MODIFIED BREGMAN ITERATION FOR PORTFOLIO
OPTIMIZATION∗

STEFANIA CORSARO† AND VALENTINA DE SIMONE‡

Abstract. We consider the $l_1$-regularized Markowitz model, where a $l_1$-penalty term is added to
the objective function of the classical mean-variance one to stabilize the solution process, promoting
sparsity in the solution and avoiding short positions. In this paper, we consider the Bregman iteration
method to solve the related constrained optimization problem. We propose an iterative algorithm
based on a modified Bregman iteration, in which an adaptive updating rule for the regularization
parameter is defined. Our main result shows that the modified scheme preserves the properties of
the original one. Numerical tests are reported, which show the effectiveness of the our approach.

Key words. Portfolio selection. Markowitz model. $l_1$-regularization. Bregman iteration.

AMS subject classifications. 91G10, 65F22, 90C30

1. Introduction. In the classical Markowitz mean-variance framework [12], port-
folio selection aims at the construction of an investment portfolio that exposes investor
to minimum risk providing him a fixed expected return. This approach was proposed
by Markowitz in his aforementioned seminal paper, where he stated that portfolio
selection strategy should provide an optimal trade-off between expected return and
risk (mean-variance approach). In a successive work [13], Markowitz reinforced his
theory arguing that, under certain, mild conditions, a portfolio from a mean-variance
efficient frontier will approximately maximize the investor’s expected utility.
Markowitz model relies on information about future, since expected returns should ac-
tually be computed discounting future flows, that are clearly not available. A common
choice is to use historical data as predictive of the future behavior of asset returns.
This practice has certain drawbacks; indeed, a limited amount of relevant historical
data is often available. Moreover, correlation between assets returns can lead to ill-
conditioned covariance matrices.
It is well known that errors in estimation of expected values affect solutions more
severely than errors on variances. For this reason, to overcome this issue some au-
thors focus on minimum-variance portfolios, which do not take into account the return
constraint. We recall [20] and references therein. Moreover, different regularization
techniques have been suggested; a review of them can be found in [6]. Among these,
penalization techniques have been considered, both for the minimum- and the mean-
variance approach. In [20] $l_1$ and squared-$l_2$ norm constraints are proposed for the
minimum-variance criterion. In [23] an algorithm for the optimal minimum-variance
portfolio selection with a weighted $l_1$ and squared-$l_2$ norm penalty is presented. In
[5] $l_p$-norm regularized models are proposed.
In this paper, we consider the $l_1$ mean-variance regularized model introduced in [10],
where a $l_1$—penalty term is added to promote sparsity in the solution. Since solu-
tions establish the amount of capital to be invested in each available security, sparsity
means that money are invested in a few securities. This allows investor to reduce both
the number of positions to be monitored and the transaction costs, particularly rele-
vant for small investors, that are not taken into account in the theoretical Markowitz
model. Another useful interpretation of $l_1$ regularization is related to the amount of
shorting in the portfolio; from the financial point of view negative solutions correspond to short sales, investment strategies where the investor sells shares of borrowed stock in the open market with the expectation that the price of the stock will decrease over time. In many markets, among which Italy, Germany and Switzerland, restrictions on short sales have been established in the last years, thus positivity is desired as well. Then, the choice of the regularization parameter is mandatory in order to provide sparse, non-negative solutions preserving fidelity to data. In this paper we propose an iterative algorithm based on a modified Bregman iteration. Bregman iteration was recently introduced with success in many fields in which problems are formulated in term of $l_1$ functionals, as image restoration [17] matrix rank minimization [16] and compressed sensing [22]. Our modification to the original scheme is related to an adaptive updating rule for the regularization parameter in the regularized model. We show that our modified scheme preserves the properties of the original one and is able to select a good value of regularization parameter within a negligible computational time. Numerical tests confirm the effectiveness of the proposed algorithm.

The paper is organized as follows. In section 2 we briefly recall Markowitz mean-variance model. In section 3 we introduce Bregman iteration for portfolio selection. Our main results are in section 4, where we introduce our algorithm, based on a modified Bregman iteration, for the $l_1$-regularized Markowitz model. In section 5 we validate our approach by means of several numerical exsperimen ts. Finally, in section 6 we give some conclusion and outline future work.

