A SHIFTED PRIMAL-DUAL INTERIOR METHOD FOR NONLINEAR OPTIMIZATION

Philip E. Gill* Vyacheslav Kungurtsev† Daniel P. Robinson‡

UCSD Center for Computational Mathematics
Technical Report CCoM-18-1
February 1, 2018

Abstract

Interior methods provide an effective approach for the treatment of inequality constraints in nonlinearly constrained optimization. A new primal-dual interior method is proposed based on minimizing a sequence of shifted primal-dual penalty-barrier functions. Certain global convergence properties are established. In particular, it is shown that every limit point is either an infeasible stationary point, or an approximate KKT point, i.e., it satisfies reasonable stopping criteria for a local minimizer and is a KKT point under a weak constraint qualification. It is shown that under suitable assumptions, the method is equivalent to the conventional path-following interior method in the neighborhood of a solution.

Key words. Nonlinear programming, nonlinear constraints, augmented Lagrangian methods, interior methods, path-following methods, regularized methods, primal-dual methods.

AMS subject classifications. 49J20, 49J15, 49M37, 49D37, 65F05, 65K05, 90C30

---

*Department of Mathematics, University of California, San Diego, La Jolla, CA 92093-0112 (`pgill@ucsd.edu`). Research supported in part by National Science Foundation grants DMS-0915220 and DMS-1318480.

†Agent Technology Center, Department of Computer Science, Faculty of Electrical Engineering, Czech Technical University in Prague. (`vyacheslav.kungurtsev@fel.cvut.cz`) Research supported by the European social fund within the framework of realizing the project "Support of inter-sectoral mobility and quality enhancement of research teams at the Czech Technical University in Prague", CZ.1.07/2.3.00/30.0034.

‡Department of Applied Mathematics and Statistics, Johns Hopkins University, Baltimore, MD 21218-2682 (`daniel.p.robinson@jhu.edu`). Research supported in part by National Science Foundation grant DMS-1217153.

1. Introduction

This paper is concerned with solving nonlinear optimization problems of the form

$$\underset{x\in\mathbb{R}^n}{\text{minimize}} \;\; f(x) \quad \text{subject to} \;\; c(x) \geq 0, \tag{NIP}$$

where $c : \mathbb{R}^n \mapsto \mathbb{R}^m$ and $f : \mathbb{R}^n \mapsto \mathbb{R}$ are twice-continuously differentiable. Barrier methods are a class of methods for solving (NIP) that involve the unconstrained minimization of a sequence of barrier functions parameterized by a scalar barrier parameter (see, e.g., Frisch [8], Fiacco and McCormick [6], and Fiacco [5]). Each barrier function includes a logarithmic barrier term that creates a positive singularity at the boundary of the feasible region and enforces strict feasibility of the barrier function minimizers. Reducing the barrier parameter to zero has the effect of allowing the barrier minimizers to approach the boundary of the feasible region in the neighborhood of a solution. However, as the barrier parameter decreases and the values of the constraints that are active at the solution approach zero, the linear equations associated with solving each barrier subproblem become increasingly ill-conditioned. Shifted or modified barrier functions were introduced to avoid this ill-conditioning by implicitly shifting the constraints so that the barrier minimizers approach the solution without the need for the barrier parameter to go to zero (see, e.g., Gill et al. [9], Polyak [14], Breitfeld and Shanno [3], Conn, Gould and Toint [4], Nash, Polyak and Sofer [12], and Goldfarb et al. [11]). Shifted or modified barrier functions are defined in terms of the Lagrange multiplier estimates and are analogous to augmented Lagrangian methods for equality constrained optimization.

The trajectory of minimizers of a conventional barrier function forms a continuous path of points that approaches a solution of (NIP) from the interior of the feasible region. Points on this path may be interpreted as satisfying a set of perturbed optimality conditions for (NIP). The direct solution of these perturbed optimality conditions as a set of nonlinear equations using Newton's method provides an alternative to solving the ill-conditioned equations associated with a conventional barrier method. This approach is the basis of the class of path-following interior methods (for a survey, see, e.g., Forsgren, Gill and Wright [7], and Wright [17]). For these methods, the barrier function often serves as a merit function for forcing the convergence of Newton's iterates from an arbitrary starting point. Moreover, the perturbation parameter associated with the early iterations serves an auxiliary role as an implicit regularization parameter in the linear equations. This regularization plays a crucial role in the robustness of interior methods on ill-conditioned and ill-posed problems.

In this paper we formulate and analyze a new primal-dual interior method based on minimizing a sequence of shifted primal-dual penalty-barrier functions. Certain global convergence properties are established. In particular, it is shown that every limit point is either an infeasible stationary point, or an approximate KKT point, i.e., it satisfies reasonable stopping criteria for a local minimizer and is a KKT point under a weak constraint qualification. In addition, it is shown that under suitable assumptions, the method is equivalent to a regularized path-following interior method in the neighborhood of a solution.

The paper is organized in four sections. The proposed primal-dual penalty-barrier function is introduced in Section 2. In this section we also propose a line-search algorithm for minimizing the penalty-barrier function for fixed penalty and barrier parameters, and establish convergence of the algorithm under certain assumptions. In Section 3, an algorithm for solving problem (NIP) that builds upon the work from Section 2 is proposed and global convergence results are established. In Section 4 we discuss implementation details of the modified Newton method for minimizing each barrier-penalty function. Finally, Section 5 gives some conclusions and topics for further work.

1.1. Notation and terminology

Unless explicitly indicated otherwise, $\|\cdot\|$ denotes the vector two-norm or its induced matrix norm. The inertia of a real symmetric matrix A, denoted by $\mathrm{In}(A)$, is the integer triple (a_+, a_-, a_0) giving the number of positive, negative and zero eigenvalues of A. Given vectors a and b with the same dimension, the vector with ith component $a_i b_i$ is denoted by $a \cdot b$. Similarly, $\min(a, b)$ is a vector with components $\min(a_i, b_i)$. The vectors e and e_j denote, respectively, the column vector of ones and the jth column of the identity matrix I. The dimensions of e, e_i and I are defined by the context. Given vectors x and y, the long vector consisting of the elements of x augmented by elements of y is denoted by (x, y). The ith component of a vector labeled with a subscript will be denoted by $[\,\cdot\,]_i$, e.g., $[\,v_F\,]_i$ is the ith component of the vector v_F. The subvector of components with indices in the index set $\mathcal{S}$ is denoted by $[\,\cdot\,]_{\mathcal{S}}$, e.g., $[\,v\,]_{\mathcal{S}}$ is the vector with components v_i for $i \in \mathcal{S}$. A local solution of an optimization problem is denoted by x^*. The vector $g(x)$ is used to denote $\nabla f(x)$, the gradient of $f(x)$. The matrix $J(x)$ denotes the $m \times n$ constraint Jacobian, which has ith row $\nabla c_i(x)^T$, the gradient of the ith constraint function $c_i(x)$. The Lagrangian function associated with (NIP) is $L(x, y) = f(x) - c(x)^T y$, where y is the m-vector of dual variables. The Hessian of the Lagrangian with respect to x is denoted by $H(x, y) = \nabla^2 f(x) - \sum_{i=1}^m y_i \nabla^2 c_i(x)$. Let $\{\alpha_j\}_{j\geq 0}$ be a sequence of scalars, vectors, or matrices and let $\{\beta_j\}_{j\geq 0}$ be a sequence of positive scalars. If there exists a positive constant γ such that $\|\alpha_j\| \leq \gamma \beta_j$, we write $\alpha_j = O\big(\beta_j\big)$. If there exists a sequence $\{\gamma_j\} \to 0$ such that $\|\alpha_j\| \leq \gamma_j \beta_j$, we say that $\alpha_j = o(\beta_j)$. If there exists a positive sequence $\{\sigma_j\} \to 0$ and a positive constant β such that $\beta_j > \beta \sigma_j$, we write $\beta_j = \Omega(\sigma_j)$.

2. Minimizing a new primal-dual penalty-barrier function

In order to avoid the need to find a strictly feasible point for the constraints of (NIP), each inequality $c_i(x) \geq 0$ is written in terms of an equality and nonnegative slack variable $c_i(x) - s_i = 0$, $s_i \geq 0$. This gives the equivalent problem

$$\underset{x\in\mathbb{R}^n, s\in\mathbb{R}^m}{\text{minimize}} \;\; f(x) \quad \text{subject to} \;\; c(x) - s = 0, \quad s \geq 0. \tag{2.1}$$

This transformation implies that the equality constraint $c_i(x) - s_i = 0$ is likely to be evaluated at a point for which $c_i(x) < 0$. The first-order KKT conditions for problem (2.1) are

$$c(x^*) - s^* = 0, \qquad s^* \geq 0, \tag{2.2a}$$
$$g(x^*) - J(x^*)^T y^* = 0, \qquad y^* - w^* = 0, \tag{2.2b}$$
$$s^* \cdot w^* = 0, \qquad w^* \geq 0. \tag{2.2c}$$

The vectors y^* and w^* constitute the Lagrange multipliers for the equality constraints $c(x) - s = 0$ and non-negativity constraints $s \geq 0$ respectively. The quantities (x, s, y, w) are said to constitute a *primal-dual estimate* of the quantities (x^*, s^*, y^*, w^*) that satisfy the optimality conditions for (2.1).

In this section, we introduce a new penalty-barrier function that is designed to be minimized with respect to both the primal and dual variables simultaneously. After defining this function, a line-search algorithm is proposed for minimizing it when all of the parameters are fixed. Throughout this discussion and beyond, the following assumption for the problem functions is assumed to hold.

Assumption 2.1. *The functions f and c are two-times continuously differentiable.*

The proposed primal-dual function is defined as

$$\begin{aligned} M(x, s, y, w\,; y^E, \mu^P, \mu^B) = \underbrace{f}_{(A)} \underbrace{- (c-s)^T y^E}_{(B)} \\ \underbrace{+ \frac{1}{2\mu^P}\|c - s\|^2}_{(C)} \underbrace{+ \frac{1}{2\mu^P}\|c - s + \mu^P(y - y^E)\|^2}_{(D)} + \varPhi(s, w\,; y^E, \mu^B), \end{aligned}$$

where $\varPhi(s, w\,; y^E, \mu^B)$ is the modified barrier function

$$\begin{aligned} &\varPhi(w, s\,; y^E, \mu^B) \\ &\quad = \underbrace{-\sum_{i=1}^m \mu^B y_i^E \ln\big(s_i + \mu^B\big)}_{(E)} \underbrace{-\sum_{i=1}^m \mu^B y_i^E \ln\big(w_i(s_i + \mu^B)\big)}_{(F)} \underbrace{+\sum_{i=1}^m w_i(s_i + \mu^B)}_{(G)}. \end{aligned}$$

The vector $y^E \in \mathbb{R}^m$ is an estimate of a Lagrange multiplier vector for the constraint $c(x) - s = 0$, the scalar $\mu^P > 0$ a penalty parameter, and the scalar $\mu^B > 0$ is a barrier parameter.

