IMA Journal of Numerical Analysis (2017) **37**, 407–443
doi: 10.1093/imanum/drw004
Advance Access publication on May 4, 2016

A stabilized SQP method: global convergence

PHILIP E. GILL
Department of Mathematics, University of California, San Diego, La Jolla, CA 92093-0112, USA
pgill@ucsd.edu

VYACHESLAV KUNGURTSEV
Department of Computer Science, Agent Technology Center, Faculty of Electrical Engineering, Czech Technical University in Prague 121 35 Praha 2, Prague, Czech Republic
vyacheslav.kungurtsev@fel.cvut.cz

AND

DANIEL P. ROBINSON*
Department of Applied Mathematics and Statistics, Johns Hopkins University, Baltimore, MD 21218-2682, USA
*Corresponding author: daniel.p.robinson@jhu.edu

[Received on 8 August 2014; revised on 11 January 2016]

Stabilized sequential quadratic programming (SQP) methods for nonlinear optimization are designed to provide a sequence of iterates with fast local convergence even when the active-constraint gradients are linearly dependent. This paper concerns the global convergence properties of a stabilized SQP method with a primal–dual augmented Lagrangian merit function. The proposed method incorporates two novel features. First, a flexible line search is used based on a direction formed from an approximate solution of a strictly convex quadratic programming (QP) subproblem and, when one exists, a direction of negative curvature for the primal–dual merit function. Second, when certain conditions hold, an approximate QP solution is computed by solving a single linear system defined in terms of an estimate of the optimal active set. We also establish two desirable convergence results. (i) It is shown that with an appropriate choice of termination condition, the method terminates in a finite number of iterations without the assumption of a constraint qualification. The method may be interpreted as an SQP method with an augmented Lagrangian safeguarding strategy. This safeguarding becomes relevant only when the iterates are converging to an infeasible stationary point of the norm of the constraint violations. Otherwise, the method terminates with a point that approximately satisfies certain second-order necessary conditions for optimality. In this situation, if all termination conditions are removed, then the limit points either satisfy the same second-order necessary conditions exactly or fail to satisfy a weak second-order constraint qualification. (ii) The global convergence analysis concerns a specific algorithm that estimates the least curvature of the merit function at each step. If negative curvature directions are omitted, the analysis still applies and establishes convergence to either first-order solutions or infeasible stationary points. The superlinear convergence of the iterates and the formal local equivalence to stabilized SQP is established in a companion paper (Report CCoM 14-01, *Center for Computational Mathematics*, University of California, San Diego, 2014).

Keywords: nonlinear programming; augmented Lagrangian; sequential quadratic programming; SQP methods; stabilized SQP; primal–dual methods; second-order optimality.

1. Introduction

The nonlinear problem under consideration has the form

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

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 $l \leq x \leq u$. For problem (NP), the vector $g(x)$ is used to denote $\nabla f(x)$, the gradient of f at x. The matrix $J(x)$ denotes the $m\times n$ constraint Jacobian, which has ith row $\nabla c_i(x)^{\mathrm{T}}$, the gradient of the ith constraint function c_i at x. The Lagrangian associated with (NP) is $L(x,y,z)=f(x)-c(x)^Ty-z^Tx$, where y and z are m- and n-vectors of dual variables associated with the equality constraints and non-negativity constraints, respectively. The Hessian of the Lagrangian with respect to x is denoted by $H(x,y)=\nabla^2 f(x)-\sum_{i=1}^m y_i\nabla^2 c_i(x)$.

Sequential quadratic programming (SQP) methods are an important class of methods for nonlinearly constrained optimization (for a survey, see, e.g., Boggs & Tolle, 1995; Gill & Wong, 2012). The kth iteration of a conventional SQP method involves the solution of a quadratic programming (QP) subproblem in which a local quadratic model of the Lagrangian is minimized subject to the linearized constraints. In this paper we focus on the properties of an SQP method that uses a merit function to ensure global convergence. Stabilized SQP methods are designed to resolve some of the numerical and theoretical difficulties associated with SQP methods when they are applied to ill-posed or degenerate nonlinear problems (see, e.g., Wright, 1998, 2002, 2005; Hager, 1999; Oberlin & Wright, 2006; Fernández & Solodov, 2010; Izmailov & Solodov, 2011). Broadly speaking, stabilized SQP methods are designed to provide a sequence of iterates with fast local convergence regardless of whether or not the active-constraint gradients are linearly dependent. There are two important issues associated with the formulation of a practical stabilized SQP method. The first is that stabilized SQP methods have no global convergence theory. One strategy that has been proposed for dealing with this issue is to start by solving the QP subproblem associated with a conventional globally convergent SQP method and switch to the stabilized QP subproblem when it is determined that the iterates are in the proximity of a solution. This strategy may require several switches between a conventional and a stabilized SQP method before the neighbourhood of a solution is identified correctly. The second issue concerns the assumptions needed to guarantee a fast local convergence rate. In general, the Hessian of the Lagrangian is not positive definite in the neighbourhood of a solution, which implies that the stabilized QP subproblem may be nonconvex and have local or unbounded solutions. Nonconvex QP is NP-hard—even for the calculation of a local minimizer (Contesse, 1980; Forsgren *et al.*, 1991). When establishing the local convergence rate of stabilized SQP methods, the potential nonconvexity of the QP subproblem implies that assumptions must be made regarding *which* solution of the QP subproblem is found. This is not a major concern for equality-constrained problems because in this case, the stabilized SQP subproblem has a unique stationary point in the neighbourhood of a solution (see Izmailov & Solodov, 2011, Lemma 2). However, in any stabilized SQP method that must solve a QP subproblem with inequality constraints it is not possible to guarantee that a stationary point found for one subproblem will be close to a stationary point found for the next.

This paper concerns the formulation and global convergence analysis of a stabilized SQP method that uses a primal–dual augmented Lagrangian merit function to ensure convergence from an arbitrary starting point. Given an estimate (x_k, y_k) of a primal–dual solution (x^*, y^*) of problem (NP), the proposed

stabilized SQP method computes a new primal–dual estimate by solving a QP subproblem of the form

$$\begin{array}{ll} \underset{x,y}{\text{minimize}} & g(x_k)^T(x-x_k) + \frac{1}{2}(x-x_k)^{\mathrm{T}}H(x_k,y_k)(x-x_k) + \frac{1}{2}\mu\|y\|^2 \\ \text{subject to} & c(x_k) + J(x_k)(x-x_k) + \mu(y-y_k) = 0, \quad x \geq 0, \end{array} \tag{1.1}$$

where μ ($\mu > 0$) is a small scalar parameter. (It must be emphasized that this QP subproblem is not the same as that used in the conventional stabilized SQP method, which regularizes both equality and inequality constraints. In (1.1) only the equality constraints are regularized.) The analysis indicates that the method has the same strong first- and second-order convergence properties that have been established for augmented Lagrangian methods, while being able to transition seamlessly to stabilized SQP with fast local convergence in the neighbourhood of a solution. This transition is possible, in part, because of the result of Izmailov & Solodov (2011, Lemma 2), which establishes that the stabilized SQP subproblem has a unique solution once the active constraints associated with a local solution have been determined.

The principal contributions of this paper are the following.

1. The method uses a flexible line search along a direction formed from an approximate solution of a strictly convex QP subproblem and, when one exists, a direction of negative curvature for the primal–dual merit function. The superlinear convergence of the iterates and the formal local equivalence to stabilized SQP is established in a companion paper (see Gill *et al.*, 2014). It is not necessary to solve an indefinite QP subproblem.
2. When certain conditions hold, an approximate QP solution is computed by solving a single linear system defined in terms of an estimate of the optimal active set. These conditions may be satisfied at any iterate, but are most likely to be satisfied in the neighbourhood of a solution. The conditions exploit the formal equivalence between the QP subproblem (1.1) and a bound-constrained QP associated with minimizing a quadratic model of the merit function.
3. Convergence to first-order Karush-Kuhn-Tucker (KKT) points is established under weaker conditions than those assumed in Gill & Robinson (2013). It is shown that with an appropriate termination condition, the method terminates in a finite number of iterations without the assumption of a constraint qualification. The method may be interpreted as an SQP method with an augmented Lagrangian safeguarding strategy. This safeguarding becomes relevant only when the iterates are converging to an infeasible stationary point of the norm of the constraint violations. Otherwise, the method terminates with a point that approximately satisfies certain second-order necessary conditions for optimality. In this situation, if all termination conditions are removed, then limit points either satisfy the same second-order necessary conditions exactly or fail to satisfy a weak second-order constraint qualification.

For the main algorithm proposed in Section 2, the cost of each iteration is dominated by the cost of factoring two matrices with so-called 'regularized KKT' structure (see Section 2). However, the algorithm is easily modified so that the negative curvature direction is omitted, in which case the line search is defined in terms of the approximate QP solution only. The resulting algorithm requires only one factorization per iteration and constitutes a globally convergent stabilized SQP method for finding first-order points.

The main algorithm is intended for applications in which some guarantee of convergence to second-order solutions is needed (see, e.g., Creveling *et al.*, 2008; Abarbanel *et al.*, 2011; Castillo & O'Neill, 2013). For other applications, the negative curvature direction may be omitted completely as described above, or may be computed in the final stages of the optimization as a check that the iterates are converging

to a point satisfying second-order optimality conditions. In the latter case the results of this paper and its companion paper Gill *et al.* (2014) imply that a direction of negative curvature, when it is computed, may be used without impeding the overall convergence rate or preventing global convergence.

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 properties of the primal–dual augmented Lagrangian function. Section 2 provides the details of the second-order primal–dual stabilized SQP method. The global convergence of the method is established in Section 3. Section 4 provides final comments and conclusions.

1.1 *Notation*

Unless explicitly indicated otherwise, $\|\cdot\|$ denotes the vector two-norm or its induced matrix norm. 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 1s and the jth column of the identity matrix I; the dimensions of e, e_j and I are defined by the context. The set of integers $\{1, 2, \ldots, n\}$ is denoted by $1 : n$. 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 value of a scalar-, vector- or matrix-valued function F with arguments x and y will be written as either $F(x, y)$ or $F(v)$, where v is the vector (x, y). The ith component of a vector will be denoted by $[\,\cdot\,]_i$, e.g., $[v]_i$ is the ith component of the vector v. For a given l-vector u and index set $\mathcal{S}$, the quantity $[u]_{\mathcal{S}}$ denotes the subvector of components u_j such that $j \in \mathcal{S} \cap \{1, 2, \ldots, l\}$. Similarly, if M is a symmetric $l \times l$ matrix, then $[M]_{\mathcal{S}}$ denotes the symmetric matrix with elements m_{ij} for $i, j \in \mathcal{S} \cap \{1, 2, \ldots, l\}$. Given a sequence $\{\alpha_j\}_{j\geq 0}$ of scalars, vectors or matrices, we write $\alpha_j = \mathcal{O}(\beta_j)$ if there exists a positive constant γ such that $\|\alpha_j\| \leq \gamma\beta_j$ for some positive scalar sequence $\{\beta_j\}_{j\geq 0}$.

1.2 *Background*

At the basic level, the method proceeds to compute a first-order KKT point of (NP), which is defined formally as follows.

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

$$r(x, y) = \left\| \left(c(x), \min\left(x, g(x) - J(x)^T y\right)\right)\right\|. \tag{1.2}$$

Any (x^*, y^*) satisfying $r(x^*, y^*) = 0$, is called a *first-order KKT pair*.

If a suitable constraint qualification holds then a local minimizer x^* for problem (NP) is a first-order KKT point. In this case, any y^* associated with a KKT pair (x^*, y^*) is a vector of Lagrange multipliers for the constraints $c(x) = 0$. Given a first-order KKT point x^*, the nonempty set of dual vectors $\mathcal{Y}(x^*)$ is defined so that

$$\mathcal{Y}(x^*) = \{y \in \mathbb{R}^m : (x^*, y) \ \text{ satisfies } \ r(x^*, y) = 0\}. \tag{1.3}$$

For any $x \geq 0$, the active set at x is given by

$$\mathcal{A}(x) = \{i : [x]_i = 0\}. \tag{1.4}$$

At a primal–dual first-order solution (x^*, y^*), second-order conditions for (NP) may be defined in terms of a partition of $\mathcal{A}(x^*)$. The index set of *strongly active* variables is given by

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

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

$$\mathcal{A}_0(x^*, y^*) = \{i \in \mathcal{A}(x^*) : [g(x^*) - J(x^*)^T y^*]_i = 0\}. \tag{1.6}$$

A set of second-order conditions is given in the following definition.

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. (1.2)), and

$$p^T H(x^*, y^*) p > 0 \quad \text{for all} \quad p \in \mathcal{C}(x^*, y^*) \setminus \{0\}, \tag{1.7}$$

where $\mathcal{C}(x^*, y^*)$ is the critical cone

$$\mathcal{C}(x^*, y^*) = \text{null}\big(J(x^*)\big) \cap \{p : p_i = 0 \ \text{ for } \ i \in \mathcal{A}_+(x^*, y^*), \ p_i \geq 0 \ \text{ for } \ i \in \mathcal{A}_0(x^*, y^*)\}.$$

The proposed method is based on the primal–dual augmented Lagrangian function of Gill & Robinson (2013), which was first used to define a 'regularized SQP method' in which the linear system of equations solved at each iteration of the QP subproblem is guaranteed to be nonsingular. The function is given by

$$M(x, y; y^E, \mu) = f(x) - c(x)^T y^E + \frac{1}{2\mu}\|c(x)\|^2 + \frac{1}{2\mu}\|c(x) + \mu(y - y^E)\|^2, \tag{1.8}$$

where μ is a positive penalty parameter and y^E is a Lagrange multiplier estimate. (The function M is a member of a one-parameter family of functions that includes the conventional augmented Lagrangian. For more details, see Robinson, 2007; Gill & Robinson, 2012.) The next result, which is proved in Gill *et al.* (2013, Theorem A.1), motivates the use of (1.8) as an SQP merit function.

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 problem of minimizing the primal–dual function $M(x, y; y^E, \mu)$ of (1.8) subject to the non-negativity constraints $x \geq 0$.