2. Portfolio selection model. We refer to the classical Markowitz mean-variance model. Given $n$ traded assets, the core of the problem is to establish the amount of capital to be invested in each available security. We assume that one unit of capital is available and define

$$w = (w_1, w_2, \ldots, w_n)^T$$

the portfolio weight vector, that is, the amount $w_i$ is invested in the $i$-th security. Asset returns are assumed to be stationary. If we denote with

$$\mu = (\mu_1, \mu_2, \ldots, \mu_n)^T$$

the expected asset returns, then the expected portfolio return is their weighted sum:

$$\sum_{i=1}^{n} w_i \mu_i.$$  \hfill (1)

We moreover denote with $\sigma_{ij}$ is the covariance between returns of securities $i$ and $j$. The portfolio risk is measured by means of its variance, given by:

$$V = \sum_{i=1}^{n} \sum_{j=1}^{n} \sigma_{ij} w_i w_j$$

Let $\rho$ be the fixed expected portfolio return and $C$ the covariance matrix of returns. Portfolio selection is formulated as the following quadratic constrained optimization problem:

$$\min_{w} w^T C w$$

s.t.

$$w^T \mu = \rho$$

$$w^T 1 = 1,$$  \hfill (2)
where \( \mathbf{1} \) is the unitary column vector. The first constraint fixes the expected return, according to (1). The second one is a budget constraint which establishes that all the available capital is invested. The non-negativity constraint is often added to avoid short positions.

Let us consider a set of \( m \) evenly spaced dates

\[
t = (t_1, t_2, \ldots, t_m)
\]

at which asset returns are estimated and build the matrix \( R \in \mathbb{R}^{m \times n} \) that contains observed historical returns of asset \( i \) on its \( i \)-th column. It can be shown that problem (2) can be stated in the following form:

\[
\begin{align*}
\min_w & \frac{1}{m} \| \rho \mathbf{1} - Rw \|_2^2 \\
\text{s.t.} & \quad w^T \mu = \rho \\
& \quad w^T \mathbf{1} = 1.
\end{align*}
\]

(3)

As the asset returns tend to be highly correlated, the matrix \( R \) has some singular values close to zero; therefore regularization techniques, that add to objective function some form of a priori knowledge about the solution, must be considered. In this paper we consider the following \( l_1 \)-regularized problem:

\[
\begin{align*}
\min_w & \| \rho \mathbf{1} - Rw \|_2^2 + \tau \| w \|_1 \\
\text{s.t.} & \quad w^T \mu = \rho \\
& \quad w^T \mathbf{1} = 1,
\end{align*}
\]

(4)

where the \( 1/m \) term has been incorporated into the regularization one. When removing the constraints, the problem is referred to as lasso regression [19].

From the second constraint in (4) it follows that the objective function can be equivalently written as:

\[
\| \rho \mathbf{1}_m - Rw \|_2^2 + 2\tau \sum_{i:w_i<0} |w_i| + \tau.
\]

This form points out that \( l_1 \) penalty is equivalent to a penalty on short positions. In the limit of very large values of the regularization parameter, we obtain a portfolio with only positive weights, as observed also in [15].

3. Bregman iteration for portfolio selection. Portfolio selection can be formulated as the constrained nonlinear optimization problem:

\[
\begin{align*}
\min_w & \quad E(w) \\
\text{s.t.} & \quad Aw = b,
\end{align*}
\]

(5)

where

\[
E(w) = \| \rho \mathbf{1} - Rw \|_2^2 + \tau \| w \|_1
\]

is strictly convex and non-smooth due to the presence of the \( l_1 \) penalty term, and

\[
A = \begin{pmatrix} \mu^T \\ \mathbf{1} \end{pmatrix} \in \mathbb{R}^{2 \times n} \quad \text{and} \quad b = (\rho, 1)^T \in \mathbb{R}^2.
\]
One way to solve (5) is to convert it into an unconstrained problem, for example by using a penalty function/continuation method, which approximates it by a sequence:

\[
\min_w E(w) + \frac{\lambda_k}{2}\|Aw - b\|_2^2, \quad \lambda_k \in \mathbb{R}^+.
\]

It is well known that, if the \(k\)-th subproblem (6) has solution \(w_k\) and \(\{\lambda_k\}\) is an increasing sequence tending to \(\infty\) as \(k \to \infty\), any limit point of \(\{w_k\}\) is a solution of (5) [11]. Therefore, in many problems it is necessary to choose very large values of \(\lambda_k\) and it makes (6) extremely difficult to solve numerically. Alternatively, Bregman iteration can be used to reduce (5) in a short sequence of unconstrained problems by using the Bregman distance associated with \(E\) [4], where, conversely, the value of \(\lambda_k\) in (6) remains constant.

