J. Appl. Math. & Informatics Vol. **27**(*2009*), *No. 5 - 6, pp. 981 - 1000*
Website: http://www.kcam.biz

SOLVING NONLINEAR ASSET LIABILITY MANAGEMENT PROBLEMS WITH A PRIMAL-DUAL INTERIOR POINT NONMONOTONE TRUST REGION METHOD[1]

NENGZHU GU* AND YAN ZHAO

ABSTRACT. This paper considers asset liability management problems when their deterministic equivalent formulations are general nonlinear optimization problems. The presented approach uses a nonmonotone trust region strategy for solving a sequence of unconstrained subproblems parameterized by a scalar parameter. The objective function of each unconstrained subproblem is an augmented penalty-barrier function that involves both primal and dual variables. Each subproblem is solved approximately. The algorithm does not restrict a monotonic decrease of the objective function value at each iteration. If a trial step is not accepted, the algorithm performs a nonmonotone line search to find a new acceptable point instead of resolving the subproblem. We prove that the algorithm globally converges to a point satisfying the second-order necessary optimality conditions.

AMS Mathematics Subject Classification : 90C30, 65K05.
Key words and phrases : Nonlinear optimization, primal-dual method; interior point method, nonmonotone technique, trust region method.

1. Introduction

Asset liability management problems have received considerable attention in financial planning under uncertainty, see, for example, Ziemba and Mulvey [19], Consiglio et al. [4], Hilli et al. [14]. Generally, these problems can be formulated as stochastic programming problems. However, the deterministic equivalent formulations of these stochastic programs have large dimensions even for moderate numbers of assets, time stages and scenarios per time stage. Therefore, most of asset liability management models have been limited to simple linear or quadratic models such that they can be solved by currently available solvers. Recently, many papers have been devoted to solve nonlinear multistage stochastic programming problems. For instance, Gondzio and Grothey [11] proposed

Received May 19, 2008. Revised March 1, 2009. Accepted March 24, 2009. * Corresponding author. [1]Research was supported by the National Natural Science Foundation of China (No.10671126).

a parallel interior point algorithm for multistage stochastic quadratic programming problems. Zhao [18] presented a algorithm based on the Lagrangian dual method for multistage stochastic convex programming problems. Berkelaar et al. [2] developed an algorithm based on the path-following interior point method combined with the homogeneous self-dual embedding technique for multistage stochastic convex programming problems. The same characteristic in these papers is that Newton step is used to solve the Newton equations linear system. Motivated by these studies, in this paper we propose a primal-dual interior point nonmonotone trust region method for solving general nonlinear multistage stochastic programming problems (general nonlinear multistage stochastic programming asset liability management models can be seen in Gondzio and Grothey [12]). Our method suggests that we only use approximate Newton step to solve the linear system. Particularly, a nonmonotone strategy is helpful to reduce computational cost, especially for large dimension optimization problems. Without loss of generality, we consider the deterministic equivalent formulation of a general nonlinear multistage asset liability management problem of the form

$$\begin{aligned} \min_{x\in\Re^n} \quad & f(x) \\ s.t. \quad & c_i(x) = 0, i \in E, \\ & c_i(x) \geq 0, i \in I, \end{aligned} \tag{1}$$

where $c(x)$ is an m-vector of nonlinear functions with i-th component $c_i(x), i = 1, 2, \cdots, m$, and E and I are nonintersecting index sets. It is assumed that f and c are twice continuously differentiable, with the gradient of $f(x)$ denoted by $g(x)$ and the $m \times n$ Jacobian of $c(x)$ denoted by $J(x)$.

Interior point methods are well-suited to solve large dimension nonlinear optimization problems. In recent years, interior point methods have received extensive attention, see, e.g., [1,3,5,7,8,17]. For a more complete survey, see Gertz and Gill [9] and Forsgren, Gill and Wright [7]. Due to Fiacco and McCormick [6] classical penalty-barrier method, an alternative method for solving problem (1) is to minimize the following unconstrained function

$$L^\mu(x) = f(x) + \frac{1}{2\mu}\sum_{i\in E} c_i(x)^2 - \mu \sum_{i\in I} lnc_i(x), \tag{2}$$

where the positive parameter $\{\mu\}$ is a decreasing sequence. The first constraint term on the right-hand side is the usual quadratic penalty function with penalty parameter $1/(2\mu)$. The second constraint term is the logarithmic barrier function, which creates a positive singularity at the boundary of the feasible region and therefore ensures strict feasibility while approaching the solution. The mechanism of primal-dual methods involves a two-level structure of outer and inner iterations. Each outer iteration is associated with an element of a decreasing positive sequence of parameters $\{\mu_j\}$ such that $\lim_{j\to\infty} \mu_j = 0$. The inner iterations correspond to an iterative process for the unconstrained minimization of $L^\mu(x)$ for a given μ.

To describe principia of primal-dual interior methods, we consider the first-order necessary optimality conditions associated with problem (1). At an optimal solution x^*, there exists an m-vector λ^* of Lagrange multipliers such that

$$\begin{cases} \nabla f(x^*) - J(x^*)^T \lambda^* = 0, \\ c_i(x^*) = 0, \quad i \in E, \\ c_i(x^*)\lambda_i^* = 0, \quad i \in I, \\ \lambda_i^* \geq 0, \quad c_i(x^*) \geq 0, \quad i \in I. \end{cases}$$

If we combine the first three terms of (3) in a compact form, then (3) can be reformulated as $F^\infty(x^*, \lambda^*) = 0$, and $\lambda_i^* \geq 0$, $c_i(x^*) \geq 0$ for $i \in I$, where F^∞ is the vector-valued function

$$F^\infty(x, \lambda) = \begin{pmatrix} \nabla f(x) - J(x)^T \lambda \\ r(x, \lambda) \end{pmatrix}, \tag{3}$$

and $r_i(x, \lambda) = c_i(x), i \in E; r_i(x, \lambda) = c_i(x)\lambda_i, i \in I$.

Primal-dual methods can be interpreted as solving a sequence of nonlinear systems in which each condition $c_i(x) = 0, i \in E$ is perturbed as $c_i(x) + \mu\lambda_i = 0$ and $c_i(x)\lambda_i = 0, i \in I$ is perturbed as $c_i(x)\lambda_i - \mu = 0$ for some small positive μ. The perturbed equation $F^\mu(x, \lambda) = 0$ can be solved by performing a form of Newton's method in which x and λ are chosen to be interior for the inequalities $c_i(x) > 0$ and $\lambda_i > 0$ for $i \in I$. The inner mechanism of primal-dual interior method is to solve equation $F^\mu(x, \lambda) = 0$ inaccurately for a positive decreasing sequence of μ such that $(x(\mu), \lambda(\mu))$ converges to (x^*, y^*) as $\mu \to 0$.

We now give a brief description for inner iteration. For any value of μ, an associated point $(x(\mu), \lambda(\mu))$ on the trajectory satisfies the $n + m$ equations

$$\begin{cases} \nabla f(x) - J(x)^T \lambda = 0, \\ c_i(x) + \mu\lambda_i = 0, \quad i \in E, \\ c_i(x)\lambda_i - \mu = 0, \quad i \in I. \end{cases}$$

These relations imply that $(x(\mu), \lambda(\mu))$ can be determined by solving $n + m$ nonlinear equations in the $n + m$ unknowns (x, λ) using an iterative method. Let ν denotes the $n + m$ vector of unknowns (x, λ) at an interior point, that is, a point such that $c_i(x) > 0$ and $\lambda_i > 0$ for $i \in I$. If $F^\mu(\nu)$ denotes the function $F^\mu(x, \lambda)$, then a Newton direction $\Delta\nu = (\Delta x, \Delta\lambda)$ is defined by the Newton equations $F^\mu(\nu)'\Delta\nu = -F^\mu(\nu)$. To describe these Newton equations conveniently, it is helpful to rewrite the second and the third terms of (4) in vector form $\Gamma^\mu(x)(\lambda - \pi^\mu(x)) = 0$, where $\pi^\mu(x)$ is given as

$$\pi_i^\mu(x) = \begin{cases} -c_i(x)/\mu & if \;\; i \in E, \\ \mu/c_i(x) & if \;\; i \in I, \end{cases}$$

and $\Gamma^\mu(x)$ is the diagonal matrix with diagonal entries

$$\gamma_i^\mu(x) = \begin{cases} \mu & if \;\; i \in E, \\ c_i(x) & if \;\; i \in I. \end{cases}$$

