A REGULARIZED SQP METHOD WITH CONVERGENCE TO SECOND-ORDER OPTIMAL POINTS

PHILIP E. GILL*, VYACHESLAV KUNGURTSEV†, AND DANIEL P. ROBINSON‡

Abstract. Regularized and stabilized sequential quadratic programming methods are two classes of sequential quadratic programming (SQP) methods designed to resolve the numerical and theoretical difficulties associated with ill-posed or degenerate nonlinear optimization problems. Recently, a regularized SQP method has been proposed that provides a strong connection between augmented Lagrangian methods and stabilized SQP methods. The method is formulated as a regularized SQP method with an implicit safeguarding strategy based on minimizing a bound-constrained primal-dual augmented Lagrangian. Each iteration involves the solution of a regularized quadratic program (QP) that is equivalent to a strictly convex bound-constrained QP based on minimizing a quadratic model of the augmented Lagrangian. The solution of the QP subproblem defines a descent direction for a flexible line search that provides a sufficient decrease in a primal-dual augmented Lagrangian merit function. Under certain conditions, the method is guaranteed to converge to a point satisfying the first-order Karush-Kuhn-Tucker (KKT) conditions. In this paper, the regularized SQP method is extended to allow convergence to points satisfying certain *second-order* KKT conditions. The method is based on performing a flexible line search along a direction formed from the solution of a strictly convex regularized quadratic programming subproblem and, when one exists, a direction of negative curvature for the primal-dual augmented Lagrangian. It is shown that with an appropriate choice of termination condition, the method terminates in a finite number of iterations. As in the first-order case, the method is formulated as a regularized SQP method with an augmented Lagrangian safeguarding strategy. It is shown that this safeguarding becomes relevant only when the iterates are converging to an infeasible stationary point of the constraint violations. Otherwise, the method terminates with a point that either satisfies the second-order necessary conditions for optimality, or fails to satisfy a weak second-order constraint qualification.

Key words. Nonlinear programming, augmented Lagrangian, sequential quadratic programming, SQP methods, stabilized SQP, regularized methods, primal-dual methods, second-order optimality.

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

1. Introduction.

This paper is concerned with computing solutions to the nonlinear optimization problem

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

where $c\colon \mathbb{R}^n \mapsto \mathbb{R}^m$ and $f\colon \mathbb{R}^n \mapsto \mathbb{R}$ are twice-continuously differentiable. This problem format assumes that all general inequality constraints have been converted to equalities by the use of slack variables. Methods for solving problem (NP) are easily extended to the more general setting with $l \leq x \leq u$.

Sequential quadratic programming (SQP) methods are a popular class of methods for nonlinearly constrained optimization (for a survey, see, e.g., [6, 14]). They are particularly

*Department of Mathematics, University of California, San Diego, La Jolla, CA 92093-0112 (pgill@ucsd.edu). Research supported in part by National Science Foundation grant DMS-1318480, and by Department of Energy grant DE-SC0002349.

†Electrical Engineering Department (ESAT-STADIUS) and the Optimization in Engineering Center (OPTEC), KU Leuven, Kasteelpark Arenberg 10, B-3001 Leuven-Heverlee, Belgium (vkungurt@esat.kuleuven.be). Research supported by Research Council KUL: PFV/10/002 Optimization in Engineering Center, GOA/10/09 MaNet, Belgian Federal Science Policy Office: IUAP P7 (DYSCO, Dynamical Systems, Control and Optimization, 2012–2017); ERC ST HIGHWIND (259 166).

‡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.

effective for solving a sequence of related problems, such as those arising in mixed-integer nonlinear programming and the optimization of functions subject to differential equation constraints. Conventional SQP methods find an approximate solution of a sequence of quadratic programming (QP) subproblems in which a quadratic model of the objective function is minimized subject to the linearized constraints. Convergence from any starting point is enforced by requiring the improvement in some merit function at each step.

Regularized and stabilized SQP methods are designed to resolve the numerical and theoretical difficulties associated with ill-posed or degenerate nonlinear optimization problems (see, e.g., [15, 23, 24, 25, 10, 16]). In recent papers, Gill and Robinson [12, 13] have proposed a regularized SQP method that establishes a strong connection between augmented Lagrangian methods and stabilized SQP methods. The method is formulated as a regularized SQP method with an implicit safeguarding strategy based on minimizing a bound-constrained primal-dual augmented Lagrangian. Each iteration involves the solution of a strictly convex bound-constrained quadratic programming (QP) subproblem that is equivalent to a regularized quadratic program in which the dual variables appear explicitly. The solution of the QP subproblem defines a descent direction for a flexible line search that provides a sufficient decrease in a primal-dual augmented Lagrangian merit function. Under certain conditions, the method is guaranteed to converge to a point satisfying the first-order Karush-Kuhn-Tucker (KKT) conditions.

A substantial benefit of SQP methods is that they are easily adapted to accept an estimate of the active set. However, if such "warm starts" are to be exploited fully, it is necessary that the second derivatives of the problem functions are available and that these derivatives are utilized by the SQP method. Unfortunately, it is difficult to formulate a conventional SQP method that uses exact second derivatives. The main difficulty stems from the possibility that the Hessian of the Lagrangian is indefinite, in which case the inequality constrained QP subproblem is nonconvex. Unfortunately, nonconvex quadratic programming is NP-hard, and even the seemingly simple task of checking for local optimality is intractable when there are zero Lagrange multipliers. In addition, a nonconvex QP is likely to have many local solutions, and may be unbounded. For these reasons, some SQP methods define a convex subproblem by using a positive-definite quasi-Newton approximation for the Hessian. Other methods rely on a nonconvex QP solver, but require the calculation of a global solution of the QP subproblem, which has the benefit of ensuring that the same local solution is found for the final sequence of related QPs.

The method of Gill and Robinson [13] is able to use exact second derivative information by incorporating a convexification algorithm that defines a strictly convex QP subproblem without requiring that the Hessian of the Lagrangian be positive definite in the neighborhood of a solution. However, the Gill-Robinson method does not exploit another important advantage of using second derivatives: the potential of being able to check if the iterates are converging to a point that satisfies the second-order conditions for optimality, and, if not, to move away from such points. This capability is useful when solving highly nonconvex problems for which some confidence is needed that a local minimizer has been found (see, e.g., [8, 1, 7]).

This paper considers the formulation and analysis of a regularized SQP method that is guaranteed to converge to a point satisfying certain second-order KKT conditions (for a precise definition, see Section 3). The method is an extension of the Gill-Robinson method that is based on performing a flexible line search along a direction formed from the solution of a strictly convex regularized quadratic programming subproblem and, when one exists, a direction of negative curvature for the primal-dual augmented Lagrangian. It is shown

that with an appropriate choice of termination condition, the method terminates in a finite number of iterations. As in the first-order case, the method is formulated as a regularized SQP method with an augmented Lagrangian safeguarding strategy. It is shown that safeguarding becomes relevant only when the iterates are converging to an infeasible stationary point of the constraint violations. Otherwise, the method terminates with a point that either satisfies the second-order necessary conditions for optimality, or fails to satisfy a weak second-order constraint qualification.

Additional contributions of this paper include a proof of convergence to first-order optimal points under weaker conditions than those assumed in [13]. Also, a new proof is given for the result that minimizers of problem (NP) are also minimizers of the bound-constrained primal-dual augmented Lagrangian function. The new proof requires a less restrictive form of the second-order sufficient conditions.

The remainder of the paper is organized as follows. This section concludes with a review of the first-order optimality conditions for (NP) and the role of the primal-dual augmented Lagrangian function in the formulation of regularized SQP methods. Section 2 provides the details of the second-order primal-dual regularized SQP method. The global convergence of the method is established in Section 3. Section 4 provides final comments and conclusions.

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. The least eigenvalue of a symmetric matrix A will be denoted by $\lambda_{\min}(A)$. 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., $[\,v_{\scriptscriptstyle F}\,]_i$ is the ith component of the vector $v_{\scriptscriptstyle F}$. 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 $i \in \mathcal{S}$. Similarly, if M is a symmetric matrix, then $[\,M\,]_{\mathcal{S}}$ denotes the symmetric matrix with elements m_{ij} for $i,\, j \in \mathcal{S}$. A local solution of problem (NP) 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 (NP) is $L(x, y, z) = f(x) - c(x)^T y - z^T x$, where y and z are m- and n-vectors of dual variables associated with the equality constraints and bounds, respectively. 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 $\alpha_j = O\big(\beta_j\big)$.

1.2. Background. If a suitable constraint qualification holds, then a local minimizer for (NP) is a first-order KKT point.

Definition 1.1 (First-order KKT point). *The vector x^* is called a* first-order KKT point *for problem* (NP) *if there exists a y^* such that $r(x^*, y^*) = 0$, where*

$$r(x, y) = \left\| \begin{pmatrix} c(x) \\ \min\big(x, g(x) - J(x)^T y\big) \end{pmatrix} \right\|. \tag{1.1}$$

Any (x^*, y^*) *satisfying* $r(x^*, y^*) = 0$*, is called a* first-order KKT pair*, with* y^* *designated as a vector of Lagrange multipliers associated with the constraints* $c(x) = 0$*.*

Given a first-order KKT point x^*, we define $\mathcal{Y}(x^*)$ to be the nonempty set of Lagrange multiplier vectors

$$\mathcal{Y}(x^*) = \{y \in \mathbb{R}^m : (x^*, y) \text{ satisfies } r(x^*, y) = 0\}. \tag{1.2}$$

It is well-known that the first-order KKT conditions may be characterized in terms of the index sets of active and free components of the primal variables x. If $x \geq 0$, the *active set* and *free set* at x are defined as

$$\mathcal{A}(x) = \{i : x_i = 0\} \quad \text{and} \quad \mathcal{F}(x) = \{1, 2, \ldots, n\} \setminus \mathcal{A}(x), \tag{1.3}$$

respectively. At a primal-dual first-order solution (x^*, y^*), computationally tractable second-order conditions may be defined in terms of the disjoint index sets associated with the zero and strictly positive components of the dual vector y^*. The index set of *strongly active* variables is given by

$$\mathcal{A}_+(x^*, y^*) = \{i : [\, g(x^*) - J(x^*)^T y^* \,]_i > 0\} \tag{1.4}$$

and, similarly, the set of *weakly active* variables is

$$\mathcal{A}_0(x^*, y^*) = \{i : [\, g(x^*) - J(x^*)^T y^* \,]_i = 0\}. \tag{1.5}$$

Definition 1.2 (Second-order sufficient conditions). *A primal-dual pair* (x^*, y^*) *satisfies the* second-order sufficient optimality conditions *for problem* (NP) *if it satisfies the first-order conditions* $r(x^*, y^*) = 0$ (cf. (1.1)) and

$$d^T H(x^*, y^*) d > 0 \text{ for all } d \in \mathcal{C}(x^*, y^*) \setminus \{0\}, \tag{1.6}$$

where $\mathcal{C}(x^*, y^*)$ is the critical cone $\mathcal{C}(x^*, y^*) := \text{null}\big(J(x^*)\big) \cap \mathcal{C}_{\scriptscriptstyle M}(x^*, y^*)$, with

$$\mathcal{C}_{\scriptscriptstyle M}(x^*, y^*) = \{d : d_i = 0 \text{ for } i \in \mathcal{A}_+(x^*, y^*),\ d_i \geq 0 \text{ for } i \in \mathcal{A}_0(x^*, y^*)\}.$$

The proposed algorithm is based on the first-order primal-dual SQP method of Gill and Robinson [13]. The line-search direction is augmented by a direction of negative curvature that facilitates convergence to points that satisfy the second-order necessary conditions for optimality. In addition to the new second-order features, the relevant aspects of the first-order algorithm are also included for clarity.

The regularized SQP line-search method is based on the properties of the primal-dual augmented Lagrangian function

$$M(x, y\,; y^{\scriptscriptstyle E}, \mu) = f(x) - c(x)^T y^{\scriptscriptstyle E} + \frac{1}{2\mu}\|c(x)\|^2 + \frac{\nu}{2\mu}\|c(x) + \mu(y - y^{\scriptscriptstyle E})\|^2, \tag{1.7}$$

where ν is a nonnegative scalar, μ is a positive penalty parameter, and $y^{\scriptscriptstyle E}$ is an estimate of a Lagrange multiplier vector y^*. Gill and Robinson [12] describe how the parameter ν may be used to define various forms of the augmented Lagrangian function. However, once ν is chosen, it remains fixed for the definition of a given algorithm and therefore the dependence on ν is omitted from the definition of $M(x, y\,; y^{\scriptscriptstyle E}, \mu)$ and its derivatives. With the notation

$c = c(x)$, $g = g(x)$, and $J = J(x)$, the gradient and Hessian of $M(x, y\,; y^{E}, \mu)$ may be written as

$$\nabla M(x, y\,; y^{E}, \mu) = \begin{pmatrix} g - J^{T}\big(\pi + \nu(\pi - y)\big) \\ \nu\mu(y - \pi) \end{pmatrix}, \tag{1.8}$$

and

$$\nabla^{2} M(x, y\,; y^{E}, \mu) = \begin{pmatrix} H\big(x, \pi + \nu(\pi - y)\big) + \frac{1}{\mu}(1 + \nu)J^{T}J & \nu J^{T} \\ \nu J & \nu\mu I \end{pmatrix}, \tag{1.9}$$

where $\pi \equiv \pi(x\,; y^{E}, \mu)$ denotes the vector-valued function

$$\pi(x\,; y^{E}, \mu) = y^{E} - \frac{1}{\mu}c(x). \tag{1.10}$$

For more details on the properties of the primal-dual augmented Lagrangian, see Robinson [21] and Gill and Robinson [12].

The next result motivates the use of (1.7) as a merit function. For a proof that requires *more restrictive* assumptions on problem (NP), see Robinson [21, Theorem 4.6.1].

THEOREM 1.3. *If (x^{*}, y^{*}) is a solution of problem* (NP) *that satisfies the second-order sufficient optimality conditions given by Definition* 1.2*, then for the choice $y^{E} = y^{*}$, there exists a positive $\bar{\mu}$ such that for all $0 < \mu < \bar{\mu}$, the point (x^{*}, y^{*}) satisfies the second-order sufficient optimality conditions for the primal-dual optimization problem*

$$\underset{x,y}{\text{minimize}} \;\; M(x, y\,; y^{E}, \mu) \;\; \text{subject to} \;\; x \geq 0 \tag{1.11}$$

for every positive value of ν used in the definition (1.7) *of M.*

Proof. As (x^{*}, y^{*}) is a primal-dual solution of problem (NP), it follows that

$$c(x^{*}) = 0, \quad x^{*} \geq 0, \quad z^{*} = g(x^{*}) - J(x^{*})^{T}y^{*} \geq 0, \quad x^{*} \cdot z^{*} = 0, \tag{1.12}$$

and

$$d^{T}H(x^{*}, y^{*})d \geq \tfrac{1}{2}\gamma\|d\|^{2} \;\; \text{for all } d \in \mathcal{C}(x^{*}, y^{*}) \text{ and some } \gamma > 0. \tag{1.13}$$

It follows from (1.12), (1.8), (1.10), and the assumption that $y^{E} = y^{*}$, that

$$\nabla M(x^{*}, y^{*}; y^{*}, \mu) = \begin{pmatrix} g(x^{*}) - J(x^{*})^{T}y^{*} \\ 0 \end{pmatrix} = \begin{pmatrix} z^{*} \\ 0 \end{pmatrix}.$$

Combining these conditions with (1.12) implies that (x^{*}, y^{*}) satisfies that first-order optimality conditions for a solution of the bound-constrained problem (1.11).

