+1)}, \hat{\mathbf{f}}^{(j+1)}; \boldsymbol{\xi}^{(j+1)}, \boldsymbol{\mu}^{(j+1)}) - \mathcal{L}_A(\mathbf{k}^{(j+1)}, \mathbf{f}^{(j+1)}, \hat{\mathbf{k}}^{(j+1)}, \hat{\mathbf{f}}^{(j+1)}; \boldsymbol{\xi}^{(j)}, \boldsymbol{\mu}^{(j)}) \\ &= \langle \boldsymbol{\xi}^{(j)} - \boldsymbol{\xi}^{(j+1)}, \hat{\mathbf{f}}^{(j+1)} - \mathbf{f}^{(j+1)} \rangle + \langle \boldsymbol{\mu}^{(j)} - \boldsymbol{\mu}^{(j+1)}, \hat{\mathbf{k}}^{(j+1)} - \mathbf{k}^{(j+1)} \rangle \\ &= \frac{1}{\omega}\|\boldsymbol{\xi}^{(j)} - \boldsymbol{\xi}^{(j+1)}\|^2 + \frac{1}{\omega}\|\boldsymbol{\mu}^{(j)} - \boldsymbol{\mu}^{(j+1)}\|^2, \end{aligned}$$

where the last step is obtained by recalling the definition of $\boldsymbol{\xi}^{(j+1)}$ and $\boldsymbol{\mu}^{(j+1)}$. We move to the second part.

As above we indicate with $\partial\mathcal{L}_A$ the subdifferential of L . Let a general element of $\partial\mathcal{L}_A$ be denoted with

$$\boldsymbol{\theta} \begin{pmatrix} \mathbf{f} \\ \hat{\mathbf{k}} \end{pmatrix} \in \partial \begin{pmatrix} \mathbf{f} \\ \hat{\mathbf{k}} \end{pmatrix} \mathcal{L}_A \left(\mathbf{k}^{(j+1)}, \mathbf{f}^{(j+1)}, \hat{\mathbf{k}}^{(j+1)}, \hat{\mathbf{f}}^{(j+1)}; \boldsymbol{\xi}^{(j)}, \boldsymbol{\mu}^{(j)} \right)$$

and

$$\boldsymbol{\theta} \begin{pmatrix} \hat{\mathbf{f}} \\ \mathbf{k} \end{pmatrix} \in \partial \begin{pmatrix} \hat{\mathbf{f}} \\ \mathbf{k} \end{pmatrix} \mathcal{L}_A \left(\mathbf{k}^{(j+1)}, \mathbf{f}^{(j+1)}, \hat{\mathbf{k}}^{(j+1)}, \hat{\mathbf{f}}^{(j+1)}; \boldsymbol{\xi}^{(j)}, \boldsymbol{\mu}^{(j)} \right).$$

