A GLOBALLY CONVERGENT STABILIZED SQP METHOD

PHILIP E. GILL* AND DANIEL P. ROBINSON†

Abstract. Sequential quadratic programming (SQP) methods are a popular class of methods for nonlinearly constrained optimization. They are particularly effective for solving a sequence of related problems, such as those arising in mixed-integer nonlinear programming and the optimization of functions subject to differential equation constraints.

Recently, there has been considerable interest in the formulation of *stabilized* SQP methods, which are specifically designed to handle degenerate optimization problems. Existing stabilized SQP methods are essentially local, in the sense that both the formulation and analysis focus on a neighborhood of a solution. We present the formulation and analysis of a new SQP method that has favorable global convergence properties yet is equivalent to a variant of the conventional stabilized SQP method in the neighborhood of a solution. The method is based on the combination of a primal-dual generalized augmented Lagrangian merit function with a *flexible* line search to obtain a sequence of improving estimates of the solution. An important feature of the method is that the quadratic programming (QP) subproblem is defined using the exact Hessian of the Lagrangian, yet has a unique bounded solution. This gives the potential for fast convergence in the neighborhood of a solution. Additional benefits of the method include: (i) each QP subproblem is regularized; (ii) the QP subproblem always has a known feasible point; and (iii) a projected gradient method may be used to identify the QP active set when far from the solution.

Key words. Nonlinear programming, nonlinear constraints, augmented Lagrangian, sequential quadratic programming, SQP methods, stabilized SQP, regularized methods, primal-dual methods.

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

1. Introduction.

In this paper we consider methods for the solution of optimization problems of the form:

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

where $c:\mathbb{R}^n \mapsto \mathbb{R}^m$ and $f:\mathbb{R}^n \mapsto \mathbb{R}$ 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 $l \le x \le u$.

Some of the most efficient algorithms for nonlinear optimization are sequential quadratic programming (SQP) methods (for a survey, see, e.g., [1, 24]). SQP methods are able to capitalize on a good initial starting point, which makes them particularly effective for solving a sequence of related problems, such as those arising in the optimization of functions subject to differential equation constraints. Conventional SQP methods find an approximate solution of a sequence of quadratic programming (QP) subproblems in which a quadratic model of the objective function is minimized subject to the linearized constraints. Convergence from any starting point is enforced by requiring the improvement in some merit function at each step. The merit function is usually a penalty or augmented Lagrangian function that defines some compromise between reducing the objective function and satisfying the constraints.

SQP methods have an inner/outer iteration structure, with the work for an inner iteration being dominated by the cost of solving a system of symmetric indefinite

*Department of Mathematics, University of California, San Diego, La Jolla, CA 92093-0112 (pgill@ucsd.edu). Research supported in part by National Science Foundation grants DMS-0511766 and DMS-0915220, and by Department of Energy grant DE-SC0002349.

†Department of Applied Mathematics and Statistics, Johns Hopkins University, Baltimore, MD 21218-2682 (daniel.p.robinson@jhu.edu).

linear equations involving a subset of the variables and constraints. Some of the most successful methods use sophisticated matrix factorization updating techniques that exploit the fact that the linear equations change by only a single row and column at each inner iteration. These updating techniques are often customized for the particular QP method being used and have the benefit of providing a uniform treatment of ill-conditioning and singularity. On the negative side, it is difficult to implement conventional SQP methods so that that the second derivatives of f and c may be used efficiently and reliably. Some of these difficulties stem from the theoretical properties of the quadratic programming subproblem, which can be nonconvex when second derivatives are used. Nonconvex quadratic programming is NP-hard—even for the calculation of a local minimizer [9, 17]. The complexity of the QP subproblem has been a major impediment to the formulation of conventional second-derivative SQP methods (although methods based on indefinite QP have been proposed [13, 14]). Over the years, algorithm developers have avoided this difficulty by using a convex QP defined in terms of a positive semidefinite approximate Hessian. In some cases, this QP is used to define the search direction directly [29, 37, 39, 40, 21, 20]; in others, the QP is used to identify the constraints for an equality constrained subproblem that uses second derivatives [26, 27, 34].