It remains to show that (x^{*}, y^{*}) also satisfies the second-order sufficient optimality conditions for the bound-constrained problem (1.11). The result is established by showing that $\mathcal{C}_{M}(x^{*}, y^{*})$ is the critical cone for the bound constraints of (1.11) at a first-order solution. In particular, we show that for $\bar{\gamma} \in (0, 1)$ chosen small enough to satisfy

$$1 + \nu - \frac{\nu}{1 - \bar{\gamma}} > 0, \tag{1.14}$$

there exists a $\bar{\mu} > 0$ such that for every $\mu \in (0, \bar{\mu})$, it holds that

$$w^T \nabla^2 M(x^*, y^*; y^*, \mu) w \geq \tfrac{1}{2}\bar{\gamma}\nu\bar{\mu}\|w\|^2 \ \text{ for all } w = \begin{pmatrix} p \\ q \end{pmatrix} \text{ with } p \in \mathcal{C}_{\scriptscriptstyle M}(x^*, y^*). \tag{1.15}$$

The proof is by contradiction. Consider sequences $\{\mu_k\}$ and $\{w_k\}$ such that

$$\lim_{k\to\infty} \mu_k = 0, \qquad \|w_k\| = \left\| \begin{pmatrix} p_k \\ q_k \end{pmatrix} \right\| = 1, \qquad p_k \in \mathcal{C}_{\scriptscriptstyle M}(x^*, y^*), \tag{1.16}$$

$$\lim_{k\to\infty} w_k = w_*, \qquad \|w_*\| = \left\| \begin{pmatrix} p_* \\ q_* \end{pmatrix} \right\| = 1, \qquad p_* \in \mathcal{C}_{\scriptscriptstyle M}(x^*, y^*), \tag{1.17}$$

and

$$w_k^T \nabla^2 M(x^*, y^*; y^*, \mu_k) w_k < \tfrac{1}{2}\bar{\gamma}\nu\mu_k. \tag{1.18}$$

In order to simplify the notation, let g, J and H denote the quantities $g(x^*)$, $J(x^*)$, and $H(x^*, y^*)$. If the first-order conditions (1.12) are used to simplify the Hessian $\nabla^2 M$ (1.9) evaluated at $(x, y) = (x^*, y^*)$ and $y^{\scriptscriptstyle E} = y^*$, then the curvature condition (1.18) may be written as

$$p_k^T \Big(H + \frac{1+\nu}{\mu_k} J^T J \Big) p_k + 2\nu q_k^T J p_k + \nu\mu_k \|q_k\|^2 < \tfrac{1}{2}\bar{\gamma}\nu\mu_k. \tag{1.19}$$

Multiplying this inequality by μ_k, and taking limits yields

$$(1+\nu) p_*^T J^T J p_* = (1+\nu)\|J p_*\|^2 \leq 0.$$

As $\nu > 0$, it follows that $J p_* = 0$, and since by definition $p_* \in \mathcal{C}_{\scriptscriptstyle M}(x^*, y^*)$, we conclude that $p_* \in \mathcal{C}(x^*, y^*)$.

Next, we re-write the curvature condition (1.19) so that

$$\begin{aligned}
\tfrac{1}{2}\bar{\gamma}\nu\mu_k &> p_k^T \Big(H + \frac{1+\nu}{\mu_k} J^T J \Big) p_k + 2\nu q_k^T J p_k + \nu\mu_k \|q_k\|^2 \\
&= p_k^T H p_k + \Big(1 + \nu - \frac{\nu}{1-\bar{\gamma}} \Big) \frac{1}{\mu_k} p_k^T J^T J p_k + \frac{\nu}{\mu_k - \mu_k\bar{\gamma}} p_k^T J^T J p_k \\
&\qquad + 2\nu q_k^T J p_k + \nu(\mu_k - \mu_k\bar{\gamma})\|q_k\|^2 + \nu\mu_k\bar{\gamma}\|q_k\|^2 \\
&\geq p_k^T H p_k + \Big\| \frac{\sqrt{\nu}}{\sqrt{\mu_k - \mu_k\bar{\gamma}}} J p_k + \sqrt{\nu}\,\sqrt{\mu_k - \mu_k\bar{\gamma}}\, q_k \Big\|^2 + \nu\mu_k\bar{\gamma}\|q_k\|^2 \\
&\geq p_k^T H p_k + \nu\mu_k\bar{\gamma}\|q_k\|^2,
\end{aligned} \tag{1.20}$$

where we have used (1.14) and the inequality $(1/\mu_k) p_k^T J^T J p_k = (1/\mu_k)\|J p_k\|^2 \geq 0$. The combination of the inequality (1.20), the assumption that $\{p_k\} \to p_* \in \mathcal{C}(x^*, y^*)$, and the second-order sufficient condition (1.13) yields

$$\tfrac{1}{2}\bar{\gamma}\nu\mu_k \geq \tfrac{1}{4}\gamma\|p_k\|^2 + \nu\mu_k\bar{\gamma}\|q_k\|^2 \quad \text{for all } k \text{ sufficiently large.} \tag{1.21}$$

Taking limits and using the properties (1.16) and (1.17), yields the limits $\|p_*\| = 0$ and $\|q_*\| = 1$. In addition, the inequality (1.21) implies that

$$\frac{\bar{\gamma}}{2} \geq \frac{\gamma}{4\nu\mu_k}\|p_k\|^2 + \bar{\gamma}\|q_k\|^2 \geq \bar{\gamma}\|q_k\|^2.$$

Taking limits in this inequality and using the properties (1.17) and $\|q_*\| = 1$, gives the inequality $\bar{\gamma}/2 \geq \bar{\gamma}$, which is a contradiction. It follows that (1.15) holds for all $\mu \in (0, \bar{\mu})$ and $w = (p, q)$ such that $p \in \mathcal{C}_M(x^*, y^*)$, which implies that (x^*, y^*) satisfies the second-order sufficient optimality conditions for the bound-constrained problem (1.11) as required. □

2. A Second-Order Primal-Dual Regularized SQP Algorithm. The proposed method is based on solving a sequence of strictly convex regularized QP subproblems in which the dual variables appear explicitly in the objective function and constraints. The solution of each subproblem defines a primal-dual search direction for improving a merit function based on the primal-dual augmented Lagrangian (1.7). The choice of penalty parameters and Lagrange multiplier estimate y^E in the merit function is motivated by the equivalence of the regularized QP to a problem defined by minimizing a quadratic model of the augmented Lagrangian function (1.7) subject to the bound constraints. At each iteration, an improved value of the merit function is found using a flexible line search along a direction composed of a nonzero descent direction defined by the QP subproblem and, when one exists, an approximate direction of negative curvature. This defines an inner/outer iteration structure in which the inner iterations are those of the algorithm used to solve the QP subproblem.

This section starts with a description of the notation, and a brief summary of the five main steps of the algorithm. References to the sections that give a detailed description of the steps are also provided. The complete algorithm is stated formally as Algorithm 3 on page 16.

Overview of the step computation. The flexible line search requires the computation of two primal-dual directions. The first is the solution of a bound-constrained quadratic program with positive-definite Hessian $\widehat{B}$. The second is a direction of negative curvature (if one exists) for an approximate $\nabla^2 M$ defined below. The computations associated with each of these directions are based extensively on the matrix $B(x, y\,; \mu)$ defined by replacing $H\big(x, \pi + \nu(\pi - y)\big)$ by $H(x, y)$ in the leading block of $\nabla^2 M$, i.e.,

$$B(x, y\,; \mu) = \begin{pmatrix} H(x, y) + \dfrac{1}{\mu}(1+\nu)J(x)^T J(x) & \nu J(x)^T \\ \nu J(x) & \nu\mu I \end{pmatrix}. \tag{2.1}$$

Both this approximation and the exact Hessian $\nabla^2 M$ have the same characteristic *doubly-augmented* structure with a $(1/\mu)J(x)^T J(x)$ term in the leading diagonal block. However, unlike $\nabla^2 M$, the matrix $B(x, y\,; \mu)$ is independent of the multiplier estimate y^E (cf. (1.9)).

For numerical stability, it is important that all computations be performed without the explicit calculation of the matrices $\nabla^2 M$ or B. It will be shown that all the relevant properties of B may be determined from the properties of either of the two matrices

$$H(x, y) + \frac{1}{\mu}(1+\nu)J(x)^T J(x) \quad \text{and} \quad \begin{pmatrix} H(x, y) & J(x)^T \\ J(x) & -\mu I \end{pmatrix}.$$

We refer to the first of these matrices as having *augmented Lagrangian form*, and the second as having *regularized KKT form*. For simplicity of exposition, the formulation and analysis of the main algorithm is given in terms of matrices in augmented Lagrangian form. However, it is crucial that any numerical method be based solely on a factorization of a matrix in KKT form. *It must be emphasized that at no point is it necessary to form or factor a matrix in doubly-augmented or augmented Lagrangian form.* All relevant information is obtained by factoring matrices in regularized KKT form.

Theorem 1.3 implies that for μ sufficiently small, a point $v^* = (x^*, y^*)$ satisfying the second-order sufficient conditions of Definition 1.2 is a local minimizer of the bound-constrained quadratic program

$$\begin{array}{ll} \underset{v}{\text{minimize}} & (v - v^*)^T \nabla M(v^*\,; y^*, \mu) + \frac{1}{2}(v - v^*)^T B(v^*\,; \mu)(v - v^*) \\ \text{subject to} & v_i \geq 0, \quad i = 1, 2, \ldots, n, \end{array} \tag{2.2}$$

where $v = (x, y)$, and $B(v\,;\mu)$ denotes the matrix $B(x, y\,;\mu)$. This result provides the motivation for defining the Hessian of the strictly convex QP subproblem in terms of a positive-definite matrix $\widehat{B} \equiv \widehat{B}(x, y\,;\mu)$ such that $\widehat{B}$ is equal to $B(x, y\,;\mu)$ when B is sufficiently positive definite. For the case where $B(x, y\,;\mu)$ is not positive definite, Gill and Robinson [13, Theorem 4.5] give a method for computing a symmetric $\widehat{H}(x, y)$, not necessarily positive definite, such that $\widehat{H}(x_k, y_k) + (1/\mu)J(x_k)^T J(x_k)$ is positive definite. Given $\widehat{H}$, it follows that the matrix

$$\widehat{B}(x, y\,;\mu) = \begin{pmatrix} \widehat{H}(x, y) + \dfrac{1}{\mu}(1 + \nu)J(x)^T J(x) & \nu J(x)^T \\ \nu J(x) & \nu\mu I \end{pmatrix} \tag{2.3}$$

is positive definite for all positive ν (see Gill and Robinson [13, Lemma 2.2]).

At the start of an outer iteration, the active and free sets of variables at a solution of problem (NP) are estimated by ϵ-active sets $\mathcal{A}_\epsilon$ and $\mathcal{F}_\epsilon$, where the size of ϵ reflects the distance of (x, y) to a first-order optimal point of problem (NP). These sets define the composition of matrices used to compute a direction of negative curvature for $B(x, y\,;\mu)$, and also the initial active-set for the quadratic programming method for the convex QP subproblem. The ϵ-active set is defined as

$$\mathcal{A}_\epsilon(x, y, \mu) = \big\{ i : x_i \leq \epsilon, \;\; \text{with} \;\; \epsilon \equiv \min\big(\, \epsilon_a,\, \max\big(\mu, r(x, y)^\gamma\big)\,\big)\big\}, \tag{2.4}$$

where γ and ϵ_a are fixed scalars satisfying $0 < \gamma < 1$ and $0 < \epsilon_a < 1$, and the function $r(x, y)$ is defined by (1.1). Similarly, the ϵ-free set is the complement of $\mathcal{A}_\epsilon$, i.e.,

$$\mathcal{F}_\epsilon(x, y, \mu) = \{1, 2, \ldots, n\} \setminus \mathcal{A}_\epsilon(x, y, \mu). \tag{2.5}$$

These definitions imply that $\mathcal{F}(x^*) \subseteq \mathcal{F}_\epsilon(x, y, \mu)$ when μ is sufficiently small and (x, y) is close to a first-order KKT pair (x^*, y^*).

Algorithm summary. For the remainder of the discussion, it is assumed that ν is a fixed positive scalar. The algorithm generates a sequence $\{v_k\}$ where $v_k = (x_k, y_k)$ is the kth estimate of a primal-dual solution of problem (NP). A key feature is that a small penalty parameter μ_k^R is used for the step computations, and a larger penalty parameter μ_k is used for the definition of the merit function. In this context, μ_k^R plays the role of a regularization parameter that guarantees that the linear system of equations that must be solved at each inner iteration is nonsingular (for other examples of this type of regularization, see [22, 2, 13]).

The computation associated with the kth iteration may be arranged into five main steps.

1. Given (x_k, y_k) and the regularization parameter μ_{k-1}^R from the previous iteration, define $\mathcal{F}_\epsilon(x_k, y_k, \mu_{k-1}^R)$ and $B(x_k, y_k\,;\mu_{k-1}^R)$. Compute the positive-definite matrix $\widehat{B}(x_k, y_k\,;\mu_{k-1}^R)$ together with a nonnegative scalar $\xi_k^{(1)}$ and vector $s_k^{(1)}$ such that if $\xi_k^{(1)} > 0$, then $(-\xi_k^{(1)}, s_k^{(1)})$ approximates the most negative eigenpair of $B(x_k, y_k\,;\mu_{k-1}^R)$ (see Section 2.1).

2. Use $\xi_k^{(1)}$ and $r(x_k, y_k)$ to define values of y_k^E and μ_k^R for the kth iteration (see Section 2.2).
3. Define a descent direction $d_k = (p_k, q_k)$ by solving a convex bound-constrained subproblem with Hessian $\widehat{B}(x_k, y_k\,; \mu_k^R)$ and gradient $\nabla M(x_k, y_k\,; \mu_k^R)$. The primal part of d_k satisfies $x_k + p_k \geq 0$ (see Section 2.3).
4. Compute a direction of negative curvature $s_k = (u_k, w_k)$ by rescaling the direction $s_k^{(1)}$. The primal part of s_k satisfies $x_k + p_k + u_k \geq 0$ (see Section 2.3).
5. Perform a flexible line search along the vector

$$\Delta v_k = s_k + d_k = \begin{pmatrix} p_k + u_k \\ q_k + w_k \end{pmatrix}$$

(see Section 2.4). Update the line-search penalty parameter μ_k.

2.1. Estimating negative curvature. The definitions of the regularization parameter μ_k^R and multipliers estimate y_k^E for the kth iteration depend on an estimate of the smallest eigenvalue of the approximate Hessian $B(x_k, y_k\,; \mu_{k-1}^R)$ of (2.1). This estimate may be defined in terms of an estimate of $\lambda_{\min}\big(H_F + (1/\mu_{k-1}^R)J_F^T J_F\big)$, where H_F and J_F are defined in terms of the index set $\mathcal{F}_\epsilon(x_k, y_k, \mu_{k-1}^R)$, i.e., H_F is the matrix of ϵ-free rows and columns of $H(x_k, y_k)$, and J_F is the matrix of ϵ-free columns of J. Algorithm 1 summarizes the principal

Algorithm 1 Computation of curvature information.

procedure CURVATURE($\mathcal{F}_\epsilon, H_k, J_k, \mu_{k-1}^R$)
  Compute H_F and J_F as submatrices of H_k and J_k associated with the set $\mathcal{F}_\epsilon$;
  if $H_F + (1/\mu_{k-1}^R)J_F^T J_F$ is positive semidefinite **then**
    Set $s_k^{(1)} \leftarrow (u_k^{(1)}, w_k^{(1)}) = (0, 0)$; $\quad \xi_k^{(1)} \leftarrow 0$;
  else
    Compute a direction $u_F \neq 0$ that satisfies (2.6);
    Set $\xi_k^{(1)} \leftarrow -u_F^T\big(H_F + (1/\mu_{k-1}^R)J_F^T J_F\big)u_F/\|u_F\|^2 > 0$;
    Set $[\,u_k^{(1)}\,]_F \leftarrow u_F$; $\quad [\,u_k^{(1)}\,]_A \leftarrow 0$;
    Set $w_k^{(1)} \leftarrow -(1/\mu_{k-1}^R)J_k u_k^{(1)}$;
    Set $s_k^{(1)} \leftarrow (u_k^{(1)}, w_k^{(1)})$; [primal-dual direction $s_k^{(1)} = (u_k^{(1)}, w_k^{(1)}) \neq 0$]
  end if
  return $(u_k^{(1)}, w_k^{(1)}, \xi_k^{(1)})$;
end procedure

calculations associated with finding a nonnegative estimate $\xi_k^{(1)}$ of $\max\{0, -\lambda_{\min}\}$, where $\lambda_{\min}$ denotes the least eigenvalue of $H_F + (1/\mu_{k-1}^R)J_F^T J_F$. If the computed $\xi_k^{(1)}$ is positive, an important by-product of the computation is a feasible direction of negative curvature for the approximate Hessian $B(x_k, y_k\,; \mu_{k-1}^R)$ of (2.1). This direction is combined with a descent direction to form the direction of search for the line search (see Section 2.3).

Algorithm 1 assumes that a procedure is available for determining whether or not $H_F + (1/\mu_{k-1}^R)J_F^T J_F$ is positive semidefinite. If $H_F + (1/\mu_{k-1}^R)J_F^T J_F$ is not positive semidefinite, the procedure must provide a direction u_F satisfying

$$u_F^T\Big(H_F + \frac{1}{\mu_{k-1}^R}J_F^T J_F\Big)u_F \leq \theta\lambda_{\min}\Big(H_F + \frac{1}{\mu_{k-1}^R}J_F^T J_F\Big)\|u_F\|^2 < 0, \tag{2.6}$$

where θ is a positive scalar that is independent of x_k and y_k. Several alternative methods may be used to find the vector $u_{\scriptscriptstyle F}$, but the details of the computation are not relevant for the proof of global convergence. A specific procedure appropriate for a matrix in KKT form is given by Forsgren and Gill [11, Section 4.3]. The inequality (2.6) allows the definition of a direction of negative curvature for the approximate primal-dual Hessian $B(x,y\,;\mu)$ (2.1) when $H_{\scriptscriptstyle F} + (1/\mu_{k-1}^{\scriptscriptstyle R})J_{\scriptscriptstyle F}^T J_{\scriptscriptstyle F}$ is not positive semidefinite (see Lemma 2.1 below).