Using these definitions, $F^\mu(x,\lambda)$ and $(F^\mu)'(x,\lambda)$ can be expressed as

$$F^\mu(x,\lambda)=\begin{pmatrix}\nabla f(x)-J(x)^T\lambda\\ \Gamma^\mu(x)(\lambda-\pi^\mu(x))\end{pmatrix},\quad (F^\mu)'(x,\lambda)=\begin{pmatrix}H(x,\lambda) & -J(x)^T\\ Z(\lambda)J(x) & \Gamma^\mu(x)\end{pmatrix},$$

where $H(x,\lambda)=\nabla^2 f(x)-\sum_{i=1}^m \lambda_i\nabla^2 c_i(x)$ denotes the Hessian of the Lagrangian, and $Z(\lambda)$ is a diagonal matrix with diagonal entries

$$z_i(\lambda)=\begin{cases}1 & if\ \ i\in E,\\ \lambda_i & if\ \ i\in I.\end{cases}$$

Thus the Newton equations have the form

$$\begin{pmatrix}H(x,\lambda) & -J(x)^T\\ Z(\lambda)J(x) & \Gamma^\mu(x)\end{pmatrix}\begin{pmatrix}\Delta x\\ \Delta\lambda\end{pmatrix}=\begin{pmatrix}\nabla f(x)-J(x)^T\lambda\\ \Gamma^\mu(x)(\lambda-\pi^\mu(x))\end{pmatrix}. \tag{4}$$

This is an unsymmetric primal-dual system, it can be symmetrized by premultiplying the last m rows by Z^{-1} and changing the sign of $\Delta\lambda$, namely

$$\begin{pmatrix}H(x,\lambda) & J(x)^T\\ J(x) & -W(x)\end{pmatrix}\begin{pmatrix}\Delta x\\ -\Delta\lambda\end{pmatrix}=\begin{pmatrix}\nabla f(x)-J(x)^T\lambda\\ W(x)(\lambda-\pi^\mu(x))\end{pmatrix}, \tag{5}$$

where $W=Z^{-1}\Gamma$. The cost of a primal-dual iteration is dominated by the cost of solving the linear system (4).

The primal-dual method we stated above requires that every iterate (x,λ) satisfies $c_i(x)>0$ and $\lambda_i>0$ for all $i\in I$, and that $(x(\mu),\lambda(\mu))$ converge to (x^*,λ^*) as μ converge to zero. To ensure global convergence it is necessary to use a merit function to force the early iterates towards the trajectory $(x(\mu),\lambda(\mu))$. Although (2) is a classical penalty-barrier function, however, if L^μ is used as as merit function, the iterative information of the dual variables can not be obtained simply because it does not have terms involving λ. An alternative approach is based on the properties of the Forsgren-Gill augmented barrier-penalty function:

$$\begin{aligned}Q^\mu(x,\lambda)&=f(x)-\mu\sum_{i\in I}lnc_i(x)+\frac{1}{2\mu}\sum_{i\in E}c_i(x)^2\\ &-\mu\sum_{i\in I}(ln(\frac{c_i(x)\lambda_i}{\mu})+\frac{\mu-c_i(x)\lambda_i}{\mu})+\frac{1}{2\mu}\sum_{i\in E}(c_i(x)+\mu\lambda_i)^2,\end{aligned} \tag{6}$$

which is the penalty-barrier function $L^\mu(x)$ augmented by a weighted proximity term that measures the distance of (x,λ) to the trajectory $(x(\mu),\lambda(\mu))$. A fundamental property of $Q^\mu(x,\lambda)$ is that it is minimized with respect to both x and λ at any point $Q^\mu(x,\lambda)$ on the trajectory.

Associated with the augmented barrier-penalty function (6), a local quadratic model is

$$\min_{s\in\Re^{n+m}}\quad g^Ts+\frac{1}{2}s^TBs=\phi(s), \tag{7}$$

where s is a direction which combines the primal and dual variables,

$$g=\begin{pmatrix}\nabla f-J^T(2\pi-\lambda)\\ W(\lambda-\pi)\end{pmatrix}\ and\ B=\begin{pmatrix}H+2J^TW^{-1}J & J^T\\ J & W\end{pmatrix}. \tag{8}$$

The vector g is the gradient of the merit function ∇Q^μ, B is an approximation of Hessian matrix $\nabla^2 Q^\mu$. If $(x, \lambda) = (x(\mu), \lambda(\mu))$ is a point on the trajectory, then $B = \nabla^2 Q^\mu$. This quadratic model gives us implications that we can solve it by many effective methods. For instance, Newton methods and trust region methods. It can be shown that if $H + J^T W^{-1} J$ is positive definite, then B is positive definite. That is, if $H + J^T W^{-1} J$ is positive definite, then the solution of the symmetric positive definite system $Bs = -g$ is the unique minimizer of $\phi(s)$. Accordingly, $s = (\Delta x, \Delta\lambda)$ is a solution of the primal-dual system (4). This implies that algorithms that solve the primal-dual equations are implicitly using an approximate Newton method to minimize $Q^\mu(x, \lambda)$. However, $H + J^T W^{-1} J$ is not necessarily positive definite, this motivate us to employ other methods which without requiring B is positive definite.

The above theoretical analysis inspire us to recall trust region methods, which without requirement that B is positive definite for solving a quadratic model. Furthermore, strong convergence results are available and softwares for them are reliable and efficient. Consider the unconstrained minimization of $Q^\mu(x, \lambda)$, trust region methods minimize a quadratic model of the objective function subject to a restriction on the length of the step. At each iteration, denoted by $\nu_j = (x_j, \lambda_j)$, a trial step s_j is generated by solving the subproblem:

$$\begin{aligned} \min_{s \in \Re^{n+m}} \quad & g_j^T s + \frac{1}{2} s^T B_j s = \phi_j(s) \\ s.t. \quad & \|s\|_{T_j} \le \Delta_j, \end{aligned} \tag{9}$$

where g_j and B_j are defined by (8), i.e.,

$$g_j = \begin{pmatrix} \nabla f_j - J_j^T (2\pi_j - \lambda_j) \\ W_j(\lambda_j - \pi_j) \end{pmatrix} \; and \; B_j = \begin{pmatrix} H_j + 2J_j^T W_j^{-1} J_j & J_j^T \\ J_j & W_j \end{pmatrix},$$

$\|\cdot\|_{T_j}$ denotes the elliptic norm $\|s\|_{T_j} = (s^T T_j s)^{1/2}$ and $\Delta_k > 0$ is a trust region radius. The matrix T_j is a block-diagonal matrix of the form $T_j = diag(M_j, N_j)$, where M_j and N_j are $n \times n$ and $m \times m$ symmetric positive-definite matrices. There are many successful monotone algorithms for computing an approximate solution of (9)(see for example [9,10,15,16]). However, monotone algorithms can not guarantee that at each iteration an acceptable step is found. The subproblem may be solved several times at an iteration, which can considerably increase the total cost of computation for large scale problems.

Gertz and Gill [9] proposed a primal-dual interior point monotone algorithm for (1) based on (6). In this paper, our purpose is to develop a primal-dual interior point nonmonotone trust region algorithm to solve (1) based on the convex combination nonmonotone technique of the form[13]

$$f(x_j + \alpha d_j) \le D_j + \delta\alpha \nabla f(x_j)^T d_j, \tag{10}$$

where D_j is a simple convex combination of the previous D_{j-1} and f_j, say

$$D_j = \begin{cases} f(x_j), & j = 1; \\ \eta D_{j-1} + (1-\eta) f(x_j) & j \geq 2 \end{cases} \tag{11}$$

for some fixed $\eta \in (0, 1)$, or a variable η_j.

The paper is organized as follows. In Section 2, we describe the algorithm. In Section 3, we establish the global convergence for our algorithm under suitable conditions. Finally, we give our brief conclusions in Section 4.

