A SUBSPACE-ACCELERATED SPLIT BREGMAN METHOD FOR SPARSE DATA RECOVERY WITH JOINT ℓ_1-TYPE REGULARIZERS *

VALENTINA DE SIMONE†, DANIELA DI SERAFINO†, AND MARCO VIOLA†

VERSION 2 – Mar 23, 2020

Abstract. We propose a subspace-accelerated Bregman method for the linearly constrained minimization of functions of the form $f(\mathbf{u}) + \tau_1 \|\mathbf{u}\|_1 + \tau_2 \|D\,\mathbf{u}\|_1$, where f is a smooth convex function and D represents a linear operator, e.g. a finite difference operator, as in anisotropic Total Variation and fused-lasso regularizations. Problems of this type arise in a wide variety of applications, including portfolio optimization, learning of predictive models from functional Magnetic Resonance Imaging (fMRI) data, and source detection problems in electroencephalography. The use of $\|D\,\mathbf{u}\|_1$ is aimed at encouraging structured sparsity in the solution. The subspaces where the acceleration is performed are selected so that the restriction of the objective function is a smooth function in a neighborhood of the current iterate. Numerical experiments on multi-period portfolio selection problems using real datasets show the effectiveness of the proposed method.

Key words. split Bregman method, subspace acceleration, joint ℓ_1-type regularizers, multi-period portfolio optimization.

AMS subject classifications. 65K05, 90C25.

1. Introduction.

We are interested in the solution of problems of the form

$$\begin{array}{ll} \min & f(\mathbf{u}) + \tau_1 \|\mathbf{u}\|_1 + \tau_2 \|D\,\mathbf{u}\|_1 \\ \text{s.t.} & A\mathbf{u} = \mathbf{b}, \end{array} \tag{1.1}$$

where $f : \mathbb{R}^n \to \mathbb{R}$ is a closed convex function at least twice continuously differentiable, $\mathbf{u} \in \mathbb{R}^n$, $D \in \mathbb{R}^{q\times n}$, $A \in \mathbb{R}^{m\times n}$, and $\mathbf{b} \in \mathbb{R}^m$. The ℓ_1 regularization term in the objective function encourages sparsity in the solution while the use of $\|D\,\mathbf{u}\|_1$ is aimed at incorporating further information about the solution. For example, in the case of discrete anisoptropic Total Variation [23, 26], D is a first-order finite-difference operator and the regularization encourages smoothness along certain directions. The combination of the two regularization terms can be seen as a generalization of the fused lasso regularization introduced in [35] in the case of least-squares regression. Problems of the form (1.1) arise, e.g., in multi-period portfolio optimization [16], in predictive modeling and classification (machine learning) on functional Magnetic Resonance Imaging (fMRI) data [2, 19], in source detection problems in electroencephalography [4], and in multiple change-point detection [30].

Methods based on Bregman iterations [6, 10, 26, 31] have proved to be efficient in the solution of this type of problems. As we will see in Section 3, the Bregman iterative scheme requires at each step the solution of an ℓ_1-regularized unconstrained optimization subproblem. For this minimization, which does not need to be performed exactly, but generally requires high accuracy (see Theorem 3.1 in Section 3), one can use iterative methods suited to deal with the ℓ_1-regularization term, such as FISTA [3], SpaRSA [36], BOSVS [14] and ADMM [5]. A possible drawback is that these methods may be inefficient when high accuracy is required.

Herein, we propose a subspace-acceleration strategy for the Bregman iterative scheme, which is aimed at replacing, at certain steps, the unconstrained minimization of the ℓ_1-regularized subproblem with the unconstrained minimization of a smooth restriction of it to a

*This work was partially supported by Gruppo Nazionale per il Calcolo Scientifico - Istituto Nazionale di Alta Matematica (GNCS-INdAM) and by the VAIN-HOPES Project, funded by the 2019 V:ALERE (VAnviteLli pEr la RicErca) Program of the University of Campania "L. Vanvitelli".

†Department of Mathematics and Physics, University of Campania "L. Vanvitelli", viale A. Lincoln, 5, Caserta (Italy), {valentina.desimone, daniela.diserafino, marco.viola}@unicampania.it

suitable subspace. The proposed strategy finds its roots in the class of orthant-based methods [1, 9, 28] for ℓ_1-regularized minimization problems, which are based on the consecutive minimization of smooth approximations to the problem over a sequence of orthants. However, by following [12], instead of considering the restriction to the full orthant, we restrict the minimization to the orthant face identified by the zero variables. Ideally, one would like to perform this subspace minimization only when there is guarantee that the subproblem solution will lie on that orthant face. However, this is unpractical to check. For this reason, starting from the work in [12], we introduce a switching criterion to decide whether to perform the subspace-acceleration step. The criterion is based on the use of some optimality measures for the current iterate with respect to the current subproblem. More specifically, it is based on a comparison between a measure of the optimality violation of the zero variables and a measure of the optimality violation of the other variables. This strategy comes from the adaptation to ℓ_1-regularized optimization of the concept of proportional iterates, developed in the case of quadratic optimization problems subject to bound constraints or to bound constraints and a single linear equality constraint [18, 20, 21, 24, 25, 29].

The idea of introducing acceleration steps over suitable subspaces to improve the performance of splitting methods for problem (1.1) is not new. An example is provided, e.g., by [11]. However, the strategy we propose in this work differs from that subspace-acceleration strategy because we focus on Bregman iterations and aim at replacing nonsmooth unconstrained subproblems with smooth smaller ones, while the algorithm in [11] is based on the introduction of a subspace-acceleration step after the minimization steps in an ADMM algorithm [5], where the subspace is spanned by directions obtained by using information from previous iterations.

This paper is organized as follows. In Section 2 we recall some convex analysis concepts that will be used later in this work. In Section 3 we briefly describe the Bregman iterative scheme for the solution of problem (1.1) and prove its convergence in the case of inexact subproblem minimization. In Section 4 we show how suitable subspace-acceleration steps can be introduced into the Bregman iterative scheme. In Section 5 we report numerical results for the solution of portfolio optimization problems modeled by (1.1). We provide some conclusions in Section 6.

Notation. Throughout this paper scalars are denoted by lightface Roman or Greek fonts, e.g., $a, \alpha \in \mathbb{R}$, vectors by boldface Roman or Greek fonts, e.g., $\mathbf{v}, \boldsymbol{\mu} \in \mathbb{R}^n$. The i-th entry of a vector $\mathbf{v} \in \mathbb{R}^n$ is denoted v_i or $[v]_i$. Given a continuously differentiable function $F(\mathbf{x}) : \mathbb{R}^n \to \mathbb{R}$, we use $\nabla_i F(\mathbf{x})$ to indicate the first derivative of F with respect to the variable x_i. We use $\mathbf{0}_n$ and $\mathbf{1}_n$ to indicate the vectors in $\mathbb{R}^n$ with all entries equal to 0 and 1, respectively; the subscript is omitted if the dimension is clear from the context. For any vectors $\mathbf{u} \in \mathbb{R}^{n_1}$ and $\mathbf{v} \in \mathbb{R}^{n_2}$ we use the notation $[\mathbf{u}; \mathbf{v}]$ to represent the vector $[\mathbf{u}^\top, \mathbf{v}^\top]^\top \in \mathbb{R}^{n_1+n_2}$. The Euclidean scalar product between $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$ is indicated as $\langle \mathbf{u}, \mathbf{v} \rangle$. Norms $\|\cdot\|$ are ℓ_2. Superscripts are used to denote the elements of a sequence, e.g., $\{\mathbf{x}^k\}$.

2. Preliminaries.

We recall some concepts that will be used in the next sections.

DEFINITION 2.1. *Given a function $Q : \mathbb{R}^n \to \mathbb{R}$, the convex conjugate Q^* of Q is defined as follows:*

$$Q^*(\mathbf{y}) = \sup_{\mathbf{x}} \langle \mathbf{y}, \mathbf{x} \rangle - Q(\mathbf{x}).$$

Note that Q^* is a closed convex function for any given Q. If Q is strictly convex, then Q^* is also continuously differentiable; moreover, if Q is a closed convex function, then

$Q^{**}(\mathbf{x}) = Q(\mathbf{x})$ [27].

DEFINITION 2.2. *Given a closed convex function $Q : \mathbb{R}^n \to \mathbb{R}$, a vector $\mathbf{p} \in \mathbb{R}^n$ is said a subgradient of Q at a point $\mathbf{x} \in \mathbb{R}^n$ if*

$$Q(\mathbf{y}) - Q(\mathbf{x}) \geq \langle \mathbf{p},\, \mathbf{y} - \mathbf{x} \rangle\,, \quad \forall\, \mathbf{y} \in \mathbb{R}^n.$$

The set of all the subgradients of Q at $\mathbf{x}$ is referred to as the subdifferential of Q at $\mathbf{x}$, and is denoted $\partial Q(\mathbf{x})$.

If Q is a closed convex function, then [27, Chapter X]

$$\mathbf{p} \in \partial Q(\mathbf{x}) \quad \text{if and only if} \quad \mathbf{x} \in \partial Q^*(\mathbf{p}). \tag{2.1}$$

Moreover, we have that $Q(\mathbf{x}) + Q^*(\mathbf{p}) = \langle \mathbf{p},\, \mathbf{x} \rangle$.

DEFINITION 2.3. *A point-to-set map $\Phi : \mathbb{R}^n \to 2^{\mathbb{R}^n}$ is said to be a monotone operator if*

$$\langle \mathbf{x} - \mathbf{y},\, \mathbf{u} - \mathbf{v} \rangle \geq 0, \qquad \textit{for all } \ \mathbf{x}, \mathbf{y} \in \mathbb{R}^n,\ \mathbf{u} \in \Phi(\mathbf{x}),\ \mathbf{v} \in \Phi(\mathbf{y}).$$

Moreover, Φ is said to be maximal monotone if it is monotone and its graph, i.e., the set

$$\{(\mathbf{x}, \mathbf{y}) \in \mathbb{R}^n \times \mathbb{R}^n \ :\ \mathbf{y} \in \Psi(\mathbf{x})\}\,,$$

is not strictly contained in the graph of any other monotone operator.

An example of maximal monotone operator is the subdifferential of a lower-semicontinuous convex function (see [33] and references therein).