Then it holds

$$\begin{aligned} & \mathcal{L}_A \left(\mathbf{k}^{(j+1)}, \mathbf{f}^{(j+1)}, \hat{\mathbf{k}}^{(j+1)}, \hat{\mathbf{f}}^{(j+1)}; \boldsymbol{\xi}^{(j)}, \boldsymbol{\mu}^{(j)} \right) - \mathcal{L}_A \left(\mathbf{k}^{(j)}, \mathbf{f}^{(j)}, \hat{\mathbf{k}}^{(j)}, \hat{\mathbf{f}}^{(j)}; \boldsymbol{\xi}^{(j)}, \boldsymbol{\mu}^{(j)} \right) \\ &= \mathcal{L}_A \left(\mathbf{k}^{(j+1)}, \mathbf{f}^{(j+1)}, \hat{\mathbf{k}}^{(j+1)}, \hat{\mathbf{f}}^{(j+1)}; \boldsymbol{\xi}^{(j)}, \boldsymbol{\mu}^{(j)} \right) - \mathcal{L}_A \left(\mathbf{k}^{(j+1)}, \mathbf{f}^{(j)}, \hat{\mathbf{k}}^{(j)}, \hat{\mathbf{f}}^{(j+1)}; \boldsymbol{\xi}^{(j)}, \boldsymbol{\mu}^{(j)} \right) \\ &+ \mathcal{L}_A \left(\mathbf{k}^{(j+1)}, \mathbf{f}^{(j)}, \hat{\mathbf{k}}^{(j)}, \hat{\mathbf{f}}^{(j+1)}; \boldsymbol{\xi}^{(j)}, \boldsymbol{\mu}^{(j)} \right) - \mathcal{L}_A \left(\mathbf{k}^{(j)}, \mathbf{f}^{(j)}, \hat{\mathbf{k}}^{(j)}, \hat{\mathbf{f}}^{(j)}; \boldsymbol{\xi}^{(j)}, \boldsymbol{\mu}^{(j)} \right) \\ &\stackrel{(a)}{\leq} \left\langle \boldsymbol{\theta} \begin{pmatrix} \mathbf{f} \\ \hat{\mathbf{k}} \end{pmatrix}, \begin{pmatrix} \mathbf{f}_{j+1} \\ \hat{\mathbf{k}}_{j+1} \end{pmatrix} - \begin{pmatrix} \mathbf{f}_j \\ \hat{\mathbf{k}}_j \end{pmatrix} \right\rangle - \frac{\rho}{2} \left\| \begin{pmatrix} \mathbf{f}_{j+1} \\ \hat{\mathbf{k}}_{j+1} \end{pmatrix} - \begin{pmatrix} \mathbf{f}_j \\ \hat{\mathbf{k}}_j \end{pmatrix} \right\|^2 \\ &+ \left\langle \boldsymbol{\theta} \begin{pmatrix} \hat{\mathbf{f}} \\ \mathbf{k} \end{pmatrix}, \begin{pmatrix} \hat{\mathbf{f}}_{j+1} \\ \mathbf{k}_{j+1} \end{pmatrix} - \begin{pmatrix} \hat{\mathbf{f}}_j \\ \mathbf{k}_j \end{pmatrix} \right\rangle - \frac{\rho}{2} \left\| \begin{pmatrix} \hat{\mathbf{f}}_{j+1} \\ \mathbf{k}_{j+1} \end{pmatrix} - \begin{pmatrix} \hat{\mathbf{f}}_j \\ \mathbf{k}_j \end{pmatrix} \right\|^2 \\ &\stackrel{(b)}{\leq} -\frac{\rho}{2} \left\| \begin{pmatrix} \mathbf{f}_{j+1} \\ \hat{\mathbf{k}}_{j+1} \end{pmatrix} - \begin{pmatrix} \mathbf{f}_j \\ \hat{\mathbf{k}}_j \end{pmatrix} \right\|^2 - \frac{\rho}{2} \left\| \begin{pmatrix} \hat{\mathbf{f}}_{j+1} \\ \mathbf{k}_{j+1} \end{pmatrix} - \begin{pmatrix} \hat{\mathbf{f}}_j \\ \mathbf{k}_j \end{pmatrix} \right\|^2 \\ (28) \quad &= -\frac{\rho}{2} \left(\left\| \mathbf{f}^{(j+1)} - \mathbf{f}^{(j)} \right\|^2 + \left\| \hat{\mathbf{f}}^{(j+1)} - \hat{\mathbf{f}}^{(j)} \right\|^2 + \left\| \mathbf{k}^{(j+1)} - \mathbf{k}^{(j)} \right\|^2 + \left\| \hat{\mathbf{k}}^{(j+1)} - \hat{\mathbf{k}}^{(j)} \right\|^2 \right), \end{aligned}$$

where (a) follows from Assumption 3(iii) and (b) comes from the optimality condition for $\mathbf{k}, \hat{\mathbf{k}}, \hat{\mathbf{f}}$, and \mathbf{f} , i.e., from the fact that we can specialize the subgradients $\boldsymbol{\theta}$ to be the one that satisfies the optimality conditions.