2. Nonmonotone trust region algorithm

We use $\|\cdot\|$ to represent the Euclidean norm or its subordinate matrix norm. The least eigenvalue of a symmetric matrix A will be denoted by $\zeta_{\min}(A)$. Let $\{\varphi_j\}_j \geq 0$ be a sequence of scalars, vectors or matrices and let $\{\chi_j\}_j \geq 0$ be a sequence of positive scalars. If there exists a positive constant γ such that $\|\varphi_j\| \leq \gamma\chi_j$, we write $\varphi_j = O(\chi_j)$. If there exist positive constants γ_1 and γ_2 such that $\gamma_1\chi_j \leq \|\varphi_j\| \leq \gamma_2\chi_j$, we write $\varphi_j = \Theta(\chi_j)$. If a vector is denoted by a lower-case letter, the same upper-case letter denotes the diagonal matrix whose elements are those of the vector, for instance, $V = diag(v)$. Let Ω be the feasible points set of $Q^\mu(x, \lambda)$ for every $\mu > 0$.

Similar to Section 5 in [9], we solve the trust region subproblem (9) inaccurately. Given a fixed tolerance $\tau \in (0, 1)$, an approximate solution for subproblem (9) only to satisfy the conditions

$$\phi_j(s_j) \leq \tau\phi_j(s_j^c) \ \ and \ \|s_j\|_{T_j} \leq \Delta_j, \tag{12}$$

where s_j^c is the scaled Cauchy point, which is defined as the solution of the problem $\min_{s,\beta}\{\phi_j(s) : T_j s = \beta g_j, \|s\|_{T_j} \leq \Delta_j\}$. The method finds a step s_j that satisfies the sufficient decrease conditions

$$\phi_j(s_j) \leq \tau\phi_j^* \ \ and \ \|s_j\|_{T_j} \leq \Delta_j, \tag{13}$$

where ϕ_j^* is the unique minimum of $\phi_j(s)$ on $\{s : \|s\|_{T_j} \leq \Delta_j\}$. The required conditions (12) on s_j then follow from the inequalities $\phi_j(s_j) \leq \tau\phi_j^* \leq \tau\phi_j(s_j^c)$.

The nonmonotone line search is given as:

$$Q^\mu(\nu_j + \alpha s_j) \leq D^\mu(\nu_j) + \delta\alpha g_j^T s_j, \tag{14}$$

where

$$D^\mu(\nu_j) = \begin{cases} Q^\mu(\nu_j), & j = 1; \\ \eta D^{\mu_{j-1}}(\nu_{j-1}) + (1-\eta) Q^\mu(\nu_j), & j \geq 2, \end{cases} \tag{15}$$

where μ_{j-1} is the parameter value of μ on the $(j-1)$-th iterate.

To determine whether a trial step will be accepted, we compute ρ_j, the ratio between the actual reduction and the predicted reduction as

$$\rho_j = \frac{D^\mu(\nu_j) - Q^\mu(\nu_j + s_j)}{\phi_j(0) - \phi_j(s_j)}, \tag{16}$$

where $D^\mu(\nu_j)$ is defined by (15). If $\rho_j \geq \varpi$, where ϖ is a constant, we accept s_j as a successful step and let $\nu_{j+1} = \nu_j + s_j$. Otherwise, we generate a new iterative point by $\nu_{j+1} = \nu_j + \alpha_j s_j$, where α_j is a steplength satisfying the line search condition (14).

We now describe a nonmonotone trust region algorithm as follows:

Algorithm 1

Step 1: Given $\Delta_1 > 0$, choose constants $c_1, c_2, \varpi, \epsilon, \xi, \delta$, such that $0 < c_1 < 1 < c_2$, $\varpi \in (0,1)$, $\epsilon = 10^{-6}$, $\xi \in (0,1)$ and $\delta \in (0, 1/2)$. Set $j := 1$.

Step 2: Compute g_j, if $\|g_j\| < \epsilon$, stop.

Step 3: Solve (9) inaccurately.

Step 4: Compute ρ_j by (16).

Step 5: If $\rho_j \geq \varpi$, go to Step 6. Otherwise, compute i_j, the minimum nonnegative integer i satisfies

$$Q^{\mu}(\nu_j + \xi^i s_j) \leq D^{\mu}(\nu_j) + \delta \xi^i g_j^T s_j. \tag{17}$$

Set $\alpha_j = \xi^{i_j}$,

$$\nu_{j+1} = \nu_j + \alpha_j s_j, \tag{18}$$

and

$$\Delta_{j+1} \in [\|\nu_{j+1} - \nu_j\|, c_1 \Delta_j], \tag{19}$$

go to Step 7.

Step 6: Set

$$\nu_{j+1} = \nu_j + s_j, \tag{20}$$

and

$$\Delta_{j+1} \begin{cases} = \Delta_j, & if \quad \|s_j\|_{T_j} < \Delta_j, \\ \in [\Delta_j, c_2 \Delta_j], & if \quad \|s_j\|_{T_j} = \Delta_j. \end{cases} \tag{21}$$

Step 7: Set $j := j + 1$, go to Step 2.

Since $\{B_j\}$ may grow without bound, it is helpful to give a transformed matrix for $\{B_j\}$, such that the new transformation is bounded. Gertz and Gill [9] give a corollary that if $\{f(x_j)\}$ is bounded below and $\{c_i(x_j)\}$ is bounded above for all $i \in I$, then the sequence $\{W_j^{-1/2}\}$ is bonded. This observation leads us to consider the transformed matrix

$$T_j^{-\frac{1}{2}} B_j T_j^{-\frac{1}{2}} = \begin{pmatrix} M_j^{-\frac{1}{2}}(H_j + 2J_j^T W_j^{-1} J_j) M_j^{-\frac{1}{2}} & M_j^{-\frac{1}{2}} J_j^T M_j^{-\frac{1}{2}} \\ W_j^{-\frac{1}{2}} J_j W_j^{-\frac{1}{2}} & I \end{pmatrix}, \tag{22}$$

which still remains bounded even $\|B_j\|$ does not, where $T_j = diag(M_j, W_j)$ with M_j is a $n \times n$ symmetric positive-definite (e.g., the identify matrix) and $w_{ij} = \mu$ for $i \in E$, $w_{ij} = c_i(x_j)/\lambda_{ij}$ for $i \in I$. Therefore, at each point $\nu = (x, \lambda)$, we consider the transformed quantities

$$\hat{g}(\nu) = T^{-\frac{1}{2}} g \quad and \quad \hat{B}(\nu) = T^{-\frac{1}{2}} B T^{-\frac{1}{2}}. \tag{23}$$

Consequently, the trust region subproblem (9) can be reformulated as

$$\min_{\hat{s}\in\Re^{n+m}} \quad \hat{g}_j^T\hat{s} + \frac{1}{2}\hat{s}^T\hat{B}_j\hat{s} = \hat{\phi}_j(\hat{s}) \qquad (24)$$
$$s.t. \quad \|\hat{s}\| \le \Delta_j,$$

where $\hat{s} = T_j^{1/2}s$. It is more convenient to use (24) for discussing the theoretical properties of the algorithm. To this end, we assume that $\{f(x_j)\}$ is bounded below and $\{c_i(x_j)\}$ is bounded above for all $i \in I$, such that matrix $\{\hat{B}_j\}$ is bounded ($\{\|\hat{B}_j\|\}$ is bounded). Associated with problem (24), line search condition (17) can be written as $Q^\mu(\nu_j + \xi^i\hat{s}_j) \le D^\mu(\nu_j) + \delta\xi^i\hat{g}_j^T\hat{s}_j$.

The properties of the approximate solution for subproblem (24) have been reported by many researchers, see e.g., [10,15,16]. It has been shown that the following two inequalities

$$\phi_j(0) - \phi_j(s_j) \ge \tau\|\hat{g}_j\|\min\{\Delta_j, \|\hat{g}_j\|/\|\hat{B}_j\|\} \qquad (25)$$

and

$$\hat{s}_j^T\hat{g}_j \le -\tau\|\hat{g}_j\|\min\{\Delta_j, \|\hat{g}_j\|/\|\hat{B}_j\|\} \qquad (26)$$

hold, where $\tau \in (0,1)$ is a constant (see [16]). In this paper, we solve (9) inaccurately such that $\|\hat{s}_j\| \le \Delta_j$, and the above two inequalities hold.

3. Global convergence

Under certain assumptions, we show that the sequence of inner iterations converges to a point satisfying the second-order necessary optimality conditions. We begin this section with the following two assumptions.

Assumption A. The function $Q^\mu(\nu)$ is bounded below on Ω for every $\mu > 0$.