DEFINITION 2.4. *Given an operator $\Phi : \mathbb{R}^n \to 2^{\mathbb{R}^n}$, the inverse of Φ is the operator $\Phi^{-1} : \mathbb{R}^n \to 2^{\mathbb{R}^n}$ defined as*

$$\Phi^{-1}(\mathbf{y}) = \{\mathbf{x} \in \mathbb{R}^n \ :\ \mathbf{y} \in \Phi(\mathbf{x})\}\,.$$

3. The split Bregman method. For the sake of simplicity and self consistency, we briefly describe the split Bregman method [26] for the solution of ℓ_1-regularized problems of type (1.1). In order to separate the two ℓ_1-regularization terms, we introduce the auxiliary variable $\mathbf{d} = D\mathbf{u}$, so that problem (1.1) can be reformulated as

$$\begin{array}{ll} \min & E(\mathbf{u}, \mathbf{d}) \equiv f(\mathbf{u}) + \tau_1 \|\mathbf{u}\|_1 + \tau_2 \|\mathbf{d}\|_1 \\ \text{s.t.} & A\mathbf{u} \qquad = \mathbf{b}, \\ & D\,\mathbf{u} - \mathbf{d} = \mathbf{0}. \end{array} \tag{3.1}$$

The split Bregman method is based on a Bregman iterative scheme for the solution of (3.1). Letting $\mathbf{u}^0 \in \mathbb{R}^n$, $\mathbf{d}^0 \in \mathbb{R}^q$, and $\mathbf{p}^0 = \left[\mathbf{p}_\mathbf{u}^0;\, \mathbf{p}_\mathbf{d}^0\right] \in \partial E(\mathbf{u}^0, \mathbf{d}^0)$, the k-th iteration of the Bregman method reads as follows:

$$\left[\mathbf{u}^{k+1};\, \mathbf{d}^{k+1}\right] = \operatorname*{argmin}_{\mathbf{u},\mathbf{d}} \mathcal{D}_E^{\mathbf{p}^k}\left([\mathbf{u};\, \mathbf{d}]\,,\, \left[\mathbf{u}^k; \mathbf{d}^k\right]\right) + \frac{\lambda}{2}\|A\,\mathbf{u} - \mathbf{b}\|^2 + \frac{\lambda}{2}\|D\,\mathbf{u} - \mathbf{d}\|^2, \tag{3.2}$$

$$\mathbf{p}_\mathbf{u}^{k+1} = \mathbf{p}_\mathbf{u}^k - \lambda A^\top(A\mathbf{u}^{k+1} - \mathbf{b}) - \lambda D^\top(D\mathbf{u}^{k+1} - \mathbf{d}^{k+1}), \tag{3.3}$$

$$\mathbf{p}_\mathbf{d}^{k+1} = \mathbf{p}_\mathbf{d}^k - \lambda(\mathbf{d}^{k+1} - D\mathbf{u}^{k+1}), \tag{3.4}$$

where $\mathbf{p}^k = \left[\mathbf{p}_\mathbf{u}^k;\, \mathbf{p}_\mathbf{d}^k\right]$ and

$$\mathcal{D}_E^{\bar{\mathbf{p}}}\left([\mathbf{u};\, \mathbf{d}],\, [\bar{\mathbf{u}};\, \bar{\mathbf{d}}]\right) = E(\mathbf{u}, \mathbf{d}) - E(\bar{\mathbf{u}}, \bar{\mathbf{d}}) - \langle \bar{\mathbf{p}}_\mathbf{u},\, \mathbf{u} - \bar{\mathbf{u}} \rangle - \left\langle \bar{\mathbf{p}}_\mathbf{d},\, \mathbf{d} - \bar{\mathbf{d}} \right\rangle,$$

with $\bar{\mathbf{p}} \in \partial E(\bar{\mathbf{u}}, \bar{\mathbf{d}})$, is the so-called Bregman distance associated with the convex function E at the point $[\bar{\mathbf{u}};\, \bar{\mathbf{d}}]$.

Following [26, 31], thanks to the linearity of the equality constraints, a simplified iteration can be used in place of the original Bregman one:

$$\left[\mathbf{u}^{k+1};\, \mathbf{d}^{k+1}\right] = \operatorname*{argmin}_{\mathbf{u},\mathbf{d}} E(\mathbf{u},\mathbf{d}) + \frac{\lambda}{2}\|A\,\mathbf{u} - \mathbf{b}_{\mathbf{u}}^{k}\|^2 + \frac{\lambda}{2}\|D\,\mathbf{u} - \mathbf{d} - \mathbf{b}_{\mathbf{d}}^{k}\|^2, \tag{3.5}$$

$$\mathbf{b}_{\mathbf{u}}^{k+1} = \mathbf{b}_{\mathbf{u}}^{k} + \mathbf{b} - A\mathbf{u}^{k+1}, \tag{3.6}$$

$$\mathbf{b}_{\mathbf{d}}^{k+1} = \mathbf{b}_{\mathbf{d}}^{k} + \mathbf{d}^{k+1} - D\mathbf{u}^{k+1}. \tag{3.7}$$

In order to simplify the notation, it is convenient to rewrite (3.1) in terms of a single variable $\mathbf{x}$ as

$$\begin{aligned} \min \quad & K(\mathbf{x}) \equiv F(\mathbf{x}) + \sum_{i=1}^{n+q} \delta_i |x_i| \\ \text{s.t.} \quad & M\,\mathbf{x} = \mathbf{s}, \end{aligned} \tag{3.8}$$

where

$$\mathbf{x} = \begin{bmatrix} \mathbf{u} \\ \mathbf{d} \end{bmatrix}, \quad F(\mathbf{x}) = f(\mathbf{u}), \quad M = \begin{bmatrix} A & 0 \\ D & -I \end{bmatrix}, \quad \mathbf{s} = \begin{bmatrix} \mathbf{b} \\ \mathbf{0} \end{bmatrix}, \tag{3.9}$$

and

$$\delta_i = \begin{cases} \tau_1, & \text{if } i \le n, \\ \tau_2, & \text{if } i > n. \end{cases}$$

We also denote $n_x = n + q$ the size of $\mathbf{x}$ and $n_s = m + q$ the number of rows of M (i.e., the size of $\mathbf{s}$), so that $M \in \mathbb{R}^{n_s \times n_x}$. Then, iteration (3.5)-(3.7) can be written as

$$\mathbf{x}^{k+1} = \operatorname*{argmin}_{\mathbf{x}} K(\mathbf{x}) + \frac{\lambda}{2}\|M\,\mathbf{x} - \mathbf{s}^{k}\|^2, \tag{3.10}$$

$$\mathbf{s}^{k+1} = \mathbf{s}^{k} + \mathbf{s} - M\,\mathbf{x}^{k+1}, \tag{3.11}$$

where $\mathbf{s}^k = \left[\mathbf{b}_{\mathbf{u}}^{k};\, \mathbf{b}_{\mathbf{d}}^{k}\right]$.

We can rewrite (3.10)-(3.11) as the augmented Lagrangian iteration

$$\mathbf{x}^{k+1} = \operatorname*{argmin}_{\mathbf{x}} K(\mathbf{x}) - \left\langle \boldsymbol{\mu}^k,\, M\mathbf{x} \right\rangle + \frac{\lambda}{2}\|M\,\mathbf{x} - \mathbf{s}\|^2, \tag{3.12}$$

$$\boldsymbol{\mu}^{k+1} = \boldsymbol{\mu}^{k} + \lambda\left(\mathbf{s} - M\,\mathbf{x}^{k+1}\right), \tag{3.13}$$

where we set $\boldsymbol{\mu}^k = \lambda(\mathbf{s}^k - \mathbf{s})$ for all k.

The following theorem, which is adapted from [22, Theorem 3], provides a general convergence result for the augmented Lagrangian scheme (3.12)-(3.13) when the minimization in (3.12) is performed inexactly.

THEOREM 3.1. *Let $K(\mathbf{x})$ be a closed convex function, and let $K(\mathbf{x}) + \|M\,\mathbf{x}\|^2$ be strictly convex. Let $\boldsymbol{\mu}^0 \in \mathbb{R}^{n_s}$ and $\mathbf{x}^0 \in \mathbb{R}^{n_x}$ be arbitrary and let $\lambda > 0$. Suppose that*

(i) $\left\| \mathbf{x}^{k+1} - \operatorname*{argmin}_{\mathbf{x}} \left(K(\mathbf{x}) - \left\langle \boldsymbol{\mu}^k,\, M\,\mathbf{x} \right\rangle + \frac{\lambda}{2}\|M\,\mathbf{x} - \mathbf{s}\|^2 \right) \right\| < \nu_k,$

(ii) $\boldsymbol{\mu}^{k+1} = \boldsymbol{\mu}^{k} + \lambda\left(\mathbf{s} - M\,\mathbf{x}^{k+1}\right),$

where $\nu_k \ge 0$ and $\sum_{k=0}^{\infty} \nu_k < +\infty$. If there exists a saddle point $(\widehat{\mathbf{x}}, \widehat{\boldsymbol{\mu}})$ of the Lagrangian function

$$\mathcal{L}(\mathbf{x}, \boldsymbol{\mu}) = K(\mathbf{x}) - \left\langle \boldsymbol{\mu},\, M\,\mathbf{x} - \mathbf{s} \right\rangle,$$

then $\mathbf{x}^k \to \widehat{\mathbf{x}}$ *and* $\boldsymbol{\mu}^k \to \widehat{\boldsymbol{\mu}}$*. If no such saddle point exists, then at least one of the sequences* $\{\mathbf{x}^k\}$ *and* $\{\boldsymbol{\mu}^k\}$ *is unbounded.*

Proof. For each k, let $\bar{\mathbf{x}}^k$ be the unique solution to the minimization problem in (i) (the uniqueness comes from the strict convexity of $K(\mathbf{x}) + \|M\,\mathbf{x}\|^2$). Since $\bar{\mathbf{x}}^k$ is a stationary point, it satisfies the necessary condition

$$\mathbf{0} \in \partial K(\bar{\mathbf{x}}^k) - M^\top \boldsymbol{\mu}^k + \lambda\, M^\top \left(M\,\bar{\mathbf{x}}^k - \mathbf{s}\right). \tag{3.14}$$

By defining $\widetilde{\boldsymbol{\mu}}^k = \boldsymbol{\mu}^k - \lambda\left(M\,\bar{\mathbf{x}}^k - \mathbf{s}\right)$, condition (3.14) can be written as

$$M^\top \widetilde{\boldsymbol{\mu}}^k \in \partial K(\bar{\mathbf{x}}^k)$$

which, by (2.1), is equivalent to

$$\bar{\mathbf{x}}^k \in \partial K^*(M^\top \widetilde{\boldsymbol{\mu}}^k).$$

Therefore,

$$M\,\bar{\mathbf{x}}^k - \mathbf{s} \in \Psi(\widetilde{\boldsymbol{\mu}}^k), \tag{3.15}$$

where $\Psi(\widetilde{\boldsymbol{\mu}}^k) \equiv M\,\partial K^*(M^\top \widetilde{\boldsymbol{\mu}}^k) - \mathbf{s}$. From the definition of $\widetilde{\boldsymbol{\mu}}^k$ and (3.15) it follows that

$$\widetilde{\boldsymbol{\mu}}^k = \boldsymbol{\mu}^k - \lambda\left(M\,\bar{\mathbf{x}}^k - \mathbf{s}\right) \in \boldsymbol{\mu}^k - \lambda\,\Psi(\widetilde{\boldsymbol{\mu}}^k),$$

that is

$$\boldsymbol{\mu}^k \in \widetilde{\boldsymbol{\mu}}^k + \lambda\,\Psi(\widetilde{\boldsymbol{\mu}}^k) = (I + \lambda\,\Psi)\,(\widetilde{\boldsymbol{\mu}}^k). \tag{3.16}$$