Recently, there has been considerable interest in the formulation of *stabilized* SQP methods, which are specifically designed to improve the convergence rate for degenerate problems [41, 28, 42, 35, 12, 32]. Existing stabilized SQP methods are essentially local, in the sense that both the formulation and analysis focus on a neighborhood of a solution. In parallel with these developments, *regularized* methods have been proposed that reduce the dependency on custom matrix factorization and updating schemes for solving the QP subproblem (see, e.g., [24]). Regularized methods are global methods in the sense that they provide for the efficient and accurate solution of the QP subproblem at every iteration.

A seemingly different approach from tackling problem (NP) directly is to replace the constrained problem by a sequence of bound-constrained problems in which the equality constraints are included in an augmented Lagrangian objective function [30, 36, 4, 6, 7, 8, 2]. These methods have strong global convergence properties that require relatively weak assumptions on the problem.

In this paper we formulate and analyze a new SQP method that smoothly combines the use of a bound-constrained augmented Lagrangian function with the three elements of conventional, regularized and stabilized SQP. In particular, the method has favorable global convergence properties, yet is equivalent to a variant of the conventional stabilized SQP method in the neighborhood of a solution. The method pairs the primal-dual generalized augmented Lagrangian merit function defined in [22] with a *flexible* line search to obtain a sequence of improving estimates of the solution. A crucial feature of the method is that the QP subproblem is defined using the exact Hessian of the Lagrangian, yet has a unique bounded solution. This gives the potential for fast convergence in the neighborhood of a solution. Additional benefits of the method include: (i) each QP subproblem is regularized; (ii) the QP subproblem always has a known feasible point; and (iii) a projected-gradient method may be used to identify the QP active set when far from the solution.

The paper is organized in six sections. Section 1.2 is a review of some of the basic properties of SQP methods. In Section 2, the steps of the proposed primal-dual SQP method are defined. Similarities with the conventional Hestenes-Powell augmented Lagrangian method are also discussed. In Section 3, we consider methods for the solution of the QP subproblem and show that in the neighborhood of a solution, the method is equivalent to a variant of the stabilized SQP method [28, 32, 41, 12]. A general global convergence result is established in Section 4 that does not make any constraint qualification or nondegeneracy assumption. Finally, in Section 5, a "convexification procedure" is proposed for obtaining a QP subproblem with a bounded unique solution even when the Hessian of the Lagrangian is not positive definite.

1.1. Notation and terminology. Unless explicitly indicated otherwise, $\|\cdot\|$ denotes the vector two-norm or its induced matrix norm. The inertia of a real symmetric matrix A, denoted by $\mathrm{In}(A)$, is the integer triple (a_+, a_-, a_0) giving the number of positive, negative and zero eigenvalues of A. Given vectors a and b with the same dimension, the vector with ith component $a_i b_i$ is denoted by $a \cdot b$. Similarly, $\min(a,b)$ is a vector with components $\min(a_i, b_i)$. The vectors e and e_j denote, respectively, the column vector of ones and the jth column of the identity matrix I. The dimensions of e, e_i and I are defined by the context. Given vectors x and y, the long vector consisting of the elements of x augmented by elements of y is denoted by (x, y). The ith component of a vector labeled with a subscript will be denoted by $[\,\cdot\,]_i$, e.g., $[\,v\,]_i$ is the ith component of the vector v. The subvector of components with indices in the index set $\mathcal{S}$ is denoted by $[\,\cdot\,]_{\mathcal{S}}$, e.g., $[\,v\,]_{\mathcal{S}}$ is the vector with components $[\,v\,]_i$ for $i \in \mathcal{S}$. Similarly, if M is a symmetric matrix, then $[\,M\,]_{\mathcal{S}}$ denotes the symmetric matrix with elements m_{ij} for $i, j \in \mathcal{S}$. A local solution of an optimization problem is denoted by x^*. The vector $g(x)$ is used to denote $\nabla f(x)$, the gradient of $f(x)$, and $H(x)$ denotes the (symmetric) Hessian matrix $\nabla^2 f(x)$. The matrix $J(x)$ denotes the $m \times n$ constraint Jacobian, which has ith row $\nabla c_i(x)^T$, the gradient of the ith constraint function $c_i(x)$. The matrix $H_i(x)$ denotes the Hessian of $c_i(x)$. The Lagrangian function associated with (NP) is $\mathcal{L}(x,y,z) = f(x) - c(x)^T y - z^T x$, where y and z are m- and n-vectors of dual variables associated with the equality constraints and bounds, respectively. The Hessian of the Lagrangian with respect to x is denoted by $H(x,y) = H(x) - \sum_{i=1}^m y_i H_i(x)$.

