THE SPLIT BREGMAN ALGORITHM APPLIED TO PDE-CONSTRAINED OPTIMIZATION PROBLEMS WITH TOTAL VARIATION REGULARIZATION

OLE LØSETH ELVETUN* AND BJØRN FREDRIK NIELSEN†

Abstract. We derive an efficient solution method for ill-posed PDE-constrained optimization problems with total variation regularization. This regularization technique allows discontinuous solutions, which is desirable in many applications. Our approach is to adapt the split Bregman technique to handle such PDE-constrained optimization problems. This leads to an iterative scheme where we must solve a linear saddle point problem in each iteration. We prove that the spectra of the corresponding saddle point operators are almost contained in three bounded intervals, not containing zero, with a very limited number of isolated eigenvalues. Krylov subspace methods handle such operators very well and thus provide an efficient algorithm. In fact, we can guarantee that the number of iterations needed cannot grow faster than $O([\ln(\alpha^{-1})]^2)$ as $\alpha \to 0$, where α is a small regularization parameters. Morover, in our numerical experiments we demonstrate that one can expect iteration numbers of order $O(\ln(\alpha^{-1}))$.

Key words. Total Variation regularization, PDE-constrained optimization, Bregman algorithm, MINRES, KKT systems

1. Introduction. In the field of PDE-constrained optimization, sophisticated algorithms and increased computing power have made it possible to compute numerical solutions of many advanced optimization problems. The use of Karush-Kuhn-Tucker (KKT) systems to solve such problems has become increasingly popular. These optimality systems are usually ill-posed, which leads to bad condition numbers for the discretized systems. Therefore, some kind of regularization technique must be invoked. The most popular method is the Tikhonov regularization technique, since this leads to linear optimality systems, provided that the state equation is also linear. In [1] the authors prove that a large class of such saddle point systems can be solved efficiently by applying the Minimal Residual (MINRES) algorithm. More specifically, they prove that the eigenvalues of the KKT system are almost contained in three bounded intervals. The number of isolated eigenvalues is only of order $O(\ln(\alpha^{-1}))$, where α is the regularization parameter. Krylov subspace methods are well suited to handle systems with such spectra.

It is known, however, that the use of a Tikhonov regularization term produces a smooth solution. In many inverse problems, the control parameter is often some physical property, like a heat source, density of a medium or an electrical potential. When we try to identify such quantities, it might be desirable to make a sharp separation between regions with different qualities of the physical property. In other words, we want "jumps" in the solution. Thus, one might argue that the smooth solutions produced with Tikhonov regularization are of limited value in such cases.

In the field of image analysis, researchers have for decades been interested in optimization problems with such discontinuous solutions. In [2] the authors proposed the famous Rudin-Osher-Fatemi (ROF) model

$$\min_{v \in BV(\Omega)} \left\{ \frac{1}{2}\rho \|v - d\|^2_{L^2(\Omega)} + \int_\Omega |Dv| \, dx \right\}, \tag{1.1}$$

*Department of Mathematical Sciences and Technology, Norwegian University of Life Sciences, Norway. (ole.elvetun@umb.no).

†Department of Mathematical Sciences and Technology, Norwegian University of Life Sciences, Norway; Simula Research Laboratory; Center for Cardiological Innovation, Oslo University Hospital.

where the Banach space of functions with bounded variation is defined by

$$BV(\Omega) = \{v \in L^1(\Omega) : \int_\Omega |Dv| \; dx < \infty\}, \tag{1.2}$$

and the regularization term in (1.1) is defined by the distribution

$$\int_\Omega |Dv| \; dx = \sup \left\{ \int_\Omega v \text{ div } p : p \in C_0^1(\Omega; \mathbb{R}^n); \quad |p|_\infty \leq 1 \right\}. \tag{1.3}$$

For elements in $W^{1,1}(\Omega)$, the distribution (1.3) is equal to the normal weak derivative, see [3].

The regularization term (1.3) is known as Total Variation (TV) regularization, and it allows for discontinuous solutions. This ability to include "jumps" in the solution has made it very popular for denoising pictures. Unfortunately, (1.1) is a very challenging problem to solve, since the TV term is highly non-linear and also non-differentiable. Nevertheless, due to the desirable denoising property, it has received much attention, and a large number of solution algorithms have been suggested. In Section 3, we will present some of the most well known methods.

The denoising case has been extended to include more sophisticated problems. In particular, the deblurring problem has been thoroughly analyzed. This problem can be written as

$$\min_{v \in BV(\Omega)} \left\{ \frac{1}{2} \rho \|Kv - d\|^2_{L^2(\Omega)} + \int_\Omega |Dv| \; dx \right\}, \tag{1.4}$$

where $K : BV(\Omega) \to L^2(\Omega)$ typically is a convolution operator, see e.g. [4]. This deblurring problem is the starting point for our PDE-constrained optimization formulation. Mathematically, an abstract form of a PDE-constrained optimization problem with TV regularization reads

$$\min_{(v,u) \in V \times U} \left\{ \frac{1}{2} \rho \|Tu - d\|^2_Z + \int_\Omega |Dv| \; dx \right\},$$

subject to

$$Au + Bv = 0, \tag{1.5}$$

where

- V is the control space, $1 \leq \dim(V) \leq \infty$,
- U is the state space, $1 \leq \dim(U) \leq \infty$, and
- Z is the observation space, $1 \leq \dim(Z) \leq \infty$.

Here, U and Z are Hilbert spaces, and (1.5) is a PDE. The control space V will be discussed thoroughly in the next section. Furthermore, d is the given observation data, the domain $\Omega \subset \mathbb{R}^n$ is bounded, and $\rho > 1$ is the regularization parameter[1]. The operators A, B and T will be discussed properly in Section 4.

For a specific kind of elliptic equation, the use of total variation regularization has been used to identify discontinuous coefficients. Basically, such problems have been solved with an augmented Lagrangian method, see e.g. [5, 6], or a level set method, see e.g. [7, 8].

[1] Often the regularization parameter is placed in front of the regularization term and not in front of the data fidelity term. In the former approach, the values will then typically be $1/\rho$.

Unfortunately, for many PDE-constrained optimization problems, the total variation term does not provide sufficient regularization. More specifically, the so-called forward operator must be injective to guarantee a unique solution. Therefore, we might consider the more general formulation

$$\min_{(v,u)\in V\times U}\left\{\frac{1}{2}\rho\|Tu-d\|_Z^2+\frac{1}{2}\kappa\|v\|_{L^2(\Omega)}^2+\int_\Omega |Dv|\ dx\right\}, \tag{1.6}$$

subject to

$$Au+Bv=0, \tag{1.7}$$

where $0\leq\kappa<<1$. As we will see in the next section, this problem has a unique solution if $\kappa>0$, or if A, B and T are all injective.

The objective of this paper is to propose and analyze an efficient algorithm for solving the general problem (1.6)-(1.7). To succeed with this objective, we must not only guarantee an efficient iterative solution of the non-linear total variation term, but also for the inner systems that we will obtain in each iteration of the outer algorithm. This will be achieved by combining the analysis in [1] with a successful method for solving (1.4), namely the split Bregman method [9]. In more detail, we outline the paper as follows:

- Section 2 contains a short introduction to the existence and uniqueness properties of TV regularization problems of the form (1.4), while some of the most popular solution techniques are presented in Section 3. These are well-known topics in the imaging community, but they are seldomly discussed for PDE-constrained optimization problems. Hence, we choose to include these topics.
- In Section 4 we show how the PDE-constrained optimization problem (1.6)-(1.7) can be modified in such a way that we can apply the split Bregman algorithm.
- In Section 5 we prove that the KKT systems that arise in each iteration of the split Bregman algorithm has a spectrum *almost* contained in three bounded intervals, with a very limited number of isolated eigenvalues. Hence, Krylov subspace algorithms will handle these systems very well. We will come back to the exact form of this spectrum in Section 5.
- Section 6 presents an alternative version of the split Bregman algorithm.
- Finally, in Section 7 we illuminate the theoretical results with some numerical experiments.

2. Existence and uniqueness. The existence and uniqueness issues for total variation regularization problems on the form

$$\min_{v\in V}\left\{\underbrace{\frac{1}{2}\rho\|Kv-d\|_Z^2+\frac{1}{2}\kappa\|v\|_{L^2(\Omega)}^2}_{J(v)}+\int_\Omega |Dv|\ dx\right\} \tag{2.1}$$

are discussed thoroughly in [10]:

THEOREM 2.1. *Let V be a closed, convex subspace of $L^2(\Omega)$ and let Z be a general Hilbert space. Assume that $V\cap BV(\Omega)$ is nonempty and that one of the following two conditions are satisfied:*

(i) V is bounded in $L^2(\Omega)$.
(ii) J is coercive in the following sense:

$$\text{If } \{v_j\}_{j=1}^{\infty} \quad \subset \quad V, \left\{ \int_\Omega |\nabla v_j| \right\}_{j=1}^{\infty} \text{ is bounded and}$$
$$\|v_j\|_{L^2(\Omega)} \to \infty \Rightarrow J(v_j) \to \infty.$$

Then (2.1) has at least one solution $v^* \in V \cap BV(\Omega)$.

Note that if $\kappa > 0$, then the second condition holds.

THEOREM 2.2. *If either $K : L^2(\Omega) \to Z$ is injective or $\kappa > 0$, then the solution v^* of (2.1) is unique.*

These theorems are well known, but they are presented here for convenience. In Section 5, we will see that, in order to guarantee efficient solution of the inner systems we obtain in each iteration of the split Bregman algorithm, we do not need the (rather) strong assumptions in Theorem 2.2.

3. Brief overview of solution algorithms. The TV denoising problem has triggered the development of several different solution algorithms for (1.4) and (2.1). Two of these, the fixed-point algorithm introduced in [11] and the primal-dual Newton method proposed in [12], can be deduced from the Euler-Lagrange equation associated with (2.1). This equation is obtained by setting the Fréchet derivative of (2.1) equal to zero. Hence, we get[2]

$$\rho K^* K v - \text{ div } \frac{Dv}{|Dv|} = \rho K^* d. \tag{3.1}$$

Clearly, this is not well-defined for $Dv = 0$. However, we can introduce the perturbation

$$|Dv|_\beta = \sqrt{|Dv|^2 + \beta^2}, 0 < \beta << 1,$$

in the denominator to make the equation well-defined, see [3] for a thorough analysis of this perturbation. We omit a further discussion of this topic, since the split Bregman algorithm does not require the use of such a perturbation. Instead, we present the first solution method; the fixed-point algorithm introduced in [11].