If we define the positive-definite matrices

$$D_P = \mu^P I \quad \text{and} \quad D_B = (S + \mu^B I)W^{-1},$$

and the auxiliary vectors

$$\pi^P = \pi^P(x, s) = y^E - \frac{1}{\mu^P}\big(c(x) - s\big) \quad \text{and} \quad \pi^B = \pi^B(s) = \mu^B(S + \mu^B I)^{-1} y^E,$$

then $\nabla M(x, s, y, w\,; y^E, \mu^P, \mu^B)$ can be written as

$$\nabla M = \begin{pmatrix} g - J^T\big(\pi^P + (\pi^P - y)\big) \\ \big(\pi^P + (\pi^P - y)\big) - \big(\pi^B + (\pi^B - w)\big) \\ -D_P(\pi^P - y) \\ -D_B(\pi^B - w) \end{pmatrix}. \tag{2.3}$$

Of course, some algebraic simplifications may be performed in the previous displayed equation, but we prefer this presentation of the gradient since it highlights quantities (e.g., $\pi^P - y$) that will be important in our analysis. Similarly, the penalty-barrier function Hessian $\nabla^2 M(x, s, y, w\,; y^E, \mu^P, \mu^B)$ may be written in the form

$$\nabla^2 M = \begin{pmatrix} H + 2J^T D_P^{-1} J & -2J^T D_P^{-1} & J^T & 0 \\ -2D_P^{-1} J & 2(D_P^{-1} + D_B^{-1} W^{-1} \Pi^B) & -I & I \\ J & -I & D_P & 0 \\ 0 & I & 0 & D_B W^{-1} \Pi^B \end{pmatrix},$$

where $H = H\big(x, \pi^P + (\pi^P - y)\big)$.

In developing algorithms, our goal is to achieve rapid convergence of the iterates to a solution of (2.1) without the need for μ^P and μ^P to go to zero. The mechanism for ensuring convergence of any method based on M is the minimization of M for fixed parameters. Thus, for the remainder of this section, we present and analyze a line search method for this purpose. To simplify notation, for the remainder of this section, we drop all notational dependence on y^E, μ^P, and μ^B, since they are fixed.

The method for minimizing M with fixed parameters is formally stated as Algorithm 1. Each iteration requires the computation of a search direction Δv as the solution of the system

$$H_k^M \Delta v_k = -\nabla M(v_k), \tag{2.4}$$

where H_k^M is a positive-definite matrix and $v_k = (x_k, s_k, y_k, w_k)$; note that the search direction Δv_k may be partitioned as $\Delta v_k = (\Delta x_k, \Delta s_k, \Delta y_k, \Delta w_k)$. (The reader should think of H_k^M as a positive-definite approximation to $\nabla^2 M(x_k, s_k, y_k, w_k)$. Details on how H_k^M is defined are discussed in Section 2.1.) A line search is then performed, as given by Algorithm 2, to obtain a step length α_k, such that the step $v_{k+1} \leftarrow v_k + \alpha_k \Delta v_k$ sufficiently decreases the function M and keeps important quantities positive (see Line 8 of Algorithm 2).

Algorithm 1 Minimizing the merit function for fixed parameters y^E, μ^P, and μ^B.

procedure `MERIT`(x_0, s_0, y_0, w_0)
  Restrictions: $s_0 + \mu^B e > 0$ and $w_0 > 0$.
  Set $v_0 \leftarrow (x_0, s_0, y_0, w_0)$.
  for $k = 0, 1, 2, \ldots,$ **do**
    Choose $H_k^M \succ 0$, and then compute the search direction Δv_k from (2.4).
    Calculate $\alpha_k =$ `LINESEARCH`$(v_k, \Delta v_k)$ from Algorithm 2.
    Set $v_{k+1} \leftarrow v_k + \alpha_k \Delta v_k$.
    Set $\widehat{s}_{k+1} \leftarrow c(x_{k+1}) - \mu^P\big(y^E + \tfrac{1}{2}(w_{k+1} - y_{k+1})\big)$.
    Perform a slack reset $s_{k+1} \leftarrow \max\{s_{k+1}, \widehat{s}_{k+1}\}$.
    Set $v_{k+1} \leftarrow (x_{k+1}, s_{k+1}, y_{k+1}, w_{k+1})$.

Algorithm 2 A line search.

procedure LINESEARCH($v_k, \Delta v_k$)
  Recall: $v_k = (x_k, s_k, y_k, w_k)$ and $\Delta v_k = (\Delta x_k, \Delta s_k, \Delta y_k, \Delta w_k)$.
  Requirements: $s_k + \mu^B e > 0$ and $w_k > 0$.
  Constants: $\{\eta, \gamma\} \in (0, 1)$.
  Set $\alpha \leftarrow 1$.
  loop
    if $M(v_k + \alpha\Delta v_k) \le M(v_k) + \eta\alpha\nabla M(v_k)^T \Delta v_k$ **then**
      if $s_k + \alpha\Delta s_k + \mu^B e > 0$ and $w_k + \alpha\Delta w_k > 0$ **then**
        return $\alpha_k \leftarrow \alpha$.
    Set $\alpha \leftarrow \gamma\alpha$.

2.1. The modified Newton equations

The matrix H_k^M used in (2.4) to compute the modified Newton direction is obtained by replacing π^P by y in the $(1,1)$ block and π^B by w in the $(4,4)$ block of $\nabla^2 M(x_k, s_k, y_k, w_k)$. This gives the matrix

$$B(x, s, y, w; \mu^P, \mu^B) = \begin{pmatrix} H + 2J^T D_P^{-1} J & -2J^T D_P^{-1} & J^T & 0 \\ -2D_P^{-1} J & 2(D_P^{-1} + D_B^{-1}) & -I & I \\ J & -I & D_P & 0 \\ 0 & I & 0 & D_B \end{pmatrix}, \tag{2.5}$$

where $H = H(x, y)$. Thus, a modified Newton direction Δv may be computed as the solution of the modified Newton equations

$$B(x, s, y, w\,; \mu^P, \mu^B)\Delta v = -\nabla M(x, s, y, w; y^E, \mu^P, \mu^B).$$

If we then define the nonsingular matrix

$$T = \begin{pmatrix} I & 0 & -2J^T D_P^{-1} & 0 \\ 0 & I & 2D_P^{-1} & -2D_B^{-1} \\ 0 & 0 & I & 0 \\ 0 & 0 & 0 & W \end{pmatrix},$$

we can conclude that the above modified Newton direction also satisfies

$$TB(x, s, y, w\,; \mu^P, \mu^B)\Delta v = -T\nabla M(x, s, y, w; y^E, \mu^P, \mu^B),$$

which, upon multiplication and use of $WD_B = S + \mu^B I$, leads to the system

$$\begin{pmatrix} H & 0 & -J^T & 0 \\ 0 & 0 & I & -I \\ J & -I & D_P & 0 \\ 0 & W & 0 & S + \mu^B I \end{pmatrix} \begin{pmatrix} \Delta x \\ \Delta s \\ \Delta y \\ \Delta w \end{pmatrix} = - \begin{pmatrix} g - J^T y \\ y - w \\ c - s + \mu^P(y - y^E) \\ s \cdot w + \mu^B(w - y^E) \end{pmatrix}. \tag{2.6}$$

These equations are identical to the path-following equations (2.8) derived in the next section. This also establishes that the direction computed from the primal-dual path-following equations will be a descent direction for M provided the matrix $B(x, s, y, w; \mu^P, \mu^B)$ is positive definite. This is discussed further in Section 4

2.2. The path-following equations

Consider the following primal-dual path following equations given by

$$F(x,s,y,w\,;y^E,\mu^P,\mu^B) = \begin{pmatrix} g(x) - J(x)^T y \\ y - w \\ c(x) - s + \mu^P(y - y^E) \\ s\cdot w + \mu^B(w - y^E) \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}. \tag{2.7}$$

Note that a zero (x,s,y,w) of F that satisfies $s > 0$ and $w > 0$ approximates a solution to problem (2.1), with the approximation becoming increasingly accurate as both $\mu^P(y - y^E) \to 0$ and $\mu^B(w - w^E) \to 0$. If $v = (x,s,y,w)$ is a given approximate primal-dual zero of F such that $s + \mu^B e > 0$ and $w > 0$, the Newton equations for the change in variables $\Delta v = (\Delta x, \Delta s, \Delta y, \Delta w)$ are given by $F'(v) = -F(v)$, i.e.,

$$\begin{pmatrix} H & 0 & -J^T & 0 \\ 0 & 0 & I & -I \\ J & -I & \mu^P I & 0 \\ 0 & W & 0 & S + \mu^B I \end{pmatrix} \begin{pmatrix} \Delta x \\ \Delta s \\ \Delta y \\ \Delta w \end{pmatrix} = - \begin{pmatrix} g - J^T y \\ y - w \\ c - s + \mu^P(y - y^E) \\ s\cdot w + \mu^B(w - y^E) \end{pmatrix}, \tag{2.8}$$

where $c = c(x)$, $g = g(x)$, $J = J(x)$, and $H = H(x,y)$. These equations are identical to the modified Newton equations (2.6) for minimizing M.

2.3. Convergence

One may suspect that the description thus far would be enough to ensure (under typical assumptions) the existence of a limit point (x^*, s^*, y^*, w^*) of the iterates generated by Algorithm 1 such that $\nabla M(x^*, s^*, y^*, w^*) = 0$. Indeed, this may be shown, but it is not sufficient for our ultimate purpose, namely, to use Algorithm 1 as the building block for an algorithm for solving problem (2.1). The slack reset used in Line 9 of Algorithm 1 has been introduced with this goal in mind. The specific update can be motivated by noting that $\widehat{s}_{k+1}$, as defined in Line 8, is the unique minimizer with respect to s of the sum of terms (B), (C), (D), and (G) in the definition of the function M. In particular, it follows from Lines 8 and 9 that the value of s_{k+1} that results from line 9 satisfies

$$s_{k+1} \ge \widehat{s}_{k+1} = c(x_{k+1}) - \mu^P\big(y^E + \tfrac{1}{2}(w_{k+1} - y_{k+1})\big),$$

which implies that, after rearrangement

$$c(x_{k+1}) - s_{k+1} \le \mu^P\big(y^E + \tfrac{1}{2}(w_{k+1} - y_{k+1})\big). \tag{2.9}$$

It may not be apparent now, but (2.9) is crucial. In particular, later when we allow modifications to μ^P and y^E, this inequality will ensure any limit point of $\{(x_k, s_k)\}$, say (x^*, s^*), satisfies $c(x^*) - s^* \le 0$ provided we keep y^E and $w_{k+1} - y_{k+1}$ bounded and allow μ^P to converge to zero. This will be necessary to handle problems that are (locally) infeasible, which is a challenge for all methods for nonconvex optimization. Finally, we note that the slack update never causes M to increase, which means that the M is monotonically decreasing. This is established in the next lemma, which requires the following assumption.

Assumption 2.2. *The set of matrices $\{H_k^M\}_{k\ge 0}$ are chosen to be uniformly positive definite and uniformly bounded in norm.*

Lemma 2.1. *The iterates $v_k = (x_k, s_k, y_k, w_k)$ satisfy $M(v_{k+1}) \le M(v_k)$ for all k.*

Proof. It is well-known that Δv_k is a strict descent direction for M at v_k, i.e., $\nabla M(v_k)^T \Delta v_k < 0$, if $\nabla M(v_k)$ is nonzero and H_k^M is positive definite (see Assumption 2.2). As $\nabla M(v_k)^T \Delta v_k = 0$ if $\nabla M(v_k) = 0$, it must hold that $\nabla M(v_k)^T \Delta v_k \le 0$ for all k. This property implies that the line search performed in Algorithm 2 produces an α_k such that $v_{k+1} = v_k + \alpha_k \Delta v_k$ satisfies $M(v_{k+1}) \le M(v_k)$. Thus, the only way the desired result can not hold is if the slack reset procedure in line 9 of Algorithm 1 causes M to increase. We complete the proof by showing that this can not happen.