1.2. Background. The vector-pair (x^*, y^*) is called a first-order solution to problem (NP) if it satisfies

$$c(x^*) = 0 \quad \text{and} \quad \min\big(x^*, z^*\big) = 0, \tag{1.1}$$

where y^* and z^* are the Lagrange multipliers associated with the constraints $c(x) = 0$ and $x \ge 0$ respectively, with $z^* = g(x^*) - J(x^*)^T y^*$.

Given an estimate (x_k, y_k) of a primal-dual solution of (NP), a *line-search* SQP method computes a search direction p_k such that $x_k + p_k$ is the solution (when it exists) of the convex quadratic program

$$\begin{array}{ll} \underset{x}{\text{minimize}} & g_k^T(x - x_k) + \tfrac{1}{2}(x - x_k)^T \bar{H}_k (x - x_k) \\ \text{subject to} & c_k + J_k(x - x_k) = 0, \quad x \ge 0, \end{array} \tag{1.2}$$

where c_k, g_k and J_k denote the quantities $c(x)$, $g(x)$ and $J(x)$ evaluated at x_k, and $\bar{H}_k$ is some positive-definite approximation to $H(x_k, y_k)$. If the Lagrange multiplier vector associated with the constraint $c_k + J_k(x - x_k) = 0$ is written in the form $y_k + q_k$, then a solution $(x_k + p_k, y_k + q_k)$ of the QP subproblem (1.2) satisfies the optimality conditions

$$c_k + J_k p_k = 0 \quad \text{and} \quad \min\big(x_k + p_k,\; g_k + \bar{H}_k p_k - J_k^T(y_k + q_k)\big) = 0,$$

which are analogous to (1.1). Given any $x \ge 0$, let $\mathcal{A}$ and $\mathcal{F}$ denote the index sets

$$\mathcal{A}(x) = \{i : x_i = 0\} \ \text{ and } \ \mathcal{F}(x) = \{1, 2, \dots, n\}/\mathcal{A}(x). \tag{1.3}$$

If x is feasible for the constraints $c_k + J_k(x - x_k) = 0$, then $\mathcal{A}(x)$ is the *active set* at x. If the set $\mathcal{A}$ associated with a solution of the subproblem (1.2) is known, then $x_k + p_k$ may be found by solving linear equations that represent the optimality conditions for an equality-constrained QP with the inequalities $x \ge 0$ replaced by $x_i = 0$ for $i \in \mathcal{A}$. In general, the optimal $\mathcal{A}$ is not known in advance, and active-set methods generate a sequence of estimates $(\widehat{p}_j, \widehat{q}_j) \approx (p_k, q_k)$ such that $(\widehat{p}_{j+1}, \widehat{q}_{j+1}) = (\widehat{p}_j, \widehat{q}_j) + \alpha_j(\Delta p_j, \Delta q_j)$, with $(\Delta p_j, \Delta q_j)$ a solution of

$$\begin{pmatrix} \bar{H}_F & -J_F^T \\ J_F & 0 \end{pmatrix} \begin{pmatrix} \Delta p_F \\ \Delta q_j \end{pmatrix} = - \begin{pmatrix} [\, g_k + \bar{H}_k \widehat{p}_j - J_k^T(y_k + \widehat{q}_j)\,]_F \\ c_k + J_k \widehat{p}_j \end{pmatrix}, \tag{1.4}$$

where $\bar{H}_F$ is the matrix of free rows and columns of $\bar{H}_k$, J_F is the matrix of free columns of J_k, and the step length α is chosen to ensure feasibility of *all* variables, not just those in the set $\mathcal{A}$.

