RESEARCH ARTICLE

The global convergence properties of an adaptive QP-free method without a penalty function or a filter for minimax optimization
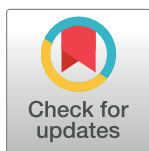
Ke Su‡, Shaohua Liu‡, Wei Lu

Key Laboratory of Machine Learning and Computational Intelligence Baoding, College of Mathematics and Information Science, Hebei University, Baoding, China

☯ These authors contributed equally to this work.
‡ KS and SL also contributed equally to this work.
* lsh6108@126.com

Abstract

In this paper, we proposed an adaptive QP-free method without a penalty function or a filter for minimax optimization. In each iteration, solved two linear systems of equations constructed from Lagrange multipliers and KKT-conditioned NCP functions. Based on the work set, the computational scale is further reduced. Instead of the filter structure, we adopt a nonmonotonic equilibrium mechanism with an adaptive parameter adjusted according to the result of each iteration. Feasibility of the algorithm are given, and the convergence under some assumptions is demonstrated. Numerical results and practical application are reported at the end.

Introduction

The minimax problem with nonlinear inequality constraints has the following form,

$$
\begin{aligned}
\min_{x \in R^n} \quad & F(x) \\
\text{s.t.} \quad & f_i(x) \leq 0, i \in \mathcal{I}_2 = \{l+1, \cdots, m\}
\end{aligned}
\tag{1}
$$

where $F(x) = \max\{f_i(x)\}, i \in \mathcal{I}_1 = \{1, 2, \cdots, l\}$.

The minimax problem is a specific class of nonsmooth optimization problems. The general optimization methods cannot be applied directly in (1) because the objective function is non-differentiable. The common approaches for (1) are gradient sampling methods [1, 2], Cutting Plane Method, and bundle method [3].subgradient methods [4].

Smoothing is one of the most popular classes among all methods for solving nonsmooth problems. There are two main approaches proposed by previous scholars to deal with this problem. First, approximating the non-differential function by a smooth exponential function with parameters (which is also called entropy function). Shor [4] proposed two smoothing algorithms with an active set strategy and a new adaptive parameter update rule. Second, introduce an artificial additional variable to transform the problem into an equivalent nonlinear

programming with smooth constraints, as follows

$$\min \quad t$$

$$\text{s.t.} \quad f_i(x) - t \quad \leq 0, i \in \mathcal{I}_1 = \{1, 2, \cdots, l\} \tag{2}$$

$$f_i(x) \quad \leq 0, i \in \mathcal{I}_2 = \{l+1, \cdots, m\},$$

. where $f_i(x) : \mathbb{R}^n \to \mathbb{R}$ are Lipschitz continuously function; $t \in \mathbb{R}$.

For (2), there are many algorithms can be used such as gradient projection [5], interior point method [6], trust-region [7], sequential quadratic programming(SQP) [8], penalty function [9], filter methods [10] or QP-free method [11], etc.

The extraordinary efficiency of SQP methods in solving nonlinear constrained optimization problems (NLP) has allowed its extension to many other problems, such as minimax problems [8–13]. But the sequences may fail to converge as the initial point lies far from the optimal point in the SQP algorithm. So penalty function methods were proposed by Courant [13] in 1943 to enhance the convergence of the algorithm. The objective function is defined as the sum of the objective function and penalty term in the penalty function method. In [9], Ma gives an exact smooth penalty function method to solve minimax problems with mixed constraints. However, the choice of penalty parameters during the iterative process is complicated. Meantime, the effectiveness of the penalty function method is significantly affected by it.

Fletcher [14] proposed the filter algorithm which can effectively avoid the choice of penalty parameters. It is inspired by the idea of multi-objective programming in 2002, where the objective function and the constraint violation function are considered separately. The combination of the filter and SQP methods has been applied to the minimax problem due to its satisfactory numerical results. [15] gave a trust-region SQP filtering method combining nonmonotonic techniques to solve the unconstrained minimization problem. Luo [10] constructed a new feasible sub-problem based on working sets and incorporated filtering techniques. Although the filter method has good numerical performance, the update of the filter set also faces the problem that the set is getting larger and the computational storage is growing. On the other hand, the feasibility restoration phase is difficult to avoid in the filter method, which more or less increases the computational effort.

To overcome the possible inconsistency in solving the sub-problem and the high computational cost, Panier [16] proposed a QP-free algorithm (SSLE algorithm) for optimization problems with inequality constraints based on the KKT conditions and Newton's method. Each iteration requires solving two systems of linear equations with the same coefficient matrix and a least-squares subproblem. Global and superlinear convergence are established without the assumption of stationary point isolate. In [11], Jian and Ma presented a new QP-free algorithm for minimax problems according to the unique structure of these problems. [17, 18] proposed two QP-free algorithms for solving constrained optimization problems respectively.

Inspired by the above study, a nonmonotonic QP-free algorithm without a penalty function or a filter is given in this paper for the minimax problem. And the global convergence, as well as the superlinear convergence under some mild conditions, is proved. This algorithm combines the NCP function to solve in each iteration two nonlinear systems of equations with the same coefficient matrix, which can be viewed as a Newton-quasi-interaction perturbation of the primal and dyadic variables of the KKT condition. An adjustable operator is introduced, which changes in each iteration according to the results of the previous iteration, thus changing the degree of influence of the objective function in this mechanism. A nonmonotonic mechanism is used to avoid the Maratos effect. The working set is introduced to reduce the computational effort further.

The paper consists of the following parts. Section 1 introduces the previous methods for solving minimax problem. In section 2, the structure of the work is described. Section 3 discusses the implementation of the algorithm. Section 4 discusses the global convergence and superlinear convergence rate of the algorithm. In section 5, numerical results and practical application are given. The article has been summarized in the final.

Description of algorithm

Preliminaries

Define the following notations:

$$X = (x^T, t)^T = (x_1, x_2, \cdots, t)^T;$$

$$\mathcal{I} = \mathcal{I}_1 \cup \mathcal{I}_2 = \{1, \cdots, l, l+1, \cdots, m\};$$

$$W_{\mathcal{I}}(X) = (\nabla h_i(X), i \in \mathcal{I});$$

$$\xi(X) = \max_{i \in \mathcal{I}} \{0, h_i(X)\};$$

$$I(X) = \{h_i(X) = \xi(X), i \in \mathcal{I}\};$$

$$\Phi_I^k = \Phi_I(X^k, \mu^k);$$

$$\omega^k = \omega(X^k).$$

The following function $h_i(X)$ is defined to represent the constraint of (2),

$$h_i(X) = \begin{cases} f_i(x) - t & , i \in \mathcal{I}_1 = \{1, 2, \cdots, l\}; \\ f_i(x) & , i \in \mathcal{I}_2 = \{l+1, \cdots, m\}. \end{cases}$$

where $h_i(X) : \mathbb{R}^{n+1} \to \mathbb{R}$ are Lipschitz continuously functions.

Let $G(X) = (h_1, \cdots, h_l, h_{l+1}, \cdots, h_m)^T : \mathbb{R}^{n+1} \to \mathbb{R}^m$.

(2) are equivalent to

$$\begin{aligned} \min \quad & \omega(X) \\ \text{s.t.} \quad & h_i(X) \leq 0, i \in \mathcal{I}, \end{aligned} \tag{3}$$

where $\omega(X) = t$.

