

AN AFFINE SCALING INTERIOR POINT BACKTRACKING ALGORITHM FOR NONLINEAR CONSTRAINED OPTIMISATION

DETONG ZHU¹

(Received 30 August, 2002; revised 7 January, 2003)

Abstract

In this paper we propose a new affine scaling interior trust region algorithm with a nonmonotonic backtracking technique for nonlinear equality constrained optimisation with nonnegative constraints on the variables. In order to deal with large problems, the general full trust region subproblem is decomposed into a pair of trust region subproblems in horizontal and vertical subspaces. The horizontal trust region subproblem in the algorithm is defined by minimising a quadratic function subject only to an ellipsoidal constraint in a null tangential subspace and the vertical trust region subproblem is defined by the least squares subproblem subject only to an ellipsoidal constraint. By adopting Fletcher's penalty function as the merit function, combining a trust region strategy and a nonmonotone line search, the mixing technique will switch to a backtracking step generated by the two trust region subproblems to obtain an acceptable step. The global convergence of the proposed algorithm is proved while maintaining a fast local superlinear convergence rate, which is established under some reasonable conditions. A nonmonotonic criterion is used to speed up the convergence progress in some highly nonlinear cases.

1. Introduction

In this paper we analyse the trust region interior point algorithm for solving the following nonlinear equality constrained optimisation problem with nonnegative constraints on the variables:

$$\min f(x) \quad \text{s.t.} \quad \begin{cases} c(x) = 0, \\ x \geq 0, \end{cases} \quad (1.1)$$

where $f : \mathbb{R}^n \rightarrow \mathbb{R}$ is a smooth nonlinear function, not necessarily convex and $c(x) : \mathbb{R}^n \rightarrow \mathbb{R}^m$, ($m \leq n$) is a vector-valued nonlinear function. There are quite a

¹Department of Mathematics, Shanghai Normal University, Shanghai 200234, China; e-mail: dtzhu@shtu.edu.cn.

© Australian Mathematical Society 2004, Serial-fee code 1446-1811/04

few articles proposing sequential quadratic programming methods with a trust region approach for this problem subject only to nonlinear equality constraints and no variable inequality constraints (see [8, 9, 11]). Most existing methods for solving optimisation problems subject only to linear inequality constraints generate sequences of points in the interior of the feasible set with the linear inequality constraints. Recently, Coleman and Li [2] presented a trust region interior point algorithm for minimisation problems with only simple bounds on the variables. Bonnans and Pola [1] also proposed an interior trust region algorithm for the linear constrained $c(x) = Ax - b$ optimisation (1.1) where the matrix $A \in \mathbb{R}^{m \times n}$ and the vector $b \in \mathbb{R}^m$. A search direction of the algorithm at the k -th iteration is determined by solving the trust region convex quadratic programming problem

$$\min q_k(d) \stackrel{\text{def}}{=} g_k^T d + \frac{1}{2} d^T B_k d \quad \text{s.t.} \quad \begin{cases} Ad = 0, \\ d^T X_k^{-2} d \leq \Delta_k^2, \\ x_k + d > 0, \end{cases} \quad (1.2)$$

where $g_k = \nabla f(x_k)$, $d = x - x_k$, $q_k(d)$ is the local quadratic approximation of f and Δ_k is the trust region radius. Here B_k is a symmetric approximation of the Hessian of the objective function and $X_k \stackrel{\text{def}}{=} \text{diag}\{x_k^1, \dots, x_k^n\}$ is a scaling matrix, where x_k^i is the i -th component of $x_k > 0$, the current interior feasible point. Let d_k be the solution of the subproblem. Then Bonnans and Pola [1] used a line search to compute $\alpha_k = \omega^{l_k}$, where l_k is the smallest nonnegative integer such that

$$f(x_k + \omega^{l_k} d_k) \leq f(x_k) + \omega^{l_k} \beta q_k(d_k), \quad (1.3)$$

with $\beta \in (0, 1/2)$ and $\omega \in (0, 1)$. Finally,

$$x_{k+1} = x_k + \alpha_k d_k.$$

However, Bonnans and Pola [1] only used the line search restricted to the trust region and assumed that $q_k(d_k)$ is a convex function in order to obtain global convergence of the proposed algorithm.

The trust region method is a well-accepted technique in nonlinear optimisation to assure global convergence. However, the search direction generated in the trust region subproblem (1.2) must satisfy strict feasibility, that is, $x_k + d > 0$, which results in computational difficulties. It is possible that the trust region subproblem with a strict feasibility constraint needs to be resolved many times before obtaining an acceptable step, and hence the total computational effort for completing one iteration might be expensive and difficult. The idea of the trust region and line search backtracking technique suggested in (1.3) (also see Nocedal and Yuan [6]) motivates us to switch to the line search technique by employing a trial step which may be unacceptable in

the trust region method, since the trial step should provide a direction of sufficient descent. Recently, the nonmonotone technique has been developed to combine the line search technique with a trust region strategy for solving unconstrained optimisation (see [3] and [4], for instance). The nonmonotonic idea also motivates further study of the affine scaling backtracking trust region interior point algorithm for problem (1.1), because monotonicity may cause a series of very small steps if the contours of the objective function f are highly nonlinear and give rise to a family of curves with large curvature. The general full trust region subproblem will be decomposed into a pair of trust region subproblems in horizontal and tangential subspaces of nonlinear equality constraints. The trust region horizontal subproblem in the proposed algorithm is defined by minimising a quadratic function subject only to an ellipsoidal constraint in the null subspace of the tangential subspace of the equality constraints, and the trust region vertical subproblem is defined by the least squares problem subject only to an ellipsoidal constraint. It is clear that the two subproblems are easy to solve.

The paper is organised as follows. In Section 2, we describe the algorithm which combines the techniques of trust region, interior point, nonmonotonic backtracking search and affine scaling. In Section 3, weak global convergence of the proposed algorithm is established, while some further convergence properties such as strong global convergence and the local convergence rate are discussed in Section 4.

2. Algorithm

In this section, by introducing Fletcher's penalty function as a merit function, we propose a new interior trust region method with a nonmonotonic backtracking technique for nonlinear constrained optimisation (1.1) with nonnegativity constraints on the variables. The trust region subproblem for the tangential space of the nonlinear equality constraints involves choosing a scaling matrix and a quadratic approximation model of the objective function. We motivate our choice of scaling matrix by examining the optimality conditions for problem (1.1).

The basic idea can be summarised as follows: Let

$$g(x) \stackrel{\text{def}}{=} \nabla f(x) \quad \text{and} \quad A(x) \stackrel{\text{def}}{=} \nabla c(x) = [\nabla c_1(x), \dots, \nabla c_m(x)] \in \mathbb{R}^{n \times m}.$$

Assume that $A(x)$ has full column rank. Let

$$l(x, \lambda, \mu) = f(x) - \lambda^T c(x) - \mu^T x$$

be the Lagrange function of problem (1.1), where the Lagrange multipliers $\lambda \in \mathbb{R}^m$, $0 \leq \mu \in \mathbb{R}^n$. Optimality conditions for problem (1.1) are well established. Assuming feasibility, first-order necessary conditions for x_* to be a local minimiser are that there

exist $0 \leq \mu_* \in \mathbb{R}^n$ and $\lambda_* \in \mathbb{R}^m$ such that

$$g_* - A_*\lambda_* - \mu_* = 0, \tag{2.1}$$

$$\mu_*^T x_* = 0, \tag{2.2}$$

$$c(x_*) = 0.$$

Equivalently, (2.1)–(2.2) imply

$$\begin{cases} (g_* - A_*\lambda_*)_i = 0, & \text{if } (x_*)_i > 0, \\ (g_* - A_*\lambda_*)_i \geq 0, & \text{if } (x_*)_i = 0, \end{cases}$$

where $(g_* - A_*\lambda_*)_i$ and $(x_*)_i$ are the i -th component of $(g_* - A_*\lambda_*)$ and x_* , respectively. Strict complementarity is said to hold at x_* if at least one of the two inequalities $(x_*)_i > 0$ and $(\mu_*)_i > 0$, that is, $|(x_*)_i| + |(\mu_*)_i| > 0, i = 1, \dots, m$, where $(\mu_*)_i$ is the i -th component of μ_* . We now define a vector function $\gamma(x) : \mathbb{R}^n \rightarrow \mathbb{R}^n$ with the i -th component of the vector function defined componentwise as follows:

$$\gamma_i(x) \stackrel{\text{def}}{=} \begin{cases} x_i, & \text{if } (g(x) - A(x)\lambda)_i \geq 0, \\ -1, & \text{if } (g(x) - A(x)\lambda)_i < 0. \end{cases}$$