If the equations (1.4) are to be used to define Δp_F and Δq_j, then it is necessary that J_F has full rank, which is probably the greatest outstanding issue associated with systems of the form (1.4). Two remedies are available.

- *Rank-enforcing active-set methods* maintain a set of indices $\mathcal{B}$ associated with a matrix of columns J_B with rank m, i.e., the rows of J_B are linearly independent. The set $\mathcal{B}$ is the complement in $(1, 2, \dots, n)$ of a "working set" of indices that estimates the set $\mathcal{A}$ at a solution of (1.2). If $\mathcal{N}$ is a subset of $\mathcal{A}$, then the system analogous to (1.4) is given by

$$\begin{pmatrix} \bar{H}_B & -J_B^T \\ J_B & 0 \end{pmatrix} \begin{pmatrix} \Delta p_B \\ \Delta q_j \end{pmatrix} = - \begin{pmatrix} [\, g_k + \bar{H}_k \widehat{p}_j - J_k^T(y_k + \widehat{q}_j)\,]_B \\ c_k + J_k \widehat{p}_j \end{pmatrix}, \tag{1.5}$$

  which is nonsingular because of the linear independence of the rows of J_B.
- *Regularized active-set methods* include a nonzero diagonal regularization term in the $(2, 2)$ block of (1.4). The magnitude of the regularization is generally based on heuristic arguments that give mixed results in practice.

2. A Regularized Primal-Dual Line-Search SQP Algorithm. In this section, we define a regularized SQP line-search method based on the primal-dual augmented Lagrangian function

$$\mathcal{M}^\nu(x, y\,; y^E, \mu) = f(x) - c(x)^T y^E + \frac{1}{2\mu}\|c(x)\|^2 + \frac{\nu}{2\mu}\|c(x) + \mu(y - y^E)\|^2, \tag{2.1}$$

where ν is a scalar, μ is the so-called penalty parameter, and y^E is an estimate of an optimal Lagrange multiplier vector y^*. This function, proposed by Robinson [38], and Gill and Robinson [22], may be derived by applying the primal-dual penalty function of Forsgren and Gill [16] to a problem in which the constraints are shifted by a constant vector (see Powell [36]). With the notation $c = c(x)$, $g = g(x)$, and $J = J(x)$, the gradient of $\mathcal{M}^\nu(x, y\,; y^E, \mu)$ may be written as

$$\nabla \mathcal{M}^\nu(x, y\,; y^E, \mu) = \begin{pmatrix} g - J^T\big((1+\nu)(y^E - \frac{1}{\mu}c) - \nu y\big) \\ \nu\big(c + \mu(y - y^E)\big) \end{pmatrix} \tag{2.2a}$$

$$= \begin{pmatrix} g - J^T\big(\pi + \nu(\pi - y)\big) \\ \nu\mu(y - \pi) \end{pmatrix}, \tag{2.2b}$$

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

$$\pi(x\,;y^E,\mu) = y^E - \frac{1}{\mu}c(x). \tag{2.3}$$

Similarly, the Hessian of $\mathcal{M}^\nu(x,y\,;y^E,\mu)$ may be written as

$$\nabla^2\mathcal{M}^\nu(x,y\,;y^E,\mu) = \begin{pmatrix} H\big(x,\pi+\nu(\pi-y)\big) + \frac{1}{\mu}(1+\nu)J^TJ & \nu J^T \\ \nu J & \nu\mu I \end{pmatrix}. \tag{2.4}$$

We use $\mathcal{M}^\nu(x,y)$, $\nabla\mathcal{M}^\nu(x,y)$, and $\nabla^2\mathcal{M}^\nu(x,y)$, to denote $\mathcal{M}^\nu$, $\nabla\mathcal{M}^\nu$, and $\nabla^2\mathcal{M}^\nu$ evaluated with parameters y^E and μ. (We note that a trust-region based method could also be given, but we leave the statement and analysis to a future paper.)