Assumption B. There exists a sufficiently small positive constant r such that

$$\hat{s}^T\hat{B}_j\hat{s} \ge r\|\hat{s}\|^2, \qquad \forall\ \hat{s} \in \Re^{n+m} \ \ \textit{and} \ \ j = 1, 2, \cdots.$$

Remark 3.1. Under Assumption A, by the fact that $Q^\mu(\nu)$ is continuously differentiable, there exists a constant $R > r$, such that

$$\|\nabla^2 Q^\mu(\nu)\| \le R, \qquad \forall\ \nu \in \Omega. \qquad (27)$$

For simplicity, we define two index sets as follows:

$$S = \{j : \rho_j \ge \varpi\} \text{ and } F = \{j : \rho_j < \varpi\}.$$

Note that parameter value μ is determinate according to every iterate. In the remainder of this paper, we will write $Q^\mu(\nu_j) = Q(\nu_j) = Q_j$ and $D^\mu(\nu_j) = D(\nu_j) = D_j$.

Compared to the Armijo-type line search condition $Q^\mu(\nu_j + \alpha s_j) \le Q^\mu(\nu_j) + \delta\alpha g_j^T s_j$, an important feature of nonmonotone technique (14) is that it does not restrict a monotonic decrease of the objective function value at each iterate.

Lemma 1. *Let $\{\nu_j\}$ be the sequence generated by Algorithm 1. Then*

$$Q_{j+1} < D_{j+1} \tag{28}$$

holds for all j.

Proof. Due to definition of D_j (15), we obtain

$$\begin{aligned} D_{j+1} - Q_{j+1} &= \eta D_j + (1-\eta)Q_{j+1} - Q_{j+1} \\ &= \eta(D_j - Q_{j+1}). \end{aligned} \tag{29}$$

Inequalities (25) and (26) imply that $\phi_j(0) - \phi_j(s_j) > 0$ and $\hat{s}_j^T \hat{g}_j < 0$ if the algorithm does not terminate or converge. If $j \in S$, i.e., $\rho_j \geq \varpi$, then $D_j - Q_{j+1} > 0$. Otherwise, the nonmonotone line search (17) is performed, this process still leads to $D_j - Q_{j+1} > 0$. Thus (28) holds. □

Next, we show that Algorithm 1 is well defined. It is enough to prove that there exists an integer i_j such that the line search (17) holds.

Lemma 2. *Assume that sequence $\{\nu_j\}$ is generated by Algorithm 1. Then line search (17) terminates in finite steps, i.e., there exists an integer i_j such that the line search (17) holds, for any $j \in F$.*

Proof. Suppose first, for the purpose of deriving a contradiction, that there exists $j \in F$ such that

$$Q(\nu_j + \xi^i \hat{s}_j) > D(\nu_j) + \delta \xi^i \hat{g}_j^T \hat{s}_j, \qquad \forall\, i.$$

Using the fact $D(\nu_j) > Q(\nu_j)$, we obtain

$$Q(\nu_j + \xi^i \hat{s}_j) > Q(\nu_j) + \delta \xi^i \hat{g}_j^T \hat{s}_j, \qquad \forall\, i.$$

This inequality leads to

$$\frac{Q(\nu_j + \xi^i \hat{s}_j) - Q(\nu_j)}{\xi^i} > \delta \hat{g}_j^T \hat{s}_j.$$

Since $Q(\nu_j)$ is differentiable, taking limit with $i \to \infty$, we have

$$\hat{g}_j^T \hat{s}_j \geq \delta \hat{g}_j^T \hat{s}_j. \tag{30}$$

Recall that $\delta \in (0, 1/2)$, thus (30) implies that $\hat{g}_j^T \hat{s}_j \geq 0$. However, we have from (26) that $\hat{g}_k^T \hat{s}_k < 0$. Therefore, for any $j \in F$, there exists $i_j > 0$ such that (17) holds. □

Now we proceed to establish a lower bound for stepsize α_j. To this end, we need to use Assumption B.

Lemma 3. *Assume that Assumptions A and B hold, sequence $\{\nu_j\}$ is generated by Algorithm 1. Then the stepsize α_j satisfies*

$$\alpha_j > \frac{(1-\delta)\xi r}{R} \tag{31}$$

for all $j \in F$.

Proof. If the line search is performed, it means that trial step is not accepted, the iterate is updated by $\nu_{j+1} = \nu_j + \alpha_j \hat{s}_j$. Using the definition of line search (17), we obtain

$$Q(\nu_j + \xi^{-1}\alpha_j \hat{s}_j) > D(\nu_j) + \delta\xi^{-1}\alpha_j \hat{g}_j^T \hat{s}_j,$$

Again using the fact $D(\nu_j) > Q(\nu_j)$, we have

$$Q(\nu_j + \xi^{-1}\alpha_j \hat{s}_j) > Q(\nu_j) + \delta\xi^{-1}\alpha_j \hat{g}_j^T \hat{s}_j. \tag{32}$$

Due to Taylor's expansion, we obtain

$$Q(\nu_j + \xi^{-1}\alpha_j \hat{s}_j) = Q(\nu_j) + \xi^{-1}\alpha_j \hat{g}_j^T \hat{s}_j + \frac{1}{2}\xi^{-2}\alpha_j^2 \hat{s}_j^T \nabla^2 Q(\iota_j)\hat{s}_j, \tag{33}$$

where $\iota_j \in (\nu_j, \nu_j + \xi^{-1}\alpha_j \hat{s}_j)$. It is clear from (32), (33) and Remark 3.1 that

$$\delta\xi^{-1}\alpha_j \hat{g}_j^T \hat{s}_j < \xi^{-1}\alpha_j \hat{g}_j^T \hat{s}_j + \frac{1}{2}\xi^{-2}\alpha_j^2 R\|\hat{s}_j\|^2,$$

this inequality yields

$$-(1-\delta)\hat{g}_j^T \hat{s}_j < \frac{1}{2}\xi^{-1}\alpha_j R\|\hat{s}_j\|^2. \tag{34}$$

On the other hand, we deduce from (25) that

$$\phi_j(0) - \phi_j(s_j) = -\hat{g}_j^T \hat{s}_j - \frac{1}{2}\hat{s}_j^T \hat{B}_j \hat{s}_j > 0. \tag{35}$$

Combining (34) and (35), we obtain

$$(1-\delta)\hat{s}_j^T \hat{B}_j \hat{s}_j < \xi^{-1}\alpha_j R\|\hat{s}_j\|^2. \tag{36}$$

Thus, we have from Assumption B and (36) that (31) holds. □

Recall that $\{\hat{B}_j\}$ is bounded, here we define a sequence

$$G_j = 1 + \max_{1 \le i \le j} \|\hat{B}_i\|. \tag{37}$$

This sequence will be used in our theoretical discussion.

In the following lemma, we find that sequence $\{D_k\}$ is monotonically decreasing.

Lemma 4. *Suppose that Assumption A holds and $\{\nu_j\}$ is a sequence generated by Algorithm 1, then sequence $\{D_k\}$ is monotonically decreasing. Furthermore, if sequence $\{\nu_j\}$ does not converge, i.e., there exists a constant $\epsilon > 0$ such that*

$$\|\hat{g}_j\| \ge \epsilon, \qquad \forall\, j. \tag{38}$$

Then sequence $\{D(\nu_j)\}$ satisfies

$$D(\nu_{j+1}) - D(\nu_j) \le -(1-\eta)\psi\epsilon \min\{\Delta_j, \epsilon/G_j\} \tag{39}$$

for all j, where $\psi = \min\{\varpi\tau, \frac{\delta\tau(1-\delta)\xi r}{R}\}$.