The vector $\widehat{s}_{k+1}$ used in the slack reset is the unique vector that minimizes the sum of terms (B), (C), (D), and (G) that define the function M, so that the sum of these terms can not increase. Also, term (A) is independent of s, so that its value does not change. Next, we observe that the slack-reset procedure only has the effect of possibly increasing the value of some of its components, which means that terms (E) and (F) in the definition of M can only decrease. In total, this implies that the slack reset can never increase the value of M. ∎

The analysis requires the following assumption on the iterates to hold.

Assumption 2.3. *The iterates $\{x_k\}$ are contained in a compact set.*

Lemma 2.2. *The iterates (x_k, s_k, y_k, w_k) computed by Algorithm 1 have the following properties.*

(i) *The sequences $\{s_k\}$, $\{c(x_k) - s_k\}$, $\{y_k\}$, and $\{w_k\}$ are uniformly bounded.*

(ii) *For all i it holds that*

$$\liminf_{k\ge 0}\, [s_k + \mu^B e]_i > 0 \;\; \text{and} \;\; \liminf_{k\ge 0}\, [w_k]_i > 0.$$

(iii) *The sequences $\{\pi^P(x_k, s_k)\}$, $\{\pi^B(s_k)\}$, and $\{\nabla M(x_k, s_k, y_k, w_k)\}$ are uniformly bounded.*

(iv) *There exists a scalar $M_{\rm low}$ so that $M(x_k, s_k, y_k, w_k) \ge M_{\rm low} > -\infty$ for all k.*

Proof. For a proof by contradiction, assume that $\{\|s_k\|\}$ is not uniformly bounded. Noting that $s_k + \mu^B e \ge 0$ by construction, means that there must exist a subsequence $\mathcal{S}$ and component i such that

$$\lim_{k\in\mathcal{S}}\, [\, s_k\,]_i = \infty \;\; \text{and} \;\; [s_k]_i \ge [s_k]_j \;\; \text{for all } j \text{ and } k \in \mathcal{S}. \tag{2.10}$$

We now show that the merit function must converge to infinity along $\mathcal{S}$. It follows from (2.10), Assumption 2.3, continuity of c, $s_k + \mu^B e \ge 0$, and $y^E \ge 0$ that terms

(A) and (B) are bounded below for all k, and that term (C) converges to ∞ along $\mathcal{S}$ at the same rate as $\|s_k\|^2$. It is also clear that (D) is bounded below by zero. On the other hand, term (E) goes to $-\infty$ along $\mathcal{S}$ at the rate $-\ln([s_k]_i + \mu^B)$. Next, note that term (G) in the definition of M is bounded below. Now, if term (F) is uniformly bounded below on $\mathcal{S}$, then the previous argument proves that M converges to infinity along $\mathcal{S}$, which contradicts Lemma 2.1. Otherwise, if term (F) goes to $-\infty$ along $\mathcal{S}$, then we may conclude the existence of a subsequence $\mathcal{S}_1 \subseteq \mathcal{S}$ and a component, say j, such that

$$\lim_{k\in\mathcal{S}_1} [s_k + \mu^B e]_j [w_k]_j = \infty \quad \text{and} \tag{2.11}$$

$$[s_k + \mu^B e]_j [w_k]_j \geq [s_k + \mu^B e]_l [w_k]_l \text{ for all } l \text{ and } k \in \mathcal{S}_1. \tag{2.12}$$

Using these properties and the definition of (G), we may conclude that (G) converges to ∞ faster than (F) converges to $-\infty$. Thus, we may conclude that M converges to ∞ along $\mathcal{S}_1$, which contradicts Lemma 2.1. We have thus proved that $\{s_k\}$ is uniformly bounded, which is the first part of result (i). The second part of (i), i.e., the uniform boundedness of $\{c(x_k) - s_k\}$, follows from the first result, the continuity of c, and Assumption 2.3.

Next, we prove the third bound in part (i), i.e., that $\{y_k\}$ is uniformly bounded. For a proof by contradiction, we may assume that there exists some subsequence $\mathcal{S}$ and component i such that

$$\lim_{k\in\mathcal{S}} |[y_k]_i| = \infty \text{ and } |[y_k]_i| \geq |[y_k]_j| \text{ for all } j \text{ and } k \in \mathcal{S}. \tag{2.13}$$

Observe, as in the previous paragraph, that terms (A), (B) and (C) are uniformly bounded below over all k, and that term (D) converges to ∞ along $\mathcal{S}$ at the rate of $[y_k]_i^2$ since we have already proved that $\{s_k\}$ is uniformly bounded. Using the uniform boundedness of $\{s_k\}$ a second time, we may also deduce that term (E) is uniformly bounded below. If term (F) is bounded below along $\mathcal{S}$, then since term (G) is bounded we would conclude, in totality, that $\lim_{k\in\mathcal{S}} M(v_k) = \infty$, which contradicts Lemma 2.1. Thus, term (F) must converge to $-\infty$, from which we are ensured of a subsequence $\mathcal{S}_1 \subseteq \mathcal{S}$ and a component, say j, that again satisfy (2.11) and (2.12). For such $k \in \mathcal{S}_1$ and j we have that term (G) converges to ∞ faster than term (F) converges to $-\infty$, so that we may again conclude that $\lim_{k\in\mathcal{S}_1} M(v_k) = \infty$ along $\mathcal{S}_1$, which contradicts Lemma 2.1. Thus, $\{y_k\}$ is uniformly bounded.

We now prove the final bound in part (i), i.e., that $\{w_k\}$ is uniformly bounded. For a prove by contraction, we assume that the set is unbounded, which implies—using the fact that $w_k > 0$ by assumption—the existence of a subsequence $\mathcal{S}$ and a component i such that

$$\lim_{k\in\mathcal{S}} [w_k]_i = \infty \text{ and } [w_k]_i \geq [w_k]_j \text{ for all } j \text{ and } k \in \mathcal{S}. \tag{2.14}$$

It follows that there exists a subsequence $\mathcal{S}_1 \subseteq \mathcal{S}$ and set $\mathcal{J} \subseteq \{1, 2, \ldots, m\}$ satisfying

$$\lim_{k\in\mathcal{S}_1} [w_k]_j = \infty \text{ for all } j \in \mathcal{J} \quad \text{and} \quad \{[w_k]_j : j \notin \mathcal{J} \text{ and } k \in \mathcal{S}_1\} \text{ is bounded.} \tag{2.15}$$

Next, using similar arguments as before and the new fact that $\{y_k\}$ is bounded, we can notice that terms (A), (B), (C), and (D) are bounded. We may also combine terms (E) and (F) to obtain

$$\text{(E)} + \text{(F)} = -\mu^B \sum_{j=1}^{m} y_j^E \big(2\ln([s_k + \mu^B e]_j) + \ln([w_k]_j)\big), \tag{2.16}$$

and in view of term (G) and Lemma 2.1, conclude that

$$[w_k]_j [s_k + \mu^B e]_j = O\big(\ln([w_k]_i)\big) \text{ for all } 1 \leq j \leq m, \tag{2.17}$$

which can be seen to hold as follows. We have from (2.14), boundedness of $\{s_k\}$, $y^E > 0$, and (2.16) that (E) + (F) is asymptotically bounded below by $-\mu^B y_i^E \ln([w_k]_i)$ along $\mathcal{S}$. Combining this with the boundedness of terms (A), (B), (C), and (D), allows us to conclude that (2.17) must hold, for otherwise the merit function M would converge to infinity (see term (G)) along $\mathcal{S}$, contradicting Lemma 2.1. We have thus established that (2.17) holds.

Using $w_k > 0$ (holds by construction) and monotonicity of $\ln(\cdot)$, it follows from (2.17) that there exists a positive constant κ_1 such that

$$\ln\big([s_k + \mu^B e]_j\big) \leq \ln\left(\frac{\kappa_1 \ln([w_k]_i)}{[w_k]_j}\right) = \ln(\kappa_1) + \ln\big(\ln([w_k]_i)\big) - \ln([w_k]_j) \tag{2.18}$$

for all $1 \leq j \leq m$ and sufficiently large k. Combining (2.16), boundedness of $\{s_k\}$, (2.15), $y^E > 0$, and (2.18) give the existence of positive constants κ_2 and κ_3 satisfying

$$\begin{aligned}
\text{(E)} + \text{(F)} &\geq -\kappa_2 - \mu^B \sum_{j \in \mathcal{J}} y_j^E \big(2\ln([s_k + \mu^B e]_j) + \ln([w_k]_j)\big) \\
&\geq -\kappa_2 - \mu^B \sum_{j \in \mathcal{J}} y_j^E \big(2\ln(\kappa_i) + 2\ln\big(\ln([w_k]_i)\big) - \ln([w_k]_j)\big) \\
&\geq -\kappa_3 - \mu^B \sum_{j \in \mathcal{J}} y_j^E \big(2\ln\big(\ln([w_k]_i)\big) - \ln([w_k]_j)\big)
\end{aligned} \tag{2.19}$$

for all sufficiently large k. We now focus on bounding the summation in (2.19). To this end, we first define

$$\alpha = \frac{[y^E]_i}{4\|y^E\|_1} > 0,$$

which is well defined since $y^E > 0$. We may then conclude from (2.14) and (2.15) that

$$2\ln\big(\ln([w_k]_i)\big) - \ln([w_k]_j) \leq \alpha \ln([w_k]_i)$$

for all $j \in \mathcal{J}$ and sufficiently large $k \in \mathcal{S}_1$. Using this bound, (2.19), and $y_k^E > 0$

allows us to conclude that

$$\begin{aligned}
&\text{(E)} + \text{(F)} \\
&\geq -\kappa_3 - \mu^B y_i^E \big(2\ln\big(\ln([w_k]_i)\big) - \ln([w_k]_i)\big) - \mu^B \sum_{j\in\mathcal{J}, j\neq i} y_j^E \big(2\ln\big(\ln([w_k]_i)\big) - \ln([w_k]_j)\big) \\
&\geq -\kappa_3 - \mu^B y_i^E \big(2\ln\big(\ln([w_k]_i)\big) - \ln([w_k]_i)\big) - \mu^B \sum_{j\in\mathcal{J}, j\neq i} y_j^E \alpha \ln([w_k]_i) \\
&\geq -\kappa_3 - \mu^B y_i^E \big(2\ln\big(\ln([w_k]_i)\big) - \ln([w_k]_i)\big) - \mu^B \alpha \ln([w_k]_i)\|y^E\|_1
\end{aligned}$$

for all sufficiently large $k \in \mathcal{S}_1$. Combing this the choice of α and the fact that

$$2\ln\big(\ln([w_k]_i)\big) - \ln([w_k]_i) \leq -\tfrac{1}{2}\ln([w_k]_i)$$

for all sufficiently large $k \in \mathcal{S}$ (this follows from (2.14)), we arrive at

$$\begin{aligned}
\text{(E)} + \text{(F)} &\geq -\kappa_3 + \tfrac{1}{2}\mu^B y_i^E \ln([w_k]_i) - \mu^B \alpha \ln([w_k]_i)\|y^E\|_1 \\
&\geq -\kappa_3 + \mu^B \left(\tfrac{1}{2}[y^E]_i - \alpha\|y^E\|_1\right)\ln([w_k]_i) \\
&= -\kappa_3 + \tfrac{1}{4}\mu^B \ln([w_k]_i),
\end{aligned}$$

for all sufficiently large $k \in \mathcal{S}_1$. In particular, this inequality and (2.14) together give

$$\lim_{k\in\mathcal{S}_1} \text{(E)} + \text{(F)} = \infty.$$

Since we already said that terms (A), (B), (C), and (D) are bounded, and it is clear that term (G) is bounded below by zero, we must conclude that the merit function M converges to infinity along $\mathcal{S}_1$. Since this contradicts Lemma 2.2, we finally conclude that $\{w_k\}$ is bounded, as claimed.