The result of Theorem 1.3 implies that problem (NP) may be replaced by a sequence of bound-constrained problems with objective function $M(x, y; y_k^E, \mu)$ defined in terms of a sequence of multiplier estimates $\{y_k^E\}$. This approach defines an inner/outer iteration structure, with the inner iterations being those of

the active-set method used to minimize a quadratic model of the merit function subject to the bound constraints. Each quadratic model is defined in terms of the gradient and Hessian of M. For given values y^E and μ, the gradient and Hessian of M at (x, y) may be written in the form

$$\nabla M(x, y\,; y^E, \mu) = \begin{pmatrix} g(x) - J(x)^T\big(\pi(x\,; y^E, \mu) + (\pi(x\,; y^E, \mu) - y)\big) \\ \mu(y - \pi(x\,; y^E, \mu)) \end{pmatrix} \tag{1.9}$$

and

$$\nabla^2 M(x, y\,; y^E, \mu) = \begin{pmatrix} H\big(x, \pi(x\,; y^E, \mu) + (\pi(x\,; y^E, \mu) - y)\big) + \frac{2}{\mu} J(x)^T J(x) & J(x)^{\mathrm{T}} \\ J(x) & \mu I \end{pmatrix}, \tag{1.10}$$

where π denotes the vector-valued function $\pi(x\,; y^E, \mu) = y^E - c(x)/\mu$. If $v_k = (x_k, y_k)$ denotes the kth estimate of a primal–dual solution of problem (NP), one possible local quadratic model of the change in $M(x, y\,; y_k^E, \mu)$ is given by

$$\mathcal{Q}_k(v\,; y_k^E, \mu) = (v - v_k)^{\mathrm{T}} \nabla M(v_k\,; y_k^E, \mu) + \tfrac{1}{2}(v - v_k)^{\mathrm{T}} B(v_k\,; \mu)(v - v_k), \tag{1.11}$$

where y_k^E estimates a vector of Lagrange multipliers, and $B(v_k\,; \mu)$ is defined by replacing $\pi(x_k\,; y_k^E, \mu)$ by y_k in the leading block of $\nabla^2 M$, i.e.,

$$B(x_k, y_k\,; \mu) = \begin{pmatrix} H(x_k, y_k) + \frac{2}{\mu} J(x_k)^T J(x_k) & J(x_k)^{\mathrm{T}} \\ J(x_k) & \mu I \end{pmatrix}. \tag{1.12}$$

Both this approximation and the exact Hessian $\nabla^2 M$ have the same characteristic *doubly augmented* structure involving the term $(2/\mu)J(x_k)^T J(x_k)$ in the leading diagonal block. However, unlike $\nabla^2 M$, the matrix $B(x_k, y_k\,; \mu)$ is independent of the multiplier estimate y_k^E (cf. (1.10)). The benefit of using $B(x_k, y_k\,; \mu)$ to define the quadratic model (1.11) is that the QP subproblem

$$\underset{v}{\text{minimize}} \;\; \mathcal{Q}_k(v\,; y_k^E, \mu) \;\; \text{subject to} \;\; [v]_i \geq 0, \quad i = 1 : n, \tag{1.13}$$

is formally equivalent to the SQP subproblem

$$\begin{array}{ll} \underset{x,y}{\text{minimize}} & g(x_k)^T(x - x_k) + \frac{1}{2}(x - x_k)^{\mathrm{T}} H(x_k, y_k)(x - x_k) + \frac{1}{2}\mu\|y\|^2 \\ \text{subject to} & c(x_k) + J(x_k)(x - x_k) + \mu(y - y_k^E) = 0, \quad x \geq 0 \end{array} \tag{1.14}$$

(see Gill & Robinson, 2013). This equivalence suggests an algorithm based on minimizing (1.13) with $\mathcal{Q}_k$ defined with a small μ analogous to the perturbation in the QP subproblem (1.1). However, if $M(x, y\,; y_k^E, \mu)$ is to serve as a line-search merit function, it should be defined with a value of μ that is not too small. Accordingly, different penalty parameters μ_k^R and μ_k are defined, with μ_k^R used for the definition of the local quadratic model, and μ_k used for the line-search merit function. In the neighbourhood of a solution, μ_k^R and μ_k are computed so that $\mu_k^R \ll \mu_k$, which implies that μ_k^R plays the role of the stabilization parameter μ in (1.1).

2. A second-order primal–dual stabilized SQP algorithm

In this section we describe a method designed to compute a point satisfying certain first- and second-order conditions for a solution of (NP). At the start of the kth iteration, the primal–dual point (x_k, y_k) is known, together with the regularization parameter μ_{k-1}^R and penalty parameter μ_{k-1} from the previous iteration. The first step is to compute y_k^E and μ_k^R for the new iteration. These parameters are defined in terms of an estimate of the optimal active set of problem (NP). This estimate involves a positive scalar ϵ that reflects the distance of (x, y) to a first-order optimal pair for problem (NP). The ϵ-active set is defined as

$$\mathcal{A}_\epsilon(x, y, \mu) = \big\{ i : x_i \le \epsilon, \ \text{ with } \ \epsilon \equiv \min\big(\epsilon_a, \max\big(\mu, r(x, y)^\gamma\big)\big)\big\}, \tag{2.1}$$

where γ and ϵ_a are fixed scalars satisfying $0 < \gamma < 1$ and $0 < \epsilon_a < 1$, and $r(x, y)$ is the non-negative scalar of (1.2). If $\mathcal{A}$ is the active set (1.4) then the set of primal–dual free variables is the complement of $\mathcal{A}$ in $\{1, 2, \ldots, n + m\}$, i.e.,

$$\mathcal{F}(x) = \{1, 2, \ldots, n + m\} \setminus \mathcal{A}(x). \tag{2.2}$$

Similarly, the ϵ-free set is defined as

$$\mathcal{F}_\epsilon(x, y, \mu) = \{1, 2, \ldots, n + m\} \setminus \mathcal{A}_\epsilon(x, y, \mu). \tag{2.3}$$

The ϵ-free set defines the composition of certain matrices used to compute a direction of negative curvature for $\mathcal{Q}_k(v\,; y_{k-1}^E, \mu_{k-1}^R)$, and the estimate of the active set for the convex QP subproblem.

The following simple argument shows 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^*). If $i \notin \mathcal{A}(x^*)$ then $x_i^* > 0$. Assume that (x, y) is sufficiently close to (x^*, y^*) and μ is sufficiently small that: (i) $r(x, y)^\gamma < \mu \le x_i^*$, (ii) $|x_i - x_i^*| < x_i^* - \mu$ and (iii) $\mu \le \epsilon_a$. Then

$$\mathcal{A}_\epsilon(x, y, \mu) = \big\{ i : x_i \le \min\big(\epsilon_a, \max\big(\mu, r(x, y)^\gamma\big)\big)\big\} = \big\{ i : x_i \le \mu \big\},$$

and the inequality $|x_i - x_i^*| < x_i^* - \mu$ implies $x_i > \mu$, in which case $i \notin \mathcal{A}_\epsilon(x, y, \mu)$. We have shown that if $i \notin \mathcal{A}(x^*)$ then $i \notin \mathcal{A}_\epsilon(x, y, \mu)$, which implies that $\mathcal{F}(x^*) \subseteq \mathcal{F}_\epsilon(x, y, \mu)$, as required.

At the kth iterate (x_k, y_k) a flexible line search is performed along a direction formed from two primal–dual directions d_k and s_k, either of which may be zero. A nonzero d_k is a descent direction for the merit function; a nonzero s_k is a direction of negative curvature for the quadratic model $\mathcal{Q}_k(v_k\,; y_{k-1}^E, \mu_{k-1}^R)$. The descent direction d_k is either a 'local descent step' designed to facilitate a fast rate of local convergence, or a 'global descent step' designed to encourage global convergence. In each case, d_k is an approximate solution of a QP subproblem with bound constraints.

Overview of the step computation. If a solution of the subproblem (1.13) is written in terms of the step d_k from the base point v_k then the QP optimality conditions are given by

$$\begin{gathered}[\nabla \mathcal{Q}_k(v_k + d_k\,; y_k^E, \mu_k^R)]_{\mathcal{F}(v_k+d_k)} = 0, \quad [\nabla \mathcal{Q}_k(v_k + d_k\,; y_k^E, \mu_k^R)]_{\mathcal{A}(v_k+d_k)} \ge 0 \ \text{ and} \\ [v_k + d_k]_i \ge 0 \ \text{ for } \ i = 1 : n,\end{gathered} \tag{2.4}$$

where the free and active sets are evaluated at the QP solution $v_k + d_k$. The *local descent step* is the solution of an equality-constrained QP subproblem defined by relaxing the optimality conditions (2.4). (See

Section 2.3 for more details.) The *global descent step* is the solution of a convex inequality-constrained QP subproblem. If $v_k = (x_k, y_k)$ is not close to a solution, the matrix $B(v_k\,;\mu_k^R)$ may not be positive definite and (1.13) is not an appropriate QP subproblem for the global descent step. Instead, a strictly convex bound-constrained QP subproblem is defined that preserves the positive curvature of the nonconvex quadratic model (1.11) on the reduced space associated with the free variables. The strictly convex QP is defined in terms of the matrix $\widehat{B}(v_k\,;\mu)$ such that

$$\widehat{B}(x_k, y_k\,;\mu) = \begin{pmatrix} \widehat{H}(x_k, y_k) + \frac{2}{\mu}J(x_k)^T J(x_k) & J(x_k)^{\mathrm{T}} \\ J(x_k) & \mu I \end{pmatrix} \tag{2.5}$$

with the symmetric matrix $\widehat{H}(x_k, y_k)$ defined so that $\widehat{H}(x_k, y_k) + (1/\mu)J(x_k)^T J(x_k)$ is positive definite (the identity (2.7) below justifies this choice). With this choice of $\widehat{H}(x_k, y_k)$, the matrix $\widehat{B}_{\mathcal{F}_\epsilon}(x_k, y_k\,;\mu_k^R)$ is equal to $B_{\mathcal{F}_\epsilon}(x_k, y_k\,;\mu_k^R)$ when $B_{\mathcal{F}_\epsilon}(x_k, y_k\,;\mu_k^R)$ is positive definite. A method for computing a suitable symmetric (but not necessarily positive-definite) matrix $\widehat{H}(x_k, y_k)$ is given by Gill & Robinson (2013, Section 4).

Once $\widehat{B}(x_k, y_k\,;\mu)$ has been computed, the global descent step is computed by solving the QP subproblem

$$\begin{aligned} &\underset{v}{\text{minimize}}\ \ \widehat{\mathcal{Q}}_k(v\,;y_k^E, \mu_k^R) = (v - v_k)^{\mathrm{T}}\nabla M(v_k\,;y_k^E, \mu_k^R) + \tfrac{1}{2}(v - v_k)^{\mathrm{T}}\widehat{B}(v_k\,;\mu_k^R)(v - v_k) \\ &\text{subject to}\ \ [v]_i \geq 0, \quad i = 1:n. \end{aligned} \tag{2.6}$$

Throughout what follows, the unique primal–dual solution of this QP subproblem is denoted by $\hat{v}_k$, which has primal and dual components $\hat{v}_k = (\hat{x}_k, \hat{y}_k)$. The global descent step is then computed as $d_k = \hat{v}_k - v_k = (\hat{x}_k - x_k, \hat{y}_k - y_k)$.

For numerical stability, it is important that all computations be performed without the explicit calculation of the matrices $\nabla^2 M$ or B. The relevant properties of B may be determined by exploiting the relationship between the three matrices

$$\begin{pmatrix} H(x,y) + \frac{2}{\mu}J(x)^T J(x) & J(x)^{\mathrm{T}} \\ J(x) & \mu I \end{pmatrix}, \quad H(x,y) + \frac{1}{\mu}J(x)^T J(x) \quad \text{and} \quad \begin{pmatrix} H(x,y) & J(x)^{\mathrm{T}} \\ J(x) & -\mu I \end{pmatrix},$$

which are said to have doubly augmented form, augmented Lagrangian form, and regularized KKT form, respectively. These matrices are related via the matrix identities

$$\begin{pmatrix} H + \frac{1+\nu}{\mu}J^T J & \nu J^{\mathrm{T}} \\ \nu J & \nu\mu I \end{pmatrix} = \begin{pmatrix} I & \frac{1}{\mu}J^{\mathrm{T}} \\ 0 & I \end{pmatrix}\begin{pmatrix} H + \frac{1}{\mu}J^T J & 0 \\ 0 & \nu\mu I \end{pmatrix}\begin{pmatrix} I & 0 \\ \frac{1}{\mu}J & I \end{pmatrix} \tag{2.7}$$

$$= \begin{pmatrix} I & \frac{1+\nu}{\nu\mu}J^{\mathrm{T}} \\ 0 & \frac{1}{\nu}I \end{pmatrix}\begin{pmatrix} H & J^{\mathrm{T}} \\ J & -\mu I \end{pmatrix}\begin{pmatrix} I & \\ & -I \end{pmatrix}, \tag{2.8}$$

which hold for all positive ν (see, e.g., Forsgren & Gill, 1998; Robinson, 2007; Gill & Robinson, 2013). For simplicity of exposition, the formulation and analysis of the main algorithm is given in terms of matrices in augmented Lagrangian form. However, in practice, all computations are performed by factoring matrices in regularized KKT form. At no point is it necessary to form a matrix in doubly augmented or augmented Lagrangian form. The term 'regularized KKT form' stems from the role of μ as a *regularization* parameter

for the matrix in (2.8). (For other examples of this type of regularization, see Saunders & Tomlin, 1996; Altman & Gondzio, 1999; Gill & Robinson, 2013.) In practice, the global descent step is found by solving the QP form (1.14) of the bound-constrained subproblem with $H(x,y_k) = \widehat{H}(x_k,y_k)$. Gill & Robinson (2013, Theorem 5.2) show that the application of a conventional active-set method to (1.14) involves the solution of a sequence of linear equations in regularized KKT form.

Summary of the main algorithm. The computation associated with the kth iteration of the main algorithm (Algorithm 5 of Section 2.5) is arranged into five principal steps.