The Euler-Lagrange fixed-point method. The equation we must solve in this algorithm can be though of as a lagged-diffusion equation, where one updates the solution according to

$$\rho K^* K v^{k+1} - \text{ div } \frac{Dv^{k+1}}{|Dv^k|_\beta} = \rho K^* d,$$

cf. (3.1). Although this method is very robust, it is also rather slow and it might cause numerical instabilities with regard to the perturbation parameter β. In addition, since we fix the denominator, we "lose control" of the coercivity as $|Dv^k| \to \infty$.

[2]The Euler-Lagrange equation is usually discussed for the case $\kappa = 0$, i.e. for (1.4). We present it here in the same manner.

The primal-dual Newton method. Initial attempts to apply Newton's method to (3.1) were unsuccessful, as the convergence basin is extremely small. In [12] the authors suggested a remedy with the introduction of the flux parameter

$$p = \frac{Dv}{|Dv|_\beta}. \tag{3.2}$$

They then obtained the optimality system

$$\begin{aligned} \rho K^* K v - \ \text{div}\ p &= \rho K^* d, \\ |Dv|_\beta p - Dv &= 0, \end{aligned}$$

where the second equation is deduced from the definition of the flux parameter (3.2). The authors showed numerically that this approach produced convergence results superior to the previous fixed-point algorithm. Unfortunately, this system also has problems with numerical instabilities as the perturbation parameter $\beta \to 0$.

Note that both the fixed-point algorithm and the primal-dual Newton method attempt to solve the Euler-Lagrange equation. This equation, however, is not well-defined without the perturbation $|Dv|_\beta$ of $|Dv|$. With the introduction of the perturbation, the non-differentiable term is approximated by an infinitely smooth quantity. In the next method, the minimization problem is solved with a technique which involves subgradients instead of the derivative; there is no longer a need for the perturbation. This turns out to be advantageous.

The split Bregman method. This method has its roots in the Bregman iteration, which is an algorithm for computing extrema of convex functionals [13]. Later, it was used in [14] as a new regularization procedure for inverse problems. In [9] the authors used this approach to find an effective solution method for L^1-regularization problems. In particular, they demonstrated why this method works well for total variation problems. The method is, however, designed for finite dimensional problems. Hence, the remainder of this article will be concerned with finite dimensional approximations. More specifically, there are two reasons for why we choose to analyze the discretized and not the continuous system:

(i) Whenever a specific minimization problem is to be solved, we must approximate the problem using some kind of discretization technique. Hence, we will ultimately work with finite dimensional problems.

(ii) The infinite dimensional analysis involves the use of the dual space of $BV(\Omega)$, which is extremely challenging to represent. In fact, since the space $BV(\Omega)$ is not separable, the dual space is not even a function space [15]. In a finite dimensional setting, however, all norms are equivalent, and we might "choose" a dual space that is easier to analyze. We will later observe that, in the finite dimensional split Bregman algorithm, we obtain the Laplace operator $-\Delta : H_h^1(\Omega) \to H_h^1(\Omega)'$, instead of the more challenging operator $D'D : BV(\Omega) \to BV(\Omega)'$. We return to this matter in Section 4.

We now briefly discuss the split Bregman algorithm presented in [9]. Let P_h^1 be the space of piecewise linear functions, and let

$$V_h = V \cap P_h^1 \text{ and } Z_h = Z \cap P_h^1$$

be finite dimensional subspaces of V and Z, respectively, cf. (2.1). The authors of [9] started by formulating a discretized approximation of the problem (2.1), and then

wrote it on the equivalent form

$$\min_{v_h,p_h\in V_h\times\mathbf{P}_h^0}\left\{\frac{1}{2}\rho\|Kv_h-d_h\|_{Z_h}^2+\frac{1}{2}\kappa\|v_h\|_{L_h^2(\Omega)}^2+\int_\Omega|p_h|\right\}, \tag{3.3}$$

subject to

$$\nabla v_h = p_h, \tag{3.4}$$

where $\mathbf{P}_h^0$ is a vector space of piecewise constant functions.

We will not go into details on how the split Bregman algorithm is derived, instead we refer to [9] and [16]. We would, however, like to highlight an interesting remark from [16]: Note that the problem (3.3)-(3.4) can be solved by sequentially solving the penalty formulation

$$\min_{v_h,p_h\in V_h\times\mathbf{P}_h^0}\left\{\frac{1}{2}\rho\|Kv_h-d_h\|_{Z_h}^2+\frac{1}{2}\kappa\|v_h\|_{L_h^2(\Omega)}^2+\int_\Omega|p_h|+\frac{1}{2}\lambda^k\|\nabla v_h-p_h\|_{L_h^2(\Omega)}^2,\right\}$$

where $\lambda^k\to\infty$. Unfortunately, such penalty methods are ineffective, and leads to numerical difficulties as λ^k grows large.

In the split Bregman algorithm, see Algorithm 1, we note that the parameter λ is fixed. Instead, it is the "data" that varies with the introduction of b^k. Hence, we obtain much better numerical stability. It was shown in [17] and [18] that the split Bregman method is equivalent to an augmented Lagrangian method [19, 20].

Algorithm 1 The split Bregman algorithm for total variation regularization

1: Choose $v_h^0=0, p_h^0=0, b_h^0=0$
2: **for** k = 0, 1, 2,... **do**
3: $\quad v_h^{k+1}=\arg\min_{v_h\in V_h}\frac{1}{2}\rho\|Kv_h-d_h\|_{Z_h}^2+\frac{1}{2}\kappa\|v_h\|_{L_h^2(\Omega)}^2+\frac{1}{2}\lambda\|\nabla v_h-p_h^k+b_h^k\|_{L_h^2(\Omega)}^2,$
4: $\quad p_h^{k+1}=\arg\min_{p_h\in\mathbf{P}_h^0}\int_\Omega|p_h|+\frac{\lambda}{2}\|\nabla v_h^{k+1}-p_h+b_h^k\|_{L_h^2(\Omega)}^2,$
5: $\quad b_h^{k+1}=b_h^k+\nabla v_h^{k+1}-p_h^{k+1}.$
6: **end for**

Before we end this section, we would like to present one important theorem from [16]:

THEOREM 3.1. *Assume that there exists at least one solution v_h^* of (2.1). Then the split Bregman algorithm satisfy*

$$\lim_{k\to\infty}\frac{1}{2}\rho\|Kv_h^k-d_h\|_{Z_h}^2+\frac{1}{2}\kappa\|v_h^k\|_{L_h^2(\Omega)}^2+\int_\Omega|\nabla v_h^k|\,dx = \frac{1}{2}\rho\|Kv_h^*-d_h\|_{Z_h}^2+\frac{1}{2}\kappa\|v_h^*\|_{L_h^2(\Omega)}^2+\int_\Omega|\nabla v_h^*|\,dx.$$

If the solution v_h^* is unique, we also have

$$\lim_{k\to\infty}\|v_h^k-v_h^*\|_{L_h^2(\Omega)}=0.$$

4. Split Bregman algorithm for PDE-constrained optimization problems. Recall that our main objective is to derive an efficient solution method for (1.6)-(1.7), i.e. for rather general PDE-constrained optimization problems subject to TV regularization. We will restrict our analysis to problems that satisfy the assumptions

$\mathcal{A}\mathbf{1}$: $A : U \to U'$ is bounded and linear.
$\mathcal{A}\mathbf{2}$: A^{-1} exists and is bounded.
$\mathcal{A}\mathbf{3}$: $B : V \to U'$ is bounded and linear.
$\mathcal{A}\mathbf{4}$: $T : U \to Z$ is bounded and linear.

Due to assumption $\mathcal{A}\mathbf{2}$, we can, in the finite dimensional setting, write (1.7) on the form

$$u_h = -A^{-1}Bv_h. \tag{4.1}$$

Consequently, we might formulate the minimization problem (1.6)-(1.7) as

$$\min_{v\in V_h}\left\{\frac{1}{2}\rho\|Kv_h - d_h\|^2_{Z_h} + \frac{1}{2}\kappa\|v_h\|^2_{L^2_h(\Omega)} + \int_\Omega |\nabla v_h|\ dx\right\} \tag{4.2}$$

where $K : V_h \to Z_h$ is defined by

$$K = -TA^{-1}B. \tag{4.3}$$

We observe that the minimization problem (4.2) is on the same form as (2.1). This motivates the use of the split Bregman algorithm. Unfortunately, however, the explicit computation of the operator K is not possible in practical applications; if (1.7) is a PDE, then the inverse of A is typically too expensive to compute explicitly. This issue has been handled, in the case of Tikhonov regularization, by solving the associated KKT system. The purpose of this paper is to adapt the KKT approach to the framework of the split Bregman algorithm. As we will see below, this yields an efficient and practical solution method for PDE-constrained optimization problems subject to TV regularization.