We define $D(x) \stackrel{\text{def}}{=} \text{diag}\{|\gamma_1(x)|^{-1/2}, \dots, |\gamma_n(x)|^{-1/2}\}$, which arises naturally from examining the first-order necessary conditions for (1.1)

$$D^{-2}(x)\{g(x) - A(x)\lambda\} = 0, \tag{2.3}$$

where the Lagrange multiplier λ is the solution vector of the least squares problem $\min_\lambda \|A(x)\lambda - g(x)\|_{D(x)^{-2}}$. We have that $\lambda(x)$ can be obtained by solving the normal equation $D(x)^{-1}A(x)\lambda = D(x)^{-1}g(x)$, that is, if we set $\bar{A}(x) \stackrel{\text{def}}{=} D(x)^{-1}A(x)$, $\bar{g}(x) \stackrel{\text{def}}{=} D(x)^{-1}g(x)$, then

$$\lambda(x) = [\bar{A}(x)^T \bar{A}(x)]^{-1} \bar{A}(x) \bar{g}(x). \tag{2.4}$$

So, (2.3) implies that

$$D^{-1}(x)(I - \bar{A}(x)[\bar{A}(x)^T \bar{A}(x)]^{-1} \bar{A}(x)^T) \bar{g}(x) = 0.$$

Setting $P(x) \stackrel{\text{def}}{=} I - \bar{A}(x)[\bar{A}(x)^T \bar{A}(x)]^{-1} \bar{A}(x)^T$, where I is the identity matrix on \mathbb{R}^n , we have that

$$D^{-1}(x)[P(x)\bar{g}(x)] = 0, \tag{2.5}$$

which implies $P(x)\bar{g}(x) = 0$. We define the following sets:

$$\mathcal{F} \stackrel{\text{def}}{=} \{x \in \mathbb{R}^n \mid c(x) = 0, x \geq 0\} \quad \text{and} \quad \mathcal{F}^0 \stackrel{\text{def}}{=} \{x \in \mathbb{R}^n \mid c(x) = 0, x > 0\}$$

so that \mathcal{F} is the set of feasible points and \mathcal{F}^0 is the set of ‘strictly feasible’ points of the variable nonnegative constraints. In what follows, we assume that \mathcal{F} is bounded and \mathcal{F}^0 is nonempty, and that the system (2.3) is continuous but not everywhere differentiable. Nondifferentiability for the system (2.3) occurs when $\gamma_i = 0$. Discontinuity of γ_i may also occur when $(g - A(x)\lambda)_i = 0$. Assume that $x_k \in \mathcal{F}^0$. Noting that $(A_k \lambda_k)^T d_k = 0$, a Newton step for (2.3) satisfies that

$$(D_k^{-2} \nabla^2 l(x_k, \lambda_k, \mu_k) + \text{diag}\{g_k - A_k \lambda_k\} J_k^\gamma) d_k = -D_k^{-2} (g_k - A_k \lambda_k), \tag{2.6}$$

where

$$\text{diag}\{g_k - A_k \lambda_k\} \stackrel{\text{def}}{=} \text{diag}\{(g_k - A_k \lambda_k)_1, \dots, (g_k - A_k \lambda_k)_n\}$$

with $(g_k - A_k \lambda_k)_i$ being the i -th component of $g_k - A_k \lambda_k$, and where $J^\gamma(x) \in \mathbb{R}^{n \times n}$ is the Jacobian matrix of $|\gamma(x)|$ whenever $|\gamma(x)|$ is differentiable. Each diagonal component of the diagonal matrix J^γ equals either zero or 1.

Based on the Newton step (2.6) for the system (2.3), we define our quadratic model as a quadratic programming problem with linearised equality constraints, that is, the quadratic model in the tangential subspace of the equality constraints is

$$\min \bar{\varphi}_k(\bar{d}) \stackrel{\text{def}}{=} \bar{g}_k^T \bar{d} + \frac{1}{2} \bar{d}^T (\bar{H}_k + \bar{C}_k) \bar{d} \quad \text{s. t.} \quad \bar{A}_k^T \bar{d} = 0,$$

where

$$\begin{aligned} \bar{g}_k &\stackrel{\text{def}}{=} D_k^{-1} g_k, & \bar{A}_k &\stackrel{\text{def}}{=} D_k^{-1} A_k, \\ \bar{H}_k &\stackrel{\text{def}}{=} D_k^{-1} (\nabla_{xx}^2 l_k) D_k^{-1}, & \bar{C}_k &\stackrel{\text{def}}{=} \text{diag}\{g_k - A_k \lambda_k\} J_k^\gamma. \end{aligned} \tag{2.7}$$

Therefore, based on (2.5), the trust region subproblem in the tangential subspace is as follows:

$$\min \bar{\varphi}_k(\bar{d}) \stackrel{\text{def}}{=} \bar{g}_k^T \bar{d} + \frac{1}{2} \bar{d}^T (\bar{H}_k + \bar{C}_k) \bar{d} \quad \text{s. t.} \quad \begin{cases} \bar{A}_k^T \bar{d} = 0, \\ \|\bar{d}\| \leq \Delta_k. \end{cases} \tag{\bar{S}_k}$$

Let $d^h = D_k^{-1} \bar{d}$, then the subproblem (\bar{S}_k) is equivalent to the following problem in the original variable space:

$$\min \varphi_k(d) \stackrel{\text{def}}{=} g_k^T d^h + \frac{1}{2} (d^h)^T B_k d^h \quad \text{s. t.} \quad \begin{cases} A_k^T d^h = 0, \\ \|D_k d^h\| \leq \Delta_k, \end{cases} \tag{S_k}$$

where B_k is $\nabla^2 l_k + D_k \text{diag}\{g_k - A_k \lambda_k\} J_k^\gamma D_k$ or its approximate and $\|\cdot\|$ is the l_2 -norm.

Based on the solution of the trust region subproblem (S_k) , we can obtain the following lemma (see [10]).

LEMMA 2.1. *The step d_k^h is a solution of subproblem (S_k) if and only if there exist $0 \leq \mu_k \in \mathbb{R}^n, \lambda_k \in \mathbb{R}^m$, such that*

$$\begin{aligned} (B_k + \mu_k D_k^2) d_k^h &= -g_k + A_k \lambda_k, \\ A_k^T d_k^h &= 0, \\ \mu_k (\Delta_k - \|D_k d_k^h\|) &= 0 \end{aligned} \tag{2.8}$$

holds and $B_k + \mu_k D_k^2$ is positive semidefinite in the null subspace $\mathcal{N}(A_k^T)$.

Lemma 2.1 establishes the necessary and sufficient conditions concerning d_k^h, μ_k and λ_k when d_k^h solves the subproblem (S_k) . Lemma 2.1 also implies that $P_k \bar{g}_k = 0$ if and only if $d_k^h = 0$ is a solution of the subproblem (S_k) .

In a current iteration x_k of the general trust region algorithm, the full quadratic subproblem will refer to solving

$$\min g_k^T p + \frac{1}{2} p^T W_k p \quad \text{s.t.} \quad \begin{cases} A_k^T p + c_k = 0, \\ \|p\| \leq \Delta_k, \end{cases}$$

where $W_k \stackrel{\text{def}}{=} \nabla_{xx}^2(x_k, \lambda_k, \mu_k)$ is the Hessian of the Lagrange function.

However, as is well known, restricting the size of a step with the trust region radius may preclude us from satisfying the linearised constraints. Therefore the full trust region subproblem is decomposed into two subproblems in horizontal and vertical subspaces in the proposed algorithm. In a current iteration x_k of the proposed algorithm, we first solve the horizontal (tangential) trust region subproblem (S_k) , and then compute a step that lies well within the trust region and satisfies the linearised equality constraint $A_k^T p + c_k = 0$ as much as possible. It motivates the solution of the linearised equality constraint which is done by defining the vertical (or normal) subproblem

$$\min \|A_k^T d^v + c_k\| \quad \text{s.t.} \quad \|d^v\| \leq \Delta_k. \tag{2.9}$$

This problem is a least squares problem with an ellipsoidal constraint. It may have many solutions, but the solution d_k^v is always chosen to lie in the range subspace $\mathcal{R}(A_k)$ of A_k . We now define the total step of our trust region strategy as follows:

$$d_k = d_k^h + d_k^v.$$

Then we set $x_{k+1} = x_k + d_k$.