The Bregman distance [4] associated with a proper convex functional \(E(w): \mathbb{R}^n \to \mathbb{R}\) at point \(v\) is defined as:

\[
D_E^p(w, v) = E(w) - E(v) - < p, w - v >,
\]

where \(p \in \partial E(v)\) is a subgradient in the subdifferential of \(E\) at point \(v\) and \(<.,.>\) denotes the canonical inner product in \(\mathbb{R}^n\). It is not a distance in the usual sense because it is not in general symmetric but it does measure closeness between \(w\) and \(v\) in the sense that if \(u\) lies on the line segment \((w, v)\), then the line segment \((w, u)\) has smaller Bregman distance than \((w, v)\) does. At each Bregman iteration \(E(w)\) is replaced by the Bregman distance so a subproblem in the form of (6) is solved according to the following iterative scheme:

\[
\begin{align*}
\text{w}_{k+1} &= \arg\min_w D_E^{pk}(w, w_k) + \frac{\lambda_k}{2}\|Aw - b\|_2^2, \\
\text{p}_{k+1} &= \text{p}_k - \lambda A^T(A\text{w}_{k+1} - b) \in \partial E(\text{w}_{k+1}).
\end{align*}
\]

The updating rule of \(\text{p}_{k+1}\) is chosen according to the first-order optimality condition for \(\text{w}_{k+1}\) and ensures that \(D_E^{pk+1}(w, w_{k+1})\) is well defined. Under suitable hypotheses the convergence of the sequence \(\{w_k\}\) to the solution of the constrained problem (5) is guaranteed in a finite number of steps [17]; furthermore, using the equivalence of Bregman iteration with the augmented Lagrangian one [8], convergence is proved also in [9]. Note that the convergence results guarantee the monotonic decrease of \(\|Aw_k - b\|_2^2\), thus for large \(k\) the constraint conditions are satisfied to an arbitrary high degree of accuracy. This yields a natural stopping criterion according to a discrepancy principle.

In [22] it is shown that (8) can be simplified as:

\[
\begin{align*}
\text{w}_{k+1} &= \arg\min_w E(w) + \frac{\lambda_k}{2}\|Aw - b\|_2^2, \\
\text{b}_{k+1} &= \text{b}_k + b - A\text{w}_{k+1}.
\end{align*}
\]

The schemes (8) and (9) are equivalent in the sense that the objective function in these two versions is the same (up to a constant) for all \(k\) [8], that is:

\[
D_E^{pk}(w, w_k) + \frac{\lambda}{2}\|Aw - b\|_2^2 = E(w) + \frac{\lambda}{2}\|Aw - b_k\|_2^2 + C.
\]

Since there is generally no explicit expression for the solution of the sub-minimization problem involved in (9), at each iteration the solution is computed inexactly using an iterative solver. So, in the last years there has been a growing interest about inexact solution of the subproblem involved in Bregman iteration. In recent papers it is proved...
that, for many applications, Bregman iterations yield very accurate solutions even if subproblems are not solved as accurately [17, 18, 22]. In [24] convergence results are obtained for piece-wise linear convex functionals. In [3] the inexactness in the inner solution is controlled by a criterion that preserves the convergence of the Bregman iteration and its features in image restoration.

4. Modified Bregman iteration. A crucial issue in the solution of (4) is the choice of a suitable value for the regularization parameter \( \tau \), as already pointed out. The aim is to select \( \tau \) so to realize a trade-off between sparsity and positivity (requiring sufficiently large values) and fidelity to data (requiring small values). While the literature offers a significative number of methods for Tikhonov regularization [21], \( l_1 \) regularization parameter selection is often based on problem-dependent criteria and related to iterative empirical estimates, that require a high computational cost. In [10] least-angle regression (LARS) algorithm proceeds by decreasing the value of \( \tau \) progressively from very large values, exploiting the fact that the dependence of the optimal weight on \( \tau \) is piecewise linear.