To prove part (ii) by contradiction, first suppose that $\{[s_k + \mu^B e]_i\} \to 0$ along some subsequence $\mathcal{S}$ and for some component i. As before, terms (A), (B), (C), and (D) are all bounded from below over all k. We may also use the fact that $\{s_k\}$ and $\{w_k\}$ were proved to be uniformly bounded in part (i) to conclude that terms (E) and (F) converge to ∞ along $\mathcal{S}$, and term (G) is bounded below. In totality, we have shown that $\lim_{k\in\mathcal{S}} M(v_k) = \infty$, which contradicts Lemma 2.1, and in turn, proves that $\liminf\, [s_k + \mu e]_i > 0$ for all i. A similar argument may be used to prove that $\liminf\, [w_k]_i > 0$ for all i, which completes the proof.

We now consider part (iii). The sequence $\{\pi^P(x_k, s_k)\}$ is uniformly bounded as a consequence of part (i) and the fact that y^E and μ^P are fixed throughout. The sequence $\{\pi^B(s_k)\}$ is uniformly bounded as a consequence of part (ii) and the fact that y^E and μ^B are fixed throughout. Lastly, the sequence $\{\nabla M(x_k, s_k, y_k)\}$ is uniformly bounded as a consequence of parts (i) and (ii), the uniform boundedness that we just proved for $\{\pi^P(x_k, s_k)\}$ and $\{\pi^B(s_k)\}$, Assumptions 2.1 and 2.3, and the fact that y^E and μ^P are fixed throughout.

To prove part (iv), we prove that all terms that define M are each uniformly bounded below. Term (A) is uniformly bounded below because of Assumptions 2.1 and 2.2. Term (B) is uniformly bounded below as a consequence of Lemma 2.2 (i) and the fact that y^E is kept fixed. Terms (C) and (D) are both nonnegative, hence, trivially uniformly bounded below. Terms (E) and (F) are uniformly bounded below because μ^B and y^E are kept fixed, and Lemma 2.2 (i). Finally, it follows from Lemma 2.2 (ii) that term (G) is positive. The desired M_{low} now follows. ∎

Certain conditions hold when the gradients of M is bounded away from zero.

Lemma 2.3. *If there exists a scalar $\epsilon > 0$ and subsequence $\mathcal{S}$ satisfying*

$$\|\nabla M(v_k)\| \geq \epsilon \;\; \text{for all} \;\; k \in \mathcal{S}, \tag{2.20}$$

then the following hold.

(i) *The set $\{\|\Delta v_k\|\}_{k\in\mathcal{S}}$ is uniformly bounded above and away from zero.*

(ii) *There exists a scalar $\delta > 0$ such that $\nabla M(v_k)^T \Delta v_k \leq -\delta$ for all $k \in \mathcal{S}$.*

(iii) *For all $k \in \mathcal{S}$, the Armijo condition in line 7 of Algorithm 2 is satisfied for all $\alpha \geq \alpha_{\min}$ and some $\alpha_{\min} > 0$.*

Proof. Part (i) follows from (2.20), Assumption 2.2, Lemma 2.2 (iii), and the fact that Δv_k is computed from (2.4).

For part (ii), first observe from (2.4) that

$$\nabla M(v_k)^T \Delta v_k = -\Delta v_k^T H_k^M \Delta v_k \leq -\lambda_{\min}(H_k^M)\|\Delta v_k\|_2^2. \tag{2.21}$$

The existence of δ in part (ii) now follows from (2.21), Assumption 2.2, and part (i).

Finally, we consider part (iii). A standard result in unconstrained optimization [13] is that the Armijo condition is satisfied for all

$$\alpha = \Omega\left(\frac{-\nabla M(v_k)^T \Delta v_k}{\|\Delta v_k\|^2}\right). \tag{2.22}$$

We note that this result requires Lipschitz continuity of $\nabla M(v)$, which holds as a consequence of Assumptions 2.1 and Lemma 2.2(ii). The existence of the $\alpha_{\min} > 0$ claimed in part (iii) now follows from (2.22), and parts (i) and (ii). ∎

We may now state the main convergence result.

Theorem 2.1. *The iterates $v_k = (x_k, s_k, y_k)$ satisfy $\lim_{k\to\infty} \nabla M(v_k) = 0$.*

Proof. The proof is by contradiction. Suppose that there exists a constant $\epsilon > 0$ and a subsequence $\mathcal{S}$ such that $\|\nabla M(v_k)\| \geq \epsilon$ for all $k \in \mathcal{S}$. It follows from Lemma 2.1 and Lemma 2.2(iv) that $\lim_{k\to\infty} M(v_k) = M_{\min} > -\infty$. Using this fact and the knowledge that the Armijo condition during the line search (see line 7 in Algorithm 2) is satisfied for all k, we must conclude that

$$\lim_{k\to\infty} \alpha_k \nabla M(v_k)^T \Delta v_k = 0,$$

which in light of Lemma 2.3(ii) means that $\lim_{k\in\mathcal{S}} \alpha_k = 0$. From this fact and Lemma 2.3(iii) we may conclude that the inequality constraints enforced in line 8 of Algorithm 2 must have restricted the step length. In particular, this implies the existence of a subsequence $\mathcal{S}_1 \subseteq \mathcal{S}$ and a component i such that either

$$[s_k + \alpha_k \Delta s_k + \mu^B e]_i > 0 \text{ and } [s_k + (1/\gamma)\alpha_k \Delta s_k + \mu^B]_i \le 0 \text{ for } k \in \mathcal{S}_1$$

or

$$[w_k + \alpha_k \Delta w_k]_i > 0 \text{ and } [w_k + (1/\gamma)\alpha_k \Delta w_k]_i \le 0 \text{ for } k \in \mathcal{S}_1, \tag{2.23}$$

where $\gamma \in (0,1)$ is used in Algorithm 2. Since the argument used for both cases is the same, we assume, without loss of generality, that (2.23) occurs. It follows from Lemma 2.2(ii) that there exists some $\epsilon > 0$ such that

$$\epsilon < w_{k+1} = w_k + \alpha_k \Delta w_k = w_k + (1/\gamma)\alpha_k \Delta w_k - (1/\gamma)\alpha_k \Delta w_k + \alpha_k \Delta w_k$$

for all sufficiently large k, so that with (2.23) we have

$$w_k + (1/\gamma)\alpha_k \Delta w_k > \epsilon + (1/\gamma)\alpha_k \Delta w_k - \alpha_k \Delta w_k = \epsilon + \alpha_k \Delta w_k (1/\gamma - 1) > 0$$

for all sufficiently large $k \in \mathcal{S}_1$, where the last inequality follows from $\lim_{k\in\mathcal{S}} \alpha_k = 0$ and Lemma 2.3(i). This contradicts (2.23) for all sufficiently large $k \in \mathcal{S}_1$. ∎

3. An algorithm for nonlinear optimization

3.1. The algorithm

The proposed method (see Algorithm 3) is essentially a combination of Algorithm 1 and conditions for adjusting the parameters that define the merit function, which were fixed in Algorithm 3. The proposed strategy uses the distinction between O-iterations, M-iterations, and F-iterations, which are described below.

The definition of an O-iteration is based on conditions that indicate convergence to an optimal solution for problem (2.1). In order to measure such progress at a point $v_{k+1} = (x_{k+1}, s_{k+1}, y_{k+1}, w_{k+1})$, we define the following feasibility, optimality, and complementarity measures:

$$\begin{aligned} \chi_{\text{feas}}(v_{k+1}) &= \|c(x_{k+1}) - s_{k+1}\| \\ \chi_{\text{opt}}(v_{k+1}) &= \max\big(\|g(x_{k+1}) - J(x_{k+1})^T y_{k+1}\|, \|y_{k+1} - w_{k+1}\|\big), \text{ and} \\ \chi_{\text{comp}}(v_{k+1}) &= \max\big(\|\min(s_{k+1}, w_{k+1})\|, \|w_{k+1} \cdot s_{k+1}\|\big), \end{aligned} \tag{3.1}$$

so that first-order optimality for problem (2.1) is characterized by

$$\chi(v_{k+1}) = \chi_{\text{feas}}(v_{k+1}) + \chi_{\text{opt}}(v_{k+1}) + \chi_{\text{comp}}(v_{k+1}) = 0. \tag{3.2}$$

We then say that the kth iteration is an O-iteration if $\chi(v_{k+1}) \le \chi_k^{\max}$, where $\{\chi_k^{\max}\}$ is a monotonically decreasing positive sequence maintained by the algorithm, and update the parameters in Step 10. (Note that these updates ensure that $\{\chi_k^{\max}\}$ converges to zero if infinitely many O-iterations occur). We call v_{k+1} an O-iterate.

If the condition for an O-iteration does not hold, a test is made to determine if $v_{k+1} = (x_{k+1}, s_{k+1}, y_{k+1}, w_{k+1})$ is an approximate first-order solution of the problem

$$\underset{v=(x,s,y,w)}{\text{minimize}} \;\; M(v\,; y_k^E, \mu_k^P, \mu_k^B). \tag{3.3}$$

In particular, if $v_{k+1} = (x_{k+1}, s_{k+1}, y_{k+1}, w_{k+1})$ satisfies

$$\|\nabla_x M(v_{k+1}\,; y_k^E, \mu_k^P, \mu_k^B)\|_\infty \le \tau_k, \tag{3.4a}$$

$$\|\nabla_s M(v_{k+1}\,; y_k^E, \mu_k^P, \mu_k^B)\|_\infty \le \tau_k, \tag{3.4b}$$

$$\|\nabla_y M(v_{k+1}\,; y_k^E, \mu_k^P, \mu_k^B)\|_\infty \le \tau_k \|D_{k+1}^P\|_\infty, \quad \text{and} \tag{3.4c}$$

$$\|\nabla_w M(v_{k+1}\,; y_k^E, \mu_k^P, \mu_k^B)\|_\infty \le \tau_k \|D_{k+1}^B\|_\infty, \tag{3.4d}$$

where τ_k is a positive tolerance, $D_{k+1}^P = \mu_k^P I$, and $D_{k+1}^B = (S_{k+1} + \mu_k^B I) W_{k+1}^{-1}$, then we say that the kth iteration is an M-iteration and that v_{k+1} is an M-iterate since it is an approximate first-order solution to (3.3). In this case, the multiplier estimate y_{k+1}^E is defined by the safeguarded value

$$y_{k+1}^E = \max\bigl(-y_{\max} e, \min(y_{k+1}, y_{\max} e)\bigr) \tag{3.5}$$

for some positive constant $y_{\max}$. Next, in Step 13 we check whether the condition

$$\chi_{\text{feas}}(v_{k+1}) \le \tau_k \tag{3.6}$$

holds, and when it does, we set $\mu_{k+1}^P \leftarrow \mu_k^P$; otherwise, we set $\mu_{k+1}^P \leftarrow \frac{1}{2}\mu_k^P$ in order to place more emphasis on decreasing the constraint violation during subsequent iterations. Similarly, in Step 17 we check whether the conditions

$$\chi_{\text{comp}}(v_{k+1}) \le \tau_k \;\; \text{and} \;\; s_{k+1} \ge -\tau_k e \tag{3.7}$$

hold, and when they do we set $\mu_{k+1}^B \leftarrow \mu_k^B$; otherwise, we set $\mu_{k+1}^B \leftarrow \frac{1}{2}\mu_k^B$ in order to place more emphasis on achieving complementarity in during subsequent iterations. This completes the updates performed during an M-iteration.