The vector $w_k^{(1)}$ defined in Step 9 of Algorithm 1 is the change in the dual variables as a function of the change $u_k^{(1)}$ in the primal variables. The definition of $w_k^{(1)}$ ensures that the equality constraints associated with the stabilized QP subproblem (2.18) are satisfied. It also facilitates a proof that the resulting direction $s_k^{(1)}$ is a direction of negative curvature for the approximate primal-dual Hessian $B(x,y\,;\mu)$ when $\xi_k^{(1)} > 0$.

The next lemma gives the properties of the quantities computed by Algorithm 1.

LEMMA 2.1. *Suppose that $\xi_k^{(1)}$ and $s_k^{(1)} = (u_k^{(1)}, w_k^{(1)})$ are defined by Algorithm* 1*, and that the matrix $H_{\scriptscriptstyle F} + (1/\mu_{k-1}^{\scriptscriptstyle R})J_{\scriptscriptstyle F}^T J_{\scriptscriptstyle F}$ is* not *positive semidefinite. Then, for all $\nu > 0$ it holds that*

(1) *$u_k^{(1)} \neq 0$, $s_k^{(1)} \neq 0$, $\xi_k^{(1)} > 0$; and*

(2) *the vector $s_k^{(1)}$ is a direction of negative curvature for the primal-dual Hessian $B(v_k\,;\mu_{k-1}^{\scriptscriptstyle R})$ of* (2.1)*. In particular, $s_k^{(1)}$ satisfies*

$$s_k^{(1)T} B(v_k\,;\mu_{k-1}^{\scriptscriptstyle R}) s_k^{(1)} \leq \bar{\theta}_k \|u_k^{(1)}\|^2 < 0,$$

where $\bar{\theta}_k = \theta\lambda_{\min}\big(H_{\scriptscriptstyle F} + (1/\mu_{k-1}^{\scriptscriptstyle R})J_{\scriptscriptstyle F}^T J_{\scriptscriptstyle F}\big)$ with θ defined in (2.6)*.*

Proof. To simplify notation, the suffix k is omitted and the quantities $\mu_{k-1}^{\scriptscriptstyle R}$, $\bar{\theta}_k$, $s_k^{(1)}$, $u_k^{(1)}$, $w_k^{(1)}$, $\xi_k^{(1)}$, $J(x_k)$, $H(x_k,y_k)$, and $B(v_k\,;\mu_{k-1}^{\scriptscriptstyle R})$ are denoted by μ, $\bar{\theta}$, s, u, w, ξ, J, H and B, respectively.

As $H_{\scriptscriptstyle F} + (1/\mu)J_{\scriptscriptstyle F}^T J_{\scriptscriptstyle F}$ is not positive semidefinite by assumption, Algorithm 1 defines a nonzero direction $u_{\scriptscriptstyle F}$ that satisfies (2.6). It follows that

$$u \neq 0, \;\; s \neq 0, \;\; \lambda_{\min}\Big(H_{\scriptscriptstyle F} + \frac{1}{\mu}J_{\scriptscriptstyle F}^T J_{\scriptscriptstyle F}\Big) < 0, \;\; \text{and} \;\; \xi > 0, \tag{2.7}$$

which proves part (1). The definition of B implies that

$$\begin{aligned} s^T B s &= \begin{pmatrix} u \\ w \end{pmatrix}^T \begin{pmatrix} H + \frac{1}{\mu}(1+\nu)J^T J & \nu J^T \\ \nu J & \nu\mu I \end{pmatrix} \begin{pmatrix} u \\ w \end{pmatrix} \\ &= u^T H u + \frac{1}{\mu}(1+\nu)u^T J^T J u + 2\nu u^T J^T w + \nu\mu\|w\|^2 = u^T\Big(H + \frac{1}{\mu}J^T J\Big)u. \end{aligned} \tag{2.8}$$

The definition of u in Step 8 of Algorithm 1, and the requirement that $u_{\scriptscriptstyle F}$ satisfies (2.6) yield the inequality

$$u^T\Big(H + \frac{1}{\mu}J^T J\Big)u \leq \theta\lambda_{\min}\Big(H_{\scriptscriptstyle F} + \frac{1}{\mu}J_{\scriptscriptstyle F}^T J_{\scriptscriptstyle F}\Big)\|u\|^2. \tag{2.9}$$

Given the definition $\bar{\theta} = \theta\lambda_{\min}(H_{\scriptscriptstyle F} + \frac{1}{\mu}J_{\scriptscriptstyle F}^T J_{\scriptscriptstyle F})$, it must hold that $\bar{\theta} < 0$, where the strict inequality follows from (2.7). The result now follows directly from (2.8), (2.9), and the definition $u = u_k^{(1)}$ as a nonzero vector in Algorithm 1. □

This section ends with a result that relates the quantity $\xi_k^{(1)}$, which estimates the magnitude of the minimum eigenvalue of $H_F + (1/\mu_{k-1}^R)J_F^T J_F$, to the minimum eigenvalue of a reduced Hessian matrix, which plays a fundamental role in the second-order conditions for optimality.

LEMMA 2.2. *If Z_F is a matrix whose columns form an orthonormal basis for the null-space of J_F, then*

$$\lambda_{\min}\big(H_F + (1/\mu_{k-1}^R)J_F^T J_F\big) \leq \lambda_{\min}\big(Z_F^T H_F Z_F\big).$$

Proof. Let $w = Z_F v$, where v is a vector satisfying $\|v\| = 1$ and $v^T Z_F^T H_F Z_F v = \lambda_{\min}(Z_F^T H_F Z_F)$. The definitions of w and Z_F imply that $J_F w = J_F Z_F v = 0$, and since Z_F has orthonormal columns, it must hold that $\|w\| = 1$. These identities and the definition of v yield

$$\begin{aligned}\lambda_{\min}\big(H_F + (1/\mu_{k-1}^R)J_F^T J_F\big) &\leq \frac{w^T\big(H_F + (1/\mu_{k-1}^R)J_F^T J_F\big)w}{w^T w}\\ &= w^T H_F w = v^T Z_F^T H_F Z_F v = \lambda_{\min}\big(Z_F^T H_F Z_F\big),\end{aligned}$$

as required. □

2.2. Updating the multiplier estimate and regularization parameter. The multiplier estimate y_k^E is updated to be the current dual vector y_k as long as there is improvement in some measure of the distance to a primal-dual second-order solution (x^*, y^*). Algorithm 3 uses the feasibility and optimality measures $\eta(x_k)$ and $\omega(x_k, y_k)$ such that

$$\eta(x_k) = \|c(x_k)\| \;\text{ and }\; \omega(x_k, y_k) = \max\left(\left\|\min(x_k, g(x_k) - J(x_k)^T y_k)\right\|,\, \xi_k^{(1)}\right), \tag{2.10}$$

where the quantity $\xi_k^{(1)}$ is defined by Algorithm 1. Given $\eta(x_k)$ and $\omega(x_k, y_k)$, weighted combinations of the feasibility and optimality measures are defined as

$$\phi_V(x_k, y_k) = \eta(x_k) + \beta\omega(x_k, y_k) \qquad \text{and} \qquad \phi_O(x_k, y_k) = \beta\eta(x_k) + \omega(x_k, y_k),$$

where β is a fixed scalar such that $0 < \beta \ll 1$. The update $y_k^E = y_k$ is performed if

$$\phi_V(v_k) \leq \tfrac{1}{2}\phi_V^{\max} \qquad \text{or} \qquad \phi_O(v_k) \leq \tfrac{1}{2}\phi_O^{\max}, \tag{2.11}$$

in which case the kth iterate is called a V-iterate or an O-iterate, respectively. The quantities $\phi_V^{\max}$ and $\phi_O^{\max}$ are positive bounds that are reduced during the solution process. At a V- or O-iterate, the regularization parameter is updated as

$$\mu_k^R = \begin{cases}\min\left(\mu_{k-1}^R,\, \max\left(r_k, \xi_k^{(1)}\right)^\gamma\right), & \text{if } \max\left(r_k, \xi_k^{(1)}\right) > 0;\\ \tfrac{1}{2}\mu_{k-1}^R, & \text{otherwise,}\end{cases} \tag{2.12}$$

for some fixed $\gamma \in (0, 1)$, and where $r_k \equiv r(x_k, y_k)$ is the first-order optimality measure (1.1).

If the conditions for a V- or O-iterate do not hold, (x_k, y_k) is checked to determine if it is an approximate second-order solution of the bound-constrained problem

$$\underset{x,y}{\text{minimize}}\;\; M(x, y\,; y_{k-1}^E, \mu_{k-1}^R) \quad \text{subject to} \quad x \geq 0. \tag{2.13}$$

Specifically, (x_k, y_k) is tested using the conditions

$$\| \min(x_k, \nabla_x M(x_k, y_k\,; y^E_{k-1}, \mu^R_{k-1}))\| \leq \tau_{k-1}, \tag{2.14a}$$

$$\|\nabla_y M(x_k, y_k\,; y^E_{k-1}, \mu^R_{k-1})\| \leq \tau_{k-1}\mu^R_{k-1}, \quad \text{and} \tag{2.14b}$$

$$\xi^{(1)}_k \leq \tau_{k-1}, \tag{2.14c}$$

where τ_{k-1} is a positive tolerance. If these conditions are satisfied, then (x_k, y_k) is called an M-iterate and the parameters are updated as in a conventional augmented Lagrangian method; i.e., the multiplier estimate y^E_{k-1} is replaced by the safeguarded value

$$y^E_k = \max\big(-y_{\max}e,\ \min(\, y_k,\ y_{\max}e\,)\big) \tag{2.15}$$

for some large positive scalar constant $y_{\max}$, and the new regularization parameter is given by

$$\mu^R_k = \begin{cases} \min\big(\, \tfrac{1}{2}\mu^R_{k-1},\ \max\big(r_k, \xi^{(1)}_k\big)^\gamma\big), & \text{if } \max(r_k, \xi^{(1)}_k) > 0; \\ \tfrac{1}{2}\mu^R_{k-1}, & \text{otherwise.} \end{cases} \tag{2.16}$$

In addition, the tolerance τ_{k-1} is decreased by a constant factor. Conditions analogous to (2.14) are used in the first-order method of Gill and Robinson [13], in which case numerical experiments indicate that M-iterates occur infrequently relative to the total number of iterations.

Finally, if neither (2.11) nor (2.14) are satisfied, then $y^E_k = y^E_{k-1}$ and $\mu^R_k = \mu^R_{k-1}$. As the multiplier estimates and regularization parameter are *fixed* at their current values in this case, the kth iterate is called an F-iterate.

2.3. Definition of the line-search direction. The flexible line search described in Section 2.4 uses a search direction Δv_k that is a combination of a descent direction and a direction of negative curvature. The inclusion of the negative curvature direction allows the algorithm to move away from points that do not satisfy the second-order optimality conditions.

The descent direction. Once μ^R_k and y^E_k have been updated, the descent direction portion d_k of the line search direction is computed by solving a strictly convex bound-constrained quadratic programming subproblem with the matrix $\widehat{B}(v\,;\mu) \equiv \widehat{B}(x, y\,;\mu)$ of (2.3) defined with $v = (x_k, y_k)$ and $\mu = \mu^R_k$. As $\widehat{B}(v_k\,;\mu^R_{k-1})$ is positive definite by construction, the matrix $\widehat{B}(v_k\,;\mu^R_k)$ must also be positive definite because the new regularization parameter satisfies $0 < \mu^R_k \leq \mu^R_{k-1}$ (see Section 2.2). The formal definition is $d_k = (\widehat{x}_k - x_k, \widehat{y}_k - y_k)$, where $\widehat{v}_k = (\widehat{x}_k, \widehat{y}_k)$ is a solution of the strictly convex QP problem

$$\begin{aligned} &\underset{v}{\text{minimize}} \quad (v - v_k)^T \nabla M(v_k\,; y^E_k, \mu^R_k) + \tfrac{1}{2}(v - v_k)^T \widehat{B}(v_k\,;\mu^R_k)(v - v_k) \\ &\text{subject to} \quad v_i \geq 0, \quad i = 1, 2, \ldots, n. \end{aligned} \tag{2.17}$$

However, as noted above, it is crucial that d_k be computed without it being necessary to form the matrix $\widehat{B}(v_k\,;\mu^R_k)$. The next result not only characterizes the solution of (2.17), but also provides the link between the primal-dual SQP method and stabilized SQP.

THEOREM 2.3 (Gill and Robinson [13, Result 2.1].). *Let ν and μ^R_k be fixed positive scalars, and define the quantities $v_k = (x_k, y_k)$, $g_k = g(x_k)$, $c_k = c(x_k)$, $J_k = J(x_k)$, and*

$\widehat{H}_k = \widehat{H}(x_k, y_k)$. *If the matrix* $\widehat{H}_k + (1/\mu_k^R) J_k^T J_k$ *is positive definite, then* $\widehat{v}_k = (\widehat{x}_k, \widehat{y}_k)$ *is a solution of the bound-constrained problem* (2.17) *if and only if it solves the stabilized SQP problem*

$$\begin{array}{ll} \underset{x,y}{\text{minimize}} & g_k^T(x - x_k) + \frac{1}{2}(x - x_k)^T \widehat{H}_k (x - x_k) + \frac{1}{2}\mu_k^R \|y\|^2 \\ \text{subject to} & c_k + J_k(x - x_k) + \mu_k^R (y - y_k^E) = 0, \quad x \geq 0. \quad \square \end{array} \tag{2.18}$$

This stabilized QP subproblem may be solved using a conventional active-set algorithm in which each iteration requires the solution of a linear system of equations with matrix in regularized KKT form. In addition, under mild conditions, Gill and Robinson [13, Theorem 5.1] show that conventional active-set methods applied to the stabilized QP (2.18) and the bound-constrained QP (2.17) generate the same sequence of iterates.

The direction of negative curvature. The vector $s_k^{(1)} = (u_k^{(1)}, w_k^{(1)})$ computed as a by-product of the computation of $\xi_k^{(1)}$ in Algorithm 1 is a direction of negative curvature for the approximate Hessian $B(x_k, y_k\,; \mu_{k-1}^R)$. After applying a rescaling procedure described below, this vector is used as the direction of negative curvature for the line search. First, an intermediate direction $s_k^{(2)}$ is computed such that

$$s_k^{(2)} \equiv (u_k^{(2)}, w_k^{(2)}) = \begin{cases} -(u_k^{(1)}, w_k^{(1)}), & \text{if } \nabla M(v_k\,; y_k^E, \mu_k^R)^T s_k^{(1)} > 0; \\ (u_k^{(1)}, w_k^{(1)}), & \text{otherwise,} \end{cases} \tag{2.19}$$

which implies that $s_k^{(2)}$ is a non-ascent direction for $M(v\,; y_k^E, \mu_k^R)$ at $v = (x_k, y_k)$. The direction $s_k^{(2)}$ is then scaled again by the factor

$$\sigma_k = \underset{\sigma \geq 0}{\text{argmax}} \left\{ \sigma : \widehat{x}_k + \sigma u_k^{(2)} \geq 0,\ \|\sigma u_k^{(2)}\| \leq \max(\xi_k^{(1)}, \|\widehat{x}_k - x_k\|) \right\}, \tag{2.20}$$

where $\widehat{x}_k$ is the solution of the stabilized SQP subproblem (2.18), and $\xi_k^{(1)}$ is defined by Algorithm 1. The direction of negative curvature used in the line search is then

$$s_k \equiv (u_k, w_k) = \sigma_k (u_k^{(2)}, w_k^{(2)}) = \sigma_k s_k^{(2)}. \tag{2.21}$$

The direction s_k may be zero in certain situations, but the next result defines one important situation in which σ_k is positive and s_k is nonzero.

LEMMA 2.4. *If* $d_k = 0$ *and* $\xi_k^{(1)} > 0$, *then* $\sigma_k > 0$ *and* $s_k \neq 0$.

Proof. The definition of a positive $\xi_k^{(1)}$ in Algorithm 1 implies that the vectors $u_{_F}$, $u_k^{(1)}$ and $u_k^{(2)}$ are nonzero. This gives a nonzero $s_k^{(1)}$ regardless of the value of $w_k^{(1)}$. The definition of $s_k^{(2)}$ in (2.19) can only change the sign of $u_k^{(1)}$, and hence $s_k^{(2)}$ must be nonzero also.