In this section, we present a numerical algorithm, based on a modified Bregman iteration with adaptive updating rule for \( \tau \). Our basic idea for defining the rule for \( \tau \) comes from the following proposition [10]:

**Proposition 1.** Let \( w_{\tau_1} \) and \( w_{\tau_2} \) be solution of the \( l_1 \)-regularized problem (4) with \( \tau_1 \) and \( \tau_2 \) respectively. If some of \( (w_{\tau_2})_i \) are negative and all the entries in \( w_{\tau_1} \) are positive or zero, we have \( \tau_1 > \tau_2 \).

We propose an updating rule for \( \tau \) that generates an increasing sequence of values; we start with a small initial value \( \tau_0 \) and increase it as long as short positions occur, so that \( \tau \) is fixed to the minimum of a discrete set of parameters that does not provide negative weights.

Let

\[
E_k(w) = \|Rw - p\|_2^2 + \tau_k \|w\|_1, \quad k = 0, 1, \ldots
\]

We now prove the main result of this paper:

**Theorem 2.** Given \((w_{k+1}, p_{k+1})\) provided by (8) applied to \( E_k \), if \( \tau_{k+1} \geq \tau_k \), then

\[
\frac{\tau_{k+1}}{\tau_k} p_{k+1} \in \partial E_{k+1}(w_{k+1}).
\]

**Proof.** It holds \( p_{k+1} \in \partial E_k(w_{k+1}) \), that is:

\[
E_k(w) \geq E_k(w_{k+1}) + < p_{k+1}, w - w_{k+1} >.
\]

We have \( E_{k+1}(w) \geq E_k(w) \) since \( \tau_{k+1} \geq \tau_k \); then:

\[
E_{k+1}(w) \geq E_k(w_{k+1}) + < p_{k+1}, w - w_{k+1} >
\]

\[
\geq \frac{\tau_{k+1}}{\tau_k} \left( E_k(w_{k+1}) + < p_{k+1}, w - w_{k+1} > \right)
\]

\[
= \frac{\tau_{k+1}}{\tau_k} \|Rw_{k+1} - p\|_2^2 + \tau_{k+1} \|w_{k+1}\|_1 + \frac{\tau_{k+1}}{\tau_k} < p_{k+1}, w - w_{k+1} >
\]

\[
\geq \|Rw_{k+1} - p\|_2^2 + \tau_{k+1} \|w_{k+1}\|_1 + < \frac{\tau_{k+1}}{\tau_k} p_{k+1}, w - w_{k+1} >
\]

\[
= E_{k+1}(w_{k+1}) + < \frac{\tau_{k+1}}{\tau_k} p_{k+1}, w - w_{k+1} >
\]
which completes the proof.

We propose the modified Bregman iteration:

\[
\begin{align*}
\mathbf{w}_{k+1} &= \arg\min_{\mathbf{w}} D_{E_k}^{P_k}(\mathbf{w}, \mathbf{w}_k) + \frac{\lambda}{2} \|A\mathbf{w} - \mathbf{b}\|_2^2, \\
\tau_{k+1} &= \phi(|W_{k+1}^-|)\tau_k, \\
P_{k+1} &= \frac{\tau_{k+1}}{\tau_k} P_k - \lambda A^T(A\mathbf{w}_{k+1} - b),
\end{align*}
\]

where $W_{k+1}^- = \{i : (\mathbf{w}_{k+1})_i < 0\}$ and $\phi : \mathbb{N}_0 \to [1, +\infty]$ such that $\phi(0) = 1$, in order to fix the value of $\tau$ to the first encountered one, that we call $\tau^+$, that produces positive solutions. All our experiments reveal that this actually occurs in a small number of steps $k^+$.

Note that one could modify $\phi$ so to control the number of short positions rather than avoiding them.

Relation (10) in Theorem 2 guarantees that the iterative scheme (11) is well defined, thus preserves properties of the original one.

Using this and the equivalence between (8) and (9) we propose the Algorithm 1.