Our approach is motivated by the following theorem, which shows that minimizers of problem (NP) are also minimizers—under certain assumptions—of the bound constrained problem

$$\underset{x,y}{\text{minimize}}\ \ \mathcal{M}^\nu(x,y\,;y^*,\mu) \quad \text{subject to} \ \ x \ge 0, \tag{2.5}$$

where y^* is a Lagrange multiplier vector for the equality constraints $c(x) = 0$.

THEOREM 2.1. *If (x^*,y^*) satisfies the second-order sufficient conditions for a solution of problem* (NP)*, then there exists a positive $\bar\mu$ such that for all $0 < \mu < \bar\mu$, the point (x^*,y^*) is a minimizer of the bound constrained problem* (2.5) *for all $\nu > 0$.* ∎

2.1. Definition of the search direction. To motivate the computation of the step, we consider a quadratic approximation to $\mathcal{M}^\nu$. Given (x,y) and fixed $\nu \ge 0$, we define

$$H_{\scriptscriptstyle M}^\nu(x,y\,;\mu) = \begin{pmatrix} \bar H(x,y) + \frac{1}{\mu}(1+\nu)J(x)^TJ(x) & \nu J(x)^T \\ \nu J(x) & \nu\mu I \end{pmatrix}, \tag{2.6}$$

where $\bar H(x,y)$ is a symmetric approximation to $H\big(x,\pi+\nu(\pi-y)\big) \approx H(x,y)$ such that $\bar H(x,y) + \frac{1}{\mu}J(x)^TJ(x)$ is positive definite. The approximation $\pi+\nu(\pi-y)\approx y$ is valid provided $\pi \approx y$. The restriction on the inertia of $\bar H$ implies that $H_{\scriptscriptstyle M}^\nu(x,y\,;\mu)$ is positive definite for $\nu > 0$ and positive semidefinite for $\nu = 0$ (see Theorem 3.1 of Section 3.2.3).

Using this definition of $H_{\scriptscriptstyle M}^\nu$ at the kth primal-dual iterate $v_k = (x_k,y_k)$, consider the convex QP subproblem

$$\underset{\Delta v=(p,q)}{\text{minimize}}\ \nabla\mathcal{M}^\nu(v_k)^T\Delta v + \tfrac{1}{2}\Delta v^T H_{\scriptscriptstyle M}^\nu(v_k)\Delta v \quad \text{subject to} \ \ x_k + p \ge 0, \tag{2.7}$$

where $\mathcal{M}^\nu(v)$ denotes the merit function evaluated at $v = (x,y)$. For any primal-dual QP solution $\Delta v_k = (p_k,q_k)$, it is shown in Theorem 3.3 of Section 3.2.3 that the first-order conditions associated with the variables in $\mathcal{F}(x_k+p_k)$ may be written in matrix form as:

$$\begin{pmatrix} \bar H_{\scriptscriptstyle F} & -J_{\scriptscriptstyle F}^T \\ J_{\scriptscriptstyle F} & \mu I \end{pmatrix}\begin{pmatrix} p_{\scriptscriptstyle F} \\ q_k \end{pmatrix} = -\begin{pmatrix} [\,g_k - J_k^Ty_k - \bar H_k s\,]_{\scriptscriptstyle F} \\ c_k + \mu(y_k - y^E) - J_k s \end{pmatrix}, \tag{2.8}$$

where c_k, g_k and J_k denote the quantities $c(x)$, $g(x)$ and $J(x)$ evaluated at x_k, and s is a nonnegative vector such that

$$s_i = \begin{cases} [\,x_k\,]_i & \text{if } i \in \mathcal{A}(x_k+p_k); \\ 0 & \text{if } i \in \mathcal{F}(x_k+p_k). \end{cases}$$

(The assumption of positive-definiteness of $\bar{H}_k + \frac{1}{\mu} J_k^T J_k$ implies that the matrix associated with the equations (2.8) is nonsingular.) It follows that if $\mathcal{A}(x_k + p_k) = \mathcal{A}(x_k)$, then (p_k, q_k) satisfies the perturbed Newton equations

$$\begin{pmatrix} H_F & -J_F^T \\ J_F & \mu I \end{pmatrix} \begin{pmatrix} p_F \\ q_k \end{pmatrix} = - \begin{pmatrix} [\, g_k - J_k^T y_k \,]_F \\ c_k + \mu(y_k - y^E) \end{pmatrix}.$$

A key property is that if $\mu = 0$ and J_F has full rank, then this equation is identical to the equation for the conventional SQP step given by (1.4). This provides the motivation to use different penalty parameters for the step computation and the merit function.