1. Given (x_k,y_k) and the regularization parameter μ_{k-1}^R from the previous iteration, compute $\mathcal{F}_\epsilon(x_k,y_k,\mu_{k-1}^R)$ and $B(x_k,y_k\,;\mu_{k-1}^R)$. Compute the non-negative scalar ξ_k and vector $s_k^{(1)}$ such that, if $\xi_k > 0$ then ξ_k approximates the magnitude of the 'most negative' or 'least' eigenvalue of $B_{\mathcal{F}_\epsilon}(v_k\,;\mu_{k-1}^R)$, and $s_k^{(1)}$ satisfies $s_k^{(1)\mathrm{T}}B(v_k\,;\mu_{k-1}^R)s_k^{(1)} < 0$ (see Section 2.1). If $\xi_k = 0$ then $s_k^{(1)} = 0$. If $B_{\mathcal{F}_\epsilon}(v_k\,;\mu_{k-1}^R)$ is positive definite then $(\xi_k,s_k^{(1)}) = 0$. (See Algorithm 1.)
2. Use ξ_k and $r(x_k,y_k)$ to compute y_k^E and μ_k^R for the kth iteration (see Section 2.2).
3. Compute a positive-definite $\widehat{B}(v_k\,;\mu_k^R)$ such that $\widehat{B}_{\mathcal{F}_\epsilon}(x_k,y_k\,;\mu_k^R) = B_{\mathcal{F}_\epsilon}(x_k,y_k\,;\mu_k^R)$ if the matrix $B_{\mathcal{F}_\epsilon}(x_k,y_k\,;\mu_k^R)$ is positive definite. Compute $d_k = (p_k,q_k)$ as either a local or global descent step. The vector p_k of primal components of d_k satisfies $x_k + p_k \geq 0$. (See Section 2.3 and Algorithm 2.)
4. Compute a direction of negative curvature $s_k = (u_k,w_k)$ by rescaling the direction $s_k^{(1)}$. The vector u_k of primal components of s_k satisfies $x_k + p_k + u_k \geq 0$. (See Section 2.3 and Algorithm 3.)
5. Perform a flexible line search along the vector $\Delta v_k = s_k + d_k = (u_k + p_k, w_k + q_k)$. (See Section 2.4 and Algorithm 4.) Update the line-search penalty parameter μ_k.

2.1 *Computing a direction of negative curvature for $\mathcal{Q}_k$*

The values of the regularization parameter μ_k^R and multiplier estimate y_k^E for the kth iteration depend on an estimate of the smallest eigenvalue of $B_{\mathcal{F}_\epsilon}(x_k,y_k\,;\mu_{k-1}^R)$ computed from the ϵ-free rows and columns of the approximate Hessian $B(x_k,y_k\,;\mu_{k-1}^R)$ given in (1.12). This estimate may be computed in terms of an estimate of $\lambda_{\min}\big(H_{\mathcal{F}_\epsilon} + (1/\mu_{k-1}^R)J_{\mathcal{F}_\epsilon}^T J_{\mathcal{F}_\epsilon}\big)$, where $H_{\mathcal{F}_\epsilon}$ is the matrix of ϵ-free rows and columns of $H(x_k,y_k)$, and $J_{\mathcal{F}_\epsilon}$ is the matrix of ϵ-free columns of $J(x_k)$.

Algorithm 1 summarizes the principal calculations associated with finding a non-negative estimate ξ_k of $\max\{0,-\lambda_{\min}\}$, where $\lambda_{\min}$ denotes the least eigenvalue of $H_{\mathcal{F}_\epsilon} + (1/\mu_{k-1}^R)J_{\mathcal{F}_\epsilon}^T J_{\mathcal{F}_\epsilon}$. If the computed ξ_k is positive, then a by-product of the computation is a feasible approximate direction of negative curvature for the merit function $M(x_k,y_k\,;y_k^E,\mu_{k-1}^R)$. 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 $H_{\mathcal{F}_\epsilon} + (1/\mu_{k-1}^R)J_{\mathcal{F}_\epsilon}^{\mathrm{T}} J_{\mathcal{F}_\epsilon}$ is positive semidefinite. If $H_{\mathcal{F}_\epsilon} + (1/\mu_{k-1}^R)J_{\mathcal{F}_\epsilon}^T J_{\mathcal{F}_\epsilon}$ is not positive semidefinite, the procedure must provide a direction $u_{\mathcal{F}_\epsilon}$ satisfying

$$u_{\mathcal{F}_\epsilon}^{\mathrm{T}}\Big(H_{\mathcal{F}_\epsilon} + \frac{1}{\mu_{k-1}^R}J_{\mathcal{F}_\epsilon}^T J_{\mathcal{F}_\epsilon}\Big)u_{\mathcal{F}_\epsilon} \leq \theta\lambda_{\min}\Big(H_{\mathcal{F}_\epsilon} + \frac{1}{\mu_{k-1}^R}J_{\mathcal{F}_\epsilon}^T J_{\mathcal{F}_\epsilon}\Big)\|u_{\mathcal{F}_\epsilon}\|^2 < 0, \tag{2.9}$$

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 $u_{\mathcal{F}_\epsilon}$, but the details of the computation are not relevant for the proof of global convergence.

Algorithm 1 Least curvature estimate of $\mathcal{Q}_k$

1: **procedure** LEAST_CURVATURE_ESTIMATE($x_k, y_k, \mu^R_{k-1}, J_k, H_k$)
2: Compute $H_{\mathcal{F}_\epsilon}$ and $J_{\mathcal{F}_\epsilon}$ as submatrices of H_k and J_k associated with $\mathcal{F}_\epsilon(x_k, y_k, \mu^R_{k-1})$;
3: **if** $H_{\mathcal{F}_\epsilon} + (1/\mu^R_{k-1})J^T_{\mathcal{F}_\epsilon}J_{\mathcal{F}_\epsilon}$ is positive semidefinite **then**
4: $\xi_k = 0; \quad u_k^{(1)} = 0; \quad w_k^{(1)} = 0;$
5: **else**
6: Compute $u_{\mathcal{F}_\epsilon} \neq 0$ such that (2.9) is satisfied;
7: $\xi_k = -u^{\mathrm{T}}_{\mathcal{F}_\epsilon}\big(H_{\mathcal{F}_\epsilon} + (1/\mu^R_{k-1})J^T_{\mathcal{F}_\epsilon}J_{\mathcal{F}_\epsilon}\big)u_{\mathcal{F}_\epsilon}/\|u_{\mathcal{F}_\epsilon}\|^2 > 0;$
8: $u_k^{(1)} = 0; \quad [u_k^{(1)}]_{\mathcal{F}_\epsilon} = u_{\mathcal{F}_\epsilon};$
9: $w_k^{(1)} = -(1/\mu^R_{k-1})J_k u_k^{(1)};$
10: **end if**
11: $s_k^{(1)} = (u_k^{(1)}, w_k^{(1)});$
12: **return** $(s_k^{(1)}, \xi_k);$
13: **end procedure**

A specific procedure appropriate for a matrix in regularized KKT form is given by Forsgren & Gill (1998, Section 4.3). Inequality (2.9) allows the definition of a direction of negative curvature for the quadratic model $\mathcal{Q}_k(v\,; y^E_{k-1}, \mu^R_{k-1})$ when $H_{\mathcal{F}_\epsilon} + (1/\mu^R_{k-1})J^{\mathrm{T}}_{\mathcal{F}_\epsilon}J_{\mathcal{F}_\epsilon}$ is not positive semidefinite (see Lemma 2.1).

The vector $w_k^{(1)}$ computed 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 $J_k u_k^{(1)} + \mu^R_{k-1} w_k^{(1)} = 0$, which implies that the step $(x_k + u_k^{(1)}, y_k + w_k^{(1)})$ does not change the residuals of the constraints of the QP subproblem (1.14). It also facilitates a proof that the resulting direction $s_k^{(1)}$ is a direction of negative curvature for the quadratic model $\mathcal{Q}_k(v\,; y^E_{k-1}, \mu^R_{k-1})$ when $\xi_k > 0$ (see Lemma 2.1).

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

LEMMA 2.1 Suppose that ξ_k and $s_k^{(1)} = (u_k^{(1)}, w_k^{(1)})$ are computed by Algorithm 1, and that the matrix $H_{\mathcal{F}_\epsilon} + (1/\mu^R_{k-1})J^T_{\mathcal{F}_\epsilon}J_{\mathcal{F}_\epsilon}$ is *not* positive semidefinite. Then

(1) $u_k^{(1)} \neq 0$, $s_k^{(1)} \neq 0$, $\xi_k > 0$; and
(2) $s_k^{(1)}$ is a direction of negative curvature for $\mathcal{Q}_k(v\,; y^E_{k-1}, \mu^R_{k-1})$, the quadratic model of (1.11). In particular, $s_k^{(1)}$ satisfies $s_k^{(1)\mathrm{T}} B(v_k\,; \mu^R_{k-1}) s_k^{(1)} \leq \bar{\theta}_k \|u_k^{(1)}\|^2 < 0$, where $\bar{\theta}_k = \theta\lambda_{\min}\big(H_{\mathcal{F}_\epsilon} + (1/\mu^R_{k-1})J^{\mathrm{T}}_{\mathcal{F}_\epsilon}J_{\mathcal{F}_\epsilon}\big)$ with θ the value associated with the bound (2.9).

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

As $H_{\mathcal{F}_\epsilon} + (1/\mu)J^T_{\mathcal{F}_\epsilon}J_{\mathcal{F}_\epsilon}$ is not positive semidefinite by assumption, Algorithm 1 computes a nonzero direction $u_{\mathcal{F}_\epsilon}$ that satisfies (2.9). In this case,

$$u \neq 0, \quad s \neq 0, \quad \lambda_{\min}\Big(H_{\mathcal{F}_\epsilon} + \frac{1}{\mu}J^T_{\mathcal{F}_\epsilon}J_{\mathcal{F}_\epsilon}\Big) < 0, \quad \text{and} \quad \xi > 0, \tag{2.10}$$

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

$$s^T Bs = \begin{pmatrix} u \\ w \end{pmatrix}^T \begin{pmatrix} H + \frac{2}{\mu} J^T J & J^T \\ J & \mu I \end{pmatrix} \begin{pmatrix} u \\ w \end{pmatrix}$$

$$= u^T Hu + \frac{2}{\mu} u^T J^T Ju + 2u^T J^T w + \mu \|w\|^2 = u^T \Big(H + \frac{1}{\mu} J^T J \Big) u. \tag{2.11}$$

The definition of u in step 8 of Algorithm 1, and the requirement that $u_{\mathcal{F}_\epsilon}$ satisfies (2.9), yield the inequality

$$u^T \Big(H + \frac{1}{\mu} J^T J \Big) u \leq \theta \lambda_{\min} \Big(H_{\mathcal{F}_\epsilon} + \frac{1}{\mu} J_{\mathcal{F}_\epsilon}^T J_{\mathcal{F}_\epsilon} \Big) \|u\|^2. \tag{2.12}$$

Given the definition $\bar{\theta} = \theta \lambda_{\min}(H_{\mathcal{F}_\epsilon} + \frac{1}{\mu} J_{\mathcal{F}_\epsilon}^T J_{\mathcal{F}_\epsilon})$, it must hold that $\bar{\theta} < 0$, where the strict inequality follows from (2.10). The result now follows directly from (2.11), (2.12) and the definition of $u = u_k^{(1)}$ as a nonzero vector in Algorithm 1. □

2.2 *Updating the multiplier estimate and regularization parameter*

Given x_k, y_k, y_{k-1}^E and μ_{k-1}^R from the previous iteration, the computation of y_k^E and μ_k^R requires the scalar ξ_k computed by Algorithm 1, and scalars $\phi_{V,k-1}^{\max}$, $\phi_{O,k-1}^{\max}$ and τ_{k-1} discussed below.

The multiplier estimate y_k^E is set to y_k if (x_k, y_k) gives an improvement in a measure of the distance to a primal–dual second-order solution (x^*, y^*). The main algorithm (Algorithm 5 of Section 2.5) uses the feasibility and optimality measures $\eta(x_k)$ and $\omega(x_k, y_k, \xi_k)$ such that

$$\eta(x_k) = \|c(x_k)\| \quad \text{and} \quad \omega(x_k, y_k, \xi_k) = \max\left(\left\| \min(x_k, g(x_k) - J(x_k)^T y_k) \right\|, \ \xi_k \right), \tag{2.13}$$

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

$$\phi_{\mathrm{V}}(x_k, y_k) = \eta(x_k) + \beta \omega(x_k, y_k, \xi_k) \qquad \text{and} \qquad \phi_{\mathrm{O}}(x_k, y_k, \xi_k) = \beta \eta(x_k) + \omega(x_k, y_k, \xi_k),$$

where β is a fixed scalar such that $0 < \beta \ll 1$. The subscripts on ϕ_{V} and ϕ_{O} reflect the preference given to reducing the violation measure or the optimality measure, as defined by the placement of the parameter β. The update $y_k^E = y_k$ is performed if

$$\phi_{\mathrm{V}}(v_k) \leq \tfrac{1}{2} \phi_{V,k-1}^{\max} \qquad \text{or} \qquad \phi_{\mathrm{O}}(v_k, \xi_k) \leq \tfrac{1}{2} \phi_{O,k-1}^{\max}, \tag{2.14}$$

in which case the kth iterate is called a V-iterate or an O-iterate, respectively. A 'V–O-iterate' is any point for which one or both of these conditions holds. The associated iteration (or iteration index) is called a 'V–O iteration'. The quantities $\phi_{V,k-1}^{\max}$ and $\phi_{O,k-1}^{\max}$ are the values of positive bounds that are reduced during the solution process. For a V–O iteration, new values are given by $\tau_k = \frac{1}{2} \tau_{k-1}$, and $\phi_{V,k}^{\max} = \frac{1}{2} \phi_{V,k-1}^{\max}$ or $\phi_{O,k}^{\max} = \frac{1}{2} \phi_{O,k-1}^{\max}$, depending on which of the inequalities (2.14) holds. In addition, the regularization parameter is computed as

$$\mu_k^R = \begin{cases} \min\left(\mu_0^R, \ \max\left(r_k, \xi_k \right)^\gamma \right) & \text{if } \max\left(r_k, \xi_k \right) > 0, \\ \frac{1}{2} \mu_{k-1}^R & \text{otherwise,} \end{cases} \tag{2.15}$$

for some fixed $\gamma \in (0,1)$, and where $r_k \equiv r(x_k, y_k)$ is the first-order optimality measure (1.2). Note that the definition (2.15) implies that the sequence $\{\mu_k^R\}$ is not necessarily monotonic.

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

$$\underset{x,y}{\text{minimize}} \;\; M(x, y\,; y_{k-1}^E, \mu_{k-1}^R) \quad \text{subject to} \quad x \geq 0. \tag{2.16}$$

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

$$\| \min(x_k, \nabla_x M(x_k, y_k\,; y_{k-1}^E, \mu_{k-1}^R)) \| \leq \tau_{k-1}, \tag{2.17a}$$

$$\| \nabla_y M(x_k, y_k\,; y_{k-1}^E, \mu_{k-1}^R) \| \leq \tau_{k-1} \mu_{k-1}^R \;\; \text{and} \tag{2.17b}$$

$$\xi_k \leq \tau_{k-1}, \tag{2.17c}$$

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_{k-1}^E is replaced by the safeguarded value

$$y_k^E = \max\big(-y_{\max} e,\; \min(y_k,\; y_{\max} e) \big) \tag{2.18}$$

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

$$\mu_k^R = \begin{cases} \min\big(\frac{1}{2}\mu_{k-1}^R,\; \max\big(r_k, \xi_k\big)^{\gamma} \big) & \text{if } \max(r_k, \xi_k) > 0, \\ \frac{1}{2}\mu_{k-1}^R & \text{otherwise.} \end{cases} \tag{2.19}$$

In addition, a new tolerance τ_k is computed such that $\tau_k = \frac{1}{2}\tau_{k-1}$. Numerical results given by Gill *et al.* (2014) indicate that M-iterates occur infrequently relative to the total number of iterations.