\textbf{Algorithm 1} Modified Bregman Iteration for portfolio selection

\begin{algorithmic}
\STATE Given $0 < \tau_0 < 1$, $\lambda$
\STATE $k := 0$
\STATE $\mathbf{w}_k := 0, b_k := 0, \tau_k = \tau_0$
\WHILE{“not convergence”}
\STATE $\mathbf{w}_{k+1} := \arg\min_{\mathbf{w}} E_k(\mathbf{w}) + \frac{\lambda}{2} \|A\mathbf{w} - \mathbf{b}_k\|_2^2$
\STATE $\text{num}_{\text{short}} := |\{i : (\mathbf{w}_{k+1})_i < 0\}|$
\STATE $\tau_{k+1} := \phi(\text{num}_{\text{short}})\tau_k$
\STATE $\mathbf{b}_{k+1} := \mathbf{b}_k + \frac{\tau_{k+1}}{\tau_k} (\mathbf{b} - A\mathbf{w}_{k+1})$
\STATE $k := k + 1$
\ENDWHILE
\end{algorithmic}

Following Theorem 2.2 in [8] it is possible to prove the next result.

\begin{theorem}
Suppose that at a certain step $\bar{k} \geq k_+$ the iterate $\mathbf{w}_{\bar{k}}$, produced by the Algorithm 1 satisfies $A\mathbf{w}_{\bar{k}} = b$. Then $\mathbf{w}_{\bar{k}}$ is a solution to the constrained problem.
\end{theorem}

\begin{align*}
\min_{\mathbf{w}} E_{k_+}(\mathbf{w}) \\
\text{s.t.} \\
A\mathbf{w} = \mathbf{b}.
\end{align*}

This result shows that if the succession provided by Algorithm 1 converges in the sense of $\lim_{k \to \infty} \|A\mathbf{w}_k - b\|_2 = 0$, then the iterates $\mathbf{w}_k$ will get arbitrarily close to a solution of the original constrained problem with a selected regularization parameter $\tau^+$.

5. Experimental results. In this section, we discuss some computational issues and show the effectiveness of Algorithm 1 for solving the regularized portfolio optimization problem.

In Algorithm 1 we set $\lambda = 10$, $\tau_0 = 2^{-5}$ and

\[\phi(m) = \begin{cases} 
1 & m = 0 \\
2 & m > 0;
\end{cases}\]
iterations are stopped as soon as $\|b_k - b_{k-1}\| \leq Tol$ with $Tol = 10^{-4}$ that, from the financial point of the view, guarantees constraints at a sufficient accuracy. We implement the Fast Proximal Gradient method with backtracking stepsize rule (FISTA) [1] to solve the unconstrained subproblem at each modified Bregman iteration in Algorithm 1. FISTA is an accelerated variant of Forward Backward (FB) algorithm, built upon the ideas of Güler [7] and Nesterov [14]. Note that FB is a first-order method for minimizing objective functions $F(x) = f(x) + g(x)$, where $g : \mathbb{R}^n \rightarrow \mathbb{R}$ is a proper, convex, lower semicontinuous function with $\text{dom}(g)$ closed and $f : \mathbb{R}^n \rightarrow \mathbb{R}$ is convex and $\nabla f$ is $L$-Lipschitz continuous. It generates a sequence $(x_n)_{n \in \mathbb{N}}$ in two separate stages; the former performs a forward (explicit) step which involves only $f$, while the latter performs a backward (implicit) step involving a proximal map associated to $g$ [2]. In our case we set $f = \|\rho 1 - R w\|^2 + \frac{\lambda}{2}\|A w - b\|^2$ and $g = \tau \|w\|_1$, then the proximal map of $g$ is the simple and explicit Soft threshold operator:

$$\text{Prox}_g(w_i) = \text{sgn}(w_i) (|w_i| - \min\{|w_i|, \tau\}).$$

Inner iterations are stopped when the relative difference in Euclidean norm between two successive iterates is less than $Tol_{inn} = 10^{-4}$. All our experiments, some of which are reported in the following, show that it is not worth to require a great accuracy to the inner solver from the computational point of view.

The tests have been performed in Matlab R2015a (v. 8.5, 64-bit) environment, on a six-core Xeon processor with 24 GB of RAM and 12 MB of cache memory, running Ubuntu/Linux 12.04.5. We compare our optimal portfolios with the evenly weighted one (the naive portfolio), usually taken as benchmark in literature. This essentially for two reasons: it is easy to implement and many investors still use such simple rule to allocate their wealth across assets.