We do this by applying Algorithm 1 to the minimization problem (4.2). Step 5 in Algorithm 1 is straightforward, and Step 4 is very cheap to solve by the shrinkage operator

$$p^{k+1}_{h,x_i}(x) = \text{shrink}\left(\nabla_{x_i} v^{k+1}_h(x) + b^k_{h,x_i}(x), \frac{1}{\lambda}\right), \tag{4.4}$$

where

$$\text{shrink}(a,b) = \frac{a}{|a|} * \max(a-b,0),$$

see [9]. Hence, the challenge is to find the minimizer of Step 3. That is, we must solve the minimization problem

$$\min_{v_h\in V_h}\left\{\frac{1}{2}\rho\|Kv_h - d_h\|^2_{Z_h} + \frac{1}{2}\kappa\|v_h\|^2_{L^2_h(\Omega)} + \frac{1}{2}\lambda\|\nabla v_h - p^k_h + b^k_h\|^2_{L^2_h(\Omega)}\right\},$$

where d_h, p^k_h and b^k_h are given quantities. By combining this minimization problem with equations (4.1) and (4.3), we get the equivalent constrained minimization problem:

$$\min_{v_h,u_h\in V_h\times U_h}\left\{\frac{1}{2}\rho\|Tu_h - d_h\|^2_{Z_h} + \frac{1}{2}\kappa\|v_h\|^2_{L^2_h(\Omega)} + \frac{1}{2}\lambda\|\nabla v_h - p^k_h + b^k_h\|^2_{L^2_h(\Omega)}\right\} \tag{4.5}$$

subject to

$$Au_h + Bv_h = 0. \tag{4.6}$$

For the sake of simplicity, we want our optimality system to be as similar as possible to the optimality system analyzed in [1]. Thus, we need to scale the cost-functional in (4.5) such that we get

$$\min_{v_h, u_h \in V_h \times U_h} \left\{ \frac{1}{2}\|Tu_h - d_h\|_{Z_h}^2 + \frac{1}{2}\gamma\|v_h\|_{L_h^2(\Omega)}^2 + \frac{1}{2}\alpha\|\nabla v_h - p_h^k + b_h^k\|_{L_h^2(\Omega)}^2 \right\} \tag{4.7}$$

subject to (4.6), where

$$\alpha = \frac{\lambda}{\rho} \text{ and } \gamma = \frac{\kappa}{\rho}. \tag{4.8}$$

Next, we can introduce the Lagrangian associated with (4.6)-(4.7):

$$\mathcal{L}(v_h, u_h, w_h) = \frac{1}{2}\|Tu_h - d_h\|_{Z_h}^2 + \frac{1}{2}\gamma\|v_h\|_{L_h^2(\Omega)}^2 + \frac{1}{2}\alpha\|\nabla v_h - p_h^k + b_h^k\|_{L_h^2(\Omega)}^2 + \langle Au_h + Bv_h, w_h\rangle.$$

The first-order optimality conditions can be found by computing the derivatives of the Lagrangian with respect to v_h, u_h and w_h. These conditions can be expressed by the KKT system

$$\underbrace{\begin{bmatrix} -\alpha\Delta + \gamma E & 0 & B' \\ 0 & T'T & A' \\ B & A & 0 \end{bmatrix}}_{\widehat{\mathcal{A}}_\alpha} \begin{bmatrix} v_h \\ u_h \\ w_h \end{bmatrix} = \begin{bmatrix} -\alpha\nabla \cdot p_h^k + \alpha\nabla \cdot b_h^k \\ T'd_h \\ 0 \end{bmatrix}, \tag{4.9}$$

where " $'$ " is used to denote dual operators, and $E : V_h \to V_h'$ is defined by

$$\langle Ev, \phi\rangle = (v, \phi)_{L_h^2(\Omega)}, \quad \phi \in V_h.$$

We have thus derived a new system of equations to be solved in Step 3 in Algorithm 1, which does not require the explicit inverse of A. Also note the form of the operator $-\Delta : V_h \to V_h'$ in the top left corner of the KKT system (4.9). In an infinite dimensional setting, this operator must be replaced with the more involved operator $D'D : BV(\Omega) \to BV(\Omega)'$, see [10] for a thorough discussion of this operator. The operator $D'D$ is much more challenging to analyze, but it coincides with the operator $-\Delta$ in a finite dimensional setting. This follows from the fact that $Dv = \nabla v$ for all elements in $W^{1,1}(\Omega)$, see [3]. Hence, we found it more convenient to discuss the matter in the finite dimensional setting. This concludes the discussion of Step 3 in Algorithm 1.

We might now formulate the full algorithm for solving the PDE-constrained optimization problem (1.6)-(1.7), see Algorithm 2.

The efficiency of the split Bregman algorithm has been demonstrated earlier, see e.g. [9, 16]. Of the three inner steps of the **for**-loop in Algorithm 2, the update of b_h^k is obviously cheap, and the update of p_h^k is accomplished by the simple shrinkage operator (4.4). What remains, however, is to analyze the spectrum of the KKT system in (4.9), see Step 3 in Algorithm 2: The efficiency of the algorithm is highly dependent on how fast we can solve these KKT systems with, e.g., Krylov subspace solvers.

Algorithm 2 The split Bregman algorithm for PDE-constrained optimization problems with TV regularization

1: Choose $v_h^0 = 0, p_h^0 = 0, b_h^0 = 0$
2: **for** k = 0, 1, 2,... **do**
3: Let (v_h^{k+1}, u_h^{k+1}, w_h^{k+1}) be the solution of (4.9).
4: $p_h^{k+1} = \arg\min_{p_h \in \mathbf{P}_0^h} \int_{\Omega_h} |p_h| + \frac{\lambda}{2} \|\nabla v_h^{k+1} - p_h + b_h^k\|^2_{L_h^2(\Omega)}$,
5: $b_h^{k+1} = b_h^k + \nabla v_h^{k+1} - p_h^{k+1}$.
6: **end for**

5. Spectrum of the KKT system. In its current form, the operator $\widehat{\mathcal{A}}_\alpha$ in (4.9) is a mapping from the product space $V_h \times U_h \times U_h$ onto the dual space $V_h' \times U_h' \times U_h'$.

Since this operator maps to the dual space, and not to the space itself, it is not possible to use the MINRES method directly. A remedy exists, however, in the form of Riesz maps. In this case, we must introduce the two Riesz maps

$$R_{V_h} : V_h \to V_h',$$
$$R_{U_h} : U_h \to U_h'.$$

This enables us to use the MINRES algorithm, since the KKT system (4.9) can be written as

$$\underbrace{\begin{bmatrix} R_{V_h}^{-1} & 0 & 0 \\ 0 & R_{U_h}^{-1} & 0 \\ 0 & 0 & R_{U_h}^{-1} \end{bmatrix}}_{\mathcal{R}^{-1}} \underbrace{\begin{bmatrix} -\alpha\Delta + \gamma E & 0 & B' \\ 0 & T'T & A' \\ B & A & 0 \end{bmatrix}}_{\widehat{\mathcal{A}}_\alpha} \begin{bmatrix} v_h \\ u_h \\ w_h \end{bmatrix}$$
$$= \begin{bmatrix} R_{V_h}^{-1} & 0 & 0 \\ 0 & R_{U_h}^{-1} & 0 \\ 0 & 0 & R_{U_h}^{-1} \end{bmatrix} \begin{bmatrix} -\alpha\nabla \cdot p_h^k + \alpha\nabla \cdot b_h^k \\ T'd_h \\ 0 \end{bmatrix}, \tag{5.1}$$

where

$$\mathcal{R}^{-1}\widehat{\mathcal{A}}_\alpha : V_h \times U_h \times U_h \to V_h \times U_h \times U_h.$$

The operator $\mathcal{R}^{-1}$ can be considered to be a preconditioner. See [21, 1] for a more thorough analysis.

We performed an experimental investigation that suggested the use of small values of α to obtain good convergence results for the outer split Bregman algorithm. That is, λ/ρ should be small, see (4.8). According to standard theory for Krylov subspace methods, the number of iterations needed by the MINRES algorithm is of the same order as the spectral condition number of the involved operator. In our case, that corresponds to iterations numbers of order $O(\alpha^{-1})$, when $\gamma = 0$. We will now prove that this estimate is very pessimistic.

Since the case $\gamma = 0$ is the most challenging, and also the most interesting, we will for the rest of the analysis assume that this is the case, i.e. $\gamma = 0$. Let us first simplify the notation in (5.1), and write the operator $\mathcal{R}^{-1}\widehat{\mathcal{A}}_\alpha$ in the form

$$\mathcal{R}^{-1}\widehat{\mathcal{A}}_\alpha = \mathcal{A}_\alpha = \begin{bmatrix} \alpha Q & 0 & \tilde{B}^* \\ 0 & T^*T & \tilde{A}^* \\ \tilde{B} & \tilde{A} & 0 \end{bmatrix}, \tag{5.2}$$

where we have the following definitions:

- $Q = -R_{V_h}^{-1}\Delta : V_h \to V_h$,
- $\tilde{B} = R_{U_h}^{-1}B : V_h \to U_h$,
- $\tilde{A} = R_{U_h}^{-1}A : U_h \to U_h$,
- $T^*T = R_{U_h}^{-1}T'T : U_h \to U_h$.

In this new form, the operator $\mathcal{A}_\alpha$ in (5.2) is very similar to the operator analysed in [1]. In fact, they analyzed the operator $\mathcal{B}_\alpha : V \times U \times U \to V \times U \times U$, defined as

$$\mathcal{B}_\alpha = \begin{bmatrix} \alpha I & 0 & \tilde{B}^* \\ 0 & T^*T & \tilde{A}^* \\ \tilde{B} & \tilde{A} & 0 \end{bmatrix}. \tag{5.3}$$

The main result in [1] is that the spectrum of $\mathcal{B}_\alpha$ is of the form

$$\text{sp}(\mathcal{B}_\alpha) \subset [-b, -a] \cup [c\alpha, 2\alpha] \cup \{\tau_1, \tau_2, ..., \tau_{N(\alpha)}\} \cup [a, b],$$

where

$$N(\alpha) = O(\ln(\alpha^{-1})$$

and the constants a, b and $c > 0$ are independent of the parameter α. The analysis is in [1] roughly performed as follows:

- The negative eigenvalues are shown to be bounded away from zero, regardless of the size of regularization parameter $\alpha \geq 0$. That is, it even holds for $\alpha = 0$. Hence, the negative eigenvalues of $\mathcal{A}_\alpha$, defined in (5.2), are bounded away from zero: The argument in [1] can be adapted to the present situation in a straightforward manner.
- For the positive eigenvalues, the Courant-Fischer-Weyl min-max principle is used to show that the difference between the eigenvalues of $\mathcal{B}_0$ and $\mathcal{B}_\alpha$ is "small", where $\mathcal{B}_0$ denotes the operator $\mathcal{B}_\alpha$ with zero regularization $\alpha = 0$. More specifically, they prove that the difference between the eigenvalues of $\mathcal{B}_0$ and $\mathcal{B}_\alpha$, properly sorted, is less than the size of the regularization parameter $0 < \alpha \ll 1$. It is easy to verify that a similar property will hold for $\mathcal{A}_0$ and $\mathcal{A}_\alpha$. More specifically, the difference between the eigenvalues of $\mathcal{A}_0$ and $\mathcal{A}_\alpha$ is less than $\tilde{c}\alpha$, where $\tilde{c} = \|Q\|$.
- Finally, the analysis in [1] requires that

  $$\alpha(v, v)_{V_h} + (T^*Tu, u)_{U_h}$$

  must be coercive whenever

  $$\tilde{A}u + \tilde{B}v = 0.$$

  It is proven in [1] that this property holds for the operator $\mathcal{B}_\alpha$. For the operator $\mathcal{A}_\alpha$, defined in (5.2), this analysis is more involved, and it will therefore be explored in detail here. More specifically, we must show, provided that suitable assumptions hold, that

  $$\alpha(Qv, v)_{V_h} + (T^*Tu, u)_{U_h}$$

  is coercive for all (v, u) satisfying

  $$\tilde{A}u + \tilde{B}v = 0.$$

To further investigate the coercivity problem associated with (5.2), we introduce the notation

$$X_h = V_h \times U_h, \|x\|_{X_h} = \|(v_h, u_h)\|_{X_h} = \sqrt{\|v_h\|_{V_h}^2 + \|u_h\|_{U_h}^2},$$

$$M_\alpha = \begin{bmatrix} \alpha Q & 0 \\ 0 & T^*T \end{bmatrix} : X_h \to X_h, \tag{5.4}$$

$$N = \begin{bmatrix} \tilde{B} & \tilde{A} \end{bmatrix} : X_h \to U_h. \tag{5.5}$$

