An Alternating Direction Algorithm for Total Variation Reconstruction of Distributed Parameters

Nuno B. Brás, *Member, IEEE,* J. Bioucas-Dias, *Member, IEEE,*, Raul C. Martins, *Member, IEEE,*, A. C. Serra, *Member, IEEE,*

***Abstract*—Augmented Lagrangian variational formulations and alternating optimization have been adopted to solve distributed parameter estimation problems. The alternating direction method of multipliers (ADMM) is one of such formulations/optimization methods. Very recently the number of applications of ADMM, or variants of it, to solve inverse problems in image and signal processing has increased at an exponential rate. The reason for this interest is that ADMM decomposes a difficult optimization problem into a sequence of much simpler problems.**

In this work we use ADMM to reconstruct piece-wise smooth distributed parameters of elliptic partial differential equations from noisy and linear (blurred) observations of the underlying field. The distributed parameters are estimated by solving an inverse problem with total variation regularization.

The proposed instance of ADMM solves, in each iteration, an ℓ_2 and a decoupled $\ell_2 - \ell_1$ optimization problems. An operator splitting is used to simplify the treatment of the TV regularizer avoiding its smooth approximation and yielding a simple, yet effective, ADMM reconstruction method compared with previously proposed approaches.

The competitiveness of the proposed method, with respect to the state-of-the-art, is illustrated in simulated 1D and 2D elliptic equation problems, which are representative of many real applications.

I. Introduction

Recovering the constitutive parameters of a partial differential equation from field vectors or scalar potential observations is usually known as distributed parameter estimation (DPE) [1]. DPEs from elliptical differential equations are frequently studied due to their wide applicability to real world problems. These problems arise from many areas such as electrical impedance tomography [2], magnetotelluric inversion [3], oil reservoir simulation [4], water resources research [5], elastic membrane behaviour [6], heat transfer problems [7], electrostatic problem [2], among others.

Consider the set of partial differential equations

$$\begin{aligned} -\nabla \cdot (\sigma \nabla u_i) &= y_i, \quad \text{in} \quad \Omega \subset \mathbb{R}^2 \\ u_i &= 0, \quad \text{on} \quad \partial\Omega, \end{aligned} \tag{1}$$

where, for $i = 1, \ldots, n_I$, u_i denotes the scalar field associated to the source y_i and $\sigma(x)$, for $x \in \Omega$, denotes the constitutive parameters to be inferred. The above set of partial differential equations may have resulted, for example, from a series of n_I experiments where, in each one, the source y_i is repositioned and linear observations, d_i, of the field u_i are carried out, *i.e.*,

$$d_i = M u_i + \eta_i, \quad i = 1, \ldots, n_I, \tag{2}$$

where M is a linear operator and η_i, for $i = 1, \ldots, n_I$, models additive perturbations.

Denoting the forward operator associated to the partial differential equation (1) by $\mathcal{A}(\sigma)$, equipped with the respective boundary condition, then the inverse problem in hand is

$$\text{find } \sigma \in \mathcal{S} \text{ compatible with } \left\{ \begin{array}{l} d = Mu + \eta \\ \text{s.t.: } \mathcal{A}(\sigma)\, u = y \end{array} \right\}, \tag{3}$$

where $\mathcal{S}$ denotes a suitable parameter space, $u \equiv [u_1, \ldots, u_{n_I}]$, $d \equiv [d_1, \ldots, d_{n_I}]$, and $y \equiv [y_1, \ldots, y_{n_I}]$.

The inverse problem (3) is known to be ill-posed [8]. A popular approach to infer the image σ consists in solving the regularized least squares minimization problem

$$\begin{aligned} (\widehat{u}, \widehat{\sigma}) \quad \in \quad & \arg\min_{u, \sigma \in \mathcal{S}} \quad (1/2)\|Mu - d\|^2 + \tau\phi(\sigma) \\ & \text{s.t.:} \quad \mathcal{A}(\sigma)\, u = y, \end{aligned} \tag{4}$$

where $\|\cdot\|$ is a norm in an appropriate space, ϕ is a regularizer (or prior information, in Bayesian inference terms) that promotes images of parameters σ with desirable properties, and $\tau > 0$ is the so-called regularization parameter setting the relative weight between the regularizer $\phi(\sigma)$ and the data term $(1/2)\|Mu - d\|^2$. The formulation (4) is termed *output least squares*.

Assuming that the inverse operators $\mathcal{A}_i^{-1}(\sigma)$, for $i = 1, \ldots, n_I$, exist, then $\mathcal{A}^{-1}$ exists and, thus, the output least approach (4) is equivalent to the unconstrained least squares formulation

$$\widehat{\sigma} \quad \in \quad \arg\min_{\sigma \in \mathcal{S}} \quad (1/2)\|\mathcal{F}(\sigma) - d\|^2 + \tau\phi(\sigma), \tag{5}$$

where $\mathcal{F}(\sigma) \equiv M\mathcal{A}^{-1}(\sigma)\, y$. Usually, this operator doesn't have a closed form expression and the computation of its gradient and Hessian is complex from the computational point of view.

The complexity involved in solving (4) or (5) and the quality of reconstructed images depend crucially on the type of regularizers used. The quadratic regularizer

$$\phi_2(\sigma) \equiv \int_\Omega (L\sigma)^2,$$

where L is a linear operator, has been widely used (see [9] and references therein), namely because it leads to smooth

Manuscript received October X, 2009. Nuno Brás was sponsored by the Portuguese Ph.D. 309 scholarship program from FCT (SFRH/BD/17405/2004). This work was sponsored by the FCT research project PTDC/EEA-ELC/105333/2008. All authors are with the Instituto de Telecomunicações (IT) and with Instituto Superior Técnico, Universidade Técnica de Lisboa, Portugal. (e-mails: nuno.bras@lx.it.pt, bioucas@lx.it.pt, rcmartins@lx.it.pt, acserra@ist.utl.pt).

objective functions which can be optimized with standard optimization tools based on gradients and Hessians. However, the quadratic regularizer yields poor performance in piece-wise smooth images because the sharper transitions are removed.

The total variation (TV) regularizer [10]

$$\phi_{\text{TV}}(\sigma) \equiv \int_{\Omega} |\nabla \sigma|$$

is arguably the best known and most often used regularizer for piece-wise smooth images. The TV ability to preserve discontinuities and "average" smooth areas is related with its non-differentiability. Aiming at exploiting the TV discontinuity preserving ability and simultaneously using smooth optimization tools based on gradients and Hessians, smooth approximations of TV have been adopted in DPE (see, *e.g.*, [1],[11]).

The interest in TV based regularization and the fact that ϕ_{TV} is convex, although not strictly, have fostered active research in convex optimization to find efficient solution to these problems. The well known ℓ_2-TV deblurring problem is such an example for which effective solutions have been recently introduced [12], [13], [14], [15], [16], [17], [18], [19].

Quadratic and TV regularization belong to the so-called *analysis approach* [20], since they are based on a regularizer that analyzes the image itself. In the *synthesis approach* [20], the original image σ is represented as a linear combination of the elements of some frame, *i.e.,* $\sigma = \mathcal{W}\theta$ where $\mathcal{W}$ is a wavelet frame, and the regularizer is defined on θ. The work [21] on electric current density imaging is representative of this strategy.

A. Related work

Several approaches to the optimization (4) compute saddle points of the respective Lagrangian. Assuming smooth regularizers, for example quadratic ones or smooth approximations to TV, most of these approaches solve the optimality equations by some variant of Newton method usually the Gauss-Newton method. See [22], [23], [24], [25] and references therein.

Computing the saddle point of the Lagrangian of (4) is usually complex because it involves a large number of non-linear coupled equations on three classes of variables: the sought image of parameters σ, the field u, and the Lagrange multipliers which are herein denoted by λ. The availability of parallel computing at accessible costs has fostered the interest in this approach to solve large and complex inverse problems, namely the one in hand. The work [9] addresses a class of overlapping Newton–Krylov–Schwarz algorithms for solving such coupled systems, obtained with a pointwise ordering of the variables.

The augmented Lagrangian method, or method of multipliers, is another direction used solve (4) [11], [26], [27], [28]. In this approach, the constraints in a constrained optimization problem are replaced by penalty terms in the objective function incorporating explicit estimates of the Lagrange multipliers to improve the conditioning of the resulting optimization problem [29, Ch. 17]. With respect to our setting, each iteration of the method consists in minimizing the augmented Lagrangian with respect to (u, σ) followed by an update of the Lagrange multipliers. The former optimization is often replaced by decoupled optimizations with respect to u and σ leading to Uzawa-type schemes termed in [30], [31], [32] alternate direction method of multipliers (ADMM). The decoupled problems are usually smaller and much simpler to solve than original joint optimization with respect to (u, σ).

The unconstrained optimization (5) has also been adopted to infer the image of parameters σ, namely by using Gauss-Newton type methods [33], [8]. However, due to the complex structure of the operator $\mathcal{F}^{-1}$, a large number of forward problems per iteration should be solved to determine the Gauss-Newton Hessian, which makes the method computationally too demanding for 2D (or 3D) systems. When possible, reciprocal methods based on adjoint fields are used to increase each Hessian calculation speed (see, *e.g.* [33], [3] for examples on the adjoint field use in Biomedical and Geophysical applications).