The Lagrangian function is defined as

$$L(X, \mu) = \omega(X) + \sum_{i \in \mathcal{I}} \mu_i h_i(X), \tag{4}$$

where $\mu = (\mu_1, \cdots, \mu_l, \mu_{l+1}, \cdots, \mu_m)^T$ is the multiplier vector.

$(\bar{X}, \bar{\mu})$ is called the KKT point for (3) if the following conditions hold,

$$
\begin{cases}
\nabla L(\bar{X}, \bar{\mu}) = (0, \cdots, 0, 1)^T + \sum_{i \in \mathcal{I}}^{m} \bar{\mu}_i \nabla h_i(\bar{X}) = 0; \\
h_i(\bar{X}) \leq 0, i \in \mathcal{I}; \\
\bar{\mu}_i^T h_i(\bar{X}) = 0, i \in \mathcal{I}; \\
\bar{\mu} \geq 0.
\end{cases} \tag{5}
$$

To construct the system of equations, we introduce the nonlinear complementarity problem (NCP) function, and $\varphi(a, b)$ is called an NCP function if the following relationship holds,

$$
\varphi(a, b) = 0 \Leftrightarrow ab = 0, a \geq 0, b \geq 0. \tag{6}
$$

The NCP function is Lipschitz continuous and differentiable except for the origin. Strong semi-smoothness holds at $(0, 0)$. The Fischer-Burmeister NCP function is a simple NCP function with good theoretical properties and numerical performance.

The Fischer-Burmeister NCP function used in this paper has the following structure:

$$
\varphi_{\mathrm{FB}}(a, b) = \sqrt{a^2 + b^2} - a - b.
$$

So we have

$$
\nabla \varphi_{\mathrm{FB}}(a, b) = \begin{cases}
\left(\dfrac{a}{\sqrt{a^2 + b^2}} - 1, \dfrac{b}{\sqrt{a^2 + b^2}} - 1 \right), (a, b) \neq (0, 0); \\
(\zeta - 1, v - 1) \mid \zeta^2 + v^2 = 1, (a, b) = (0, 0).
\end{cases} \tag{7}
$$

Then the NCP function Φ_i is defined by

$$
\Phi_I = (\phi_1, \cdots, \phi_m)^T,
$$

where

$$
\phi_i(X, \mu) = \varphi_{\mathrm{FB}}(a, b) \big|_{a=-h_i(X), b=\mu_i, i \in \mathcal{I}}.
$$

According to (5), define

$$
\Phi(X, \mu) = (\nabla_X L(X, \mu)^T, \Phi_I^T)^T. \tag{8}
$$

Clearly, KKT condition (5) holds equivalent to $\Phi(X, \mu) = 0$.
Similar to (7), if $(h_i(X), \mu_i) \neq (0, 0)$, then

$$
\nabla_X \phi_i = \left(\frac{-h_i(X)}{\sqrt{(h_i(X))^2 + \mu_i^2}} + 1 \right) \nabla h_i(X), i \in \mathcal{I};
$$

$$
\nabla_\mu \phi_i = \left(\frac{\mu_i}{\sqrt{(h_i(X))^2 + \mu_i^2}} - 1 \right) e_i, i \in \mathcal{I},
$$

where $e_i = (0, \cdots, 0, 1, 0, \cdots, 0)^T$, and

$$\nabla h_i(X) = \begin{cases} \left(\dfrac{\partial f_i}{\partial x_1}, \cdots, \dfrac{\partial f_i}{\partial x_n}, -1\right)^T, i \in \mathcal{I}_1; \\ \left(\dfrac{\partial f_i}{\partial x_1}, \cdots, \dfrac{\partial f_i}{\partial x_n}, 0\right)^T, i \in \mathcal{I}_2. \end{cases}$$

If $(h_i(X), \mu_i) = (0, 0)$, introduce the following notations:

$$\zeta_j^k = \begin{cases} \dfrac{-h_i^k(X)}{\sqrt{\left(h_j^k(X)\right)^2 + \left(\mu_j^k\right)^2}} + 1, (h^k(X), \mu^k) \neq (0, 0); \\ 1 + \dfrac{\sqrt{2}}{2} \qquad\qquad\qquad , else. \end{cases}$$

$$v_j^k = \begin{cases} \dfrac{\mu_i^k}{\sqrt{\left(h_j^k(X)\right)^2 + \left(\mu_j^k\right)^2}} + 1, (h^k(X), \mu^k) \neq (0, 0); \\ -1 + \dfrac{\sqrt{2}}{2} \qquad\qquad\qquad , else. \end{cases}$$

It holds that

$$\nabla_X \phi_i(X, \mu) = (\zeta + 1)\nabla h_i(X)| - 1 \leq \zeta \leq 1;$$

$$\nabla_\mu \phi_i(X, \mu) = (\zeta - 1))| - 1 \leq \zeta \leq 1.$$

Let $G_W^k = (h_j)^T : \mathbb{R}^{n+1} \to \mathbb{R}^m, j \in W$. ∇G_W^k is the Hessian matrix of $G_W^k = \{h_i^k | i \in W\}$ and W is active set.

Define the coefficient matrix V^k according to (8)

$$V^k := \begin{pmatrix} B^k & \nabla G_W^k \\ diag(\zeta_W^k)(\nabla G_W^k)^T & diag(v_W^k - c^k)^T \end{pmatrix},$$

where B^k is a approximation of the Hessian matrix of $L(X, \mu)$, and

$$c_j^k = c \min\{1, \|\Phi^k\|^\tau\}$$

. where $c > 0$ and $\tau > 1$ are given parameters.

In this work, to increase the convergence and flexibility of the algorithm, we substitute the objective function with the following equation:

$$\Theta(X) = \omega(X) + \delta\|\Phi_l\|^2. \tag{9}$$

We introduce such an adjustable operator, which is not a penalty parameter, but is changed in each iteration based on the effect of the iteration results. Fig 1 shows the initial situation of taking $\delta = 1$.
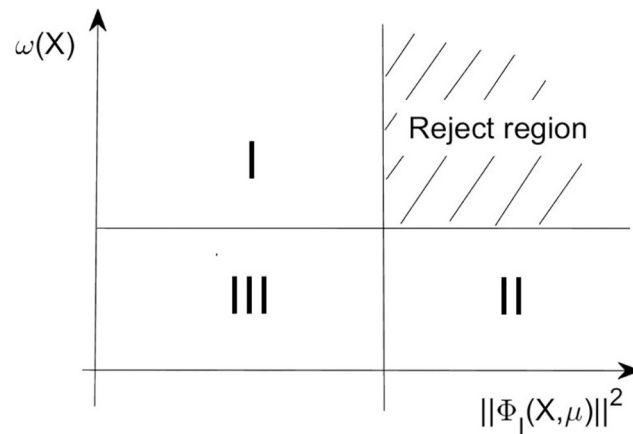
Fig 1. Reject region for iteration points.

We give the following filter equivalence mechanism:

$$\|\Phi_I(\hat{X}^{k+1}, \hat{\mu}^{k+1})\| \leq \theta \max_{0 \leq r \leq q(k)-1} \|\Phi_I^{k-r}\| \tag{10}$$

or both

$$\|\Phi_I(\hat{X}^{k+1}, \hat{\mu}^{k+1})\| \leq \max\left\{\frac{l_k + 1}{2}, \theta_1\right\} h_{max}^k; \tag{11}$$

$$\Theta(\hat{X}^{k+1}) \leq \Theta(X^k) - \alpha^k \theta_2 \|\Phi_I^k\|. \tag{12}$$

There are three regions in the first quadrant. If the trial point X_k is located in region I at the kth iteration, i.e., (10) is satisfied at the current trial point, but (11) and (12) are not satisfied. Then this point is accepted. If the trial point X_k is located in reject region, i.e., the function value and constraint violation are not yielded a satisfactory decrease. So the point is rejected. If the iteration point lies in region II, i.e., (10) is not hold, but (11) and (12) are satisfied. It means that the objective function is improved, but the constraint violation function does not reach the sufficient descent condition, so we need to tighten up our acceptance region (see Fig 2).