Finally, if neither (2.14) nor (2.17) are satisfied then $y_k^E = y_{k-1}^E$, $\mu_k^R = \mu_{k-1}^R$, $\phi_{V,k}^{\max} = \phi_{V,k-1}^{\max}$, $\phi_{Q,k}^{\max} = \phi_{Q,k-1}^{\max}$ and $\tau_k = \tau_{k-1}$. 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*

Once μ_k^R and y_k^E have been computed, the line-search directions s_k and d_k are computed. This section provides more details of the computation of the local and global descent directions, and direction of negative curvature introduced in the overview of the step computation. The computations are summarized in Algorithms 2 and 3 below.

2.3.1 *The local descent direction.* A local descent direction is computed if $B_{\mathcal{F}_\epsilon}(x_k, y_k\,; \mu_k^R)$ is positive definite and v_k is a V–O-iterate (in which case $y_k^E = y_k$). Under these conditions, $d_k = \hat{v}_k - v_k$, where $\hat{v}_k$ is the unique solution of the QP subproblem

$$\underset{v}{\text{minimize}} \;\; \mathcal{Q}_k(v\,; y_k^E, \mu_k^R) \quad \text{subject to} \;\; [v]_{\mathcal{A}_\epsilon} = 0, \tag{2.20}$$

where $\mathcal{Q}_k(v\,;y_k^E,\mu_k^R)$ is the quadratic model (1.11). Given a feasible point $\hat{v}_k^{(0)}$ such that

$$[\hat{v}_k^{(0)}]_{\mathcal{A}_\epsilon}=0 \quad \text{and} \quad [\hat{v}_k^{(0)}]_{\mathcal{F}_\epsilon}=[v_k]_{\mathcal{F}_\epsilon}, \tag{2.21}$$

the vector $\hat{v}_k$ is computed in the form $\hat{v}_k^{(0)}+\Delta\hat{v}_k^{(0)}$, where $[\Delta\hat{v}_k^{(0)}]_{\mathcal{A}_\epsilon}=0$ and $[\Delta\hat{v}_k^{(0)}]_{\mathcal{F}_\epsilon}$ satisfy the equations $B_{\mathcal{F}_\epsilon}[\Delta\hat{v}_k^{(0)}]_{\mathcal{F}_\epsilon}=-[\nabla\mathcal{Q}_k(\hat{v}_k^{(0)}\,;y_k^E,\mu_k^R)]_{\mathcal{F}_\epsilon}$. This direction is used in the line search if it satisfies the conditions

$$[v_k+d_k]_i\geq 0,\ i=1:n,\ \ [\nabla\mathcal{Q}_k(v_k+d_k\,;y_k^E,\mu_k^R)]_{\mathcal{A}_\epsilon}\geq -t_k e \ \text{ and } \ \nabla M_k^T d_k<0, \tag{2.22}$$

where t_k is a small positive scalar computed in Algorithm 2. The condition on the ϵ-active components of $\nabla\mathcal{Q}_k$ relaxes the analogous gradient condition on the active components of $\nabla\mathcal{Q}_k$ in (2.4) associated with solving (1.13), and serves to reduce the possibility of unnecessary QP iterations in the neighbourhood of a solution. In the companion paper Gill *et al.* (2014, Theorem 3.1) it is shown that, in the limit, a local descent direction is used at every iteration, and that the iterates are equivalent to those of a conventional stabilized SQP method. These properties facilitate a proof of superlinear convergence under mild assumptions.

If the conditions for computing the local descent step are not satisfied, or the local descent step does not satisfy the conditions in (2.22), then a global descent direction is found by solving the convex QP problem (2.6). If the local descent step has been computed, then the QP solver for the global descent step can reuse the factorization needed to solve the equations for $[\Delta\hat{v}_k^{(0)}]_{\mathcal{F}_\epsilon}$.

2.3.2 *The global descent direction.* The global descent direction is $d_k=\hat{v}_k-v_k$, where $\hat{v}_k=(\hat{x}_k,\hat{y}_k)$ is the unique solution of the strictly convex QP problem (2.6). This 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. Let $\widehat{\mathcal{Q}}_k(v)$ denote the objective of the QP (2.6) defined with parameters y_k^E and μ_k^R. Given an initial feasible point $\hat{v}_k^{(0)}$ for problem (2.6), active-set methods generate a feasible sequence $\{\hat{v}_k^{(j)}\}_{j>0}$ such that $\widehat{\mathcal{Q}}_k(\hat{v}_k^{(j)})\leq\widehat{\mathcal{Q}}_k(\hat{v}_k^{(j-1)})$ and $\hat{v}_k^{(j)}$ minimizes $\widehat{\mathcal{Q}}_k(v)$ on a 'working set' $\mathcal{W}_j$ of the non-negativity constraints. The first working set $\mathcal{W}_0$ is the ϵ-active set $\mathcal{A}_\epsilon(x_k,y_k,\mu_k^R)$, which defines the initial feasible point $\hat{v}_k^{(0)}$ as in (2.21).

2.3.3 *The direction of negative curvature.* The nonzero vector $s_k^{(1)}=(u_k^{(1)},w_k^{(1)})$ computed as a by-product of the computation of ξ_k in Algorithm 1 is a direction of negative curvature for the quadratic model $\mathcal{Q}_k(v\,;y_{k-1}^E,\mu_{k-1}^R)$. In Algorithm 3, $s_k^{(1)}$ is rescaled to provide the direction s_k used in the line search. First, an intermediate direction $s_k^{(2)}$ is computed such that

$$s_k^{(2)}\equiv(u_k^{(2)},w_k^{(2)})=\begin{cases}-(u_k^{(1)},w_k^{(1)}) & \text{if } \nabla M(v_k\,;y_k^E,\mu_k^R)^{\mathrm{T}}s_k^{(1)}>0;\\ \ \ (u_k^{(1)},w_k^{(1)}) & \text{otherwise,}\end{cases} \tag{2.23}$$

which implies that $s_k^{(2)}$ is a nonascent 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 a positive σ_k to give a direction s_k such that: (i) $v_k+s_k+d_k$ is feasible for the non-negativity constraints, and (ii) the primal portions of s_k and d_k have comparable norms. (Only the primal part of s_k is influential in the scaling because the dual part $w_k^{(1)}$ of the base vector $s_k^{(1)}=(u_k^{(1)},w_k^{(1)})$ is a function of the primal part $u_k^{(1)}$, see Section 2.1.) The choice of σ_k not only ensures that s_k and d_k have approximately equal weight in the definition of $v_k+s_k+d_k$ but also gives a sensibly scaled s_k when $\|s_k^{(1)}\|$ is arbitrarily

Algorithm 2 Computation of the primal–dual descent direction d_k

procedure DESCENT_DIRECTION($x_k, y_k, y_k^E, \mu_k^R, J_k, H_k$)
  Constants: $0 < \lambda < \min\{\gamma, 1-\gamma\} < 1$;
  $B = B(x_k, y_k; \mu_k^R)$; Compute a positive-definite $\widehat{B}$ from B;
  $\nabla M_k = \nabla M(x_k, y_k; y_k^E, \mu_k^R)$; $t_k = r(x_k, y_k)^\lambda$;
  $[\hat{v}_k^{(0)}]_{\mathcal{A}_\epsilon} = 0$; $[\hat{v}_k^{(0)}]_{\mathcal{F}_\epsilon} = [v_k]_{\mathcal{F}_\epsilon}$;
  if ($B_{\mathcal{F}_\epsilon}$ is positive definite **and** v_k is a V–O-iterate) **then**
    $[\Delta\hat{v}_k^{(0)}]_{\mathcal{A}_\epsilon} = 0$; Solve $B_{\mathcal{F}_\epsilon}[\Delta\hat{v}_k^{(0)}]_{\mathcal{F}_\epsilon} = -[\nabla\mathcal{Q}_k(\hat{v}_k^{(0)})]_{\mathcal{F}_\epsilon}$; $\hat{v}_k = \hat{v}_k^{(0)} + \Delta\hat{v}_k^{(0)}$;
    $d_k = \hat{v}_k - v_k$;
    if ($v_k + d_k$ is feasible **and** $\nabla M_k^T d_k < 0$ **and** $[\nabla\mathcal{Q}_k(v_k + d_k)]_{\mathcal{A}_\epsilon} \geq -t_k e$) **then**
      return d_k; [local descent direction]
    end if
  end if
  $[\Delta\hat{v}_k^{(0)}]_{\mathcal{A}_\epsilon} = 0$; Solve $\widehat{B}_{\mathcal{F}_\epsilon}[\Delta\hat{v}_k^{(0)}]_{\mathcal{F}_\epsilon} = -[\nabla\widehat{\mathcal{Q}}_k(\hat{v}_k^{(0)})]_{\mathcal{F}_\epsilon}$;
  Compute $\widehat{\alpha}_0 \geq 0$ and $\hat{v}_k^{(1)}$ such that $\hat{v}_k^{(1)} = \hat{v}_k^{(0)} + \widehat{\alpha}_0\Delta\hat{v}_k^{(0)}$ is feasible for the bounds;
  Solve the convex QP (2.6) for $\hat{v}_k$, starting at $\hat{v}_k^{(1)}$;
  $d_k = \hat{v}_k - v_k$; [global descent direction]
  return d_k;
end procedure

large or small. As only the primal variables are subject to non-negativity constraints, the scalar σ_k must be chosen to satisfy

$$[x_k + \sigma_k u_k^{(2)} + p_k]_i = [x_k + \sigma_k u_k^{(2)} + \hat{x}_k - x_k]_i = [\hat{x}_k + \sigma_k u_k^{(2)}]_i \geq 0,$$

where $\hat{x}_k$ is the primal solution of the subproblem (2.6). The value of σ_k is chosen as large as possible subject to the restriction that the resulting primal step $\|\sigma_k u_k^{(2)}\|$ is no greater than the primal step $\|\hat{x}_k - x_k\|$ ($= \|p_k\|$) from the QP subproblem. The curvature estimate ξ_k computed by Algorithm 1 is included in the bound to give an appropriately scaled direction when $\|p_k\|$ is small or zero. These considerations lead to the definition

$$\sigma_k = \underset{\sigma \geq 0}{\operatorname{argmax}} \left\{ \sigma : \hat{x}_k + \sigma u_k^{(2)} \geq 0,\ \|\sigma u_k^{(2)}\| \leq \max(\xi_k, \|\hat{x}_k - x_k\|) \right\}. \tag{2.24}$$

The direction of negative curvature used in the line search is then

$$s_k \equiv (u_k, w_k) = \sigma_k(u_k^{(2)}, w_k^{(2)}) = \sigma_k s_k^{(2)}. \tag{2.25}$$

The computation of the vector s_k is summarized in Algorithm 3.

The direction s_k is zero if no direction of negative curvature exists for $\mathcal{Q}_k$, in which case the value of ξ_k computed by Algorithm 1 is zero. However, the next result considers one important situation in which σ_k is positive and s_k is nonzero.

LEMMA 2.2 If $d_k = 0$ and $\xi_k > 0$ then $\sigma_k > 0$ and $s_k \neq 0$.

Algorithm 3 Feasible direction of negative curvature for the merit function

procedure NEGATIVE_CURVATURE_DIRECTION($s_k^{(1)}, \xi_k, x_k, d_k, J_k, H_k$)

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

Let p_k and $u_k^{(2)}$ be the first n components of d_k and $s_k^{(2)}$;

$\sigma_k = \operatorname{argmax}_{\sigma \geq 0} \left\{ \sigma : x_k + p_k + \sigma u_k^{(2)} \geq 0,\ \|\sigma u_k^{(2)}\| \leq \max(\xi_k, \|p_k\|) \right\}$;

$s_k = \sigma_k s_k^{(2)}$; [scaled curvature direction]

return s_k;

end procedure

Proof. A positive ξ_k in Algorithm 1 implies that the vectors $u_{\mathcal{F}_\epsilon}$, $u_k^{(1)}$ 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 (2.23) can change only 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 $\hat{x}_k$, the solution of the QP subproblem (1.14), is identical to x_k and $\|\hat{x}_k - x_k\|$ is zero. The positive value of ξ_k 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 (2.1), and the fact that $\mu_{k-1}^R > 0$. □

2.4 *The flexible line search and penalty-parameter update*

The flexible line search defined in Algorithm 4 is a generalization of the line search used in the first-order primal–dual method of Gill & Robinson (2013). (The idea of a flexible line search was proposed by Curtis & Nocedal, 2008 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 line-search model function

$$\psi_k(\alpha\,; \mu, \ell_k) = \Psi_k(0\,; \mu) + \alpha \Psi_k'(0\,; \mu) + \tfrac{1}{2}(\ell_k - 1)\alpha^2 \min\left(0, \Delta v_k^T B(v_k\,; \mu_{k-1}^R) \Delta v_k\right), \tag{2.26}$$

where Ψ_k' denotes the derivative with respect to α. The scalar ℓ_k is either 1 or 2, depending on the required order of the line-search model. The value $\ell_k = 1$ implies that Ψ_k is an affine function, which gives a first-order line-search model. The value $\ell_k = 2$ defines a quadratic Ψ_k and second-order line-search model. The choice of line-search model for Algorithm 4 is motivated by superlinear convergence considerations. In particular, it allows the acceptance of a unit step in the neighbourhood of a solution, which is a key result in establishing the asymptotic equivalence between Algorithm 5 and a conventional stabilized SQP method (see Gill *et al.*, 2014, Theorem 3.4).