B. Proposed approach and contribution

In this paper, we adopt TV regularization to estimate piece-wise smooth images of parameters, *i.e.*, we assume that the images of parameters are smooth by regions with abrupt transitions between neighboring regions. We follow the constrained formulation (4) and use an instance of ADMM to solve the resulting optimization problem. In line with [14], we convert a complex ℓ_2-TV optimization into two very simple decoupled $\ell_2 - \ell_1$ optimization steps yielding a very simple and fast algorithm. Furthermore, we use the exact TV definition in contrast with most works on DPE that use a smooth TV approximation.

The line of attack we follow, based on ADMM and related to proximal splittings [34] and with split Bregman methods [14], has been increasingly adopted to solve many inverse problems in signal and image processing [16], [35], [17], [36], [18].

Although ADMM is not a new method, the way it has been recently exploited is a novelty: the central idea is to convert an unconstrained optimization problem into a constrained one where the initial variables have been split into new ones. The ADMM method runs an alternate optimization with respect to the new variables. The effectiveness of the obtained algorithm relies on the wisdom in the choice of the new variables.

The resulting algorithm is applied to the particular case of a 2D elliptical equation inverse problem with discontinuous image of parameters, noisy observations, and relatively small observation data compared with the parameter space. In comparison with an ADMM approach using a smooth approximation to TV and smooth optimization tools, we obtain better and faster reconstructions.

The organization of this paper is as follows: Section II formulates the DPE as a constrained optimization problem in a finite dimensional setting. A characterization of the optimization problem in hand is given. Section III presents two instances of the ADMM to solve the optimization problem formulated in Section II. In the first ADMM instance a TV-l2 subproblem is directly solved, whereas in the second instance the TV regularizer is decoupled using a proper variable splitting (proposed version). In Section IV-A a simple 1D

case is implemented and studied. In IV-B a 2D electrical tomographic problem is solved. The reconstruction results are shown and the performance issues are analysed. Finally, Section V presents final remarks and conclusions.

II. Problem Characterization and Formulation

We now formalize problem (4) in a finite dimensional setting. We assume that all the involved entities have been accurately discretized on the same grid with n elements. More details of the discretization are given in Section IV.

The discretized images are arranged in column lexicographic ordering. We address the DPE problem in terms of the following entities, for $i = 1, \ldots, n_I$:

$\mathcal{S}$	$\equiv$	$[k_{\min}, k_{\max}]^n,\ 0 < k_{\min} < k_{\max}$ set of parameter images
$\sigma \in \mathcal{S}$	$\equiv$	image of parameters
$u_i \in \mathbb{R}^n$	$\equiv$	fields u_i
$y_i \in \mathbb{R}^n$	$\equiv$	sources y_i
$M \in \mathbb{R}^{p\times n}$	$\equiv$	observation matrix
$d_i \in \mathbb{R}^p$	$\equiv$	linear and noisy observations
$B \in \mathbb{R}^{q\times n}$	$\equiv$	boundary (for u) selecting matrix.
$C \in \mathbb{R}^{r\times n}$	$\equiv$	boundary (for σ) selecting matrix.
$A(\sigma) \in \mathbb{R}^{n\times n}$	$\equiv$	operator $\mathcal{A}$ for a given σ.
$\overline{A}(u) \in \mathbb{R}^{n\times n}$	$\equiv$	operator $\mathcal{A}$ for a given u.

For compactness, we will denote matrices $A(\sigma) \equiv A_\sigma$ and $\overline{A}(u) \equiv \overline{A}_u$. We may then write $A_\sigma\, u_i = \overline{A}_{u_i}\, \sigma$. As a consequence we have

$$A_\sigma\, u = y \quad \Leftrightarrow \quad \overline{A}_u\, \sigma = y\!:, \tag{6}$$

where $y\!: \equiv \mathrm{vec}(y)$, *i.e.,* $y\!:$ is the column-wise lexicographic ordering of y, and $\overline{A}_u \equiv [\overline{A}_{u_1}^T, \overline{A}_{u_2}^T, \ldots, \overline{A}_{u_{n_I}}^T]^T$.

The main source of difficulty in DPE problems is that the operators A_σ and $\overline{A}_{u_i}$ are either ill-posed or ill-conditioned. To give an idea of these difficulties, we built the matrices A_σ and $\overline{A}_{u_i}$ corresponding to a 1D problem with $n = 256$, $n_I = 1$, and $(y)_i = 1$ for $i = 128$ and $(y)_i = 0$ for $i \neq 128$ (the notation $(y)_i$ stands for the ith component of vector y. The derivatives were discretized using finite differences[1] and the boundary conditions $(\sigma)_1 = 1$ and $(\sigma)_{256} = 1$ were enforced. The plot on the top of Fig. 1 is the field obtained by solving the linear system $A_\sigma u = y$ for the original piecewise constant vector of parameters σ shown in the bottom of the same figure. The original vector σ is exactly recovered by solving the system $\overline{A}_u \sigma = y$ using the field u obtained before (with the boundary conditions above referred). However, if u is perturbed, even very slightly, the obtained values of σ by solving the same problem are of little use. The noisy estimate of σ shown in the bottom of Fig. 1 is obtained by solving $\overline{A}_{u+\delta u}\sigma = y$, where δu is a noise vector of independent and identical distributed (i.i.d.) uniform samples in the interval $[-a, a]$ with

$$a \equiv \varepsilon \frac{\|u\|_F}{\sqrt{n_I}}, \tag{7}$$

[1]We have dropped the index 1 in u_1 because we have just one set of observations in this example.

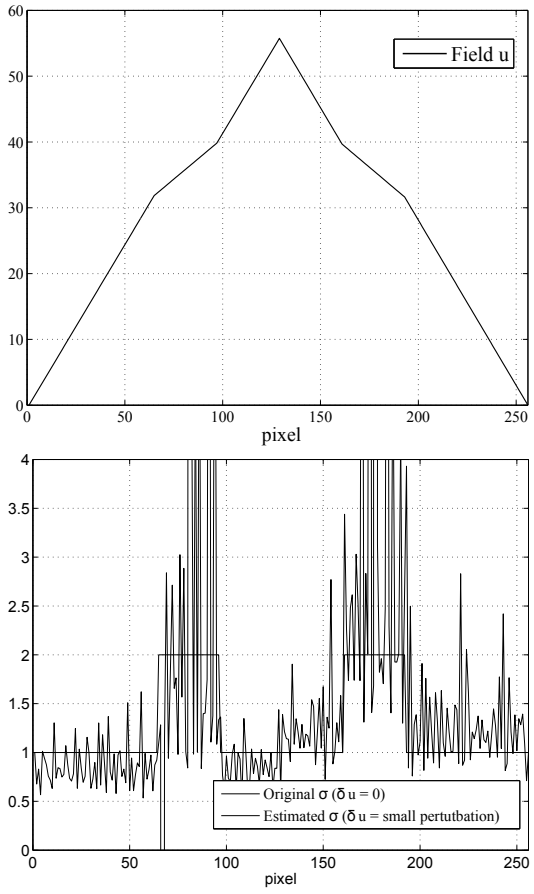

Fig. 1. Illustration of the effect of a small perturbation in the estimation of σ. Top: field u obtained by solving the system $A_\sigma u = y$. Bottom: vector of Paremeters σ obtained by solving $\overline{A}_{(u+\delta u)}\sigma = y$ for $\delta u = 0$ and δu corresponding to zero-mean iid uniformly distributed samples in the interval [-a a] where $a = 0.003\,\|u\|_F/\sqrt{n}$.

where $\|\cdot\|_F$ stands for Frobenious norm and $\varepsilon > 0$ controls the relative power of noise with respect to the power of u, as in [11]. In Fig. 1 we set $\varepsilon = 0.003$.

To further study the results plotted in Fig.1, we define $\delta\sigma \equiv \widehat{\sigma} - \sigma$, where $\widehat{\sigma}$ and σ denote, respectively, the estimated and the original vector of parameters, and $\delta A \equiv \overline{A}_{u+\delta u} - \overline{A}_u$. Neglecting second order variations, we have $\overline{A}_u \delta\sigma = -\delta A\, \sigma$. Therefore, the size of $\delta\sigma$ is determined by the amplification introduced by the inverse of the singular values of $\overline{A}_u$. For the considered example, these singular values decay at an approximate linear rate from 1 to 10^{-3}. We may then anticipate large perturbations $\delta\sigma$ even for small field perturbations δu. The described ill-conditioning of the DPE inverse problem is, very often, even worse because we have access to u only through a low rank measurement matrix M.

As in many inverse problems, we use regularization to promote solutions with desirable characteristics and, in this way, mitigate the ill-posedness or ill-conditioning of the DPE problem. As already referred to, we are particularly interested in piece-wise smooth images of parameters to which the TV regularizer is very well suited. Hence, we are interested in

solving the optimization problem

$$(\widehat{u},\widehat{\sigma}) \in \arg\min_{u,\sigma\in\mathcal{S}} (1/2)\|Mu-d\|_F^2+\tau\phi_{\text{TV}}(\sigma) \quad (8)$$
$$\text{s.t.:} \quad A_\sigma\, u = y, \quad B\, u = 0, \quad C\,\sigma = \sigma_0.$$

where $u \equiv [u_1,\ldots,u_{n_I}]$, $d \equiv [d_1,\ldots,d_{n_I}]$, and $y \equiv [y_1,\ldots,y_{n_I}]$. The constrains $B\,u=0$ and $C\sigma=\sigma_0$ impose, respectively, $u=0$ and $\sigma=\sigma_0$ at the boundary.