Combining (27) and (28) with (26) and using Lemma 8, we obtain

$$\begin{aligned} & \mathcal{L}_A \left(\mathbf{k}^{(j+1)}, \mathbf{f}^{(j+1)}, \hat{\mathbf{k}}^{(j+1)}, \hat{\mathbf{f}}^{(j+1)}; \boldsymbol{\xi}^{(j+1)}, \boldsymbol{\mu}^{(j+1)} \right) - \mathcal{L}_A \left(\mathbf{k}^{(j)}, \mathbf{f}^{(j)}, \hat{\mathbf{k}}^{(j)}, \hat{\mathbf{f}}^{(j)}; \boldsymbol{\xi}^{(j)}, \boldsymbol{\mu}^{(j)} \right) \\ &\leq \frac{1}{\omega} \left\| \boldsymbol{\xi}^{(j)} - \boldsymbol{\xi}^{(j+1)} \right\|^2 + \frac{1}{\omega} \left\| \boldsymbol{\mu}^{(j)} - \boldsymbol{\mu}^{(j+1)} \right\|^2 - \frac{\rho}{2} \left(\left\| \mathbf{f}^{(j+1)} - \mathbf{f}^{(j)} \right\|^2 \right. \\ &\quad \left. + \left\| \hat{\mathbf{f}}^{(j+1)} - \hat{\mathbf{f}}^{(j)} \right\|^2 + \left\| \mathbf{k}^{(j+1)} - \mathbf{k}^{(j)} \right\|^2 + \left\| \hat{\mathbf{k}}^{(j+1)} - \hat{\mathbf{k}}^{(j)} \right\|^2 \right) \\ &\leq \left(\frac{C^2}{\omega} - \frac{\rho}{2} \right) \left(\left\| \mathbf{f}^{(j+1)} - \mathbf{f}^{(j)} \right\|^2 + \left\| \hat{\mathbf{k}}^{(j+1)} - \hat{\mathbf{k}}^{(j)} \right\|^2 \right) - \frac{\rho}{2} \left(\left\| \hat{\mathbf{f}}^{(j+1)} - \hat{\mathbf{f}}^{(j)} \right\|^2 + \left\| \mathbf{k}^{(j+1)} - \mathbf{k}^{(j)} \right\|^2 \right) \quad \square \end{aligned}$$

We now show that the iterates generated by Algorithm 3 converges to a limit, i.e., we prove Lemma 10.

Proof (Lemma 10). We observe that since we assumed that $\frac{C^2}{\omega} - \frac{\rho}{2} < 0$, from Proposition 9 it holds that the succession $\left\{ \mathcal{L}_A \left(\mathbf{k}^{(j)}, \mathbf{f}^{(j)}, \hat{\mathbf{k}}^{(j)}, \hat{\mathbf{f}}^{(j)}; \boldsymbol{\xi}^{(j)}, \boldsymbol{\mu}^{(j)} \right) \right\}_j$ is monotonically decreasing.

We now prove that the succession is bounded from below. \mathcal{L}_A can be rewritten as