Proof. We will finish the proof by considering two cases.
Case 1. $j \in S$, we have from (16) and (25) that

$$D(\nu_j) - Q(\nu_{j+1}) \geq \varpi(\phi_j(0) - \phi_j(\hat{s}_j)) \geq \varpi\tau\|\hat{g}_j\| \min\{\Delta_j, \|\hat{g}_j\|/\|\hat{B}_j\|\}.$$

Thus

$$Q(\nu_{j+1}) \leq D(\nu_j) - \varpi\tau\|\hat{g}_j\| \min\{\Delta_j, \|\hat{g}_j\|/\|\hat{B}_j\|\}. \tag{40}$$

Case 2. $j \in F$, it follows from (17), (26) and (31) that

$$\begin{aligned} Q(\nu_{j+1}) &\leq D(\nu_j) - \delta\tau\alpha_j\|\hat{g}_j\| \min\{\Delta_j, \|\hat{g}_j\|/\|\hat{B}_j\|\} \\ &\leq D(\nu_j) - \frac{\delta\tau(1-\delta)\xi r}{R}\|\hat{g}_j\| \min\{\Delta_j, \|\hat{g}_j\|/\|\hat{B}_j\|\}. \end{aligned} \tag{41}$$

Let $\psi = \min\{\varpi\tau, \frac{\delta\tau(1-\delta)\xi r}{R}\}$. Combining (40) and (41), we have

$$Q(\nu_{j+1}) \leq D(\nu_j) - \psi\|\hat{g}_j\| \min\{\Delta_j, \|\hat{g}_j\|/\|\hat{B}_j\|\}. \tag{42}$$

Due to the definition of D_{j+1} (15) and inequality (42), we have

$$\begin{aligned} D_{j+1} &= \eta D_j + (1-\eta)Q_{j+1} \\ &\leq \eta D_j + (1-\eta)(D_j - \psi\|\hat{g}_j\| \min\{\Delta_j, \|\hat{g}_j\|/\|\hat{B}_j\|\}) \\ &= D_j - (1-\eta)\psi\|\hat{g}_j\| \min\{\Delta_j, \|\hat{g}_j\|/\|\hat{B}_j\|\}. \end{aligned} \tag{43}$$

Inequality (43) implies that sequence $\{D_j\}$ is monotonically decreasing. Thus, (43), (38) and (37) give

$$\begin{aligned} D_{j+1} - D_j &\leq -(1-\eta)\psi\|\hat{g}_j\| \min\{\Delta_j, \|\hat{g}_j\|/\|\hat{B}_j\|\} \\ &\leq -(1-\eta)\psi\epsilon \min\{\Delta_j, \epsilon/G_j\}. \end{aligned}$$

This completes the proof. □

The monotonicity of $\{D_j\}$ leads to the following important lemma.

Lemma 5. *Suppose that Assumption A holds, if sequence $\{\nu_j\}$ does not converge, i.e., there exists a constant $\epsilon > 0$ such that (38) holds. Then*

$$\lim_{j\to\infty} \min\{\Delta_j, \epsilon/G_j\} = 0. \tag{44}$$

Proof. Since $\{Q_j\}$ is bounded below, we have from Lemma 1 that $\{D_j\}$ is also bounded below. Inequality (39) leads to

$$\sum_{j=1}^{\infty}(D(\nu_{j+1}) - D(\nu_j)) \leq \sum_{j=1}^{\infty} -(1-\eta)\psi\epsilon \min\{\Delta_j, \epsilon/G_j\}.$$

Consequently,

$$\sum_{j=1}^{\infty}(1-\eta)\psi\epsilon \min\{\Delta_j, \epsilon/G_j\} < \infty.$$

It is obviously that (44) holds. □

In the following lemma, we show that there exists a lower bound for Δ_j, for $j \in F$.

Lemma 6. *Suppose that sequence $\{\nu_j\}$ is generated by Algorithm 1, if sequence $\{\nu_j\}$ does not converge, that is, there exists a constant $\epsilon > 0$ such that (38) holds. Then*

$$\Delta_j > \epsilon/G_j \tag{45}$$

is satisfied for $j \in F$ sufficiently large .

Proof. Using (32), the Taylor's expansion (33) and Remark 3.1, we obtain

$$\begin{aligned} 0 &> Q(\nu_j) - Q(\nu_j + \xi^{-1}\alpha_j\hat{s}_j) + \delta\xi^{-1}\alpha_j\hat{g}_j^T\hat{s}_j \\ &\geq -\xi^{-1}(1-\delta)\alpha_j\hat{g}_j^T\hat{s}_j - \frac{1}{2}R\xi^{-2}\alpha_j^2\|\hat{s}_j\|^2. \end{aligned}$$

Which, together with (26) and (37), we have

$$\begin{aligned} 0 &> \xi^{-1}(1-\delta)\tau\epsilon\alpha_j\min\{\Delta_j, \epsilon/G_j\} - \frac{1}{2}R\xi^{-2}\alpha_j\|\hat{s}_j\|\|\nu_{j+1}-\nu_j\| \\ &\geq \xi^{-1}\alpha_j[(1-\delta)\tau\epsilon\min\{\Delta_j, \epsilon/G_j\} - \frac{1}{2}R\xi^{-1}\Delta_j\|\nu_{j+1}-\nu_j\|] \\ &= \xi^{-1}\alpha_j\Delta_j[(1-\delta)\tau\epsilon\min\{1, \epsilon/(\Delta_jG_j)\} - \frac{1}{2}R\xi^{-1}\|\nu_{j+1}-\nu_j\|]. \end{aligned}$$

This inequality leads to

$$\|\nu_{j+1}-\nu_j\| > \frac{2(1-\delta)\tau\xi\epsilon}{R}\min\{1, \epsilon/(\Delta_jG_j)\}. \tag{46}$$

Assume that (45) does not hold, i.e.,

$$\Delta_j \leq \epsilon/G_j. \tag{47}$$

Then it follows from (46) that

$$\|\nu_{j+1}-\nu_j\| > 2(1-\delta)\tau\epsilon\xi/R. \tag{48}$$

Due to Step 5 of Algorithm 1, we have

$$\|\nu_{j+1}-\nu_j\| = \alpha_j\|\hat{s}_j\| \leq \Delta_j, \qquad \forall\ j \in F. \tag{49}$$

Now, (47), (48) and (49) imply that

$$\epsilon/G_j \geq \Delta_j > 2(1-\delta)\tau\epsilon\xi/R,$$

this relation contradicts (44). Thus (45) holds. □

Based on Lemma 5 and Lemma 6, we will show that there exists a lower bound for the trust region radius Δ_j, for all sufficiently large j.

Lemma 7. *Suppose that sequence $\{\nu_j\}$ is generated by Algorithm 1, if there exists a constant $\epsilon > 0$ such that $\|\hat{g}_j\| \geq \epsilon$. Then*

$$\Delta_j > \epsilon/G_j \tag{50}$$

for all sufficiently large j.

Proof. If there are an finite number of unsuccessful iterations, namely, F is a finite set, then there exists a positive constant Δ^* such that $\Delta_j > \Delta^*$ for all j. Note that (44) implies that $\lim_{j\to\infty} 1/G_j = 0$. Hence (50) holds for all large j.

Now, we assume that F is an infinite set.

(1): If $j \in F$, by Lemma 6 there exists a $\bar{j} \in F$ such that (50) holds for $j \in F$ and $j \geq \bar{j}$.

(2): If $j \in S$ and $j \geq \bar{j}$, let $\tilde{j} = \max\{i : i \in F \ \ and \ \ i \leq j\}$. The definition of $\tilde{j}$ leads to

$$\Delta_{\tilde{j}} > \epsilon/G_{\tilde{j}} \tag{51}$$

and

$$\tilde{j} + s \in S \tag{52}$$

for all $s = 1, 2, \cdots j - \tilde{j}$. Using (51), (52) and the rules for updating trust region radius (21), we have

$$\Delta_{\tilde{j}} \leq \Delta_{\tilde{j}+1} \leq \Delta_{\tilde{j}+2} \leq \cdots \leq \Delta_j. \tag{53}$$

Relations (53) and (51) suggest that

$$\Delta_j > \epsilon/G_{\tilde{j}}.$$

Since $j > \tilde{j}$, by the monotonicity of $\{G_j\}$, we have $G_j \geq G_{\tilde{j}}$ and hence $\epsilon/G_{\tilde{j}} \geq \epsilon/G_j$. Thus (50) holds. □

Based on these lemmas, we now give an important theorem for Algorithm 1. Since $\{\hat{B}_j\}$ is bounded, it follows that $\{G_j\}$ satisfies

$$\sum_{j=1}^{\infty} 1/G_j = \infty, \tag{54}$$

where G_j is defined by (37).

Theorem 1. *Suppose that Assumption A holds. Then the sequence $\{\nu_j\}$ generated by Algorithm 1 satisfies*

$$\liminf_{j\to\infty} \|\hat{g}_j\| = 0, \tag{55}$$

or some $\{\nu_j\}$ satisfies the termination criterion and the algorithm terminates.