With respect to the discretization of the TV regularizer, we adopt the discretized TV defined as [12]

$$\phi_{\text{TV}}(\sigma) \equiv \sum_{i=1}^{n}\sqrt{(D_1\,\sigma)_i^2+(D_2\,\sigma)_i^2}, \quad (9)$$

where D_1 and D_2 are first order finite difference matrices operating on two orthogonal directions, which enforces piece-wise constant solutions. We note, however, that higher order differences may be used. For example, second order differences enforces piece-wise linear solutions.

Assuming that the inverse forward operator A^{-1} exists, then the feasible set

$$\{(\sigma,u)\in\mathcal{S}^n\times\mathbb{R}^{n_I}\,|\,A_\sigma\,u=y,\ B\,u=0,\ C\,\sigma=\sigma_0\}$$

in the DPE problem (8), is compact. This is consequence of the compactness of $\mathcal{S}$ and of the continuity of A^{-1}, thus mapping closed and bounded sets into closed and bounded sets. Therefore, Weierstrass's theorem [37] guarantees the existence of an optimum solution for the DPE problem (8).

III. Proposed ADMM method

In this section, we present two instances of the ADMM method to solve the DPE problem (8). The first, ADMM (version 1), results naturally from the structure of problem (8). This algorithm, here included for comparison purposes, belongs to a family of state-of-the-art ADMM methods introduced in [11], [26], [38]. The second, DIPESAL, besides faster, is able to better enforce the regularization. It results from rewriting the TV regularizer in an equivalent constrained form.

The augmented Lagrangian plays a central role in the ADMM method [30], [31], [32]. For the problem (8), the augmented Lagrangian can be written as

$$\mathcal{L}(u,\sigma,w,\lambda) \equiv \psi(u)+\tau\phi_{\text{TV}}(\sigma)+\iota_{\mathcal{S}}(w) - \sum_{i=1}^{4}\lambda_i^T c_i(u,\sigma,w)+(\mu/2)\|c_i(u,\sigma,w)\|^2, \quad (10)$$

where $\psi(u)\equiv(1/2)\|Mu-d\|_F^2$, $\iota_{\mathcal{S}}:\mathbb{R}_n\to\{0,\infty\}$ is the indicator function of set $\mathcal{S}$ defined as $\iota_{\mathcal{S}}(w)=0$ if $w\in\mathcal{S}$ and $\iota_{\mathcal{S}}(w)=\infty$ if $w\notin\mathcal{S}$, λ_i, for $i=1,\ldots,4$, are the Lagrange multiplier vectors for the equality constraints c_i given by

$$\begin{aligned} c_1(u,\sigma,w) &\equiv A_\sigma u - y \Leftrightarrow \overline{A}_u\sigma - y\text{:}\\ c_2(u,\sigma,w) &\equiv Bu\\ c_3(u,\sigma,w) &\equiv C\sigma-\sigma_0\\ c_4(u,\sigma,w) &\equiv \sigma - w, \end{aligned} \quad (11)$$

and $\mu>0$ controls the weight of the quadratic penalty terms. Notice that the constraint $\sigma\in\mathcal{S}$ is enforced by the inclusion of the indicator function $\iota_{\mathcal{S}}(w)$ in (10) and the equality constraint $\sigma-w$.

Completing the square in (10) allows to redefine this function as

$$\mathcal{L}(u,\sigma,w,e) = \psi(u)+\tau\phi_{\text{TV}}(\sigma)+\iota_{\mathcal{S}}(w) + \sum_{i=1}^{4}(\mu/2)\|c_i(u,\sigma,w)-e_i\|^2+a(e_i), \quad (12)$$

where $e_i=\lambda_i/\mu$ and $a(e_i)=-{e_i}^2$ are terms not depending on (u,σ,w).

Algorithm 1 ADMM (version 1)

Require: $k=0$, choose $\mu>0$, u^0, w^0, and e_i^0, for $i=1,\ldots,4$

1: **Repeat**
2: $\quad\sigma^{k+1}\in\arg\min_{\sigma}\mathcal{L}(u^k,\sigma,w^k,e^k)$
3: $\quad u^{k+1}\in\arg\min_{u}\mathcal{L}(u,\sigma^{k+1},w^k,e^k)$
4: $\quad w^{k+1}\in\arg\min_{w}\mathcal{L}(u^{k+1},\sigma^{k+1},w,e^k)$
5: $\quad e_i^{k+1}\leftarrow e_i^k-c_i(u^{k+1},\sigma^{k+1},w^{k+1})$, for $i=1,\ldots,4$
6: $\quad k\leftarrow k+1$
7: **Until** some stopping criterion is satisfied.

Algorithm 1 shows the pseudo-code for an augmented Lagrangian method with alternating direction minimization, aimed at computing the solution of (8). Step 5 corresponds to the Lagrange multipliers vector update. If all the constraints were linear, this algorithm would belong to the ADMM class[2] [30], [39], [40], [35].

However, the conbstraint $A_\sigma\,u=y$ in 8 is bilinear and thus, our problem does not belong the canonical ADMM class. Anyway, given the strong similarity between the proposed and the ADMM algorithms, we still term the former as ADMM, as in the recent work [40], were the authors use the term ADMM Bi-convex Problems.

Each iteration of Algorithm 1 implements alternating direction optimization with respect to u, σ, and w, and updates the estimate of the Lagrange multipliers. The optimization with respect to σ (line 2) amounts to solve the ℓ_2-TV regularization problem

$$\sigma^{k+1}=\arg\min_{\sigma}\left\{(1/2)\,\sigma^T F^k\sigma-(\nu_\sigma^k)^T\sigma+(\tau/\mu)\,\phi_{\text{TV}}(\sigma)\right\}, \quad (13)$$

where

$$\begin{aligned} F^k &= \overline{A}_{u^k}^T\overline{A}_{u^k}+C^TC+I\\ \nu_\sigma^k &= \overline{A}_{u^k}^T(y+e_1^k)+C^T(\sigma_0+e_3^k)+(w^k+e_4^k). \end{aligned} \quad (14)$$

Problem (13), although convex, is non-smooth, preventing the use of off-the-shelf gradient and Hessian based tools. Very

[2]ADMM algorithms solve the problem

$$(\widehat{x},\widehat{z}) \in \arg\min_{x,z} f(x)+g(z) \quad \text{s.t.:} \quad Ax+Bz=c.$$

often, this difficulty has been circumvented by replacing the TV regularizer with the smooth approximation

$$\sum_{i=1}^{n} \sqrt{(D_1\,\sigma)_i^2 + (D_2\,\sigma)_i^2 + \gamma}, \tag{15}$$

where γ is a small positive number. This has been followed, for example, in [11], [38], where an approximated solution of (13) is computed by a one step lagged diffusivity fixed point iteration [41]. In the ADMM version 1 algorithm, this was the considered approach to minimize σ.

The optimization with respect to u (line 3) is a quadratic problem, with solution given by

$$u^{k+1} \;=\; \left(M^T M + \mu A_{\sigma^{k+1}}^T A_{\sigma^{k+1}} + \mu B^T B\right)^{-1} \nu_u^k \tag{16}$$

where

$$\nu_u^k \;\equiv\; \left(M^T d + \mu A_{\sigma^{k+1}}^T (y + e_1^k) + \mu B^T e_2^k\right). \tag{17}$$

The optimization with respect to w (line 4) amounts to compute

$$w^{k+1} = \arg\min_{w} \left\{\iota_{\mathcal{S}}(w) + (\mu/2)\|w - \nu_w^k\|^2\right\}, \tag{18}$$

where

$$\nu_w^k = \sigma^{k+1} - e_4^k.$$

Problem (18) is the projection of ν_w^k on the convex set $\mathcal{S}$, denoted by $P_{\mathcal{S}}(\nu_w^k)$, where $P_{\mathcal{S}} : \mathbb{R}^n \to \mathcal{S}$ is defined as

$$P_{\mathcal{S}} : x \mapsto \max\{k_{min}, \min\{x, k_{max}\}\} \tag{19}$$

and the max and min are componentwise.

A. Two more constraints. A simpler problem

In DPE problems, the ℓ_2-TV optimization (13) has been frequently attacked by the lagged diffusivity fixed point iteration. In recent years, large research efforts have been devoted to the ℓ_2-TV problem and many other convex non-smooth problems. We mention the seminal [12], which applies when F^k is the identity. The work [13], which elaborates on [12], solves (13) for data terms $\|K\sigma - \nu\|^2$, where K is a generic matrix. These iterative algorithms are closely related with the so-called forward-backward splitting and Douglas-Rachford splitting methods [34], which are known to have low speed of convergence when the operator K is ill-posed or bad-conditioned. This aspect have been actively researched leading to faster ℓ_2-TV algorithms [15], [16], [14], [17], [18], [19]. We note, however, that solving exactly the sequence of sub-problems (13) may not be a wise approach for two reasons: firstly, it will make the algorithm too slow; secondly, these problems vary from ADMM iteration to iteration.