$$\begin{aligned} & \mathcal{L}_A \left(\mathbf{k}^{(j)}, \mathbf{f}^{(j)}, \hat{\mathbf{k}}^{(j)}, \hat{\mathbf{f}}^{(j)}; \boldsymbol{\xi}^{(j)}, \boldsymbol{\mu}^{(j)} \right) \\ &= \left\| B \left(\mathbf{k}^{(j)}, \mathbf{f}^{(j)} \right) - \mathbf{g}_\delta \right\|^2 + \alpha^E \left\| \mathbf{f}^{(j)} \right\|^2 + \alpha^{\text{TV}} \left\| \hat{\mathbf{f}}^{(j)} \right\|_{\text{TV}} + \gamma \left\| \mathbf{k}^{(j)} - \mathbf{k}_\epsilon \right\|^2 + \beta \left\| \hat{\mathbf{k}}^{(j)} \right\|_{\text{TV}} \\ &+ \frac{\omega}{2} \left\| \hat{\mathbf{f}}^{(j)} - \mathbf{f}^{(j)} \right\|^2 - \left\langle \boldsymbol{\xi}^{(j)}, \hat{\mathbf{f}}^{(j)} - \mathbf{f}^{(j)} \right\rangle + \frac{\omega}{2} \left\| \hat{\mathbf{k}}^{(j)} - \mathbf{k}^{(j)} \right\|^2 - \left\langle \boldsymbol{\mu}^{(j)}, \hat{\mathbf{k}}^{(j)} - \mathbf{k}^{(j)} \right\rangle \\ &\in \left\| B \left(\mathbf{k}^{(j)}, \mathbf{f}^{(j)} \right) - \mathbf{g}_\delta \right\|^2 + \alpha^E \left\| \mathbf{f}^{(j)} \right\|^2 + \alpha^{\text{TV}} \left\| \hat{\mathbf{f}}^{(j)} \right\|_{\text{TV}} + \gamma \left\| \mathbf{k}^{(j)} - \mathbf{k}_\epsilon \right\|^2 + \beta \left\| \hat{\mathbf{k}}^{(j)} \right\|_{\text{TV}} \\ &+ \frac{\omega}{2} \left\| \hat{\mathbf{f}}^{(j)} - \mathbf{f}^{(j)} \right\|^2 - \left\langle \nabla_{\mathbf{f}} \phi \left(\mathbf{k}^{(j)}, \mathbf{f}^{(j)} \right), \hat{\mathbf{f}}^{(j)} - \mathbf{f}^{(j)} \right\rangle + \frac{\omega}{2} \left\| \hat{\mathbf{k}}^{(j)} - \mathbf{k}^{(j)} \right\|^2 - \left\langle \partial\psi_\beta \left(\mathbf{k}^{(j)} \right), \hat{\mathbf{k}}^{(j)} - \mathbf{k}^{(j)} \right\rangle. \end{aligned}$$

Using the fact that $\nabla_{\mathbf{f}}\phi$ and all the elements in $\partial\psi_\beta$ are Lipschitz continuous and considering that $\frac{\alpha^2}{\omega} - \frac{\rho}{2} < 0$ by assumption we get

$$\begin{aligned} \mathcal{L}_A \left(\mathbf{k}^{(j)}, \mathbf{f}^{(j)}, \hat{\mathbf{k}}^{(j)}, \hat{\mathbf{f}}^{(j)}; \boldsymbol{\xi}^{(j)}, \boldsymbol{\mu}^{(j)} \right) &\geq \left\| B \left(\mathbf{k}^{(j)}, \mathbf{f}^{(j)} \right) - \mathbf{g}_\delta \right\|^2 \\ &\quad + \alpha^E \left\| \mathbf{f}^{(j)} \right\|^2 + \alpha^{TV} \left\| \mathbf{f}^{(j)} \right\|_{TV} + \gamma \left\| \mathbf{k}^{(j)} - \mathbf{k}_\epsilon \right\|^2 + \beta \left\| \mathbf{k}^{(j)} \right\|_{TV} \\ &\geq \nu \end{aligned}$$

where in the last step we have used the fact that ν is the global minimum of $J_{\alpha,\beta}^{\delta,\epsilon}(\mathbf{k}, \mathbf{f})$.

Since the succession $\left\{ \mathcal{L}_A \left(\mathbf{k}^{(j)}, \mathbf{f}^{(j)}, \hat{\mathbf{k}}^{(j)}, \hat{\mathbf{f}}^{(j)}; \boldsymbol{\xi}^{(j)}, \boldsymbol{\mu}^{(j)} \right) \right\}_j$ is monotonically decreasing and bounded we have that it converges. \square

Finally we show the proof of our main result, Theorem 11.