It remains to show that σ_k is positive. The assumption that d_k is zero implies that $\widehat{x}_k$, the solution of the QP subproblem (2.18), is identical to x_k and $\|\widehat{x}_k - x_k\|$ is zero. The positive value of $\xi_k^{(1)}$ provides a strictly positive upper bound on σ_k that will be achieved provided there is no index $i \in \mathcal{F}_\epsilon(x_k, y_k, \mu_{k-1}^R)$ such that $[\,x_k\,]_i = 0$. This is assured by the definition of the ϵ-active set from (2.4), and the fact that $\mu_{k-1}^R > 0$. $\square$

2.4. The flexible line search and penalty-parameter update. The flexible line search is a generalization of the line search used in the first-order primal-dual method of Gill and Robinson [13]. (The idea of a flexible line search was proposed by Curtis and Nocedal [9] in the context of minimizing an ℓ_1 penalty function.)

Given a primal-dual search direction $\Delta v_k = d_k + s_k$ and a line-search penalty parameter μ, consider the univariate function $\Psi_k(\alpha\,;\mu) = M(v_k + \alpha\Delta v_k\,; y_k^E, \mu)$. (The multiplier estimate y_k^E remains fixed during the line search and is omitted as a parameter of Ψ_k.) The line search is based on approximating $\Psi_k(\alpha\,;\mu)$ by the quadratic model

$$\psi_k(\alpha\,;\mu) = \Psi_k(0\,;\mu) + \alpha\Psi_k'(0\,;\mu) + \tfrac{1}{2}\alpha^2 \min\big(0, \Delta v_k^T B(v_k\,;\mu_{k-1}^R)\Delta v_k\big), \tag{2.22}$$

where Ψ_k' denotes the derivative with respect to α. A conventional line search requires that the step α_k produces a ratio of the actual and predicted reduction in Ψ_k that is at least γ_s, where γ_s is a scalar satisfying $0 < \gamma_s < 1$. This requirement may be written in terms of the familiar Armijo condition

$$\Psi_k(0\,;\mu) - \Psi_k(\alpha_k\,;\mu) \geq \gamma_s\big(\,\psi_k(0\,;\mu) - \psi_k(\alpha_k\,;\mu)\,\big).$$

The flexible line search defined in Algorithm 2 requires that α_k satisfies the modified Armijo condition

$$\Psi_k(0\,;\mu_k^F) - \Psi_k(\alpha_k\,;\mu_k^F) \geq \gamma_s\big(\,\psi_k(0\,;\mu_k^R) - \psi_k(\alpha_k\,;\mu_k^R)\,\big) \tag{2.23}$$

for some value of μ^F such that $\mu_k^F \in [\mu_k^R, \mu_k]$. In practice, the step may be found by reducing α_k by a constant factor until either $\rho_k(\alpha_k\,;\mu_k) \geq \gamma_s$ or $\rho_k(\alpha_k\,;\mu_k^R) \geq \gamma_s$, where

$$\rho_k(\alpha\,;\mu) = \big(\Psi_k(0\,;\mu) - \Psi_k(\alpha\,;\mu)\big)/\big(\,\psi_k(0\,;\mu_k^R) - \psi_k(\alpha\,;\mu_k^R)\,\big).$$

Algorithm 2 Flexible line-search.

Input (d_k, s_k);
Choose constant $\gamma_s \in (0,1)$;
Set $\Delta v_k = d_k + s_k$;
if $s_k = 0$ **and** $d_k = 0$ **then**
    Set $\alpha_k \leftarrow 1$;
else if $d_k \neq 0$ **or** $\nabla M(v_k\,; y_k^E, \mu_k^R)^T s_k < 0$ **or** $\mu_k^R = \mu_{k-1}^R$ **then**
    Set $\alpha_k \leftarrow 1$;
    while $\rho_k(\alpha_k\,;\mu_k^R) < \gamma_s$ **and** $\rho_k(\alpha_k\,;\mu_k) < \gamma_s$ **do**
        Set $\alpha_k \leftarrow \frac{1}{2}\alpha_k$;
    end while
else $[d_k = 0, s_k \neq 0, \xi_k^{(1)} > 0]$
    Set $\xi^R \leftarrow -s_k^T \nabla^2 M(v_k\,; y_k^E, \mu_k^R) s_k / \|u_k\|^2$;
    if $\xi^R > \gamma_s \xi_k^{(1)}$ **then**
        Set $\alpha_k \leftarrow 1$;
        while $\rho_k(\alpha_k\,;\mu_k^R) < \gamma_s$ **and** $\rho_k(\alpha_k\,;\mu_k) < \gamma_s$ **do**
            Set $\alpha_k \leftarrow \frac{1}{2}\alpha_k$;
        end while
    else
        Set $\alpha_k = 0$;
    end if
end if
return $\alpha_k \geq 0$

The Armijo procedure is not executed if $\Delta v_k = 0$ or $d_k = 0$ and the curvature of $M(v_k\,; y_k^E, \mu_k^R)$ along the vector s_k is not sufficiently large compared to the curvature of the quadratic model. The condition for negligible curvature involves the calculation of the scalar $\xi^R = -s_k^T \nabla^2 M(v_k\,; y_k^E, \mu_k^R) s_k / \|u_k\|^2$, which satisfies

$$\xi^R = -\frac{s_k^T \nabla^2 M(v_k\,; y_k^E, \mu_k^R) s_k}{\|u_k\|^2} = -\frac{u_k^T \Big(H(x_k, \widehat{y}_k) + (1/\mu_k^R) J(x_k)^T J(x_k) \Big) u_k}{\|u_k\|^2}.$$

with $\widehat{y}_k = \pi_k + \nu(\pi_k - y_k)$.

On completion of the flexible line search, the iteration ends with the definition of the line-search penalty parameter μ_{k+1} for the next iteration. The strategy described below is motivated by the goal of decreasing the penalty parameter only when the trial step indicates that the merit function defined with μ_k has not been sufficiently reduced. In particular, μ_{k+1} is defined as

$$\mu_{k+1} = \begin{cases} \mu_k, & \text{if } \rho_k(\alpha_k; \mu_k) \geq \gamma_s \text{ or } \Delta v_k = 0 \text{ or } \alpha_k = 0, \\ \max\left(\frac{1}{2}\mu_k, \mu_k^R\right), & \text{otherwise.} \end{cases} \tag{2.24}$$

This completes the description of the kth iteration.

2.5. The algorithm. The regularized second-order primal-dual SQP algorithm is formally stated as Algorithm 3 below.

The next lemma establishes some properties associated with the key steps of the computation.

LEMMA 2.5. *Assume that $\nu > 0$. Let d_k and s_k be vectors defined by Algorithm 3.*

(1) *If $d_k = 0$, then*
   (a) $\min(x_k, g(x_k) - J(x_k)^T y_k) = 0$, *and* $\pi(x_k, y_k^E, \mu_k^R) = y_k$;
   (b) *if the kth iterate is either a V- or O-iterate, then $r(x_k, y_k) = 0$;*
   (c) *if the kth iterate is an M-iterate and $\|y_k\|_\infty \leq y_{\max}$, then $r(x_k, y_k) = 0$.*

(2) *If $d_k = s_k = 0$, then $\xi_k^{(1)} = 0$, the kth iterate is not an F-iterate, and $\mu_k^R < \mu_{k-1}^R$.*

Proof. For part (1a), if d_k is zero, the optimality conditions for the bound-constrained QP subproblem (2.17) give

$$0 = \begin{pmatrix} \min\left(x_k, \nabla_x M(v_k\,; y_k^E, \mu_k^R)\right) \\ \nabla_y M(v_k\,; y_k^E, \mu_k^R) \end{pmatrix} = \begin{pmatrix} \min(x_k, g_k - J_k^T\left(\pi_k + \nu(\pi_k - y_k)\right) \\ \nu\mu_k^R(y_k - \pi_k) \end{pmatrix},$$

where $\pi_k = \pi(x_k, y_k^E, \mu_k^R)$. As ν and μ_k^R are both positive, it follows that

$$y_k = \pi_k \quad \text{and} \quad \min(x_k, g_k - J_k^T y_k) = 0. \tag{2.25}$$

This completes the proof of part (1a).

If d_k is zero and the kth iterate is either a V- or an O-iterate, then it follows from the update for y_k^E given by Algorithm 3, the definition of π_k, and part (1a) that

$$y_k^E = y_k = \pi_k = y_k^E - c(x_k)/\mu_k^R, \tag{2.26}$$

which implies that $c(x_k) = 0$. Combining this with (1.1) and (2.25) gives $r(x_k, y_k) = 0$, which proves part (1b).

Algorithm 3 Regularized second-order primal-dual SQP algorithm **pdSQP2**.

Input (x_1, y_1);
Set parameters $\{\alpha_{\min}, \tau_{\text{stop}}, \gamma_s\} \subset (0,1)$ and $\{k_{\max}, \nu\} > 0$;
Initialize y_0^E, $\tau_0 > 0$, $\mu_0^R > 0$, and $\mu_1 \in [\mu_0^R, \infty)$;
for $k = 1 : k_{\max}$ **do**
  Compute the ϵ-free set $\mathcal{F}_\epsilon = \mathcal{F}_\epsilon(x_k, y_k, \mu_{k-1}^R)$ given by (2.5);
  Set $J_k \leftarrow J(x_k)$ and $H_k \leftarrow H(x_k, y_k)$;
  Compute $\left(u_k^{(1)}, w_k^{(1)}, \xi_k^{(1)}\right) = \text{CURVATURE}(\mathcal{F}_\epsilon, H_k, J_k, \mu_{k-1}^R)$ using Algorithm 1;
  Set $s_k^{(1)} \leftarrow \left(u_k^{(1)}, w_k^{(1)}\right)$;
  Compute $\widehat{H}_k$ from H_k such that $\widehat{H}_k + (1/\mu_{k-1}^R) J_k^T J_k$ is positive definite;
  Compute $r_k = r(x_k, y_k)$ from (1.1);
  if $r_k \le \tau_{\text{stop}}$ **and** $\xi_k^{(1)} \le \tau_{\text{stop}}$ **and** $\mu_{k-1}^R \le \tau_{\text{stop}}$ **then**
    exit with the approximate second-order KKT point (x_k, y_k).
  end if
  if $\phi_V(x_k, y_k) \le \frac{1}{2}\phi_V^{\max}$ **then** [V-iterate]
    Set $\phi_V^{\max} \leftarrow \frac{1}{2}\phi_V^{\max}$, $y_k^E \leftarrow y_k$, and $\tau_k \leftarrow \tau_{k-1}$;
    Set μ_k^R using (2.12);
  else if $\phi_O(x_k, y_k) \le \frac{1}{2}\phi_O^{\max}$ **then** [O-iterate]
    Set $\phi_O^{\max} \leftarrow \frac{1}{2}\phi_O^{\max}$, $y_k^E \leftarrow y_k$, and $\tau_k \leftarrow \tau_{k-1}$;
    Set μ_k^R using (2.12);
  else if (x_k, y_k) satisfies (2.14a)–(2.14c) **then** [M-iterate]
    if $\min\left(\|c_k\|, \tau_{\text{stop}}\right) > \mu_k^R$ **and** $\min\left(x_k, J_k^T c_k\right) \le \tau_{\text{stop}}$ **then**
      exit with the approximate infeasible stationary point x_k.
    end if
    Set $\tau_k \leftarrow \frac{1}{2}\tau_{k-1}$;
    Set y_k^E as in (2.15) and μ_k^R as in (2.16);
  else [F-iterate]
    Set $y_k^E \leftarrow y_{k-1}^E$, $\mu_k^R \leftarrow \mu_{k-1}^R$, and $\tau_k \leftarrow \tau_{k-1}$;
  end if
  Solve the QP (2.17) for $d_k = (p_k, q_k) = (\widehat{x}_k - x_k, \widehat{y}_k - y_k)$;
  Compute $s_k^{(2)} = \left(u_k^{(2)}, w_k^{(2)}\right)$ from (2.19); [ensure descent]
  Compute $\sigma_k \ge 0$ from (2.20); [scale factor]
  Compute $s_k = (u_k, w_k)$ from (2.21); [scaled curvature direction]
  Set $\Delta v_k = d_k + s_k$;
  Compute $\alpha_k \ge 0$ from Algorithm 2 so that $(\Delta v_k, \alpha_k)$ satisfies (2.23);
  Set μ_{k+1} using (2.24);
  Set $v_{k+1} \leftarrow (x_{k+1}, y_{k+1}) = v_k + \alpha_k d_k + \alpha_k s_k$;
end for

If the kth iterate is an M-iterate and $\|y_k\|_\infty \le y_{\max}$, it follows from the update for y_k^E in (2.15), and the definition of π_k that (2.26) holds. As in part (1b), this proves that $c(x_k) = 0$. Combining this result with (1.1) and (2.25) yields $r(x_k, y_k) = 0$, which establishes part (1c).

The assumption for part (2) to hold is that both d_k and s_k are zero. Suppose that the curvature result does not hold, i.e., assume that $\xi_k^{(1)} > 0$. If d_k is zero, then x_k is optimal for the QP subproblem (2.17) and $\widehat{x}_k = x_k$. If this is the case, then the assumption that $\xi_k^{(1)} > 0$ in the definitions of $u_k^{(2)}$ and its associated scale factor σ_k in (2.20) imply that

$\sigma_k > 0$. However, if $\sigma_k > 0$ and $s_k = 0$ in the definition of $s_k^{(2)}$ (2.21), then not only is $s_k^{(2)} = 0$, but also $s_k^{(1)} = 0$ from (2.19). This gives the required contradiction because $s_k^{(1)}$ must be nonzero for $\xi_k^{(1)}$ to be positive in Algorithm 1. It follows that $\xi_k^{(1)} = 0$, as required.

The proof that the kth iterate cannot be an F-iterate when both d_k and s_k are zero is also by contradiction. For an F-iterate, the parameters μ_k^R and y_k^E used in the definition of the QP subproblem are

$$\mu_k^R = \mu_{k-1}^R \quad \text{and} \quad y_k^E = y_{k-1}^E. \tag{2.27}$$

If the solution of the subproblem is $d_k = 0$, then part (1a) and the result that $\xi_k^{(1)} = 0$ imply that (x_k, y_k) satisfies the conditions (2.14) for an M-iterate. This is a contradiction because a point is classified as an F-iterate if it is not a V-iterate or O-iterate and the conditions for an M-iterate fail to hold.

It remains to show that the regularization parameter must be decreased. Assume to the contrary that $\mu_k^R = \mu_{k-1}^R$. It has already been shown that x_k cannot be an F-iterate. Moreover, if x_k were an M-iterate then the update (cf. (2.16)) would imply that $\mu_k^R < \mu_{k-1}^R$. It follows that x_k must be either a V-iterate or an O-iterate. The properties of the update (2.12) associated with a V- or O-iterate, and the assumption that $\mu_k^R = \mu_{k-1}^R$ imply that $\max\big(r_k, \xi_k^{(1)}\big) > 0$. This is a contradiction because $r_k = 0$ follows from part (1b), and it has been shown above that $\xi_k^{(1)} = 0$. As all possible classifications have been exhausted, it must hold that $\mu_k^R < \mu_{k-1}^R$. □

The next lemma summarizes the principal properties of the flexible line-search computation and the penalty parameter update scheme. In particular, the result of part (1) implies that the line search is guaranteed to terminate in a finite number of steps, and as a consequence, the algorithm is well-defined.

LEMMA 2.6. *If f and c are twice continuously differentiable, then the following properties hold.*

(1) *The while-loops given by Steps* 8 *and* 15 *of Algorithm* 2 *terminate with $\alpha_k > 0$.*
(2) *If $\mu_k < \mu_{k-1}$ for some $k \geq 1$, then either the while loop given by Step* 8 *or the while loop given by Step* 15 *of Algorithm* 2 *was executed.*
(3) *If $\alpha_k = 0$, then the kth iterate is not an F-iterate.*

Proof. To prove part (1), it is sufficient to prove that $\rho_k(\alpha; \mu_k^R) \geq \gamma_s$ for all $\alpha > 0$ sufficiently small, which is equivalent to proving that $\eta(\alpha) > 0$, where

$$\eta(\alpha) = \Psi_k(0\,; \mu_k^R) - \Psi_k(\alpha\,; \mu_k^R) - \gamma_s\big(\,\psi_k(0\,; \mu_k^R) - \psi_k(\alpha\,; \mu_k^R)\,\big).$$

Substituting the definition of the quadratic model from (2.22) into the definition of $\eta(\alpha)$ and performing some trivial rearrangement yields

$$\eta(\alpha) = \Psi_k(0\,; \mu_k^R) - \Psi_k(\alpha\,; \mu_k^R) + \gamma_s\alpha\Psi_k'(0\,; \mu_k^R) + \tfrac{1}{2}\gamma_s\alpha^2 \min\big(0, \Delta v_k^T B(v_k\,; \mu_{k-1}^R)\Delta v_k\big).$$

Substituting the Taylor-series expansion

$$\Psi_k(\alpha\,; \mu_k^R) = \Psi_k(0\,; \mu_k^R) + \alpha\Psi_k'(0\,; \mu_k^R) + \tfrac{1}{2}\alpha^2\Psi_k''(0\,; \mu_k^R) + O(|\alpha|^3)$$

gives

$$\eta(\alpha) = \alpha(\gamma_s - 1)\Psi_k'(0\,; \mu_k^R) + \tfrac{1}{2}\alpha^2\omega_k - O(|\alpha|^3) \tag{2.28}$$

where ω_k is the scalar

$$\omega_k = \gamma_s \min\big(0, \Delta v_k^T B(v_k\,;\mu_{k-1}^R)\Delta v_k\big) - \Psi_k''(0\,;\mu_k^R) \tag{2.29}$$

$$= \gamma_s\big(\min(0, \Delta v_k^T B(v_k\,;\mu_{k-1}^R)\Delta v_k) - \Psi_k''(0\,;\mu_k^R)\big) + (\gamma_s - 1)\Psi_k''(0\,;\mu_k^R). \tag{2.30}$$

Case 1: Consider the computation associated with satisfying the Armijo condition in Step 8 in Algorithm 2. If this loop is executed, it must be the case that the combined direction Δv_k is nonzero, and at least one of the conditions $d_k \neq 0$, $\nabla M(v_k\,;y_k^E,\mu_k^R)^T s_k \neq 0$, and $\mu_k^R = \mu_{k-1}^R$ must hold. Based on the values of these quantities, two subcases are considered.