In order to replace the $\ell_2 - TV$ with simple closed-form subproblems, in each ADMM iteration, we now replace the initial DPE problem (8) with the equivalent one

$$\begin{aligned}(\widehat{u}, \widehat{\sigma}, \widehat{z}) \;\in\; & \arg\min_{u,\sigma\in\mathcal{S},z} \quad (1/2)\|Mu - d\|_F^2 + \tau\|z\|_1 \\ & \text{s.t.: } A_\sigma\, u = y \quad B\, u = 0, \quad C\,\sigma = \sigma_0 \\ & \quad D_1\,\sigma = z_1, \quad D_2\,\sigma = z_2\end{aligned} \tag{20}$$

where $z \equiv [z_1 z_2] \in \mathbb{R}^{n\times 2}$ and

$$\|z\|_1 \equiv \sum_{i=1}^{n} |(z)_i| = \sum_{i=1}^{n} \sqrt{z_{1,i}^2 + z_{2,i}^2},$$

i.e., $\|z\|_1$ is the ℓ_1 norm of z. Comparing (20) with (8), the TV regularizer applied to σ was replaced with the ℓ_1 norm of z and two more constraints were introduced. The augmented Lagrangian for the new problem (20) can be written as

$$\begin{aligned}\mathcal{L}(u,\sigma,w,z,e) \;=\; & \psi(u) + \tau\|z\|_1 + \iota_{\mathcal{S}}(w) \\ & + \sum_{i=1}^{6} (\mu/2)\|c_i(u,\sigma,w,z) - e_i\|^2 + a(e_i),\end{aligned} \tag{21}$$

where the equality constraints c_i are given by

$$\begin{aligned} c_1(u,\sigma,w,z) &\equiv A_\sigma u - y \Leftrightarrow \overline{A}_u \sigma - y\text{:} \\ c_2(u,\sigma,w,z) &\equiv Bu \\ c_3(u,\sigma,w,z) &\equiv C\sigma - \sigma_0 \\ c_4(u,\sigma,w,z) &\equiv \sigma - w \\ c_5(u,\sigma,w,z) &\equiv D_1\sigma - z_1 \\ c_6(u,\sigma,w,z) &\equiv D_2\sigma - z_2. \end{aligned} \tag{22}$$

Algorithm 2 DIPESAL

Require: $k = 0$, choose $\mu > 0$, u^0, w^0, z^0, and e_i^0, for $i = 1,\ldots,6$

1: **Repeat**
2: $\quad \sigma^{k+1} \in \arg\min_{\sigma} \mathcal{L}(u^k, \sigma, w^k, z^k, e^k)$
3: $\quad u^{k+1} \in \arg\min_{u} \mathcal{L}(u, \sigma^{k+1}, w^k, z^k, e^k)$
4: $\quad w^{k+1} \in \arg\min_{w} \mathcal{L}(u^{k+1}, \sigma^{k+1}, w, z^k, e^k)$
5: $\quad z^{k+1} \in \arg\min_{z} \mathcal{L}(u^{k+1}, \sigma^{k+1}, w^{k+1}, z, e^k)$
6: $\quad e_i^{k+1} \leftarrow e_i^k - c_i(u^{k+1}, \sigma^{k+1}, w^{k+1}, z^{k+1})$, for $i = 1,\ldots,6$
7: $\quad k \leftarrow k+1$
8: **Until** some stopping criterion is satisfied.

Algorithm 2 shows the pseudo code for the proposed ADMM algorithm, called distributed parameter estimation via splitting and augmented Lagrangian (DIPESAL) to solve (20). Each iteration implements alternating direction optimization with respect to σ, u, w, and z and updates the estimate of the Lagrange multipliers e_i, for $i = 1,\ldots,6$. The optimization with respect to σ (line 2) is a quadratic problem with solution

$$\sigma^{k+1} = \left(\overline{A}_{u^k}^T \overline{A}_{u^k} + C^T C + I + D_1^T D_1 + D_2^T D_2\right)^{-1} \nu_\sigma^k, \tag{23}$$

with

$$\begin{aligned}\nu_\sigma^k \;\equiv\; & \left(\overline{A}_{u^k}^T (y\text{:} + e_1^k\text{:}) + C^T(\sigma_0 + e_3^k) + (w^k + e_4^k)\right. \\ & \left. + D_1^T(z_1^k + e_5^k) + D_2^T(z_2^k + e_6^k)\right).\end{aligned}$$

The optimization with respect to u (line 3) is also a quadratic problem, whose solution is (16).

The optimization with respect to w (line 4) is, as in ADMM (version 1), the projection

$$w^{k+1} = P_{\mathcal{S}}(\nu_w^k), \tag{24}$$

where $P_{\mathcal{S}}$ is defined in (19) and

$$\nu_w^k \equiv \sigma^{k+1} - e_4^k.$$

The optimization with respect to z (line 5) amounts to solve the $\ell_2 - \ell_1$ decoupled problem

$$z^{k+1} = \arg\min_z \left\{ (1/2)\|z - \nu_z^k\|_F^2 + (\tau/\mu)\|z\|_1 \right\}, \quad (25)$$

where

$$\nu_z^k \equiv [\nu_{z_1}^k \; \nu_{z_2}^k] \in \mathbb{R}^{n\times 2}$$

and

$$\begin{array}{rcl} \nu_{z_1}^k & \equiv & D_1\,\sigma^{k+1} - e_5^k \\ \nu_{z_2}^k & \equiv & D_2\,\sigma^{k+1} - e_6^k. \end{array} \quad (26)$$

The solution for (25) is [14]

$$z^{k+1} = \text{soft}(\nu_z^k, \tau/\mu), \quad (27)$$

where $\text{soft}(\nu_z^k, \tau/\mu)$ operates pixel-wise and, for $x \in \mathbb{R}^2$ and $\delta > 0$, we have

$$\begin{array}{ll} \text{soft}(x,\delta) \equiv \frac{x}{|x|}\max(|x| - \delta, 0) & \Leftarrow x \neq (0,0) \\ \text{soft}(x,\delta) \equiv 0 & \Leftarrow x = (0,0). \end{array} \quad (28)$$

Each iteration of the DIPESAL algorithm implements two quadratic optimizations, with respect to σ and to u, and two very simple pixel-wise decoupled optimizations with respect w and z . In conclusion, each iteration of the DIPESAL is lighter than that of the original ADMM, version 1 and has a closed description.

B. Over-Relaxation

It is known that it should be possible for over-relaxation to accelerate convergence in split based algorithms for convex programming scenarios. Such acceleration has been experimentally confirmed in some applications of the method of multipliers (see [30] and references therein). It consists on an over-evaluation of the updated parameter in relation to what was prescribed by each sub-problem. An over-relaxation of the parameter σ is considered here, simply by imposing a new update rule given by

$${\sigma_*}^{k+1} = \sigma^k + \zeta\Delta\sigma^{k+1} \quad (29)$$

where $\zeta > 1$ and $\Delta\sigma^{k+1}$ is the difference between the current estimate σ^k and the resolved σ^{k+1}. The over-relaxation can theoretically be applied to each sub-problem although the best results were found for the σ sub-problem (see section IV).

C. Complexity

The computational complexity involved on each iteration of DIPESAL id the following:

step 2: $\mathcal{O}(n^2)$ complexity for computing and approximate solution for linear system of size n;

step 3: $\mathcal{O}(n^2)$ complexity for computing and approximate solution for linear system of size n;

steps 4, 5, 6: $\mathcal{O}(n)$ complexity associated with the application of the operators A, $\overline{A}_u$ D_1 and D_2.

We have used a fixed number of PCG iterations to compute the approximated solutions of the linear system in steps 2 and 3. The total complexity per iteration is thus $\mathcal{O}(n^2)$.

D. ADMM convergence

For convex problems with linear constraints, the convergence of the ADMM method was proved in [30], [39], [42]. In the presented problem, the constraint $A_\sigma u = y$ is not linear, and therefore, those convergence results cannot be invoked. For a while let us assume, however, that $A_\sigma u$ is linear. Note that $A_\sigma u$ is indeed approximately linear for small variations of u and σ because it is C^∞. For small variations, with respect to σ' and u', and considering the number of sources $n_I = 1$, just for keeping the notation light, we have then

$$A_\sigma u \simeq a + A_1\sigma + A_2 u,$$

where $a = A_{\sigma'}u'$, $A_1 = \overline{A}_{u'}$, and $A_2 = A_{\sigma'}$. The set of equality constraints (22) $c_i(u,\sigma,w,z) = 0$, for $i = 1,\ldots,6$, are then equivalent to

$$G\sigma + Hv = b,$$

where $v = [u^T, w^T, z_1^T, z_2^T]^T$ and

$$G \equiv \begin{bmatrix} A_1 \\ 0 \\ C \\ I \\ D_1 \\ D_2 \end{bmatrix}, H \equiv \begin{bmatrix} A_2 & 0 & 0 & 0 \\ B & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & -I & 0 & 0 \\ 0 & 0 & -I & 0 \\ 0 & 0 & 0 & -I \end{bmatrix}, b \equiv \begin{bmatrix} y - a \\ 0 \\ \sigma_0 \\ 0 \\ 0 \\ 0 \end{bmatrix}.$$

Matrix G has full column rank because it is block column and the identity is one of its blocks. Matrix H is also full column rank because the block $[A_2^T B^T]^T$ corresponding, respectively, to the discretized differential equation (1) and to the boundary conditions, uniquely defines the field u. As a consequence, and considering that ψ, $\|\cdot\|_1$ and $\iota_{\mathcal{S}}$ are closed proper convex functions, Theorem 3.1 of [39] (a small modification of Theorem 8 of [30]) ensures that, if there exists a saddle point of the augmented Lagrangian $L(u,\sigma,w,z,e)$ (20), then $u^k \to u^*$, $\sigma^k \to z^*$, $w^k \to w^*$, $z^k \to z^*$, and $e^k \to e^*$, where $(u^*,\sigma^*,w^*,z^*,e^*)$ is such a saddle point. On the other hand, if no such saddle point exists, then at least one of the sequences $\{u^k\}$, $\{w^k\}$, $\{z^k\}$, or $\{e^k\}$ is unbounded.