We evaluate our approach observing the out-of-sample performances of optimal portfolios as in [20]. This means that for each T-years period of asset returns, we use historical series to solve (4); the target return $\rho$ is fixed to the average return provided by the naive portfolio in those years. The optimal solution obtained in this way is used to build a portfolio that is retained for 1 year. We continue this process by moving one year ahead until we reach the end of the period, ending with a series of out-of-sample portfolios. We then compare the so obtained average return $\hat{\rho}$ and standard deviation values $\hat{\sigma}$ with the corresponding ones of the naive portfolio. We moreover compute the sharp ratio $SR = \hat{\rho}/\hat{\sigma}$: since one would desire great return and small variance values, the sharp ratio can be taken as reference value for the comparison. We present the results on three test problems; the first and the second one come from Fama and French database\textsuperscript{1}. The last one is built on data from Italian market.

### 5.1. Test 1: FF48

We consider the first database - FF48 - which contains monthly returns of 48 industry sector portfolios from July 1926 to December 2015. Using data from 1970 to 2015, we construct optimal portfolios and analyze their out-of-sample performance. Starting from July 1970, we use the $T = 5$–years so 40 optimal portfolios are built, until June 2015. Portfolios in FF48 exhibit moderate correlation, indeed $\text{cond}(R^T R) = O(10^4)$ for all simulations. We tested difference values of $Tol_{inn}$; all our experiments, show that lower values of $Tol_{inn}$ do not improve results, thus we show results obtained for $Tol_{inn} = 10^{-4}$.

\textsuperscript{1}data available at http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html#BookEquity
In table 1, for both optimal and naive portfolio, expected return, standard deviation and sharp ratio are reported, all expressed on annual basis. Values refer to average values computed over 8 years, grouped as described in the first column of the table. The first row contains average values computed over the all 40–years period of simulation. Note that, since one would desire great return and small variance values, the sharp ratio can be taken as reference value for the comparison. In all cases, optimal portfolio exhibits greater values of sharp ratio than the naive one. In figure 1

<table>
<thead>
<tr>
<th>Period</th>
<th>Optimal portfolio</th>
<th>Naive portfolio</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\rho$</td>
<td>$\sigma$</td>
</tr>
<tr>
<td>1975/07 – 2015/06</td>
<td>14%</td>
<td>39%</td>
</tr>
<tr>
<td>1975/07 – 1983/06</td>
<td>23%</td>
<td>45%</td>
</tr>
<tr>
<td>1983/07 – 1991/06</td>
<td>14%</td>
<td>40%</td>
</tr>
<tr>
<td>1991/07 – 1999/06</td>
<td>15%</td>
<td>30%</td>
</tr>
<tr>
<td>1999/07 – 2007/06</td>
<td>13%</td>
<td>35%</td>
</tr>
<tr>
<td>2007/07 – 2015/06</td>
<td>7%</td>
<td>44%</td>
</tr>
</tbody>
</table>

Table 1

Comparison between optimal and naive portfolio for FF48. Reported return and standard deviation are average values over 40 years (first line) and over groups of 8 years (lines 2 – 6).

we report the number of active positions in optimal no-short portfolios (top) and the number of modified Bregman iterations (bottom) for each year of simulation. We note the fast convergence of the Algorithm 1, with an average of the number of iterations around 12. Values of $\tau_+$ range between $2^{-3}$ and $2^{-1}$, showing that the our adaptive rule for $\tau$ in a small number of iteration selects a small regularization parameter able to promote sparsity (the percentage of sparsity varies from 6% to 23%) and positivity.

5.2. Test 2: FF100. We here show results on the second database by Fama and French - FF100 - containing data of 100 portfolios which are the intersections of 10 portfolios formed on size and 10 portfolios formed on the ratio of book equity to market equity. Also FF100 contains monthly returns from from July 1926 to December 2015. We consider 96 portfolios of the 100 available, selected with a preprocessing procedure which eliminates elements with highest volatilities.

We apply the same strategy as in FF48 ($T = 5$–years, 40 optimal portfolios constructed). Correlation values observed in FF100 are higher than in the previous test, the conditioning of $R^TR$ is $O(10^{18})$.

In table 2, we report the same analysis done before: expected return, standard deviation and sharp ratio expressed on annual are basis are reported. On the overall period, optimal portfolio outperforms the naive one. Values of $\tau_+$ range between $2^{-4}$ and $2^{-2}$, the percentage of sparsity varies from 4% to 27% (Fig. 3). Note that, looking at details on each year of simulation, we observe negative returns for both optimal and naive portfolio. For instance, in the 8th year of simulation, optimal portfolio produces a loss of 4%, the naive one of 12%. In the 10th year the losses are of 1% and 10% respectively. This happens because almost all components in portfolios show decreased returns. Finally, we note that in the period 07/1975 – 06/1983 naive portfolio outperforms the optimal one. For instance, in the 3rd year of simulation the optimal portfolio, which contains 5 assets (56, 90, 91, 93, 95), produces a gain of 8%, versus a gain of 18% of the naive one. This behavior is essentially due to a drastic change in asset returns with respect to historical data. To see this, in Fig. 2 we report the average 5–years returns in the period from July 1972 to June 1977, used to build the
Fig. 1. FF48. Top: active positions in optimal portfolios. Bottom: number of modified Bregman iterations.