Subcase 1: Suppose that $d_k \neq 0$ or $\nabla M(v_k\,;y_k^E,\mu_k^R)^T s_k \neq 0$. It follows that

$$\Psi_k'(0\,;\mu_k^R) = \nabla M(v_k\,;y_k^E,\mu_k^R)^T \Delta v_k = \nabla M(v_k\,;y_k^E,\mu_k^R)^T (d_k + s_k) < 0, \tag{2.31}$$

because $\nabla M(v_k\,;y_k^E,\mu_k^R)^T s_k \leq 0$ by construction (see Steps 30 and 32 of Algorithm 3) and d_k is the solution of the strictly convex problem (2.17). As $\gamma_s \in (0,1)$, it follows from (2.28) and (2.31) that $\eta(\alpha) > 0$ for all positive α sufficiently small.

Subcase 2: Suppose that $d_k = 0$, $\nabla M(v_k\,;y_k^E,\mu_k^R)^T s_k = 0$, and $\mu_k^R = \mu_{k-1}^R$. It follows immediately that $\Psi_k'(0\,;\mu_k^R) = 0$, and we may invoke part (1a) of Lemma 2.5 to give $y_k = \pi_k = \pi(x_k, y_k^E, \mu_k^R)$. This result, when combined with the assumption that $\mu_k^R = \mu_{k-1}^R$ and $d_k = 0$, gives

$$\Psi_k''(0\,;\mu_k^R) = s_k^T \nabla^2 M(v_k\,;y_k^E,\mu_k^R) s_k = s_k^T B(v_k\,;\mu_{k-1}^R) s_k. \tag{2.32}$$

As d_k is zero, the vector s_k must be nonzero. It then follows from part (2) of Lemma 2.1, (2.19), and (2.21) that $s_k^T B(v_k\,;\mu_{k-1}^R) s_k < 0$. Combining this result with the two identities (2.32) and $\Delta v_k = s_k$, the definition of ω_k in (2.30), and the assumption that $\gamma_s \in (0,1)$, gives the result that $\omega_k = (\gamma_s - 1)\Psi_k''(0\,;\mu_k^R) > 0$. This result, together with the fact that $\Psi_k'(0\,;\mu_k^R) = 0$, and the expression for $\eta(\alpha)$ given in (2.28) implies that $\eta(\alpha) > 0$ for all positive α sufficiently small.

Case 2: Consider the computation associated with satisfying the Armijo condition in Step 15. In this case it must hold that

$$d_k = 0, \quad s_k \neq 0, \quad \nabla M(v_k\,;y_k^E,\mu_k^R)^T s_k = 0, \quad \text{and} \quad \xi^R > \gamma_s \xi_k^{(1)} > 0. \tag{2.33}$$

As a consequence, the identity $\Psi_k'(0\,;\mu_k^R) = 0$ holds, and the expression (2.28) for $\eta(\alpha)$ may be written in the form

$$\eta(\alpha) = \tfrac{1}{2}\alpha^2 \omega_k - O(|\alpha|^3). \tag{2.34}$$

Combining the conditions (2.33), the values $(u_k^{(1)}, w_k^{(1)}, \xi_k^{(1)})$ returned by Algorithm 1, the curvature expression (2.8), and the expressions (2.19) and (2.21) defining the direction of

negative curvature, gives

$$
\begin{aligned}
-\frac{s_k^T \nabla^2 M(v_k\,; y_k^E, \mu_k^R) s_k}{\|u_k\|^2} = \xi^R &> \gamma_s \xi_k^{(1)} \\
&= -\gamma_s \, \frac{u_k^{(1)T} \left[H(x_k, y_k) + (1/\mu_{k-1}^R) J(x_k)^T J(x_k) \right] u_k^{(1)}}{\|u_k^{(1)}\|^2} \\
&= -\gamma_s \, \frac{s_k^{(1)T} B(v_k\,; \mu_{k-1}^R) s_k^{(1)}}{\|u_k^{(1)}\|^2} \\
&= -\gamma_s \, \frac{s_k^T B(v_k\,; \mu_{k-1}^R) s_k}{\|u_k\|^2}.
\end{aligned}
$$

The last of these inequalities is equivalent to

$$
\gamma_s s_k^T B(v_k\,; \mu_{k-1}^R) s_k - s_k^T \nabla^2 M(v_k\,; y_k^E, \mu_k^R) s_k = \gamma_s s_k^T B(v_k\,; \mu_{k-1}^R) s_k - \Psi_k''(0\,; \mu_k^R) > 0.
$$

Combining this inequality with $s_k^T B(v_k\,; \mu_{k-1}^R) s_k < 0$, which follows from (2.8) and the definition of $u_k^{(1)}$ in Algorithm 1, shows that the scalar ω_k is positive in (2.29), which, in turn, implies that the function $\eta(\alpha)$ of (2.34) satisfies $\eta(\alpha) > 0$ for all $\alpha > 0$ sufficiently small. This completes the proof of part (1).

Part (2) follows directly from the definition of the penalty parameter update (2.24) and the structure of Algorithm 2.

For part (3), if $\alpha_k = 0$, then it follows from the **elseif** statement in Step 6 of Algorithm 2 that $\mu_k^R \neq \mu_{k-1}^R$. It follows that the kth iterate cannot be an F-iterate. □

The principal focus of the next section is the global convergence of Algorithm 3 under the assumption that the finite termination conditions given by Steps 11 and 21 are omitted, and that $k_{\max} = \infty$. This allows the discussion of the properties associated with the *infinite* set of iterations generated by the algorithm. The only finite termination result is Theorem 3.12, which establishes that Algorithm 3 will terminate after a finite number of iterations if the algorithm includes a check for an approximate second-order KKT point satisfying the termination conditions

$$
r(x_k, y_k) \leq \tau_{\text{stop}}, \quad \xi_k^{(1)} \leq \tau_{\text{stop}}, \quad \text{and} \quad \mu_{k-1}^R \leq \tau_{\text{stop}}, \tag{2.35}
$$

together with a check for an approximate infeasible stationary point (ISP) satisfying the conditions

$$
\min\left(\eta(x_k), \tau_{\text{stop}}\right) > \mu_k^R, \;\; \|\min\left(x_k, J(x_k)^T c(x_k)\right)\| \leq \tau_{\text{stop}}, \;\; \text{with } k \text{ an M-iterate.} \tag{2.36}
$$

In these conditions, τ_{stop} is a given positive stopping tolerance, r is defined by (1.1), $\xi_k^{(1)}$ is computed in Algorithm 1, and η is the infeasibility measure (2.10). If (2.36) holds, then x_k is regarded as an approximate ISP since it approximately solves the problem

$$
\underset{x \in \mathbb{R}^n}{\text{minimize}} \;\; \tfrac{1}{2}\|c(x)\|_2^2 \;\; \text{subject to } x \geq 0, \tag{2.37}
$$

but the constraint violation is bounded away from zero by $\mu_k^R > 0$. A compact representation of the first-order optimality conditions for problem (2.37) is:

$$
\left\|\min\left(x, J(x)^T c(x)\right)\right\| = 0. \tag{2.38}
$$

In the termination conditions (2.36), the requirements that (x_k, y_k) is an M-iterate and that μ_k^R is small relative to $\min\big(\eta(x_k), \tau_{\text{stop}}\big)$ have the practical benefit of reducing the likelihood of premature termination when progress towards an approximate second-order KKT point is still possible. The termination conditions (2.35) and (2.36) constitute Steps 11 and 21 of Algorithm 3.

3. Global convergence. As in the first-order convergence analysis of Gill and Robinson [13], three standard assumptions are made about the iterates and the properties of the problem functions.

ASSUMPTION 3.1. *The sequence of matrices $\{\widehat{H}(x_k, y_k)\}_{k\geq 0}$ is chosen to satisfy*

$$\|\widehat{H}(x_k, y_k)\| \leq \widehat{H}_{\max} \quad \textit{and} \quad \lambda_{\min}\Big(\widehat{H}(x_k, y_k) + (1/\mu_k^R)J(x_k)^T J(x_k)\Big) \geq \underline{\lambda}_{\min},$$

for some positive $\widehat{H}_{\max}$ and $\underline{\lambda}_{\min}$, and all $k \geq 0$.

ASSUMPTION 3.2. *The functions f and c are twice continuously differentiable.*

ASSUMPTION 3.3. *The sequence $\{x_k\}_{k\geq 0}$ is contained in a compact set.*

Unless there is an explicit statement to the contrary, every result given in this section requires that the Assumptions 3.1–3.3 hold.

The following theorem extends a result of Gill and Robinson [13, Theorem 3.1] to allow for the computation of directions of negative curvature.

THEOREM 3.1. *Let $\{v_k\} = \{(x_k, y_k)\}$ denote the sequence of primal-dual iterates generated by Algorithm 3. Suppose that there exists some $\widehat{k}$ such that the kth iterate is an F-iterate for every k in the set $\mathcal{S} = \{k : k \geq \widehat{k}\}$. The following results hold.*

(1) *There exist positive constant scalars τ, μ^R, and μ, and a constant vector y^E such that, for all $k \in \mathcal{S}$, it holds that*

$$\tau_k = \tau, \quad \mu_k^R = \mu^R, \quad \mu_k = \mu, \quad \textit{and} \quad y_k^E = y^E. \tag{3.1}$$

(2) *The sequences $\{d_k\}$, $\{y_k\}$, and $\{\pi(x_k, y^E, \mu^R)\}$ are uniformly bounded for all $k \in \mathcal{S}$.*

(3) *The sequences $\{\xi_k^{(1)}\}$, $\{s_k\}$, $\{\Delta v_k\}$ $(= \{d_k + s_k\})$, $\{B(v_k\,;\mu^R)\}$, $\{\nabla^2 M(v_k\,; y^E, \mu^R)\}$, $\{\Delta v_k^T B(v_k\,;\mu^R)\Delta v_k\}$, and $\{\Delta v_k^T \nabla^2 M(v_k\,; y_k^E, \mu^R)\Delta v_k\}$ are uniformly bounded for all $k \in \mathcal{S}$.*

(4) *The sequence $\{\|d_k\| + \|s_k\|\}_{k\in\mathcal{S}}$ is bounded away from zero.*

(5) *If $\lim_{k\in\mathcal{S}'} \nabla M(v_k\,; y^E, \mu^R)^T d_k = 0$ for any subsequence $\mathcal{S}' \subseteq \mathcal{S}$, then*

$$\lim_{k\in\mathcal{S}'} d_k = \lim_{k\in\mathcal{S}'} \|\pi(x_k, y^E, \mu^R) - y_k\| = \lim_{k\in\mathcal{S}'} \|B(v_k\,;\mu^R) - \nabla^2 M(v_k\,; y^E, \mu^R)\| = 0,$$

and $\xi_k^{(1)} > \tau$ for all $k \in \mathcal{S}'$ sufficiently large.

(6) *There exists a positive ϵ such that*

$$\nabla M(v_k\,; y^E, \mu^R)^T d_k \leq -\epsilon \quad \textit{or} \quad s_k^T B(v_k\,;\mu^R) s_k \leq -\epsilon \quad \textit{for all } k \in \mathcal{S}.$$

Proof. Part (1) follows from the definition of the set $\mathcal{S}$, the structure of Algorithm 3, and (2.24).

Part (2) follows from the statement and proof of Theorem 3.1 of Gill and Robinson [13].

The first step in the proof of part (3) is to show that the set $\{\xi_k^{(1)}\}_{k\in\mathcal{S}}$ is uniformly

bounded. From the definition of $\xi_k^{(1)}$ in Algorithm 1, it must hold that

$$\begin{aligned} 0 \leq \xi_k^{(1)} &= -\frac{u_k^{(1)T}\big(H(x_k,y_k) + (1/\mu^R)J(x_k)^TJ(x_k)\big)u_k^{(1)}}{\|u_k^{(1)}\|^2} \\ &\leq -\frac{u_k^{(1)T}H(x_k,y_k)u_k^{(1)}}{\|u_k^{(1)}\|^2} \leq -\lambda_{\min}(H(x_k,y_k)) \;\text{ for } k \in \mathcal{S}. \end{aligned} \tag{3.2}$$

From the result of part (2), the set of multipliers $\{y_k\}_{k\in\mathcal{S}}$ is uniformly bounded, which, together with the inequality (3.2) and Assumptions 3.2 and 3.3, imply the required uniform boundedness of $\{\xi_k^{(1)}\}_{k\in\mathcal{S}}$.

The proof that the sequence $\{s_k\}_{k\in\mathcal{S}}$ is uniformly bounded involves showing that the vectors u_k and w_k that define $s_k = (u_k, w_k)$ are bounded. The result for $\{u_k\}_{k\in\mathcal{S}}$ follows from the uniform boundedness of the sequences $\{\xi_k^{(1)}\}_{k\in\mathcal{S}}$ and $\{d_k\}_{k\in\mathcal{S}}$ shown in part (2), and the definition of u_k from (2.21) and (2.20). For the sequence $\{w_k\}_{k\in\mathcal{S}}$, the expressions (2.19), (2.20), and (2.21) that define w_k in terms of $w_k^{(1)}$, and the definition of $w_k^{(1)}$ in Algorithm 1 give

$$\begin{aligned} \|w_k\| &= \|\sigma_k w_k^{(2)}\| = \|\sigma_k w_k^{(1)}\| = \|(\sigma_k/\mu^R)J(x_k)u_k^{(1)}\| \\ &= \frac{1}{\mu^R}\|\sigma_k J(x_k)u_k^{(2)}\| = \frac{1}{\mu^R}\|J(x_k)u_k\| \leq \frac{1}{\mu^R}\|J(x_k)\|\|u_k\|. \end{aligned}$$

The combination of this inequality with Assumptions 3.2–3.3, and the uniform boundedness of $\{u_k\}_{k\in\mathcal{S}}$ implies that $\{w_k\}_{k\in\mathcal{S}}$ is also uniformly bounded. This completes the proof that $\{s_k\}_{k\in\mathcal{S}}$ is uniformly bounded.

The uniform boundedness of $\{\nabla^2 M(v_k\,; y^E, \mu^R)\}_{k\in\mathcal{S}}$ and $\{B(v_k\,;\mu^R\}_{k\in\mathcal{S}}$ follow directly from their respective definitions (1.9) and (2.1), together with Assumptions 3.2–3.3, and part (2) of this theorem. This result and the boundedness of $\{d_k\}_{\in\mathcal{S}}$, $\{s_k\}_{\in\mathcal{S}}$, implies the uniform boundedness of $\Delta v_k^T B(v_k\,;\mu^R)\Delta v_k$ and $\Delta v_k^T \nabla^2 M(v_k\,; y_k^E, \mu^R)\Delta v_k$ on the subsequence $\mathcal{S}$. This completes the proof of part (3).