Since we work with finite dimensional spaces, we employ the space V_h with the norm

$$\|\cdot\|_{V_h}^2 = \|\cdot\|_{L_h^2(\Omega)}^2 + |\cdot|_{H_h^1(\Omega)}^2, \tag{5.6}$$

i.e. in the finite dimensional setting we can guarantee that a subspace of $L^2(\Omega)$ is also a subspace of $H^1(\Omega)$, provided that proper elements are used. In fact, it is shown in [22] that the approximation of functions with piecewise linear elements converge (in $W^{1,1}$-norm) to the exact function of bounded variation as the mesh size goes to zero. Thus, with these elements we get the well defined mapping $-\Delta : H_h^1(\Omega) \to H_h^1(\Omega)'$. Note that, for the analysis presented below, we must assume that the operator B satisfies assumption $\mathcal{A}\mathbf{3}$ with the norm (5.6), i.e. that

$$B : V_h \to U_h'$$

is bounded, which along with assumptions $\mathcal{A}\mathbf{2}$ and $\mathcal{A}\mathbf{4}$ imply that

$$K = -TA^{-1}B = -T\tilde{A}^{-1}\tilde{B}$$

is bounded.

We are now ready to formulate the result concerning the coercivity issue for the operator $\mathcal{A}_\alpha$ defined in (5.2).

LEMMA 5.1. *Let M_α and N be defined as in (5.4) and (5.5), respectively. Assume that V_h is a closed, convex, finite dimensional subspace of $L^2(\Omega)$, and let U_h and Z_h be finite dimensional subspaces of the two Hilbert spaces U and Z, respectively. Furthermore, assume that $\mathcal{A}\mathbf{1} - \mathcal{A}\mathbf{4}$ hold and that*

$$K = -T\tilde{A}^{-1}\tilde{B} : V_h \to Z_h$$

does not annihilate constants, i.e. the constant function $k \notin \mathcal{N}(K)$. Then the operator M_α is coercive on the kernel of N, i.e. for $\alpha \in (0, 1)$:

$$(M_\alpha x, x)_{X_h} \geq c\alpha \|x\|_{X_h}^2 \tag{5.7}$$

for all $x = (v_h, u_h) \in X_h$ satisfying

$$\tilde{A}u_h + \tilde{B}v_h = 0. \tag{5.8}$$

Proof. We will first show that, if K does not annihilate constants, then there exists a constant $c \in (0,1)$ such that

$$(Kv_h, Kv_h)_{Z_h} \geq (c-1)(\nabla v_h, \nabla v_h)_{L^2_h(\Omega)} + c(v_h, v_h)_{L^2_h(\Omega)}, \quad \forall v_h \in V_h. \tag{5.9}$$

Thereafter, we will use this result to prove the coercivity.

Assume that there does not exist any constant $c \in (0,1)$ such that (5.9) holds. We will show that this implies that the constant function k must belong to the null-space of $K = -T\tilde{A}^{-1}\tilde{B}$. If inequality (5.9) is not satisfied for any constant $c \in (0,1)$, it follows that there exists a sequence

$$\{v_h^i\}, \quad \|v_h^i\|^2_{L^2_h(\Omega)} = (v_h^i, v_h^i)_{L^2_h(\Omega)} = 1,$$

such that

$$\begin{aligned} 0 \leq (Kv_h^i, Kv_h^i)_{Z_h} &< \left(\frac{1}{i} - 1\right) |v_h^i|^2_{H^1_h(\Omega)} + \frac{1}{i}(v_h^i, v_h^i)_{L^2_h(\Omega)} \\ &= \left(\frac{1}{i} - 1\right) |v_h^i|^2_{H^1_h(\Omega)} + \frac{1}{i}. \end{aligned}$$

We may choose a sequence with the property $\|v_h^i\|^2_{L^2_h(\Omega)} = 1$ because the operator K is linear. Since $(1/i - 1) \to -1$ as $i \to \infty$, we can conclude that

$$|v_h^i|_{H^1_h(\Omega)} \to 0 \quad \text{as } i \to \infty, \tag{5.10}$$

$$\|Kv_h^i\|_{Z_h} \to 0 \quad \text{as } i \to \infty. \tag{5.11}$$

We will now show that $\{v_h^i\}$ has a limit in $H^1_h(\Omega)$. Let

$$S_h = \left\{ s_h \in H^1_h(\Omega) : \int_\Omega s_h = 0 \right\}.$$

It is well known that $H^1_h(\Omega) = S_h \oplus \mathbb{R}$, i.e. every function in $H^1_h(\Omega)$ can be (uniquely) expressed as a sum of a function in S_h and a constant. Hence,

$$\begin{aligned} &v_h^i = s_h^i + r^i, \quad \text{where} \\ &s_h^i \in S_h, \\ &r^i \in \mathbb{R} \text{ is a constant.} \end{aligned}$$

From this splitting, we obtain

$$0 \leq |s_h^i|_{H^1_h(\Omega)} = |s_h^i + r^i|_{H^1_h(\Omega)} = |v_h^i|_{H^1_h(\Omega)} \to 0 \text{ as } i \to \infty, \tag{5.12}$$

see (5.10).

This enables us to use the Poincaré inequality to conclude that

$$0 \leq \|s_h^i\|_{L^2_h(\Omega)} \leq C|s_h^i|_{H^1_h(\Omega)} \to 0 \quad \text{as } i \to \infty,$$

i.e.

$$\|s_h^i\|_{L^2_h(\Omega)} \to 0 \quad \text{as } i \to \infty. \tag{5.13}$$

Furthermore, recall that $\|v_h^i\|^2_{L_h^2(\Omega)} = 1$ and that $\int_\Omega s_h^i = 0$. Thus, it follows that

$$\begin{aligned}
1 &= \|v_h^i\|^2_{L_h^2(\Omega)} \\
&= \|s_h^i + r^i\|^2_{L_h^2(\Omega)} \\
&= \|s_h^i\|^2_{L_h^2(\Omega)} + 2(s_h^i, r^i)_{L_h^2(\Omega)} + \|r^i\|^2_{L_h^2(\Omega)} \\
&= \|s_h^i\|^2_{L_h^2(\Omega)} + 2r_i \int_\Omega s_h^i \, dx + \|r^i\|^2_{L_h^2(\Omega)} \\
&= \|s_h^i\|^2_{L_h^2(\Omega)} + |\Omega|^2 (r^i)^2,
\end{aligned}$$

which yields

$$(r^i)^2 = \frac{1}{|\Omega|^2}\left(1 - \|s_h^i\|^2_{L_h^2(\Omega)}\right).$$

By using (5.13) we get

$$r^i = \frac{1}{|\Omega|}\sqrt{1 - \|s_h^i\|^2_{L_h^2(\Omega)}} \to r^* = \frac{1}{|\Omega|} \quad \text{as } i \to \infty.$$

We claim that also the sequence $\{v_h^i\}$ converges toward r^* in $H_h^1(\Omega)$:

$$v_h^i \to r^* = \frac{1}{|\Omega|} \quad \text{in } H_h^1(\Omega).$$

This follows from the fact that $s_h^i = v_h^i - r^i$ and (5.12)-(5.13):

$$\begin{aligned}
\|v_h^i - r^*\|_{H_h^1(\Omega)} &= \|v_h^i - r^i + r^i - r^*\|_{H_h^1(\Omega)} \\
&\le \|v_h^i - r^i\|_{H_h^1(\Omega)} + \|r^i - r^*\|_{H_h^1(\Omega)} \\
&= \|s_h^i\|_{H_h^1(\Omega)} + \|r^i - r^*\|_{H_h^1(\Omega)} \\
&= \|s_h^i\|_{H_h^1(\Omega)} + \|r^i - r^*\|_{L_h^2(\Omega)} \xrightarrow{i\to\infty} 0.
\end{aligned}$$

Here, we have used that $\{r^i\}$ is a sequence of constants and that r^* is a constant, which implies that $\|r^i - r^*\|_{H_h^1(\Omega)} = \|r^i - r^*\|_{L_h^2(\Omega)}$ and that

$$r^i \to r^* \text{ in } \mathbb{R} \Rightarrow \|r^i - r^*\|_{L_h^2(\Omega)} \to 0,$$

provided that Ω has finite measure.

Recall that we assumed that $K : V_h \subset H^1(\Omega) \to Z$ is bounded. Thus,

$$\lim_{i\to\infty} \|Kv_h^i\|^2_{Z_h} = \|K \lim_{i\to\infty} v_h^i\|^2_{Z_h} \tag{5.14}$$

$$= \|Kr^*\|^2_{Z_h} \tag{5.15}$$

$$= (Kr^*, Kr^*)_{Z_h}.$$

By combining these observations with (5.11), we conclude that

$$r^* \in \mathcal{N}(K).$$

To summarize, if (5.9) does not hold, then K annihilates constants. Conversely, if K does not annihilate constants, inequality (5.9) must hold.