Proof. For the purpose of deriving a contradiction, assume that (55) does not hold, i.e., there exists a constant ϵ such that $\|\hat{g}_j\| \geq \epsilon$. It follows from (50) that

$$\Delta_j > \epsilon/G_j \tag{56}$$

for sufficiently large j. On the other hand, It is clear from (39) that

$$(1-\eta)\psi\epsilon\min\{\Delta_j, \epsilon/G_j\} \le D_j - D_{j+1}. \tag{57}$$

Combining (56) and (57), we obtain

$$\sum_{j=1}^{\infty}(1-\eta)\psi\epsilon^2/G_j \le \sum_{j=1}^{\infty}(D_j - D_{j+1}).$$

Noticing that $Q_1 = D_1$, $Q_j < D_j$ and Q_j is bounded below, we have

$$\sum_{j=1}^{\infty} 1/G_j < \infty. \tag{58}$$

Relation (58) contradicts (54). Hence (55) holds. □

We now prove that the algorithm globally converges to a point satisfying the second-order necessary optimality conditions. We first present three important propositions, they are similar to the corollaries given in [9].

Proposition 1. *Assume that the iterates $\{x_j\}$ generated by Algorithm 1 lie in a compact region, and that the sequences $\{M_j\}$ and $\{M_j^{-1}\}$ are bounded. Then*
(*a*) *$\{W_j\}$, $\{W_j^{-1}\}$, $\{T_j\}$ and $\{T_j^{-1}\}$ are bounded;*
(*b*) *$\hat{g}_j = \Theta(\|g_j\|)$; and*
(*c*) *$\hat{B}_j = \Theta(\|B_j\|)$ with $\zeta_{\min}(\hat{B}_j) = \Theta(\zeta_{\min}(B_j))$.*

Proof. Similar to the proof of Corollary 4.2 in [9]. □

Proposition 2. *Assume that the iterates $\{x_j\}$ generated by Algorithm 1 lie in a compact region. Then the sequence $\{(x_j, \lambda_j)\}$ lies in the interior of a region within which $g(x,\lambda)$ is uniformly continuous. Moreover, $Q(x,\lambda)$ is also uniformly continuous in the same region and hence $\{B_j\}$ is bounded.*

Proof. Similar to the proof of Lemma 4.6 in [9]. □

Proposition 3. *If s_j satisfies the trust region termination condition (13) then $-\phi(s_j) \ge \frac{1}{2}\tau\sigma_j\Delta_j^2$.*

Proof. Similar to the proof of Corollary 4.3 in [9]. □

Under the assumption that the iterations lie in a compact region, Proposition 1 implies that it is no longer necessary to distinguish between g_j and $\hat{g}_j$ or B_j and $\hat{B}_j$ in the convergence results.

Theorem 2. *Assume that sequence $\{x_j\}$ generated by Algorithm 1 lies in a compact region, and that $\lim_{j\to\infty}\|B_j - \nabla^2 Q(\nu_j)\| = 0$. Suppose that for each j, the step s_j satisfies the termination criteria (13). Then either some ν_j satisfies the termination criteria or $\limsup_{j\to\infty}\zeta_{\min}(\nabla^2 Q(\nu_j)) \ge 0$.*

Proof. It follows from (54) that $\lim_{j\to\infty} 1/G_j \neq 0$. Thus we deduce from Lemma 5 that $\lim_{j\to\infty} \Delta_j = 0$, this relation implies that $\rho_j < \varpi$ for all j sufficiently large. Otherwise, by the updating rules of Algorithm 1, it follows $\Delta_{j+1} \geq \Delta_j$, it is not possible for $\Delta_j \to 0$.

Due to the definition of ρ_j and the parameter $\varpi \in (0,1)$, relation $\rho_j < \varpi$ leads to $D(\nu_j) - Q(\nu_j + s_j) \leq -\phi(s_j)$, that is, $0 \leq Q(\nu_j + s_j) - D(\nu_j) - \phi(s_j)$. Using the monotonicity of $D(\nu_j)$ and the fact that $Q(\nu_j) < D(\nu_j)$, we obtain that $Q(\nu_j + s_j) - D(\nu_j) \leq Q(\nu_j + s_j) - Q(\nu_j)$. Consequently, we have $0 \leq Q(\nu_j + s_j) - D(\nu_j) - \phi(s_j) \leq Q(\nu_j + s_j) - Q(\nu_j) - \phi(s_j)$.

Suppose, for the purpose of deriving a contradiction, that $\{\sigma_j\}$ is bounded away from zero. By Taylor's theorem,

$$\begin{aligned}
&|Q(\nu_j + s_j) - D(\nu_j) - \phi(s_j)| \\
&\leq |Q(\nu_j + s_j) - Q(\nu_j) - \phi(s_j)| \\
&= |Q(\nu_j + s_j) - Q(\nu_j) - g_j^T s_j - \frac{1}{2} s_j^T B_j s_j| \\
&\leq \frac{1}{2}\|s_j\|^2 (\max_{0\leq\varsigma\leq 1} \|\nabla^2 Q(\nu_j + \varsigma s_j) - \nabla^2 Q(\nu_j)\| + \|B_j - \nabla^2 Q(\nu_j)\|) \\
&\leq \frac{1}{2}\|T_j^{-1}\|^2 \Delta_j^2 (\max_{0\leq\varsigma\leq 1} \|\nabla^2 Q(\nu_j + \varsigma s_j) - \nabla^2 Q(\nu_j)\| + \|B_j - \nabla^2 Q(\nu_j)\|).
\end{aligned}$$

Dividing both sides of this expression by $-\phi(s_j)$ and using Proposition 3 yields

$$|\rho_j - 1| \leq \frac{\|T_j^{-1}\|^2}{\tau\sigma_j} (\max_{0\leq\varsigma\leq 1} \|\nabla^2 Q(\nu_j + \varsigma s_j) - \nabla^2 Q(\nu_j)\| + \|B_j - \nabla^2 Q(\nu_j)\|),$$

here we use equality $|-\rho_j + 1| = |\rho_j - 1|$. Using the assumption that sequence $\{\sigma_j\}$ is bounded away from zero, the fact that $\{T_j^{-1}\}$ is bounded and $\nabla^2 Q(\nu)$ is uniformly continuous, and noticing that $\lim_{j\to\infty} \|B_j - \nabla^2 Q(\nu_j)\| = 0$, we can deduce from the above inequality that

$$\lim_{j\to\infty} |\rho_j - 1| = 0.$$

This result contradict $\rho_j < \varpi$ for all j sufficiently large. Thus the assumption that $\{\sigma_j\}$ is bounded away from zero is not true. Since subproblem (9) is solved inaccurately, $\sigma_j \geq -\zeta_{\min}(B_j)$, it must hold that $\limsup_{j\to\infty} \zeta_{\min}(B_j) \geq 0$. Due to $\|B_j - \nabla^2 Q(\nu_j)\| \to 0$, it follows that $\limsup_{j\to\infty} \zeta_{\min}(\nabla^2 Q(\nu_j)) \geq 0$. □

Theorem 1 is a crucial property of Algorithm 1, which gives us major implications to establish the convergence of g_j. Before we establish the main result, we give an important proposition concerning the convergence of F^μ.

Proposition 4. *Assume that the sequences $\{M_j\}$ and $\{M_j^{-1}\}$ are bounded. If $\{f(x_j)\}$ is bounded below, $\{c_i(x_j)\}$ is bounded above for all $i \in I$, and $J_j^T(\lambda - \pi_j) = O(\|\lambda_j - \pi_j\|)$, then $\hat{g}_j = \Theta(\|F^\mu(x_j, \lambda_j)\|)$.*

Proof. Similar to the proof of Theorem 4.1 in [9]. □

Now, our purpose is to show the main convergence result that $\lim_{j\to\infty}\|g_j\| = 0$ holds based on the result $\liminf_{j\to\infty}\|\hat{g}_j\| = 0$. To this end, we give further lemmas, partial proof techniques motivated by [9].

Lemma 8. *Let $\{\nu_j\}$ be the sequence of iterates generated by Algorithm 1 and let $\{\alpha_j\}$ be the corresponding sequence of step lengths. Then*

$$D(\nu_j) - D(\nu_j + \alpha_j s_j) \geq \delta\alpha_{\min}\tau\|\hat{g}_j\|\min\{\Delta_j, \|\hat{g}_j\|/\|\hat{B}_j\|\},$$

where $\alpha_{\min} = (1-\delta)\xi r/R$.