Given an iterate $v_k = (x_k, y_k)$ and Lagrange multiplier estimate y_k^E, the primal-dual search direction $\Delta v_k = (p_k, q_k)$ is defined such that $v_k + \Delta v_k = (x_k + p_k, y_k + q_k)$ is a solution of the convex bound-constraint QP problem:

$$\begin{array}{ll} \underset{v=(x,y)}{\text{minimize}} & (v - v_k)^T \nabla \mathcal{M}^\nu(v_k\,; y_k^E, \mu_k^R) + \frac{1}{2}(v - v_k)^T H_M^\nu(v_k\,; \mu_k^R)(v - v_k) \\ \text{subject to} & x \geq 0, \end{array} \tag{2.9}$$

where μ_k^R is a small parameter, and $H_M^\nu(v_k\,; \mu_k^R)$ is the matrix (2.6) written in terms of the composite variables $v_k = (x_k, y_k)$. In this context, μ_k^R plays the role of a *regularization* parameter rather than a *penalty* parameter, thereby providing an $O(\mu_k^R)$ estimate of the conventional SQP direction. This approach is nonstandard because a small "penalty parameter" μ_k^R is used by design, whereas other augmented Lagrangian-based methods attempt to keep μ as large as possible [3, 20].

Finally, we note that if $v = v_k$ is a solution of the QP (2.9), then v_k is a first-order solution of

$$\underset{v=(x,y)}{\text{minimize}} \;\; \mathcal{M}^\nu(v\,; y_k^E, \mu_k^R) \quad \text{subject to} \quad x \geq 0. \tag{2.10}$$

In Section 3 it is shown that, under certain conditions, the primal-dual vector $v_k + \Delta v_k = (x_k + p_k, y_k + q_k)$ is a solution of problem (2.9) if and only if it solves

$$\begin{array}{ll} \underset{x,y}{\text{minimize}} & g_k^T(x - x_k) + \frac{1}{2}(x - x_k)^T \bar{H}(x_k, y_k)(x - x_k) + \frac{1}{2}\mu_k^R \|y\|^2 \\ \text{subject to} & c_k + J_k(x - x_k) + \mu_k^R(y - y_k^E) = 0, \;\; x \geq 0, \end{array} \tag{2.11}$$

which is often referred to as the "stabilized" SQP subproblem because of its calming effect on multiplier estimates for degenerate problems (see, e.g., [28, 41]). Therefore, the proposed method provides a natural link between the stabilized SQP methods (which employ a subproblem appropriate for degenerate problems), conventional SQP methods (which are highly efficient in practice), and augmented Lagrangian methods (which have desirable global convergence properties).

2.2. Definition of the new iterate. Once the search direction Δv_k has been determined, a "flexible" backtracking line search is performed on the primal-dual augmented Lagrangian. A conventional backtracking line search defines $v_{k+1} = v_k + \alpha_k \Delta v_k$, where $\alpha_k = 2^{-j}$ and j is the smallest nonnegative integer such that

$$\mathcal{M}^\nu(v_k + \alpha_k \Delta v_k\,; y_k^E, \mu_k) \leq \mathcal{M}^\nu(v_k\,; y_k^E, \mu_k) + \alpha_k \eta_S \Delta v_k^T \nabla \mathcal{M}^\nu(v_k\,; y_k^E, \mu_k)$$

for a given scalar $\eta_S \in (0, 1)$. However, this approach would suffer from the Maratos effect [33] simply because the penalty parameter μ_k and the regularization parameter