So we adjust the parameter δ_k as follows,

$$\delta_{k+1} = \min\left\{\bar{\delta}, \delta_k + \left|\frac{\omega^k - \omega^{k+1}}{\|\Phi_I^k\|^2 - \|\Phi_I^{k+1}\|^2}\right|\right\}, \tag{13}$$

where $\bar{\delta}$ is a constant.

If X_k is located in region III, which means the algorithm makes a good improvement in the objective function and the constraint violation function, So we intend to relax the acceptance criteria and expect further improvements. This means increasing the value of δ_k to make the rejected region narrower. (see Fig 3).
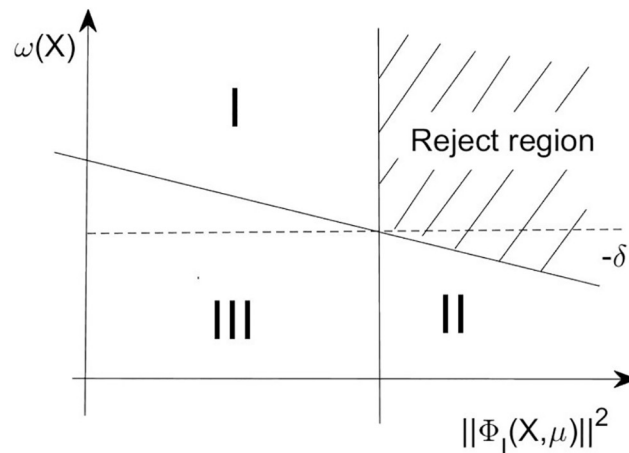
Fig 2. The relaxed reject region for iteration points.
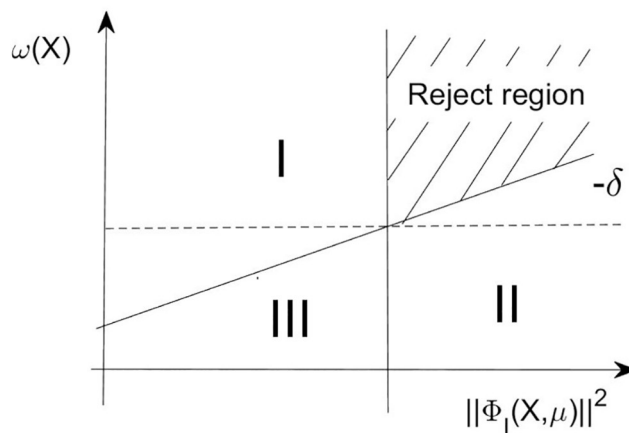
Fig 3. The tightened reject region for iteration points.

Adjust δ_k as follows.

$$\delta_{k+1} = \max\left\{ -\bar{\delta}, \delta_k - \left| \frac{\omega^k - \omega^{k+1}}{\|\Phi_I^k\|^2 - \|\Phi_I^{k+1}\|^2} \right| \right\}. \tag{14}$$

Algorithm A

Based on the above analysis, we give algorithm A for the problem (1).

Step0: (Initialization.)

Choose an initial point

$X^0 \in \mathbb{R}^n; 0 < \beta, \gamma, \gamma_1, \gamma_2, \gamma_3, \theta, \theta_1, \theta_2, \bar{\delta} < 1; h_{max}^0 = 0; l_0 = 1, B^0 = \mathcal{I} \in \mathbb{R}^{n \times n}$.

Step1: (Working set.)

Step1.1 Set $i = 0$, $\epsilon_{k,i} = \epsilon_{k-1}$.

Step1.2 Compute $\mathcal{I}_{k,i} = \{j \in \mathcal{I} \mid -\epsilon_{k,i} \leq h_j(X^k) - \xi(X^k) \leq 0\}$.

Step1.3 If $det\left(W_{\mathcal{I}_{k,i}}^{\mathrm{T}}(X^k)W_{\mathcal{I}_{k,i}}(X^k)\right) \geq \epsilon_{k,i}$, set $\mathcal{I}_k = \mathcal{I}_{k,i}$, $W_k = W_{\mathcal{I}_k}(X^k)$, and $\epsilon_k = \epsilon_{k,i}$, go to Step 2. Otherwise, let $\epsilon_{k,i+1} := \beta\epsilon_{k,i}$, let $i := i+1$, go to Step 1.2.

Step2:(Search directions.)

Step2.1 If $\Phi_I^k \neq 0$, compute $(d^{k0}, \bar{\lambda}^{k0})$ by solving (15) in (d, λ):

$$V^k\begin{pmatrix} d \\ \lambda \end{pmatrix} = \begin{pmatrix} -\nabla\omega^k \\ 0 \end{pmatrix}. \tag{15}$$

Let

$$\lambda_j^{k0} = \begin{cases} 0, j \in \mathcal{I} \setminus \mathcal{I}_k; \\ \lambda_j^{k0} = v_j^k\bar{\lambda}_j^{k0}/(-v_j^k + c_j^k); j \in \mathcal{I}, v_j^k \neq 0; \\ \lambda_j^{k0} = \bar{\lambda}_j^{k0}; j \in \mathcal{I}, v_j^k = 0. \end{cases} \tag{16}$$

Step2.2 Compute $(d^{k1}, \bar{\lambda}^{k1})$ by solving (17) in (d, λ):

$$V^k\begin{pmatrix} d \\ \lambda \end{pmatrix} = \begin{pmatrix} -\nabla L^k \\ -\Phi_I^k \end{pmatrix}. \tag{17}$$

Let

$$\lambda_j^{k1} = \begin{cases} 0, j \in \mathcal{I} \setminus \mathcal{I}_k; \\ \lambda_j^{k1} = v_j^k\bar{\lambda}_j^{k1}/(-v_j^k + c_j^k); j \in \mathcal{I}, v_j^k \neq 0; \\ \lambda_j^{k1} = \bar{\lambda}_j^{k1}; j \in \mathcal{I}, v_j^k = 0. \end{cases} \tag{18}$$

If $\Phi_I^k = 0$, then $d^k = d^{k0}$, $\lambda^k = \lambda^{k0}$; else, if $d^k = 0$, then we let $d^k = d^{k1}$, $\lambda^k = \lambda^{k1}$. If neither is satisfied, then compute ρ^k according to the following definition:

$$\rho_1^k = \begin{cases} 1, \langle\nabla\omega(X^k), d^{k1}\rangle \leq \gamma < \langle\nabla\omega^k, d^{k0}\rangle; \\ (1-\gamma)\dfrac{\langle\nabla\omega^k, d^{k0}\rangle}{\langle\nabla\omega^k, d^{k0} - d^{k1}\rangle}, otherwise. \end{cases}$$

If $-\gamma_1\langle d^{k0}, \nabla f^k\rangle \geq h_{\max}^k$, $\|\Phi_I^k\| \leq \gamma_2 h_{\max}^k$, then $\rho^k = \rho_1^k$, else $\rho^k = \gamma_3$. Denote

$$d^k = (1-\rho^k)d^{k0} + \rho^k d^{k1};$$

$$\lambda^k = (1-\rho^k)\lambda^{k0} + \rho^k\lambda^{k1}.$$

Step3: (Adjustment nonmonotonous line search.)