The above convergence results are valid only in a small neighborhood of a saddle point. A more general result, assuming that $M = I$ *i.e.*, all the field is observed, is stated in [11]. The authors of [11] note that if a uniform finite element mesh is used in the discretization, the TV-norm is equivalent to Sobolev norms $H^k(\Omega)$, $k = 1, 2$, and then the techniques used in [43] can be employed to show that the augmented Lagrangians $\mathcal{L}$ (10) and (21) have unique saddle points, if the observation error $\|u - d\|$ is sufficiently small. Furthermore, the ADMM iterates converge to the saddle point at a linear rate. However, if the observation error $\|u - d\|$ is large, only a subsequence of the sequence generated by the ADMM method converges to a saddle point.

In this paper, we are interested in DPE problems where very few observations are available, *i.e.*, where matrix M has much less rows than columns. In this case, the convergence results stated in [11] can not be invoked. However, we have always observed convergence as long as the parameter μ, weighting the augmented Lagrangian term, is larger than the

minimum threshold. This behavior is in line with the Proposition 4.2.3, [37], in the context of augmented Lagrangian methods, which gives conditions under which there is a minimizer of $\mathcal{L}(\sigma, w, u, \lambda)$ that lies close to a local solution (σ^*, w^*, u^*) and gives error bounds on both (σ^k, w^k, u^k) and the update multiplier λ^{k+1}. This proposition assumes however, smoothness of the augmented Lagrangian such that the penalty parameter can be chosen large enough to make its Hessian positive definite.

IV. Results

In this section, the effectiveness of the proposed ADMM algorithm, DIPESAL, is illustrated by solving 1D and 2D simulated DPE problems with piecewise smooth parameters. Different levels of noise, number and type of observations provided by the observation matrix M are considered.

A. 1D Experiments

Consider an 1D elliptic partial differential equation similar to (1) but defined in $\Omega \equiv [0, 1]$. We sample Ω using $n = 256$ uniformly spaced samples, thus with a sampling interval of $h = 1/(n-1)$. The derivatives are approximated with finite differences; The TV regularizer is given by (9) with $D_2 = 0$, *i.e.*,

$$\phi_{\text{TV}}(\sigma) \equiv \sum_{i=1}^{n} |(D_1\,\sigma)_i| = \|D_1\,\sigma\|_1. \tag{30}$$

All the steps of Algorithm 2 apply to 1D problems with minor modifications, namely in the optimization with respect to z (step 5) which is still given by (28) with $\nu_{z_1}^k$ in place of ν_z^k.

To evaluate the robustness of the proposed algorithm to perturbations in the measurements, noise is added to the field u, using the same setup as in [11]: the noise is independent and identical distributed (i.i.d.) with uniform distribution in the interval $[-a, a]$ where (7) is used.

In all examples the regularization parameter is hand tuned for optimal performance measured by the reconstruction error

$$\text{Rerr} \equiv \frac{\|\widehat{\sigma} - \sigma\|}{\|\sigma\|}. \tag{31}$$

1) Full measurements: In this subsection, we consider the case where $M = I$, *i.e.*, we have measurements of the field u at every discretized coordinate.

Figure (2) left shows a piecewise smooth parameter function with values ranging between 1 and 4. The right part of the same figure shows three sources y_i, for $i = 1, 2, 3$, scaled by h for representation purposes, and the respective field u_i obtained by solving the system

$$A_\sigma u = y, \quad Bu = 0.$$

Figure (3) top shows the the original parameter function and its estimate for $\varepsilon = 0.001$. The regularization parameter and the reconstruction error are $\tau = 0.01$ and Rerr $= 0.02$, respectively. The bottom part of the same figure shows the norms, as a function of the algorithm iterations, of the constraints c_1 and c_5 and of the $\nabla_u \mathcal{L}$.

Figure 4 top is as Fig. 3 top for $\varepsilon = 0.01$ and for $\varepsilon = 0.1$. As expected, the parameter estimates degrade as ε increases.

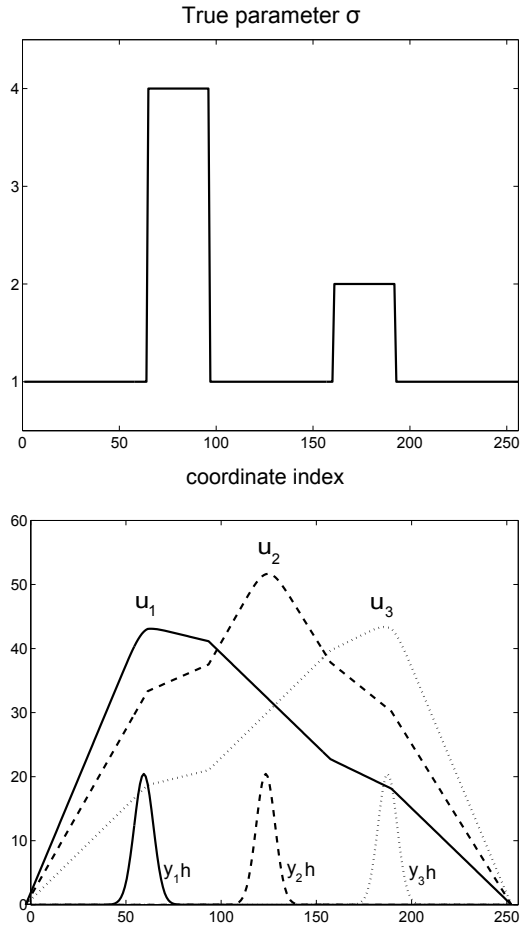

Fig. 2. Top: Piecewise smooth parameter function with values ranging between 1 and 4. Bottom: Scaled (for representation purposes) sources y_i, for $i = 1, 2, 3$ and the respective fields u_i.

This degradation can be mitigated by increasing the number of sources n_I. This is illustrated with the parameter estimate shown in the same figure for $\varepsilon = 0.1$ and $n_I = 9$, which yields Rerr $= 0.130$, whereas the estimate for $\varepsilon = 0.1$ and $n_I = 3$ yields Rerr $= 0.173$.

Figure 4 bottom plots the noisy measurements d_i, the estimated fields u_i, and the original fields u_i for $\varepsilon = 0.1$ and $n_I = 3$. In spite of the highly noisy measurements, the estimated fields are quite accurate.

2) Partial measurements: In this section we take $n_I = 3$, $\varepsilon = 0.001$ but consider partial measurements. This is a scenario of practical relevance because in most DPE problems we have only access to very few field measurements. The degree of sparsity in the measurements is controlled by the number of rows, p, of matrix $M \in \mathbb{R}^{p \times n}$.

Figure (5) shows estimation results, using measurement matrices using Gaussian shaped rows. The parameter function has more discontinuities than in the previous examples and, therefore, it is difficult to reconstruct from partial measurements. As expected, the quality of the parameter estimates degrades as the number of observations decreases. Nevertheless, we call attention to the quality of the estimation obtained with just $p = n/10 \simeq 25$ measurements. For $p = n/20$, there a clear loss of quality in the corresponding estimate. Notice,

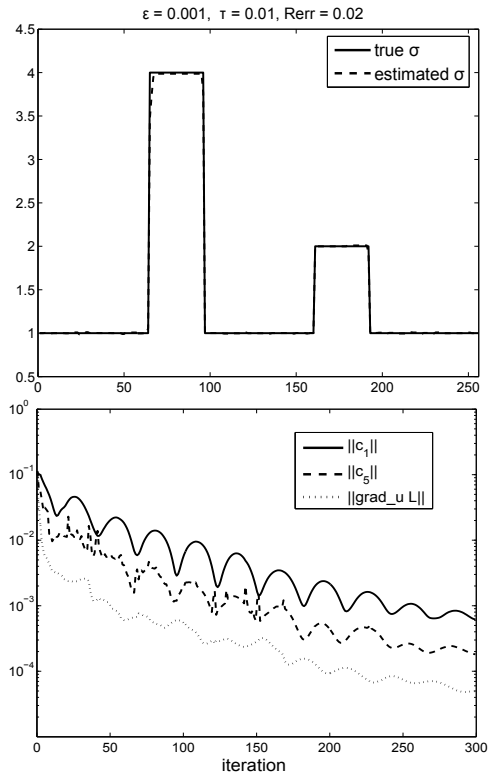

Fig. 3. Top: Original (solid line) and estimated (dashed line) parameters for $\varepsilon = 0.001$ (noise controlling parameter), $\tau = 0.01$ (regularization parameter), leading to a reconstruction error of Rerr = 0.02. Bottom: Evolution of the norm constraints c_1 and c_5 and the norm of $\nabla_u \mathcal{L}$ along the algorithm iterations.

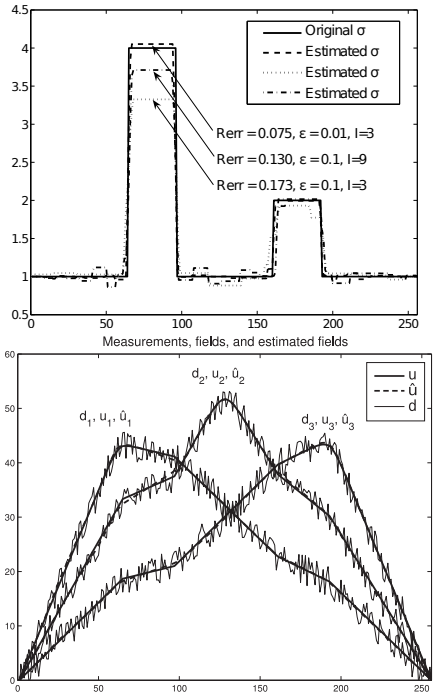

Fig. 4. Top: Original (solid line) and estimated (dashed lines) parameters for ($\varepsilon = 0.01$, $\tau = 0.2$, $n_I = 3$), ($\varepsilon = 0.1$, $\tau = 8$, $n_I = 3$), and ($\varepsilon = 0.1$, $\tau = 8$, $n_I = 9$). The reconstruction errors of this estimated parameters are, respectively, 0.075, 0. 173, and 0.130. Bottom: For $\varepsilon = 0.1$ and $n_I = 3$, and for $i = 1, 2, 3$, the noisy measurements d_i, the estimated fields u_i, and the original fields u_i.