Proof (Theorem 11). We only prove part (a), we omit the proof of part (b) since it can be copied with no significant modification from [25, Theorem 2.4 part 3].

From Proposition 9 and Lemma 10 for $j \rightarrow \infty$ we have that

$$\left\| \mathbf{f}^{(j+1)} - \mathbf{f}^{(j)} \right\| \rightarrow 0, \quad \left\| \hat{\mathbf{f}}^{(j+1)} - \hat{\mathbf{f}}^{(j)} \right\| \rightarrow 0, \quad \left\| \mathbf{k}^{(j+1)} - \mathbf{k}^{(j)} \right\| \rightarrow 0, \quad \left\| \hat{\mathbf{k}}^{(j+1)} - \hat{\mathbf{k}}^{(j)} \right\| \rightarrow 0.$$

Moreover, in force of Lemma 8 it holds

$$(29) \quad \left\| \boldsymbol{\xi}^{(j+1)} - \boldsymbol{\xi}^{(j)} \right\| \rightarrow 0, \quad \left\| \boldsymbol{\mu}^{(j+1)} - \boldsymbol{\mu}^{(j)} \right\| \rightarrow 0.$$

Let ϕ be the functional defined in (10) and recall that $\mathbf{p}^* = (\mathbf{f}_*, \mathbf{k}_*, \hat{\mathbf{f}}_*, \hat{\mathbf{k}}_*, \boldsymbol{\xi}_*, \boldsymbol{\mu}_*)$ is the limit point (that exists in virtue of Lemma 10) generated by the iterations of Algorithm 3. Observing that, from (29), we obtain that $\mathbf{f}_* = \hat{\mathbf{f}}_*$ and $\mathbf{k}_* = \hat{\mathbf{k}}_*$, which proves (i).

We move now to the proof of (ii). From (25) we have that $\mathbf{0} = \nabla_{\mathbf{f}}\phi(\mathbf{f}^{(j+1)}, \mathbf{k}^{(j+1)}) + \boldsymbol{\xi}^{(j+1)}$ and by taking the limit for $j \rightarrow \infty$ we get that $\mathbf{0} = \nabla_{\mathbf{f}}\phi(\mathbf{f}_*, \mathbf{k}_*) + \boldsymbol{\xi}_*$. Consider now $\mathbf{k}^{(j+1)}$, imposing the optimality condition for $\mathbf{k}^{(j+1)}$ to (9), it holds

$$\begin{aligned} \mathbf{0} &= \nabla_{\mathbf{k}}\phi \left(\mathbf{f}^{(j)}, \mathbf{k}^{(j+1)} \right) + \boldsymbol{\mu}^{(j)} - \omega \left(\hat{\mathbf{k}}^{(j)} - \mathbf{k}^{(j+1)} \right) \\ &= \nabla_{\mathbf{k}}\phi \left(\mathbf{f}^{(j)}, \mathbf{k}^{(j+1)} \right) + \boldsymbol{\mu}^{(j)} - \omega \left(\hat{\mathbf{k}}^{(j)} - \mathbf{k}^{(j+1)} \right) - \omega \hat{\mathbf{k}}^{(j+1)} + \omega \hat{\mathbf{k}}^{(j+1)} \\ &= \nabla_{\mathbf{k}}\phi \left(\mathbf{f}^{(j)}, \mathbf{k}^{(j+1)} \right) + \boldsymbol{\mu}^{(j+1)} - \omega \left(\hat{\mathbf{k}}^{(j)} - \hat{\mathbf{k}}^{(j+1)} \right), \end{aligned}$$

where we added and subtracted $\omega \hat{\mathbf{k}}^{(j+1)}$ in order to make $\boldsymbol{\mu}^{(j+1)}$ to appear.