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 $0 < \gamma_S < 1$. The reduction requirement may be written in terms of the Armijo condition

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

Algorithm 4 Flexible line search

procedure FLEXIBLE_LINE_SEARCH($d_k, s_k, y_k^E, \mu_k, \mu_k^R, \mu_{k-1}^R, \ell_k, J_k, H_k$)
  Constant: $\gamma_S \in (0, \frac{1}{2})$;
  Compute $\nabla M = \nabla M(x_k, y_k\,; y_k^E, \mu_k^R)$;
  if $s_k = 0$ **and** $d_k = 0$ **then**
    $\alpha_k = 1$;
  else if ($d_k \neq 0$ **or** $\nabla M^T s_k < 0$ **or** $\mu_k^R = \mu_{k-1}^R$) **then**
    $\alpha_k = 1$;
    while $\rho_k(\alpha_k\,; \mu_k^R, \ell_k) < \gamma_S$ **and** $\rho_k(\alpha_k\,; \mu_k, \ell_k) < \gamma_S$ **do**
      $\alpha_k = \frac{1}{2}\alpha_k$;
    end while
  else $[d_k = 0, s_k \neq 0, \xi_k > 0]$
    $\xi_k^R = -s_k^{\mathrm{T}} \nabla^2 M(v_k\,; y_k^E, \mu_k^R) s_k / \|u_k\|^2$; [by definition, $s_k = (u_k, w_k)$]
    if $\xi_k^R > \gamma_S \xi_k$ **then**
      $\alpha_k = 1$;
      while $\rho_k(\alpha_k\,; \mu_k^R, \ell_k) < \gamma_S$ **and** $\rho_k(\alpha_k\,; \mu_k, \ell_k) < \gamma_S$ **do**
        $\alpha_k = \frac{1}{2}\alpha_k$;
      end while
    else
      $\alpha_k = 0$;
    end if
  end if
  return $\alpha_k \geq 0$
end procedure

The flexible line search defined in Algorithm 4 requires that α_k satisfies the modified Armijo condition

$$\Psi_k(0\,; \mu_k^F) - \Psi_k(\alpha_k\,; \mu_k^F) \geq \gamma_S \big(\psi_k(0\,; \mu_k^R, \ell_k) - \psi_k(\alpha_k\,; \mu_k^R, \ell_k) \big) \tag{2.27}$$

for some value of μ_k^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 $\rho_k(\alpha_k\,; \mu_k, \ell_k) \geq \gamma_S$ or $\rho_k(\alpha_k\,; \mu_k^R, \ell_k) \geq \gamma_S$, where

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

The Armijo procedure is not executed if $d_k = s_k = 0$, or $d_k = 0$ and the curvature of the merit function $M(v_k\,; y_k^E, \mu_k^R)$ along the vector s_k is not sufficiently large compared with the curvature of the quadratic model. The condition for negligible curvature involves the calculation of the scalar

$$\xi_k^R = -s_k^{\mathrm{T}} \nabla^2 M(v_k\,; y_k^E, \mu_k^R) s_k / \|u_k\|^2 = -u_k^{\mathrm{T}} \big(H(x_k, \hat{y}_k) + (1/\mu_k^R) J(x_k)^T J(x_k) \big) u_k / \|u_k\|^2 \tag{2.28}$$

(cf. (2.11)), where $\hat{y}_k = \pi_k + (\pi_k - y_k)$ and π_k denotes the vector $\pi(x_k\,; y_k^E, \mu_k^R)$. If the Armijo procedure is not executed (i.e., $\alpha_k = 0$, or $\alpha_k = 1$ with $d_k = s_k = 0$) then $v_{k+1} = v_k$. In this case, it must hold that $\mu_k^R < \mu_{k-1}^R$ (see Lemmas 2.3(2) and 2.4(3)). The analysis will show that the property $\lim_{k \to \infty} \mu_k^R = 0$ and the definitions of V-, O- and M-iterates are key in establishing the convergence results for Algorithm 5.

On completion of the line search, the line-search penalty parameter μ_{k+1} is computed and the iteration ends. The choice of μ_{k+1} is motivated by the goal of decreasing the penalty parameter only when the trial step indicates that the merit function has not been sufficiently reduced. In particular, μ_{k+1} is computed as

$$\mu_{k+1} = \begin{cases} \mu_k & \text{if } \rho_k(\alpha_k; \mu_k, \ell_k) \geq \gamma_S, \text{ or } d_k = s_k = 0, \text{ or } \alpha_k = 0, \\ \max\left(\frac{1}{2}\mu_k, \mu_k^R\right) & \text{otherwise.} \end{cases} \tag{2.29}$$

2.5 *The main algorithm*

The stabilized second-order primal–dual SQP algorithm is formally stated as Algorithm 5. The algorithm is terminated if one of two sets of conditions holds. The first is

$$r(x_k, y_k) \leq \tau_{\text{stop}}, \quad \xi_k \leq \tau_{\text{stop}} \quad \text{and} \quad \mu_{k-1}^R \leq \tau_{\text{stop}}, \tag{2.30}$$

where τ_{stop} is a given positive stopping tolerance, $r(x, y)$ is defined in (1.2) and ξ_k is computed in Algorithm 1. If the conditions in (2.30) hold, then (x_k, y_k) is an approximate second-order KKT point for (NP). The motivation for the second set of termination conditions is to identify convergence to an infeasible stationary point (ISP), i.e., a minimizer of

$$\underset{x \in \mathbb{R}^n}{\text{minimize}} \ \ \tfrac{1}{2}\|c(x)\|_2^2 \quad \text{subject to} \quad x \geq 0 \tag{2.31}$$

at which the constraint $c(x) = 0$ is violated. The first-order optimality conditions for (2.31) may be written in the form

$$\left\| \min\left(x, J(x)^T c(x)\right) \right\| = 0, \tag{2.32}$$

which motivates the second set of termination conditions:

$$\min\left(\|c(x_k)\|, \tau_{\text{stop}}\right) > \mu_k^R, \quad \| \min\left(x_k, J(x_k)^T c(x_k)\right) \| \leq \tau_{\text{stop}}, \quad \text{with } k \text{ an M-iteration.} \tag{2.33}$$

If these conditions hold then the constraint violation is bounded below by μ_k^R ($\mu_k^R > 0$) and $v_k = (x_k, y_k)$ is an M-iterate that approximately satisfies (2.32). The requirements that k is an M-iteration and μ_k^R is small relative to $\min\left(\|c(x_k)\|, \tau_{\text{stop}}\right)$ 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 (2.30) and (2.33) constitute the conditions used in steps 10 and 25 of Algorithm 5.

Discussion. Each iteration of Algorithm 5 requires the solution of equations involving the ϵ-free rows and columns of $B_{\mathcal{F}_\epsilon}(v_k\,; \mu_{k-1}^R)$ and $B_{\mathcal{F}_\epsilon}(v_k\,; \mu_k^R)$. The solve with $B_{\mathcal{F}_\epsilon}(v_k\,; \mu_{k-1}^R)$ is used to compute the least-curvature estimate ξ_k in step 8. Solves with $B_{\mathcal{F}_\epsilon}(v_k\,; \mu_k^R)$ are required to compute the local descent step and the solution of the QP subproblem. This implies that if $\mu_k^R \neq \mu_{k-1}^R$, it is necessary to compute the factorizations of two matrices in regularized KKT form at each iteration. If the computation of ξ_k is omitted, and the line search is defined in terms of the descent step d_k only, then one KKT factorization is required at each iteration. Moreover, the analysis of Gill *et al.* (2014) shows that this variant of Algorithm 5 is a globally convergent stabilized SQP method for finding first-order solutions.

Algorithm 5 Second-order primal–dual SQP algorithm

procedure PDSQP2(x_1, y_1)
  Constants: $\{\tau_{\text{stop}}, \gamma_S\} \subset (0, \frac{1}{2})$, $0 < \gamma < 1$ and $0 < \epsilon_a \ll 1$;
  Choose $y_0^E \in \mathbb{R}^m$, $\{\tau_0, \phi_{V,0}^{\max}, \phi_{O,0}^{\max}\} \subset (0, \infty)$ and $0 < \mu_0^R \leq \mu_1 < \infty$;
  $k = 1$;
  loop
    Compute the ϵ-free set $\mathcal{F}_\epsilon(x_k, y_k, \mu_{k-1}^R)$ from (2.3);
    $J_k = J(x_k)$; $H_k = H(x_k, y_k)$;
    $\left(s_k^{(1)}, \xi_k\right)$ = LEAST_CURVATURE_ESTIMATE$(x_k, y_k, \mu_{k-1}^R, J_k, H_k)$; [Algorithm 1]
    Compute $r(x_k, y_k)$ from (1.2);
    if (termination condition (2.30) holds) **then**
      return the approximate second-order KKT point (x_k, y_k);
    end if
    if $(\phi_{\text{V}}(x_k, y_k) \leq \frac{1}{2}\phi_{V,k-1}^{\max})$ **then** [V-iterate]
      $\phi_{V,k}^{\max} = \frac{1}{2}\phi_{V,k-1}^{\max}$; $y_k^E = y_k$; $\tau_k = \tau_{k-1}$;
      Set μ_k^R as in (2.15);
    else if $(\phi_{\text{O}}(x_k, y_k, \xi_k) \leq \frac{1}{2}\phi_{O,k-1}^{\max})$ **then** [O-iterate]
      $\phi_{O,k}^{\max} = \frac{1}{2}\phi_{O,k-1}^{\max}$; $y_k^E = y_k$; $\tau_k = \tau_{k-1}$;
      Set μ_k^R as in (2.15);
    else if $((x_k, y_k)$ satisfies (2.17a–c) **then** [M-iterate]
      Set y_k^E as in (2.18); $\tau_k = \frac{1}{2}\tau_{k-1}$;
      Set μ_k^R as in (2.19);
    else [F-iterate]
      $y_k^E = y_{k-1}^E$; $\tau_k = \tau_{k-1}$; $\mu_k^R = \mu_{k-1}^R$;
    end if
    if (termination condition (2.33) holds) **then**
      exit with the approximate infeasible stationary point x_k.
    end if
    Compute the ϵ-free set $\mathcal{F}_\epsilon(x_k, y_k, \mu_k^R)$ from (2.3);
    d_k = DESCENT_DIRECTION$(x_k, y_k, y_k^E, \mu_k^R, J_k, H_k)$; [Algorithm 2]
    s_k = CURVATURE_DIRECTION$(s_k^{(1)}, \xi_k, x_k, d_k, J_k, H_k)$; [Algorithm 3]
    if $(d_k \neq 0$ **and** $s_k = 0$ **and** (x_k, y_k) is a V–O-iterate) **then**
      $\ell_k = 1$;
    else
      $\ell_k = 2$;
    end if
    $\mu_k = \max(\mu_k^R, \mu_k)$;
    α_k = FLEXIBLE_LINE_SEARCH$(d_k, s_k, y_k^E, \mu_k, \mu_k^R, \mu_{k-1}^R, \ell_k, J_k, H_k)$; [Algorithm 4]
    Set μ_{k+1} as in (2.29);
    $v_{k+1} = (x_{k+1}, y_{k+1}) = v_k + \alpha_k d_k + \alpha_k s_k$;
    $k = k + 1$;
  end loop
end procedure

2.6 *Properties of the main algorithm*

The next lemma establishes some properties associated with the key computational steps of Algorithm 5.

LEMMA 2.3 Let d_k and s_k be vectors computed by Algorithm 5.

(1) If $d_k = 0$ then
   (a) $\min(x_k, g(x_k) - J(x_k)^{\mathrm{T}} y_k) = 0$, and $\pi(x_k, y_k^E, \mu_k^R) = y_k$;
   (b) if the kth iterate is a V–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 $\xi_k = 0$, the kth iterate is not an F-iterate, and $\mu_k^R < \mu_{k-1}^R$.

Proof. For all parts of this lemma it is assumed that $d_k = 0$, in which case it must hold that $\nabla M(v_k; y_k^E, \mu_k^R)^T d_k = 0$. It follows from step 9 of Algorithm 2 that d_k is a global descent step.

For part (1a), as d_k is zero, the optimality conditions for the bound-constrained QP subproblem (2.6) give

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

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

$$y_k = \pi_k \quad \text{and} \quad \min(x_k, g_k - J_k^T y_k) = 0. \tag{2.34}$$

This completes the proof of part (1a).

As d_k is zero and the kth iterate is a V–O-iterate by assumption, it follows from the update for y_k^E given by Algorithm 5, the definition of π_k, and part (1a) that

$$y_k^E = y_k = \pi_k = y_k^E - c(x_k)/\mu_k^R, \tag{2.35}$$

which implies that $c(x_k) = 0$. Combining this with (1.2) and (2.34) gives $r(x_k, y_k) = 0$, which proves part (1b).

If the kth iterate is an M-iterate and $\|y_k\|_\infty \leq y_{\max}$, it follows from the update for y_k^E in (2.18), and the definition of π_k that (2.35) holds. As in part (1b), this proves that $c(x_k) = 0$. Combining this result with (1.2) and (2.34) 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 $\xi_k > 0$. As d_k is zero, x_k is optimal for the QP subproblem (2.6) and $\hat{x}_k = x_k$. In this case, the assumption that $\xi_k > 0$ in the definitions of $u_k^{(2)}$ and its associated scale factor σ_k in (2.24) implies that $\sigma_k > 0$. However, if $\sigma_k > 0$ and $s_k = 0$ in the definition of $s_k^{(2)}$ (2.25) then not only is $s_k^{(2)} = 0$ but also $s_k^{(1)} = 0$ from (2.23). This gives the required contradiction because $s_k^{(1)}$ must be nonzero for ξ_k to be positive in Algorithm 1. It follows that $\xi_k = 0$, 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

$$\mu_k^R = \mu_{k-1}^R \quad \text{and} \quad y_k^E = y_{k-1}^E. \tag{2.36}$$

As the solution of the subproblem is $d_k = 0$ by assumption, part (1a) and the result that $\xi_k = 0$ imply that (x_k, y_k) satisfies the conditions (2.17) for an M-iterate. This is a contradiction because a point is classified as an F-iterate if it is not a V–O-iterate and the conditions for an M-iterate fail to hold.

It remains to show that the regularization parameter is 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. (2.19)) would imply that $\mu_k^R < \mu_{k-1}^R$. It follows that x_k must be a V–O-iterate. The properties of the update (2.15) associated with a V–O-iterate, and the assumption that $\mu_k^R = \mu_{k-1}^R$, imply that $\max\big(r_k, \xi_k\big) > 0$. This is a contradiction because $r_k = 0$ follows from part (1b), and it has been shown above that $\xi_k = 0$. As all possible cases have been exhausted, it holds 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 2.4 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 4 terminate with $\alpha_k > 0$.

(2) If $\mu_k < \mu_{k-1}$ for some $k \geq 1$ then the while loop given by either step 8 or step 15 of Algorithm 4 was executed.

(3) If $\alpha_k = 0$ then the kth iterate is not an F-iterate and $\mu_k^R < \mu_{k-1}^R$.