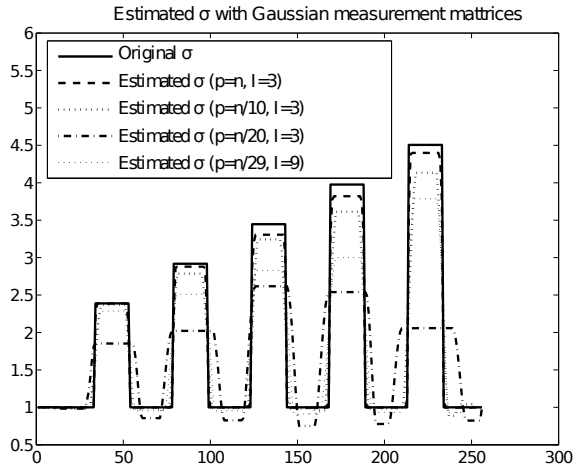

Fig. 5. Parameter estimates based on measurements obtained with measurement matrices using Gaussian shaped rows. The reconstruction errors are Rerr = 0.06 for $p = n, n_I = 3$ (p denotes the number of measurements), Rerr = 0.17 for $p = n/10, n_I = 3$, Rerr = 0.43 for $p = n/20, n_I = 3$, and Rerr = 0.23 for $p = n/20, n_I = 9$; the parameters (τ, μ) are set to, respectively, ($\tau = 0.1$, $\mu = 100\tau$), ($\tau = 0.01$, $\mu = 100\tau$), ($\tau = 0.01$, $\mu = 100\tau$), ($\tau = 0.005$, $\mu = 100\tau$).

however, the large improvement obtained by increasing the number of sources from $n_I = 3$ to $n_I = 9$, in the case we have $p = n/20$; the reconstruction error increases from Rerr = 0.43 to Rerr = 0.23, illustrating that the lack of measurements may be mitigated by increasing the number of sources, which can be achieved carrying out more experiments.

3) Convergence: Concerning convergence, and in line with convergence results stated in the previous Section, we have observed that the proposed algorithm always converges, provided that the parameter μ controlling the weight of quadratic terms $\|c_i - e_i\|^2$ in the augmented Lagrangian is large enough. Furthermore, the speed with which the constraints c_i approach to zero increases with μ. However, if μ is too large, the convergence speed of the algorithm is slow. We have verified experimentally that μ should vary linearly with the regularization parameters, *i.e.*, $\mu = \alpha\tau$, where α depends on h and on the values of the fields on of the parameters. Lately, in [40] it is proposed an adaptive updating of μ based on the primal and dual residuals. Its application to this problem should be subject of further research. For the 1D experiments we set $\alpha \in [50, 500]$.

B. The 2D Scenario

In this section, a real world 2D elliptic DPE problem is presented to assess Algorithms 1 with a fixed point iteration to obtain an approximated solution of (13) and the presented DIPESAL, following the knowledge obtained from the 1D experiments.

We consider an electrostatic elliptic DPE problem with homogeneous boundary conditions, where the underlying forward problem is stated by,

$$\begin{aligned} \nabla \cdot \sigma \nabla u_i &= -\dot{\rho}_i \text{ in } \Omega \subset \mathbb{R}^2 \quad i = 1, \ldots, n_I \\ u_i &= 0 \text{ on } \partial\Omega, \end{aligned} \tag{32}$$

where u_i is the (electric) scalar potential, $\dot{\rho}_i$ are the n_I source currents and σ is the inhomogeneous conductivity map.

Although boundary conditions and sources could be arranged in different manner depending on the application, the forward problem is generically the same. Actually, the choice of other boundary conditions and sources is straightforward.

In the 2D case, the discretization is done using a staggered finite "volume" formulation (see e.g. [44]). The involved fields and parameters are defined in the cell centre and in the cell edges, as presented in Fig. 6:

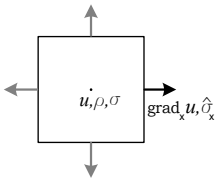

Fig. 6. The used 2D finite "volume" stencil and the space definition of each used variable.

The resulting discretized equation is given by

$$\texttt{div}\ (P\sigma \texttt{grad}\ u_i) = -\dot{\rho}_i, \tag{33}$$

where σ is considered to be isotropic, represented by a diagonal matrix. P is an average matrix such that $P\sigma = \tilde{\sigma}$, defined over the edges. The div matrix corresponds to a discrete divergence matrix and $\texttt{grad}_k = \texttt{div}_k{}^T$ where k is the coordinates x and y. The following conditions define the forward simulation context:

1) The problem is defined in a square with side length 100 points (10000 simulations points) with unitary length in both x and y directions.
2) Dirichlet homogeneous boundary conditions for u_i and no boundary conditions for σ are used.
3) 36 Dirac "sources" placed as described in Fig. 7, with unitary values and switched on and off independently.
4) M, the observation matrix given again by $M \in \mathbb{R}^{p\times n}$, has $p = 80$, acquired according to Fig. 7.
5) $d = Mu + \eta$, where η is defined as white noise with 40 dB to 26 dB SNR (between 1 % to 5 % of the mean value of d).
6) σ is depicted in Fig. 7.

The reconstruction is obtained from $p = 80$ and $n_I = 36$, used to invert a parameter space (n) of 10000 points ($p = n/125$ and $n_I = 36$) with several noise levels from ($\varepsilon = 0.01$ to $\varepsilon = 0.05$). Moreover, none of the sensing points is placed in the interior of the heterogeneous zone of the σ map, which gives a tomographical flavour to the process. In practice, it was seen that this makes the problem even more difficult to invert.

The described DPE problem is indeed a complex scenario, joining the ill-posedness of the underlying parameter estimation problem with the sparseness of the retrieved information from the fields. The algorithms ran during a fixed amount of iterations (200). This was the stopping criterion used in both cases.

In Fig. 8, the reconstruction with the ADMM (version 1) to obtain an approximated solution of (13) is presented for $\varepsilon = 0.05$. The following parameters were used:

1) μ (Constrains penalty parameters)=5 $\times$ 10^{-11}.

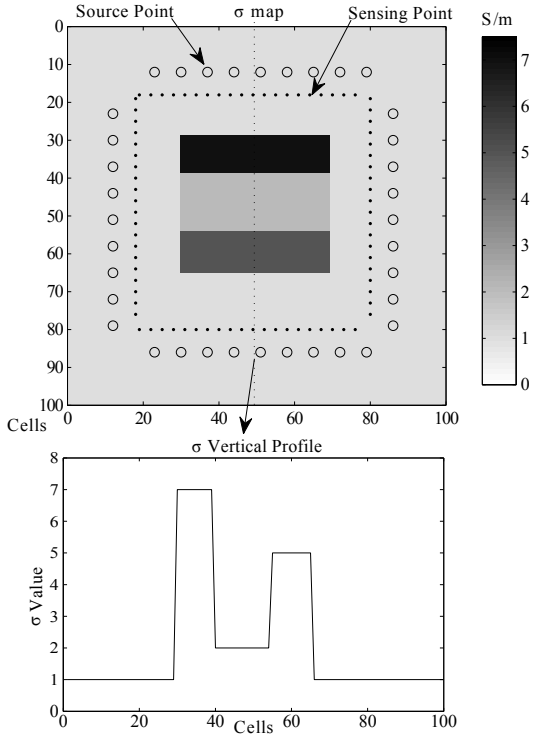

Fig. 7. Top: Sensing and source points used in a 100 points square side piecewise smooth σ map ranging from 1 to 7 S/m. Bottom: Vertical σ profile

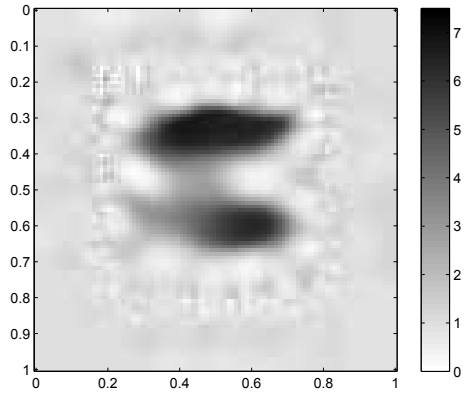

Fig. 8. ADMM (version 1) reconstruction in the 2D square with 100 points square side for $\varepsilon = 0.05$

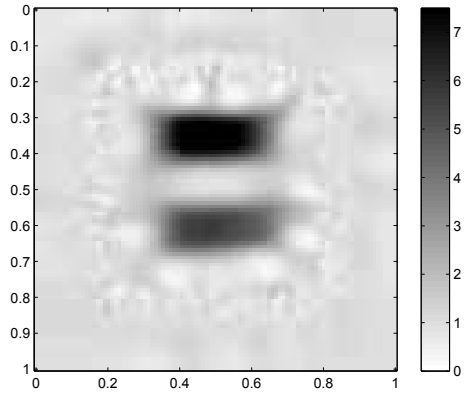

Fig. 9. DIPESAL method reconstruction in the 2D square with 100 points square side $\varepsilon = 0.05$

2) τ (Regularization parameter)=5×10^{-4}.
3) u_{i0} (field array initial estimation)= $A(\sigma_0)^{-1} y_i$.
4) σ_0 (initial σ map) constant and equal to 1 S/m.
5) λ_i (Lagrangian Multipliers) = 0.