Step3.1 If

$$\|\Phi(\hat{X}^{k+1}, \hat{\mu}^{k+1})\| \leq \theta\|\Phi(X^k, \mu^k)\|. \tag{19}$$

and (10) or (12) at least one is satisfied, update δ_k by (14), Let

$$\hat{X}^{k+1} = X^k + d^k, \hat{\mu}^{k+1} = \mu^k + \lambda^k. \tag{20}$$

Step3.2 Let M be a non-negative integer, and for each $k \geq 1$, let $q(k)$ satisfy

$$q(0) = 0, 0 \leq q(k) \leq \min\{q(k-1) + 1, M\},$$

Defined

$$\hat{X}^{k+1} = X^k + \alpha^k d^k, \hat{\mu}^{k+1} = \mu^k + \alpha^k \lambda^k. \tag{21}$$

where $\alpha^k = \tau^j$, j is the smallest non-negative integer that satisfying

$$\|\Phi_I(\hat{X}^{k+1}, \hat{\mu}^{k+1})\| \leq \theta \max_{0 \leq r \leq q(k)-1} \|\Phi_I^{k-r}\| \tag{22}$$

or both

$$\|\Phi_I(\hat{X}^{k+1}, \hat{\mu}^{k+1})\| \leq \max\left\{\frac{\iota_k + 1}{2}, \theta_1\right\} h_{max}^k; \tag{23}$$

$$\Theta(\hat{X}^{k+1}) \leq \Theta(X^k) - \alpha^k \theta_2 \|\Phi_I^k\|. \tag{24}$$

Step4: (Update.)
If (10)-(12) hold, update δ_k by (14), and let

$$X^{k+1} = \hat{X}^{k+1}; \mu^{k+1} = \hat{\mu}^{k+1}. \tag{25}$$

Else, If (10) holds, but (11) and (12) are not all satisfied, then update X^k and μ^k by (25). Otherwise, update δ_k by (13), and let

$$X^{k+1} = X^k, \mu^{k+1} = \mu^k. \tag{26}$$

Compute B^{k+1} by BFGS updated formula.
If (10) holds at X^{k+1} but not at X^k, $h_{max}^{k+1} = \Phi_I^k$; else, $h_{max}^{k+1} = h_{max}$.
If (10) holds, update ι_k by

$$\iota_{k+1} = \max_{0 \leq r \leq q(k+1)-1} \|\Phi_I^{k+1-r}\| / \max_{0 \leq r \leq q(k)-1} \|\Phi_I^{k-r}\|.$$

Else, $\iota_{k+1} = \iota_k$. $k = k + 1$, go back to step 1.

Remark 2.1 The Eqs (10) and (12) are composed of Lagrange multipliers and the KKT condition. The solution of the system satisfies the first-order optimality condition of the original problem.

Remark 2.2 For convenience, if (10) holds, then the iterative step is called is called Φ-step; if (11) and (12) hold, it is called Θ-step.

Remark 2.3 Obviously, δ_k is bounded. And the Lagrangian function is Lipschitz continuous.

Implementation of algorithm

Assumption A1 B^k is positive definite and there exists positive numbers m_1 and m_2 such that

$$m_1 \|d\|^2 \leq d^T B^k d \leq m_2 \|d\|^2 \tag{27}$$

for all $d \in \mathbb{R}^n$ and all k.

Assumption A2 $\{X \mid \omega(X) \leq \omega(X^0)\}$ and $\|\mu^k + \lambda^k\|$ are bounded as k is sufficiently large.

Assumption A3 $\omega(X) : \mathbb{R}^{n+1} \to \mathbb{R}$ and $h(X) : \mathbb{R}^{n+1} \times \Omega \to \mathbb{R}$ are Lipschitz continuously differentiable. For all $a, b \in \mathbb{R}^{n+m}$,

$$\|\nabla\Omega(a) - \nabla\Omega(b)\| \leq m_0 \|a - b\|, \|\Phi(a) - \Phi(b)\| \leq m_0 \|a - b\|,$$

where $m_0 > 0$ is the Lipschitz constant.

Assumption A4 The Mangasarian-Fromovitz qualification condition (MFCQ) is satisfied at X_i^*, $i \in W(X)$.

Assumption A5 The sequence of $\{B^k\}$ satisfies

$$\frac{\|(B^k - \nabla_X^2 L(X^k, \mu^k))d^{k1}\|}{\|d^{k1}\|} \to 0.$$

Assumption A6 The strict complementarity condition holds at (X^*, μ^*).

Remark 3.1 It follows from A3 that the Lagrangian function (4) is Lipschitz continuous.

The following lemmas show the algorithm is well defined.

Lemma 3.1 Step 1 is finitely terminated.

proof Assume that the conclusion is not valid, then step 1 will run an infinite number of times.

$$\det\left[W_{J_{k,i} \cup L}^{\mathrm{T}} W_{J_{k,i} \cup L}\right] < \frac{1}{2^i}\epsilon_0. \tag{28}$$

From the definition of W_k, we know that $W_{k,i+1} \subseteq W_{k,i}$. As i is large enough, $W_{k,i} = W_{k,i+1}$ marked as W_k^*.

Then we have $\det\left[W_{J_{k,i} \cup L}^{\mathrm{T}} W_{J_{k,i} \cup L}\right] = 0$ and $W_k^* = I(X^k)$.

That is in contradiction to A4.

Lemma 3.2 If $\Phi^k \neq 0$, V^k is nonsingular for all k.

proof Let $V^k(u^k, v^k) = 0$, where $u = (u_1, \ldots, u_m)^T, v = (v_1, \ldots, v_m)^T$. From (17), we have

$$B^k u + \nabla G_W^k v = 0 \tag{29}$$

$$\mathrm{diag}(\zeta_W^k)(\nabla G^k)^T u + \mathrm{diag}(v_W^k)^T v = 0. \tag{30}$$

Due to $\Phi^k \neq 0$, there has $\zeta_{W,j}^k \neq 0$ and $v_{W,j}^k \neq 0$ for any j by their definitions.

$$v = -(\mathrm{diag}(v_W^k)^T)^{-1}(\mathrm{diag}(\zeta_W^k)(\nabla G^k)^T u. \tag{31}$$

Taking (31) to (29), then pre multiplying (29) by u^T, we get

$$u^T B^k u + u^k(\nabla G^k)^T)(-(\mathrm{diag}(v_W^k)^T)^{-1}(\mathrm{diag}(\zeta_W^k)(\nabla G^k)^T u) = 0,$$

where B^k is positive definite, and $(\nabla G^k)^T)(-(\mathrm{diag}(v_W^k)^T)^{-1}(\mathrm{diag}(\zeta_W^k)(\nabla G^k)^T$ is semi-definite.

So we got $u = 0$. Submitting $u = 0$ to (31), then $v = 0$. Since $(u, v) = 0$ is the unique solution of $V^k(u^T, v^T)^T = 0$, V^k is nonsingular, the conclusion holds.