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

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

First, assume that $\ell_k = 2$ in Algorithm 4. Substituting the definition of the quadratic model from (2.26) 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

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

gives

$$\eta(\alpha) = \alpha(\gamma_S - 1)\Psi_k'(0\,; \mu_k^R) + \tfrac{1}{2}\alpha^2\omega_k - \mathcal{O}\big(|\alpha|^3\big), \tag{2.37}$$

where ω_k is the scalar

$$\omega_k = \gamma_S \min\big(0, \Delta v_k^T B(v_k\,; \mu_{k-1}^R)\Delta v_k\big) - \Psi_k''(0\,; \mu_k^R) \tag{2.38}$$

$$= \gamma_S\big(\min(0, \Delta v_k^T B(v_k\,; \mu_{k-1}^R)\Delta v_k) - \Psi_k''(0\,; \mu_k^R)\big) + (\gamma_S - 1)\Psi_k''(0\,; \mu_k^R). \tag{2.39}$$

Case 1. Consider the computation associated with satisfying the Armijo condition in step 8 in Algorithm 4. If this loop is executed, it must be the case that the direction $\Delta v_k = s_k + d_k$ is nonzero, and at least one of the conditions $d_k \neq 0$, $\nabla M(v_k\,; y_k^E, \mu_k^R)^T s_k < 0$ 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 $\nabla M(v_k\,; y_k^E, \mu_k^R)^T s_k < 0$. It follows that

$$\Psi_k'(0\,; \mu_k^R) = \nabla M(v_k\,; y_k^E, \mu_k^R)^T \Delta v_k = \nabla M(v_k\,; y_k^E, \mu_k^R)^T (d_k + s_k) < 0, \tag{2.40}$$

because $\nabla M(v_k\,; y_k^E, \mu_k^R)^T s_k \leq 0$ by construction (see steps 2 and 5 of Algorithm 5), and either d_k is the solution of the strictly convex problem (2.6) if it is a global descent step or d_k satisfies $\nabla M(v_k\,; y_k^E, \mu_k^R)^T d_k < 0$, which is a condition for the acceptance of a local descent step (see step 9 in Algorithm 2). As $\gamma_S \in (0, 1)$, it follows from (2.37) and (2.40) 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 Lemma 2.3(1a) 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

$$\Psi_k''(0\,; \mu_k^R) = s_k^T \nabla^2 M(v_k\,; y_k^E, \mu_k^R) s_k = s_k^T B(v_k\,; \mu_{k-1}^R) s_k. \tag{2.41}$$

As d_k is zero, the vector s_k must be nonzero. It then follows from Lemma 2.1(2), (2.23) and (2.25) that $s_k^T B(v_k\,; \mu_{k-1}^R) s_k < 0$. Combining this result with the two identities (2.41) and $\Delta v_k = s_k$, the definition of ω_k in (2.39) 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 (2.37) imply 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

$$d_k = 0, \quad s_k \neq 0, \quad \nabla M(v_k\,; y_k^E, \mu_k^R)^T s_k = 0 \quad \text{and} \quad \xi_k^R > \gamma_S \xi_k > 0, \tag{2.42}$$

where ξ_k^R is the scalar defined in (2.28). As a consequence, the identity $\Psi_k'(0\,; \mu_k^R) = 0$ holds, and the expression (2.37) for $\eta(\alpha)$ may be written in the form

$$\eta(\alpha) = \tfrac{1}{2}\alpha^2 \omega_k - \mathcal{O}\big(|\alpha|^3\big). \tag{2.43}$$

Combining the conditions (2.42), the values $(u_k^{(1)}, w_k^{(1)}, \xi_k)$ returned by Algorithm 1, the curvature expression (2.11) and the expressions (2.23) and (2.25) defining the direction of negative curvature, gives

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

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 (2.11), and the definition of $u_k^{(1)}$ in Algorithm 1, shows that the scalar ω_k is positive in (2.38). This result implies that the function $\eta(\alpha)$ of (2.43) satisfies $\eta(\alpha) > 0$ for all $\alpha > 0$ sufficiently small. This completes the proof of part (1) when $\ell_k = 2$.

Now assume that $\ell_k = 1$. As a consequence, Algorithm 4 gives $d_k \neq 0$ and $s_k = 0$, which implies that (2.40) holds, as above. The same argument used for Subcase 1 of Case 1 may be repeated, which proves part (1) for the case $\ell_k = 1$.

Part (2) follows directly from the definition of the penalty-parameter update (2.29) and the structure of Algorithm 4.

For part (3), if $\alpha_k = 0$ then it follows from the **elseif** statement in step 6 of Algorithm 4 that $\mu_k^R \neq \mu_{k-1}^R$. It follows that the kth iterate cannot be an F-iterate and $\mu_k^R < \mu_{k-1}^R$. □

The principal focus of the next section is the global convergence of Algorithm 5 under the assumption that the finite termination conditions given by steps 10 and 25 are omitted. 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 3.8, which establishes that Algorithm 5 will terminate after a finite number of iterations if the algorithm includes the termination condition (2.30) for an approximate KKT point and condition (2.33) for an approximate infeasible stationary point.

3. Global convergence

The following four assumptions are made about the iterates, the properties of the problem functions and Algorithm 5. Unless there is an explicit statement to the contrary, every result given in this section requires that Assumptions 3.1–3.4 hold.

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

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

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.

ASSUMPTION 3.4 The termination conditions of steps 10 and 25 are omitted from Algorithm 5.

The global analysis proceeds as follows. First, certain properties of the iterates are established under Assumptions 3.1–3.4. Second, these properties are used to show that without Assumption 3.4, finite termination of Algorithm 5 occurs for any positive stopping tolerance. This result does not require the assumption of a constraint qualification. Finally, under the assumption of a weak constraint qualification, the properties of the limit points of an infinite sequence of iterates are investigated.

The next theorem extends the result by Gill & Robinson (2013, Theorem 3.1) to allow for the computation of directions of negative curvature. The proof gives properties of the iterates that must hold when every iterate is an F-iterate for sufficiently large k. These properties will be used to show that there must exist an infinite set of V–O-iterates or M-iterates.

THEOREM 3.5 Let $\{v_k\} = \{(x_k, y_k)\}$ denote the sequence of iterates generated by Algorithm 5. 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 constants τ, μ^R and μ, and a constant vector y^E such that

$$\tau_k = \tau, \quad \mu_k^R = \mu^R, \quad \mu_k = \mu \quad \text{and} \quad y_k^E = y^E \ \text{ for all } k \in \mathcal{S}. \tag{3.1}$$

(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 $\{\xi_k\}$, $\{s_k\}$, $\{\Delta v_k\}$ $(= \{d_k + s_k\})$, $\{B(v_k\,; \mu^R)\}$, $\{\nabla^2 M(v_k\,; y^E, \mu^R)\}$, $\{\Delta v_k^{\mathrm{T}} B(v_k\,; \mu^R)\Delta v_k\}$ and $\{\Delta v_k^{\mathrm{T}} \nabla^2 M(v_k\,; y_k^E, \mu^R)\Delta v_k\}$ 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 $\xi_k > \tau$ for all $k \in \mathcal{S}'$ sufficiently large.

(6) There exists a positive ϵ_F such that

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

Proof. First note that since k is assumed to be an F-iterate for all $k \geq \widehat{k}$, it may be assumed without loss of generality that every descent direction d_k is of a global descent step. This is permitted because a requirement for computing a local descent step is that iteration k is a V–O-iterate.

Part (1) follows from the definition of $\mathcal{S}$, the structure of Algorithm 5 and (2.29).

Part (2) follows from the statement and proof of Gill & Robinson (2013, Theorem 3.1).

The first step in the proof of part (3) is to show that the set $\{\xi_k\}_{k\in\mathcal{S}}$ is uniformly bounded. From the definition of ξ_k in Algorithm 1, it must hold that

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

From 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, implies the required uniform boundedness of $\{\xi_k\}_{k\in\mathcal{S}}$.

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 constitute $s_k = (u_k, w_k)$ are bounded. The result for $\{u_k\}_{k\in\mathcal{S}}$ follows from the uniform boundedness of the sequences $\{\xi_k\}_{k\in\mathcal{S}}$ and $\{d_k\}_{k\in\mathcal{S}}$ shown in part (2), and the definition of u_k from (2.25) and (2.24). For the sequence $\{w_k\}_{k\in\mathcal{S}}$, the expressions (2.23), (2.24) and (2.25) 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 and 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 $\{B(v_k\,;\mu^R)\}_{k\in\mathcal{S}}$ follows from their respective definitions (1.10) and (1.12), Assumptions 3.2 and 3.3, and part (2) of this theorem. This result and the boundedness of $\{d_k\}_{\epsilon\mathcal{S}}$, $\{s_k\}_{\epsilon\mathcal{S}}$ imply 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 proved 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 $\hat{v}_k$ of the bound-constrained QP subproblem (2.6) can be written in terms of the vector $d_k=\hat{v}_k-v_k=(\hat{x}_k-x_k,\hat{y}_k-y_k)$. 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

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

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 (2.17a) and (2.17b) must be satisfied for $k\in\mathcal{S}_1$ sufficiently large. If there existed a subsequence of $\mathcal{S}_1$ for which $\xi_k\le\tau$ then eventually condition (2.17c) would also be satisfied, which would violate the assumption that all iterates are F-iterates for $k\ge\widehat{k}$. As a consequence, it may be inferred that $\xi_k>\tau$ for all $k\in\mathcal{S}_1$ sufficiently large.

If $\xi_k>\tau>0$ for all $k\in\mathcal{S}_1$ sufficiently large then Algorithm 1 will compute a nonzero $u_k^{(1)}$ for all $k\in\mathcal{S}_1$ sufficiently large. As $u_k^{(2)}$ is $\pm u_k^{(1)}$ from (2.23), $u_k^{(2)}$ is also nonzero for the same values of k. Let ϵ_k denote the value of ϵ that defines the ϵ-active set (2.1) at (x_k,y_k). From this definition of ϵ_k, it must hold that $\epsilon_k\ge\min(\mu_{k-1}^R,\epsilon_a)=\min(\mu^R,\epsilon_a)>0$. 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 $[\hat{x}_k]_j\ge\frac{1}{2}\epsilon_k\ge\frac{1}{2}\min(\mu^R,\epsilon_a)$ must hold for all $k\in\mathcal{S}_1$ sufficiently large. If this lower bound is combined with the inequality $\xi_k>\tau>0$ 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 (2.24) that there must exist some positive δ_1 such that $\|u_k\|=\|\sigma_k u_k^{(2)}\|\ge\delta_1$ 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 $U_k=\begin{pmatrix} I & -(1/\mu^R)J_k^{\mathrm{T}} \\ 0 & I \end{pmatrix}$. As $d=0$ is feasible for (2.6) and $d_k=(p_k,q_k)=(\hat{x}_k-x_k,\hat{y}_k-y_k)$ is computed from the unique solution $\hat{v}_k=(\hat{x}_k,\hat{y}_k)$ of the strictly convex QP (2.6), it follows that

$$\begin{aligned} -\nabla M(v_k\,;y^E,\mu^R)^T d_k &\ge \tfrac{1}{2} d_k^T \widehat{B}(v_k\,;\mu^R)d_k \\ &= \tfrac{1}{2} d_k^T U_k^{-1} U_k \widehat{B}(v_k\,;\mu^R) U_k^{\mathrm{T}} U_k^{-\mathrm{T}} d_k \end{aligned}$$

$$= \frac{1}{2}\begin{pmatrix} p_k \\ q_k + \frac{1}{\mu^R} J_k p_k \end{pmatrix}^T \begin{pmatrix} \widehat{H}_k + \frac{1}{\mu^R} J_k^{\mathrm{T}} J_k & 0 \\ 0 & \mu^R \end{pmatrix} \begin{pmatrix} p_k \\ q_k + \frac{1}{\mu^R} J_k p_k \end{pmatrix} \quad \text{(cf. (2.7))}$$

$$\geq \tfrac{1}{2}\big(\underline{\lambda}_{\min}\|p_k\|^2 + \mu^R\|q_k + (1/\mu^R)J_k p_k\|^2\big),$$

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 $\lim_{k\in\mathcal{S}'} q_k = 0$ because the sequence $\{J_k\}$ is uniformly bounded by Assumptions 3.2 and 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 (1.9) 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 and 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 $\xi_k \leq \tau$ then some subsequent iterate is an M-iterate, which would be a contradiction. This implies that $\xi_k > \tau$ for all $k \in \mathcal{S}'$ 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

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

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 (2.25) and $s_k^{(2)}$ (2.23), and Lemma 2.1(2) yields

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

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 $\lambda_{\min}(H_{\mathcal{F}_\epsilon} + (1/\mu^R)J_{\mathcal{F}_\epsilon}^T J_{\mathcal{F}_\epsilon})$ converges to zero on some subsequence of $\mathcal{S}_1$, which would imply that ξ_k 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 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 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 3.6 Consider the infinite sequence $\{v_k\}$, where $v_k = (x_k, y_k)$ is generated by Algorithm 5. Then

(1) the union of the index sets of V–O- and M-iterates is infinite; and

(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 Theorem 3.5(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 and 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 Theorem 3.5(1) gives $\mu_k^R = \mu^R$ Algorithm 5 gives $\ell_k = 2$, and Algorithm 2 gives d_k is a global descent step for every $k \in \mathcal{S}$. The inequality (3.11) is established by considering the Armijo acceptance condition (2.27) associated with the flexible line search described in Section 2.4. Suppose that the values $\mu^F = \mu^R$ and $\ell_k = 2$ are used in condition (2.27), 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

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

where $\Psi_k(\alpha\,; \mu) = M(v_k + \alpha\Delta v_k\,; y_k^E, \mu)$, $\Delta v_k = d_k + s_k$ and $\psi_k(\alpha\,; \mu^R, \ell_k)$ is the quadratic model (2.26). Expanding the left-hand side of (3.12) using a Taylor-series expansion and performing some rearrangement yields

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

There are two cases to consider.