If the kth iteration is neither an O- or an M-iteration, we say that it is an F-iteration. During an F-iteration none of the merit function parameters are changed, so that progress during that iteration is solely on reducing the merit function.

3.2. Convergence

We establish convergence of the iterates by making use of a complementary approximate KKT condition, which was first introduced in [2].

Definition 3.1. (Complementary Approximate KKT Condition)
A feasible point (x^*, s^*)*, i.e., it satisfies* $s^* \ge 0$ *and* $c(x^*) - s^* = 0$*, is said to*

Algorithm 3 A primal-dual barrier method.

procedure PDB(x_0, s_0, y_0, w_0)
  Restrictions: $s_0 > 0$ and $w_0 > 0$.
  Constants: $\{\eta, \gamma\} \subset (0, 1)$.
  Choose $y_0^E > 0$, $\chi_0^{\max} > 0$, and $\{\mu_0^P, \mu_0^B\} \subset (0, \infty)$.
  Set $v_0 \leftarrow (x_0, s_0, y_0, w_0)$.
  for $k = 0, 1, 2, \ldots$ **do**
    Set $y^E \leftarrow y_k^E$, $\mu^P \leftarrow \mu_k^P$, and $\mu^B \leftarrow \mu_k^B$.
    Compute $v_{k+1} = (x_{k+1}, s_{k+1}, y_{k+1}, w_{k+1})$ from Steps 5–10 of Algorithm 1.
    if $\chi(v_{k+1}) \le \chi_k^{\max}$ **then** ▷ O-iterate
      Set $\chi_{k+1}^{\max} = \frac{1}{2}\chi_k^{\max}$, $y_{k+1}^E = y_{k+1}$, $\mu_{k+1}^P \leftarrow \mu_k^P$, $\mu_{k+1}^B \leftarrow \mu_k^B$, $\tau_{k+1} \leftarrow \tau_k$.
    else if v_{k+1} satisfies (3.4) **then** ▷ M-iterate
      Set $\chi_{k+1}^{\max} = \chi_k^{\max}$, y_{k+1}^E using (3.5), and $\tau_{k+1} = \frac{1}{2}\tau_k$.
      if (3.6) holds **then** set $\mu_{k+1}^P \leftarrow \mu_k^P$ **else** set $\mu_{k+1}^P \leftarrow \frac{1}{2}\mu_k^P$.
      if (3.7) holds **then**
        Set $\mu_{k+1}^B \leftarrow \mu_k^B$.
      else
        Set $\mu_{k+1}^B \leftarrow \frac{1}{2}\mu_k^B$ and reset s_{k+1} so that $s_{k+1} + \mu_{k+1}^B e > 0$.
    else ▷ F-iterate
      Set $\chi_{k+1}^{\max} = \chi_k^{\max}$, $y_{k+1}^E = y_k^E$ $\mu_{k+1}^P \leftarrow \mu_k^P$, $\mu_{k+1}^B \leftarrow \mu_k^B$, and $\tau_{k+1} = \tau_k$.

satisfy the complementary approximate-KKT (CAKKT) *condition if there exists* $\{(x_j, s_j, u_j, z_j)\}$ *with* $\{x_j\} \to x^*$ *and* $\{s_j\} \to s^*$ *such that*

$$\{g(x_j) - J(x_j)^T u_j\} \to 0, \tag{3.8a}$$

$$\{u_j - z_j\} \to 0, \tag{3.8b}$$

$$\{z_j\} \ge 0, \quad \textit{and} \tag{3.8c}$$

$$\{z_j \cdot s_j\} \to 0. \tag{3.8d}$$

We say that such an (x^*, s^*) *is a CAKKT point.*

This condition is stronger than the approximate KKT (AKKT) condition introduced in [15]. Its usefulness is made clear by the following result, which uses the cone continuity property [1]. The CCP is the weakest constraint qualification that can be associated with sequential optimality conditions.

Lemma 3.1. (Andreani et al. [1]) *If* (x^*, s^*) *is an CAKKT point that satisfies the cone continuity property (CCP), then* (x^*, s^*) *is a first-order KKT point.*

Our analysis now aims to establish the existence of such CAKKT points. To facilitate referencing of the iterations, we define

$$\begin{aligned} \mathcal{O} &= \{k : \text{iteration } k \text{ is an O-iteration}\}, \\ \mathcal{M} &= \{k : \text{iteration } k \text{ is an M-iteration}\}, \quad \text{and} \\ \mathcal{F} &= \{k : \text{iteration } k \text{ is an F-iteration}\}. \end{aligned}$$

We first show that limit points of the sequence of O-iterates are CAKKT points.

Lemma 3.2. *If* $|\mathcal{O}| = \infty$*, then there exists at least one limit point* (x^*, s^*) *of the infinite sequence* $\{(x_{k+1}, s_{k+1})\}_{k\in\mathcal{O}}$ *and any such limit point is a CAKKT point.*

Proof. By Assumption 2.3 there exists a limit point of $\{x_{k+1}\}_{k\in\mathcal{O}}$, so let x^* be any such limit point. Combining this with Assumption 2.1 gives the existence of $\mathcal{K} \subseteq \mathcal{O}$ such that $\{x_{k+1}\}_{k\in\mathcal{K}} \to x^*$ and $\{c(x_{k+1})\}_{k\in\mathcal{K}} \to c(x^*)$. Next, note that $\{\chi_k^{\max}\} \to 0$ because of $|\mathcal{O}| = \infty$ and the updating strategy used in Algorithm 3. Since we also have $\chi(v_{k+1}) \le \chi_k^{\max}$ for all $k \in \mathcal{K} \subseteq \mathcal{O}$ and $\chi_{\text{feas}}(v_{k+1}) \le \chi(v_{k+1})$ for all k, it follows that $\{\chi_{\text{feas}}(v_{k+1})\}_{k\in\mathcal{K}} \to 0$, i.e., that $\{c_{k+1} - s_{k+1}\}_{k\in\mathcal{K}} \to 0$, and thus $\{s_{k+1}\}_{k\in\mathcal{K}} \to \lim_{k\in\mathcal{K}} c(x_{k+1}) = c(x^*) =: s^*$. Moreover, (x^*, s^*) is feasible for the general constraints since $c(x^*) - s^* = 0$. The remaining feasibility condition $s^* \ge 0$ follows from $\{\chi_{\text{comp}}(v_{k+1})\}_{k\in\mathcal{K}} \le \{\chi(v_{k+1})\}_{k\in\mathcal{K}} \le \{\chi_k^{\max}\}_{k\in\mathcal{K}} \to 0$.

It remains to verify that conditions (3.8a)–(3.8d) hold. It follows from our previously established limit $\{\chi(v_{k+1})\}_{k\in\mathcal{K}} \to 0$ and the definition of $\chi(\cdot)$ that $\{\chi_{\text{opt}}(v_{k+1}) + \chi_{\text{comp}}(v_{k+1})\}_{k\in\mathcal{K}} \le \{\chi(v_{k+1})\}_{k\in\mathcal{K}} \to 0$, which in turn gives $\{g_{k+1} - J_{k+1}^T y_{k+1}\}_{k\in\mathcal{K}} \to 0$, $\{y_{k+1} - w_{k+1}\}_{k\in\mathcal{K}} \to 0$, and $\{w_{k+1} \cdot s_{k+1}\}_{k\in\mathcal{K}} \to 0$ and verifies that conditions (3.8a), (3.8b), and (3.8d) are satisfied. Line 8 of Algorithm 2 (this is called in Line 6 of Algorithm 1, which in turn is called in Line 8 of Algorithm 3) implies the nonnegativity of w_{k+1} for all k, so that (3.8c) is satisfied. Thus (x^*, s^*) satisfies the properties of CAKKT with $\{v_{k+1}\}_{k\in\mathcal{K}}$ corresponding to $\{(x_j, s_j, u_j, z_j)\}$. ∎

The majority of our remaining analysis concerns the the sequence of M-iterates, which is an infinite set anytime the set of O-iterates is finite.

Lemma 3.3. *If* $|\mathcal{O}| < \infty$*, then* $|\mathcal{M}| = \infty$*.*

Proof. For a proof by contradiction, suppose that $|\mathcal{O} \cup \mathcal{M}| < \infty$. It follows from the construction of Algorithm 3 that $k \in \mathcal{F}$ for all sufficiently large k. In turn, this means that there exists k_F such that

$$k \in \mathcal{F}, \ \ \tau_k = \tau > 0, \ \ \mu_k^P = \mu^P > 0, \ \ \text{and} \ \ \mu_k^B = \mu^B > 0 \ \ \text{for all } k \ge k_F. \tag{3.9}$$

This means that the iterates computed by Algorithm 3 are the same as those computed by Algorithm 1 for all $k \ge k_F$. We may therefore apply Theorem 2.1, Lemma 2.2(i), and Lemma 2.2(ii) to conclude that (3.4) is satisfied for all $k \in \mathcal{K}$ sufficiently large. This would mean, in view of Step 11 of Algorithm 3, that $k \in \mathcal{M}$ for all $k \ge k_F$, which contradicts (3.9) since $\mathcal{F} \cap \mathcal{M} = \emptyset$. ∎

Before establishing conditions that ensure CAKKT points exist when the set $\mathcal{M}$ is infinite, we use the next lemma to justify the quantities used in the right-hand-side of (3.4). We use the quantities

$$\pi_{k+1}^P = y_k^E - \frac{1}{\mu_k^P}\big(c(x_{k+1}) - s_{k+1}\big) \ \ \text{and} \ \ \pi_{k+1}^B = \mu_k^B (s_{k+1} + \mu_k^B I)^{-1} y_k^E \tag{3.10}$$

that arise in the definition of the gradient of the merit function (see (2.3)).