Part (4) is established by contradiction. Assume that there exists a subsequence $\mathcal{S}_1 \subseteq \mathcal{S}$ such that

$$\lim_{k\in\mathcal{S}_1} d_k = 0 \quad \text{and} \quad \lim_{k\in\mathcal{S}_1} s_k = 0. \tag{3.3}$$

The solution $\widehat{v}_k$ of the bound-constrained QP subproblem (2.17) can be written in terms of the vector $d_k = \widehat{v}_k - v_k = (\widehat{x}_k - x_k, \widehat{y}_k - y_k)$. As $\mu_k^R = \mu^R$ for $k \in \mathcal{S}_1 \subseteq \mathcal{S}$ from part (1), the QP optimality conditions can be written in the form

$$\begin{pmatrix} z_k \\ 0 \end{pmatrix} = \widehat{B}(v_k\,;\mu^R)d_k + \nabla M(v_k\,; y^E, \mu^R), \quad \text{with} \quad \min(\widehat{x}_k, z_k) = 0. \tag{3.4}$$

From part (3), the matrix $\widehat{B}(v_k\,;\mu^R)$ is uniformly bounded on $\mathcal{S}$, which implies that with the assumption (3.3) and the limit $\lim_{k\in\mathcal{S}_1} d_k = 0$, the optimality conditions (2.14a) and (2.14b) must be satisfied for $k \in \mathcal{S}_1$ sufficiently large. If there existed a subsequence of $\mathcal{S}_1$ for which $\xi_k^{(1)} \leq \tau$, then eventually condition (2.14c) would also be satisfied, which would violate the assumption that all iterates are F-iterates for $k \geq \widehat{k}$. Thus, it may be inferred that $\xi_k^{(1)} > \tau$ for all $k \in \mathcal{S}_1$ sufficiently large.

If $\xi_k^{(1)} > \tau > 0$ for all $k \in \mathcal{S}_1$ sufficiently large, then Algorithm 1 will define a nonzero $u_k^{(1)}$ for all $k \in \mathcal{S}_1$ sufficiently large. As $u_k^{(2)}$ is $\pm u_k^{(1)}$ from (2.19), $u_k^{(2)}$ is also nonzero for the same values of k. Let ϵ_k denote the value of ϵ that defines the ϵ-active set (2.4) at (x_k, y_k). From this definition of ϵ_k, it must hold that $\epsilon_k \geq \min(\mu_{k-1}^R, \epsilon_a) = \min(\mu^R, \epsilon_a) > 0$. Moreover, the assumption (3.3) that $\lim_{k\in\mathcal{S}_1} d_k = 0$ implies that for every $j \in \mathcal{F}_\epsilon(x_k, y_k, \mu^R)$, the lower bound $[\widehat{x}_k]_j \geq \frac{1}{2}\epsilon_k \geq \frac{1}{2}\min(\mu^R, \epsilon_a)$ must hold for all $k \in \mathcal{S}_1$ sufficiently large. If this lower bound is combined with the inequality $\xi_k^{(1)} > \tau > 0$ and the property that $[u_k^{(2)}]_j = 0$ for $j \in \mathcal{A}_\epsilon(x_k, y_k, \mu^R)$, it follows from the definition of σ_k in (2.20) that there must exist some positive δ_1 such that $\|u_k\| = \|\sigma_k u_k^{(2)}\| \geq \delta_1$ for all $k \in \mathcal{S}_1$ sufficiently large. This contradicts the assumption (3.3) that $\lim_{k\in\mathcal{S}_1} s_k = 0$ because u_k forms the first n components of s_k. This contradiction implies that (3.3) cannot hold, which proves part (4).

For the proof of part (5), assume that there exists a subsequence $\mathcal{S}' \subseteq \mathcal{S}$ such that

$$\lim_{k\in\mathcal{S}'} \nabla M(v_k\,; y^E, \mu^R)^T d_k = 0, \tag{3.5}$$

and define the nonsingular matrix

$$L_k = \begin{pmatrix} I & 0 \\ -\frac{1}{\mu^R} J_k & I \end{pmatrix}.$$

As $d = 0$ is feasible for (2.17) and $d_k = (p_k, q_k) = (\widehat{x}_k - x_k, \widehat{y}_k - y_k)$ is the unique solution of the strictly convex QP (2.17), it follows that

$$\begin{aligned}
-\nabla M(v_k\,; y^E, \mu^R)^T d_k &\geq \tfrac{1}{2} d_k^T \widehat{B}(v_k\,; \mu^R) d_k \\
&= \tfrac{1}{2} d_k^T L_k^{-T} L_k^T \widehat{B}(v_k\,; \mu^R) L_k L_k^{-1} d_k \\
&= \frac{1}{2} \begin{pmatrix} p_k \\ q_k + \frac{1}{\mu^R} J_k p_k \end{pmatrix}^T \begin{pmatrix} \widehat{H}_k + \frac{1}{\mu^R} J_k^T J_k & 0 \\ 0 & \nu\mu^R \end{pmatrix} \begin{pmatrix} p_k \\ q_k + \frac{1}{\mu^R} J_k p_k \end{pmatrix} \\
&\geq \tfrac{1}{2}\big(\underline{\lambda}_{\min} \|p_k\|^2 + \nu\mu^R \|q_k + (1/\mu^R) J_k p_k\|^2\big),
\end{aligned}$$

where $\underline{\lambda}_{\min} > 0$ is defined in Assumption 3.1. Combining this inequality with (3.5) yields

$$\lim_{k\in\mathcal{S}'} p_k = 0 \quad \text{and} \quad \lim_{k\in\mathcal{S}'} \big(q_k + (1/\mu^R) J_k p_k\big) = 0,$$

in which case $\lim_{k\in\mathcal{S}'} q_k = 0$ since $\{J_k\}$ is uniformly bounded by Asumptions 3.2–3.3. As p_k and q_k are the primal and dual components of d_k, it must be the case that

$$\lim_{k\in\mathcal{S}'} d_k = 0, \tag{3.6}$$

which is the first result of part (5). If the result (3.6) that $\lim_{k\in\mathcal{S}'} d_k = 0$ is combined with the QP optimality conditions (3.4) and the definition of $\nabla_y M(v_k\,; y^E, \mu^R)$ in (1.8), then

$$\lim_{k\in\mathcal{S}'} \|\pi(x_k, y^E, \mu^R) - y_k\| = 0,$$

which is the second result. Combining this result with the definitions of $\nabla^2 M$ and B and Assumptions 3.2–3.3 gives

$$\lim_{k\in\mathcal{S}'} \|\nabla^2 M(v_k\,; y^E, \mu^R) - B(v_k\,; \mu^R)\| = 0,$$

which is the third result.

The proof of the first result of part (5) establishes the limit $\lim_{k\in\mathcal{S}'} d_k = 0$ (see (3.6)). An argument analogous to that used in the proof of part (4) may be used to show that if there is a subsequence of $\mathcal{S}'$ such that $\xi_k^{(1)} \leq \tau$, then some subsequent iterate is an M-iterate, which would be a contradiction. This implies that

$$\xi_k^{(1)} > \tau \quad \text{for all } k \in \mathcal{S}' \text{ sufficiently large,}$$

which completes the proof of part (5).

Part (6) is established by contradiction. If the result does not hold, there must exist a subsequence $\mathcal{S}_1 \subseteq \mathcal{S}$ such that for all $k \in \mathcal{S}_1$, the regularization parameter μ_k^R is fixed at μ^R, with

$$\lim_{k\in\mathcal{S}_1} \nabla M(v_k\,; y^E, \mu^R)^T d_k = 0 \quad \text{and} \quad \lim_{k\in\mathcal{S}_1} s_k^T B(v_k\,; \mu^R) s_k = 0, \tag{3.7}$$

The proofs of parts (4) and (5) establish the existence of a positive δ_1 such that

$$\|s_k\| \geq \delta_1 \quad \text{for } k \in \mathcal{S}_1 \text{ sufficiently large.} \tag{3.8}$$

The combination of this bound, the definitions of s_k (2.21) and $s_k^{(2)}$ (2.19), and part (2) of Lemma 2.1 yields

$$\begin{aligned} s_k^T B(v_k\,; \mu^R) s_k &= \sigma_k^2 s_k^{(2)T} B(v_k\,; \mu^R) s_k^{(2)} = \sigma_k^2 s_k^{(1)T} B(v_k\,; \mu^R) s_k^{(1)} \\ &\leq \sigma_k^2 \bar{\theta}_k \|u_k^{(1)}\|^2 = \sigma_k^2 \bar{\theta}_k \|u_k^{(2)}\|^2 \\ &= \bar{\theta}_k \|u_k\|^2 < 0, \quad \text{for } k \in \mathcal{S}_1 \text{ sufficiently large.} \end{aligned} \tag{3.9}$$

The scalars $\{\bar{\theta}_k\}$ must be bounded away from zero on $\mathcal{S}_1$. Otherwise, the definition of $\bar{\theta}_k$ would imply that $\lambda_{\min}(H_F + (1/\mu^R) J_F^T J_F)$ converges to zero on some subsequence of $\mathcal{S}_1$, which would imply that $\xi_k^{(1)}$ converges to zero on the same subsequence. This contradicts the result of part (5) above, and implies that $\{\bar{\theta}_k\}$ must be bounded away from zero on $\mathcal{S}_1$. Combining this with (3.9) and (3.7) gives

$$\lim_{k\in\mathcal{S}_1} u_k = 0. \tag{3.10}$$

The following simple argument shows that this result leads to the required contradiction. Under the assumption (3.7) that $\lim_{k\in\mathcal{S}_1} \nabla M(v_k\,; y^E, \mu^R)^T d_k = 0$, the result of part (5) implies that $\lim_{k\in\mathcal{S}_1} d_k = 0$. However, if $\lim_{k\in\mathcal{S}_1} d_k = 0$, then the same arguments used to show that $\|s_k\|$ is bounded away from zero in the proof of part (5) may be used to show that that there exists a $\delta_2 > 0$ such that $\|u_k\| \geq \delta_2$ for all $k \in \mathcal{S}_1$ sufficiently large. This contradicts the implication (3.10). It follows that the initial assumption (3.7) cannot hold, which completes the proof of part (6). □

THEOREM 3.2. *Consider the infinite sequence $\{v_k\}$, where $v_k = (x_k, y_k)$ is generated by Algorithm 3.*

(1) *The union of the index sets of V-, O- and M-iterates is infinite.*

(2) *The sequence of regularization parameters satisfies $\{\mu_k^R\} \to 0$.*

Proof. The proof of part (1) is by contradiction. If the number of V-, O-, and M-iterates is finite, then there must exist an index $\widehat{k}$ such that the kth iterate is an F-iterate for all $k \geq \widehat{k}$. In this case, the result of part (1) of Theorem 3.1 may be invoked to conclude that

(3.1) holds for the set $\mathcal{S} = \{k : k \geq \widehat{k}\}$. It will be shown that this result implies the existence of an infinite subsequence $\mathcal{S}_1 \subseteq \mathcal{S}$ and a fixed positive values μ^R and κ such that

$$M(v_{k+1}\,; y^E, \mu^R) \leq M(v_k\,; y^E, \mu^R) - \kappa \quad \text{for } k \in \mathcal{S}_1. \tag{3.11}$$

This may be combined with (3.1) to conclude that M decreases monotonically for every F-iterate. Moreover, it must hold that $\lim_{k\to\infty} M(v_k\,; y^E, \mu^R) = -\infty$. This is impossible given Assumptions 3.2–3.3, and the contradiction implies that the theorem must hold.

If every iterate is an F-iterate for the set $\mathcal{S} = \{k : k \geq \widehat{k}\}$, then part (1) of Theorem 3.1 gives $\mu_k^R = \mu^R$ for every $k \in \mathcal{S}$. The inequality (3.11) is established by considering the Armijo acceptance condition (2.23) associated with the flexible line search described in Section 2.4. Suppose that the value $\mu^F = \mu^R$ is used in the condition (2.23), in which case the line search is equivalent to a conventional Armijo backtracking line search. For all $k \in \mathcal{S}$, this condition may be written in the form

$$\Psi_k(\alpha\,; \mu^R) \leq \Psi_k(0\,; \mu^R) + \gamma_s\big(\psi_k(\alpha\,; \mu^R) - \psi_k(0\,; \mu^R)\big), \tag{3.12}$$

where $\Psi_k(\alpha\,; \mu) = M(v_k + \alpha \Delta v_k\,; y_k^E, \mu)$, $\Delta v_k = d_k + s_k$, and $\psi_k(\alpha\,; \mu^R)$ is the quadratic model (2.22). Expanding the left-hand side of (3.12) using a Taylor-series expansion and performing some rearrangement yields

$$(1-\gamma_s)\Psi_k'(0\,; \mu^R) + \tfrac{1}{2}\alpha\big(\Psi_k''(0\,; \mu^R) - \gamma_s\psi_k''(0\,; \mu^R)\big) \leq -O\big(|\alpha|^2\|\Delta v_k\|^3\big). \tag{3.13}$$

There are two cases to consider.

Case 1: $\Psi_k'(0\,; \mu^R) = \nabla M(v_k\,; y^E, \mu^R)^T \Delta v_k \leq -\delta$ for some $\delta > 0$ and all k large.
It follows from parts (2) and (3) of Theorem 3.1 that $\{\Psi_k''(0\,; \mu^R)\}_{k\in\mathcal{S}}$, $\{\psi_k''(0\,; \mu^R)\}_{k\in\mathcal{S}}$, $\{d_k\}_{k\in\mathcal{S}}$, and $\{s_k\}_{k\in\mathcal{S}}$ are uniformly bounded. These results and the assumption that $\Psi_k'(0\,; \mu^R) \leq -\delta$ imply the existence of a positive $\widehat{\alpha}$ such that (3.13) is satisfied for all $0 < \alpha \leq \widehat{\alpha}$ and all $k \in \mathcal{S}$ sufficiently large. Thus, a conventional backtracking line search would terminate with an $\alpha_k \geq \alpha_{\min}$ for some $\alpha_{\min} > 0$ and all k sufficiently large. However, the use of a value of μ^F ($\mu^F \geq \mu^R$) in the flexible line search allows for the early termination of the backtracking loop, and it must hold that the resulting value must satisfy $\alpha_k \geq \alpha_{\min}$. The combination of part (1) of Theorem 3.1, the acceptance criterion (3.12), the definition of μ_k implied by (2.24), the assumption that $\Psi_k'(0\,; \mu^R) \leq -\delta$, and the bound $\alpha_k \geq \alpha_{\min}$, yields the inequality

$$\begin{aligned}\Psi_k(\alpha_k\,; \mu) &\leq \Psi_k(0\,; \mu) + \gamma_s\alpha_k\Psi_k'(0\,; \mu^R) + \gamma_s\alpha_k^2\psi_k''(0\,; \mu^R) \\ &\leq \Psi_k(0\,; \mu) - \gamma_s\alpha_{\min}\delta, \quad \text{for all } k \text{ sufficiently large.}\end{aligned}$$

If $\mathcal{S}_1$ is the set of k sufficiently large such that $\Psi_k'(0\,; \mu^R) \leq -\delta$, then the final inequality is equivalent to (3.11) with $\kappa = \gamma_s\alpha_{\min}\delta$.

Case 2: A subsequence $\mathcal{S}_1 \subseteq \mathcal{S}$ exists such that $\lim_{k\in\mathcal{S}_1} \Psi_k'(0\,; \mu^R) = 0$.
Algorithm 3 provides directions s_k and d_k that satisfy the inequalities $\nabla M(v_k\,; y^E, \mu^R)^T s_k \leq 0$ and $\nabla M(v_k\,; y^E, \mu^R)^T d_k \leq 0$. Under the given assumption that $\lim_{k\in\mathcal{S}_1} \Psi_k'(0\,; \mu^R) = 0$, it must hold that

$$\lim_{k\in\mathcal{S}_1} \nabla M(v_k\,; y^E, \mu^R)^T d_k = 0. \tag{3.14}$$

This result, when combined with part (6) of Theorem 3.1, indicates that there must exist an $\epsilon > 0$ such that

$$s_k^T B(v_k\,; \mu^R)^T s_k \leq -\epsilon \quad \text{for all } k \in \mathcal{S}_1 \text{ sufficiently large.}$$

This inequality, in combination with (3.14) and parts (3) and (5) of Theorem 3.1 implies that

$$\Delta v_k^T B(v_k\,;\mu^R)\Delta v_k \leq -\tfrac{1}{2}\epsilon \quad \text{for } k \in \mathcal{S}_1 \text{ sufficiently large.} \tag{3.15}$$

As $\Psi_k'(0\,;\mu^R) \leq 0$ by construction, and $\gamma_s \in (0,1)$, a sufficient condition for the inequality (3.13) (or the equivalent Armijo condition (3.12)) to hold is that

$$\tfrac{1}{2}\Big(\Psi_k''(0\,;\mu^R) - \gamma_s\psi_k''(0\,;\mu^R)\Big) \leq -O\Big(|\alpha|\,\|\Delta v_k\|^2\Big). \tag{3.16}$$

However, the definitions of ψ_k, and Ψ_k, and parts (1)–(3), and (5) of Theorem 3.1 imply that

$$\begin{aligned}\Psi_k''(0\,;\mu^R) - \gamma_s\psi_k''(0\,;\mu^R) &\leq \tfrac{1}{2}(1-\gamma_s)\Delta v_k^T B(v_k\,;\mu^R)\Delta v_k \\ &\leq -\tfrac{1}{4}(1-\gamma_s)\epsilon \quad \text{for } k \in \mathcal{S}_1 \text{ sufficiently large,}\end{aligned}$$

where $\gamma_s \in (0,1)$. This inequality allows the sufficient condition (3.16) to be restated as

$$-\tfrac{1}{8}(1-\gamma_s)\epsilon \leq -O(|\alpha|\,\|\Delta v_k\|^2) \quad \text{for } k \in \mathcal{S}_1 \text{ sufficiently large.}$$

Parts (2) and (3) of Theorem 3.1 imply that $\{d_k\}$ and $\{s_k\}$ are uniformly bounded on $\mathcal{S}_1$. It follows that there exists a positive scalar $\widehat{\alpha}$ independent of k such that (3.13) is satisfied for all $0 < \alpha \leq \widehat{\alpha}$ and all $k \in \mathcal{S}_1$ sufficiently large. As in the previous case, this implies that $\alpha_k \geq \alpha_{\min}$ for some $\alpha_{\min} > 0$ and all $k \in \mathcal{S}_1$ sufficiently large, where α_k is the step length defined by the flexible line search. It remains to show that (3.11) holds for some positive κ. A combination of the descent condition (3.12), part (1) of Theorem 3.1, the line-search penalty parameter update (2.24), the descent property $\Psi_k'(0\,;\mu^R) \leq 0$, the bound $\alpha_k \geq \alpha_{\min}$, and the uniform bound $\Delta v_k^T B(v_k\,;\mu^R)\Delta v_k \leq -\frac{1}{2}\epsilon$ of (3.15), gives

$$\begin{aligned}\Psi_k(\alpha_k\,;\mu) &\leq \Psi_k(0\,;\mu) + \gamma_s\alpha_k\Psi_k'(0\,;\mu^R) + \tfrac{1}{2}\gamma_s\alpha_k^2\psi_k''(0\,;\mu^R) \\ &\leq \Psi_k(0\,;\mu) - \tfrac{1}{4}\gamma_s\alpha_{\min}^2\epsilon \quad \text{for all } k \in \mathcal{S}_1 \text{ sufficiently large,}\end{aligned}$$

which is equivalent to (3.11) with $\kappa = \frac{1}{4}\gamma_s\alpha_{\min}^2\epsilon$.