Lemma 3.3 If $d^{k0} \neq 0$,then$>$

$$(d^{k0})^T B^k d^{k0} \leq -(d^{k0})^T \nabla \omega^k,$$

where B^k is an approximation of the Hessian matrix of $L(X, \mu)$; $\nabla \omega^k$ is the gradient vector of ω (X^k, μ^k).

proof From (15), we have

$$B^k d^{k0} + \nabla G_W^k \lambda^{k0} = -\nabla \omega^k; \tag{32}$$

$$\mathrm{diag}(\zeta_W^k)(\nabla G_W^k)^T d^{k0} + \mathrm{diag}(v_W^k - c^k)\hat{\lambda}^{k0} = 0. \tag{33}$$

Then

$$\bar{\lambda}^{k0} = -(\mathrm{diag}(v_W^k - c^k))^{-1}\mathrm{diag}(\zeta_W^k)(\nabla G_W^k)^T d^{k0}. \tag{34}$$

Taking (34) to (32), then

$$(d^{k0})^T(B^k d^{k0} + \nabla G_W^k \lambda^{k0})$$

$$= (d^{k0})^T B^k d^{k0} - (d^{k0})^T \nabla G_W^k \mathrm{diag}(\zeta_W^k)(\mathrm{diag}(v_W^k - c^k))^{-1}(\nabla G_W^k)^T d^{k0}.$$

$$= -(d^{k0})^T \nabla \omega^k.$$

The matrix $-\nabla G_W^k \mathrm{diag}(\zeta_W^k)(\mathrm{diag}(v_W^k - c^k))^{-1}(\nabla G_W^k)^T d^{k0}$ is positive semi-definite, so we have

$$(d^{k0})^T B^k d^{k0} \le -(d^{k0})^T \nabla \omega^k.$$

$$\begin{aligned}
(d^k)^T \nabla \omega^k &= (1 - \rho^k)(d^{k0})^T \nabla \omega^k + \rho^k(d^{k1})^T \nabla \omega^k \\
&= (d^{k0})^T \nabla \omega^k \left[1 - (1 - \gamma)\frac{\langle \nabla \omega^k, d^{k0} \rangle}{\langle \nabla \omega^k, d^{k0} - d^{k1} \rangle} + (1 - \gamma)\frac{\langle \nabla \omega^k, d^{k0} \rangle}{\langle \nabla \omega^k, d^{k0} - d^{k1} \rangle} \right] \\
&= \theta(d^{k0})^T \nabla \omega^k \\
&\le -\theta(d^{k0})^T B^k d^{k0}.
\end{aligned}$$

Hence the conclusion holds.

Lemma 3.4 For any $0 < \alpha \le 1$, there exists $t_1 > 0$ such that

$$\|\Phi_I(X^k + \alpha d^{k0}, \mu^k + \alpha \lambda^{k0})\|^2 - \|\Phi_I^k\|^2 \le t_1 \alpha^2 \|(d^{k0}, \lambda^{k0})\|^2 \tag{35}$$

proof If $\Phi_I^k = 0$, then from (31), there exists $t_1 > 0$ such that for any $0 < \alpha \le 1$,

$$\begin{aligned}
\|\Phi_I(X^k + \alpha d^{k0}, \mu^k + \alpha \lambda^{k0})\|^2 &= \|\Phi_I(X^k + \alpha d^{k0}, \mu^k + \alpha \lambda^{k0}) - \Phi_I^k\|^2 \\
&\le \alpha^2 m_2^2 \|(d^{k0}, \lambda^{k0})\|^2.
\end{aligned}$$

Therefore the conclusion holds for $\Phi_I^k = 0$.

If $\Phi_I^k = 0$, then $\varphi_i(0, 0) = 0$ indicates

$$\begin{aligned}
\|\Phi_I^k + \alpha(\mathrm{diag}(\zeta^k)(\nabla G_W^k)^T d^{k0} + \mathrm{diag}(v^k)\lambda^k)\|^2 &= \|\Phi_I^k\|^2 \\
&\quad + \alpha^2 \|\mathrm{diag}(\zeta)^k)(\nabla G_W^k)^T d^k + \mathrm{diag}(v)^k)\lambda^k\|^2,
\end{aligned} \tag{36}$$

where $\mathrm{diag}(\zeta^k)$ and $\mathrm{diag}(v^k)$ are the diagonal matrices with diagonal elements ζ_j^{k0} and $(v_j^k - cj^k)$.

So we have

$$\|\Phi_I(X^k + \alpha d^{k0}, \mu^k + \alpha \lambda^{k0})\|^2 = \|\Phi_I^k\|^2 + O(\alpha^2). \tag{37}$$

Hence the conclusion holds.

Lemma 3.5 If $\Phi_I^k \neq 0$, for any $\varepsilon > 0$, there exists $\bar{\alpha} > 0$, such that for any $0 < \alpha \leq \bar{\alpha}$,

$$\left\|\Phi_I^k\right\|^2 - \left\|\Phi_I(X^k + \alpha d^{k1}, \mu^k + \alpha \lambda^{k1})\right\|^2 \geq (2 - \varepsilon)\alpha\left\|\Phi_I^k\right\|^2. \tag{38}$$

proof If $\Phi_I^k \neq 0$, by (SSLE) we have

$$\text{diag}(\zeta^k)(\nabla G_W^k)^T d^{k1} + \text{diag}(v^k - c^k)\lambda^{k1} = -\Phi_I^k. \tag{39}$$

For any $i \neq 0$,

$$\phi_i(X^k + \alpha d^{k1}, \mu^k + \alpha \lambda^{k1}) = \phi_i^k + \alpha(\text{diag}(\zeta^k)(\nabla G_W^k)^T, \text{diag}(v^k - c^k)\lambda^{k1}) + O(\alpha^2). \tag{40}$$

By (36) and (40)

$$\left\|\Phi_I(X^k + \alpha d^{k0}, \mu^k + \alpha \lambda^{k0})\right\|^2$$

$$= \left\|\Phi_I^k\right\|^2 + \alpha^2 \left\|\text{diag}(\zeta^k)(\nabla G_W^k)^T d^{k1} + \text{diag}(v^k - c^k)\lambda^{k1}\right\|^2 \tag{41}$$

$$+ 2\alpha(\Phi_I^k)^T\left(\text{diag}(\zeta^k)(\nabla G^k)^T d^{k1} + \text{diag}(v^k - c^k)\lambda^{k1}\right) + O(\alpha^2).$$

Since $c_i^k \neq 0$, by the definitions of c_i^k and v_i^k and $\phi_i^k = 0$, the following equation holds

$$\left\|\Phi_I^k + \alpha\left(\text{diag}(\zeta^k)(\nabla G_W^k)^T d^{k1} + \text{diag}(v^k - c^k)\lambda^{k1}\right)\right\|^2 = (1 - 2\alpha)\left\|\Phi_I^k\right\|^2$$

$$+ \alpha^2 \left\|\text{diag}(\zeta^k)(\nabla G_W^k)^T d^{k1} + \text{diag}(v^k - c^k)\lambda^{k1}\right\|^2. \tag{42}$$

By (41) and (42), given any $\varepsilon > 0$, there is $\bar{\alpha} > 0$ such that, for any $0 < \alpha \leq \bar{\alpha}$,

$$\left\|\Phi_I^k\right\|^2 - \left\|\Phi_I(X^k + \alpha d^{k1}, \mu^k + \alpha \lambda^{k1})\right\|^2 \geq (2 - \varepsilon)\alpha\left\|\Phi_I^k\right\|^2. \tag{43}$$

Hence the conclusion holds.

From Lemmas 3.4–3.5, we can obtain the following Lemma 3.6.