Observe that $\Psi(\boldsymbol{\mu}) = \partial\left(K^*(M^\top \boldsymbol{\mu}) - \langle \mathbf{s}, \boldsymbol{\mu}\rangle\right)$, i.e., it is the subdifferential of a closed convex function. From [32, Corollary 31.5.2] we have that Ψ is a maximal monotone operator. Thus, by [22, Corollary 2.2], for any $c > 0$ the operator $J_{c\Psi} \equiv (I + c\,\Psi)^{-1}$ is single valued and has full domain. By (3.16), we have

$$\widetilde{\boldsymbol{\mu}}^k = (I + \lambda\,\Psi)^{-1}\,(\boldsymbol{\mu}^k) = J_{\lambda\Psi}(\boldsymbol{\mu}^k).$$

Thus, by hypothesis (i) we get

$$\left\|\boldsymbol{\mu}^{k+1} - (I + \lambda\,\Psi)^{-1}\,(\boldsymbol{\mu}^k)\right\| = \left\|\boldsymbol{\mu}^{k+1} - \widetilde{\boldsymbol{\mu}}^k\right\| \le \lambda\|M\|\,\left\|\mathbf{x}^{k+1} - \bar{\mathbf{x}}^k\right\| < \lambda\|M\|\nu_k \equiv \beta_k, \tag{3.17}$$

with $\sum_{k=0}^{\infty} \beta_k < +\infty$. By [22, Theorem 3] we have that the sequence $\{\boldsymbol{\mu}^k\}$ satisfies one of the two following conditions:

1) if Ψ has a zero, i.e., there exists a vector $\widehat{\boldsymbol{\mu}}$ such that

$$\Psi(\widehat{\boldsymbol{\mu}}) = M\,\partial K^*(M^\top \widehat{\boldsymbol{\mu}}) - \mathbf{s} = \mathbf{0},$$

then $\boldsymbol{\mu}^k \to \widehat{\boldsymbol{\mu}}$;

2) if Ψ has no zeros, then the sequence is unbounded.

Now we prove that in case 1) the sequence $\{\mathbf{x}^k\}$ converges to a point $\widehat{\mathbf{x}}$. To this aim, we consider the minimization problem in (i). By defining $Z(\mathbf{x}) \equiv K(\mathbf{x}) + \dfrac{\lambda}{2}\|M\,\mathbf{x} - \mathbf{s}\|^2$, which is a strictly convex function by hypothesis, we can write the stationarity condition for $\bar{\mathbf{x}}^k$ as

$$\mathbf{0} \in \partial Z(\bar{\mathbf{x}}^k) - M^\top \boldsymbol{\mu}^k,$$

or equivalently as

$$\bar{\mathbf{x}}^k \in \partial Z^*(M^\top \boldsymbol{\mu}^k).$$

The strict convexity of Z implies that Z^* is a continuously differentiable function and hence

$$\bar{\mathbf{x}}^k = \nabla Z^*(M^\top \boldsymbol{\mu}^k),$$

which implies

$$\bar{\mathbf{x}}^k \to \widehat{\mathbf{x}} \equiv \nabla Z^*(M^\top \widehat{\boldsymbol{\mu}}).$$

This, together with $\left\|\mathbf{x}^{k+1} - \bar{\mathbf{x}}^k\right\| < \nu_k \to 0$, yields $\mathbf{x}^k \to \widehat{\mathbf{x}}$.

Now we show that the pair $(\widehat{\mathbf{x}}, \widehat{\boldsymbol{\mu}})$ is a saddle point of the Lagrangian function $\mathcal{L}(\mathbf{x}, \boldsymbol{\mu})$, i.e., it satisfies

a) $\mathbf{0} \in \partial_{\mathbf{x}} \mathcal{L}(\widehat{\mathbf{x}}, \widehat{\boldsymbol{\mu}}) = \partial K(\widehat{\mathbf{x}}) - M^\top \widehat{\boldsymbol{\mu}}$ or, equivalently, $M^\top \widehat{\boldsymbol{\mu}} \in \partial K(\widehat{\mathbf{x}})$;

b) $\mathbf{0} = \nabla_{\boldsymbol{\mu}} \mathcal{L}(\widehat{\mathbf{x}}, \widehat{\boldsymbol{\mu}}) = M\,\widehat{\mathbf{x}} - \mathbf{s}$.

The proof of b) follows by noting that $M\,\mathbf{x}^{k+1} - \mathbf{s} = \dfrac{1}{\lambda}(\boldsymbol{\mu}^k - \boldsymbol{\mu}^{k+1}) \to \mathbf{0}$. In order to prove a) we observe that $\bar{\mathbf{x}}^k \to \widehat{\mathbf{x}}$ implies $\widetilde{\boldsymbol{\mu}}^k \to \widehat{\boldsymbol{\mu}}$; moreover, $M^\top \widetilde{\boldsymbol{\mu}}^k \in \partial K(\bar{\mathbf{x}}^k)$. The thesis comes from the limit property of maximal monotone operators [7] applied to ∂K. ∎

REMARK 3.2. Because of the equivalence between (3.10)-(3.11) and (3.12)-(3.13), the previous theorem implies that if $\boldsymbol{\mu}^k \to \widehat{\boldsymbol{\mu}}$, then the sequence $\{\mathbf{s}^k\}$ generated in (3.10)-(3.11) converges to $\widehat{\mathbf{s}} = \frac{1}{\lambda}\widehat{\boldsymbol{\mu}} + \mathbf{s}$.

4. Subspace acceleration for the split Bregman subproblems. Let us introduce, for each $\mathbf{x} \in \mathbb{R}^{n_x}$, the sets

$$\mathcal{A}_+(\mathbf{x}) = \{i : x_i > 0\}, \qquad \mathcal{A}_-(\mathbf{x}) = \{i : x_i < 0\},$$
$$\mathcal{A}_0(\mathbf{x}) = \{i : x_i = 0\}, \qquad \mathcal{A}_\pm(\mathbf{x}) = \mathcal{A}_+(\mathbf{x}) \cup \mathcal{A}_-(\mathbf{x}).$$

This partitioning of the variables has been used in [12, 34] to extend some ideas developed in the context of active-set methods for bound-constrained optimization [20, 21, 25] to the case of ℓ_1-regularized optimization. In the case of bound-constrained quadratic problems, suitable measures of optimality with respect to the *active* variables (i.e., the variables that are on their bounds) and the *free* variables (i.e., the variables that are not active) are used to establish whether the set of active variable is "promising". If this is the case, then a restricted version of the problem, obtained by fixing the active variables to their values, is solved with high accuracy. This results in very efficient algorithms in practice, able to outperform standard gradient projection schemes [18, 29]. The extension of this strategy to the case of ℓ_1-regularized optimization comes from the observation that zero and nonzero variables can play the role of active and free variables, respectively.

The results contained in this Section require a further assumption on the function $f(\mathbf{u})$ in (1.1).

ASSUMPTION 4.1. *The gradient of f is Lipschitz continuous with constant L over $\mathbb{R}^n$, i.e., for all $\mathbf{u}_1, \mathbf{u}_2 \in \mathbb{R}^n$*

$$\|\nabla f(\mathbf{u}_1) - \nabla f(\mathbf{u}_2)\| \le L\,\|\mathbf{u}_1 - \mathbf{u}_2\|.$$

Note that $F(\mathbf{x})$ defined in (3.9) has Lipschitz continuous gradient with the same constant L.

In order to ease the description of our acceleration strategy, we reformulate the minimization problem in (3.10) as follows:

$$\mathbf{x}^{k+1} = \operatorname*{argmin}_{\mathbf{x}} H^k(\mathbf{x}) \equiv G^k(\mathbf{x}) + \sum_{i=1}^{n_x} \delta_i |x_i|, \tag{4.1}$$

where

$$G^k(\mathbf{x}) = F(\mathbf{x}) + \frac{\lambda}{2}\|M\,\mathbf{x} - \mathbf{s}^k\|^2.$$

In this way we separate the smooth part of the objective function from the ℓ_1 regularization term. Recall that a point $\mathbf{x} \in \mathbb{R}^{n_x}$ is a solution to (4.1) if and only if it satisfies the stationarity condition $\mathbf{0} \in \partial H^k(\mathbf{x})$, i.e.,

$$\begin{cases} \nabla_i G^k(\mathbf{x}) + \delta_i = 0, & \text{if } i \in \mathcal{A}_+(\mathbf{x}), \\ \nabla_i G^k(\mathbf{x}) - \delta_i = 0, & \text{if } i \in \mathcal{A}_-(\mathbf{x}), \\ \left|\nabla_i G^k(\mathbf{x})\right| \le \delta_i, & \text{otherwise.} \end{cases} \tag{4.2}$$

Consider the pair $(\widehat{\mathbf{x}}, \widehat{\mathbf{s}})$ defined in Theorem 3.1 and in Remark 3.2. Let us define the scalars

$$\theta_1 = \frac{1}{2} \min_{i \in \mathcal{A}_\pm(\widehat{\mathbf{x}})} |\widehat{x}_i| \quad \text{and} \quad \theta_2 = \frac{1}{2} \min_{i \in \mathcal{A}_0(\widehat{\mathbf{x}})} \left(\delta_i - \left|\nabla_i \widehat{G}(\widehat{\mathbf{x}})\right|\right),$$

where

$$\widehat{G}(\mathbf{x}) = F(\mathbf{x}) + \frac{\lambda}{2}\|M\,\mathbf{x} - \widehat{\mathbf{s}}\|^2.$$

We make the following assumptions, which imply that $\theta_1, \theta_2 > 0$.

ASSUMPTION 4.2. *The solution $\widehat{\mathbf{x}}$ to problem (3.8) satisfies $\widehat{\mathbf{x}} \neq \mathbf{0}$.*

ASSUMPTION 4.3. *The solution $(\widehat{\mathbf{x}}, \widehat{\mathbf{s}})$ to problem (3.8) is nondegenerate, i.e.*

$$\min_{i \in \mathcal{A}_0(\widehat{\mathbf{x}})} \left(\delta_i - \left|\nabla_i \widehat{G}(\widehat{\mathbf{x}})\right|\right) > 0.$$

From Assumption 4.1 and the definition of $\widehat{G}(\mathbf{x})$ we have that $\nabla\widehat{G}(\mathbf{x})$ is Lipschitz continuous. Indeed, a Lipschitz constant for $\nabla\widehat{G}(\mathbf{x})$ is

$$\widehat{L} = L + \lambda\,\|M\|^2.$$

Since, for any $\mathbf{x} \in \mathbb{R}^{n_x}$ and $k \in \mathbb{N}$,

$$\nabla G^k(\mathbf{x}) = \nabla F(\mathbf{x}) + \lambda\, M^\top (M\,\mathbf{x} - \mathbf{s}^k) \quad \text{and} \quad \nabla\widehat{G}(\mathbf{x}) = \nabla F(\mathbf{x}) + \lambda\, M^\top (M\,\mathbf{x} - \widehat{\mathbf{s}}),$$

we have that for any $\mathbf{y}, \mathbf{z} \in \mathbb{R}^{n_x}$

$$\left\|\nabla G^k(\mathbf{y}) - \nabla G^k(\mathbf{z})\right\| = \left\|\nabla\widehat{G}(\mathbf{y}) - \nabla\widehat{G}(\mathbf{z})\right\| \le \widehat{L}\|\mathbf{y} - \mathbf{z}\|.$$

i.e., $\widehat{L}$ is also a Lipschitz constant for $\nabla G^k(\mathbf{x})$.
The following lemma shows that when $\mathbf{x}^k$ is sufficiently close to $\widehat{\mathbf{x}}$, then some entries of $\mathbf{x}^k$ and $\widehat{\mathbf{x}}$ have the same sign (see [13, Lemma 3.1]).