We are now ready to show that (5.7)-(5.8) does indeed hold. Note that (5.9) can be written on the form: There exists $c \in (0,1)$ such that

$$\begin{aligned}(\nabla v_h, \nabla v_h)_{L^2_h(\Omega)} + (Kv_h, Kv_h)_{Z_h} \\ \geq c[(\nabla v_h, \nabla v_h)_{L^2_h(\Omega)} + (v_h, v_h)_{L^2_h(\Omega)}], \quad \forall v_h \in V_h.\end{aligned} \tag{5.16}$$

Assume that $x = (v_h, u_h) \in X_h$ satisfies the state equation, i.e.

$$\tilde{A}u_h + \tilde{B}v_h = 0.$$

Then,

$$u_h = -\tilde{A}^{-1}\tilde{B}v_h, \tag{5.17}$$

and since $\tilde{A}^{-1}\tilde{B}$ is bounded

$$\|u_h\|_{U_h} \leq \bar{c}\|v_h\|_{V_h}. \tag{5.18}$$

In addition,

$$Tu_h = -T\tilde{A}^{-1}\tilde{B}v_h = Kv_h. \tag{5.19}$$

Therefore, see (5.4), for $\alpha \in (0,1)$,

$$\begin{aligned}(M_\alpha x, x)_{X_h} &= \alpha(\nabla v_h, \nabla v_h)_{L^2_h(\Omega)} + (T^*Tu_h, u_h)_{U_h} \\ &\geq \alpha[(\nabla v_h, \nabla v_h)_{L^2_h(\Omega)} + (Tu_h, Tu_h)_{Z_h}] \\ &= \alpha[(\nabla v_h, \nabla v_h)_{L^2_h(\Omega)} + (Kv_h, Kv_h)_{Z_h}],\end{aligned}$$

where we have used (5.19). Next, by invoking (5.16) and (5.18) we can conclude that

$$\begin{aligned}(M_\alpha x, x)_{X_h} &\geq \alpha[(\nabla v_h, \nabla v_h)_{L^2_h(\Omega)} + (Kv_h, Kv_h)_{Z_h}], \\ &\geq \alpha c[(\nabla v_h, \nabla v_h)_{L^2_h(\Omega)} + (v_h, v_h)_{L^2_h(\Omega)}], \\ &\geq \alpha c[0.5(\nabla v_h, \nabla v_h)_{L^2_h(\Omega)} + 0.5(v_h, v_h)_{L^2_h(\Omega)} + 0.5\bar{c}^{-2}\|u_h\|^2_{U_h}], \\ &\geq \alpha c \min\{0.5, 0.5\bar{c}^{-2}\}\|(v_h, u_h)\|^2_{X_h}.\end{aligned} \tag{5.20}$$

That is, M_α is coercive on the kernel of N, cf. (5.5). □

We will now use this lemma to establish the main result of this section. First, however, we need two more assumptions:

$\mathcal{A}$5 : The inf-sup condition holds, i.e.

$$\inf_{w \in U_h} \sup_{(v,u) \in V_h \times U_h} \frac{(\tilde{B}v, w)_{U_h} + (\tilde{A}u, w)_{U_h}}{\sqrt{\|v\|^2_{V_h} + \|u\|^2_{U_h}}\|w\|_{U_h}} \geq c > 0. \tag{5.21}$$

$\mathcal{A}$6 : The spectrum of the KKT system $\mathcal{A}_\alpha = \mathcal{R}^{-1}\widehat{\mathcal{A}}_\alpha$ in (5.1) is discrete/countable for every $\alpha \geq 0$. In addition, we assume that (5.1) is severely ill posed for $\alpha = 0$ and $\gamma = 0$, i.e. without regularization,

$$|\tau_i(\mathcal{A}_0)| \leq ce^{-Ci},$$

where $\tau_i(\mathcal{A}_0)$ denotes the ith eigenvalue of $\mathcal{A}_0$ sorted in decreasing order according to their absolute value, and $c, C > 0$ are positive constants.

Let us state the theorem:

THEOREM 5.2. *Assume that all assumptions of Lemma 5.1 hold. In addition, assume that* $\mathcal{A}\mathbf{5}-\mathcal{A}\mathbf{6}$ *hold. Then there exist constants* a, b, $c>0$ *such that, for every* $\alpha>0$, *the spectrum of* $\mathcal{A}_\alpha$, *defined in (5.2), satisfies*

$$\mathrm{sp}(\mathcal{A}_\alpha)\subset[-b,-a]\cup[c\alpha,2\alpha]\cup\{\tau_1,\tau_2,...,\tau_{N(\alpha)}\}\cup[a,b], \tag{5.22}$$

where $N(\alpha)=O(\ln(\alpha^{-1}))$. *The constants* a, b, c *are independent of* α.

Proof. The theorem follows from Lemma 5.1 and the analysis presented in [1]. ∎

Since the spectrum of $\mathcal{A}_\alpha$ is of the form (5.22), we can conclude that the MINRES method will handle the KKT systems (5.1) excellently. More precisely, the number of iterations needed by the MINRES scheme to solve (5.1) can not grow faster than $O([\ln(\alpha^{-1})]^2)$ as $\alpha\to 0$, see [1]. In fact, in practice, iterations counts of order $O(\ln(\alpha^{-1}))$ will in many situations occur, which is also explained in [1].

Note that, while the optimality system (1.6)-(1.7) requires that either $K=-TA^{-1}B:V\to Z$ is injective or that $\gamma>0$ to obtain a unique solution, the inner MINRES algorithm only requires that the constant k does not belong to the null-space of K.

6. Constrained split Bregman algorithm. The split Bregman algorithm we have analysed is in [9] referred to as the *unconstrained* split Bregman method. For some applications, the related *constrained* split Bregman algorithm, also introduced in [9], produces better convergence rates. In order to discuss the latter method, we observe that the problem (2.1) can be formulated on the related, constrained form

$$\min_{v\in V}\left\{\frac{1}{2}\kappa\|v\|^2_{L^2(\Omega)}+\int_\Omega|p|\,dx\right\},$$

subject to

$$\begin{aligned}Kv&=d\quad\text{on }\Omega_{\text{observe}},\\ Dv&=p\quad\text{on }\Omega.\end{aligned}$$

Here, Ω_{observe} is the domain on which the observation data d is defined. The constraints are "implicit" in the sense that they are not necessarily satisfied in each step of the split Bregman algorithm, see [9]. Instead, the scheme generates approximations which converge toward functions satisfying these constraints, and a natural stopping criterion is thus

$$\|Kv^k-d\|_Z<\text{TOL}.$$

Details about the constrained split Bregman algorithm associated with this problem can be found in [16].

It turns out that this constrained approach can also be applied to a PDE-constrained optimization problem, and an experimental investigation gave us better convergence results with this latter approach. We will therefore present the constrained split Bregman algorithm for PDE-constrained optimization problems on the form

$$\min_{v\in V}\left\{\frac{1}{2}\kappa\|v\|^2_{L^2(\Omega)}+\int_\Omega|p|\,dx\right\},$$

subject to

$$\begin{aligned} Au + Bv &= 0, \\ Tu &= d \quad \text{on } \Omega_{\text{observe}}, \\ Dv &= p \quad \text{on } \Omega. \end{aligned}$$

Note that the first constraint here is "explicit", i.e. it must be satisfied in each step of the algorithm. The latter two constraints are "implicit".

Recall the KKT system (4.9) that we derived in connection with Algorithm 2. For the constrained split Bregman method, we get the very similar optimality system

$$\underbrace{\begin{bmatrix} -\alpha\Delta + \gamma E & 0 & B' \\ 0 & T'T & A' \\ B & A & 0 \end{bmatrix}}_{\widehat{\mathcal{A}}_\alpha} \begin{bmatrix} v_h \\ u_h \\ w_h \end{bmatrix} = \begin{bmatrix} -\alpha\nabla \cdot p_h^k + \alpha\nabla \cdot b_h^k \\ T'd_h - T'c_h^k \\ 0 \end{bmatrix}, \tag{6.1}$$

where " $'$ " is used to denote dual operators, and $E : V_h \to V_h'$ is defined by

$$\langle Ev, \phi \rangle = (v, \phi)_{L_h^2(\Omega)}, \quad \phi \in V_h.$$

Compared with (4.9), only the term $-T'c_h^k$ has been added to the second row of the right hand side of (6.1). The operator $\widehat{\mathcal{A}}_\alpha$ on the left hand side is unchanged, and our analysis of the MINRES method, presented above, also applies to this KKT system. The associated algorithm is, of course, similar to Algorithm 2, see Algorithm 3.

Algorithm 3 The *constrained* split Bregman for PDE-constrained optimization problems with TV regularization

1: Choose $v_h^0 = 0, p_h^0 = 0, b_h^0 = 0$
2: **for** k = 0, 1, 2,... **do**
3: Let v_h^{k+1}, u_h^{k+1} and w_h^{k+1} be the solution of (6.1).
4: $p_h^{k+1} = \arg\min_{p_h \in \mathbf{P}_0^h} \int_{\Omega_h} |p_h| + \frac{\lambda}{2}\|\nabla v_h^{k+1} - p_h + b_h^k\|_{L_h^2(\Omega)}^2$,
5: $b_h^{k+1} = b_h^k + \nabla v_h^{k+1} - p_h^{k+1}$,
6: $c_h^{k+1} = c_h^k + Tu_h^{k+1} - d_h$.
7: **end for**

We observe that Algorithm 3 only requires one more simple update compared to Algorithm 2: The update for c_h^{k+1}. This extra computer effort is diminishingly small, and since we obtain better convergence results, we will present numerical experiments with the use of Algorithm 3 only.

7. Numerical experiments.

7.1. Example 1. Let $\Omega = (0,1) \times (0,1)$. We consider the standard example in PDE-constrained optimization, but with TV regularization instead of Tikhonov regularization. That is,

$$\min_{(v,u) \in L^2(\Omega) \times H^1(\Omega)} \left\{ \frac{1}{2}\rho\|Tu - d\|_{L^2(\Omega)}^2 + \int_\Omega |Dv| \right\}, \tag{7.1}$$

subject to

$$-\Delta u + u = v \text{ in } \Omega, \tag{7.2}$$

$$\nabla u \cdot n = 0 \text{ on } \partial\Omega. \tag{7.3}$$

Here, the operator T is the embedding $T : H^1(\Omega) \hookrightarrow L^2(\Omega)$. It is well known that this operator is bounded. Hence, assumption $\mathcal{A}\mathbf{4}$ is satisfied. The control space V, the state space U and the observation space Z are

$$V = BV(\Omega), \tag{7.4}$$

$$U = H^1(\Omega), \tag{7.5}$$

$$Z = L^2(\Omega), \tag{7.6}$$

respectively. Recall that our objective is to solve this system with Algorithm 3. The main challenge is the efficient solution of the KKT systems (6.1). To derive this optimality system, we need the weak formulation of the boundary value problem (7.2)-(7.3).