Taking the limit for $j \rightarrow \infty$, since $\left\| \hat{\mathbf{k}}^{(j+1)} - \hat{\mathbf{k}}^{(j)} \right\| \rightarrow 0$, we have that $\mathbf{0} = \nabla_{\mathbf{k}}\phi(\mathbf{f}_*, \mathbf{k}_*) + \boldsymbol{\mu}_*$, completing the proof of (ii).

For the proof of (iii) consider the optimality condition of (9) for $\hat{\mathbf{f}}^{(j+1)}$, we have that there exists $\boldsymbol{\theta} \in \partial\psi_\alpha(\hat{\mathbf{f}}^{(j+1)})$, where ψ_α is defined in (11), such that $\left\langle \hat{\mathbf{f}} - \hat{\mathbf{f}}^{(j+1)}, \boldsymbol{\theta} - \left(\boldsymbol{\xi}^{(j)} - \omega \left(\hat{\mathbf{f}}^{(j+1)} - \hat{\mathbf{f}}^{(j)} \right) \right) \right\rangle \geq 0 \forall \hat{\mathbf{f}}$.

By convexity of ψ_α it holds $\psi_\alpha(\hat{\mathbf{f}}) - \psi_\alpha(\hat{\mathbf{f}}^{(j+1)}) + \left\langle \hat{\mathbf{f}} - \hat{\mathbf{f}}^{(j+1)}, -\boldsymbol{\xi}^{(j)} + \omega \left(\hat{\mathbf{f}}^{(j+1)} - \hat{\mathbf{f}}^{(j)} \right) \right\rangle \geq 0 \forall \hat{\mathbf{f}}$. Taking to the limit for $j \rightarrow \infty$ we have that $\psi_\alpha(\hat{\mathbf{f}}) - \psi_\alpha(\hat{\mathbf{f}}_*) - \left\langle \hat{\mathbf{f}} - \hat{\mathbf{f}}_*, \boldsymbol{\xi}_* \right\rangle \geq 0 \forall \hat{\mathbf{f}}$.

Using the fact that $\hat{\mathbf{f}}_* = \mathbf{f}_*$ we have

$$\psi_\alpha(\hat{\mathbf{f}}) + \left\langle \mathbf{f}_* - \hat{\mathbf{f}}, \boldsymbol{\xi}_* \right\rangle - \left(\psi_\alpha(\hat{\mathbf{f}}_*) + \left\langle \mathbf{f}_* - \hat{\mathbf{f}}_*, \boldsymbol{\xi}_* \right\rangle \right) \geq 0 \quad \forall \hat{\mathbf{f}}.$$

Similarly we can prove, recalling the definition of ψ_β in (12), that

$$\psi_\beta(\hat{\mathbf{k}}) + \left\langle \mathbf{k}_* - \hat{\mathbf{k}}, \boldsymbol{\mu}_* \right\rangle - \left(\psi_\beta(\hat{\mathbf{k}}_*) + \left\langle \mathbf{k}_* - \hat{\mathbf{k}}_*, \boldsymbol{\mu}_* \right\rangle \right) \geq 0 \quad \forall \hat{\mathbf{k}},$$

which concludes the proof of (iii). We have then proven that the limit point of the succession is a stationary point for the unconstrained problem. \square

Appendix B. ADMM algorithm for multiple constrained minimization. We now discuss how to use the ADMM algorithm for solving a constrained minimization problem. Suppose that we want to constrain a minimization problem, so that our minimizer lies in some closed and convex set $\Omega \subset \mathcal{D}(f)$

$$(30) \quad \begin{aligned} \mathbf{x} &= \arg \min_{\mathbf{x} \in \Omega} f(\mathbf{x}), \\ &\text{subject to } A\mathbf{x} = \mathbf{b}. \end{aligned}$$