For the proof of part (2) it is necessary to show that the sequence of regularization parameters satisfies $\{\mu_k^R\} \to 0$. The sequence $\{\mu_k^R\}$ decreases monotonically by construction. Moreover, the update rule (2.16) will force $\{\mu_k^R\} \to 0$ if there are infinitely many M-iterates. It remains to consider the case in which the number of M-iterates is finite, in which case, part (1) implies that the union of the sets of V- and O-iterates is infinite. Let $\mathcal{S}_2$ denote the set of V-iterates, and without loss of generality, assume that $\mathcal{S}_2$ is an infinite set. The form of the regularization parameter update (2.12) implies that if infinitely many $k \in \mathcal{S}_2$ satisfy $\max(r_k, \xi_k^{(1)}) = 0$, then $\{\mu_k^R\} \to 0$. Thus, it remains to consider the case that $\max(r_k, \xi_k^{(1)})$ remains positive for all $k \in \mathcal{S}_2$ sufficiently large. However, as the set $\mathcal{S}_2$ is infinite, it follows from the bound $\phi_V(v_k) \leq \frac{1}{2}\phi_V^{\max}$ of (2.11) and the associated update to $\phi_V^{\max}$ in Algorithm 3 that $\lim_{k\in\mathcal{S}} \phi_V(x_k, y_k) = 0$. The definition of ϕ_V then implies that

$$\lim_{k\in\mathcal{S}_2} c(x_k) = 0, \quad \lim_{k\in\mathcal{S}_2} \|\min\big(x_k, g(x_k) - J(x_k)^T y_k\big)\| = 0, \quad \text{and} \quad \lim_{k\in\mathcal{S}_2} \xi_k^{(1)} = 0.$$

These identities imply that the first-order condition (1.1) holds for all $k \in \mathcal{S}_2$ sufficiently large, giving $\lim_{k\in\mathcal{S}} r(x_k, y_k) = 0$. This will cause the regularization parameter update (2.12) to give $\{\mu_k^R\} \to 0$, which completes the proof. □

The analysis of convergence to second-order solutions involves a number of alternative constraint qualifications. The definitions and properties are reviewed below. The conditions involve the sets $\{\nabla c_i(x) : i = 1, 2, \ldots, m\}$ and $\{e_j : j = 1, 2, \ldots, n\}$ consisting of the gradients of the equality constraints $c(x) = 0$ and inequalities $x \geq 0$.

The next two definitions are used in the formulation of the constant positive generator constraint qualification of Andreani et al. [4, Definition 3.1], which is used in the proof of convergence to first-order KKT points.

DEFINITION 3.3. *Let $\mathcal{I}$ and $\mathcal{J}$ denote the index sets $\mathcal{I} \subseteq \{1, 2, \ldots, m\}$ and $\mathcal{J} \subseteq \{1, 2, \ldots, n\}$.*

(1) *A* positive linear combination *of the vectors $\{\nabla c_i(x)\}_{i\in\mathcal{I}}$ and $\{e_i\}_{i\in\mathcal{J}}$ is a vector of the form*

$$\sum_{i\in\mathcal{I}} \alpha_i \nabla c_i(x) + \sum_{j\in\mathcal{J}} \beta_j e_j, \quad \text{with } \beta_j \geq 0 \text{ for all } j \in \mathcal{J}.$$

(2) *The set of all such positive linear combinations is called the* positive linear span *of the set $\big\{\{\nabla c_i(x)\}_{i\in\mathcal{I}},\ \{e_i\}_{i\in\mathcal{J}}\big\}$.*

(3) *The vectors $\{\nabla c_i(x)\}_{i\in\mathcal{I}}$ and $\{e_i\}_{i\in\mathcal{J}}$ are said to be* positively linearly independent *if the only way to write the zero vector using positive linear combinations is to use all trivial coefficients. Otherwise, the vectors are said to be* positively linearly dependent*.*

DEFINITION 3.4 (CPGCQ). *Assume that x is a feasible point for problem* (NP) *and let $\mathcal{I} = \{1, 2, \ldots, m\}$. The* constant positive generator constraint qualification (CPGCQ) *holds at x if there exist sets $\mathcal{I}' \subseteq \mathcal{I}$ and $\mathcal{J} \subseteq \mathcal{A}(x)$ and a neighborhood $\mathcal{B}(x)$ of x such that the following two properties hold:* (i) *the vectors $\big\{\{\nabla c_i(x)\}_{i\in\mathcal{I}'}, \{e_i\}_{i\in\mathcal{J}}\big\}$ are positively linearly independent, with positive linear span equal to the positive linear span of the set $\big\{\{\nabla c_i(x)\}_{i\in\mathcal{I}},\ \{e_i\}_{i\in\mathcal{A}(x)}\big\}$; and* (ii) *for every $\bar{x} \in \mathcal{B}(x)$, any vector in the positive linear span of $\big\{\{\nabla c_i(\bar{x})\}_{i\in\mathcal{I}},\ \{e_i\}_{i\in\mathcal{A}(x)}\big\}$ is in the positive linear span of the vectors $\big\{\{\nabla c_i(\bar{x})\}_{i\in\mathcal{I}'},\ \{e_i\}_{i\in\mathcal{J}}\big\}$.*

A constraint qualification in common use is the Mangasarian-Fromovitz constraint qualification (see [19, 18]).

DEFINITION 3.5 (MFCQ). *Assume that x is a feasible point for problem* (NP)*. The* Mangasarian-Fromovitz constraint qualification (MFCQ) *holds at x if $J(x)$ has full row rank and there exists a vector p such that $J(x)p = 0$ and $p_j > 0$ for every $j \in \mathcal{A}(x)$. The equivalent* dual form of the MFCQ *is that the vectors $\big\{\{\nabla c_i(x)\}_{i=1}^m,\ \{e_i\}_{i\in\mathcal{A}(x)}\big\}$ are positively linearly independent.*

The Mangasarian-Fromovitz constraint qualification is a stronger condition than the constant positive generator constraint qualification of Definition 3.4.

The Mangasarian-Fromovitz constraint qualification is not a second-order constraint qualification. The *weak constant rank condition* defined below was introduced by Andreani et al. [5, pg. 532] in the context of combining the Mangasarian-Fromovitz constraint qualification with the weak constant rank condition to define a second-order constraint-qualification.

DEFINITION 3.6 (WCRC). *Assume that x is a feasible point for problem* (NP)*. Given any $\bar{x}$, let $J_{\mathcal{F}}(\bar{x})$ denote the submatrix of columns of $J(\bar{x})$ associated with $\mathcal{F}(x)$, the set of free variables at x* (see (1.3))*. The* weak constant rank condition (WCRC) *holds at x if there exists a neighborhood $\mathcal{B}(x)$ for which the rank of $J_{\mathcal{F}}(\bar{x})$ is constant for all $\bar{x} \in \mathcal{B}(x)$.*

The equivalence of the following definition of an approximate KKT sequence with the definition of Qi and Wei [20, Definition 2.5] is established by Kungurtsev [17, Result 8.5.1].

DEFINITION 3.7 (Approximate KKT sequence). *The sequence* $\{(x_k, y_k)\}$ *with each* x_k *nonnegative is an* approximate-KKT (AKKT) sequence *if* $\{x_k\} \to x_*$ *and* $\{r(x_k, y_k)\} \to 0$, *where* r *is the norm of the residual* (1.1) *associated with the definition of a first-order KKT point.*

The next result, which concerns the properties of an AKKT sequence, is required for the subsequent convergence analysis.

THEOREM 3.8. *Suppose that* $\{(x_k, y_k)\}$ *is an* AKKT *sequence with* $\{x_k\} \to x_*$. *The following results hold.*

(1) *If the* CPGCQ *holds at* x_*, *then* x_* *is a first-order KKT point for problem* (NP), *i.e., the set of multipliers* $\mathcal{Y}(x_*)$ *of* (1.2) *is nonempty.*
(2) *In addition, if the MFCQ holds at* x_*, *then the sequence* $\{y_k\}$ *is uniformly bounded, contains at least one limit point, and every limit point is contained in* $\mathcal{Y}(x_*)$.

Proof. The proof of part (1) is given by Andreani et al. [4, Theorem 3.3].

The assumptions for part (2) include that both the CPGCQ and the MFCQ hold at x_*. Let $\{\delta_k\}$ and $\{z_k\}$ denote sequences such that

$$\delta_k = \|y_k\|_\infty \quad \text{and} \quad z_k = \max(0, g_k - J_k^T y_k) \geq 0.$$

These definitions and the assumption that $\{(x_k, y_k)\}$ is an AKKT sequence imply that

$$\lim_{k\to\infty} \big(g_k - J_k^T y_k - z_k\big) = 0. \tag{3.17}$$

Suppose that the dual sequence $\{y_k\}$ is unbounded, in which case there must exist some subsequence $\mathcal{S}$ such that $\lim_{k\in\mathcal{S}} \|y_k\| = \infty$. Then, it must hold that the sequences $\{y_k/\delta_k\}_{k\in\mathcal{S}}$ and $\{z_k/\delta_k\}_{k\in\mathcal{S}}$ are bounded from the definitions of δ_k and z_k ($z_k \geq 0$), the assumption that $\{x_k\} \to x_*$, and Assumption 3.2. It follows that there exists a subsequence $\mathcal{S}_1 \subseteq \mathcal{S}$ and vectors $(\bar{y}, \bar{z})$ such that $\lim_{k\in\mathcal{S}_1}(y_k, z_k)/\delta_k = (\bar{y}, \bar{z})$ with $\bar{z} \geq 0$. Combining this inequality with the limit (3.17), $\{\delta_k\}_{k\in\mathcal{S}_1} \to \infty$ and $\{x_k\} \to x_*$, and the Assumption 3.2 gives

$$0 = \lim_{k\in\mathcal{S}_1} \big(g_k - J_k^T y_k - z_k\big)/\delta_k = -J(x_*)^T\bar{y} - \bar{z} \quad \Longrightarrow \quad J(x_*)^T\bar{y} + \bar{z} = 0. \tag{3.18}$$

It now follows from $J(x_*)^T\bar{y} + \bar{z} = 0$, $\bar{z} \geq 0$, and the dual form of the MFCQ that $\bar{y} = \bar{z} = 0$. However, the definition of δ_k and the fact that $\{y_k\}$ is unbounded on $\mathcal{S}$ imply that $\lim_{k\in\mathcal{S}} \|y_k\|/\delta_k = 1$. As $\mathcal{S}_1 \subseteq \mathcal{S}$, the definition of $\bar{y}$ implies that $\|\bar{y}\| = 1$, which contradicts the fact that $\bar{y} = 0$. Thus, it must hold that the sequence $\{y_k\}$ is uniformly bounded with at least one limit point. Finally, the result that every limit point y_* of $\{y_k\}$ is contained in $\mathcal{Y}(x_*)$ follows as a direct consequence of the result that $\{(x_k, y_k)\}$ is an AKKT sequence and the definition (1.1) of a first-order KKT point. □

There are a number of alternative definitions of a second-order KKT point for problem (NP). The proofs below use the following definition (see, e.g., Andreani et al. [3, pg. 211]).

DEFINITION 3.9 (Second-order KKT point). *The primal-dual pair* (x^*, y^*) *is a* second-order KKT point *for problem* (NP) *if* (x^*, y^*) *is a first-order KKT pair* (cf. (1.1)), *and*

$$d^T H(x^*, y^*) d \geq 0 \text{ for all } d \in \mathcal{C}_A(x^*), \tag{3.19}$$

where $\mathcal{C}_A(x)$ *is the set of directions*

$$\mathcal{C}_A(x) = \{d : J(x)d = 0,\ d_i = 0 \ \text{ for } \ i \in \mathcal{A}(x)\}. \tag{3.20}$$

The next result establishes global convergence to second-order KKT points. In particular, it is shown that if there are infinitely many V- or O-iterates, and both the Mangasarian-Fromovitz constraint qualification and weak constant rank condition hold at a primal limit point, then any primal-dual limit point of the same sequence is a second-order KKT point. It follows directly that the limit point must satisfy the second-order necessary conditions for optimality because the Mangasarian-Fromovitz constraint qualification and the weak constant rank condition together define a second-order constraint qualification.

THEOREM 3.10. *Let $\{v_k\} = \{(x_k, y_k)\}$ denote the sequence of primal-dual iterates generated by Algorithm 3. Assume that the algorithm generates infinitely many V- or O-iterates, i.e., $|\mathcal{S}| = \infty$, where $\mathcal{S}$ is the index set*

$$\mathcal{S} = \{k : \textit{iteration } k \textit{ is either a V- or O-iterate}\}.$$

(1) *There exists a subsequence $\mathcal{S}_1 \subseteq \mathcal{S}$ and a limit point x_* such that $\lim_{k\in\mathcal{S}_1} x_k = x_*$.*
(2) *Either x_* fails to satisfy the CPGCQ, or x_* is a first-order KKT point for problem* (NP).
(3) *If x_* a first-order KKT point for problem* (NP)*, then the following results hold.*
   (a) *If the MFCQ holds at x_*, then the sequence $\{y_k\}_{k\in\mathcal{S}_1}$ is bounded, and every limit point y_* defines a first-order KKT pair (x_*, y_*) for problem* (NP).
   (b) *If, in addition, the WCRC holds at x_*, then (x_*, y_*) is a second-order KKT point* (see Definition 3.9).

Proof. For part (1), Assumptions 3.2 and 3.3 imply that there exists a vector x_* and subsequence $\mathcal{S}_1 \subseteq \mathcal{S}$ such that

$$\lim_{k\in\mathcal{S}_1} x_k = x_*. \tag{3.21}$$

If the CPGCQ is not satisfied at x_*, then the first alternative of part (2) holds and there is nothing to prove. For the remainder of the proof it is assumed that CPGCQ holds at x_*. It follows from the properties of the subsequence $\mathcal{S}_1$, the definition of Algorithm 3, and the existence of the limit (3.21) that

$$\{(x_k, y_k)\}_{k\in\mathcal{S}_1} \text{ is an AKKT sequence with } \lim_{k\in\mathcal{S}_1} x_k = x_*. \tag{3.22}$$

If this result is combined with the nonnegativity of x_k imposed by Algorithm 3 and the result of part (1) of Theorem 3.8, it follows that x_* is a first-order KKT point for problem (NP), which proves part (2).

For the proof of part (3a), the assumptions that x_* is a first-order KKT point and the MFCQ holds at x_*, together with the result of part (2) of Theorem 3.8 imply that the set $\{y_k\}_{k\in\mathcal{S}_1}$ is bounded, with every limit point y_* defining a KKT pair (x_*, y_*) for problem (NP).