Lemma 3.4. *If* $|\mathcal{M}| = \infty$ *and* $\mathcal{K} \subseteq \mathcal{M}$, *then*

$$\lim_{k\in\mathcal{K}} |\pi^P_{k+1} - y_{k+1}| = \lim_{k\in\mathcal{K}} |\pi^B_{k+1} - w_{k+1}| = \lim_{k\in\mathcal{K}} |\pi^P_{k+1} - \pi^B_{k+1}| = \lim_{k\in\mathcal{K}} |y_{k+1} - w_{k+1}| = 0.$$

Proof. It follows from (3.4c) and (3.4d) that

$$|\pi^P_{k+1} - y_{k+1}| \le \tau_k \quad \text{and} \quad |\pi^B_{k+1} - w_{k+1}| \le \tau_k. \tag{3.11}$$

Since $|\mathcal{M}| = \infty$ by assumption, we have from the structure of Algorithm 3 that $\lim_{k\to\infty} \tau_k = 0$. Combining this with (3.11) establishes the first two desired results. We may then combine $\lim_{k\to\infty} \tau_k = 0$, (3.11), and (3.4b) to deduce that

$$\lim_{k\in\mathcal{K}} |\pi^P_{k+1} - \pi^B_{k+1}| = 0, \tag{3.12}$$

which is the third desired result. Finally, it follows from (3.12) and (3.11) that

$$\begin{aligned} 0 &= \lim_{k\in\mathcal{K}} |\pi^P_{k+1} - \pi^B_{k+1}| = \lim_{k\in\mathcal{K}} |(\pi^P_{k+1} - y_{k+1}) + (y_{k+1} - w_{k+1}) + (w_{k+1} - \pi^B_{k+1})| \\ &= \lim_{k\in\mathcal{K}} |y_{k+1} - w_{k+1}|, \end{aligned}$$

where we have again used $\lim_{k\to\infty} \tau_k = 0$. This is the fourth desired result. ∎

We now show that when the index set of O-iterations is finite, any limit point of the primal sequence of M-iterates (which must be infinite) is a CAKKT point, provided the limit point is feasible with respect to the general constraints.

Lemma 3.5. *If* $|\mathcal{O}| < \infty$ *and there exists an infinite sequence* $\mathcal{K} \subseteq \mathcal{M}$ *such that* $\lim_{k\in\mathcal{K}}(x_{k+1}, s_{k+1}) = (x^*, s^*)$, *then* $c(x^*) - s^* \neq 0$.

Proof. For a proof by contradiction, suppose that $c(x^*) - s^* = 0$. Under this assumption, we now proceed to show that (x^*, s^*) is a CAKKT point by showing that the sequence $\{(x_{k+1}, s_{k+1}, u_{k+1}, z_{k+1})\}_{k\in\mathcal{K}} = \{(x_{k+1}, s_{k+1}, y_{k+1}, w_{k+1})\}_{k\in\mathcal{K}}$ has the properties required in Definition 3.1. Note that $|\mathcal{O}| < \infty$, Lemma 3.3, and the updates in Algorithm 3 together imply that $\lim_{k\to\infty} \tau_k = 0$ and

$$\chi^{\max}_k = \chi^{\max} > 0 \;\text{ for all sufficiently large } k \in \mathcal{K}. \tag{3.13}$$

The suppositions of this lemma, Lemma 3.4, and $\mathcal{K} \subseteq \mathcal{M}$ imply that

$$\lim_{k\in\mathcal{K}}(u_{k+1} - z_{k+1}) = \lim_{k\in\mathcal{K}}(y_{k+1} - w_{k+1}) = 0,$$

so that condition (3.8b) holds. Also, condition (3.8c) holds for $z_{k+1} = w_{k+1}$ since $w_{k+1} > 0$ for all k (see Step 8 of Algorithm 2, which is computed when the line search in Step 6 of Algorithm 1 is performed, which is called in Step 8 of Algorithm 3).

We next show that (3.8a) holds. We have

$$\begin{aligned} g_{k+1} - J^T_{k+1} y_{k+1} &= g_{k+1} - J^T_{k+1}(2\pi^P_{k+1} + y_{k+1} - 2\pi^P_{k+1}) \\ &= g_{k+1} - J^T_{k+1}(2\pi^P_{k+1} - y_{k+1}) - 2J^T_{k+1}(y_{k+1} - \pi^P_{k+1}) \\ &= \nabla_x M(v_{k+1}; y^E_k, \mu^P_k, \mu^B_k) - 2J^T_{k+1}(y_{k+1} - \pi^P_{k+1}). \end{aligned}$$

Combining this with Assumption 2.1, $\{x_{k+1}\}_{k\in\mathcal{K}} \to x^*$, $\lim_{k\to\infty} \tau_k = 0$, (3.4a), and Lemma 3.4 that $\lim_{k\in\mathcal{K}}(g_{k+1} - J_{k+1}^T y_{k+1}) = 0$, which is (3.8a) since $y_{k+1} = u_{k+1}$.

Now, since we are still under the assumption that $c(x^*) - s^* = 0$, it remains to show that $s^* \ge 0$ in order to establish that (x^*, s^*) is feasible. To this end, note that the line search Algorithm 2 gives $s_{k+1} + \mu_k^B e > 0$ for all k. Thus, if $\lim_{k\to\infty} \mu_k^B = 0$ we must have $s^* = \lim_{k\in\mathcal{K}} s_{k+1} \ge -\lim_{k\in\mathcal{K}} \mu_k^B e = 0$. On the other hand, suppose that $\lim_{k\in\mathcal{K}} \mu_k^B \ne 0$. It follows from Step 17 of Algorithm 3 that $\mu_k^B = \mu^B > 0$ and that (3.7) holds for all $k \in \mathcal{M}$ sufficiently large, i.e., $s_{k+1} \ge -\tau_k e$ for all $k \in \mathcal{M}$ sufficiently large. Since $\lim_{k\to\infty} \tau_k = 0$, we again have after taking limits that $s^* \ge 0$.

Lastly, we show that (3.8d) holds by considering two cases.

Case 1: $\lim_{k\to\infty} \mu_k^B \ne 0$. In this case we have that $\mu_k^B = \mu^B > 0$ for all sufficiently large k. Combining this with $|\mathcal{M}| = \infty$, we know in light of Step 17 of Algorithm 3 that (3.7) holds for all sufficiently large $k \in \mathcal{K}$. This means that $\|s_{k+1} \cdot w_{k+1}\| \le \chi_{\text{comp}}(v_{k+1}) \le \tau_k$ for all sufficiently large $k \in \mathcal{K}$, and since $\lim_{k\to\infty} \tau_k = 0$, we have that $\lim_{k\in\mathcal{K}}(s_{k+1} \cdot w_{k+1}) = 0$.

Case 2: $\lim_{k\to\infty} \mu_k^B = 0$. To establish (3.8d), we note that it is sufficient to show that $\lim_{k\in\mathcal{K}}[w_{k+1}]_i = 0$ for all i such that $[s^*]_i \ne 0$. Let i be such an index. We have from Lemma 3.4 that $\lim_{k\in\mathcal{K}}(\pi_{k+1}^B - w_{k+1}) = 0$. Since the matrix $S_{k+1} + \mu_k^B I$ is bounded as a result of $\{\mu_k^B\}$ being a positive monotonically decreasing sequence and $\lim_{k\in\mathcal{K}} s_k = s^*$, we also have that

$$0 = \lim_{k\in\mathcal{K}}(S_{k+1} + \mu_k^B I)(\pi_{k+1}^B - w_{k+1}) = \lim_{k\in\mathcal{K}}(\mu_k^B y_k^E - (S_{k+1} + \mu_k^B I)w_{k+1}. \tag{3.14}$$

Also, since $|\mathcal{O}| < \infty$, we know that the updating strategy for y_k^E in Algorithm 3 ensures (see (3.5)) that $\{y_k^E\}$ is uniformly bounded over all k. It then follows from (3.14), uniform boundedness of $\{y_k^E\}$, and $\lim_{k\to\infty} \mu_k^B = 0$ that

$$0 = \lim_{k\in\mathcal{K}}([s_{k+1}]_i + \mu_k^B)[w_{k+1}]_i.$$

Since $\lim_{k\in\mathcal{K}}[s_{k+1}]_i = [s^*]_i > 0$ and $\lim_{k\to\infty} \mu_k^B = 0$, we have $\lim_{k\in\mathcal{K}}[w_{k+1}]_i = 0$.

We have now established that (x^*, s^*) is a CAKKT point by using the sequence $\{(x_{k+1}, s_{k+1}, u_{k+1}, z_{k+1})\}_{k\in\mathcal{K}} = \{(x_{k+1}, s_{k+1}, y_{k+1}, w_{k+1})\}_{k\in\mathcal{K}}$ in Definition 3.1. We can then note that the CAKKT conditions (3.8a) and (3.8b) show that $\{\chi_{\text{opt}}(v_{k+1}\}_{k\in\mathcal{K}} \to 0$, while $\{s_{k+1}\}_{k\in\mathcal{K}} \to s^* \ge 0$, (3.8c), and (3.8d) show that $\{\chi_{\text{comp}}(v_{k+1}\}_{k\in\mathcal{K}} \to 0$. Since we also have $\{\chi_{\text{feas}}(v_{k+1})\}_{k\in\mathcal{K}} \to 0$ as a consequence of $c(x^*) - s^* = 0$, we may conclude that $\{\chi(v_{k+1})\}_{k\ni\mathcal{K}} \to 0$. Using this fact, (3.13), and the condition checked in Step 9 in Algorithm 3, we must conclude that $k \in \mathcal{O}$ for all sufficiently large $k \in \mathcal{K} \subseteq \mathcal{M}$. This is a contradiction since $\mathcal{O} \cap \mathcal{M} = \emptyset$, which establishes the desired result that $c(x^*) - s^* \ne 0$. ∎

We now investigate the fact that limit points of M-iterates are not feasible for the general constraints. As we will see, this involves consideration of limit points that are infeasible for the general constraints but solve the problem

$$\underset{x,s}{\text{minimize}} \;\; \tfrac{1}{2}\|c(x) - s\|_2^2 \;\; \text{subject to} \;\; s \ge 0. \tag{3.15}$$

The optimality conditions for problem (3.15) are

$$J(x^*)^T\big(c(x^*)-s^*\big)=0, \qquad s^*\ge 0, \tag{3.16a}$$
$$s^*\cdot\big(c(x^*)-s^*\big)=0, \qquad c(x^*)-s^*\le 0. \tag{3.16b}$$

We may now define an infeasible stationary point.

Definition 3.2. *The pair* (x^*,s^*) *is an* infeasible stationary point *if* $c(x^*)-s^*\neq 0$ *and* (x^*,s^*) *satisfies* (3.16)*, i.e.,* (x^*,s^*) *solves problem* (3.15).

The next result shows that when the number of O-iterations is finite and limit points of M-iterates are infeasible, they must be infeasible stationary points.

Lemma 3.6. *If* $|\mathcal{O}|<\infty$ *and there is an infinite sequence* $\mathcal{K}\subseteq\mathcal{M}$ *such that* $\lim_{k\in\mathcal{K}}(x_{k+1},s_{k+1})=(x^*,s^*)$*, then* (x^*,s^*) *is an infeasible stationary point.*

Proof. Since Lemma 3.5 establishes that $c(x^*)-s^*\neq 0$, we know that condition (3.6) when checked in Step 13 of Algorithm 3 will not hold for all $k\in\mathcal{K}\subseteq\mathcal{M}$ sufficiently large, and therefore it follows from the subsequent updates that $\{\mu_k^P\}\to 0$. Also, since $|\mathcal{O}|<\infty$, the updates to y_k^E in Algorithm 3 guarantee the boundedness of $\{y_k^E\}$ (in particular, see the update (3.5) used during M-iterations). Combining (2.9), boundedness of $\{y_k^E\}$, $\{\mu_k^P\}\to 0$, and Lemma 3.4 shows that

$$\{c(x_{k+1)})-s_{k+1})\}_{k\in\mathcal{K}}\le\{\mu_k^P(y_k^E+\tfrac{1}{2}(w_{k+1}-y_{k+1})\}_{k\in\mathcal{K}}\to 0$$

so that $c(x^*)-s^*\le 0$. Thus, the second condition in (3.16b) holds.