In order to decide the acceptance of the new point at each iteration and to adjust the trust region radius, introducing a merit function is necessary, since the nonlinear equality constraints may be infeasible while the nonnegative constraints are feasible. Fletcher’s penalty function is introduced as follows:

$$F(x, \rho) = f(x) - \lambda(x)^T c(x) + \rho \|c(x)\|^2, \tag{2.10}$$

where ρ is a penalty parameter. Since each new point at each iteration is always a feasible interior point for the nonnegativity constraints, the penalty function does not refer to the penalty of nonnegative constraints. Since this function is differentiable we have

$$\nabla_x F(x, \rho) = \nabla f(x) - \nabla \lambda(x)^T c(x) - A(x)\lambda(x) + 2\rho A(x)c(x), \quad (2.11)$$

where $\nabla \lambda(x)$ is the $m \times n$ matrix whose entries are the gradients of the Lagrange multiplier estimates. Thus, at the k -th iteration, in order to avoid $\nabla \lambda(x)$ in (2.11), we used an approximate formula, suggested by Powell and Yuan in [7]. For the direction p_k and the size of the step t_k , the approximate formula $\nabla_x F(x_k, \rho_k)^T d_k \approx \phi'_k(t_k)$ is defined as follows:

$$\phi'_k(t_k) \equiv g_k^T p_k - \lambda_k^T A_k^T d_k - \frac{1}{t_k} [\lambda(x_k + t_k d_k) - \lambda_k]^T c_k + 2\rho_k (A_k^T d_k) c_k. \quad (2.12)$$

In our algorithm the penalty parameter ρ should be updated after each iteration by the formulae below. Choose $\tau \in (0, 1/2)$ and for all k take

$$\begin{aligned} \psi_k &\stackrel{\text{def}}{=} \|A_k(A_k^T A_k)^{-1}\|, \quad b_k \stackrel{\text{def}}{=} \|D_k^{-1} B_k D_k^{-1}\|, \quad w_k \stackrel{\text{def}}{=} \|W(x_k, \lambda_k, \mu_k)\|, \\ L'_k &= \begin{cases} \frac{\|\lambda(x_k + t_k d_k) - \lambda_k\|}{\|t_k d_k\|}, & \text{if } t_k d_k \neq 0, \\ 0, & \text{otherwise,} \end{cases} \\ L''_k &= \begin{cases} \frac{\|\lambda(x_k + t_k d_k^h) - \lambda_k\|}{\|t_k d_k^h\|}, & \text{if } t_k d_k^h \neq 0, \\ 0, & \text{otherwise,} \end{cases} \\ L_k &= \max\{L'_k, L''_k\}, \quad l_k = \max\{4L_k, 1\}, \quad \hat{l}_k = (1 - \tau)l_k + 2w_k, \\ \nu_k &= \max \left\{ \psi_k l_k, \frac{l_k^2}{2\omega_2 b_k}, \frac{l_k^2}{\omega_1 b_k}, \frac{\hat{l}_k^2}{(1 - 2\tau)^2} + \frac{\hat{l}_k \psi_k}{(1 - 2\tau)} \right\}. \end{aligned} \quad (2.13)$$

Then the update for the penalty parameter ρ_k is given by

$$\rho_k = \begin{cases} \rho_{k-1}, & \text{if } \rho_{k-1} \geq \nu_k, \\ \max\{\rho_{k-1} + \kappa, \nu_k\}, & \text{otherwise,} \end{cases} \quad (2.14)$$

where κ is a positive constant.

In order to use and adjust the trust region strategy, we consider the actual change of the merit function,

$$\text{Ared}_k(d_k) = F(x_k + d_k, \rho_k) - F(x_k, \rho_k), \quad (2.15)$$

and an estimate of this change,

$$\begin{aligned} \text{Pred}_k(d_k) &= g_k^T d_k^h + \frac{1}{2}(d_k^h)^T B_k d_k^h + \frac{1}{2}[\lambda(x_k + d_k^h) - \lambda_k]^T A_k^T d_k \\ &\quad - [\lambda(x_k + d_k) - \lambda_k]^T \left(c_k + \frac{1}{2} A_k^T d_k \right) \\ &\quad + \rho_k (\|c_k + A_k^T p_k\|^2 - \|c_k\|). \end{aligned} \tag{2.16}$$

Next we develop an affine scaling interior trust region algorithm which incorporates a nonmonotonic line search technique based on the trust region subproblems (S_k) and (2.9).

Algorithm. Initialisation step. Choose parameters $\beta \in (0, 1/2)$, $\tau \in (0, 1/2)$, $\omega \in (0, 1)$, $0 < \eta_1 < \eta_2 < 1$, $0 < \gamma_1 < \gamma_2 < 1 < \gamma_3$, $0 < \kappa_2 \leq 1/4 < \kappa_1 \leq 1/2$, $\epsilon \geq 0$, a positive integer M and a positive constant κ . Let $m(0) = 0$. Choose a symmetric matrix B_0 . Select an initial trust region radius $\Delta_0 > 0$ and a maximal trust region radius $\Delta_{\max} \geq \Delta_0$, give a starting point $x_0 \in \mathbb{R}^n$ and a starting penalty weight vector $\rho_0 \in \mathbb{R}$. Set $k = 0$, go to the main step.

Main step.

- (1) Evaluate $f_k = f(x_k)$, $g_k = \nabla f(x_k)$, c_k , D_k , A_k and $\bar{A}_k = D_k^{-1} A_k$.
- (2) If $\|P_k \bar{g}_k\| + \|c_k\| \leq \epsilon$, stop with the approximate solution x_k , where $P_k = I - \bar{A}_k (\bar{A}_k^T \bar{A}_k)^{-1} \bar{A}_k^T$ is the projection map in the null subspace of $\mathcal{N}(\bar{A}_k^T)$.
- (3) Compute the multiplier $\lambda_k = [\bar{A}_k^T \bar{A}_k]^{-1} \bar{A}_k^T \bar{g}_k$. Calculate the penalty parameter ρ_{k+1} by the updating formula (2.14).
- (4) Solve the horizontal subproblem (S_k) and (2.9). Denote by d_k^h the solution of the subproblem (S_k) and by d_k^v the solution of vertical subproblem (2.9), respectively.
- (5) Set $d_k = d_k^h + d_k^v$.
- (6) Calculate the predicted reduction $\text{Pred}_k(d_k)$, given in (2.16), and the approximate directional derivative $\phi'_k(t_k)$, given in (2.12), of the merit function $F(x, \rho)$ at the point x_k along the direction d_k .
- (7) If at least one of the two following conditions:

$$\phi'_k(t_k) \leq -\kappa_1 \|P_k \bar{g}_k\| \min \left\{ \frac{\|P_k \bar{g}_k\|}{b_k}, \Delta_k \right\} - \kappa \min \left\{ \|c_k\|, \frac{\Delta_k}{\psi_k} \right\} \tag{2.17}$$

and

$$\text{Pred}_k(d_k) \leq -\kappa_2 \|P_k \bar{g}_k\| \min \left\{ \frac{\|P_k \bar{g}_k\|}{b_k}, \Delta_k \right\} - \kappa \min \left\{ \|c_k\|, \frac{\Delta_k}{\psi_k} \right\}$$

does not hold, then set $x_{k+1} = x_k$ and $\Delta_k = \gamma_1 \Delta_k$.

- (8) Choose $t_k = 1, \omega, \omega^2, \dots$, until the following inequality is satisfied:

$$F(x_k + t_k d_k, \rho_k) \leq F(x_{l(k)}, \rho_{l(k)}) + \beta t_k \phi'_k(t_k), \quad \text{with } x_k + t_k d_k \geq 0, \tag{2.18}$$

where $F(x_{l(k)}, \rho_{l(k)}) = \max_{0 \leq j \leq m(k)} \{F(x_{k-j}, \rho_{k-j})\}$.

(9) Set

$$h_k = \begin{cases} t_k d_k, & \text{if } x_k + t_k d_k > 0, \\ \theta_k t_k d_k, & \text{otherwise,} \end{cases} \quad x_{k+1} = x_k + h_k.$$

Here assume that for some constant $\theta_l \in (0, 1)$, $\theta_k \in [\phi_l, 1)$, $\theta_k - 1 = O(\|d_k\|)$.

(10) Calculate

$$\begin{aligned} \text{Pred}(h_k) &= g_k^T h_k + \frac{1}{2} (h_k^h)^T B_k (h_k^h) + \frac{1}{2} [\lambda(x_k + h_k) - \lambda_k]^T A_k^T h_k \\ &\quad - [\lambda(x_k + h_k) - \lambda_k]^T \left(c_k + \frac{1}{2} A_k^T h_k \right) + \rho_k (\|c_k + A_k^T h_k\|^2 - \|c_k\|^2) \end{aligned}$$

with

$$h_k^h = \begin{cases} t_k d_k^h, & \text{if } x_k + t_k d_k^h > 0, \\ \theta_k t_k d_k^h, & \text{otherwise,} \end{cases}$$

$$\widehat{\text{Ared}}(h_k) = F(x_{l(k)}, \rho_{l(k)}) - F(x_k + h_k, \rho_{k+1}), \quad \hat{\eta}_k = \frac{\widehat{\text{Ared}}(h_k)}{\text{Pred}(h_k)},$$

and take

$$\Delta_{k+1} = \begin{cases} [\gamma_1 \Delta_k, \gamma_2 \Delta_k], & \text{if } \hat{\eta}_k \leq \eta_1, \\ (\gamma_2 \Delta_k, \Delta_k], & \text{if } \eta_1 < \hat{\eta}_k < \eta_2, \\ (\Delta_k, \min\{\gamma_3 \Delta_k, \Delta_{\max}\}], & \text{if } \hat{\eta}_k \geq \eta_2. \end{cases}$$

(11) Take $m(k+1) = \min\{m(k) + 1, M\}$, and update B_k to obtain B_{k+1} . Then set $k \leftarrow k + 1$ and go to Step (2).

REMARK 1. In the subproblem (S_k) , $\varphi_k(d)$ is a local quadratic approximate model of the Lagrange function l around x_k in $\mathcal{N}(A_k^T)$. A candidate iterative direction d is generated by minimising $\varphi_k(d)$ along $d \in \mathcal{N}(A_k^T)$ only within the ellipsoidal ball centred at x_k with radius Δ_k .