Lemma 3.6 If $\Phi_I^k \neq 0$, then given any $\varepsilon > 0$, there exists $\bar{\alpha} > 0$ such that, for any $0 < \alpha \leq \bar{\alpha}$,

$$\left\|\Phi_I^k\right\|^2 - \left\|\Phi_I(X^k + \alpha d^k, \mu^k + \alpha \lambda^k)\right\|^2 \geq \rho^k \alpha\left\|\Phi_I^k\right\|^2. \tag{44}$$

Theorem 3.1 If Algorithm A does not terminate at X^k, then there exits an $\alpha_{min} > 0$, such to $\alpha_k \geq \alpha_{min} > 0$, we have either (10) holds, or both (11) and (12) hold.

proof If $\rho^k = \gamma_3$, for all k, $\Phi_I^k \neq 0$ and any $\alpha \leq \min\{(1 - \theta)/\gamma_3, \bar{\alpha}\}$, there have

$$\left\|\Phi_I(X^k + \alpha d^k, \mu^k + \alpha \lambda^k)\right\|^2 \leq (1 - \rho^k \alpha)\|\Phi^k\| \leq \theta\|\Phi_1^k\|^2 \leq \theta \max_{0 \leq r \leq m(k)-1}\|\Phi_I^{k-r}\|.$$

So (10) holds.

If $\rho^k = \rho_1^k$, by the definition of ρ^k, $-\gamma_1 (d^{k0})^T \nabla \omega^k \geq h_{max}^k$, $\|\Phi_I^k\| \leq \theta h_{max}^k$. From Lemma 6, define $\bar{\delta} > \dfrac{1}{\gamma_3} - \theta_2$, for all sufficiently large k, we have

$$
\begin{aligned}
\Theta(X^k) - \Theta(X^{k+1}) &\geq -\alpha^k (d^k)^T \nabla \omega^k + \delta_k \|\Phi_I^k\| - \delta_{k+1} \|\Phi_I^{k+1}\| \\
&\geq -\alpha^k \gamma_1 (d^{k0})^T \nabla \omega^k - \bar{\delta}(\|\Phi_I^k\| - \|\Phi_I^{k+1}\|) \\
&\geq \alpha^k h_{max}^k - \bar{\delta}\alpha\rho^k \|\Phi_I^k\| \\
&\geq \alpha^k \theta_2 \|\Phi_I^k\|,
\end{aligned}
\tag{45}
$$

and

$$
\|\Phi_I(X^k + \alpha d^k, \mu^k + \alpha\lambda^k)\|^2 \leq \theta_1 \|\Phi_1^k\|^2 \leq \gamma_1 h_{max}^k.
$$

The proof is complete.

Convergence of algorithm

In this section, the global and superlinear convergence rates of this algorithm have been discussed.

If $\Phi^k = 0$, then (X^k, μ^k) is KKT point. From Lemma 3.3, if $d^{k0} \neq 0$, then d^k is the decreasing direction of ω^k, (d^k, λ^k) is the decreasing direction of $\|\Phi_I^k\|$. If $\Phi_I^k = 0$ and $d^{k0} = 0$, (X^k, μ^k) satisfies the KKT condition. In the following, without loss of generality, it is assumed that the algorithm does not terminate finitely.

Lemma 4.1 Assume A1-A4 hold. $\Phi_I(X^k, \mu^k) \to 0$, as k is large enough.

proof Case 1. If (10) holds for all k sufficiently large, one has

$$
\begin{aligned}
\|\Phi_I^{l(k+1)}\| &\leq \max_{0 \leq r \leq q(k)} \|\Phi_I^{k+1-r}\| \\
&= \max\{\Phi_I^{k+1}, \Phi_I^{l(k)}\} \leq \|\Phi_I^{l(k)}\|.
\end{aligned}
$$

The sequence $\|\Phi_I^{l(k)}\|$ is monotone decreasing, which implies it is convergent. From (10), we obtain

$$
\|\Phi_I(X^{k+1}, \mu^{k+1})\| \leq \theta_1 \|\Phi_I^{l(k)}\|.
$$

So $\{\|\Phi_I(X^{k+1}, \mu^{k+1})\|\}$ is convergent. Together with $\theta_1 \in (0, 1)$, we have

$$
\|\Phi_I(X^{k+1}, \mu^{k+1})\| \leq \|\Phi_I^{l(k)}\| \to 0 (k \to \infty),
$$

which illustrates $\Phi_I(X^k, \mu^k) \to 0$.

case 2. If (11) and (43) hold for all k sufficiently large, we prove the conclusion by contradiction. Then there exists $\epsilon_1 > 0$, such that

$$
\|\Phi_I(X^{k+1}, \mu^{k+1})\| \geq \varepsilon > 0,
$$

$$
h_{max}^k \geq \varepsilon > 0.
$$

and

$$
\Theta(X^k) - \Theta(X^{k+1}) = \omega^k + \delta_k \|\Phi_I^k\|^2 - (\omega^{k+1} + \delta_{k+1} \|\Phi_I^{k+1}\|^2).
\tag{46}
$$

By (12), (43) and the definition of θ and $\bar{\delta}$, we have

$$\omega^k - \omega^{k+1} \geq \alpha^k \theta_2 \|\Phi_I^k\| - \bar{\delta}(\|\Phi_I^k\|^2 - \|\Phi_I^{k+1}\|^2)$$

$$\geq \alpha^k \theta_2 \epsilon_1 - \bar{\delta}\epsilon_1 \alpha^k > 0.$$

Because $\{\omega^k\}$ is monotonically decreasing. Then $\omega^k \to -\infty$ as $k \to +\infty$ which is contradictory to the hypothesis.

Case 3. If the Φ-step and Θ-step iterative appear alternately.

According to Remark 2.2, the iterations from k_t to k_{t+1} and $k_{t+2} + 1$ to k_{t+3} are Θ-steps. And the iteration from $k_{t+1} + 1$ to k_{t+2} iteration are Φ-steps.

$h_{\max}^k$ is updated only when the transition from the Θ-step to the Φ-step, which indicates that the constraint violation is still large enough to decrease. From step 4, we know if $h_{\max}^{k+1}$ is updated, then

$$h_{\max}^{k+1} = \theta_1 \|\Phi_I^{l(k+1)}\| \leq \theta_1 \|\Phi_I^{l(k+1)}\| = \theta_1 h_{\max}^{l(k)} \leq h_{\max}^k.$$

Thus $h_{\max}^k(X)$ is monotonous descent and $\|\Phi_I^k\| \leq \|\Phi_I^{k+1}\| \leq h_{\max}^k$. Meantime, we have

$$h_{\max}^{k_t} = h_{\max}^{k_t+1} = \cdots, = h_{\max}^{k_{t+1}};$$

$$h_{\max}^{k_{t+1}+1} = \max_{0 \leq r \leq q(k_{t+1}+1-r)-1} \|\Phi_I^{k_{t+1}+1}\| = h_{\max}^{k_{t+1}+2} = \cdots = h_{\max}^{k_{t+2}} = \cdots = h_{\max}^{k_{t+3}}.$$

Donate $\max\{(r_k + 1)/2, \theta_1\} = \bar{\theta}$, then we have

$$h_{\max}^{k_{t+1}+1} \leq \max_{0 \leq r \leq q(k_{t+2}-r)-1} \bar{\theta} h_{\max}^{k_{t+1}+1-r} \leq \theta_1 h_{\max}^{k_{t+1}}.$$

The sequence $\left\{ \cdots, h_{\max}^{k_t+1}, \cdots, h_{\max}^{k_{t+2}}, \cdots, h_{\max}^{k_{t+3}}, \cdots \right\}$ is monotonous descent for all $\alpha \geq 0$, and

$$h_{\max}^{k_{t+1}+1}/h_{\max}^{k_t+1} = \bar{\theta} < 1.$$

So $h_{\max}^{k_{t+1}+1} \to 0, t \to \infty$. Considering the non-negativity of $h_{\max}^k$, we have $h_{\max}^k \to 0, k \to \infty$. Along with (19), $\|\Phi_I^k\| \leq \max\{\theta_1, \bar{\theta}\}h_{\max}^k$. Therefore $\lim_{k\to\infty}\|\Phi_1^k\| = 0$.