The result of part (3b) assumes that the WCRC holds at x_* in addition to the MFCQ. It will be shown that, under these conditions, (x_*, y_*) is a second-order KKT point and therefore satisfies the second-order necessary conditions for optimality. Let d_* be any vector such that $\|d_*\| = 1$ and $d_* \in \mathcal{C}_A(x_*)$, where $\mathcal{C}_A$ is defined by (3.20). As the MFCQ and the WCRC hold at x_*, it follows from [5, Lemma 3.1] that there exists a sequence $\{d_k\}$ such that

$$\lim_{k\in\mathcal{S}_1} d_k = d_*, \text{ with } d_k \in \widetilde{C}(x_k) = \{d : J(x_k)d = 0 \text{ and } d_i = 0 \text{ for } i \in \mathcal{A}(x_*)\},$$

Without loss of generality, the elements of the sequence $\{d_k\}_{k\in\mathcal{S}_1}$ may be scaled so that $\|d_k\| = 1$. Consider the set

$$\widehat{C}(x_k, y_k) = \left\{d : J(x_k)d = 0 \text{ and } d_i = 0 \text{ for } i \in \mathcal{A}_\epsilon(x_k, y_k, \mu^R_{k-1})\right\}.$$

As $\{x_k\}_{k\in\mathcal{S}_1}$ is an AKKT sequence from (3.22), and $\{\mu^R_k\} \to 0$ from part (2) of Theorem 3.2, the definition of an ϵ-active set in (2.4) implies that, for all $k \in \mathcal{S}_1$ sufficiently large,

$$\mathcal{A}_\epsilon(x_k, y_k, \mu^R_{k-1}) \subseteq \mathcal{A}(x_*) \quad \text{and} \quad \widetilde{C}(x_k) \subseteq \widehat{C}(x_k, y_k). \tag{3.23}$$

The definition of the set $\mathcal{S}_1$ in (3.21), and the updates made to $\phi_V(x_k, y_k)$ and $\phi_O(x_k, y_k)$ during V- and O-iterates in Algorithm 3, imply that the optimality measure $\omega(x, y)$ of (2.10) satisfies $\omega(x_k, y_k) \to 0$ on $\mathcal{S}_1$. This implies that

$$\lim_{k\in\mathcal{S}_1} \xi^{(1)}_k = 0. \tag{3.24}$$

If $d_{\scriptscriptstyle F}$ is the vector of components of d_k associated with the ϵ-free set $\mathcal{F}_\epsilon(x_k, y_k, \mu^R_{k-1})$, then the definition of d_k as a vector of unit norm in the set $\widetilde{C}(x_k) \subseteq \widehat{C}(x_k, y_k)$ implies that $\|d_{\scriptscriptstyle F}\| = 1$. In addition, the property that $J(x_k)d_k = 0$ implies that

$$\begin{aligned} d_k^T H(x_k, y_k) d_k &= d_k^T\Big(H(x_k, y_k) + (1/\mu^R_{k-1})J_k^T J_k\Big)d_k \\ &= d_{\scriptscriptstyle F}^T\Big(H_{\scriptscriptstyle F} + (1/\mu^R_{k-1})J_{\scriptscriptstyle F}^T J_{\scriptscriptstyle F}\Big)d_{\scriptscriptstyle F} \\ &\geq \lambda_{\min}\Big(H_{\scriptscriptstyle F} + (1/\mu^R_{k-1})J_{\scriptscriptstyle F}^T J_{\scriptscriptstyle F}\Big)\|d_{\scriptscriptstyle F}\|^2 \\ &\geq \frac{u_{\scriptscriptstyle F}^T\big[H_{\scriptscriptstyle F} + (1/\mu^R_{k-1})J_{\scriptscriptstyle F}^T J_{\scriptscriptstyle F}\big]u_{\scriptscriptstyle F}}{\theta\|u_{\scriptscriptstyle F}\|^2} = -\frac{1}{\theta}\xi^{(1)}_k, \end{aligned} \tag{3.25}$$

where the last inequality follows from the definition of $\xi^{(1)}_k$ from Algorithm 1 and the inequality (2.6).

Let $\mathcal{S}_2$ denote the subsequence $\mathcal{S}_2 \subseteq \mathcal{S}_1$ such that $\lim_{k\in\mathcal{S}_2} y_k = y_*$. (The existence of this subsequence is guaranteed by the result of part (3a).) Taking limits of (3.25) over $k \in \mathcal{S}_2 \subseteq \mathcal{S}_1$ and using (3.24) gives

$$d_*^T H(x_*, y_*) d_* = \lim_{k\in\mathcal{S}_2} d_k^T H(x_k, y_k) d_k \geq \lim_{k\in\mathcal{S}_2} -\frac{1}{\theta}\xi^{(1)}_k = 0,$$

which completes the proof. □

THEOREM 3.11. *If the set of V- and O-iterates is finite, then the set $\mathcal{M}$ of M-iterates is infinite and every limit point x_* of $\{x_k\}_{k\in\mathcal{M}}$ satisfies $c(x_*) \neq 0$ and is a KKT point for the feasibility problem*

$$\underset{x\in\mathbb{R}^n}{\text{minimize}} \ \tfrac{1}{2}\|c(x)\|^2 \ \text{ subject to } x \geq 0. \tag{3.26}$$

Proof. The assumptions of this theorem and part (1) of Theorem 3.2 imply that the set

$$\mathcal{M} = \{k : \text{iteration } k \text{ is an M-iterate}\}$$

is infinite and that all iterates are either M- or F-iterate for k sufficiently large. Let x_* be any limit point of $\{x_k\}_{k\in\mathcal{M}}$, which must exist as a consequence of Assumptions 3.3. It must be the case that

$$\lim_{k\in\mathcal{M}_1} x_k = x_* \quad \text{for some } \mathcal{M}_1 \subseteq \mathcal{M}. \tag{3.27}$$

It then follows from the updating scheme used for y_k^E during M- and F-iterates that the sequence $\{y_{k-1}^E\}_{k\in\mathcal{M}_1}$ is bounded. As a consequence, there must exist a vector y_*^E such that

$$\lim_{k\in\mathcal{M}_2} y_{k-1}^E = y_*^E \quad \text{for some } \mathcal{M}_2 \subseteq \mathcal{M}_1. \tag{3.28}$$

The next part of the proof involves showing that the point x_* solves problem (3.26). The proof involves the limits

$$\lim_{k\in\mathcal{M}_2} \left\| \min\left(x_k,\ g_k - J_k^T y_{k-1}^E + \frac{1}{\mu_{k-1}^R} J_k^T c_k\right)\right\| = 0, \tag{3.29a}$$

$$\lim_{k\in\mathcal{M}_2} \|\pi_k - y_k\| = 0, \ \text{ and} \tag{3.29b}$$

$$\lim_{k\in\mathcal{M}_2} \xi_k^{(1)} = 0, \tag{3.29c}$$

where π_k denotes the vector $\pi(x_k, y_{k-1}^E, \mu_{k-1}^R)$. These limits follow from the definition of an M-iterate (2.14a)–(2.14c), and the fact that $\{\tau_k\} \to 0$ is enforced by Algorithm 3 when there are infinitely many M-iterates.

Part (2) of Theorem 3.2 implies that $\{\mu_{k-1}^R\} \to 0$ and the denominator of the second term in the minimization (3.29a) becomes arbitrarily small for $k \in \mathcal{M}_2$ sufficiently large. There are two cases to consider. First, let i be any index such that $[\,x_*\,]_i = 0$. In this case, the limits $\lim_{k\in\mathcal{M}_2} x_k = x_*$ from (3.27) and $\lim_{k\in\mathcal{M}_2} y_{k-1}^E = y_*^E$ from (3.28) imply that for any given positive δ, there is no infinite subsequence of $\mathcal{M}_2$ such that $[\,J_k^T c_k\,]_i \le -\delta$. It follows that $[\,J(x_*)^T c(x_*)\,]_i \ge 0$ for all i such that $[\,x_*\,]_i = 0$.

The second case concerns the indices i such that $[\,x_*\,]_i > 0$. The bounded limits $\lim_{k\in\mathcal{M}_2} x_k = x_*$ and $\lim_{k\in\mathcal{M}_2} y_{k-1}^E = y_*^E$ imply that $g(x_*) - J(x_*)^T y_*^E$ is bounded. It follows that if $[\,x_*\,]_i > 0$, then the property $\{\mu_{k-1}^R\} \to 0$ and the limit (3.29a) imply $[\,J(x_*)^T c(x_*)\,]_i = 0$ and $[\,g(x_*) - J(x_*)^T y_*^E\,]_i = 0$. Combining these two cases and using the fact that every iterate x_k of Algorithm 3 is nonnegative yields

$$\min\left(x_*, J(x_*)^T c(x_*)\right) = 0, \tag{3.30}$$

i.e., x_* is a first-order solution to the feasibility problem (3.26).

It remains to show that $c(x_*) \neq 0$. The proof is by contradiction. If $c(x_*) = 0$, then it follows from (3.29), and the definitions of the feasibility and optimality measures (2.10), that x_k must be a V- or O-iterate for all $k \in \mathcal{M}_2$ sufficiently large, which contradicts the assumption that the number of V- or O-iterates is finite. It follows that $c(x_*) \neq 0$, as required. □

The next result shows that, in practice, Algorithm 3 will terminate in a finite number of iterations.

THEOREM 3.12. *If Algorithm* 3 *is implemented with* $\tau_{\text{stop}} > 0$ *in the termination criteria* (2.35) *and* (2.36) *in Steps* 11 *and* 21*, respectively, the algorithm will terminate in a finite number of iterations.*

Proof. Suppose that there are infinitely many V- or O-iterates given by the sequence $\mathcal{S}$. It then follows from (2.11), the updates used for V- and O-iterates in Algorithm 3, (2.10), and part (2) of Theorem 3.2 that (2.35) will be satisfied for all $k \in \mathcal{S}$ sufficiently large. Thus, Algorithm 3 terminates after a finite number of iterations in Step 11.

Now suppose that the union of the set of V- and O-iterates is a finite set. It then follows from Assumption 3.3 and Theorem 3.11 that there exists a subsequence $\mathcal{M}$ of M-iterates such that

$$\lim_{k \in \mathcal{M}} x_k = x_*, \quad \left\| \min(x_*, J(x_*)^T c(x_*)) \right\| = 0, \quad \text{and} \quad c(x_*) \neq 0. \tag{3.31}$$

It also follows from part (2) of Theorem 3.2, (3.31), and (2.10) that

$$\min(\eta(x_k), \tau_{\text{stop}}) > \mu_k^R \quad \text{for all } k \in \mathcal{M} \text{ sufficiently large.} \tag{3.32}$$

The combination of this bound with the limits (3.31) and the definition of $\mathcal{M}$ implies that (2.36) must be satisfied for all $k \in \mathcal{M}$ sufficiently large, i.e., Algorithm 3 terminates in a finite number of iterations at Step 21. □

4. Conclusions. A regularized SQP method has been proposed based on solving a sequence of strictly convex regularized QP subproblems in which the dual variables appear explicitly in the objective function and constraints. The solution of each subproblem defines a primal-dual search direction for improving a merit function based on a primal-dual augmented Lagrangian. The choice of penalty parameters and Lagrange multiplier estimate in the merit function is motivated by the equivalence of the regularized QP to a problem defined by minimizing a quadratic model of the augmented Lagrangian function subject to the bound constraints. The method is able to use exact second derivative information by incorporating a convexification algorithm that defines a strictly convex QP subproblem without requiring that the Hessian of the Lagrangian be positive definite in the neighborhood of a solution. In addition, the use of a direction of negative curvature in the flexible line search makes it possible to prevent convergence to nonoptimal stationary points. It has been shown that the method terminates in a finite number of iterations with an appropriate choice of termination condition. The method is formulated as a regularized SQP method with an augmented Lagrangian safeguarding strategy. It is shown that safeguarding becomes relevant only when the iterates are converging to an infeasible stationary point of the constraint violations. Otherwise, the method terminates with a point that either satisfies the second-order necessary conditions for optimality, or fails to satisfy a weak second-order constraint qualification.

Additional results have included a proof of convergence to first-order optimal points under weaker conditions than those assumed in [13]. Also, a new proof is given for the result that minimizers of problem (NP) are also minimizers of the bound-constrained primal-dual augmented Lagrangian function. The new proof requires a less restrictive form of the second-order sufficient conditions.

REFERENCES

[1] H. D. I. Abarbanel, P. Bryant, P. E. Gill, M. Kostuk, J. Rofeh, Z. Singer, B. Toth, and E. Wong. Dynamical parameter and state estimation in neuron models. In D. Glanzman and M. Ding, editors, *An Exploration of Neuronal Variability and its Functional Significance*, pages 139–180. Oxford University Press, New York and Oxford, 2011.

[2] A. Altman and J. Gondzio. Regularized symmetric indefinite systems in interior point methods for linear and quadratic optimization. *Optim. Methods Softw.*, 11/12(1-4):275–302, 1999.

[3] R. Andreani, E. G. Birgin, J. M. Martínez, and M. L. Schuverdt. Second-order negative-curvature methods for box-constrained and general constrained optimization. *Comput. Optim. Appl.*, 45(2):209–236, 2010.

[4] R. Andreani, G. Haeser, M. L. Schuverdt, and P. J. S. Silva. Two new weak constraint qualifications and applications. *SIAM J. Optim.*, 22(3):1109–1135, 2012.

[5] R. Andreani, J. M. Martínez, and M. L. Schuverdt. On second-order optimality conditions for nonlinear programming. *Optimization*, 56:529–542, 2007.

[6] P. T. Boggs and J. W. Tolle. Sequential quadratic programming. *Acta Numer.*, 4:1–51, 1995.

[7] A. Castillo and R. P. O'Neill. Computational performance of solution techniques applied to the ACOPF. Optimal Power Flow Paper 5, Federal Energy Regulatory Commission, 2013.

[8] D. R. Creveling, P. E. Gill, and H. D. I. Abarbanel. State and parameter estimation in nonlinear systems as an optimal tracking problem. *Physics Letters A*, 372:2640–2644, 2008.

[9] F. E. Curtis and J. Nocedal. Flexible penalty functions for nonlinear constrained optimization. *IMA J. Numer. Anal.*, 28(4):749–769, 2008.

[10] D. Fernández and M. Solodov. Stabilized sequential quadratic programming for optimization and a stabilized Newton-type method for variational problems. *Math. Program. Ser. A*, 125:47–73, 2010.

[11] A. Forsgren and P. E. Gill. Primal-dual interior methods for nonconvex nonlinear programming. *SIAM J. Optim.*, 8:1132–1152, 1998.

[12] P. E. Gill and D. P. Robinson. A primal-dual augmented Lagrangian. *Computational Optimization and Applications*, pages 1–25, 2010.

[13] P. E. Gill and D. P. Robinson. A globally convergent stabilized SQP method. *SIAM J. Optim.*, 23(4):1983–2010, 2013.

[14] P. E. Gill and E. Wong. Sequential quadratic programming methods. In J. Lee and S. Leyffer, editors, *Mixed Integer Nonlinear Programming*, volume 154 of *The IMA Volumes in Mathematics and its Applications*, pages 147–224. Springer New York, 2012. 10.1007/978-1-4614-1927-3_6.

[15] W. W. Hager. Stabilized sequential quadratic programming. *Comput. Optim. Appl.*, 12(1-3):253–273, 1999. Computational optimization—a tribute to Olvi Mangasarian, Part I.

[16] A. F. Izmailov and M. V. Solodov. On attraction of linearly constrained Lagrangian methods and of stabilized and quasi-Newton SQP methods to critical multipliers. *Math. Program.*, 126(2, Ser. A):231–257, 2011.

[17] V. Kungurtsev. *Second-Derivative Sequential Quadratic Programming Methods for Nonlinear Optimization*. PhD thesis, Department of Mathematics, University of California San Diego, La Jolla, CA, 2013.

[18] O. L. Mangasarian. *Nonlinear programming*. Classics in Applied Mathematics, 10. Society for Industrial and Applied Mathematics (SIAM) Publications, Philadelphia, PA, 1994.

[19] O. L. Mangasarian and S. Fromovitz. The Fritz John necessary optimality conditions in the presence of equality and inequality constraints. *J. Math. Anal. Appl.*, 17:37–47, 1967.

[20] 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.

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

[22] M. A. Saunders and J. A. Tomlin. Solving regularized linear programs using barrier methods and KKT systems. Report SOL 96-4, Dept of EESOR, Stanford University, 1996.

[23] S. J. Wright. Superlinear convergence of a stabilized SQP method to a degenerate solution. *Comput. Optim. Appl.*, 11(3):253–275, 1998.

[24] S. J. Wright. Modifying SQP for degenerate problems. *SIAM J. Optim.*, 13(2):470–497, 2002.

[25] S. J. Wright. An algorithm for degenerate nonlinear programming with rapid local convergence. *SIAM J. Optim.*, 15(3):673–696, 2005.