REMARK 2. The scalar $t_k = \alpha_k$, given in Step (5), denotes the step size along d_k to the nonnegative boundary $x_k + \alpha_k d_k \geq 0$, that is,

$$\alpha_k \stackrel{\text{def}}{=} \min\{(x_{k,i}/d_{k,i}) \mid d_{k,i} < 0, i = 1, \dots, m\} \tag{2.19}$$

and $(x_{k,i}/d_{k,i}) \stackrel{\text{def}}{=} +\infty$ if $d_{k,i} = 0$, where $x_{k,i}$ and $d_{k,i}$ are the i -th components of vectors x_k and d_k , respectively. A key property of this scalar $\alpha_k d_k$ is that an arbitrary step to the point $x_k + \alpha_k d_k$ does not violate any nonnegativity constraints.

3. Convergence analysis

We make the following assumptions in this section.

ASSUMPTION H1. The sequence of points $\{x_k\}$ and $\{x_k + d_k\}$ generated by the algorithm is contained in a compact set \mathcal{X} , $f(x)$ and $c(x)$ are twice continuously differentiable on \mathcal{X} , the matrix $A(x)$ has full column rank over \mathcal{X} and $\{B_k\} \subset \mathbb{R}^{n \times n}$ is a set of bounded symmetric matrices.

ASSUMPTION H2. $D_k^{-1}B_kD_k^{-1}$ is bounded, that is, there exists $b > 0$ such that $\|D_k^{-1}B_kD_k^{-1}\| \leq b$, for every k .

According to the assumptions, there are some positive constants ψ and w such that $\psi_k \stackrel{\text{def}}{=} \|A_k(A_k^T A_k)^{-1}\| \leq \psi$ and $w_k \stackrel{\text{def}}{=} \|W(x_k, \lambda_k, \nu_k)\| \leq w$, for every k .

LEMMA 3.1. *Under Assumption H1, we have that*

$$\|c_k\|^2 - \|c_k + A_k^T d_k\|^2 \geq \|c_k\| \min\{\|c_k\|, \Delta_k/\psi_k\}, \tag{3.1}$$

$$2(A_k^T d_k)c_k \leq -\|c_k\| \min\{\|c_k\|, \Delta_k/\psi_k\}. \tag{3.2}$$

PROOF. Consider the vertical trust region subproblem (2.9) and set

$$\bar{d}_k^v = -A_k(A_k^T A_k)^{-1}c_k.$$

(1) If $\|\bar{d}_k^v\| \leq \Delta_k$, and hence $c_k + A_k^T d_k = c_k + A_k^T \bar{d}_k^v = 0$, then \bar{d}_k^v is a solution of subproblem (2.9), that is, $d_k^v = \bar{d}_k^v$ satisfies $\|A_k^T d_k + c_k\|^2 = 0$. This means that (3.1) holds.

(2) If $\|\bar{d}_k^v\| > \Delta_k$, then $(\Delta_k/\|\bar{d}_k^v\|)\bar{d}_k^v$ is a feasible solution of subproblem (2.9). We have that

$$\begin{aligned} \|c_k\|^2 - \|c_k + A_k^T d_k\|^2 &\geq \|c_k\|^2 - \left\| A_k^T \left(\frac{\Delta_k \bar{d}_k^v}{\|\bar{d}_k^v\|} \right) + c_k \right\|^2 \\ &\geq \|c_k\|^2 - \left(1 - \frac{\Delta_k}{\|\bar{d}_k^v\|} \right)^2 \|c_k\|^2 \\ &\geq \frac{\Delta_k \|c_k\|^2}{\|A_k(A_k^T A_k)^{-1}\| \|c_k\|} \geq \frac{\Delta_k}{\psi_k} \|c_k\|, \end{aligned}$$

where the second last inequality follows because $1 - (1 - \Delta_k/\|\bar{d}_k^v\|)^2 \geq \Delta_k/\|\bar{d}_k^v\|$. From (1) and (2), the first conclusion of Lemma 3.1 is true.

For (3.2), we also consider the two cases below:

(i) If $d_k^v = \bar{d}_k^v$, then $A_k^T d_k = A_k^T \bar{d}_k^v = -c_k$. So (3.2) is true.

(ii) If $\|\bar{d}_k^v\| > \Delta_k$, then

$$\|A_k^T d_k^v + c_k\|^2 \leq \left\| A_k^T \left(\frac{\Delta_k d_k^v}{\|\bar{d}_k^v\|} \right) + c_k \right\|^2 = \left(1 - \frac{\Delta_k}{\|\bar{d}_k^v\|} \right)^2 \|c_k\|^2 \leq \left(1 - \frac{\Delta_k}{\|\bar{d}_k^v\|} \right) \|c_k\|^2.$$

Then from

$$\|A_k^T d_k + c_k\|^2 = \|A_k^T d_k\|^2 + \|c_k\|^2 + 2(A_k^T d_k)c_k \geq \|c_k\|^2 + 2(A_k^T d_k)c_k,$$

we have

$$\left(1 - \frac{\Delta_k}{\|\bar{d}_k^v\|} \right) \|c_k\|^2 \geq \|c_k\|^2 + 2(A_k^T d_k)c_k,$$

which implies that (3.2) is also true.

It is well known from applying trust region algorithms that in order to assure the global convergence of the proposed algorithm, it is sufficient to show that, at the k -th iteration, the predicted reduction defined by $-\varphi_k(d_k^h)$, which is obtained by the step d_k^h from trust region subproblem (S_k) , satisfies a sufficient descent condition.

LEMMA 3.2. *Let the step d_k^h be the solution of the trust region subproblem (S_k) . Then there exists $\tau > 0$ such that the step d_k^h satisfies the following sufficient descent condition:*

$$-\varphi_k(d_k) \geq \tau \|P_k \bar{g}_k\| \min \left\{ \Delta_k, \frac{\|P_k \bar{g}_k\|}{\|D_k^{-1} B_k D_k^{-1}\|} \right\} \quad (3.3)$$

for all g_k , B_k and Δ_k , where $P_k = I - \bar{A}_k (\bar{A}_k^T \bar{A}_k)^{-1} \bar{A}_k^T$ and $\tau = 1/2$.

PROOF. See [12, Lemma 2.1].

The following lemma shows the relation between the gradient \bar{g}_k of the objective function and the step d_k^h generated by the proposed algorithm. We can see from the lemma that the direction of the trial step is a sufficient descent direction. The proofs can be also obtained in [12, Lemma 3.2].

LEMMA 3.3. *At the k -th iteration, let d_k be generated from the trust region subproblem (S_k) . Then, there exists a constant $\tau_1 > 0$ such that*

$$g_k^T d_k^h \leq -\tau_1 \|P_k \bar{g}_k\| \min \left\{ \Delta_k, \frac{\|P_k \bar{g}_k\|}{\|D_k^{-1} B_k D_k^{-1}\|} \right\}, \quad (3.4)$$

where $\tau_1 = 1/4$.

LEMMA 3.4. *Suppose that Assumption H1 holds. Then we have that*

$$\begin{aligned} \phi'_k(t_k) &\leq -\kappa_2 \|P_k \bar{g}_k\| \min \left\{ \frac{\|P_k \bar{g}_k\|}{b_k}, \Delta_k \right\} \\ &\quad + \|c_k\| (l_k \Delta_k - \rho_k \min \{ \|c_k\|, \Delta_k / \psi_k \}), \end{aligned} \tag{3.5}$$

where $b_k \stackrel{\text{def}}{=} \|D_k^{-1} B_k D_k^{-1}\|$, $\kappa_2 = \tau_2$.

PROOF. As $\|\lambda(x_k + t_k d_k) - \lambda_k\| \leq l_k t_k \|d_k\|$ and

$$(g_k - A_k \lambda_k)^T d_k = (g_k - \bar{A}_k (\bar{A}_k^T \bar{A}_k)^{-1} \bar{A}_k^T \bar{g}_k)^T \bar{d}_k = (P_k \bar{g}_k)^T d_k = (P_k \bar{g}_k)^T d_k^h,$$

it is easy to see that (2.12) and (3.4) imply that (3.5) holds.

LEMMA 3.5. *Assume that Assumption H1 holds. Then we have*

$$\begin{aligned} \text{Pred}_k(d_k) &\leq -\kappa_1 \|P_k \bar{g}_k\| \min \left\{ \frac{\|P_k \bar{g}_k\|}{b_k}, \Delta_k \right\} \\ &\quad + \|c_k\| (l_k \Delta_k - \rho_k \min \{ \|c_k\|, \Delta_k / \psi_k \}), \end{aligned} \tag{3.6}$$

where $b_k \stackrel{\text{def}}{=} \|D_k^{-1} B_k D_k^{-1}\|$ and $\kappa_1 = \tau_1$.

PROOF. By the definition of $\text{Pred}_k(d_k)$, (3.1) and (3.7), it is clear that (3.6) holds.