We next show that $s^*\ge 0$, which will verify that the second condition in (3.16a) holds. We first observe that $\{\tau_k\}\to 0$ since $\mathcal{K}\subseteq\mathcal{M}$ is infinite (on such iterations $\tau_{k+1}\leftarrow\frac{1}{2}\tau_k$). We can also see that the update in Step 17 of Algorithm 3 is tested infinitely often since $|\mathcal{M}|=\infty$. In particular, it follows from the condition (3.7) that is checked for each $k\in\mathcal{M}$ and the subsequent update to the barrier parameter that either $\{\mu_{k+1}^B\}\to 0$ or $\{\min(s_{k+1},0)\}_{k\in\mathcal{K}}\to 0$ (we used $\{\tau_k\}\to 0$ to deduce the second possible outcome). Given that $s_{k+1}+\mu_k^B e>0$ is enforced by Line 8 of Algorithm 2, regardless of which of these two outcomes occurs, we have $s^*\ge 0$.

We now prove the first part in (3.16a). To this end, we first note that the limit $\{\nabla M(v_{k+1}\,;y_k^E,\mu_k^P,\mu_k^B)\}_{k\in\mathcal{K}}\to 0$ holds as a result of condition (3.4) holding for all $k\in\mathcal{M}$ (see Step 11 of Algorithm 3). Using this fact, multiplying $\nabla_x M(v_{k+1}\,;y_k^E,\mu_k^P,\mu_k^B)$ by μ_k^P, and using the definitions in (3.10) gives

$$\big\{\mu_k^P g_{k+1}-J_{k+1}^T\big(\mu_k^P\pi_{k+1}^P+\mu_k^P(\pi_{k+1}^P-y_{k+1})\big)\big\}_{k\in\mathcal{K}}\to 0.$$

Combining this with $\{x_{k+1}\}_{k\in\mathcal{K}}\to x^*$, $\mu_k^P\to 0$, and Lemma 3.4 we arrive at

$$\{-J_{k+1}^T(\mu_k^P\pi_{k+1}^P)\}_{k\in\mathcal{K}}=\{-J_{k+1}^T(\mu_k^P y_k^E-c_{k+1}+s_{k+1})\}_{k\in\mathcal{K}}\to 0.$$

Using this limit, boundedness of $\{y_k^E\}$, $\{\mu_k^P\}\to 0$, and $\{(x_{k+1},s_{k+1}\}_{k\in\mathcal{K}}\to(x^*,s^*)$ establishes that the first condition in (3.16a) holds.

Lastly, we show that the first condition in (3.16b) holds. From Lemma 3.4 we know that $\{\pi^B_{k+1} - \pi^P_{k+1}\}_{k\in\mathcal{K}} \to 0$. Moreover, the limiting value does not change if we multiple the previous sequence (term by term) by the bounded sequence $\{\mu^P_k(s_{k+1} + \mu^B_k)\}$ (recall that $\{s_{k+1}\}_{k\in\mathcal{K}} \to s^*$), which results in

$$\{\mu^B_k\mu^P_k y^E_k - \mu^P_k(s_{k+1} + \mu^B_k)y^E_k + (s_{k+1} + \mu^B_k)(c_{k+1} - s_{k+1})\}_{k\in\mathcal{K}} \to 0.$$

From this limit, $\{\mu^P_k\} \to 0$, $\{s_{k+1}\}_{k\in\mathcal{K}} \to s^*$, and bounded $\{y^E_k\}$ we find that

$$\{(s_{k+1} + \mu^B_k)(c_{k+1} - s_{k+1})\}_{k\in\mathcal{K}} \to 0. \tag{3.17}$$

By assumption, there exists i such that $[c(x^*) - s^*]_i \neq 0$. Combining this with $\{(x_{k+1}, s_{k+1})\}_{k\in\mathcal{K}} \to (x^*, s^*)$ and (3.17) shows that $\{[s_{k+1} + \mu^B_k e]_i\}_{k\in\mathcal{K}} \to 0$. Since we have already established that $s^* \geq 0$, we must therefore conclude that $\{\mu^B_k\}_{k\in\mathcal{K}} \to 0$, but since $\{\mu^B_k\}$ is a monotonically decreasing sequence, we in fact know that $\{\mu^B_k\} \to 0$. Using this fact, (3.17), and $\{(x_{k+1}, s_{k+1})\}_{k\in\mathcal{K}} \to (x^*, s^*)$ it follows that $s^* \cdot \big(c(x^*) - s^*\big) = 0$ as desired. ∎

We now state our overall convergence result.

Theorem 3.1. *One of the following outcomes occurs.*

(i) $|\mathcal{O}| = \infty$, *limit points of* $\{(x_{k+1}, s_{k+1})\}_{k\in\mathcal{O}}$ *exist, and every such limit point* (x^*, s^*) *is a CAKKT point for problem* (2.1). *If, in addition, the CCP holds at* (x^*, s^*), *then* (x^*, s^*) *is a KKT point for problem* (2.1).

(ii) $|\mathcal{O}| < \infty$, $|\mathcal{M}| = \infty$, *limit points of* $\{(x_{k+1}, s_{k+1})\}_{k\in\mathcal{M}}$ *exist, and every such limit point* (x^*, s^*) *is an infeasible stationary point as given by Definition* 3.2.

Proof. We consider two cases.
Case 1: $|\mathcal{O}| = \infty$. It follows from Lemmas 3.2 and 3.1 that part (i) occurs.
Case 2: $|\mathcal{O}| < \infty$. In this case, it follows from Lemma 3.3 that $|\mathcal{M}| = \infty$, and from the updating strategy in Algorithm 3 (specifically, see (3.5)) that $\{y^E_k\}$ is bounded. We next show that $\{s_{k+1}\}_{k\in\mathcal{M}}$ is bounded. For a proof by contradiction, suppose that $\{s_{k+1}\}_{k\in\mathcal{M}}$ is unbounded. It follows that there exists a component i and a subsequence $\mathcal{K} \subseteq \mathcal{M}$ such that $\{[s_{k+1}]_i\}_{k\in\mathcal{K}} \to \infty$. This implies that $\{[\pi^B_{k+1}]_i\}_{k\in\mathcal{K}} \to 0$ (see (3.10)) since $\{y^E_k\}$ is bounded and $\{\mu^B_k\}$ is positive and monotonically decreasing, which in turn implies in view of Lemma 3.4 that $\{[\pi^P_{k+1}]_i\}_{k\in\mathcal{K}} \to 0$. However, this latter limit, $\{[s_{k+1}]_i\}_{k\in\mathcal{K}} \to \infty$, and $\{y^E_k\}$ bounded implies that $\{[c(x_{k+1})]_i\}_{k\in\mathcal{K}} \to \infty$, which is impossible when Assumptions 2.3 and 2.1 hold. Thus, it must hold that $\{s_{k+1}\}_{k\in\mathcal{M}}$ is bounded. Combining this with Assumption 2.3 ensures the existence of at least one limit point of $\{(x_{k+1}, s_{k+1})\}_{k\in\mathcal{M}}$. Let (x^*, s^*) be any such limit point so that there exists $\mathcal{K} \subseteq \mathcal{M}$ such that $\{(x_{k+1}, s_{k+1})\}_{k\in\mathcal{K}} \to (x^*, s^*)$. Since $|\mathcal{O}| < \infty$, it follows from Lemma 3.6 that (x^*, s^*) is an infeasible stationary point. ∎

4. Implementation details

We first discuss the choice of the matrix H_k^M in (2.4). Motivated by the discussion in Section 2.2, a natural choice would be to use the matrix $B(x,s,w\,;\pi^P,\pi^B)$ defined in (2.5). With this choice for H_k^M, however, the direction Δv resulting from solving $H_k^M\Delta v = -\nabla M(x,s,y,w\,;y^E,\mu^P,\mu^B)$ would only be guaranteed to be a decent direction for M at $v=(x,s,y,w)$ if $B(x,s,w\,;\pi^P,\pi^B)$ is positive definite. Since this is not guaranteed to be the case, we propose the choice

$$H_k^M = \begin{pmatrix} \widehat{H} + 2J^TD_P^{-1}J & -2J^TD_P^{-1} & J^T & 0 \\ -2D_P^{-1}J & 2(D_P^{-1}+D_B^{-1}) & -I & I \\ J & -I & D_P & 0 \\ 0 & I & 0 & D_B \end{pmatrix}, \tag{4.1}$$

where the symmetric matrix $\widehat{H}$ is chosen to ensure that H_k^M is positive definite. The next result uses the following definitions

$$\bar{H} = \begin{pmatrix} \widehat{H} & 0 \\ 0 & 0 \end{pmatrix}, \quad \bar{J} = \begin{pmatrix} J & -I \\ 0 & I \end{pmatrix}, \quad \bar{D} = \begin{pmatrix} D_P & 0 \\ 0 & D_B \end{pmatrix} \text{ and } H_{\text{ind}} = \begin{pmatrix} \bar{H} & \bar{J}^T \\ \bar{J} & -\bar{D} \end{pmatrix}$$

to shed light on how the matrix $\widehat{H}$ may be chosen.

Theorem 4.1. *The following inertia relationships hold when H_k^M is defined by (4.1):*

$$\text{In}(H_k^M) = \text{In}(\bar{H} + \bar{J}^T\bar{D}^{-1}\bar{J}) + (2m,0,0) \quad \textit{and} \tag{4.2}$$

$$\text{In}(H_{\text{ind}}) = \text{In}(\bar{H} + \bar{J}^T\bar{D}^{-1}\bar{J}) + (0,2m,0). \tag{4.3}$$

Moreover, the matrix H_k^M is positive definite if and only if $\text{In}(H_{\text{ind}}) = (m+n, 2m, 0)$.

Proof. By defining the nonsingular matrices T_1 and T_2 by

$$T_1 = \begin{pmatrix} I & 0 \\ -\bar{D}^{-1}\bar{J} & I \end{pmatrix} \quad \text{and} \quad T_2 = \begin{pmatrix} I & 0 \\ \bar{D}^{-1}\bar{J} & I \end{pmatrix}$$

and using Sylvesters law of inertia, we have that

$$\text{In}(H_k^M) = \text{In}(T_1^TH_k^MT_1) = \text{In}\begin{pmatrix} \bar{H} + \bar{J}^T\bar{D}^{-1}\bar{J} & 0 \\ 0 & \bar{D} \end{pmatrix} = \text{In}(\bar{H} + \bar{J}\bar{D}^{-1}\bar{J}) + (2m,0,0)$$

and

$$\text{In}(H_{\text{ind}}) = \text{In}(T_2^TH_k^MT_2) = \text{In}\begin{pmatrix} \bar{H} + \bar{J}^T\bar{D}^{-1}\bar{J} & 0 \\ 0 & -\bar{D} \end{pmatrix} = \text{In}(\bar{H}+\bar{J}^T\bar{D}^{-1}\bar{J})+(0,2m,0),$$

which establishes (4.2) and (4.3). The final conclusion about when H_k^M is positive definite follows from (4.2) and (4.3). ∎

The next result shows that the choice $\widehat{H} = H(x,y)$ is allowed in the neighborhood of a solution satisfying second-order sufficient conditions.