Let us write (30) in an equivalent way

$$(31) \quad \begin{aligned} (\mathbf{x}, \tilde{\mathbf{x}}) &= \arg \min_{\tilde{\mathbf{x}} \in \Omega, \mathbf{x}} f(\mathbf{x}), \\ &\text{subject to } A\mathbf{x} = \mathbf{b} \text{ and } \tilde{\mathbf{x}} = \mathbf{x}. \end{aligned}$$

The Augmented Lagrangian associated with (31) is

$$\mathcal{L}_A(\tilde{\mathbf{x}}, \mathbf{x}; \boldsymbol{\lambda}, \boldsymbol{\xi}) = f(\mathbf{x}) - \langle \boldsymbol{\lambda}, \tilde{\mathbf{x}} - \mathbf{x} \rangle + \frac{\omega_1}{2} \|\tilde{\mathbf{x}} - \mathbf{x}\|^2 - \langle \boldsymbol{\xi}, \mathbf{b} - A\mathbf{x} \rangle + \frac{\omega_2}{2} \|\mathbf{b} - A\mathbf{x}\|^2.$$

The ADMM applied to (31) leads to Algorithm 6.

Algorithm 6

Let $\mathbf{x}^{(0)}$, $\boldsymbol{\lambda}^{(0)}$, and $\boldsymbol{\xi}^{(0)}$ be initial guesses for \mathbf{x} , $\boldsymbol{\lambda}$, and $\boldsymbol{\xi}$, respectively

```

for  $j = 0, 1, \dots$ 
 $\tilde{\mathbf{x}}^{(j+1)} = \arg \min_{\tilde{\mathbf{x}} \in \Omega} \mathcal{L}_A(\tilde{\mathbf{x}} \mid \mathbf{x}^{(j)}; \boldsymbol{\lambda}^{(j)}, \boldsymbol{\xi}^{(j)})$ 
 $\mathbf{x}^{(j+1)} = \arg \min_{\mathbf{x}} \mathcal{L}_A(\mathbf{x} \mid \tilde{\mathbf{x}}^{(j+1)}; \boldsymbol{\lambda}^{(j)}, \boldsymbol{\xi}^{(j)})$ 
 $\boldsymbol{\lambda}^{(j+1)} = \boldsymbol{\lambda}^{(j)} - \omega_1 (\tilde{\mathbf{x}}^{(j+1)} - \mathbf{x}^{(j+1)})$ 
 $\boldsymbol{\xi}^{(j+1)} = \boldsymbol{\xi}^{(j)} - \omega_2 (\mathbf{b} - A\mathbf{x}^{(j+1)})$ 
end
    
```

It is easy to show that $\tilde{\mathbf{x}}^{(j+1)}$ in Algorithm 6 is obtained by

$$\tilde{\mathbf{x}}^{(j+1)} = P_{\Omega} \left(\mathbf{x}^{(j)} + \frac{\boldsymbol{\lambda}^{(j)}}{\omega_1} \right).$$

In this way we are able to easily deal with the constrained optimization problem.

This approach is possible if the projection P_{Ω} is easily performed. On the other hand if the projection into Ω is too complicated the algorithm above might not be attractive. Nevertheless, whenever that Ω can be written as the intersection of two or more closed and convex sets, i.e.,

$$\Omega = \bigcap_{m=1}^M \Omega^{(m)},$$

if $P_{\Omega^{(m)}}$ is easily performed, then we can still use ADMM to solve the constrained minimization problem.