LEMMA 3.6. *Suppose that Assumption H1 holds. Then there exist $\kappa_4, \kappa_5 > 0$ such that if*

$$\|c_k\| + \|P_k \bar{g}_k\| \geq (b_k + 1/\psi_k) \Delta_k, \tag{3.7}$$

we have

$$\begin{aligned} \phi'_k(t_k) &\leq -\kappa_4 \|P_k \bar{g}_k\| \min \left\{ \frac{\|P_k \bar{g}_k\|}{b_k}, \Delta_k \right\} - \kappa_4 \min \left\{ \|c_k\|, \frac{\Delta_k}{\psi_k} \right\}, \\ \text{Pred}_k(d_k) &\leq -\kappa_5 \|P_k \bar{g}_k\| \min \left\{ \frac{\|P_k \bar{g}_k\|}{b_k}, \Delta_k \right\} - \kappa_5 \min \left\{ \|c_k\|, \frac{\Delta_k}{\psi_k} \right\}, \end{aligned} \tag{3.8}$$

where $b_k \stackrel{\text{def}}{=} \|D_k^{-1} B_k D_k^{-1}\|$ and $\kappa_4 = \kappa_2/2, \kappa_5 = \kappa_1/2$.

PROOF. Let

$$\begin{aligned} \phi'_k(t_k) &+ \frac{\kappa_2}{2} \|P_k \bar{g}_k\| \min \left\{ \frac{\|P_k \bar{g}_k\|}{b_k}, \Delta_k \right\} + \frac{\rho_k}{2} \min \left\{ \|c_k\|, \frac{\Delta_k}{\psi_k} \right\} \\ &\leq -\frac{\kappa_2}{2} \min \left\{ \frac{\|P_k \bar{g}_k\|}{b_k}, \Delta_k \right\} + \|c_k\| \left(l_k \Delta_k - \frac{\rho_k}{2} \min \left\{ \|c_k\|, \frac{\Delta_k}{\psi_k} \right\} \right) \equiv \chi_k. \end{aligned} \tag{3.9}$$

We only prove that the right-hand side of (3.9) is nonpositive, and then take $\kappa_4 = \min\{\kappa_2/2, \rho_k/2\} = \kappa_2/2$. Here without loss of generality $\rho_k \geq \kappa_2$ so that (3.9) holds.

If $\|c_k\| > \Delta_k/\psi_k$, then taking $\rho_k \geq l_k\psi_k$, the right-hand side of (3.9) is nonpositive.

If $\|c_k\| \leq \Delta_k/\psi_k$, by the condition of the lemma, and

$$\|c_k\| + \|P_k \bar{g}_k\| \geq (b_k + 1/\psi_k)\Delta_k,$$

noting that $\Delta_k \leq \|P_k \bar{g}_k\|/b_k$, we have

$$\begin{aligned} \chi_k &\leq -\frac{\kappa_2}{2} \|P_k \bar{g}_k\| \Delta_k + l_k \Delta_k \|c_k\| - \frac{\rho_k}{2} \|c_k\|^2 \\ &\leq -\frac{\kappa_2 b_k}{2} \Delta_k^2 + l_k \Delta_k \|c_k\| - \frac{\rho_k}{2} \|c_k\|^2 = -\frac{1}{\|c_k\|^2} \left[\frac{\kappa_2 b_k}{2} \epsilon_k^2 - l_k \epsilon_k - \frac{\rho_k}{2} \right] \equiv \xi_k(\epsilon_k), \end{aligned}$$

where we let $\|c_k\| \neq 0$ and $\epsilon_k = \Delta_k/\|c_k\|$. If $\|c_k\| = 0$, the conclusion that the right-hand side of (3.9) is nonpositive is obvious.

The maximum point of the quadratic function $\xi_k(\epsilon_k)$ occurs when $\epsilon_k^* = l_k/(\kappa_2 b_k)$ and the maximum value is $\epsilon_k^* = l_k^2/(2\kappa_2 b_k) - \rho_k$. Hence we have $l_k^2/(2\kappa_2 b_k) \leq \rho_k$, and $\epsilon_k^* \leq 0$. Thus $\chi_k \leq 0$.

Taking $\kappa_4 = \min\{\kappa_2/2, \rho_k/2\} = \kappa_2/2$, the conclusion holds.

Similar to the proof of (3.9), without loss of generality $\rho_k \geq \kappa_1$. Taking $\kappa_5 = \min\{\kappa_1/2, \rho_k/2\} = \kappa_1/2$ and $l_k^2/(2\kappa_1 b_k) \leq \rho_k$, we have that the conclusion holds.

Let

$$\Omega_\epsilon \stackrel{\text{def}}{=} \{x \mid \|P(x)\bar{g}(x)\| + \|c(x)\| \leq \epsilon\}$$

with

$$P(x) \stackrel{\text{def}}{=} I - \bar{A}(x)[\bar{A}(x)^T \bar{A}(x)]^{-1} \bar{A}(x)^T$$

and call it the ϵ -stationary point set of problem (1.1). Obviously, if $\epsilon = 0$ then Ω_ϵ is just the set of the Karush-Kuhn-Tucker points.

LEMMA 3.7. *For every $\epsilon > 0$, there is a $\bar{\Delta}(\epsilon) > 0$ such that, for any $\xi_k \in \chi \setminus \Omega_\epsilon$, when $\Delta_k \leq \bar{\Delta}$,*

$$\phi'_k(t_k) \leq -\hat{\beta}\epsilon\Delta_k, \tag{3.10}$$

$$\text{Pred}_k(d_k) \leq -\hat{\beta}\epsilon\Delta_k, \tag{3.11}$$

where $\hat{\beta} = \kappa_4/2$.

PROOF. For every $\epsilon > 0$, if $x_k \in \chi \setminus \Omega_\epsilon$, then there exists $\epsilon > 0$ such that

$$\|c(x_k)\| + \|P_k \bar{g}_k\| \geq \epsilon, \quad k = 1, 2, \dots$$

This means that at least one of the inequalities $\|P_k \bar{g}_k\| \geq \epsilon/2$ and $\|c_k\| \geq \epsilon/2$ is true.

If $\|P_k \bar{g}_k\| \geq \epsilon/2$, then from (3.1), taking $\bar{\Delta}(\epsilon) = (b + 1/\psi)^{-1} \epsilon/2$, we have

$$\Delta_k \leq \frac{1}{2} \left(b_k + \frac{1}{\psi_k} \right)^{-1} \epsilon \leq \frac{\epsilon}{2b_k} \leq \frac{\|P_k \bar{g}_k\|}{b_k}.$$

Hence, taking $\hat{\beta} = \kappa_4/2$ in (3.8), we find that (3.10) holds.

On the other hand, if $\|c_k\| \geq \epsilon/2$, then also taking

$$\bar{\Delta}(\epsilon) = \frac{1}{2} \left(b + \frac{1}{\psi} \right)^{-1} \leq \frac{1}{2} \left(b_k + \frac{1}{\psi_k} \right)^{-1} \epsilon,$$

we get

$$\Delta_k \leq \frac{1}{2} \left(b_k + \frac{1}{\psi_k} \right)^{-1} \epsilon \leq \psi_k \frac{\epsilon}{2} \leq \psi_k \|c_k\|.$$

Hence, taking $\hat{\beta} = \kappa_4$ and $\epsilon = 1$ in (3.8), we have that (3.10) holds.

Similar to the proof of (3.10), we can show that (3.11) also holds. From the above, we have that the conclusion of the theorem holds.

From Assumption H1 and (2.4), we know that the sequences of vectors λ_k, A_k, g_k and \bar{A}_k are bounded, and hence v_k is bounded. The formula (2.14) shows that every time when the penalty weights change, their values have to be increased by at least κ . Therefore, for sufficiently large k , $\rho_{k+1} \geq v_k + \kappa$ must hold. This means that, for sufficiently large k , the penalty weights will remain the same, that is, $\rho_{k+1} = \rho_k \equiv \bar{\rho}$, for every sufficiently large k . Therefore, without loss of generality, we can assume that the penalty function $F(x, \rho)$ defined by (2.10) is independent of k . We use $\phi(x)$ to represent it, that is, $F(x, \bar{\rho}) = f(x) - \lambda(x)^T c(x) + \bar{\rho} \|c(x)\|^2$.