Case 1. $\Psi_k'(0\,;\mu^R) = \nabla M(v_k\,;y^E,\mu^R)^T\Delta v_k \le -\delta$ for some $\delta > 0$ and all k large.
It follows from Theorem 3.5(2),(3) that $\{\Psi_k''(0\,;\mu^R)\}_{k\in\mathcal{S}}$, $\{\psi_k''(0\,;\mu^R,\ell_k)\}_{k\in\mathcal{S}}$, $\{d_k\}_{k\in\mathcal{S}}$ and $\{s_k\}_{k\in\mathcal{S}}$ are uniformly bounded. These results together with the assumption that $\Psi_k'(0\,;\mu^R) \le -\delta$ imply the existence of a positive $\widehat{\alpha}$ such that (3.13) is satisfied for all $0 < \alpha \le \widehat{\alpha}$ and all $k \in \mathcal{S}$ sufficiently large. Thus, a conventional backtracking line search would terminate with an $\alpha_k \ge \alpha_{\min}$ for some $\alpha_{\min} > 0$ and all k sufficiently large. However, the use of a value of μ^F ($\mu^F \ge \mu^R$) in the flexible line search allows for the early termination of the backtracking loop, and it must hold that the resulting value must satisfy $\alpha_k \ge \alpha_{\min}$. The combination of Theorem 3.5(1), the acceptance criterion (3.12), the definition of μ_k implied by (2.29), the assumption that $\Psi_k'(0\,;\mu^R) \le -\delta$ and the bound $\alpha_k \ge \alpha_{\min}$, yields the inequality

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

If $\mathcal{S}_1$ is the set of k sufficiently large such that $\Psi_k'(0\,;\mu^R) \le -\delta$ 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 5 provides directions s_k and d_k that satisfy the inequalities $\nabla M(v_k\,;y^E,\mu^R)^{\mathrm{T}}s_k \le 0$ and $\nabla M(v_k\,;y^E,\mu^R)^{\mathrm{T}}d_k \le 0$. Under the given assumption that $\lim_{k\in\mathcal{S}_1}\Psi_k'(0\,;\mu^R) = 0$, it must hold that

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

The combination of this result with Theorem 3.5(6) indicates that there must exist an $\epsilon_F > 0$ such that

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

This inequality, in combination with (3.14) and Theorem 3.5(3),(5), implies that

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

As $\Psi_k'(0\,;\mu^R) \le 0$ 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

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

However, the definitions of ψ_k and Ψ_k, and Theorem 3.5(1)–(3),(5) imply that

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

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

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

Theorem 3.5(2),(3) 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 given 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), Theorem 3.5(1), the line-search penalty-parameter update (2.29), the descent property $\Psi_k'(0\,;\mu^R) \leq 0$, the bound $\alpha_k \geq \alpha_{\min}$ and the uniform bound $\Delta v_k^T B(v_k\,;\mu^R)\Delta v_k \leq -\frac{1}{2}\epsilon_F$ of (3.15) gives

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

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

For the proof of part (2) it is necessary to show that the sequence of regularization parameters satisfies $\{\mu_k^R\} \to 0$. The update rule (2.19) decreases μ_k^R by a fraction for M-iterations and μ_k^R remains unchanged during F-iterations. It follows from these properties and part (1) that if the number of V–O iterations were finite then the result of part (2) would hold. Therefore, for the remainder of the proof it is assumed that the number of V–O iterations 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. As the set $\mathcal{S}_2$ is infinite, it follows from the bound $\phi_{\mathrm{V}}(v_k) \leq \frac{1}{2}\phi_{V,k-1}^{\max}$ of (2.14) and the associated value of $\phi_{V,k}^{\max}$ in Algorithm 5 that $\lim_{k\in\mathcal{S}}\phi_{\mathrm{V}}(x_k,y_k) = 0$. The definition of ϕ_{V} then implies that

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

These identities imply that the first-order condition (1.2) holds for all $k \in \mathcal{S}_2$ sufficiently large, giving $\lim_{k\in\mathcal{S}} r(x_k,y_k) = 0$. This will force the regularization parameter update (2.15) to give $\{\mu_k^R\} \to 0$, which completes the proof. □

The previous result establishes that the union of V–O and M-iterations is infinite. The next result establishes that the number of V–O iterations is finite only when there is an infeasible stationary limit point associated with the constraint violation.

THEOREM 3.7 If the set of V–O iterates is finite, then the set 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

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

Proof. The assumptions of this theorem and Theorem 3.6(1) 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-iterates for k sufficiently large. Let x_* be any limit point of $\{x_k\}_{k\in\mathcal{M}}$, which must exist as a consequence of Assumption 3.3. It must be the case that

$$\lim_{k\in\mathcal{M}_1} x_k = x_* \quad \text{for some } \mathcal{M}_1 \subseteq \mathcal{M}. \tag{3.17}$$

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

$$\lim_{k\in\mathcal{M}_2} y_{k-1}^E = y_*^E \quad \text{for some } \mathcal{M}_2 \subseteq \mathcal{M}_1. \tag{3.18}$$

The next part of the proof involves showing that the point x_* solves problem (2.31). The proof involves the limits

$$\lim_{k\in\mathcal{M}_2} \left\| \min\left(x_k,\ g_k - J_k^T y_{k-1}^E + \frac{1}{\mu_{k-1}^R} J_k^T c_k\right)\right\| = 0, \tag{3.19a}$$

$$\lim_{k\in\mathcal{M}_2} \|\pi_k - y_k\| = 0 \ \text{ and} \tag{3.19b}$$

$$\lim_{k\in\mathcal{M}_2} \xi_k = 0, \tag{3.19c}$$

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 (2.17a-c), and the fact that $\{\tau_k\} \to 0$ is enforced by Algorithm 5 when there are infinitely many M-iterates.

Theorem 3.6(2) implies that $\{\mu_{k-1}^R\} \to 0$ and the denominator of the second term in the minimization (3.19a) 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 (3.17) and $\lim_{k\in\mathcal{M}_2} y_{k-1}^E = y_*^E$ from (3.18) imply that for any given positive δ, there is no infinite subsequence of $\mathcal{M}_2$ such that $[J_k^{\mathrm{T}} c_k]_i \le -\delta$. 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 (3.19a) imply $[J(x_*)^{\mathrm{T}} c(x_*)]_i = 0$ 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 5 is non-negative yields

$$\min\left(x_*, J(x_*)^T c(x_*)\right) = 0, \tag{3.20}$$

i.e., x_* is a first-order solution to the feasibility problem (2.31) (cf. (2.32)).

It remains to show that $c(x_*) \neq 0$. The proof is by contradiction. If $c(x_*) = 0$ then it follows from (3.19), and the definitions of the feasibility and optimality measures (2.13), that x_k must be a V–O iterate for all $k \in \mathcal{M}_2$ sufficiently large, which contradicts the assumption that the number of V–O iterates is finite. It follows that $c(x_*) \neq 0$, as required. □

The previous results are now used to show that Algorithm 5 terminates in a finite number of iterations when the finite termination conditions are present. Note that this theorem does not require the assumption of a constraint qualification because the proof uses the properties of the sequence of iterates and not the properties of the limit points of the sequence.

THEOREM 3.8 If Algorithm 5 is implemented with a positive value of τ_{stop} in the termination conditions (2.30) and (2.33) then the algorithm will terminate in a finite number of iterations.

Proof. The proof is by contradiction. Assume that finite termination does not occur. There are two cases to be considered.

Suppose that there are infinitely many V–O iterates given by the sequence $\mathcal{S}$. It then follows from (2.14), the updates used for V–O iterates in Algorithm 5, (2.13) and Theorem 3.6(2) that (2.30) will be satisfied for all $k \in \mathcal{S}$ sufficiently large. This implies that Algorithm 5 terminates at step 10 after a finite number of iterations. This contradiction completes the proof in this case.

Now consider the case that the set of V–O iterates is finite. It follows from Assumption 3.3 and Theorem 3.7 that there must exist a subsequence $\mathcal{M}$ of M-iterates such that

$$\lim_{k\in\mathcal{M}} x_k = x_*, \quad \left\|\min(x_*, J(x_*)^T c(x_*))\right\| = 0 \quad \text{and} \quad c(x_*) \neq 0. \tag{3.21}$$

It also follows from Theorem 3.6(2), (3.21) and (2.13) that

$$\min(\eta(x_k), \tau_{\text{stop}}) > \mu_k^R \quad \text{for all } k \in \mathcal{M} \text{ sufficiently large.} \tag{3.22}$$

The combination of this bound with the limits (3.21) and the definition of $\mathcal{M}$ implies that (2.33) must be satisfied for all $k \in \mathcal{M}$ sufficiently large, i.e., Algorithm 5 terminates in a finite number of iterations at step 25. This contradiction completes the proof. □

The analysis of convergence to second-order solutions involves a number of alternative constraint qualifications. Constraint qualifications are crucial in this analysis because the properties of *limit points* of the iterates are considered. (Compare the result to Theorem 3.8, in which finite termination is established without the need for a constraint qualification when $\tau_{\text{stop}} > 0$.) The definitions and properties are reviewed below and involve the sets $\{\nabla c_i(x) : i = 1:m\}$ and $\{e_j : j = 1:n\}$ consisting of the gradients of the equality constraints $c(x) = 0$ and inequalities $x \geq 0$.

The next two definitions are required for the formulation of the constant positive generator constraint qualification of Andreani *et al.* (2012, Definition 3.1), which is used in the proof of convergence to first-order KKT points.

DEFINITION 3.9 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 $\sum_{i\in\mathcal{I}} \alpha_i \nabla c_i(x) + \sum_{j\in\mathcal{J}} \beta_j e_j$, with $\beta_j \geq 0$ for all $j \in \mathcal{J}$.

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

(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 3.10 (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 neighbourhood $\mathcal{B}(x)$ of x such that the following two properties hold: (i) the vectors $\{\{\nabla c_i(x)\}_{i\in\mathcal{I}'}, \{e_i\}_{i\in\mathcal{J}}\}$ are positively linearly independent, with positive linear span equal to the positive linear span of the set $\left\{\{\nabla c_i(x)\}_{i\in\mathcal{I}},\ \{e_i\}_{i\in\mathcal{A}(x)}\right\}$, and (ii) for every $\bar{x} \in \mathcal{B}(x)$, any vector in the positive linear span of $\left\{\{\nabla c_i(\bar{x})\}_{i\in\mathcal{I}},\ \{e_i\}_{i\in\mathcal{A}(x)}\right\}$ is in the positive linear span of the vectors $\left\{\{\nabla c_i(\bar{x})\}_{i\in\mathcal{I}'},\ \{e_i\}_{i\in\mathcal{J}}\right\}$.

A constraint qualification in common use is the Mangasarian–Fromovitz constraint qualification (see Mangasarian & Fromovitz, 1967; Mangasarian, 1994).

DEFINITION 3.11 (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 $\left\{\{\nabla c_i(x)\}_{i=1}^m,\ \{e_i\}_{i\in\mathcal{A}(x)}\right\}$ are positively linearly independent.

The Mangasarian–Fromovitz constraint qualification is a stronger condition than the constant positive generator constraint qualification of Definition 3.10. It is not, however, a second-order constraint qualification. The *weak constant rank condition* defined below was introduced by Andreani *et al.* (2007, p. 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 3.12 (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 (2.2)). The *weak constant rank condition (WCRC)* holds at x if there exists a neighbourhood $\mathcal{B}(x)$ 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 & Wei (2000, Definition 2.5) is established by Kungurtsev (2013, Result 8.5.1).

DEFINITION 3.13 (Approximate KKT sequence) The sequence $\{(x_k, y_k)\}$ with each x_k non-negative is an *approximate-KKT (AKKT) sequence* if $\{x_k\} \to x_*$ for some x_* and $\{r(x_k, y_k)\} \to 0$, where r is the norm of the residual (1.2) 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 3.14 Suppose that $\{(x_k, y_k)\}$ is an AKKT sequence with $\{x_k\} \to x_*$ for some 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 dual vectors $\mathcal{Y}(x_*)$ of (1.3) 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.* (2012, Theorem 3.3). The proof of part (2) is a modification of the proof that MFCQ is equivalent to the boundedness of the multiplier set. For a complete proof see Gill *et al.* (2013, Theorem 3.10). □

A second-order KKT point is defined as a point that satisfies a set of second-order necessary optimality conditions for problem (NP). However, there are a number of alternative second-order conditions that may be used. These conditions vary in the definition of the set of directions on which the Hessian of the Lagrangian is required to be positive semidefinite. A standard definition is similar to the second-order sufficient conditions given in Definition 1.2, except that the quadratic form is required only to be non-negative. However, for this condition to be verified, it is necessary to find the minimum of a quadratic form

over a cone, which is a computationally intractable problem (see, e.g., Cottle *et al.*, 1970; Majthay, 1971). For computational purposes, we use an alternative second-order condition, which is sometimes referred to as the weak reduced positive semidefiniteness property (see, e.g., Facchinei & Lucidi, 1998; Moguerza & Prieto, 2003; Andreani *et al.*, 2008). The proofs use the following definition of a second-order KKT point (see, e.g., Andreani *et al.*, 2010a, p. 211).

DEFINITION 3.15 (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. (1.2)), and

$$p^T H(x^*, y^*)p \geq 0 \quad \text{for all } p \in \mathcal{C}_{\mathcal{A}}(x^*), \tag{3.23}$$

where $\mathcal{C}_{\mathcal{A}}(x)$ is the set of directions

$$\mathcal{C}_{\mathcal{A}}(x) = \{p : J(x)p = 0,\ p_i = 0 \ \text{ for } \ i \in \mathcal{A}(x)\}. \tag{3.24}$$

The next result establishes global convergence to second-order KKT points. In particular, it is shown that if there are infinitely many V–O iterations (Theorem 3.7 shows this occurs whenever infeasible stationary points are avoided), 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 that sequence is a second-order KKT point. To the best of our knowledge at the time of writing, the Mangasarian-Fromovitz constraint qualification and the weak constant rank condition together represent the weakest second-order constraint qualification used as part of a convergence analysis (see Andreani *et al.*, 2010b). Consider the simple example of the nonlinear constraint $(x_1 + 1)^2 - x_2 - 1 = 0$, with bounds $x_1, x_2, x_3 \geq 0$. At the feasible point $x = (0, 0, 1)$, the Mangasarian–Fromovitz constraint qualification and weak constant rank condition hold, but the linear independence constraint qualification does not.

THEOREM 3.16 Let $\{v_k\} = \{(x_k, y_k)\}$ denote the sequence of primal–dual iterates generated by Algorithm 5. Assume that the algorithm generates infinitely many V–O iterates, i.e., $|\mathcal{S}| = \infty$, where $\mathcal{S}$ is the index set $\mathcal{S} = \{k : \text{iteration } k \text{ is a V–O iterate}\}$.

(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_* is 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 3.15).