Proof. It follows from (17) and (31) that $D(\nu_j) - Q(\nu_j + \alpha_j s_j) \geq -\delta\alpha_{\min}\hat{g}_j^T\hat{s}_j$, which together with the fact $D(\nu_{j+1}) > Q(\nu_{j+1})$ suggest that $D(\nu_j) - D(\nu_j + \alpha_j s_j) \geq -\delta\alpha_{\min}\hat{g}_j^T\hat{s}_j$. Thus, it follows from (26) that conclusion holds. □

Lemma 9. *Assume that sequence $\{x_j\}$ generated by Algorithm 1 lies in a compact region. Choose $\varrho_1 > \varrho_2 > 0$. For every ϱ_3 there is an l_0 sufficiently large that $\sum_{j=p}^{q-1}\|\alpha_j s_j\| < \varrho_3$ for all indices $q > p > l_0$ for which $\|g(\nu_p)\| > \varrho_1$, $\|g(\nu_j)\| > \varrho_2$ for consecutive indices $j = p, p+1, \cdots, q-1$, and $\|g(\nu_q)\| < \varrho_2$.*

Proof. Note that if $\|g(\nu_j)\| > \varrho_1$ holds only finitely often, then the lemma is trivially true. Furthermore, as $\liminf_{j\to\infty}\|g(\nu_j)\| = 0$, it must hold that for every iteration p such that $\|g(\nu_p)\| > \varrho_1$, there must exist a subsequent iteration q such that $\|g(\nu_q)\| < \varrho_2$.

Let $\bar{\alpha} = 2\alpha_{\min}$. Lemma 8 implies that for every index in the set $A = \{j : \alpha_j \geq \bar{\alpha}\}$, we obtain$D(\nu_j) - D(\nu_j + \alpha_j s_j) \geq \frac{1}{2}\delta\bar{\alpha}\tau\|\hat{g}_j\|\min\{\Delta_j, \|\hat{g}_j\|/\|\hat{B}_j\|\}$. Let $C(p) = \{j : j \geq p \ \text{and} \ \|g(\nu_j)\| > \varrho_2\}$. Because $\sum_{j=1}^{\infty}(D(\nu_j) - D(\nu_{j+1})) < \infty$. Therefore, $\sum_{j\in A\cap C(p)}\|\hat{g}(\nu_j)\|\min\{\Delta_j, \|\hat{g}_j\|/\|\hat{B}_j\|\} < \infty$. By proposition 1, $\hat{g}_j = \Theta(\|g_j\|)$ and $\hat{B}_j = \Theta(\|B_j\|)$. As $\{B_j\}$ is bounded, and $\|g(\nu_j)\| > \varrho_2$ for $j \in C(p)$, it follows that

$$\sum_{j\in A\cap C(p)}\Delta_j < \infty. \tag{59}$$

Let J denote the sequence of iteration indices $\{p, p+1, \cdots, q-1\}$. Let $\{j_k\}_{k=1}^{v}$ denote the subsequence of J with indices in A. We partition J into $v+1$ nonoverlapping subsequences $P_0, P_1, \cdots, P_v$ with $P_0 = \{p, p+1, \cdots, j_1 - 1\}$, $P_k = \{j_k, j_k+1, \cdots, j_{k+1}-1\}$, $k = 1, 2, \cdots, v-1$ and $P_v = \{j_v, j_v+1, \cdots, q-1\}$. Note that if the first index p is in A, then P_0 is empty. Otherwise none of the indices of P_0 will be in A. For $k > 0$, the sequence P_k starts with an index in A, followed by a (possibly empty) sequence of indices that are not in A. These definitions allow us to write the quantity to be bounded as

$$\sum_{j=p}^{q-1}\|\alpha_j s_j\| = \sum_{j\in P_0}\|\alpha_j s_j\| + \sum_{k=1}^{v}\sum_{j\in P_k}\|\alpha_j s_j\|. \tag{60}$$

We now estimate the quantity $\sum_{j\in P_0}\|\alpha_j s_j\|$. If the set P_0 is empty, then $\sum_{j\in P_0}\|\alpha_j s_j\| = 0$. Otherwise, $\alpha_j < \bar{\alpha}$ for every $j \in P_0$ and the rules for updating the trust region radius give $\Delta_{j+1} \le c_1\Delta_j$. This gives the sequence of inequalities

$$\sum_{j\in P_0}\|\alpha_j s_j\| \le \sum_{j\in P_0}\bar{\alpha}\Delta_j \le \bar{\alpha}\Delta_p\sum_{i=0}^{\infty}(c_1)^i \le \bar{\alpha}(\frac{1}{1-c_1})\Delta_p \le \bar{\alpha}(\frac{1}{1-c_1})c_2\Delta_{\max},$$

where $\Delta_{\max} = \max_{1\le i\le j_1-1}\Delta_i$. Since $\bar{\alpha} = 2\alpha_{\min} = 2(1-\delta)\xi r/R$, we can choose r sufficiently small and R sufficiently large, such that

$$\bar{\alpha}(\frac{1}{1-c_1})c_2\Delta_{\max} \le \frac{1}{2}\varrho_3, \tag{61}$$

for any $\varrho_3 > 0$.

We proceed to estimate the quantity $\sum_{k=1}^{v}\sum_{j\in P_k}\|\alpha_j s_j\|$, we first consider the terms involving the indices in P_k for $k > 0$. For $j \in P_k\backslash\{j_k\}$, that is, for every element of P_k except the first, $\alpha_j < \bar{\alpha}$ and the rules for updating the trust region radius give $\Delta_{j+1} \le c_1\Delta_j$. Thus

$$\sum_{j\in P_k\backslash\{j_k\}}\Delta_j \le \Delta_{j_k+1}\sum_{i=0}^{\infty}(c_1)^i = \frac{1}{1-c_1}\Delta_{j_k+1}. \tag{62}$$

At iteration j_k, it is possible that the trust region radius will increase, which implies that the appropriate bound on Δ_{j_k+1} is $\Delta_{j_k+1} \le c_2\Delta_{j_k}$. Therefore, we have from (62) that

$$\sum_{j\in P_k}\Delta_j = \Delta_{j_k} + \sum_{j\in P_k\backslash\{j_k\}}\Delta_j \le \Delta_{j_k} + \frac{c_2}{1-c_1}\Delta_{j_k} = (1+\frac{c_2}{1-c_1})\Delta_{j_k}. \tag{63}$$

From the definition of P_k we obtain

$$\sum_{j\in P_k}\|\alpha_j s_j\| \le \sum_{j\in P_k}\Delta_j \le (1+\frac{c_2}{1-c_1})\Delta_{j_k}. \tag{64}$$

Note that $j_k \in A\cap C(p)$, (64) gives

$$\sum_{k=1}^{v}\sum_{j\in P_k}\|\alpha_j s_j\| \le (1+\frac{c_2}{1-c_1})\sum_{k=1}^{v}\Delta_{j_k} \le (1+\frac{c_2}{1-c_1})\sum_{j\in A\cap C(p)}\Delta_j.$$

Consider the bound (59), p can be chosen sufficiently large such that

$$(1+\frac{c_2}{1-c_1})\sum_{j\in A\cap C(p)}\Delta_j < \frac{1}{2}\varrho_3. \tag{65}$$

Thus, we have from (61) and (65) that

$$\sum_{j=p}^{q-1}\|\alpha_j s_j\| = \sum_{j\in P_0}\|\alpha_j s_j\| + \sum_{k=1}^{v}\sum_{j\in P_k}\|\alpha_j s_j\| \le \frac{1}{2}\varrho_3 + \frac{1}{2}\varrho_3 = \varrho_3,$$

where ϱ_3 may be chosen to be arbitrarily small. □

The following lemma concerns the limiting behavior of the gradient of the barrier-penalty function.

Lemma 10. *Assume that the sequence $\{x_j\}$ generated by Algorithm 1 lies in a compact region. Then $lim_{j\to\infty}\|\nabla L(x_j)\| = 0$.*

Proof. Similar to the proof of Lemma 4.12 in [9]. □

The merit function $Q(x, \lambda)$ may be written as $Q(x, \lambda) = L(x) + \Psi(x, \lambda)$, where $L(x)$ is the barrier-penalty function (2) and $\Psi(x, \lambda)$ is the proximity term

$$\Psi(x, \lambda) = -\mu \sum_{i\in I}(ln(\frac{c_i(x)\lambda_i}{\mu}) + \frac{\mu - c_i(x)\lambda_i}{\mu}) + \frac{1}{2\mu}\sum_{i\in E}(c_i(x) + \mu\lambda_i)^2.$$

The next lemma concerns the behavior of this proximity term when the norm of the merit function gradient behaves nonmonotonically in the limit.