THEOREM 3.8. *When $x_{k+1} = x_k + d_k$, we have $|Ared_k(d_k) - Pred_k(d_k)| = O(\|d_k\|^2)$, where $Ared_k(d_k)$ and $Pred_k(d_k)$ are defined in (2.15) and (2.16), respectively.*

PROOF. By the definition of $Ared_k(d_k)$, we see that

$$\begin{aligned} Ared_k(d_k) &= F(x_k + d_k, \rho_k) - F(x_k, \rho_k) \\ &= (g_k - A_k \lambda_k)^T d_k - [\lambda(x_k + d_k) - \lambda_k]^T (c_k + A_k^T d_k) \\ &\quad + \rho_k (\|c_k + A_k^T d_k\|^2 - \|c_k\|^2) + O(\|d_k\|^2). \end{aligned}$$

Hence

$$\begin{aligned} Pred_k(d_k) - Ared_k(d_k) &= \frac{1}{2} (d_k^h)^T B_k d_k^h + \frac{1}{2} [\lambda(x_k + d_k^h) - \lambda_k]^T A_k^T d_k \\ &\quad + \frac{1}{2} [\lambda(x_k + d_k) - \lambda_k]^T A_k^T d_k + O(\|d_k\|^2) = O(\|d_k\|^2). \end{aligned}$$

COROLLARY 3.9. *Suppose that Assumption H1 holds. Then there exists a positive constant $\zeta > 0$ such that, for all k ,*

$$|\text{Ared}_k(d_k) - \text{Pred}_k(d_k)| \leq \zeta \Delta_k^2. \tag{3.12}$$

PROOF. It is not difficult to see that if x_k is contained in the compact set \mathcal{X} and f and c_i are twice continuously differentiable on \mathcal{X} , then the $O(\|d_k\|^2)$ term in (3.12) is independent of k , and so is the term $O(\|d_k\|^2)$. From the definitions of d_k^h , d_k^v and the way we calculate d_k^h and d_k^v , we have $\|d_k\|^2 = \|d_k^h\|^2 + \|d_k^v\|^2 \leq 2\Delta_k^2$. Hence there exists $\zeta > 0$ such that (3.12) holds.

We are now ready to state one of our main results.

THEOREM 3.10. *Assume that Assumptions H1–H2 hold and that the strict complementarity of problem (3.1) at every limit point holds. Let $\{x_k\} \in \mathbb{R}^n$ be a sequence generated by the proposed algorithm. Then $\liminf_{k \rightarrow \infty} \{\|P_k \bar{g}_k\| + \|c_k\|\} = 0$.*

PROOF. According to the acceptance rule in Step (8), for sufficiently large k ,

$$F_k(x_{l(k)}, \bar{\rho}) - F_k(x_k + t_k d_k, \bar{\rho}) \geq -t_k \beta \phi'_k(t_k).$$

Taking into account that $m(k + 1) \leq m(k) + 1$, we have

$$F(x_{l(k+1)}, \bar{\rho}) \leq \max_{0 \leq j \leq m(k)+1} \{F(x_{k+1-j}, \bar{\rho})\} = F(x_{l(k)}, \bar{\rho}).$$

This means that the sequence $\{F(x_{l(k)}, \bar{\rho})\}$ is nonincreasing for all large k and hence $\{F(x_{l(k)}, \bar{\rho})\}$ is convergent.

By (2.18) and (3.8), for all large $k > M$, we get

$$\begin{aligned} F(x_{l(k+1)}, \bar{\rho}) &\leq \max_{0 \leq j \leq m(l(k)-1)} \{F(x_{l(k)-j-1}, \bar{\rho})\} + t_{l(k)-1} \beta \phi'_{l(k)-1}(t_{l(k)-1}) \\ &\leq \max_{0 \leq j \leq m(l(k)-1)} \{F_k(x_{l(k)-j-1}, \bar{\rho})\} \\ &\quad - t_{l(k)-1} \beta \kappa_4 \|P_{l(k-1)} \bar{g}_{l(k-1)}\| \min \left\{ \Delta_{l(k-1)}, \frac{\|P_{l(k-1)} \bar{g}_{l(k-1)}\|}{b} \right\} \\ &\quad - t_{l(k)-1} \beta \kappa_4 \min \left\{ \|c_{l(k-1)}\|, \frac{\Delta_{l(k-1)}}{\psi} \right\}. \end{aligned} \tag{3.13}$$

If the conclusion of the theorem were not true, there would exist some $\epsilon > 0$ such that

$$\|P_k \bar{g}_k\| + \|c_k\| \geq 2\epsilon, \quad k = 1, 2, \dots \tag{3.14}$$

This implies that, for each k , at least one of the two following inequalities, $\|P_k \bar{g}_k\| \geq \epsilon$ and $\|c_k\| \geq \epsilon$, is true. By (3.13), either

$$F_k(x_{l(k)}, \bar{\rho}) \leq F_k(x_{l(k)-1}, \bar{\rho}) - t_k \beta \kappa_4 \epsilon \min \{ \Delta_{l(k)-1}, \epsilon/b \}$$

or

$$F(x_{l(k)}, \bar{\rho}) \leq F(x_{l(k)-1}, \bar{\rho}) - t_k \beta \kappa_4 \min \{ \|c_{l(k)-1}\|, \Delta_{l(k)-1}/\psi \}$$

holds. As $\{F(x_{l(k)}, \bar{\rho})\}$ is convergent, from (3.14) we find that $\lim_{k \rightarrow \infty} t_{l(k)-1} \Delta_{l(k)-1} = 0$ which implies that either

$$\lim_{k \rightarrow \infty} t_{l(k)-1} = 0 \tag{3.15}$$

or

$$\lim_{k \rightarrow \infty} \Delta_{l(k)-1} = 0. \tag{3.16}$$

By the updating formula of Δ_k , we have $\gamma_1^{M+1} \Delta_{l(k)-1} \leq \Delta_k \leq \gamma_2^{M+1} \Delta_{l(k)-1}$, which means that from (3.16),

$$\liminf_{k \rightarrow \infty} \Delta_k = 0. \tag{3.17}$$

Similarly [10, Theorem 3.9], we can show that $\hat{\eta}_k \geq \eta_k \geq \eta$, which implies that the trust region radius will be bounded away from zero. This contradicts (3.17). This means that (3.15) holds.

Assume that t_k given in Step (8) is the step size to the boundary of nonnegativity constraints along d_k . From (2.19), $t_k \stackrel{\text{def}}{=} \min \{x_{k,i}/d_{k,i} \mid d_{k,i} < 0, i = 1, \dots, m\}$.

By the definition of t_k , $t_k = x_{k,i}/d_{k,i}$, $t_k \rightarrow 0$ means that there exists a subset $\mathcal{K} \subset \{k\}$ such that $x_{*,i} = 0$ for some i and for all $k_j \in \mathcal{K}$, $t_{k_j} = x_{k_j,i}/d_{k_j,i} \rightarrow 0$. Since the strict complementarity of problem (1.1) holds, we have that $x_{*,i} = 0$ implies $(P_* \bar{g}_*)_i > 0$. Since $x_{k_j,i} > 0$ for a sufficiently large $k_j \in \mathcal{K}$, we find that $(P_{k_j} \bar{g}_{k_j})_i = 0$. This means $(P_{k_j} \bar{g}_{k_j})_i \rightarrow (P_* \bar{g}_*)_i = 0$. It contradicts $(P_* \bar{g}_*)_i > 0$.

If (3.15) holds, we find that

$$t_k \not\rightarrow 0, \tag{3.18}$$

where t_k is determined by the step size to the boundary of the nonnegativity constraints along d_k .

Furthermore, the acceptance rule (2.18) means that, for large k ,

$$F\left(x_k + \frac{t_k}{\omega} d_k, \bar{\rho}\right) - F(x_k, \bar{\rho}) \geq F\left(x_k + \frac{t_k}{\omega} d_k, \bar{\rho}\right) - F(x_{l(k)}, \bar{\rho}) \geq \beta \frac{t_k}{\omega} \phi'_k(t_k).$$

Since

$$F\left(x_k + \frac{t_k}{\omega} d_k, \bar{\rho}\right) - F(x_k, \bar{\rho}) = \frac{t_k}{\omega} \phi'_k(t_k) + o\left(\frac{t_k}{\omega} \|d_k\|\right),$$

we have

$$(1 - \beta) \frac{t_k}{\omega} \phi'_k(t_k) + o\left(\frac{t_k}{\omega} \|d_k\|\right) \geq 0. \tag{3.19}$$

Dividing (3.19) by $t_k \|d_k\|/\omega$ and noting $\phi'_k(t_k) \leq 0$, we have

$$\lim_{k \rightarrow +\infty} \frac{\phi'_k(t_k)}{\|d_k\|} = 0. \tag{3.20}$$

From (3.8) and (3.20), we have that either

$$\phi'_k(t_k) \leq -\kappa_4 \epsilon \min\{\Delta_k, \epsilon/b\} \quad \text{or} \quad \phi'_k(t_k) \leq -\kappa_4 \min\{\Delta_k/\psi, \epsilon\}.$$

In either case, $\lim_{k \rightarrow +\infty} (\Delta_k/\|d_k\|) = 0$, which contradicts $\|d_k\| \leq \|d_k^h\| + \|d_k^v\| \leq (\|D_k^{-1}\| + 1)\Delta_k$ and hence the conclusion of the theorem is true.

4. Local convergence

Theorem 3.10 indicates that at least one limit point of $\{x_k\}$ is a stationary point. In this section we shall first extend this theorem to a stronger result and determine the local convergent rate, but more assumptions are required.