Computational details. The weak formulation reads: Find $u \in U = H^1(\Omega)$ such that

$$\langle Au, \psi\rangle = -\langle Bv, \psi\rangle \quad \forall \psi \in U,$$

where

$$A : U \to U', \quad u \to \int_\Omega \nabla u \cdot \nabla\psi + u\psi \; dx, \; \forall \phi \in U, \tag{7.7}$$

$$B : V \to U', \quad v \to \int_\Omega v\psi \; dx, \forall \phi \in U. \tag{7.8}$$

From standard PDE theory, we find that A and A^{-1} are bounded. The boundedness of

$$B : V_h \to U_h'$$

in the discrete setting, where one employs the H^1-topology (5.6) on V_h, follows from the inequalities

$$\begin{aligned}\int_\Omega v\psi \; dx &\le \|v\|_{L_h^2(\Omega)} \cdot \|\psi\|_{L_h^2(\Omega)} \\ &\le \sqrt{\|v\|^2_{L_h^2(\Omega)} + |v|^2_{H_h^1(\Omega)}} \cdot \sqrt{\|\psi\|^2_{L_h^2(\Omega)} + |\psi|^2_{H_h^1(\Omega)}} \\ &= \|v\|_{V_h} \cdot \|\psi\|_{H_h^1(\Omega)}.\end{aligned}$$

We conclude that assumptions $\mathcal{A}\mathbf{1}$, $\mathcal{A}\mathbf{2}$ and $\mathcal{A}\mathbf{3}$ are satisfied.

The KKT system to be solved in Algorithm 3 now takes the form

$$\begin{aligned}&\underbrace{\begin{bmatrix} R_{V_h}^{-1} & 0 & 0 \\ 0 & R_{U_h}^{-1} & 0 \\ 0 & 0 & R_{U_h}^{-1} \end{bmatrix}}_{\mathcal{R}^{-1}} \underbrace{\begin{bmatrix} -\alpha\Delta & 0 & B' \\ 0 & T'T & A' \\ B & A & 0 \end{bmatrix}}_{\widehat{\mathcal{A}}_\alpha} \begin{bmatrix} v_h \\ u_h \\ w_h \end{bmatrix} \\ &\qquad = \begin{bmatrix} R_{V_h}^{-1} & 0 & 0 \\ 0 & R_{U_h}^{-1} & 0 \\ 0 & 0 & R_{U_h}^{-1} \end{bmatrix} \begin{bmatrix} -\alpha\nabla \cdot p_h^k + \alpha\nabla \cdot b_h^k \\ T'd_h - T'c_h^k \\ 0 \end{bmatrix}.\end{aligned} \tag{7.9}$$

Recall that $\alpha = \lambda/\rho$, where ρ is the regularization parameter in (7.1) and λ is the parameter employed in the Bregman scheme, see the discussion of (4.5)-(4.8).

The discretization of the operator $\mathcal{R}$ in (7.9) is rather straightforward. Recall from Theorem 4 that the finite dimensional space V_h was equipped with the norm $\|\cdot\|_{H_h^1(\Omega)}$. Furthermore, since $U = H^1(\Omega)$ in this particular example, it follows that the discretization of both of the Riesz maps R_{V_h} and R_{U_h} yields the sum of the mass matrix M and stiffness matrix S.

For the operator $\widehat{\mathcal{A}}_\alpha$ in (7.9), the discretization is more challenging, but a general recipe can be found in [21]. The end result can be summarized as follows:

- A, defined in (7.7), yields the matrix $M + S$, which is the sum of the mass and stiffness matrices associated with the domain Ω.
- B, defined in (7.8), yields the mass matrix M.
- $-\Delta$ yields the stiffness matrix S.
- $T'T = R_{U_h}^{-1} T^* T$ yields the mass matrix M.
- The functions v_h, u_h, w_h, p_h^k, b_h^k, c_h^k and d_h yields the corresponding vectors $\bar{v}$, $\bar{u}$, $\bar{w}$, $\bar{p}^k$, $\bar{b}^k$, $\bar{c}^k$ and $\bar{d}$, respectively.

Hence, the discretized optimality system associated with (7.9) is

$$\underbrace{\begin{bmatrix} (M+S)^{-1} & 0 & 0 \\ 0 & (M+S)^{-1} & 0 \\ 0 & 0 & (M+S)^{-1} \end{bmatrix}}_{\bar{\mathcal{R}}^{-1}} \underbrace{\begin{bmatrix} \alpha S & 0 & M \\ 0 & M & M+S \\ M & M+S & 0 \end{bmatrix}}_{\bar{\mathcal{A}}_\alpha} \underbrace{\begin{bmatrix} \bar{v}^{k+1} \\ \bar{u}^{k+1} \\ \bar{w}^{k+1} \end{bmatrix}}_{\bar{q}^{k+1}}$$

$$= \begin{bmatrix} (M+S)^{-1} & 0 & 0 \\ 0 & (M+S)^{-1} & 0 \\ 0 & 0 & (M+S)^{-1} \end{bmatrix} \underbrace{\begin{bmatrix} -\alpha \nabla \cdot \bar{p}^k + \alpha \nabla \cdot \bar{b}^k \\ M\bar{d} - M\bar{c}^k \\ 0 \end{bmatrix}}_{\bar{g}^k}. \tag{7.10}$$

The preconditioner thus reads

$$\begin{bmatrix} (M+S)^{-1} & 0 & 0 \\ 0 & (M+S)^{-1} & 0 \\ 0 & 0 & (M+S)^{-1} \end{bmatrix}, \tag{7.11}$$

and involves the inverse of the matrix $M+S$. This inverse is computed approximately by using algebraic multigrid (AMG). We discuss this in some more detail in the numerical setup.

Numerical setup.

- We wrote the code using cbc.block, which is a FEniCS-based Python implemented library for block operators. See [23] for details.
- The PyTrilinos package was used to compute an approximation of the preconditioner (7.11). We approximated the inverse using AMG with a symmetric Gauss-Seidel smoother with three smoothing sweeps. All tables containing iteration counts for the MINRES method were generated with this approximate inverse Riesz map. On the other hand, the eigenvalues of the KKT systems $[\bar{\mathcal{R}}]^{-1}\bar{\mathcal{A}}_\alpha$, see (7.10), were computed with an *exact* inverse $[\bar{\mathcal{R}}^k]^{-1}$ computed in Octave.
- To discretize the domain, we divided $\Omega = (0,1) \times (0,1)$ into $N \times N$ squares, and each of these squares were divided into two triangles.

- The MINRES iteration process was stopped as soon as

$$\frac{\|r_n^k\|}{\|r_0^k\|} = \left[\frac{(\ \bar{\mathcal{A}}_\alpha \bar{q}_n^k - \bar{g}^k, [\bar{\mathcal{R}}]^{-1}[\bar{\mathcal{A}}_\alpha \bar{q}_n^k - \bar{g}^k]\ )}{(\ \bar{\mathcal{A}}_\alpha \bar{q}_0^k - \bar{g}^k, [\bar{\mathcal{R}}]^{-1}[\bar{\mathcal{A}}_\alpha \bar{q}_0^k - \bar{g}^k]\ )}\right]^{1/2} < \epsilon. \tag{7.12}$$

  Here, ϵ is a small positive parameter. Note that the superindex k is the iteration index for the "outer" split Bregman method, while the subindex n is the iteration index for the "inner" MINRES algorithm (at each step of the split Bregman method).
- No noise was added to the input data d, see (7.1).

Results. We are now ready to solve the problem (7.1)-(7.3). The synthetic data d was produced by setting

$$v(x) = \begin{cases} -5 & \text{if } x_2 < 0.5, \\ 7 & \text{if } x_2 > 0.5, \end{cases} \tag{7.13}$$

and then we solved the boundary value problem (7.2)-(7.3) with (7.13) as input. The data d was thereafter set equal to the solution u throughout the entire domain $\Omega = (0,1) \times (0,1)$.

Theorem 4 states that the KKT system (5.1)-(5.2) arising in each iteration of the split Bregman iteration has a spectrum of the form (5.22). In Figure 7.1, we see a spectrum of such a KKT system, and it is clearly of the form (5.22). Hence, we should expect the MINRES algorithm to solve the problem efficiently.

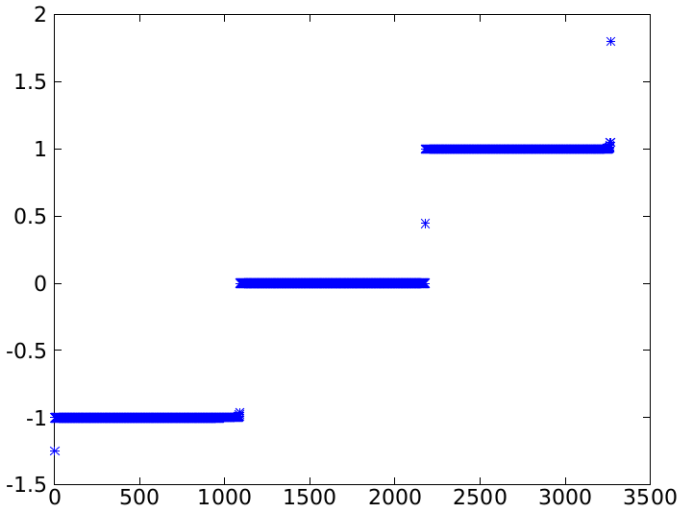

FIG. 7.1. *The eigenvalues of* $[\bar{\mathcal{R}}^k]^{-1}\bar{\mathcal{A}}_\alpha$ *in Example 1. Here,* $\alpha = 0.0001$ *and* $N = 32$, *i.e.* $h = 1/32$. *(*$[\bar{\mathcal{R}}^k]^{-1}$ *denotes the exact inverse of the preconditioner - not its AMG approximation).*

Table 7.3 illuminates the theoretically established convergence behavior of the MINRES algorithm. As previously mentioned, in [1] the authors proved that the number of iterations can not grow faster than $O([\ln(\alpha^{-1})]^2)$, and showed why iteration growth of $O(\ln(\alpha^{-1}))$ often occur in practice. For $\epsilon = 10^{-6}$, see (7.12), and $N = 256$, we get the following estimate for the iteration growth

$$40.2 - 21.6 \log_{10}(\alpha),$$

where the coefficients are computed by the least squares method. The growth is very well modeled by this formula. Similarly, for $\epsilon = 10^{-10}$ and $N = 256$, we can model the growth by the formula

$$57.6 - 43.5 \log_{10}(\alpha).$$

N\α	1	.1	.01	.001	.0001
32	22	37	47	59	73
64	31	51	63	81	102
128	26	42	59	75	97
256	39	62	84	108	124

TABLE 7.1
Stopping criterion $\epsilon = 10^{-6}$.