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

$$\lim_{k \in \mathcal{S}_1} x_k = x_*. \tag{3.25}$$

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 5 and the existence of the limit (3.25) that

$$\{(x_k, y_k)\}_{k\in\mathcal{S}_1} \text{ is an AKKT sequence with } \lim_{k\in\mathcal{S}_1} x_k = x_*. \tag{3.26}$$

If this result is combined with the non-negativity of x_k imposed by Algorithm 5 and the result of Theorem 3.14(1), 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 Theorem 3.14(2) 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 p_* be any vector such that $\|p_*\| = 1$ and $p_* \in \mathcal{C}_{\mathcal{A}}(x_*)$, where $\mathcal{C}_{\mathcal{A}}$ is defined by (3.24). As the MFCQ and the WCRC hold at x_*, it follows from Andreani *et al.* (2007, Lemma 3.1) that there exists a sequence $\{p_k\}$ such that

$$\lim_{k\in\mathcal{S}_1} p_k = p_*, \quad \text{with } \; p_k \in \widetilde{C}(x_k) = \{p : J(x_k)p = 0 \;\text{ and }\; p_i = 0 \text{ for } i \in \mathcal{A}(x_*)\}.$$

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

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

As $\{x_k\}_{k\in\mathcal{S}_1}$ is an AKKT sequence from (3.26), and $\{\mu^R_k\} \to 0$ from Theorem 3.6(2), the definition of an ϵ-active set in (2.1) implies that, for all $k \in \mathcal{S}_1$ sufficiently large,

$$\mathcal{A}_\epsilon(x_k, y_k, \mu^R_{k-1}) \subseteq \mathcal{A}(x_*) \quad \text{and} \quad \widetilde{C}(x_k) \subseteq \widehat{C}(x_k, y_k). \tag{3.27}$$

The definition of the set $\mathcal{S}_1$ in (3.25), and the updates made to $\phi_{\mathrm{V}}(x_k, y_k)$ and $\phi_{\mathrm{O}}(x_k, y_k, \xi_k)$ in a V–O iterate of Algorithm 5, imply that the optimality measure $\omega(x, y.\xi)$ of (2.13) satisfies $\omega(x_k, y_k, \xi_k) \to 0$ on $\mathcal{S}_1$. This implies that

$$\lim_{k\in\mathcal{S}_1} \xi_k = 0. \tag{3.28}$$

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

$$\begin{aligned} p_k^T H(x_k, y_k) p_k &= p_k^T \left(H(x_k, y_k) + (1/\mu^R_{k-1}) J_k^T J_k\right) p_k \\ &= p_{\mathcal{F}_\epsilon}^T \left(H_{\mathcal{F}_\epsilon} + (1/\mu^R_{k-1}) J_{\mathcal{F}_\epsilon}^T J_{\mathcal{F}_\epsilon}\right) p_{\mathcal{F}_\epsilon} \end{aligned}$$

$$\geq \lambda_{\min}\Big(H_{\mathcal{F}_\epsilon} + (1/\mu_{k-1}^R)J_{\mathcal{F}_\epsilon}^T J_{\mathcal{F}_\epsilon}\Big)\|p_{\mathcal{F}_\epsilon}\|^2$$
$$\geq \frac{u_{\mathcal{F}_\epsilon}^T\big[H_{\mathcal{F}_\epsilon} + (1/\mu_{k-1}^R)J_{\mathcal{F}_\epsilon}^T J_{\mathcal{F}_\epsilon}\big]u_{\mathcal{F}_\epsilon}}{\theta\|u_{\mathcal{F}_\epsilon}\|^2} = -\frac{1}{\theta}\xi_k, \tag{3.29}$$

where the last inequality follows from the definition of ξ_k in Algorithm 1 and the inequality (2.9).

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 (3.29) over $k \in \mathcal{S}_2 \subseteq \mathcal{S}_1$ and using (3.28) gives

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

which completes the proof. □

Discussion. The termination conditions (2.30) and (2.33) used in Algorithm 5 are precisely the conditions that define an AKKT sequence. For a positive, but arbitrarily small, termination tolerance τ_{stop}, the algorithm generates an AKKT sequence for either problem (NP) or the infeasibility problem (2.31). As a consequence, Theorem 3.8 implies that the algorithm terminates finitely. Analogous to a set-theoretic analysis of constraint qualifications (i.e., some constraint qualifications are stronger than others), there exist many alternative sets of stopping criteria, with some stronger than others. For example, Andreani *et al.* (2010b) define the complementary approximate KKT sequence (CAKKT), which defines the strongest sequential condition in the literature—i.e., there exist convergent sequences with limit points that are not local minimizers that satisfy AKKT and other sequential conditions, but do not satisfy CAKKT (see the references in Andreani *et al.*, 2010b). In this sense the conditions that define a CAKKT sequence provide for a stronger stopping criterion. Algorithm 5 may be adapted to generate CAKKT sequences by modifying the measures ϕ_{V} and ϕ_{O} of (2.13) so that $\eta(x_k) = \|c(x_k)\|$ and

$$\omega(x_k, y_k, \xi_k) = \max\big(\big\|\min(x_k, g(x_k) - J(x_k)^T y_k)\big\|,\ \|x_k \cdot (g(x_k) - J(x_k)^T y_k)\|,\ \|c(x_k) \cdot y_k\|,\ \xi_k\big),$$

and by changing the conditions that characterize an M-iterate to

$$\begin{aligned}
\|\min(x_k, \nabla_x M(x_k, y_k; y_{k-1}^E, \mu_{k-1}^R))\| &\leq \tau_{k-1},\\
\|x_k \cdot \nabla_x M(x_k, y_k; y_{k-1}^E, \mu_{k-1}^R)\| &\leq \tau_{k-1},\\
\|\nabla_y M(x_k, y_k; y_{k-1}^E, \mu_{k-1}^R)\| &\leq \tau_{k-1}\mu_{k-1}^R \quad \text{and}\\
\xi_k &\leq \tau_{k-1}.
\end{aligned}$$

If Algorithm 5 is modified to reflect these changes, it can be shown that either a CAKKT sequence is obtained for problem (NP) or an AKKT sequence is obtained for the feasibility problem (2.31).

4. Conclusions

A stabilized SQP method has been proposed that uses a primal–dual augmented Lagrangian merit function to ensure convergence from an arbitrary starting point. The method has the same strong first- and second-order convergence properties that have been established for augmented Lagrangian methods, while being able to transition seamlessly to stabilized SQP with fast local convergence in the neighbourhood of a

solution. The method uses a flexible line search along a direction formed from an approximate solution of a strictly convex QP subproblem and, when one exists, a direction of negative curvature for the primal–dual merit function. The superlinear convergence of the iterates and the formal local equivalence to stabilized SQP is established in a companion paper (see Gill *et al.*, 2014). It is not necessary to solve a nonconvex QP subproblem, and no assumptions are necessary about the quality of each subproblem solution. When certain conditions hold, an approximate QP solution is computed by solving a single linear system defined in terms of an estimate of the optimal active set. These conditions may be satisfied at any iterate, but are most likely to be satisfied in the neighbourhood of a solution. The conditions exploit the formal equivalence between a stabilized SQP subproblem and a bound-constrained QP associated with minimizing a quadratic model of the merit function.

Convergence to first-order KKT points is established under weaker conditions than those assumed in Gill & Robinson (2013). It is shown that with an appropriate choice of termination condition, the method terminates in a finite number of iterations without the assumption of a constraint qualification. The method may be interpreted as an SQP method with an augmented Lagrangian safeguarding strategy. This safeguarding becomes relevant only when the iterates are converging to an infeasible stationary point of the norm of the constraint violations. Otherwise, the method terminates with a point that approximately satisfies certain second-order necessary conditions for optimality. In this situation, if all termination conditions are removed then limit points either satisfy the same second-order necessary conditions exactly or fail to satisfy a weak second-order constraint qualification.

The main algorithm is intended for applications in which some guarantee of convergence to second-order solutions is needed. For other applications, the negative curvature direction may be omitted completely, or may be computed in the final stages of the optimization as a check that the iterates are converging to a point satisfying second-order optimality conditions. In the latter case the analysis implies that a direction of negative curvature, when it is computed, may be used without impeding the overall convergence rate or preventing global convergence.

Acknowledgements

The authors would like to thank the referees for constructive comments that significantly improved the presentation.

Funding

The research of P.E.G. was supported in part by National Science Foundation grants DMS-1318480 & DMS-1361421. The research of V.K. was supported by the European social fund within the framework of realizing the project “Support of inter-sectoral mobility and quality enhancement of research teams at the Czech Technical University in Prague”, CZ.1.07/2.3.00/30.0034. The research of D.P.R. was supported in part by National Science Foundation grant DMS-1217153.

References

Abarbanel, H. D. I., Bryant, P., Gill, P. E., Kostuk, M., Rofeh, J., Singer, Z., Toth, B. & Wong, E. (2011) Dynamical parameter and state estimation in neuron models. *An Exploration of Neuronal Variability and its Functional Significance* (D. Glanzman & M. Ding eds.) New York: Oxford University Press, pp. 139–180.

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

ANDREANI, R., BIRGIN, E. G., MARTÍNEZ, J. M. & SCHUVERDT, M. L. (2008) Augmented Lagrangian methods under the constant positive linear dependence constraint qualification. *Math. Program.*, **111**, 5–32.

ANDREANI, R., BIRGIN, E. G., MARTÍNEZ, J. M. & SCHUVERDT, M. L. (2010a) Second-order negative-curvature methods for box-constrained and general constrained optimization. *Comput. Optim. Appl.*, **45**, 209–236.

ANDREANI, R., HAESER, G., SCHUVERDT, M. L. & SILVA, P. J. S. (2012) Two new weak constraint qualifications and applications. *SIAM J. Optim.*, **22**, 1109–1135.

ANDREANI, R., MARTÍNEZ, J. M. & SCHUVERDT, M. L. (2007) On second-order optimality conditions for nonlinear programming. *Optimization*, **56**, 529–542.

ANDREANI, R., MARTÍNEZ, J. M. & SVAITER, B. F. (2010b) A new sequential optimality condition for constrained optimization and algorithmic consequences. *SIAM J. Optim.*, **20**, 3533–3554.

BOGGS, P. T. AND TOLLE, J. W. (1995) Sequential quadratic programming. *Acta Numer.*, **4**, 1–51.

CASTILLO, A. & O'NEILL, R. P. (2013) Computational performance of solution techniques applied to the ACOPF. *Optimal Power Flow Paper 5*, Federal Energy Regulatory Commission.

CONTESSE, L. B. (1980) Une caractérisation complète des minima locaux en programmation quadratique. *Numer. Math.*, **34**, 315–332.

COTTLE, R. W., HABETLER, G. J. & LEMKE, C. E. (1970) On classes of copositive matrices. *Linear Algebra Appl.*, **3**, 295–310.

CREVELING, D. R., GILL, P. E. & ABARBANEL, H. D. I. (2008) State and parameter estimation in nonlinear systems as an optimal tracking problem. *Phys. Let. A*, **372**, 2640–2644.

CURTIS, F. E. & NOCEDAL, J. (2008) Flexible penalty functions for nonlinear constrained optimization. *IMA J. Numer. Anal.*, **28**, 749–769.

FACCHINEI, F. & LUCIDI, S. (1998) Convergence to second order stationary points in inequality constrained optimization. *Math. Oper. Res.*, **23**, 746–766.

FERNÁNDEZ, D. & SOLODOV, M. (2010) Stabilized sequential quadratic programming for optimization and a stabilized Newton-type method for variational problems. *Math. Program. Ser. A*, **125**, 47–73.

FORSGREN, A. & GILL, P. E. (1998) Primal-dual interior methods for nonconvex nonlinear programming. *SIAM J. Optim.*, **8**, 1132–1152.

FORSGREN, A., GILL, P. E. & MURRAY, W. (1991) On the identification of local minimizers in inertia-controlling methods for quadratic programming. *SIAM J. Matrix Anal. Appl.*, **12**, 730–746.

GILL, P. E., KUNGURTSEV, V. & ROBINSON, D. P. (2013) *A Stabilized SQP Method: Global Convergence*. San Diego: Center for Computational Mathematics Report CCoM 13-04, University of California (Revised June 2015).

GILL, P. E., KUNGURTSEV, V. & ROBINSON, D. P. (2014) *A Stabilized SQP Method: Superlinear Convergence*. San Diego: Center for Computational Mathematics Report CCoM 14-01, University of California (Revised June 2015).

GILL, P. E. & ROBINSON, D. P. (2012) A primal-dual augmented Lagrangian. *Comput. Optim. Appl.*, **51**, 1–25.

GILL, P. E. & ROBINSON, D. P. (2013) A globally convergent stabilized SQP method. *SIAM J. Optim.*, **23**, 1983–2010.

GILL, P. E. & WONG, E. (2012) Sequential quadratic programming methods. *Mixed Integer Nonlinear Programming* (J. Lee & S. Leyffer eds.), vol. 154 IMA Volumes in Mathematics and its Applications. New York: Springer, pp. 147–224. 10.1007/978-1-4614-1927-3_6.

HAGER, W. W. (1999) Stabilized sequential quadratic programming. *Comput. Optim. Appl.*, **12**, 253–273.

IZMAILOV, A. F. & SOLODOV, M. V. (2011) On attraction of linearly constrained Lagrangian methods and of stabilized and quasi-Newton SQP methods to critical multipliers. *Math. Program.*, **126**, 231–257.

IZMAILOV, A. F. & SOLODOV, M. V. (2012) Stabilized SQP revisited. *Math. Program.*, **133**, 93–120.

KUNGURTSEV, V. (2013) Second-derivative sequential quadratic programming methods for nonlinear optimization. *Ph.D. Thesis*, Department of Mathematics, University of California San Diego, La Jolla, CA.

MAJTHAY, A. (1971) Optimality conditions for quadratic programming. *Math. Program.*, **1**, 359–365.

MANGASARIAN, O. L. (1994) *Nonlinear Programming*. Classics in Applied Mathematics, 10. Philadelphia, PA: Society for Industrial and Applied Mathematics (SIAM) Publications.

MANGASARIAN, O. L. & FROMOVITZ, S. (1967) The Fritz John necessary optimality conditions in the presence of equality and inequality constraints. *J. Math. Anal. Appl.*, **17**, 37–47.

MOGUERZA, J. M. & PRIETO, F. J. (2003) An augmented Lagrangian interior-point method using directions of negative curvature. *Math. Program.*, **95**, 573–616.

OBERLIN, C. & WRIGHT, S. J. (2006) Active set identification in nonlinear programming. *SIAM J. Optim.*, **17**, 577–605.

QI, L. & WEI, Z. (2000) On the constant positive linear dependence condition and its application to SQP methods. *SIAM J. Optim.*, **10**, 963–981.

ROBINSON, D. P. (2007) Primal–dual methods for nonlinear optimization. *Ph.D. Thesis*, Department of Mathematics, University of California San Diego, La Jolla, CA.

SAUNDERS, M. A. & TOMLIN, J. A. (1996) Solving regularized linear programs using barrier methods and KKT systems. *Report SOL 96-4*. Department of EESOR, Stanford University.

WRIGHT, S. J. (1998) Superlinear convergence of a stabilized SQP method to a degenerate solution. *Comput. Optim. Appl.*, **11**, 253–275.

WRIGHT, S. J. (2002) Modifying SQP for degenerate problems. *SIAM J. Optim.*, **13**, 470–497.

WRIGHT, S. J. (2005) An algorithm for degenerate nonlinear programming with rapid local convergence. *SIAM J. Optim.*, **15**, 673–696.