We denote the set of active constraints by $I(x) \stackrel{\text{def}}{=} \{i \mid x_i = 0, i = 1, \dots, n\}$. With any $I \subset \{1, \dots, n\}$ we associate the optimisation subproblem

$$\min f(x) \quad \text{s. t.} \quad \begin{cases} c(x) = 0, \\ x_I = 0. \end{cases} \tag{P}_I$$

ASSUMPTION H3. For all $I \subset \{1, \dots, n\}$, the first order optimality system associated with $(P)_I$ has no non-isolated solutions and the strict complementary of problem (1.1) holds.

ASSUMPTION H4. The constraints of (1.1) are qualified in the sense that $(A_k \lambda_k)_i = 0$, for all $i \notin I(x_k)$ implies that $\lambda = 0$.

Assuming that $(\bar{\lambda}, \bar{v})$ is a unique pair associated with \bar{x} satisfies Assumption H3. Define the set of strictly active constraints as $J(\bar{x}) \stackrel{\text{def}}{=} \{i \mid \bar{v}_i > 0, i = 1, \dots, n\}$ and the extended critical cone as $\mathcal{S}(x) \stackrel{\text{def}}{=} \{d \in \mathbb{R}^n \mid A(\bar{x})^T d = 0, d_i = 0, i \in J(\bar{x})\}$.

ASSUMPTION H5. The solution x_* of problem (1.1) satisfies the strong second-order condition, that is, there exists $\alpha > 0$ such that

$$p^T W_* p \geq \alpha \|p\|^2, \quad \forall p \in \mathcal{S}(x_*), \tag{4.1}$$

where $W_* = \nabla_{xx}^2 l(x_*, \lambda_*, \mu_*)$. This is a sufficient condition for strong regularity.

ASSUMPTION H6. $x_k \rightarrow x_*$.

ASSUMPTION H7.

$$\lim_{k \rightarrow \infty} \frac{\|(B_k - W_k)h_k^h\|}{\|h_k\|} = 0. \tag{4.2}$$

REMARK. This condition is commonly used in discussing superlinear convergence for variations of quasi-Newton methods and secant methods (see [5] and [9], for example).

Following the proof of Lemma 4.3 in [10], we obtain the following theorem.

THEOREM 4.1. Assume that Assumptions H1–H7 hold. If $x_{k+1} = x_k + d_k$, then $|\text{Ared}_k(d_k) - \text{Pred}_k(d_k)| = o(\|d_k^h\|^2) + o(\|d_k^v\|^2)$.

LEMMA 4.2. If Assumptions H1–H7 hold, then for sufficiently large k , $t_k \equiv 1$ in (2.18).

PROOF. By Assumption H7, we have $(d_k^h)^T B_k d_k^h = (d_k^h)^T W_k d_k^h + (\|d_k^h\| \|d_k\|)$. Hence for sufficiently large k , by (4.1)

$$(d_k^h)^T B_k d_k^h \leq -\frac{\alpha}{2} \|d_k^h\|^2 + o(\|d_k^h\| \|d_k\|).$$

By the expansions $f(x_k + d_k) - f_k = g_k^T d_k + \frac{1}{2} d_k^T (\nabla_{xx}^2 f_k) d_k + o(\|d_k^h\| \|d_k\|)$ and

$$\begin{aligned} & \lambda(x_k + d_k)^T c(x_k + d_k) - \lambda_k^T c_k \\ &= \lambda_k^T A_k^T d_k + \frac{1}{2} d_k^T \left(\sum_{i=1}^m \lambda_{k,i} \nabla_{xx}^2 c_i(x_k) \right) d_k + [\lambda(x_k + d_k) - \lambda_k]^T (c_k + A_k^T d_k) \\ & \quad + o(\|d_k\|^2) \\ &= \lambda_k^T A_k^T d_k + \frac{1}{2} d_k^T \sum_{i=1}^m \lambda_{k,i} \nabla_{xx}^2 c_i(x_k) d_k + [\lambda(x_k + d_k) - \lambda_k]^T (c_k + A_k^T d_k) + o(\|d_k\|^2), \end{aligned}$$

by (4.1), and by $d_k^T W_k d_k = (d_k^h)^T W_k d_k^h + 2(d_k^v)^T W_k d_k^h + (d_k^v)^T W_k d_k^v$, we have

$$\|d_k^T W_k d_k - (d_k^h)^T B_k d_k^h\| \leq 2\omega_k \|d_k^h\| \|d_k\| + o(\|d_k\|^2),$$

where ω_k is given by (3.1). By the definitions of $F(x, \rho)$ and $\varphi'_k(1)$,

$$\begin{aligned} & F(x_k + d_k, \rho_k) - F(x_k, \rho_k) - \beta \varphi'_k(1) \\ &= g_k^T d_k - \lambda_k^T A_k^T d_k + [\lambda(x_k + d_k) - \lambda_k]^T (c_k + A_k^T d_k) \end{aligned}$$

$$\begin{aligned}
 & + \frac{1}{2}d_k^T W_k d_k + \|c_k + A_k^T d_k\|^2 - \|c_k\|^2 + o(\|d_k\|^2) \\
 & - \beta\{g_k^T d_k - \lambda_k^T A_k^T d_k + [\lambda(x_k + d_k) - \lambda_k]^T c_k + 2\rho_k(A_k^T d_k)c_k\} \\
 \leq & \left(\frac{1}{2} - \beta\right) g_k^T d_k^h + (1 - \beta)l_k \|c_k\| \|d_k\| + \frac{1}{2}[g_k^T d_k^h + (d_k^h)^T B_k d_k^h] \\
 & + \frac{1}{2}[d_k^T W_k d_k - (d_k^h)^T B_k d_k^h] - \rho_k[1 - 2\beta]\|c_k\|^2 + o(\|d_k\|^2) \\
 \leq & -\frac{\alpha}{4}\|d_k^h\|^2 + (1 - \beta)\|c_k\| \|d_k\| \\
 & + 2\omega_k \|d_k^v\| \|d_k\| - \rho_k[1 - 2\beta]\|c_k\|^2 + o(\|d_k\|^2) \\
 \leq & -\left(\frac{1}{2} - \beta\right) \frac{\alpha}{2}\|d_k^h\|^2 + \hat{l}_k \|c_k\| \|d_k^h\| \\
 & - [\rho_k(1 - 2\beta) - \psi_k l_k]\|c_k\|^2 + o(\|d_k\|^2), \tag{4.3}
 \end{aligned}$$

where $\hat{l}_k = (1 - \beta)l_k + 2\omega_k$ and the last inequality holds because of

$$\|d_k^v\| \leq \|A_k(A_k^T A_k)^{-1}\|(\|c_k\| + \|A_k^T d_k + c_k\|) \leq 2\psi_k \|c_k\|.$$

Further, by (2.8), we find that, for all large k ,

$$\begin{aligned}
 g_k^T d_k^h & = -(g_k - A_k \lambda_k)^T d_k^h = -(d_k^h)^T (B_k + \mu_k D_k^2) d_k^h \\
 & \leq -(d_k^h)^T B_k d_k^h \leq -\frac{\alpha}{2}\|d_k^h\|^2 + o(\|d_k^h\|^2).
 \end{aligned}$$

Hence $g_k^T d_k^h + (d_k^h)^T B_k d_k^h \leq 0$ by (4.3).

We consider the following two possible cases, $\|\bar{d}_k^v\| \leq \Delta_k$ and $\|\bar{d}_k^v\| > \Delta_k$, where $\bar{d}_k^v \stackrel{\text{def}}{=} -A_k(A_k^T A_k)^{-1}c_k$.

(a) If $\|\bar{d}_k^v\| \leq \Delta_k$, we have that, from (4.3),

$$\begin{aligned}
 & F(x_k + d_k, \rho_k) - F(x_k, \rho_k) - \beta\phi'_k(1) \\
 & \leq -\frac{\alpha(1/2 - \beta)}{2}\|d_k^h\|^2 + \hat{l}_k \|c_k\| \|d_k^h\| - \left[\rho_k \left(\frac{1}{2} - 2\beta\right) - \psi_k \hat{l}_k\right] \|c_k\|^2 \\
 & + o(\|d_k\| \|d_k^h\|). \tag{4.4}
 \end{aligned}$$

We now show that the sum of the first three terms on the right-hand side is nonpositive.

If $\|c_k\| = 0$, this conclusion is obvious. Assume that $\|c_k\| \neq 0$. Dividing the first three terms by $\|c_k\|^2$ and letting $s_k = \|d_k^v\|/\|c_k\|$, we have

$$q(s_k) \stackrel{\text{def}}{=} -\frac{(1/2 - \beta)\tau}{2}s_k^2 + \hat{l}_k s_k - [\rho(1 - 2\beta) - \psi_k \hat{l}_k].$$

When $s_k^{\max} = 2\hat{l}_k/(1 - 2\beta)\tau$, this function attains its maximum value:

$$q_{\max} = \frac{\hat{l}_k^2}{(1 - 2\beta)\tau} - [\rho_k(1 - 2\beta) - \psi_k\hat{l}_k].$$

As now $\rho_k \geq \psi_k\hat{l}_k/(1 - 2\beta) + \psi_k\hat{l}_k^2/(1 - 2\beta)$, $q(\cdot)$ is always nonpositive.

We now consider the second possible case.