LEMMA 4.4. *If $\left\|\mathbf{x}^k - \widehat{\mathbf{x}}\right\| \le \dfrac{\theta_1}{2}$ then*

$$\operatorname{sign}(x_i^k) = \operatorname{sign}(\widehat{x}_i), \qquad \forall i \in \mathcal{A}_\pm(\widehat{\mathbf{x}}) \cup \left(\mathcal{A}_0(\widehat{\mathbf{x}}) \cap \mathcal{A}_0(\mathbf{x}^k)\right).$$

We recall that $\mathbb{R}^{n_x}$ can be splitted into 2^{n_x} orthants, and introduce the following definition.

DEFINITION 4.5. *Given any n_x-ple $\sigma \in \{-1,1\}^{n_x}$, the orthant associated with σ is defined as*

$$\Omega_\sigma = \{\mathbf{x} \in \mathbb{R}^{n_x} : \ (x_i \geq 0 \text{ if } \sigma_i = 1) \ \wedge \ (x_i \leq 0 \text{ if } \sigma_i = -1)\}.$$

REMARK 4.6. Lemma 4.4 suggests that when the current iterate $\mathbf{x}^k$ is close to the solution $\widehat{\mathbf{x}}$, the nonzero entries of $\mathbf{x}^k$ have the same sign as the corresponding entries of the solution $\widehat{\mathbf{x}}$, i.e., $\mathbf{x}^k$ and $\widehat{\mathbf{x}}$ lie in the same orthant of $\mathbb{R}^{n_x}$. Therefore one could think of restricting the current subproblem (4.1) to the orthant containing $\mathbf{x}^k$. The restriction of $H^k(\mathbf{x})$ to an orthant Ω_σ has the form

$$H^k_{|\Omega_\sigma}(\mathbf{x}) = G^k_{|\Omega_\sigma}(\mathbf{x}) + \langle \boldsymbol{\nu}_\sigma, \mathbf{x} \rangle,$$

where we set for all i

$$[\boldsymbol{\nu}_\sigma]_i = \begin{cases} \delta_i, & \text{if } \sigma_i = \ \ 1, \\ -\delta_i, & \text{if } \sigma_i = -1. \end{cases}$$

Since $H^k_{|\Omega_\sigma}(\mathbf{x})$ is a smooth function, if we knew that the current orthant contained the solution to (4.1), then we could choose to solve the subproblem with high accuracy by using techniques suited for smooth bound-constrained optimization problems. Similar ideas have been exploited in the solution of unconstrained ℓ_1-regularized nonlinear problems, giving rise to the family of the so-called "orthant-based algorithms" [9, 28].

We aim at introducing subspace-acceleration steps into the Bregman framework. This means that, at suitable Bregman iterations, we want to replace the minimization of H^k with the minimization of its restriction to the orthant face determined by $\mathcal{A}_0(\mathbf{x}^k)$, i.e., the set

$$\left\{\mathbf{y} \in \mathbb{R}^{n_x} \ : \ \left(y_i = 0, \ i \in \mathcal{A}_0(\mathbf{x}^k)\right) \ \wedge \ \left(\operatorname{sign}(y_i) = \operatorname{sign}(x_i^k), \ i \in \mathcal{A}_\pm(\mathbf{x}^k)\right)\right\}. \tag{4.3}$$

When $\mathcal{A}_0(\mathbf{x}^k)$ is large, this could result in a significant reduction of the computational cost of determining the next iterate.

Recall that the optimality of a given point $\mathbf{x}$ with respect to problem (4.1) can be measured in terms of the minimum norm subgradient of H^k at a given point $\mathbf{x}$, i.e., the vector $\mathbf{g}^k(\mathbf{x})$ defined componentwise as

$$[\mathbf{g}^k(\mathbf{x})]_i = \begin{cases} \nabla_i G^k(\mathbf{x}) + \delta_i, & \text{if } i \in \mathcal{A}_+(\mathbf{x}) \text{ or } (i \in \mathcal{A}_0(\mathbf{x}) \text{ and } \nabla_i G^k(\mathbf{x}) + \delta_i < 0), \\ \nabla_i G^k(\mathbf{x}) - \delta_i, & \text{if } i \in \mathcal{A}_-(\mathbf{x}) \text{ or } (i \in \mathcal{A}_0(\mathbf{x}) \text{ and } \nabla_i G^k(\mathbf{x}) - \delta_i > 0), \\ 0, & \text{otherwise.} \end{cases}$$

By following [12, 34], we split $\mathbf{g}^k(\mathbf{x})$ into the vectors $\boldsymbol{\beta}^k(\mathbf{x})$ and $\boldsymbol{\varphi}^k(\mathbf{x})$, which measure the optimality of $\mathbf{x}$ with respect to the zero and nonzero variables respectively. The two vectors are defined componentwise as

$$\begin{aligned} \left[\boldsymbol{\beta}^k(\mathbf{x})\right]_i &= \begin{cases} \nabla_i G^k(\mathbf{x}) + \delta_i, & \text{if } i \in \mathcal{A}_0(\mathbf{x}) \text{ and } \nabla_i G^k(\mathbf{x}) + \delta_i < 0, \\ \nabla_i G^k(\mathbf{x}) - \delta_i, & \text{if } i \in \mathcal{A}_0(\mathbf{x}) \text{ and } \nabla_i G^k(\mathbf{x}) - \delta_i > 0, \\ 0, & \text{otherwise,} \end{cases} \\ \left[\boldsymbol{\varphi}^k(\mathbf{x})\right]_i &= \begin{cases} 0, & \text{if } i \in \mathcal{A}_0(\mathbf{x}), \\ \min\{\nabla_i G^k(\mathbf{x}) + \delta_i, \max\{x_i, \nabla_i G^k(\mathbf{x}) - \delta_i\}\}, & \text{if } i \in \mathcal{A}_+(\mathbf{x}), \\ \max\{\nabla_i G^k(\mathbf{x}) - \delta_i, \min\{x_i, \nabla_i G^k(\mathbf{x}) + \delta_i\}\}, & \text{if } i \in \mathcal{A}_-(\mathbf{x}). \end{cases} \end{aligned} \tag{4.4}$$

It is straightforward to check that if $\boldsymbol{\beta}^k(\bar{\mathbf{x}}) = \mathbf{0}$ and $\boldsymbol{\varphi}^k(\bar{\mathbf{x}}) = \mathbf{0}$ at any point $\bar{\mathbf{x}} \in \mathbb{R}^{n_x}$, then $\bar{\mathbf{x}}$ is a stationary point for $H^k(\mathbf{x})$. It is worth noting that the vector $\boldsymbol{\varphi}^k(\mathbf{x})$ also takes into account how much nonzero variables can move before becoming zero, i.e., before $\mathbf{x}$ enters another orthant [12].

Now we can prove a bound on the components of $\nabla G^k(\mathbf{x}^k)$ corresponding to indices in $\mathcal{A}_0(\widehat{\mathbf{x}})$ when $(\mathbf{x}^k, \mathbf{s}^k)$ is "sufficiently close" to $(\widehat{\mathbf{x}}, \widehat{\mathbf{s}})$. The result extends to the case of Bregman iterations for problem (3.8) a similar result proved in [13] for the solution of ℓ_1-regularized unconstrained minimization problems.

THEOREM 4.7. *If* $\left\|\mathbf{x}^k - \widehat{\mathbf{x}}\right\| \leq \min\left\{\dfrac{\theta_1}{2}, \dfrac{\theta_2}{2\widehat{L}}\right\}$ *and* $\left\|\mathbf{s}^k - \widehat{\mathbf{s}}\right\| \leq \dfrac{\theta_2}{2\lambda\|M\|}$, *then*

i) $\left|\nabla_i G^k(\mathbf{x}^k)\right| \leq \delta_i - \theta_2, \qquad \forall i \in \mathcal{A}_0(\widehat{\mathbf{x}})$,

ii) $\boldsymbol{\beta}^k(\mathbf{x}^k) = \mathbf{0}$.

Proof. In order to prove i), let us consider an index k satisfying the hypotheses. For all i, we can write

$$\left|\left|\nabla_i G^k(\mathbf{x}^k)\right| - \left|\nabla_i \widehat{G}(\widehat{\mathbf{x}})\right|\right| \leq \left|\nabla_i G^k(\mathbf{x}^k) - \nabla_i \widehat{G}(\widehat{\mathbf{x}})\right| = \tag{4.5}$$

$$= \left|\nabla_i \widehat{G}(\mathbf{x}^k) + [\lambda M^\top(\widehat{\mathbf{s}} - \mathbf{s}^k)]_i - \nabla_i \widehat{G}(\widehat{\mathbf{x}})\right| \leq \tag{4.6}$$

$$\leq \left\|\nabla \widehat{G}(\mathbf{x}^k) + \lambda M^\top(\widehat{\mathbf{s}} - \mathbf{s}^k) - \nabla \widehat{G}(\widehat{\mathbf{x}})\right\| \leq \tag{4.7}$$

$$\leq \widehat{L}\,\|\mathbf{x}^k - \widehat{\mathbf{x}}\| + \lambda\,\|M\|\,\|\mathbf{s}^k - \widehat{\mathbf{s}}\| \leq \frac{\theta_2}{2} + \frac{\theta_2}{2} = \theta_2. \tag{4.8}$$

This implies that

$$\left|\nabla_i G^k(\mathbf{x}^k)\right| \leq \left|\nabla_i \widehat{G}(\widehat{\mathbf{x}})\right| + \theta_2 \tag{4.9}$$

for all i. Recall that $\delta_i = \tau_1$ for $i \leq n$ and $\delta_i = \tau_2$ otherwise. Without loss of generality we analyze the case $i \leq n$; the case $i > n$ can be proved in the same way. By defining

$$c_1 = \max_{l \in \mathcal{A}_0(\widehat{\mathbf{x}}) \cap \{1,\ldots,n\}} \left|\nabla_l \widehat{G}(\widehat{\mathbf{x}})\right|,$$

we have that

$$\theta_2 \leq (\tau_1 - c_1)/2.$$

Let $i \in \mathcal{A}_0(\widehat{\mathbf{x}}) \cap \{1, \ldots, n\}$. From (4.9) and the previous inequality we get

$$\left|\nabla_i G^k(\mathbf{x}^k)\right| \leq \left|\nabla_i \widehat{G}(\widehat{\mathbf{x}})\right| + \theta_2 \leq c_1 + \frac{\tau_1 - c_1}{2} =$$

$$= \tau_1 - \frac{\tau_1 - c_1}{2} \leq \tau_1 - \theta_2 = \delta_i - \theta_2.$$

This completes the proof of i).