N\α	1	.1	.01	.001	.0001
32	32	61	81	98	116
64	43	82	115	143	173
128	40	74	110	142	170
256	54	103	152	182	232

TABLE 7.2
Stopping criterion $\epsilon = 10^{-10}$.

TABLE 7.3
The average number of MINRES iterations required to solve the KKT systems arising in the first ten steps of the split Bregman algorithm in Example 1. The two panels display the iteration counts for two different choices of ϵ, see (7.12).

In Figure 7.4, two approximate solutions of the optimization problem (7.1)-(7.3) are displayed. One after 10 Bregman iterations and the other after 100 Bregman iterations. From this figure, we observe that the jump is "found" after the first 10 iterations, cf. (7.13). The subsequent iterations merely "tightens" the jump and levels out the other parts of the solution. This behavior is similar to the one described for the image denoising algorithm in [9], where the authors also gave an explanation for why the split Bregman algorithm would quickly localize the jump(s).

Remark. As mentioned above, the problem (7.1)-(7.3), with Tikhonov regularization instead of TV regularization, has been analyzed by many scientists. In fact, for Tikhonov regularization a number of numerical schemes that are completely robust with respect to the size of the regularization parameter have been developed [24, 25, 26]: Even logarithmic growth in iterations counts is avoided. As far as the authors knows, it is not know whether these techniques can be adapted to the saddle point problem (7.9).

7.2. Example 2. We will now explore a more challenging problem. Let the domain Ω still be the unit square. Furthermore, define

$$\tilde{\Omega} = (1/4, 3/4) \times (1/4, 3/4).$$

The problem we want to study is

$$\min_{(v,u)\in L^2(\tilde{\Omega})\times H^1(\Omega)} \left\{ \frac{1}{2}\rho \|Tu - d\|^2_{L^2(\partial\Omega)} + \int_{\tilde{\Omega}} |Dv| \right\}, \tag{7.14}$$

subject to

$$-\Delta u + u = \begin{cases} -v & \text{if } x \in \tilde{\Omega}, \\ 0 & \text{if } x \in \Omega \setminus \tilde{\Omega}, \end{cases} \tag{7.15}$$

$$\nabla u \cdot n = 0 \text{ on } \partial\Omega. \tag{7.16}$$

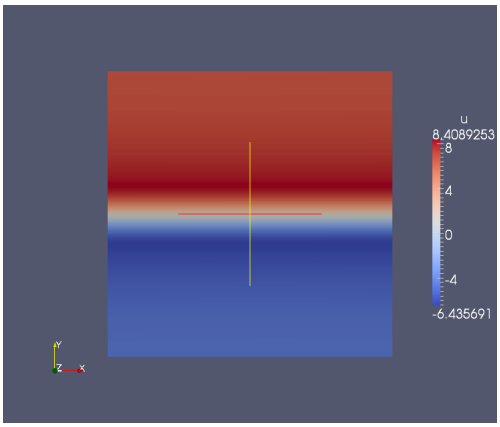

FIG. 7.2. *Approximate inverse solution* v_h^{10} *after 10 split Bregman iterations.*

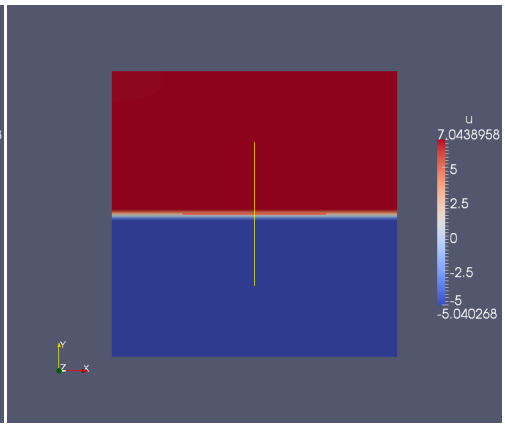

FIG. 7.3. *Approximate inverse solution* v_h^{100} *after 100 split Bregman iterations.*

FIG. 7.4. *The solution of the problem (7.1)-(7.3). Here,* $\epsilon = 10^{-6}$, $\alpha = 10^{-6}$ *and* $N = 128$ *(*$h = 1/128$*).*

We observe two differences between examples 1 and 2. First, the control domain $\tilde{\Omega}$ is now a subdomain of the entire domain Ω, bounded strictly away from the boundary $\partial\Omega$. Secondly, the observation domain is reduced from the entire domain Ω to the boundary $\partial\Omega$. In total, this results in a severely ill-posed problem for $\alpha = \lambda/\rho = 0$, see [1]. Information about λ, ρ and α is provided in connection with the discussion of equations (4.5)-(4.8).

Here, the operator $T : H^1(\Omega) \to L^2(\partial\Omega)$ is the continuous trace operator. Hence, assumption $\mathcal{A}\mathbf{4}$ is satisfied. The control space V, the state space U and the observation space Z are

$$V = BV(\tilde{\Omega}), \tag{7.17}$$

$$U = H^1(\Omega), \tag{7.18}$$

$$Z = L^2(\partial\Omega), \tag{7.19}$$

respectively. Since the discretization of (7.14)-(7.16) is very similar to the discretization of (7.1)-(7.3), we do not enter into all the details. Instead, we only focus on the differences.

The weak formulation of the state equations (7.15)-(7.16) reads: Find $u \in U = H^1(\Omega)$ such that

$$\langle Au, \psi\rangle = -\langle Bv, \psi\rangle \quad \forall \psi \in U,$$

where the operator A is still defined as in (7.7). The operator B, however, is no longer as in (7.8), but is here defined by

$$B : V \to U', \quad v \to \int_{\tilde{\Omega}} v\psi, \forall \phi \in U, \tag{7.20}$$

and in the discrete setting we employ the norm

$$\|\cdot\|_{V_h}^2 = \|\cdot\|_{L_h^2(\tilde{\Omega})}^2 + |\cdot|_{H_h^1(\tilde{\Omega})}^2$$

on the control space V_h. From standard PDE theory, we can guarantee that A and A^{-1} are bounded, and the boundedness of B is verified in a manner very similar to the argument presented in connection with Example 1:

$$\begin{aligned}\int_{\tilde{\Omega}} v\psi \, dx &\leq \|v\|_{V_h} \cdot \|\psi\|_{H_h^1(\tilde{\Omega})} \\ &\leq \|v\|_{V_h} \cdot \|\psi\|_{H_h^1(\Omega)}\end{aligned}$$

because $\tilde{\Omega}$ is a subdomain of Ω. We conclude that assumptions $\mathcal{A}\mathbf{1}$, $\mathcal{A}\mathbf{2}$ and $\mathcal{A}\mathbf{3}$ are satisfied.

The new control domain $\tilde{\Omega}$ and the redefined operators B and T lead to some changes in the discretization of the optimality system (6.1), which must be solved repeatedly in Algorithm 3. These can be summarized as follows:

- B, defined in (7.20), yields the mass matrix $\tilde{M}$ associated with the subdomain $\tilde{\Omega}$.
- $-\Delta$ yields the stiffness matrix $\tilde{S}$ associated with the subdomain $\tilde{\Omega}$.
- $T'T = R_{U_h}^{-1} T^* T$ yields the "boundary" mass matrix M_∂.
- The Riesz map R_{V_h} now yields the sum of the mass matrix $\tilde{M}$ and stiffness matrix $\tilde{S}$.

All other operators are discretized in the same fashion as in Example 1. Hence, the discretized optimality system in Algorithm 3, associated with (7.14)-(7.16), takes the form

$$\begin{aligned}&\underbrace{\begin{bmatrix}(\tilde{M}+\tilde{S})^{-1} & 0 & 0 \\ 0 & (M+S)^{-1} & 0 \\ 0 & 0 & (M+S)^{-1}\end{bmatrix}}_{\bar{\mathcal{R}}^{-1}} \underbrace{\begin{bmatrix}\alpha\tilde{S} & 0 & \tilde{M} \\ 0 & M_\partial & M+S \\ \tilde{M} & M+S & 0\end{bmatrix}}_{\bar{\mathcal{A}}_\alpha} \underbrace{\begin{bmatrix}\bar{v}^{k+1} \\ \bar{u}^{k+1} \\ \bar{w}^{k+1}\end{bmatrix}}_{\bar{q}^{k+1}} \\ &\qquad = \begin{bmatrix}(\tilde{M}+\tilde{S})^{-1} & 0 & 0 \\ 0 & (M+S)^{-1} & 0 \\ 0 & 0 & (M+S)^{-1}\end{bmatrix} \underbrace{\begin{bmatrix}-\alpha\nabla\cdot\bar{p}^k + \alpha\nabla\cdot\bar{b}^k \\ M_\partial\bar{d} - M_\partial\bar{c}^k \\ 0\end{bmatrix}}_{\bar{g}^k}.\end{aligned} \tag{7.21}$$

The preconditioner thus reads

$$\begin{bmatrix}(\tilde{M}+\tilde{S})^{-1} & 0 & 0 \\ 0 & (M+S)^{-1} & 0 \\ 0 & 0 & (M+S)^{-1}\end{bmatrix}. \tag{7.22}$$

Results. The synthetic data d was produced in the same manner as in Example 1. We computed the synthetic data from the function $v \in L^2(\tilde{\Omega})$, where

$$v(x) = \begin{cases} 5 & \text{if } x_1 < 0.5, \\ -5 & \text{if } x_1 > 0.5. \end{cases} \tag{7.23}$$

Note that the forward operator $K = -TA^{-1}B$ does not guarantee a unique solution of (7.14)-(7.16), since the trace operator is not injective, see Theorem 2. Nevertheless, the forward operator K does not annihilate constants, and from Theorem 4, it then follows that the MINRES algorithm should handle the KKT systems, arising in each Bregman iteration, very well.

Figure 7.5 shows the spectrum of $[\bar{\mathcal{R}}^k]^{-1}\bar{\mathcal{A}}_\alpha$ for this example. This eigenvalue distribution is clearly on the form (5.22). Hence, in accordance with Theorem 4, we obtain such a spectrum even though $K = -TA^{-1}B$ is not injective (and $\kappa = 0$ in these computations).