(b) $\|\bar{d}_k^v\| > \Delta_k$. We have $\psi_k\|c_k\| = \|A_k(A_k^T A_k)^{-1}\| \|c_k\| \geq \|\bar{d}_k^v\| > \Delta_k$. By (4.3), we have

$$\begin{aligned} & F(x_k + d_k, \rho_k) - F(x_k, \rho_k) - \beta\phi'(1) \\ & \leq -\frac{(1 - \beta)\tau}{2} \|d_k^h\|^2 + \hat{k}_k\|c_k\|\|d_k\| - [\rho_k(1 - 2\beta)\alpha_k - \psi_k\hat{l}_k] \frac{\Delta_k}{\psi_k} \|c_k\| \\ & \quad + o(\|d_k\|\|d_k^h\|). \end{aligned} \tag{4.5}$$

As now $\rho_k \geq \psi_k\hat{l}_k/(1 - 2\beta)$, the right-hand side of (4.5) is always nonpositive.

Combining (4.4) and (4.5), we have proved that $F(x_k + d_k, \rho_k) - F(x_k, \rho_k) \leq \beta\phi'(1)$.

Because

$$F(x_k + d_k, \rho_k) - F(x_{l(k)}, \rho_{l(k)}) \leq F(x_k + d_k, \rho_k) - F(x_k, \rho_k) \leq \beta\phi'(1),$$

we find that, for sufficiently large k , $t_k \equiv 1$ in (2.18).

Similar to the proof of (3.18), we also have that $x_{k,i} \neq 0$, for all i , and for large k . By Assumption H6, we have that $d_k \rightarrow 0$. Therefore, from the definition of t_k given in (2.19), $t_k \rightarrow +\infty$. From the above arguments, the theorem holds.

LEMMA 4.3.

$$\begin{aligned} \text{Pred}_k(d_k) & \leq -\frac{\alpha}{4} \|d_k^h\|^2 + l_k\|c_k\|(\|d_k^h\| + \|d_k^v\|) \\ & \quad - \rho\|c_k\| \min\{\|c_k\|, \Delta_k/\psi_k\} + o(\|d_k\|\|d_k^h\|). \end{aligned} \tag{4.6}$$

PROOF. By (2.7) and (4.1)–(4.2), we find that, for all large k ,

$$g_k^T d_k^h = -(g_k - A_k\lambda_k)^T d_k^h \leq -\frac{\alpha}{2} \|d_k^h\|^2 + o(\|d_k^h\|^2). \tag{4.7}$$

On the other hand, by the definitions of L_k and l_k , we have

$$\begin{aligned} & \frac{1}{2} [\lambda(x_k + d_k^h) - \lambda_k]^T A_k^T d_k - [\lambda(x_k + d_k) - \lambda_k]^T \left(c_k + \frac{1}{2} A_k^T d_k \right) \\ & \leq 2L_k\|d_k\| \cdot \|c_k\| \leq l_k\|c_k\|(\|d_k^h\| + \|d_k^v\|). \end{aligned}$$

Substituting this result and (4.7) into (2.16), we obtain (4.6).

LEMMA 4.4. *There exists $\xi > 0$ such that $\text{Pred}_k(p_k) \leq \xi \|p_k\|^2$.*

PROOF. The proof follows that of Theorem 4.6 in [10].

We now discuss the convergence rate of the proposed algorithm. For this purpose, it is necessary to show that for a sufficiently large k , the step size $t_k \equiv 1$, and there exists $\hat{\Delta} > 0$ such that $\Delta_k \geq \hat{\Delta}$.

THEOREM 4.5. *Assume that Assumptions H2–H7 hold. For sufficiently large k , then the step $t_k \equiv 1$ and the trust region constraint is inactive, that is, there exists $\hat{\Delta} > 0$ such that $\Delta_k \geq \hat{\Delta}$, for all $k \geq K'$, where K' is a sufficiently large index.*

PROOF. Let the step size scalar t_k be given in (2.17) along d_k to the nonnegative boundary (2.13). Since the strict complementarity of the problem (1.1) holds at every limit point of $\{x_k\}$, we can show that $t_k \rightarrow 1$. Similarly, we can also show that $\theta_k \rightarrow 1$, as $d_k \rightarrow 0$.

On the other hand, from Lemma 4.2, we can also obtain that, at the k -th iteration, $t_k \equiv 1$, given in (2.17), which means that the step size $t_k \equiv 1$, that is, $h_k = d_k$ for large enough k .

By Assumptions H2–H7, we can obtain from Theorem 4.1 and Lemma 4.4 that

$$|\eta_k - 1| = \frac{|\text{Ared}(h_k) - \text{Pred}(h_k)|}{|\text{Pred}(h_k)|} = \frac{O(\|d_k^h\|^2 + \|d_k^v\|^2)}{\xi_k \|d_k\|^2}. \quad (4.8)$$

Hence (4.8) and Assumption H6 mean that $\eta_k \rightarrow 1$. Hence there exists $\hat{\Delta} > 0$ such that when $\|D_k d_k^h\| \leq \hat{\Delta}$, $\|d_k^v\| \leq \hat{\Delta}$, $\hat{\eta}_k \geq \eta_k \geq \eta_2$, and therefore $\Delta_{k+1} \geq \Delta_k$. As $h_k \rightarrow 0$, there exists an index K' such that $\|D_k d_k^h\| \leq \hat{\Delta}$ and $\|d_k^v\| \leq \hat{\Delta}$, whenever $k \geq K'$. Thus $\Delta_k \geq \hat{\Delta}$, for all $k \geq K'$. The conclusion of the theorem holds even if the quasi-Newton step is used instead of the Newton step.

Theorem 4.5 means that the local convergence rate for the proposed algorithm depends on the reduced Hessian of the objective function at x_* and the local convergence rate of the step d_k . If d_k becomes the projected quasi-Newton step, then the sequence $\{x_k\}$ generated by the algorithm converges to x_* superlinearly [5] (see also [8, 9]).

Conclusions

We have introduced a new interior affine scaling trust region algorithm with a nonmonotonic backtracking technique for nonlinear equality constrained optimisation with nonnegative constraints on the variables. This approach generates strictly feasible iterates for the nonnegative constraints on the variables and possesses strong

convergence characteristics. Finally, we remark that the interior affine scaling trust region algorithm with a nonmonotonic backtracking technique can be extended to the case where there are also nonlinear equality constraints present with box constraints on the variables, that is, $\min\{f(x) \mid c(x) = 0, l \leq x \leq u\}$, where vectors $l, u \in \{\mathbb{R} \cup \{\infty\}\}^n$, $l < u$. This generalisation is also the subject of current research.

Acknowledgements

The author gratefully acknowledges the partial support of this research by the National Science Foundation of China (under Grant 10071050), the Science Foundation Grant (02ZA14070) of the Shanghai Technical Sciences Committee and the Science Foundation Grant (02DK06) of the Shanghai Education Committee. The author would like to thank an anonymous referee for his helpful comments and suggestions.

References

- [1] J. F. Bonnans and C. Pola, "A trust region interior point algorithm for linear constrained optimization", *SIAM J. Optimization* **7** (1997) 717–731.
- [2] T. F. Coleman and Y. Li, "An interior trust region approach for minimization subject to bounds", *SIAM J. Optimization* **6** (1996) 418–445.
- [3] N. Y. Deng, Y. Xiao and F. J. Zhou, "A nonmonotonic trust region algorithm", *J. Optim. Theory Appl.* **76** (1993) 259–285.
- [4] L. Grippo, F. Lampariello and S. Lucidi, "A nonmonotonic line search technique for Newton's methods", *SIAM J. Numer. Anal.* **23** (1986) 707–716.
- [5] J. Nocedal and M. L. Overton, "Projectal Hessian updating algorithms for nonlinearly constrained optimization", *SIAM J. Numer. Anal.* **22** (1985) 821–850.
- [6] J. Nocedal and Y. Yuan, "Combining trust region and line search techniques", in *Advances in Nonlinear Programming* (ed. Y. Yuan), (Kluwer, Dordrecht, 1998) 153–175.
- [7] M. J. D. Powell and Y. Yuan, "A recursive quadratic programming algorithm that uses differentiable exact penalty functions", *Math. Programming* **35** (1986) 265–278.
- [8] J. Zhang and D. Zhu, "A projective quasi-Newton method for nonlinear optimization", *J. Comput. Appl. Math.* **53** (1990) 291–307.
- [9] J. Zhang and D. Zhu, "A trust region type doglog method for nonlinear optimization", *Optimization* **21** (1990) 543–557.
- [10] D. Zhu, "Nonmonotonic projected algorithm with both trust region and line search for constrained optimization", *J. Comput. Appl. Math.* **117** (2000) 35–60.
- [11] D. Zhu, "Curvilinear paths and trust region methods with nonmonotonic back tracking technique for unconstrained optimization", *J. Comput. Math.* **19** (2001) 241–258.
- [12] D. Zhu, "An interior point algorithm with nonmonotonic backtracking technique for linear constrained optimization", *J. Comput. Math.* (2002) submitted.