To prove ii), we observe that $\beta_i^k(\mathbf{x}^k)$ can be nonzero only for $i \in \mathcal{A}_0(\mathbf{x}^k)$ and that $\mathcal{A}_0(\mathbf{x}^k)$ can be written as

$$\mathcal{A}_0(\mathbf{x}^k) = \left(\mathcal{A}_0(\mathbf{x}^k) \cap \mathcal{A}_0(\widehat{\mathbf{x}})\right) \cup \left(\mathcal{A}_0(\mathbf{x}^k) \cap \mathcal{A}_\pm(\widehat{\mathbf{x}})\right).$$

From Lemma 4.4 it follows that $\mathcal{A}_0(\mathbf{x}^k) \cap \mathcal{A}_\pm(\widehat{\mathbf{x}}) = \emptyset$. For $i \in \mathcal{A}_0(\mathbf{x}^k) \cap \mathcal{A}_0(\widehat{\mathbf{x}})$ we have

$$\left|\nabla_i G^k(\mathbf{x}^k)\right| \leq \delta_i - \theta_2 < \delta_i,$$

which concludes the proof. □

The previous theorem suggests that when $(\mathbf{x}^k, \mathbf{s}^k)$ is in a neighborhood of the solution $(\widehat{\mathbf{x}}, \widehat{\mathbf{s}})$, the only variables that violate the optimality conditions are the nonzero ones.

By Remark 4.6, the orthant containing the solution is identified as the iterates converge to the solution. Therefore, one could think of introducing into the general inexact Bregman framework (3.10)-(3.11) an automatic criterion to decide whether the solution to (3.10) can be searched in the current orthant face by means of a more efficient algorithm. Inspired by similar conditions introduced in the framework of bound-constrained quadratic problems [18, 20, 25, 29], we propose to perform subspace-acceleration steps whenever

$$\left\| \boldsymbol{\beta}^k(\mathbf{x}^k) \right\| \leq \gamma \, \left\| \boldsymbol{\varphi}^k(\mathbf{x}^k) \right\|, \tag{4.10}$$

where $\gamma > 0$ is a suitable constant. The idea is based on the observation that when the optimality violation with respect to the zero variables is smaller than the violation with respect to the nonzero ones, restricting the minimization to the latter could be more beneficial.

Moreover, since one could expect that for $(\mathbf{x}^k, \mathbf{s}^k)$ "sufficiently close" to $(\widehat{\mathbf{x}}, \widehat{\mathbf{s}})$ the minimizer of problem (4.1) lies in the same orthant face as $\mathbf{x}^k$, it is possible to replace the minimization of $H^k(\mathbf{x})$ over the orthant face containing $\mathbf{x}^k$ with the minimization over the affine closure of the orthant face, i.e.,

$$\mathcal{F}^k = \left\{ \mathbf{y} \in \mathbb{R}^{n_x} : \; y_i = 0, \; i \in \mathcal{A}_0(\mathbf{x}^k) \right\}. \tag{4.11}$$

This results in replacing the nonsmooth unconstrained minimization problem (4.1) with the smooth optimization problem

$$\mathbf{z}^{k+1} = \operatorname*{argmin}_{x_i = 0, \, i \in \mathcal{A}_0^k} H^k_{|\mathcal{F}^k}(\mathbf{x}) \equiv G^k_{|\mathcal{F}^k}(\mathbf{x}) + \sum_{i \in \mathcal{A}_\pm^k} \nu_i^k \, x_i, \tag{4.12}$$

where we set $\mathcal{A}_\pm^k \equiv \mathcal{A}_\pm(\mathbf{x}^k)$, $\mathcal{A}_0^k \equiv \mathcal{A}_0(\mathbf{x}^k)$ and for all $i \in \mathcal{A}_\pm^k$

$$\nu_i^k = \begin{cases} \delta_i, & \text{if } \operatorname{sign}(\mathbf{x}^k) = +1, \\ -\delta_i, & \text{if } \operatorname{sign}(\mathbf{x}^k) = -1. \end{cases}$$

It is worth noting that, by fixing the zero variables, problem (4.12) can be equivalently rewritten as an unconstrained minimization over $\mathbb{R}^{|\mathcal{A}_\pm^k|}$. Therefore, efficient algorithms for unconstrained smooth optimization can be exploited for its solution. Since criterion (4.10) does not guarantee that $\mathbf{z}^{k+1}$ lies in the same orthant as $\mathbf{x}^k$, we select the iterate $\mathbf{x}^{k+1}$ by a projected backtracking line search ensuring a sufficient decrease in H^k, i.e.,

$$H^k(\mathbf{x}^{k+1}) - H^k(\mathbf{x}^k) \leq \eta \langle \nabla H^k(\mathbf{x}^k), \, \mathbf{x}^{k+1} - \mathbf{x}^k \rangle, \tag{4.13}$$

where η is a small positive constant. Note that the orthogonal projection $\operatorname{proj}(\mathbf{z}; \mathbf{x})$ of a point $\mathbf{z}$ onto the orthant face containing $\mathbf{x}$ can be easily computed componentwise as

$$[\operatorname{proj}(\mathbf{z}; \mathbf{x})]_i = \begin{cases} \max\{0, z_i\}, & \text{if } i \in \mathcal{A}_+(\mathbf{x}), \\ \min\{0, z_i\}, & \text{if } i \in \mathcal{A}_-(\mathbf{x}), \\ 0, & \text{if } i \in \mathcal{A}_0(\mathbf{x}). \end{cases}$$

The resulting method, which we call *Split Bregman with Subspace Acceleration* (SBSA) is outlined in Algorithm 1.

Algorithm 1 Split Bregman with Subspace Acceleration (SBSA)

1: Choose $\mathbf{x}^0 = \mathbf{0} \in \mathbb{R}^{n_x}$, $\mathbf{s}^0 = \mathbf{0} \in \mathbb{R}^{n_s}$, $\lambda > 0$, $\gamma > 0$;
2: $\mathbf{x}^1 \approx \operatorname{argmin}_{\mathbf{x}} H^0(\mathbf{x})$;
3: **for** $k = 1, 2, \ldots$ **do**
4: $\quad \mathbf{s}^k = \mathbf{s}^{k-1} + \mathbf{s} - M\,\mathbf{x}^k$;
5: $\quad$ **if** $\|\boldsymbol{\beta}^k(\mathbf{x}^k)\| \leq \gamma \|\boldsymbol{\varphi}^k(\mathbf{x}^k)\|$ **then**
6: $\quad\quad \mathbf{z}^{k+1} \approx \operatorname{argmin} \left\{ H^k_{|\mathcal{F}^k}(\mathbf{x}) \ : \ x_i = 0,\ i \in \mathcal{A}^k_0 \right\}$;
7: $\quad\quad \mathbf{x}^{k+1} = \operatorname{proj}\left(\mathbf{x}^k + \alpha^k(\mathbf{z}^{k+1} - \mathbf{x}^k);\ \mathbf{x}^k\right)$ with α^k obtained by backtracking line search;
8: $\quad\quad$ **if** $\mathbf{x}^{k+1}$ not sufficiently accurate **then** ▷ SAFEGUARD
9: $\quad\quad\quad \mathbf{x}^{k+1} \approx \operatorname{argmin}_{\mathbf{x}} H^k(\mathbf{x})$;
10: $\quad\quad$ **end if**
11: $\quad$ **else**
12: $\quad\quad \mathbf{x}^{k+1} \approx \operatorname{argmin}_{\mathbf{x}} H^k(\mathbf{x})$;
13: $\quad$ **end if**
14: **end for**

The following theorem, which is an adaptation of [28, Theorem A.3], shows that the line search procedure at step 7 of Algorithm 1 is well defined.

THEOREM 4.8. *The backtracking projected line search in the acceleration phases of SBSA terminates in a finite number of iterations.*

Proof. Consider the k-th iteration of algorithm SBSA and suppose an acceleration step is taken. Let $\mathbf{z}^{k+1}$ be the point computed at line 6 of Algorithm 1 and $\mathbf{d}^k = \mathbf{z}^{k+1} - \mathbf{x}^k$. By construction we have that $\mathbf{d}^k_i = 0$ for all $i \in \mathcal{A}_0(\mathbf{x}^k)$. By following the first part of the proof of [28, Theorem A.3], it is easy to show that there exists $\bar{\alpha} > 0$ such that $\operatorname{proj}(\mathbf{x}^k + \alpha\mathbf{d}^k; \mathbf{x}^k) = \mathbf{x}^k + \alpha\mathbf{d}^k$ for all $\alpha \in (0, \bar{\alpha}]$, i.e., $\mathbf{x}^k + \alpha\mathbf{d}^k$ lies in the same orthant face as $\mathbf{x}^k$. Since $\mathbf{z}^{k+1}$ is an approximate minimizer of $H^k_{\mathcal{F}^k}$ and $H^k_{\mathcal{F}^k}$ is convex, $\mathbf{d}^k$ is a local descent direction for $H^k_{\mathcal{F}^k}$ in $\mathbf{x}^k$. This ensures that in a finite number of steps the backtracking procedure can find a value of α guaranteeing sufficient decrease for $H^k_{\mathcal{F}^k}$. By observing that $H^k(\mathbf{x}) = H^k_{\mathcal{F}^k}(\mathbf{x})$ for each $\mathbf{x}$ lying in the same orthant face as $\mathbf{x}^k$, we conclude that the value of α obtained with backtracking guarantees sufficient decrease of $H^k(\mathbf{x})$. ∎

According to Theorem 3.1, the convergence of the inexact scheme is only guaranteed when the solution of the subproblem in (4.1) is sufficiently accurate. For this reason, a safeguard has been considered at lines 8-10 of Algorithm 1. This could be inefficient in practice, because the output of the subspace acceleration is likely to be rejected when the iterate is far from the solution. In our implementation of Algorithm 1 we use a heuristic criterion to decide whether to accept the iterate generated by the subspace-acceleration step (see Section 5.2). We recall that for the exact Bregman scheme applied to problem (3.8) it can be proved that (see [31, Proposition 3.2])

$$\|M\mathbf{x}^{k+1} - \mathbf{s}\| \leq \|M\mathbf{x}^k - \mathbf{s}\| \tag{4.14}$$

for all k. Based on this observation, we decided to accept the iterate produced by lines 6-7 of Algorithm 1 if (4.14) is satisfied. Numerical experiments showed the effectiveness of this choice.

5. Application: multi-period portfolio selection. Portfolio selection is central to financial economics and is the building block of the capital asset pricing model. It aims to find an optimal allocation of capital among a set of assets by rational financial targets. For medium- and long-time horizons, a multi-period investment policy is considered: the investors can change the allocation of the wealth among the assets over time by the end of the investment,

taking into account the time evolution of available information. In a multi-period setting the investment period is partitioned into sub-periods, delimited by the rebalancing dates at which decisions are taken. More precisely, if m is the number of sub-periods and $t_j = 1, \ldots, m+1$ denote the rebalancing dates, then a decision taken at time t_j is kept in the j-th sub-period $[t_j, t_{j+1})$ of the investment. The optimal portfolio is defined by the vector

$$\mathbf{u} = [\mathbf{u}_1;\, \mathbf{u}_2;\, \ldots;\, \mathbf{u}_m],$$

where $\mathbf{u}_j \in \mathbb{R}^{n_a}$ is the portfolio of holdings at the beginning of period j and n_a is the number of assets.