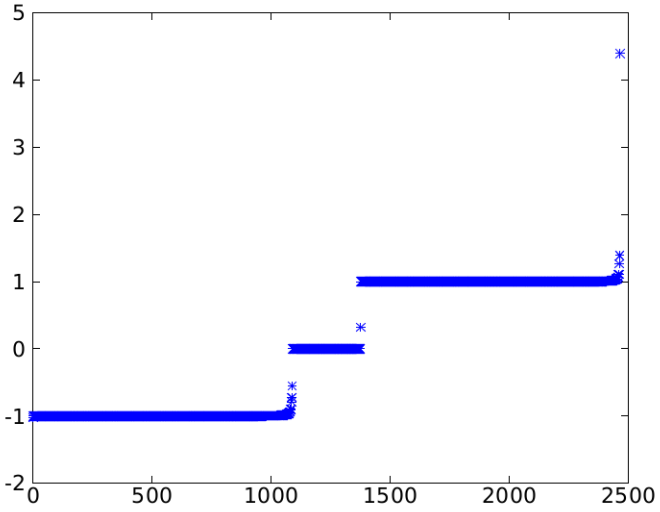

FIG. 7.5. *The eigenvalues of* $[\bar{\mathcal{R}}^k]^{-1}\bar{\mathcal{A}}_\alpha$ *in Example 2. Here,* $\alpha = 0.0001$ *and* $N = 32$. *(*$[\bar{\mathcal{R}}^k]^{-1}$ *denotes the exact inverse of the preconditioner - not its AMG approximation).*

Table 7.6 displays the iteration counts for Example 2. We see that the growth in the iteration numbers, as α decreases, is handled well by the MINRES algorithm. For example, for the case of $N = 256$ and $\epsilon = 10^{-6}$, the growth can be modeled by the formula

$$40.8 - 16.2\log_{10}(\alpha).$$

Similarly, for $N = 256$ and $\epsilon = 10^{-10}$, the least squares method gives us the formula

$$58.2 - 35.6\log_{10}(\alpha),$$

as the best logarithmic fit of iteration growth.

The approximate solutions, seen in Figure 7.8, are very close to the "input solution" (7.23). We thus get very good approximations even though we can not guarantee a unique solution ($\kappa = 0$, see Theorem 2).

8. Conclusions. We have studied PDE-constrained optimization problems subject to TV regularization. The main purpose of this text was to adapt the split Bregman algorithm, frequently used in imaging analysis, to this kind of problems.

In each iteration of the split Bregman scheme, a large KKT system

$$\mathcal{A}_\alpha q = g \tag{8.1}$$

must be solved. Here, $0 < \alpha \ll 1$ is a regularization parameter, and the spectral condition number of $\mathcal{A}_\alpha$ tends to ∞ as $\alpha \to 0$. We investigated the performance of

N\α	1	.1	.01	.001	.0001
32	29	44	49	55	63
64	34	48	58	67	82
128	36	52	59	69	84
256	41	60	71	84	110

TABLE 7.4
Stopping criterion $\epsilon = 10^{-6}$.

N\α	1	.1	.01	.001	.0001
32	41	65	82	100	109
64	47	76	104	126	154
128	50	84	112	144	169
256	57	95	131	163	201

TABLE 7.5
Stopping criterion $\epsilon = 10^{-10}$.

TABLE 7.6
The average number of MINRES iterations required to solve the KKT systems arising in the first ten steps of the split Bregman algorithm in Example 2. The two panels display the iteration counts for two different choices of ϵ*, see (7.12).*

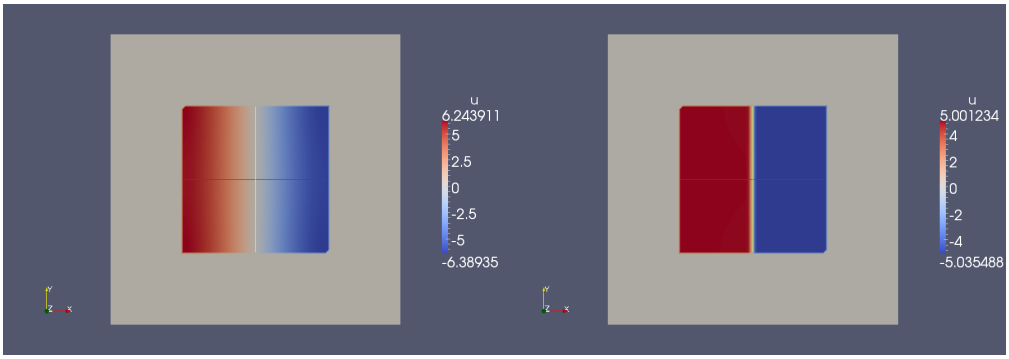

FIG. 7.6. *Approximate inverse solution* v_h^{10} *after 10 split Bregman iterations.*

FIG. 7.7. *Approximate inverse* v_h^{100} *solution after 100 split Bregman iterations.*

FIG. 7.8. *The solution of the problem (7.14)-(7.16). Here* $\epsilon = 10^{-6}$, $\alpha = 1000$, $\lambda = 0.0005$ *and* $N = 128$.

the MINRES algorithm applied to these indefinite systems. In particular, we analyzed the spectrum of $\mathcal{A}_\alpha$, and our main result shows that this spectrum is almost contained in three bounded intervals, with a small number of isolated eigenvalues. More specifically, we found that

$$\text{sp}(\mathcal{A}_\alpha) \subset [-b, -a] \cup [c\alpha, 2\alpha] \cup \{\tau_1, \tau_2, ..., \tau_{N(\alpha)}\} \cup [a, b], \tag{8.2}$$

where $N(\alpha) = O(\ln(\alpha^{-1}))$. Krylov subspace solvers therefore handle (8.1) very well: The number of iterations required by the MINRES method can not grow faster than $O([\ln(\alpha^{-1})]^2)$ as $\alpha \to 0$, and in practice one will often encounter growth rates of order $O(\ln(\alpha^{-1}))$.

Our theoretical findings were illuminated by numerical experiments. In both examples we observed approximately logarithmic growth in iteration numbers as $\alpha \to 0$. This is in accordance with our theoretical results.

Acknowledgements. The authors would like to thank the FEniCS community for their work on the automatic solution of PDEs.

REFERENCES

[1] Nielsen BF, Mardal KA. Analysis of the minimal residual method applied to ill-posed optimality systems. *SIAM Journal on Scientific Computing* 2013; **35**(2):785–814.
[2] Rudin L, Osher S, Fatemi E. Nonlinear total variation based noise removal algorithms. *Physica D* 1992; **60**:259–268.
[3] Acar R, Vogel CR. Analysis of bounded variation penalty methods for ill-posed problems. *Inverse Problems* 1994; **10**:1217–1229.
[4] Vogel CR, Oman ME. Fast, robust total variation based reconstruction of noisy, blurred images. *Image Processing, IEEE Transactions on* 1998; **7**(6):813–824.
[5] Chan TF, Tai XC. Augmented Lagrangian and total variation methods for recovering discontinuous coefficients from elliptic equations. *Computational Science for the 21st Century*. John Wiley and Sons, 1997; 597–607.
[6] Chen Z, Zou J. An augmented Lagrangian method for identifying discontinuous parameters in elliptic systems. *SIAM Journal on Control and Optimization* 1999; **37**(3):892–910.
[7] Ito K, Kunisch K, Li Z. Level-set function approach to an inverse interface problem. *Inverse Problems* 2001; .
[8] Chan TF, Tai XC. Level set and total variation regularization for elliptic inverse problems with discontinuous coefficients. *Journal of Computational Physics* 2004; **193**:40–66.
[9] Goldstein T, Osher S. The split Bregman method for L1-regularized problems. *SIAM Journal on Imaging Sciences* 2002; **2**:323–343.
[10] Chavent G, Kunish K. Regularization of linear least squares problems by total bounded variation. *ESAIM: Control, Optimisation and Calculus of Variations* 1997; **2**:359–376.
[11] Vogel CR, Oman ME. Iterative methods for total variation denoising. *SIAM Journal on Scientific Computing* 1996; **17**:227–238.
[12] Chan TF, Golub GH, Mulet P. A nonlinear primal-dual method for total variation-based image restoration. *SIAM Journal on Scientific Computing* 1999; **20**(6):1964–1977.
[13] Brgman LM. A relaxation method of finding a common point of convex sets and its application to the solution of problems in convex programming. *Zh. vychisl. Mat. mat. Fiz* 1967; **7**:620–631.
[14] Osher S, Burger M, Goldfarb D, Xu J, Yin W. An iterative regularization method for total variation-based image restoration. *Multiscale Model. Simul* 2005; **4**:460–489.
[15] Meyer Y. *Oscillating Patterns in Image Processing and Nonlinear Evolution Equations*. American Mathematical Society: Providence, Rhode Island, 2001.
[16] Cai JF, Osher S, Shen Z. Split Bregman methods and frame based image restoration. *Multiscale modeling and simulation* 2009; :337–369.
[17] Yin W, Osher S, Goldfarb D, Darbon J. Bregman iterative algorithms for L1-minimization with applications to compressed sensing. *SIAM Journal on Imaging Sciences* 2008; **1**(1):143–168.
[18] Wu C, Tai XC. Augmented Lagrangian method, dual method and split-Bregman iterations for ROF, vectorial TV and higher order models. *SIAM Journal on Imaging Sciences* 2010; **3**(3):300–339.
[19] Hestenes MR. Multiplier and gradient methods. *Journal of Optimization Theory and Applications* 1969; **4**:302–320.
[20] Powell MJD. A method for nonlinear constraints in minimization problems. *In Optimization ed. by R. Fletcher* 1969; :283–298.
[21] Mardal KA, Winther R. Preconditioning discretizations of systems of partial differential equations. *Numerical Linear Algebra with Applications* 2011; **18**(1):1–40.
[22] Fitzpatrick BG, Keeling SL. On approximation in total variation penalization for image reconstruction. *Numerical Functional Analysis and Optimization* 1997; **18**:941–958.
[23] Mardal KA, Haga JB. Block preconditioning of systems of PDEs. *Automated Solution of Differential Equations*, Logg A, Mardal KA, Wells G (eds.). Springer, 2012; 635–654.
[24] Schöberl J, Zulehner W. Symmetric indefinite preconditioners for saddle point problems with applications to PDE-constrained optimization problems. *SIAM Journal on Matrix Analysis and Applications* 2007; **29**(3):752–773.
[25] Schöberl J, Simon R, Zulehner W. A robust multigrid method for elliptic optimal control problems. *SIAM Journal on Numerical Analysis* 2011; **49**(4):1482–1503.
[26] Pearson JW, Wathen AJ. A new approximation of the Schur complement in preconditioners for PDE-constrained optimization. *Numerical Linear Algebra with Applications* 2012; **19**(5):816–829.