In Fig. 9, the reconstruction with the DIPESAL method is presented for the same noise variable value. The same list of parameters is presented for this method:

1) μ (Constrains penalty parameters)=5×10^{-11}.
2) τ (Regularization parameter)=5×10^{-4}.
3) u_{i0} (field array initial estimation)= $A(\sigma_0)^{-1} y_i$.
4) σ_0 (initial σ map) constant and equal to 1 S/m.
5) λ_i (Lagrangian Multipliers) = 0.

In Fig. 10 a sequence of reconstruction images are shown for three different levels of noise, where lower values were also used.

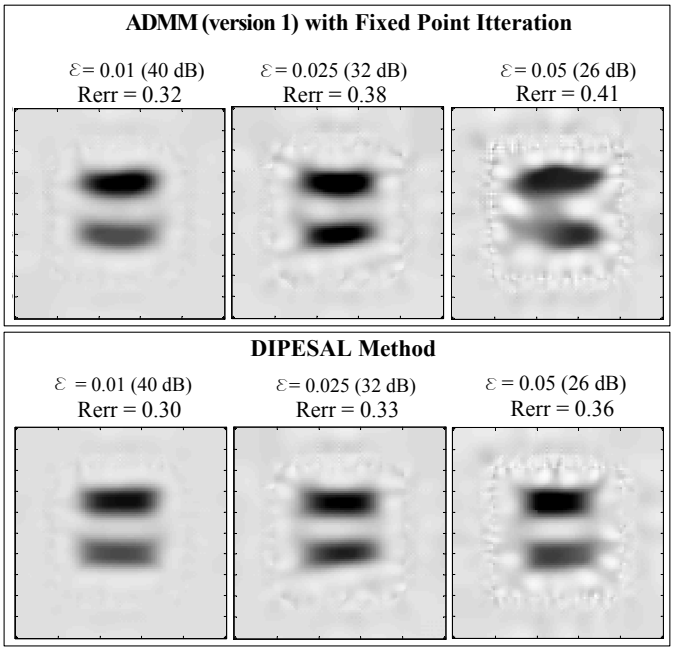

Fig. 10. Comparison between ADMM version 1 with fixed point iteration and DIPESAL for several noise levels (ε). The obtained relative error (Rerr) for each reconstruction is shown, revealing the enhanced performance of the DIPESAL method. The same grey scale shown in Fig. 9 is considered.

The relative error (Rerr) evolution with time for each method is compared in Fig. 11 for the $\varepsilon = 0.01$ scenario.

Both methods work well given the inverse problem intrinsic ill-posedeness. It is however clear the advantage of the DIPESAL method in the reconstruction figures compared with the ADMM (version 1) in terms of final relative error and analysing the obtained parameter map. In the example with $\varepsilon = 0.05$, the DIPESAL method yields the better relative performance. It leads to lower minimum value of relative residual of σ (5%) and it is faster than the version 1 method (In this problem approximately 7% faster).

1) Over-relaxation and Performance Assessment: In Fig. 12 three levels of over-relaxation were applied to the DIPESAL method.

For ζ too high the method does not converge, as seen for $\zeta = 2.0$. A fine tune ζ allows to increase convergence speed without changing convergence process. Near its limit, a ringing

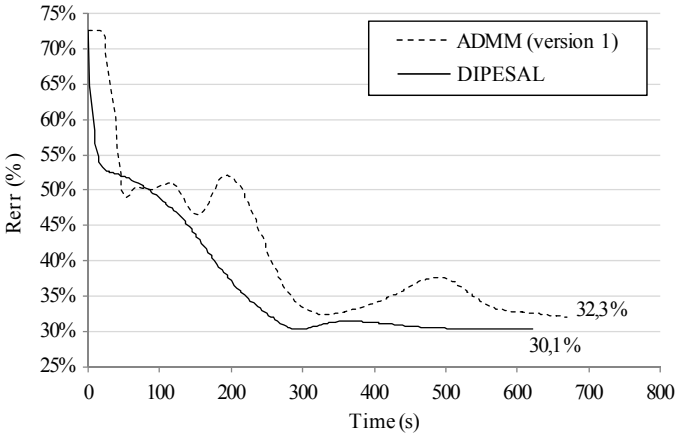

Fig. 11. Relative error (Rerr) evolution vs. time comparison figure for ADMM (version 1) and DIPESAL method for $\varepsilon = 0.01$ and 200 iterations.

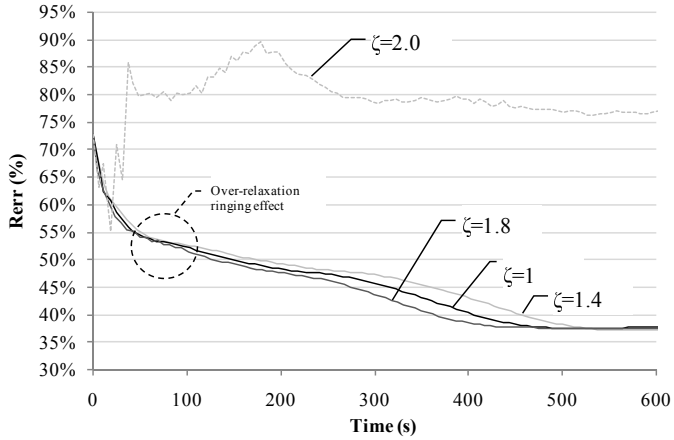

Fig. 12. Relative error (Rerr) evolution vs. time comparison for 4 over-relaxation scenarios with $\varepsilon = 0.05$. It is shown that for $\zeta \geq 2.0$ the over-relaxation is excessive.

in the convergence can be seen (see Fig. 12, inside the dotted circumference).

The best tuned method with over-relaxation was able to reduce the time to achieve the same minimum relative residual in approximately 18% of the time comparing with the same method without over-relaxation.

C. The μ Parameter

As seen in the 1D case, the choice of μ, the parameter that controls the weight of quadratic terms, influences the performance and convergence stability. Again, it was seen that large values of μ, corresponding to a higher imposition on the forward problem, take more time to reach the vicinity of the solution but its convergence is more stable to a reliable solution. For low values, the convergence is faster, but the final solution can be affected by numerical instabilities. In Fig. 13 a performance reconstruction example for 200 iterations of the ADMM (version 1), for three values of μ illustrates this behaviour:

V. Conclusions and Further Work

A new ADMM method was proposed to solve distributed parameter estimation problems with discontinuous parameter

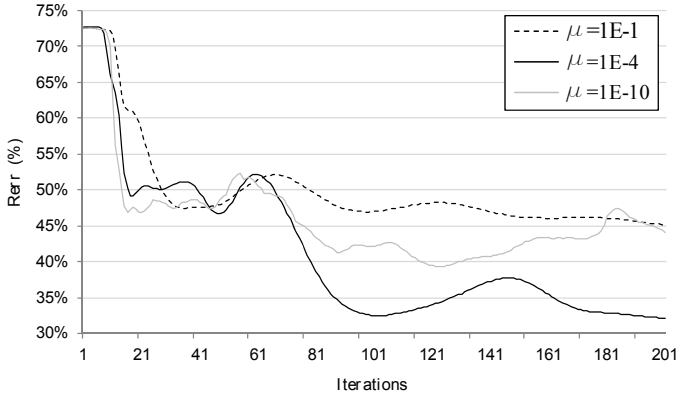

Fig. 13. Relative Error (Rerr) evolution using several μ values, 1E-1, 1E-4 aand 1E-10.

maps, in particular the case of elliptic PDEs.

The proposed algorithm implements an instance of ADMM, DIPESAL, consisting of a decoupled optimization with respect to u and σ leading to an Uzawa type scheme, to solve the resulting optimization problem, implementing the exact TV definition, in contrast with most works on DPE using TV. The method was tested for 1D simple scenario and a 2D real-world tomographical kind scenario, with partial measurements, where the number of measurements is $p = n/125$ and the number of independent sources is $n_I = 36$.

The DIPESAL was compared with ADMM (version 1) method, which is an ADMM method with an inexact TV solver. Exact TV accelerated solvers such as TwIST [15] and FISTA [45] yielded an overall slow scheme. So, although aware of the unfairness of the comparison in terms of error since different optimization problems are solved, it was decided to use a faster inexact TV solver. Furthermore this solver has been used by state-of-the-art algorithms.

In the simple 1D scenario, the feasibility of the method was shown. Its behaviour in noisy environments and for partial measurements scenarios was studied. It was clear that the number of field sources can clearly compensate partial acquisition matrices.

The developed method, DIPESAL, was shown to be more efficient in solving the proposed real-world scenario of 2D elliptical DPE when compared with the ADMM (version 1) method. The regularization was conveniently enforced and the relative residual of σ reached a minimum value 5% smaller and 7% faster. The resulting qualitative image was seen to be more accurate than the simple ADMM (version 1).

An over-relaxation strategy was tested in the 2D scenario with 18% performance improvements.

Finally the convergence dependence on the parameter that controls the weight of quadratic terms (μ) was analysed in both 1D and 2D cases, showing the importance of its balanced choice.