In a multi-period mean variance Markowitz framework, the optimal portfolio is obtained by simultaneously minimizing the risk and maximizing the return of the investment. A common strategy to estimate the parameters of the Markowitz model is to use historical data as predictive of the future behavior of asset returns. This typically leads to ill-conditioned numerical problems. Different regularization techniques have been suggested with the aim of improving the problem conditioning. In the last years, ℓ_1 regularization techniques have been considered to obtain sparse solutions in both the single and multi-period cases, with the aim of reducing costs [8, 15, 17]. Another useful interpretation of the ℓ_1 regularization is related to the amount of shorting in the portfolio. From the financial point of view, negative solutions correspond to short sales, which are generally transactions in which an investor sells borrowed securities in anticipation of a price decline. A suitable tuning of the regularization parameter permits short controlling in the solution [8].

We focus on the fused lasso portfolio selection model [16], where an additional ℓ_1 penalty term on the variation is added to the classical ℓ_1 model in order to reduce the transaction costs. Indeed, in the multi-period case, the sparsity of the solution reduces the holding costs, but it does not guarantee low transaction costs if the pattern of nonzeros positions completely changes across periods. The fused lasso term shrinks toward zero the differences of values of the wealth allocated across the assets between two contiguous rebalancing dates, thus encouraging smooth solutions that reduce transactions.

Let $\mathbf{r}_j \in \mathbb{R}^{n_a}$ and $C_j \in \mathbb{R}^{n_a \times n_a}$ contain respectively the expected return vector and the covariance matrix, assumed to be positive definite, estimated at time t_j, $j = 1, \ldots, m$. The fused lasso portfolio selection model reads:

$$\begin{aligned} \min \quad & \sum_{j=1}^{m} \langle \mathbf{u}_j,\, C_j \mathbf{u}_j \rangle + \tau_1 \|\mathbf{u}\|_1 + \tau_2 \sum_{j=1}^{m-1} \|\mathbf{u}_{j+1} - \mathbf{u}_j\|_1 \\ \text{s.t.} \quad & \langle \mathbf{u}_1,\, \mathbf{1}_{n_a} \rangle = \xi_{ini}, \\ & \langle \mathbf{u}_j,\, \mathbf{1}_{n_a} \rangle = \langle \mathbf{1}_{n_a} + \mathbf{r}_{j-1},\, \mathbf{u}_{j-1} \rangle, \quad j = 2, \ldots, m, \\ & \langle \mathbf{1}_{n_a} + \mathbf{r}_m,\, \mathbf{u}_m \rangle = \xi_{fin}, \end{aligned} \tag{5.1}$$

where $\tau_1,\ \tau_2 > 0$, ξ_{ini} is the initial wealth, ξ_{fin} is the target expected wealth resulting from the overall investment. The first constraint is the budget constraint. The strategy is assumed to be self-financing, as required by constraints from 2 to m, where it is established that at the end of each period the wealth is given by the revaluation of the previous one. The $(m+1)$-st constraint defines the expected final wealth. Problem (5.1) can be equivalently formulated as

$$\begin{aligned} \min \quad & \langle \mathbf{u},\, C\mathbf{u} \rangle + \tau_1 ||\mathbf{u}||_1 + \tau_2 \|D\,\mathbf{u}\|_1 \\ \text{s.t.} \quad & A\mathbf{u} = \mathbf{b}, \end{aligned} \tag{5.2}$$

where $C \in \mathbb{R}^{n \times n}$, with $n = m \cdot n_a$, is the symmetric positive definite block-diagonal matrix

$$C = \text{diag}(C_1, C_2, \ldots, C_m),$$

$D \in \mathbb{R}^{(n-n_a)\times n}$ is the discrete difference matrix defined by

$$d_{ij} = \begin{cases} -1, & \text{if } j = i, \\ 1, & \text{if } j = i + n_a, \\ 0, & \text{otherwise}, \end{cases}$$

$A \in R^{(m+1)\times n}$ can be regarded as a $(m+1) \times m$ lower block-bidiagonal matrix, with blocks of dimension $1 \times n_a$, defined by

$$a_{ij} = \begin{cases} \mathbf{1}_{n_a}^\top, & i = j, \\ -(\mathbf{1}_{n_a} + \mathbf{r}_{i-1})^\top, & j = i + 1, \\ \mathbf{0}_{n_a}^\top, & \text{otherwise}, \end{cases}$$

and $\mathbf{b} = (\xi_{ini}, 0, 0, ..., \xi_{fin})^\top \in \mathbb{R}^{m+1}$.

5.1. Testing environment. The SBSA algorithm has been tested on three real datasets. Two of them use a universe of investments compiled by Fama and French*. Specifically, the FF48 dataset contains monthly returns of 48 portfolios representing different industrial sectors, and the FF100 dataset includes monthly returns of 100 portfolios on the basis of size and book-to-market ratio. Both datasets consist of data ranging from July 1926 to December 2015. We consider a preprocessing procedure that eliminates the elements with the highest volatilities, so that the number of portfolios in FF100 is reduced to 96. In our experiments we use data during periods of 10, 20 and 30 years with annual rebalancing, i.e., we consider the periods July 2005 - June 2015, July 1995 - June 2015, and July 1985 - June 2015. The corresponding test problems are called FF48-10y, FF48-20y and FF48-30y, respectively. The third dataset, denoted ES50, contains the daily returns of stocks included in the EURO STOXX 50 Index Europe's leading blue-chip index for the Eurozone. The index covers the 50 largest companies among the 19 supersectors in terms of free-float market cap in 11 Eurozone countries. The dataset contains daily returns for each stock in the index from January 2008 to December 2013. For this test case we consider both annual ($m = 6$ years) and quarterly ($m = 22$ quarters) rebalancing. The corresponding test problems are referred to as ES50-A and ES50-Q, respectively.

Following [16], a rolling window for setting up the model parameters is considered. For each dataset, the length of the rolling windows is fixed in order to build positive definite covariance matrices and ensure statistical significance. Different datasets require different lengths for the rolling windows. FF100 requires ten-year data; for FF48 five years are sufficient; one-year data are used for ES50.

Our portfolio is compared with the benchmark one that is based on the strategy where the total amount is equally divided among the assets at each rebalancing date. The portfolio built following this strategy is referred to as the multi-period *naive* portfolio and it is commonly used as a benchmark by investors because it is a simple rule that reduces risk enough to make a profit. We assume that the investor has one unit of wealth at the beginning of the planning horizon, i.e., $\xi_{ini} = 1$. In order to compare the optimal portfolio with the naive one, we set the expected final wealth equal to that of the naive one, i.e., $\xi_{fin} = \xi_{naive}$, where

$$\xi_{naive} = \frac{1}{n_a}\left(\dots\left(\frac{1}{n_a}\left(\frac{\xi_{ini}}{n_a}\langle \mathbf{1}_{n_a} + \mathbf{r}_1,\, \mathbf{1}_{n_a}\rangle\right)\langle \mathbf{1}_{n_a} + \mathbf{r}_2,\, \mathbf{1}_{n_a}\rangle\right)\dots\right)\langle \mathbf{1}_{n_a} + \mathbf{r}_m,\, \mathbf{1}_{n_a}\rangle.$$

Following [16], we consider some performance metrics that take into account the risk and the cost of the portfolio. The next metric measures the risk reduction factor of the optimal

*Data available from http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html#BookEquity

strategy with respect to the benchmark one:

$$\text{ratio} = \frac{\langle \mathbf{u}_{naive},\, C\,\mathbf{u}_{naive}\rangle}{\langle \mathbf{u}_{opt},\, C\,\mathbf{u}_{opt}\rangle} \tag{5.3}$$

where $\mathbf{u}_{naive}$ and $\mathbf{u}_{opt}$ are the naive portfolio and the optimal one, respectively.

Another metric gives the percentage of active positions in portfolio, which is an estimate of the holding costs:

$$\text{density} = \frac{card\left(\{|[\mathbf{u}_j]_i| \geq \epsilon_1, i = 1, ..., n_a, j = 1, ..., m\}\right) \cdot 100}{n}\ \%, \tag{5.4}$$

where $card(S)$ denotes the cardinality of the set S. The threshold ϵ_1 is aimed at avoiding too small wealth allocations, since they make no sense in financial terms. We note that the density of the naive portfolio is $\text{density}_{naive} = 100\%$, so we have holding costs in each period for all assets. Finally, we use a metric giving information about the total number of variations in weights across periods, which are a measure of the transaction costs:

$$\mathcal{T} = trace(V^\top V) \tag{5.5}$$

where $V \in \mathbb{R}^{n_a \times (m-1)}$ with

$$v_{ij} = \begin{cases} 1, & \text{if } |[\mathbf{u}_j]_i - [\mathbf{u}_{j+1}]_i| \geq \epsilon_2, \\ 0, & \text{otherwise.} \end{cases} \tag{5.6}$$

Note that (5.5) is a pessimistic estimate of the transaction costs because weights could also change for effect of revaluation. In order to provide more detailed information about the investment, it is convenient to refer also to $||V||_1$, which is the maximum number of variations over the periods, and to $||V||_\infty$, which is the maximum number of variations over the assets.

The choice of the regularization parameters τ_1 and τ_2 in (5.1) plays a key role in obtaining solutions that meet the financial requirements. Starting from the numerical results in [16], we selected parameters in $\{10^{-4}, 10^{-3}, 10^{-2}\}$, guaranteeing a good tradeoff between the performance metrics and the number of short positions for FF48-20y, FF100-20y and ES50-Q problems. More in detail, we first set τ_1 as the smallest value producing at most 4% of short positions in the solution and then set τ_2 as the value associated with the maximum ratio as defined in (5.3). We set $\tau_1 = \tau_2 = 10^{-2}$ for FF48-20y, $\tau_1 = 10^{-3}$ and $\tau_2 = 10^{-4}$ for FF100-20y, and $\tau_1 = 10^{-3}$ and $\tau_2 = 10^{-4}$ for ES50-Q. For the tests with different horizon times, we decided to keep the same parameter setting if it provided reasonable portfolios. However, since the values of the parameters corresponding to FF100-20y produced a number of shorts greater than 4% for FF100-30y, we increased them as $\tau_1 = 10^{-2}$ and $\tau_2 = 10^{-3}$.

5.2. Implementation details and numerical results. We developed a MATLAB implementation of Algorithm 1 specifically suited to take into account that problem (5.2) is quadratic. The stopping criterion used for both the standard Bregman iterations and the accelerated ones is based on the violation of the equality constraints, i.e., the execution is halted when

$$\left\|A\mathbf{u}^k - \mathbf{b}\right\| \leq tol_B, \quad \text{and} \quad \left\|D\mathbf{u}^k - \mathbf{d}^k\right\| \leq tol_B,$$

with $tol_B = 10^{-4}$, which guarantees a sufficient accuracy in financial terms. A maximum number of Bregman iterations, equal to 10000, is also set. The parameter λ, penalizing the linear constraint violation in (3.10), is set to 1.