Theorem 4.2. *The matrix H_k^M in (4.1) with the choice $\widehat{H} = H(x,y)$ is positive definite for all $u = (x, s, y, w, y^E, \mu^P, \mu^B)$ sufficiently close to $u^* = (x^*, s^*, y^*, w^*, y^*, 0, 0)$, when (x^*, s^*, y^*, w^*) is a solution to (2.1) that satisfies second-order sufficient optimality conditions and strict complementarity.*

Proof. The inertia of the indefinite matrix $H_{\rm ind}$ may be determined from the factorization $K_{\rm ind} = S\Omega S^T$, where

$$S = \begin{pmatrix} 0 & 0 & I & 0 \\ I & 0 & 0 & 0 \\ -D_B & -I & 0 & I \\ 0 & I & 0 & 0 \end{pmatrix} \quad \text{and} \quad \Omega = \begin{pmatrix} 0 & I & 0 & 0 \\ I & -D_B & 0 & 0 \\ 0 & 0 & H & J^T \\ 0 & 0 & J & -(D_B + D_P) \end{pmatrix},$$

and we have used $J = J(x)$ and $H = H(x,y)$. As S is nonsingular, Sylvester's law of inertia and [16, Theorem 1.5.1 and Lemma 1.5.3] imply that

$$\begin{aligned} \mathrm{In}(H_{\rm ind}) &= \mathrm{In}\begin{pmatrix} 0 & I \\ I & -D_B \end{pmatrix} + \mathrm{In}\begin{pmatrix} H & J^T \\ J & -(D_P + D_B) \end{pmatrix} \\ &= (m, m, 0) + \mathrm{In}\big(H + J^T(D_P + D_B)^{-1}J^T\big) + (0, m, 0) \\ &= \mathrm{In}\big(H + J^T(D_P + D_B)^{-1}J^T\big) + (m, 2m, 0). \end{aligned} \tag{4.4}$$

We now show that $H + J^T(D_P + D_B)^{-1}J \succ 0$ under the assumptions in the theorem.

Let (x^*, s^*, y^*, w^*) be a solution to problem (2.1) that satisfies strict complementarity and second-order sufficiency optimality condition. Among other conditions, this means that

$$\max\{s^*, w^*\} > 0 \quad \text{and} \tag{4.5}$$

$$p^T H(x^*, y^*) p > 0 \quad \text{for all } p \neq 0 \text{ satisfying } J_{\mathcal{A}}(x^*)p = 0, \tag{4.6}$$

where

$$\mathcal{A} = \{i : [\, s^* \,]_i = 0\} \equiv \{i : [\, c(x^*) \,]_i = 0\}. \tag{4.7}$$

We may now write

$$\begin{aligned} H + J^T(D_P + D_B)^{-1}J = H &+ J_{\mathcal{A}}^T \left(\mu^P I + (S_{\mathcal{A}} + \mu^B I)W_{\mathcal{A}}^{-1}\right)^{-1} J_{\mathcal{A}} \\ &+ J_{\mathcal{I}}^T \left(\mu^P I + (S_{\mathcal{I}} + \mu^B I)W_{\mathcal{I}}^{-1}\right)^{-1} J_{\mathcal{I}}, \end{aligned} \tag{4.8}$$

where $\mathcal{I} = \{1, 2, \ldots, m\} \setminus \mathcal{A}$. Note that the following limits hold, where u and u^* are defined in the statement of the theorem:

$$\lim_{u \to u^*} J(x) = J(x^*), \qquad \lim_{u \to u^*} H(x,y) = H(x^*, y^*), \quad \text{and} \tag{4.9}$$

$$\lim_{u \to u^*} \left[\mu^P I + (S_{\mathcal{A}} + \mu^B I)W_{\mathcal{A}}^{-1}\right]_{ii}^{-1} = \infty. \tag{4.10}$$

for all $1 \le i \le m$. Using (4.9), (4.10), (4.6), and the same proof as in [10, Theorem 3.1], we can conclude that

$$H + J_{\mathcal{A}}^T \left(\mu^P I + (S_{\mathcal{A}} + \mu^B I)W_{\mathcal{A}}^{-1}\right)^{-1} J_{\mathcal{A}} \succ 0 \quad \text{for all } u \text{ sufficiently close to } u^*.$$

Combining this with the fact that the matrix $J_{\mathcal{I}}^T\left(\mu^P I + (S_{\mathcal{I}} + \mu^B I)W_{\mathcal{I}}^{-1}\right)^{-1} J_{\mathcal{I}} \succeq 0$ and (4.8), we know that the matrix $H + J^T(D_P + D_B)^{-1}J \succ 0$ for all u sufficiently close to u^*. Finally, we may combine this result with (4.4) to say that $\mathrm{In}(H_{\mathrm{ind}}) = (n+m, 2m, 0)$ for all u sufficiently close to u^*, which with Theorem 4.2 gives that H_k^M is positive definite for all u sufficiently close to u^*. ∎

Finally, we show that the cost of solving the modified Newton equations (2.6) is dominated by the cost of solving a symmetric system of order $n+m$. The equations (2.6) may be written in the symmetric form

$$\begin{pmatrix} H & 0 & J^T & 0 \\ 0 & 0 & -I & I \\ J & -I & -D_P & 0 \\ 0 & I & 0 & -D_B \end{pmatrix} \begin{pmatrix} \Delta x \\ \Delta s \\ -\Delta y \\ -\Delta w \end{pmatrix} = -\begin{pmatrix} g - J^T y \\ y - w \\ c - s + \mu^P(y - y^E) \\ W^{-1}\big(s \cdot w + \mu^B(w - y^E)\big) \end{pmatrix}, \tag{4.11}$$

where D_P and D_B are the positive-definite diagonal matrices $D_P = \mu^P I$ and $D_B = (S + \mu^B I)W^{-1}$. The solution of (4.11) is given by

$$\Delta w = y - w + \Delta y \quad \text{and} \quad \Delta s = -W^{-1}\big(s \cdot (y + \Delta y) + \mu^B(y + \Delta y - y^E)\big),$$

where Δx and Δy satisfy the equations

$$\begin{pmatrix} H & -J^T \\ J & D_P + D_B \end{pmatrix} \begin{pmatrix} \Delta x \\ \Delta y \end{pmatrix} = -\begin{pmatrix} g - J^T y \\ c - s + \mu^P(y - y^E) + W^{-1}\big(s \cdot y + \mu^B(y - y^E)\big) \end{pmatrix}.$$

The identity $w + \Delta w = y + \Delta y$ implies that if the initial values y_0 and w_0 are chosen to be identical, then all subsequent iterates will satisfy $w = y$, so that the path-following equations simplify to

$$\Delta w = \Delta y \quad \text{and} \tag{4.12}$$
$$\Delta s = -Y^{-1}\big(s \cdot (y + \Delta y) + \mu^B(y + \Delta y - y^E)\big), \tag{4.13}$$

where Δx and Δy satisfy the (symmetrized) equations

$$\begin{pmatrix} H & J^T \\ J & -(D_P + D_B) \end{pmatrix} \begin{pmatrix} \Delta x \\ -\Delta y \end{pmatrix} = -\begin{pmatrix} g - J^T y \\ c - s + \mu^P(y - y^E) + Y^{-1}\big(s \cdot y + \mu^B(y - y^E)\big) \end{pmatrix}.$$

5. Conclusions

In this paper a new primal-dual shifted penalty-barrier function has been proposed for solving inequality constrained nonlinear optimization problems. An algorithm has been presented in which at each step, a single linear system is solved to obtain a descent direction for this merit function, and a set of updating procedures ensure successive progress towards a solution.

References

[1] R. Andreani, J. M. Martinez, A. Ramos, and P. J. Silva. A cone-continuity constraint qualification and algorithmic consequences. *Optimization Online*, 2015. 15

[2] R. Andreani, J. M. Martínez, and B. F. Svaiter. A new sequential optimality condition for constrained optimization and algorithmic consequences. *SIAM J. Optim.*, 20(6):3533–3554, 2010. 14

[3] M. G. Breitfeld and D. F. Shanno. Computational experience with penalty-barrier methods for nonlinear programming. *Ann. Oper. Res.*, 62:439–463, 1996. 2

[4] A. R. Conn, N. I. M. Gould, and Ph. L. Toint. A globally convergent Lagrangian barrier algorithm for optimization with general inequality constraints and simple bounds. *Math. Comput.*, 66(217):261–288, S1–S11, 1997. 2

[5] A. V. Fiacco. Barrier methods for nonlinear programming. In A. Holzman, editor, *Operations Research Support Methodology*, pages 377–440, New York, 1979. Marcel Dekker. 2

[6] A. V. Fiacco and G. P. McCormick. *Nonlinear Programming: Sequential Unconstrained Minimization Techniques*. John Wiley and Sons, Inc., New York-London-Sydney, 1968. 2

[7] A. Forsgren, P. E. Gill, and M. H. Wright. Interior methods for nonlinear optimization. *SIAM Rev.*, 44:525–597, 2002. 2

[8] K. R. Frisch. The logarithmic potential method of convex programming. Memorandum of May 13, University Institute of Economics, Oslo, Norway, 1955. 2

[9] P. E. Gill, W. Murray, M. A. Saunders, and M. H. Wright. Shifted barrier methods for linear programming. Report SOL 88-9, Department of Operations Research, Stanford University, Stanford, CA, 1988. 2

[10] P. E. Gill and D. P. Robinson. A primal-dual augmented Lagrangian. *Comput. Optim. Appl.*, 51:1–25, 2012. 22

[11] D. Goldfarb, R. Polyak, K. Scheinberg, and I. Yuzefovich. A modified barrier-augmented Lagrangian method for constrained minimization. *Comput. Optim. Appl.*, 14(1):55–74, 1999. 2

[12] S. G. Nash, R. Polyak, and A. Sofer. Numerical comparison of barrier and modified-barrier methods for large-scale bound-constrained optimization. In D. W. Hearn and P. M. Pardalos, editors, *Large-Scale Optimization: State of the Art*, pages 319–338. Kluwer, Dordrecht, 1994. 2

[13] J. M. Ortega and W. C. Rheinboldt. *Iterative solution of nonlinear equations in several variables*. Academic Press, New York, 1970. 12

[14] R. Polyak. Modified barrier functions (theory and methods). *Math. Program.*, 54(2, Ser. A):177–222, 1992. 2

[15] L. Qi and Z. Wei. On the constant positive linear dependence condition and its application to SQP methods. *SIAM J. Optim.*, 10(4):963–981, 2000. 15

[16] D. P. Robinson. *Primal-Dual Methods for Nonlinear Optimization*. PhD thesis, Department of Mathematics, University of California San Diego, La Jolla, CA, 2007. 22

[17] M. H. Wright. Interior methods for constrained optimization. In *Acta Numerica, 1992*, pages 341–407. Cambridge University Press, New York, USA, 1992. 2