Lemma 4.2 Assume A1-A4 hold. If k is large enough, then $d^{k0} \to \mathbf{0}, d^k \to \mathbf{0}$.

The proof is similar to Lemma 5.4 in [20].

With Lemma 4.1 and 4.2, we can obtain the global convergence of the algorithm.

Theorem 4.1 The accumulation point (X^*, μ^*) of the sequence $\{(X^k, \mu^k)\}$ is the KKT point pair of problem (15).

The convergence rate of the algorithm is discussed next. To make the algorithm converge superlinear, assumptions 5–6 are added.

Remark 4.1 A5 illustrates (X^k, μ^k) is a Newton direction with a high order perturbation. A6 shows that Φ is continuously differentiable at each KKT point (X^*, μ^*).

Lemma 4.3 The above assumptions hold, then $\{\|(V^k)^{-1}\|\}$ and $\{\|(\hat{V}^k)^{-1}\|\}$ are uniformly bounded. The accumulation matrix V^* of $\{V^k\}$ is nonsingular.

The proof is similar to Lemma 5.5 in [19].

Lemma 4.3 implies that the Theorem 4.2 holds.

Theorem 4.2 Assume A1-A6 hold, the algorithm is superlinearly convergent, i.e., (X^k, μ^k) converges to (X^*, μ^*) superlinearly.

Numerical results

In this section, the numerical results are shown. Problem 1 the section is taken from [19]. The BFGS formula proposed by Broyden et al. is used to update B^{k+1} as [20].

Tables 1 and 2 show the numerical experimental results of two test questions, where NIT is the number of iterations, NF is the number of times the objective and constraint functions are computed, and NG is the number of times the gradient is computed. and TIME(Unit in seconds) is the CPU runtime. The numerical experiments were computed by Matlab R2020 on a computer with 16.0GB RAM and Intel 11th Gen Intel(R) Core(TM) i5–11320H @ 3.20GHz.

The parameters involved in the algorithm are chosen as follows: $B^0 = I \in \mathbb{R}^{n \times n}, \tau = 2.5$, $c = 0.2, \bar{\delta} = 2, \gamma = 0.5, \gamma_1 = 0.1, \gamma_2 = 0.95, \gamma_3 = 0.5, \theta = 0.9, \theta_1 = 0.95, \theta_2 = 0.1, h_{max}^0 = 0, \iota_0 = 1, \mu = [1, 0, \ldots, 0, 1]$.

Test problem 1 [21].

$$Figure1 \quad \min \quad F(x)$$

$$s.t. \quad f_i(x) = \sum_{i=1}^{n-1} (x_i^2 + x_{i+1}^2 + x_i x_{i+1} - 1),$$

where

$$F(x) = \max\{f_i(x) = x_i^2, i = 1, \ldots, n\}.$$

Test problem 2 [21].

$$\min \quad F(x)$$

$$s.t. \quad f_i(x) = (3.0 - 0.5x_{i+1})x_{i+1} - x_i - 2.0x_{i+2} + 1, \quad i \in [1, n-2],$$

where

$$F(x) = \max\{f_i(x) = x_i^2, i = 1, \ldots, n\}.$$

Table 1. Test Problem 1.

n	*NIT*	*NF*	*NG*	*TIME(s)*
2	2	2	2	0.045
3	9	9	16	0.079
4	8	71	78	0.107
5	9	88	96	0.109
6	9	88	96	0.113
7	8	103	110	0.127
8	9	92	100	0.131

https://doi.org/10.1371/journal.pone.0274497.t001

Table 2. Test Problem 2.

n	*NIT*	*NF*	*NG*	*TIME(s)*
3	7	38	44	0.083
4	4	19	22	0.065
5	13	172	184	0.173

https://doi.org/10.1371/journal.pone.0274497.t002

Then we discuss the application of algorithms to investment portfolios. In the investment problem proposed by Markowitz, there are two objective functions to be considered. One is to maximize the return of the portfolio, and the other is to reduce the risk. In the traditional model, the latter is to minimize the risk (variance) of a set of feasible portfolios for a given level of expected returns. By varying the expected return level as the two objectives for a set of non-dominant portfolios, the efficient frontier on returns is determined by the variance and average of the yields in the Markowitz model. Investors can get a suitable portfolio by analyzing the expected investment and return.

In the traditional Markowitz mean-variance model, it is assumed that the investor has some wealth and is ready to invest in a set of securities, which is recorded as a set P. R_k represents the return value of each security k, which is a random variable. The mean value of R_k can be calculated from historical data. Define the expected return of a security as $\mu_k = E(R_k)$, $k = 1, \cdots, P$. x_k is the proportion allocated to a certain security. The weight vector x_k needs to satisfy the following constraints,

$$\sum_{k=1}^{P} x_k e^T x = 1$$

where $e = \{1, \cdots, 1\}^T \in R^P$.

The expected returns of the portfolio are as follows:

$$E(R) = E(x_k R_1 + \cdots + x_p R_P) = x_1 \mu_1 + \cdots x_p \mu_p = \mu^T x$$

where $\mu = (\mu_1, \cdots, \mu_P)^T$.

The variance of the portfolio is

$$v(R) = E\left(\left[\sum_{k=1}^{P} x_k R_k - E\left(\sum_{k=1}^{P} x_k R_k\right)\right]^2\right)$$
$$= \sum_{k=1}^{P}\sum_{j=1}^{P} E[(R_k - \mu_k)(R_j - \mu_j)] x_k x_j$$

where covariance matrix $Q = \{Q_{i,j}\}$ is a symmetric positive semi-definite matrix with uncertain information. Assume that the short-term investment value is uncertain, but there are obviously $x_k \geq 0$, $k = 1, \cdots, n$. we let B_k and A_k be the upper and lower bounds for x_k, that is, $B_k \leq x_k \leq A_k$, $k = 1, \cdots, P$.

According to the above definition, the base/mean-variance dual-objective optimization problem [22] can be described as

$$\begin{aligned}
\min_{x \in R^P} \quad & -\frac{\mu^T x}{x^T Q x} \\
\min_{x \in R^P} \quad & T(x) \\
\text{s.t.} \quad & e^T x = 1, \\
& B_k \leq x_k \leq A_k, k = 1, \cdots, P,
\end{aligned}$$

where $T(x) = \|x\|_0 = \sum_{k=1}^{P} \text{sign}(|x_k|)$, and

$$\text{sign} = \begin{cases} -1, & x_k < 0, \\ 0, & x_k = 0, \\ 1, & x_k > 0. \end{cases}$$

Note that $sign(\|x_k\|)$ is discontinuous, so we introduce the following approximation function which is locally lipschitz continuity,

$$w_k(x_k, \delta) = \begin{cases} \dfrac{\delta}{x_k - 1 + \delta} + \delta + 1, & x_k \leq -\delta, \\ \dfrac{|x_k|}{\delta}, & |x_k| \leq; \delta, \\ \dfrac{-\delta}{x_k - 1 + \delta} + \delta + 1, & x_k \geq \delta, \end{cases} \tag{47}$$

where $\delta > 0$. For any $x_k \in R$, $\lim\limits_{\delta \to 0^+} \omega(x_k, \delta) = sign(\|x_k\|)$.