The inner minimization in the standard Bregman iterations, i.e., for the ℓ_1-regularized problems at lines 2, 9 and 12 of Algorithm 1, is performed by means of the FISTA algorithm

from the FOM package.[†] We recall that Theorem 3.1 requires the error in the solution of the subproblems to satisfy hypothesis (i). This condition cannot be used in practice, not only because the solution to the subproblem in (i) is unknown, but also because the required tolerance becomes too small after a few steps. However, as noted in [22, 33], the criterion can be replaced by more practical ones. We decided to stop the minimization when

$$\left\|\mathbf{z}^{l+1} - \mathbf{z}^{l}\right\| \leq tol_F,$$

where $\mathbf{z}^l$ is the l-th FISTA iterate and tol_F is a fixed tolerance. In our tests we set $tol_F = 10^{-5}$ for FF48 and FF100, while for ES50 it is necessary to set $tol_F = 10^{-6}$ to ensure convergence of SB within the maximum number of outer iterations. The maximum number of FISTA iterations is set to 5000.

Regarding the subspace-acceleration steps (line 6 of Algorithm 1), since they can be easily reformulated as unconstrained quadratic optimization problems, we use the conjugate gradient (CG) method. In this case the minimization is stopped when

$$\left\|\boldsymbol{\rho}^{l}\right\| \leq \left\|\boldsymbol{\rho}^{0}\right\| \, tol_{CG},$$

where $\boldsymbol{\rho}^l$ denotes the residual at the l-th CG iteration and tol_{CG} a fixed tolerance. In the tests we set $tol_{CG} = 10^{-2}$; we also choose a maximum number of CG steps equal to half the size of the subproblem to be solved. In the sufficient decrease condition (4.13) we set $\eta = 10^{-1}$.

Concerning criterion (4.10) for switching between the standard Bregman iterations and the subspace-acceleration steps, we observed that small values of γ tend to penalize the execution of acceleration steps, leading to no improvement in the performance of the algorithm. Thus, in order to favor the use of subspace-acceleration steps, we decided to initialize the parameter γ equal to 10 and to update it during the algorithm with an automatic adaptation strategy similar to that used in [18]. In particular, the value of γ is reduced by a factor 0.9 when (4.10) holds, i.e., when subspace-acceleration steps are performed, and is increased by a factor 1.1 otherwise. To warmstart the algorithm, we perform 5 standard Bregman iterations before allowing acceleration.

As regards the safeguard at lines 8-10 of Algorithm 1, by numerical experiments we found that if $\|M\mathbf{x}^{k+1} - \mathbf{s}\| > \|M\mathbf{x}^{k} - \mathbf{s}\|$, then it is generally convenient to accept $\mathbf{x}^{k+1}$, compute $\mathbf{s}^{k+1}$ according to line 4 and solve by FISTA the subproblem involving H^{k+1}.

In order to assess the performance of SBSA, we compared it with two state-of-the-art methods for the solution of problem (1.1):

- the split Bregman iteration in [26, Section 3], which we denote SB;
- the accelerated ADMM algorithm proposed in [11], called AL_SOP.

We note that the SB algorithm is equal to the SBSA algorithm without subspace acceleration. In SB we made the same choices as in SBSA for the solution of the ℓ_1-regularized subproblems with FISTA and the stopping criteria, to make the effect of the acceleration steps clearer. Regarding AL_SOP we observe that, by introducing suitable auxiliary variables, problem (5.2) can be equivalently written as

$$\begin{array}{ll} \min & \langle \mathbf{u}, C\mathbf{u}\rangle + \tau_1 \|\mathbf{v}\|_1 + \tau_2 \|\mathbf{d}\|_1 \\ \text{s.t.} & A\mathbf{u} \quad\;\; = \mathbf{b}, \\ & \mathbf{u} - \mathbf{v} \;\; = \mathbf{0}, \\ & D\,\mathbf{u} - \mathbf{d} = \mathbf{0}. \end{array} \tag{5.7}$$

Given $\mathbf{y}^k = [\mathbf{u}^k; \mathbf{v}^k; \mathbf{d}^k]$, the $(k+1)$-st iteration of the ADMM scheme applied to problem (5.7) consists of the minimization of a quadratic function to determine $\mathbf{u}^{k+1}$ and the

[†] https://sites.google.com/site/fomsolver/

TABLE 5.1
Execution times (seconds) and outer iterations of the four algorithms. "—" indicates that the algorithm does not satisfy the stopping criterion within the maximum number of iterations.

	SBSA		SBSA-LSA		SB		AL_SOP	
Problem	time	outit	time	outit	time	outit	time	outit
FF48-10y	2.61	7	2.61	7	9.38	156	2.31	2235
FF48-20y	6.06	11	6.06	11	9.29	53	10.16	4353
FF48-30y	9.12	14	9.12	14	93.30	693	38.47	8889
FF100-10y	6.63	13	6.63	13	35.81	121	4.52	1502
FF100-20y	17.16	10	17.31	11	19.87	19	19.07	2385
FF100-30y	42.10	9	42.10	9	46.08	21	—	—
ES50-Q	30.80	209	30.96	210	59.59	195	8.06	2743
ES50-A	5.05	304	5.05	305	14.87	269	0.87	1377

application of two soft-thresholding operators to determine $\mathbf{v}^{k+1}$ and $\mathbf{d}^{k+1}$. By adapting the strategy proposed in [11], we introduced at the end of each iteration an acceleration step over the subspace spanned by $\mathbf{y}^{k+1} - \mathbf{y}^k$. The choice of the subspace and the parameter ε in the acceleration step was made by following the choice in [11, Section 4]. In order to make a fair comparison between SBSA and AL_SOP, we decided to use in AL_SOP the same stopping criterion as in SBSA, with the additional requirement $\left\|\mathbf{u}^k - \mathbf{v}^k\right\| \leq tol_B$. Moreover, at each iteration the solution of the quadratic programming problem for computing $\mathbf{u}^{k+1}$ was performed by CG with the same stopping criterion used for the subproblems in SBSA. The maximum number of outer iterations for AL_SOP was set to 25000.

Finally, we also carried out a comparison with a version of SBSA where the last iterate was forced to be a subspace acceleration. In the following this strategy is denoted SBSA-LSA (LSA: Last Step is an Acceleration).

All the tests were performed with MATLAB R2018b on a 2.5 GHz Intel Core i7-6500U with 12 GB RAM, 4 MB L3 Cache, and Windows 10 Pro (ver. 1909) operating system.

The results of the tests are summarized in Tables 5.1 and 5.2. In Table 5.1 we report, for each problem and each of the four algorithms, the number of outer iterations and the execution time in seconds. The number of outer iterations shows that SBSA-LSA performed a further (final) subspace-acceleration step only for the three problems FF100-20y, ES50-Q and ES50-A, without a practical increase of the execution time. We see that the SBSA versions of the split Bregman algorithm are able to outperform SB for all the test problems. The reduction of the total time obtained with SBSA and SBSA-LSA varies between 9% for FF100-30y and 90% for FF48-30y. We note that the cost per iteration of AL_SOP is far smaller than the one of the other algorithms; however, the proposed accelerated method outperforms AL_SOP in terms of time on problems FF48-20y, FF48-30y, FF100-20y and FF100-30y. In particular, for FF100-30y AL_SOP is not able to converge in 25000 iterations, corresponding to more that 360 seconds. AL_SOP requires about the same execution time as SBSA for FF48-10y, while it is much faster for ES50-Q and ES50-A; however, as we will see later, the quality of the solutions of the ES50 problems computed by AL_SOP is worse.

In Table 5.2 we report the values of the quality metrics described in Section 5.1 for the portfolios obtained by using the four algorithms. These metrics are computed before and after thresholding the solution with $\epsilon_1 = \epsilon_2 = 10^{-4}$ (see (5.4) and (5.6)). The values before thresholding are in brackets. For each algorithm we report a single value for the ratio, since it is not practically affected by thresholding (we obtained the same results up to the fourth or fifth significant digit). The table shows that the portfolios produced by SBSA, SBSA-LSA and SB are equivalent in financial terms, since the corresponding thresholded solutions produce the same ratios, numbers of short positions, densities and transaction costs. The

TABLE 5.2
Comparison among the portfolios computed by the four considered algorithms. The values in brackets correspond to solutions without thresholding. $\mathcal{T}_{naive}$ denotes the transaction cost for the naive solution.

Problem	ratio	density	# shorts	$\mathcal{T}$	$\|V\|_1$	$\|V\|_\infty$	$\mathcal{T}_{naive}$
SBSA							
FF48-10y	2.32	15% [19.2%]	0 [0]	30 [104]	6 [10]	8 [11]	480
FF48-20y	2.28	12.6% [14.4%]	0 [0]	55 [148]	11 [20]	7 [8]	960
FF48-30y	4.64	16.3% [17.6%]	29 [29]	109 [274]	14 [30]	15 [16]	1440
FF100-10y	2.94	10.5% [10.5%]	18 [18]	82 [110]	7 [10]	14 [15]	960
FF100-20y	9.08	14.1% [15.7%]	81 [82]	217 [351]	16 [17]	34 [37]	1920
FF100-30y	7.07	7.2% [8.3%]	51 [51]	174 [279]	16 [20]	18 [21]	2880
ES50-Q	2.48	17.9% [29.8%]	0 [0]	45 [380]	10 [22]	9 [32]	1100
ES50-A	2.25	18.3% [32.3%]	0 [0]	17 [114]	3 [6]	10 [27]	300
SBSA-LSA							
FF48-10y	2.32	15% [19.2%]	0 [0]	30 [104]	6 [10]	8 [20]	
FF48-20y	2.28	12.6% [14.4%]	0 [0]	55 [148]	11 [20]	7 [11]	
FF48-30y	4.64	16.3% [17.6%]	29 [29]	109 [274]	14 [30]	15 [16]	
FF100-10y	2.94	10.5% [10.5%]	18 [18]	82 [110]	7 [10]	14 [15]	
FF100-20y	9.08	14.1% [15.7%]	81 [82]	217 [349]	16 [17]	34 [37]	
FF100-30y	7.07	7.2% [8.3%]	51 [51]	174 [279]	16 [20]	18 [21]	
ES50-Q	2.48	17.5% [26.6%]	0 [0]	47 [332]	10 [22]	9 [26]	
ES50-A	2.25	18.3% [28.7%]	0 [0]	17 [104]	3 [6]	10 [25]	
SB							
FF48-10y	2.32	15% [17.5%]	0 [0]	30 [93]	6 [10]	8 [16]	
FF48-20y	2.28	12.6% [14.9%]	0 [0]	55 [165]	11 [20]	7 [17]	
FF48-30y	4.64	16.3% [18.0%]	29 [41]	109 [286]	14 [30]	15 [24]	
FF100-10y	2.94	10.5% [10.6%]	18 [18]	82 [112]	7 [10]	14 [15]	
FF100-20y	9.08	14.1% [15.7%]	81 [89]	217 [339]	16 [18]	34 [40]	
FF100-30y	7.07	7.2% [8.8%]	51 [51]	175 [312]	16 [20]	18 [33]	
ES50-Q	2.48	15.5% [28.3%]	0 [0]	48 [355]	11 [22]	10 [26]	
ES50-A	2.27	18.3% [33.3%]	0 [0]	16 [105]	3 [6]	10 [27]	
AL_SOP							
FF48-10y	2.32	15% [100%]	0 [194]	31 [480]	7 [10]	8 [48]	
FF48-20y	2.28	12.6% [100%]	0 [420]	55 [960]	11 [20]	7 [48]	
FF48-30y	4.64	16.4% [100%]	29 [699]	113 [1440]	14 [30]	15 [48]	
FF100-10y	2.91	12.9% [100%]	18 [488]	107 [960]	9 [10]	17 [96]	
FF100-20y	8.95	17.8% [100%]	89 [560]	307 [1920]	17 [20]	45 [96]	
FF100-30y	7.07	7.8% [100%]	59 [1465]	201 [2880]	18 [30]	18 [96]	
ES50-Q	0.85	100% [100%]	0 [0]	698 [1100]	18 [22]	50 [50]	
ES50-A	2.00	45.3% [100%]	0 [124]	35 [300]	3 [6]	21 [50]	