Lemma 11. *Assume that* $\liminf_{j\to\infty}\|g(\nu_j)\| = 0$ *and that there exists a positive number ϱ_1 such that the relation $\|g(\nu_j)\| > \varrho_1$ holds infinitely often. Then there exists an $\bar{\varrho} > 0$, a positive ϱ_2 sufficiently small and an index l_0 sufficiently large that if $q > p \geq l_0$ with $\|g(\nu_q)\| > \varrho_1 > \varrho_2 > \|g(\nu_p)\|$, then $\Psi(x_q, \lambda_q) \geq \Psi(x_p, \lambda_p) + \bar{\varrho}$.*

Proof. Similar to the proof of Lemma 4.13 in [9]. □

Theorem 3. *Suppose that the iterates $\{x_j\}$ generated by Algorithm 1 lies in a compact region. Then either some ν_j satisfies the termination criteria and the algorithm terminates or* $\lim_{j\to\infty}\|g(\nu_j)\| = 0$.

Proof. We first show that $Q(\nu_j)$ is close to $D(\nu_j)$ for j sufficiently large. Due to the definition of $D(\nu_j)$, we obtain

$$D(\nu_{j+1}) - Q(\nu_{j+1}) = \eta(D(\nu_j) - D(\nu_{j+1})). \tag{66}$$

Since $D(\nu_j)$ decreases monotonically and bounded below, equality (66) suggests that $Q(\nu_j)$ will not bounded away from $D(\nu_j)$ too much for j sufficiently large. That is, $Q(\nu_j)$ is almost forced to decrease monotonically for j sufficiently large.

Now, assume that $\|g(\nu_j)\|$ does not converge to zero. Let $\varrho_1 > 0$ be a constant for which $\|g(\nu_j)\| > \varrho_1$ infinitely often. Given any ϱ_2 such that $\varrho_1 > \varrho_2 > 0$, let p be any index such that $\|g(\nu_p)\| < \varrho_2$, and let q denote the next index greater than p with satisfying $Q(v_q) < Q(v_p)$ and $\|g(\nu_q)\| > \varrho_1$. Similarly, let r be the next index greater than q such that $Q(v_r) < Q(v_q)$ and $\|g(\nu_r)\| < \varrho_2$. Based on this definitions, we can apply Lemma 11 to assert that for p sufficiently large, the proximity term satisfies $\Psi(x_q, \lambda_q) \geq \Psi(x_p, \lambda_p) + \bar{\varrho}$. Note that $Q(v_q) < Q(v_p)$, it must hold that $L(x_q) < L(x_p) - \bar{\varrho}$. By Lemma 9, we may choose p large enough such that $\sum_{j=q}^{r} = \|\alpha_j s_j\|$ arbitrarily small. Thus, since $L(x)$ is uniformly continuous, it must hold that for all sufficiently large choices of p, $|L(x_r) - L(x_q)| < \frac{1}{2}\bar{\varrho}$ and hence $L(x_r) < L(x_p) - \frac{1}{2}\bar{\varrho}$. This implies that each time $\|g(\nu_j)\|$ increases from a value less than ϱ_2 to a value greater than ϱ_1 and then decreases again to a value less than ϱ_2, the barrier-penalty function $L(x)$ must decrease by at least a constant factor. As $L(x)$ is bounded below in a compact region, the value of $\|g(\nu_j)\|$ can exceed ϱ_1 only finitely often. This leads to $\lim_{j\to\infty} \|g(\nu_j)\| = 0$ as ϱ_1 is an arbitrary positive constant. □

4. Concluding remarks

We presented a primal-dual interior point nonmonotone trust region algorithm for solving asset liability management model whose deterministic equivalent problem formulated as (1). A primal-dual interior point algorithm for solving multistage stochastic programming problems through their corresponding unconstraint optimization problems has been discussed in Zhao [18], but their algorithm was designed for multistage stochastic convex programming. Our algorithm was designed for general nonlinear multistage stochastic programming problems. We focused on discussing the theoretical results of the inner iterations. We will further study the rules for updating μ_k and the convergence of the outer iterations, as well as numerical tests on real large-scale asset liability problems.

References

1. S. Bakhtiari and A. L. Tits, *A simple primal-dual feasible interior-point method for nonlinear programming with monotone descent*, Comput. Optim. Appl. **25** (2003), 17-38.
2. A. Berkelaar, J. A. S. Gromicho, R. Kouwenberg and S. Zhang, *A primal-dual decomposition algorithm for multistage stochastic convex programming*, Math. Prog. **104** (2005), 153-177.
3. A. R. Conn, N. I. M. Gould, D. Orban and P. L. Toint, *A primal-dual trust-region algorithm for non-convex nonlinear programming*, Math. Prog. **87** (2000), 215-249.
4. A. Consiglio, D. Saunders and S. A. Zenios, *Asset and liability management for insurance products with minimum guarantees: the UK case*, J. Bank. Finan. **30** (2006), 645-667.
5. J. Dominguez and M. D. González-Lima, *A primal-dual interior-point algorithm for quadratic programming*, Numer. Algor. **42** (2006), 1-30.
6. A. V. Fiacco and G. P. McCormick, *Nonlinear programming: sequential unconstrained minimization techniques*, John Wiley and Sons, New York, (1968).
7. A. Forsgren, P. E. Gill and M. H. Wright, *Interior methods for nonlinear optimization* , SIAM Rev. **44** (2002), 525-597.
8. A. Forsgren and P. E. Gill, *Primal-dual interior methods for nonconvex nonlinear programming* , SIAM J. Optim. **8** (1998), 1132-1152.
9. E. M. Gertz and P. E. Gill, *A primal-dual trust region algorithm for nonlinear optimization*, Math. Prog. **100** (2004), 49-94.
10. E. M. Gertz, *Combination trust-region line search methods for unconstrained optimization*, PhD thesis, University of California, San Diego, (1999).
11. J. Gondzio and A. Grothey, *Parallel interior-point solver for structured quadratic programs: application to financial planning problems*, Ann. Oper. Res. **152** (2007), 319-339.
12. J. Gondzio and A. Grothey, *Solving non-linear portfolio optimization problems with the primal-dual interior point method*, Euro. J. Oper. Res. **181** (2007), 1019-1029.

13. N. Gu and J. Mo, *Incorporating nonmonotone strategies into trust region method for unconstrained optimization*, Comput. Math. Appl. **55** (2008), 2158-2172.
14. P. Hilli, M. Koivu, T. Pennanen and A. Ranne, *A stochastic programming model for asset liability management of a finnish pension company*, Ann. Oper. Res. **152** (2007), 115-139.
15. J. J. Moré and D. C. Sorensen, *Computing a trust region step*, SIAM J. Sci. Statist. Comput. **4** (1983), 553-572.
16. J. Nocedal and Y. Yuan, *Combining trust region and line search techniques*, In: Y. Yuan (Eds)*Advances in Nonlinear Programming*, Kluwer Academic, Dordrecht, (1996), 153-175.
17. M. Ulbrich, S. Ulbrich and L. N. Vicente, *A globally convergent primal-dual interior-point filter method for nonlinear programming*, Math. Prog. **100** (2004), 379-410.
18. G. Zhao, *A lagrangian dual method with self-concordant barriers for multi-stage stochastic convex programming*, Math. Prog. **102** (2005), 1-24.
19. W. T. Ziemba and J. M. Mulvey (Ed.), *Worldwide asset and liability modeling*, Cambridge University Press, Cambridge, (1998).

Nengzhu Gu received M.Sc. in Guangxi University. Ph.D of School of Business, University of Shanghai for Science and Technology. His research interest is financial computation.

School of Business, University of Shanghai for Science and Technology, Shanghai 200093, China

e-mail: gnzemail@hotmail.com

Yan Zhao Ph.D of School of Science, University of Shanghai for Science and Technology. Her research interest is nonlinear science and its application.

e-mail: zhaoyanem@hotmail.com