If the parameter estimation problem was convex, we could ensure that the DIPESAL would provide the exact solution whereas ADMM (version 1) method would provide an approximate one. In our nonconvex scenario, we cannot take this conclusion. Nevertheless, we have observed experimentally that DIPESAL yields systematically better and faster results.

Comparing with the standard Gauss-Newton approach referred in the introduction and normally considered as a standard approach to solve this kind of problem [8], the ADMM method, in particular the proposed one, is much faster, mainly in contexts where the problem is large and sparse. This constitutes an interesting area for further investigation since the presented method could be a state of the art candidate for this kind of inverse problem, for instance when the number of measurements and sources is large.

REFERENCES

[1] C. R. Vogel, *Computational Methods For Inverse Problems*. Siam, 2002.

[2] M. Cheney, D. Isaacson, and J. C. Newell, "Electrical impedance tomography," *SIAM REVIEW c*, vol. Vol. 41, No. 1, 1999.

[3] R. M. William Rodi, "Nonlinear conjugate gradients algorithm for 2d magnetotelluric inversion," *Geophysics*, vol. 66, pp. 174–187, 2001.

[4] B. F. Nielsen and A. Tveito, "An upscaling method for one-phase flow in heterogeneous reservoirs: A weightedoutputleastsquaresapproach," *Tech.Report8,Informatic Institute, University of Oslo, Oslo, Norway*, 1996.

[5] W. W. G. Yeh, "Review of parameter identification procedures in groundwater hydrology: The inverse problems," *Water Resources Research*, vol. 22, pp. 95–108, 1986.

[6] W. Bangerth, "Adaptive finite element methods for the identification of distributed parameters in partial differential equations," Ph.D. dissertation, Ruprecht Karls University, 2002.

[7] J. T. Han, "A computational method to solve nonlinear elliptic equations for natural convection in enclosures," *Numerical Heat Transfer, Part A: Applications,*, vol. Volume 2, Issue 2, pp. 165–175, 1979.

[8] C. R. Vogel, "Sparse matrix computations arising in the distributed parameter identification," *SIAM J. Matrix Anal. Appl*, vol. 20, pp. 1027–1037, 1999.

[9] X. Cai, S. Liu, and J. Zou, "Parallel overlapping domain decomposition methods for coupled inverse elliptic problems," *COMM. APP. MATH. AND COMP. SCI.*, vol. 4, no. 1, 2009.

[10] L. I. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Physica D: Nonlinear Phenomena*, vol. 60, no. 1, pp. 259 – 268, 1992.

[11] T. F. Chan and X.-C. Tai, "Identification of discontinuous coefficients in elliptic problems using total variation regularization," *SIAM J. Sci. Comput. 25*, vol. 25,3, 2003.

[12] A. Chambolle, "An algorithm for total variation minimization and applications," *Journal of Mathematical Imaging and Vision*, vol. 20, no. 1, pp. 89–97, 2004.

[13] J. Bect, L. Blanc-Feraud, G. Aubert, and A. Chambolle, "A-unified variational framework for image restoration," in *Proc. ECCV*, vol. 3024. Springer, 2004, pp. 1–13.

[14] T. Goldstein and S. Osher, "The split bregman method for l1-regularized problems," *SIAM J. Img. Sci.*, vol. 2, pp. 323–343, April 2009.

[15] J. Bioucas-Dias and M. Figueiredo, "A new twist: two-step iterative shrinkage/thresholding algorithms for image restoration," *IEEE Trans. on Image Proc.*, vol. 16, pp. 2992–3004, 2007.

[16] M. V. Afonso, J. Bioucas-Dias, and M. Figueiredo, "Fast image recovery using variable splitting and constrained optimization," *IEEE Transactions on Image Processing*, vol. 19, no. 9, pp. 2345–2356, 2010.

[17] Y. Wang, J. Yang, W. Yin, and Y. Zhang, "A new alternating minimization algorithm for total variation image reconstruction," *SIAM Journal on Imaging Sciences*, vol. 1, no. 3, pp. 248–272, 2008.

[18] J. Yang, Y. Zhang, and W. Yin., "A fast alternating direction method for tvl1-l2 signal reconstruction from partial fourier data." *To appear in IEEE Journal of Selected Topics in Signal Processing Special Issue on Compressed Sensing*, vol. b, 2008.

[19] X.-C. Tai and C. Wu, "Augmented lagrangian method, dual methods and split bregman iteration for rof model," in *Proceedings of the Second International Conference on Scale Space and Variational Methods in Computer Vision*, ser. SSVM '09. Berlin, Heidelberg: Springer-Verlag, 2009, pp. 502–513.

[20] M. Elad, P. Milanfar, and R. Rubinstein, "Analysis versus synthesis in signal priors," *Inverse Problems*, vol. 23, pp. 947–968, 2007.

[21] G. Bretti, M. Fornasier, and F. Pitolli, "Electric current density imaging via an accelerated iterative algorithm with joint sparsity constraintsconstraints," in *SPARS'09 - Signal Processing with Adaptive Sparse Structured Representations*, 2009.

[22] E. Haber and U. Ascher, "Preconditioned all-at-once methods for large, sparse parameter estimation problems," *Inverse Problems*, vol. 17, 2001.

[23] U. M. Ascher and E. Haber, "A multigrid method for distributed parameter estimation problems," *Electron. Trans. Numer. Anal.*, vol. 15, pp. 1–17, 2003.

[24] G. Kuruvila, S. Tarasan, and M. D. Salas, "Airfoil optimization by the one-shot method," *Agard report 803*, 1994.

[25] A. Shenoy, M. Heinkenschloss, and E. M. Cliff, "Airfoil design by an all-at-once method, int." *J. Comput. Fluid Dyn.*, vol. 11, pp. 3–25, 1998.

[26] K. Ito and K. Kunisch, "The augmented lagrangian method for parameter estimation in elliptic systems," *SIAM J. Control Optim.*, vol. 1, pp. 113–136, 1990.

[27] Y. L. Keung and J. Zou, "An efficient linear solver for nonlinear parameter identification problems," *SIAM J. Sci. Comput. 22*, vol. 5, pp. 1511–1526, 2000.

[28] Z. Chen and J. Zou, "An augmented lagrangian method for identifying discontinuous parameters in elliptic systems," *SIAM J. Control Optim.*, vol. 37,3, pp. 892–910, 1999.

[29] S. W. Jorge Nocedal, *Numerical Optimization*. Springer, 2000.

[30] J. Eckstein and D. P. Bertsekas, "On the douglas-rachford splitting method and the proximal point algorithm for maximal monotone operators," *Mathematical Programming*, vol. 55, pp. 293–318, 1992.

[31] R. Glowinski and A. Marroco, "Sur líapproximation, par elements finis de ordre un, et la resolution, par penalisation-dualite dune classe de problemes de dirichlet non lineares," *Revue Française d'Automatique, Informatique et Recherche Operationelle*, vol. 9, pp. 41–76, 1975.

[32] D. Gabay and B. Mercier, "A dual algorithm for the solution of nonlinear variational problems via finite-element approximations," *Computers and Mathematics with Applications*, vol. 2, pp. 17–40, 1976.

[33] M. Soleimani and W. R. Lionheart, "Absolute conductivity reconstruction in magnetic induction tomography using a nonlinear method," *IEEE Transactions on Medical Imaging*, vol. 25, no. 12, pp. 1521–1530, 2006.

[34] P. L. Combettes and J.-C. Pesquet, "Proximal splitting methods in signal processing," Tech. Rep. arXiv:0912.3522, Dec 2009.

[35] M. V. Afonso, J. B. Dias, and M. Figueiredo, "An augmented lagrangian approach to the constrained optimization formulation of imaging inverse problems," *IEEE Transactions on Image Processing*, vol. 20, no. 3, pp. 681–695, 2011.

[36] J. Bioucas-Dias and M. Figueiredo, "Multiplicative noise removal using variable splitting and constrained optimization," *Submitted to the IEEE Transactions on Image Processing*, 2009.

[37] D. P. Bertsekas, *Nonlinear programming*. Athena Scientific, 1999.

[38] T. F. Chan and Xue-Cheng, "Augmented lagrangian and total variation methods for recovering discontinuous coefficients from elliptic equations," *Computational Science for the 21st Century - J. Wiley & Sons*, pp. 597–607, 1997.

[39] E. Esser, "Applications of lagrangian-based alternating direction methods and connections to split bregman," UCLA CAM, Tech. Rep., 2009.

[40] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Foundations and Trends in Machine Learning*, vol. 3, no. 1, pp. 1–122, 2010.

[41] C. R. Vogel and M. E. Oman, "Iterative methods for total variation denoising," *SIAM J. Sci. Comput.*, vol. 17, pp. 227–238, January 1996.

[42] S. Setzer, "Split bregman algorithm, douglas-rachford splitting and frame shrinkage," in *Proceedings of the Second International Conference on Scale Space and Variational Methods in Computer Vision*, ser. SSVM '09. Berlin, Heidelberg: Springer-Verlag, 2009, pp. 464–476.

[43] K. Kunisch and X.-C. Tai, "Sequential and parallel splitting methods for bilinear control problems in hilbert spaces," *SIAM J Numer. Anal.*, vol. 34, pp. 911–914, 1997.

[44] E. Haber and S. Heldmann, "An octree multigrid method for quasi-static maxwell's equations with highly discontinuous coefficients," *Journal of Computational Physics*, vol. 223(2), pp. 783–796, 2007.

[45] A. Beck and M. Teboulle, "A fast iterative shrinkage-thresholding algorithm for linear inverse problems," *SIAM J. Imag. Sci.*, vol. 2, pp. 183–202, 2009.