Let $y = (x, y_{P+1})^T \in R^{P+1}$, we get the continuous approximation problem as follows:

$$\min_{y \in R^{P+1}} \mathbf{w}(y) = (w_1(x_1, \delta), \cdots, w_P(x_P, \delta), y_{P+1})^T$$

$$\text{s.t.} \quad -\frac{\mu^T x}{x^T Q x}, \tag{48}$$

$$e^T x = 1,$$

$$B_k \leq x_k \leq A_k, k = 1, \cdots, P.$$

where $\mathbf{w}(y) = (\omega_1(x_1, \delta), \cdots, \omega(x_p, \delta), y_{P+1})^T$. (48) can be regarded as a minimax problem, which can be solved using the algorithm 1. We take the case of $P = 2$, the optimal portfolio obtained is [0.5;0.5]. This result illustrates that under the condition that the risk and return of each stock are equal, the proportion of each stock in the optimal strategy is determined by the investor's investment willingness.

Conclusion

In this paper, we give the properties of global convergence and the global convergence of an adaptive QP-free method for solving the minimization problem. Combining the NCP function and the multiplier, in each iteration two systems of linear equations with the same coefficient matrix are solved, which can be viewed as a perturbation of the primal Variables and dual variables of the KKT condition by the Newton-quasi interaction. A new filter substitution mechanism is given, which retains the advantages of the filter method, avoids the selection of penalty parameters, and eliminates potential storage problems that may arise from the filter. And the objective function is tuned by introducing a flexible operator. A non-monotonic mechanism is used to avoid the Maratos effect to some extent, and the introduction of working set further reduces the workload. The effectiveness and convergence of the intensity algorithm are demonstrated under the assumption of no stability point isolation.

Author Contributions

Conceptualization: Ke Su, Shaohua Liu.

Software: Wei Lu.

Writing – original draft: Ke Su, Shaohua Liu.

References

1. Hare W., Nutini J. A derivative-free approximate gradient sampling algorithm for finite minimax problems[J]. Comput Optim Appl, 2013, 56: 1–38. https://doi.org/10.1007/s10589-013-9547-6

2. Burke J.V., Curtis F.E., Lewis A.S., Overton M.L., Sim?es L.E.A. Gradient Sampling Methods for Nonsmooth Optimization[M]. Numerical Nonsmooth Optimization, 2020: 201–225.

3. Antonio F., Manlio G., Giovanni G., Giovanna M., A partially inexact bundle method for convex semi-infinite minmax problems[J]. Communications in Nonlinear Science and Numerical Simulation, 2015, 21 (1-3): 172–180. https://doi.org/10.1016/j.cnsns.2014.07.033

4. Shor N.Z. The Subgradient Method. In: Minimization Methods for Non-Differentiable Functions[J]. Springer Series in Computational Mathematics, 1985, 3: 22–47. https://doi.org/10.1007/978-3-642-82118-9_3

5. Ma G, Zhang Y, Liu M. A generalized gradient projection method based on a new working set for minimax optimization problems with inequality constraints[J], J Inequal Appl, 2017, 51: 1–14. https://doi.org/10.1186/s13660-017-1321-3 PMID: 28298875

6. Obasanjo E., Tzallas-Regas G. Rustem B. An Interior-Point Algorithm for Nonlinear Minimax Problems [J]. J Optim Theory Appl, 2010, 144: 291–318. https://doi.org/10.1007/s10957-009-9599-z

7. EL-Sobky B.,Aboutahoun A.W. An active-set algorithm and a trust-region approach in constrained minimax problem[J]. Comp. Appl. Math, 2018, 37: 2605–2631. https://doi.org/10.1007/s40314-017-0468-3

8. Jian J, Mo X, Qiu L et al. Simple Sequential Quadratically Constrained Quadratic Programming Feasible Algorithm with Active Identification Sets for Constrained Minimax Problems[J]. J Optim Theory Appl, 2014, 160: 158–188. https://doi.org/10.1007/s10957-013-0339-z

9. Ma C, Lin X, Yao J, et al. New exact penalty function for solving constrained finite min-max problems[J]. Applied Mathematics and Mechanics, 2012, 33(2): 253–270. https://doi.org/10.1007/s10483-012-1548-6

10. Luo Z, Wang Z, Li R. Improved Filter-SQP Algorithm with Active Set for Constrained Minimax Problems [J]. Journal of Applied Mathematics. 2014, 2: 1–7.

11. Jian J, Ma G. A Globally Convergent QP-Free Algorithm for Inequality Constrained Minimax Optimization[J]. Acta Mathematica Scientia, 2020, 40(6): 1723–1738. https://doi.org/10.1007/s10473-020-0608-5

12. HU Q, HU J. A sequential quadratic programming algorithm for nonliear minimax problems[J]. Bulletin of the Australian Mathematical Society[J], 2007, 76(3): 353–368. https://doi.org/10.1017/S0004972700039745

13. Courant R. Variational methods for the solution of problems of equilibrium and vibrations[J]. Bull. Amer. Math. Soc. 1943, 49, 1–23. https://doi.org/10.1090/S0002-9904-1943-07818-4

14. Fletcher R. and Leyyfer S.. Nonlinear programming without a penalty function[J]. Mathemacal Programming, 2002, 91: 239–269.

15. Zhao Q, Guo N. A Nonmonotone Filter Method for Minimax Problems[J]. Applied Mathematics, 2011, 2: 1372–1377. https://doi.org/10.4236/am.2011.211193

16. Panier ER, Tits AL, Herskovits JN: A QP-free, globally convergent, locally superlinearly convergent algorithm for inequality constrained optimization[J]. SIAM J. Control Optim, 1988, 26: 788–811. https://doi.org/10.1137/0326046

17. Pu D, Liu A, Shang Y, et al. QP-free infeasible method without a penalty function and a filter [J]. Operations Research Transactions, 2013, 17(1): 25–27.

18. Shang Y, Jin Z, Pu D. A new filter QP-free method for the nonlinear inequality constrained optimization problem[J]. J Inequal Appl, 2018, 287: 1–14. https://doi.org/10.1186/s13660-018-1851-3 PMID: 30839771

19. Chen L, Wang Y, He G. A feasible active set QP-free method for nonlinear programming[J]. SIAM J. Optim, 2006, 17: 401–429. https://doi.org/10.1137/040605904

20. Yang Y, Li D, Qi L: A feasible sequential linear equation method for inequality constrained optimization [J]. SIAM J. Optim, 2003, 13: 1222–1244. https://doi.org/10.1137/S1052623401383881

21. Karmitsa, N. Test problems for large-scale nonsmooth minimization. Reports of the Department of Mathematical Information Technology, Series B.

22. Brito R. P., Vicente N. L. Efficient cardinality/mean-variance portfolios[J]. System Modeling and Optimization, 2014, 443: 52–73. https://doi.org/10.1007/978-3-662-45504-3_6