densities, transaction costs and numbers of shorts obtained before thresholding give further information on the quality of the optimal solution found by the algorithms. The additional subspace-acceleration step performed by SBSA-LSA on the ESQ50 problems allows us to obtain solutions with slightly smaller densities and smaller transaction costs. Inspection of the non-thresholded solutions of SBSA-LSA and SB shows that in general our subspace-accelerated algorithm is able to compute solutions comparable with those of SB in terms of objective function values. On the other hand, the non-thresholded solutions obtained by SBSA-LSA may have slightly smaller densities or transaction costs. By looking at the thresholded solutions obtained with AL_SOP we observe that for the FF100 problems they produce portfolios with slightly poorer qualities since they have a little higher densities, shorts and transaction costs. Regarding the ES50 problems, for which AL_SOP outperformed SBSA

and SBSA-LSA in terms of time, we see that the portfolio computed for ES50-A has a smaller ratio and a much greater density and transaction cost as compared with the other methods, while almost all the metrics concerning the portfolio produced for ES50-Q are worse than those obtained with the other algorithms. In particular, for ES50-Q the ratio is smaller than 1 and hence the computed portfolio it is not able to satisfy the financial requirements.

In our opinion, the results suggest that the proposed split Bregman method with subspace acceleration is not only efficient in terms of computational cost, but is also better than the SB and AL_SOP methods in enforcing structured sparsity in the solution, especially when no thresholding is applied. This behavior seems to depend on the backtracking projected line search performed at each acceleration step, which allows us to set variables exactly to zero.

6. Conclusions. A Split Bregman method with Subspace Acceleration (SBSA) has been proposed for sparse data recovery with joint ℓ_1-type regularizers. The acceleration technique, inspired by orthant-based methods, consists in replacing ℓ_1-regularized subproblems at certain iterations with smooth unconstrained optimization problems over orthant faces identified by zero variables. These smooth problems can be solved by fast methods. Suitable optimality measures are used to decide whether to perform subspace acceleration. Numerical experiments show that SBSA is effective in solving multi-period portfolio optimization problems and outperforms the Split Bregman method and the Accelerated ADMM algorithm proposed in [11] in terms of computational time and quality of the solution.

Future work will focus on the solution of problems where the function f in (1.1) is not quadratic, such as those arising in some classification tasks on fMRI data [19].

Acknowledgments. The authors thanks the reviewers for their useful comments, which helped them improve this article. Marco Viola also thanks Daniel Robinson for useful discussions about orthant-based methods for ℓ_1-regularized optimization problems.

REFERENCES

[1] G. ANDREW AND J. GAO, *Scalable training of L1-regularized log-linear models*, in International Conference on Machine Learning, January 2007.

[2] L. BALDASSARRE, J. MOURAO-MIRANDA, AND M. PONTIL, *Structured sparsity models for brain decoding from fMRI data*, in 2012 Second International Workshop on Pattern Recognition in NeuroImaging, July 2012, pp. 5–8.

[3] A. BECK AND M. TEBOULLE, *A fast iterative shrinkage-thresholding algorithm for linear inverse problems*, SIAM Journal on Imaging Sciences, 2 (2009), pp. 183–202.

[4] H. BECKER, L. ALBERA, P. COMON, J.-C. NUNES, R. GRIBONVAL, J. FLEUREAU, P. GUILLOTEL, AND I. MERLET, *SISSY: An efficient and automatic algorithm for the analysis of EEG sources based on structured sparsity*, NeuroImage, 157 (2017), pp. 157–172.

[5] 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, 3 (2011), pp. 1–122.

[6] L. M. BREGMAN, *The relaxation method of finding the common point of convex sets and its application to the solution of problems in convex programming*, USSR Computational Mathematics and Mathematical Physics, 7 (1967), pp. 200–217.

[7] H. BREZIS, *Opérateurs maximaux monotones et semi-groupes de contractions dans les espaces de Hilbert. Number 5 in North Holland Math*, Studies. North-Holland, Amsterdam, (1973).

[8] J. BRODIE, I. DAUBECHIES, C. DE MOL, D. GIANNONE, AND I. LORIS, *Sparse and stable Markowitz portfolios*, Proceedings of the National Academy of Sciences, 106 (2009), pp. 12267–12272.

[9] R. H. BYRD, G. M. CHIN, J. NOCEDAL, AND Y. WU, *Sample size selection in optimization methods for machine learning*, Mathematical Programming, 134 (2012), pp. 127–155.

[10] R. CAMPAGNA, S. CRISCI, S. CUOMO, L. MARCELLINO, AND G. TORALDO, *Modification of TV-ROF denoising model based on split Bregman iterations*, Applied Mathematics and Computation, 315 (2017), pp. 45–467.

[11] D.-Q. CHEN, L.-Z. CHENG, AND F. SU, *Restoration of images based on subspace optimization accelerating augmented Lagrangian approach*, Journal of Computational and Applied Mathematics, 235 (2011), pp. 2766–2774.

[12] T. CHEN, F. E. CURTIS, AND D. P. ROBINSON, *A reduced-space algorithm for minimizing ℓ_1-regularized convex functions*, SIAM Journal on Optimization, 27 (2017), pp. 1583–1610.

[13] ———, *FaRSA for $\ell 1$-regularized convex optimization: local convergence and numerical experience*, Optimization Methods and Software, 33 (2018), pp. 396–415.

[14] Y. CHEN, W. W. HAGER, M. YASHTINI, X. YE, AND H. ZHANG, *Bregman operator splitting with variable stepsize for total variation image reconstruction*, Computational Optimization and Applications, 54 (2013), pp. 317–342.

[15] S. CORSARO AND V. DE SIMONE, *Adaptive l_1-regularization for short-selling control in portfolio selection*, Computational Optimization and Applications, 72 (2019), pp. 457–478.

[16] S. CORSARO, V. DE SIMONE, AND Z. MARINO, *Fused lasso approach in portfolio selection*, Annals of Operations Research, (2019). DOI: 10.1007/s10479-019-03289-w.

[17] S. CORSARO, V. DE SIMONE, Z. MARINO, AND F. PERLA, *L_1-regularization for multi-period portfolio selection*, Annals of Operations Research, (2019). DOI: 10.1007/s10479-019-03308-w.

[18] D. DI SERAFINO, G. TORALDO, M. VIOLA, AND J. BARLOW, *A two-phase gradient method for quadratic programming problems with a single linear constraint and bounds on the variables*, SIAM Journal on Optimization, 28 (2018), pp. 2809–2838.

[19] E. D. DOHMATOB, A. GRAMFORT, B. THIRION, AND G. VAROQUAUX, *Benchmarking solvers for TV-ℓ_1 least-squares and logistic regression in brain imaging*, in 2014 International Workshop on Pattern Recognition in Neuroimaging, June 2014, pp. 1–4.

[20] Z. DOSTÁL, *Box constrained quadratic programming with proportioning and projections*, SIAM Journal on Optimization, 7 (1997), pp. 871–887.

[21] Z. DOSTÁL AND J. SCHÖBERL, *Minimizing quadratic functions subject to bound constraints with the rate of convergence and finite termination*, Computational Optimization and Applications, 30 (2005), pp. 23–43.

[22] J. ECKSTEIN AND D. P. BERTSEKAS, *On the Douglas-Rachford splitting method and the proximal point algorithm for maximal monotone operators*, Mathematical Programming, 55 (1992), pp. 293–318.

[23] S. ESEDOĞLU AND S. J. OSHER, *Decomposition of images by the anisotropic Rudin-Osher-Fatemi model*, Communications on Pure and Applied Mathematics, 57 (2004), pp. 1609–1626.

[24] A. FRIEDLANDER AND J. M. MARTÍNEZ, *On the numerical solution of bound constrained optimization problems*, RAIRO - Operations Research, 23 (1989), pp. 319–341.

[25] ———, *On the maximization of a concave quadratic function with box constraints*, SIAM Journal on Optimization, 4 (1994), pp. 177–192.

[26] T. GOLDSTEIN AND S. OSHER, *The split Bregman method for L1-regularized problems*, SIAM Journal on Imaging Sciences, 2 (2009), pp. 323–343.

[27] J. HIRIART-URRUTY AND C. LEMARECHAL, *Convex Analysis and Minimization Algorithms II: Advanced Theory and Bundle Methods*, Grundlehren der mathematischen Wissenschaften, Springer Berlin Heidelberg, 1996.

[28] N. KESKAR, J. NOCEDAL, F. ÖZTOPRAK, AND A. WÄCHTER, *A second-order method for convex ℓ_1-regularized optimization with active-set prediction*, Optimization Methods and Software, 31 (2016), pp. 605–621.

[29] H. MOHY-UD-DIN AND D. P. ROBINSON, *A solver for nonconvex bound-constrained quadratic optimization*, SIAM Journal on Optimization, 25 (2015), pp. 2385–2407.

[30] Y. S. NIU, N. HAO, AND H. ZHANG, *Multiple change-point detection: a selective overview*, Statistical Science, 31 (2016), pp. 611–623.

[31] S. OSHER, M. BURGER, D. GOLDFARB, J. XU, AND W. YIN, *An iterative regularization method for total variation-based image restoration*, Multiscale Modeling & Simulation, 4 (2005), pp. 460–489.

[32] R. T. ROCKAFELLAR, *Convex Analysis*, vol. 28, Princeton University Press, 1970.

[33] ———, *Monotone operators and the proximal point algorithm*, SIAM journal on control and optimization, 14 (1976), pp. 877–898.

[34] S. SOLNTSEV, J. NOCEDAL, AND R. H. BYRD, *An algorithm for quadratic $\ell 1$-regularized optimization with a flexible active-set strategy*, Optimization Methods and Software, 30 (2015), pp. 1213–1237.

[35] R. TIBSHIRANI, M. SAUNDERS, S. ROSSET, J. ZHU, AND K. KNIGHT, *Sparsity and smoothness via the fused lasso*, Journal of the Royal Statistical Society: Series B (Statistical Methodology), 67 (2005), pp. 91–108.

[36] S. J. WRIGHT, R. D. NOWAK, AND M. A. T. FIGUEIREDO, *Sparse reconstruction by separable approximation*, IEEE Transactions on Signal Processing, 57 (2009), pp. 2479–2493.