The optimal portfolio, and the ones observed in the successive year, that actually determine the optimal portfolio return. Graphic reveals that the most of the assets improve their performance, but the last ones, among which Algorithm 1 selected assets, do not. This situation could be controlled by a dynamic asset allocation strategies, for which at the beginning of each period during the investment horizon, the investor can freely rearrange the portfolio, but it isn’t the aim of this paper.
Table 2

Comparison between optimal and naive portfolio for FF100. Reported return and standard deviation are average values over 40 years (first line) and over groups of 8 years (lines 2–6).

<table>
<thead>
<tr>
<th></th>
<th>Optimal portfolio</th>
<th>Naive portfolio</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Av. ret.</td>
<td>St. dev.</td>
</tr>
<tr>
<td>07/1975 – 06/2015</td>
<td>15%</td>
<td>50%</td>
</tr>
<tr>
<td>07/1975 – 06/1983</td>
<td>19%</td>
<td>54%</td>
</tr>
<tr>
<td>07/1983 – 06/1991</td>
<td>15%</td>
<td>54%</td>
</tr>
<tr>
<td>07/1991 – 06/1999</td>
<td>19%</td>
<td>38%</td>
</tr>
<tr>
<td>07/1999 – 06/2007</td>
<td>13%</td>
<td>47%</td>
</tr>
<tr>
<td>07/2007 – 06/2015</td>
<td>8%</td>
<td>56%</td>
</tr>
</tbody>
</table>

5.3. Test 3: IT72. We here consider a portfolio constructed on real data from Italian market. It considers the monthly returns of 72 equities, from September 2009 to August 2016. Assets are reported in table 3; 25 assets are included in the FTSE MIB index computation. The FTSE MIB is the primary benchmark Index for the Italian equity markets. The Index is comprised of highly liquid, leading companies across different sectors, indeed it captures about the 80% of the domestic market capitalization. The FTSE MIB is computed on 40 Italian equities and seeks to replicate the broad sector weights of the Italian stock market.

Starting from September 2009, we use the $T = 6$–years data to build the optimal port-
folio from September 2015 until August 2016. The conditioning of $R^T R$ is $O(10^9)$. In figure 4 we graphically show the composition of the optimal portfolio we constructed. The optimization strategy allocates the investor wealth on 14 equities, with weights represented as percentage in the figure, among which 5 belong to the FTSE MIB set. The result is obtained in 10 Bregman iterations, with $\tau_+ = 2^{-3}$. We note that the optimal portfolio has return and standard deviation, on annual basis, given by 11% and 34% respectively. The same values for the naive portfolio are $-14\%$ and $60\%$, thus the latter provides a loss to the investor.

6. Conclusions. We have proposed an algorithm, which exploits the Bregman iteration method, for the portfolio selection problem formulated as an $l_1$-regularized
mean-variance model. We introduced an adaptive rule for the estimation of the regularization parameter based on a problem-dependent criterium; the basic idea is to generate an increasing sequence of values and fix it when no short positions occur, in order to realize a trade-off between sparse positive solutions and fidelity data. The algorithm could be slightly modified in order to produce solutions with a limited number of short positions rather than forbidding them. We thus defined a modified

**Fig. 4.** Optimal portfolio on Italian market equities. Built on monthly historical returns of 72 equities from September 2009 to August 2016.
Bregman scheme, in which the regularization parameter takes on varying values in a sequence bounded from above, that preserves the original convergence. Numerical experiments confirm the effectiveness of the proposed algorithm.

We saw in our experiments that sometimes the effectiveness of the optimization strategy can be affected by changes in market conditions. Future work could consider dynamic asset allocation, which involves frequent portfolio adjustments.

Acknowledgments. We would like to acknowledge Dr. Carlo Lanzara for his useful comments about the Italian market test.

REFERENCES