For the sake of simplicity we fix $M = 2$. Consider the minimization problem

$$\begin{aligned} &\min_{\mathbf{x}} f(\mathbf{x}) \\ &\text{s.t. } \mathbf{x} \in \Omega^{(1)} \cap \Omega^{(2)}, \end{aligned}$$

which is equivalent to the following

$$(32) \quad \begin{aligned} &\min_{\mathbf{x}, \mathbf{x}^{(1)}, \mathbf{x}^{(2)}} f(\mathbf{x}) \\ &\text{s.t. } \mathbf{x}^{(1)} \in \Omega^{(1)}, \mathbf{x}^{(2)} \in \Omega^{(2)}, \mathbf{x} = \mathbf{x}^{(1)}, \mathbf{x} = \mathbf{x}^{(2)}. \end{aligned}$$

In this way we have separated the two constraints on $\Omega^{(1)}$ and $\Omega^{(2)}$; hopefully the projection on each set is easier to compute than the projection over the intersection.

The Augmented Lagrangian of the new minimization problem (32) is

$$\mathcal{L}_A(\mathbf{x}, \mathbf{x}_{(1)}, \mathbf{x}_{(2)}; \boldsymbol{\lambda}, \boldsymbol{\theta}) = f(\mathbf{x}) - \langle \boldsymbol{\lambda}, \mathbf{x}_{(1)} - \mathbf{x} \rangle + \frac{\omega_1}{2} \|\mathbf{x}_{(1)} - \mathbf{x}\|^2 - \langle \boldsymbol{\theta}, \mathbf{x}_{(2)} - \mathbf{x} \rangle + \frac{\omega_2}{2} \|\mathbf{x}_{(2)} - \mathbf{x}\|^2.$$

We can now write the ADMM iterations for this \mathcal{L}_A in Algorithm 7.

Algorithm 7

Given $\mathbf{x}^{(0)}$, $\boldsymbol{\lambda}^{(0)}$, and $\boldsymbol{\theta}^{(0)}$ initial guesses for \mathbf{x} , $\boldsymbol{\lambda}$, and $\boldsymbol{\theta}$, respectively. Let $\omega_1, \omega_2 > 0$ be real constant numbers.

```

for  $j = 0, 1, \dots$ 
     $\begin{pmatrix} \mathbf{x}_{(1)}^{(j+1)} \\ \mathbf{x}_{(2)}^{(j+1)} \end{pmatrix} = \arg \min_{\mathbf{x}_{(1)}, \mathbf{x}_{(2)}} \mathcal{L}_A(\mathbf{x}_{(1)}, \mathbf{x}_{(2)} | \mathbf{x}^{(j)}; \boldsymbol{\lambda}^{(j)}, \boldsymbol{\theta}^{(j)})$ 
     $\mathbf{x}^{(j+1)} = \arg \min_{\mathbf{x}} \mathcal{L}_A(\mathbf{x} | \mathbf{x}_{(1)}^{(j+1)}, \mathbf{x}_{(2)}^{(j+1)}; \boldsymbol{\lambda}^{(j)}, \boldsymbol{\theta}^{(j)})$ 
     $\boldsymbol{\lambda}^{(j+1)} = \boldsymbol{\lambda}^{(j)} - \omega_1 (\mathbf{x}_{(1)}^{(j+1)} - \mathbf{x}^{(j+1)})$ 
     $\boldsymbol{\theta}^{(j+1)} = \boldsymbol{\theta}^{(j)} - \omega_2 (\mathbf{x}_{(2)}^{(j+1)} - \mathbf{x}^{(j+1)})$ 
end

```

The minimization problem related to the auxiliary variables $\mathbf{x}_{(1)}$ and $\mathbf{x}_{(2)}$ decouples and the solution is simply obtained by

$$\mathbf{x}_{(1)}^{(j+1)} = P_{\Omega^{(1)}} \left(\mathbf{x}^{(j)} + \frac{\boldsymbol{\lambda}^{(j)}}{\omega_1} \right), \quad \mathbf{x}_{(2)}^{(j+1)} = P_{\Omega^{(2)}} \left(\mathbf{x}^{(j)} + \frac{\boldsymbol{\theta}^{(j)}}{\omega_2} \right).$$

We do not consider the minimization in respect to \mathbf{x} since it is not relevant for our scope.