Journal of Computational Mathematics
Vol.42, No.6, 2024, 1452–1501.

http://www.global-sci.org/jcm
doi:10.4208/jcm.2207-m2021-0349

AN INEXACT PROXIMAL DC ALGORITHM FOR THE LARGE-SCALE CARDINALITY CONSTRAINED MEAN-VARIANCE MODEL IN SPARSE PORTFOLIO SELECTION*

Mingcai Ding, Xiaoliang Song[1) and Bo Yu
School of Mathematical Sciences, Dalian University of Technology, Dalian 116081, China
Emails: Dingmc_dlut@mail.dlut.edu.cn, songxiaoliang@dlut.edu.cn, yubo@dlut.edu.cn

Abstract

Optimization problem of cardinality constrained mean-variance (CCMV) model for sparse portfolio selection is considered. To overcome the difficulties caused by cardinality constraint, an exact penalty approach is employed, then CCMV problem is transferred into a difference-of-convex-functions (DC) problem. By exploiting the DC structure of the gained problem and the superlinear convergence of semismooth Newton (ssN) method, an inexact proximal DC algorithm with sieving strategy based on a majorized ssN method (siPDCA-mssN) is proposed. For solving the inner problems of siPDCA-mssN from dual, the second-order information is wisely incorporated and an efficient mssN method is employed. The global convergence of the sequence generated by siPDCA-mssN is proved. To solve large-scale CCMV problem, a decomposed siPDCA-mssN (DsiPDCA-mssN) is introduced. To demonstrate the efficiency of proposed algorithms, siPDCA-mssN and DsiPDCA-mssN are compared with the penalty proximal alternating linearized minimization method and the CPLEX(12.9) solver by performing numerical experiments on real-word market data and large-scale simulated data. The numerical results demonstrate that siPDCA-mssN and DsiPDCA-mssN outperform the other methods from computation time and optimal value. The out-of-sample experiments results display that the solutions of CCMV model are better than those of other portfolio selection models in terms of Sharp ratio and sparsity.

Mathematics subject classification: 65K05, 90C06, 90C26, 91G80.
Key words: Sparse portfolio selection, Cardinality constrained mean-variance model, Inexact proximal difference-of-convex-functions algorithm, Sieving strategy, Decomposed strategy.

1. Introduction

The classical Markowitz model [39], also called mean-variance (MV) model, was proposed to find the optimal portfolio selection between different assets in a frictionless market. Based on the MV model, researchers have conducted a large number of studies on the out-of-sample performance and sparsity of the portfolios. On one hand, some researches [4, 10, 22, 28, 43] have been carried out to improve the out-of-sample performance of allocation. On the other hand, many researches [5,13,16,18,20,38] focused on constructing new MV models to find sparse portfolios, which can greatly reduce the administrative and transaction costs. One common class of the approaches for obtaining sparse solutions is introducing sparse regularization strategies

* Received December 10, 2021 / Revised version received March 16, 2022 / Accepted July 11, 2022 /
1) Corresponding author

in the MV model [13,18,20]. Another popular class of the methods for getting sparse portfolios is introducing cardinality constraint into the MV model [5,16,38]. The cardinality constrained MV (CCMV) model can be expressed as

$$\begin{aligned} \min_{\boldsymbol{x}\in\mathbb{R}^n} & F(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{x}^\top \mathbf{Q}\boldsymbol{x} \\ \text{s.t. } & \boldsymbol{e}^\top \boldsymbol{x} - 1 = 0, \\ & r - \boldsymbol{R}^\top \boldsymbol{x} \le 0, \\ & \mathbf{0} \le \boldsymbol{x} \le \boldsymbol{b}, \\ & \|\boldsymbol{x}\|_0 \le K, \end{aligned} \tag{1.1}$$

where $\|\boldsymbol{x}\|_0$ denotes the number of the non-zero entries of $\boldsymbol{x} \in \mathbb{R}^n$ and K is the upper bound of the number of investments. $\mathbf{Q}$ is the covariance matrix of the n assets, which is symmetric positive semidefinite. $\boldsymbol{e} \in \mathbb{R}^n$ is the vector with all components one. $\boldsymbol{R} \in \mathbb{R}^n$ is the vector of the expected return of the n assets. $r \in \mathbb{R}$ is the minimum profit target. The box constraint $\mathbf{0} \le \boldsymbol{x} \le \boldsymbol{b}$ means that the short-selling is prohibited and the investment proportion of i-th asset has an upper bound $\boldsymbol{b}_i$.

The CCMV model in (1.1) belongs to the class of the cardinality constrained quadratic optimization problems, which have been widely studied, see [6,7,17,29,44,51]. Due to the combinatorial nature of cardinality function, the cardinality constrained optimization problems are NP-hard in most cases [6]. The common optimization methods for the cardinality constrained problem mainly fall into two main categories: integer programming methods [5,6,19,33,44] and heuristic methods [17,24,38,43]. In the class of integer programming methods, the cardinality constrained problem was reformulated into a mixed integer programming, then the branch-and-bound framework was used to solve it. The heuristic algorithms for solving cardinality constrained problem mainly include genetic algorithm, simulated annealing algorithm and tabu search algorithm. However, both these two kinds of methods are computationally expansive and time consuming, especially when the scale of problem is large.

Recently, by using the alternating iteration method and sparse projection approach, some novel and competitive methods were proposed to solve the cardinality constrained optimization problems. In 2010, Lu *et al.* [36,37] proposed a penalty decomposed (PD) method for solving the cardinality constrained problem, in which, the quadratic penalty subproblems were solved by a block coordinate descent method. By making full use of the advantages of PD method and a proximal alternating linearized minimization (PALM) method [9], Teng *et al.* [47] proposed a penalty proximal alternating linearized minimization (PPALM) method for large-scale sparse portfolio problems. As shown in [47], PPALM method can efficiently find the support set of the local optimal solution of CCMV problem, and PPALM outperforms PD method from computational time and the performance of solutions. However, the parameter associated with the proximal term of PPALM method need to be larger than the Lipschitz constant of penalty subproblem to guarantee its convergence, which would limit its convergence rate severely. In addition, the penalty parameter is difficult to set because the large penalty parameter usually leads to ill-conditioned penalty subproblem.

In addition to the above methods those directly dealing with the cardinality constraint, another common method is to transfer the cardinality constraint problems into the l_0-norm regularization problems. For the l_0-norm regularization optimization problem, many convex/nonconvex relaxation approaches are usually used to approximate the l_0-norm. On one hand, the convex relaxation of the l_0-norm is one of the most commonly used methods in