μ_k^R generally have different values. Thus, we use a "flexible penalty function" based on the work of Curtis and Nocedal [10] and define $\alpha_k = 2^{-j}$, where j is the smallest nonnegative integer such that

$$\mathcal{M}^\nu(v_k + \alpha_k \Delta v_k\,; y_k^E, \mu_k^F) \le \mathcal{M}^\nu(v_k\,; y_k^E, \mu_k^F) + \alpha_k \eta_S N_k \tag{2.12}$$

for some value $\mu_k^F \in [\mu_k^R, \mu_k]$, and where

$$N_k \triangleq \max\big(\Delta v_k^T \nabla \mathcal{M}^\nu(v_k\,; y_k^E, \mu_k^R), -10^{-3}\|\Delta v_k\|^2\big) \le 0 \tag{2.13}$$

is a sufficiently negative scalar whose value is relevant to the proof of global convergence. Once an appropriate value for α_k is found, the new primal-dual solution estimate is given by

$$x_{k+1} = x_k + \alpha_k p_k \quad \text{and} \quad y_{k+1} = y_k + \alpha_k q_k.$$

We note that the step acceptance criterion is well-defined because the *weakened* Armijo condition (2.12) will be satisfied for $\mu_k^F = \mu_k^R$ and all α sufficiently small.

2.3. Updating the multiplier estimate. The preliminary numerical results presented in [22] indicate that the method outlined thus far is robust with respect to updating y_k^E. In particular, the numerical results generated in that paper updated y_k^E at *every* iteration. Consequently, we seek a strategy that allows for frequent updates to y_k^E. To this end, we use the (merit) functions

$$\phi_S(v) = \eta(x) + 10^{-5}\omega(v) \quad \text{and} \quad \phi_L(v) = 10^{-5}\eta(x) + \omega(v), \tag{2.14}$$

where $\eta(x)$ and $\omega(v)$ are the feasibility and stationarity measures

$$\eta(x) = \|c(x)\| \quad \text{and} \quad \omega(x,y) = \big\|\min\big(x, g(x) - J(x)^T y\big)\big\| \tag{2.15}$$

at the point $v = (x, y)$. These functions measure the accuracy of (x, y) as an approximate solution of problem (NP) rather than as an approximate minimizer of $\mathcal{M}^\nu$. Both measures are bounded below by zero, and are equal to zero if v is a first-order solution to problem (NP). Such conditions are appropriate because trial steps are regularized SQP steps that should converge rapidly to a solution of problem (NP).

The estimate y_k^E is updated when any iterate v_k satisfies either $\phi_S(v_k) \le \frac{1}{2}\phi_S^{\max}$ or $\phi_L(v_k) \le \frac{1}{2}\phi_L^{\max}$, where $\phi_S^{\max}$ and $\phi_L^{\max}$ are bounds that are updated throughout the solution process. To ensure global convergence, the update to y_k^E is accompanied by a decrease in either $\phi_S^{\max}$ or $\phi_L^{\max}$.

Finally, y_k^E is also updated if an approximate first-order solution of the problem

$$\underset{x,y}{\text{minimize}}\;\; \mathcal{M}^\nu(x, y\,; y_k^E, \mu_k^R) \quad \text{subject to} \;\; x \ge 0 \tag{2.16}$$

has been found. The test for optimality is

$$\|\nabla_y \mathcal{M}^\nu(v_{k+1}\,; y_k^E, \mu_k^R)\| \le \tau_k \;\; \text{and} \;\; \big\|\min\big(x_{k+1}, \nabla_x \mathcal{M}^\nu(v_{k+1}\,; y_k^E, \mu_k^R)\big)\big\| \le \tau_k \tag{2.17}$$

for some small tolerance $\tau_k > 0$. This condition is rarely satisfied in practice, but the test is required for the proof of convergence. Nonetheless, if the condition is satisfied, y_k^E is updated with the *safeguarded* estimate

$$y_{k+1}^E = \text{mid}\big(-10^6, y_{k+1}, 10^6\big).$$

2.4. Updating the penalty parameters. As we only want to decrease μ_k^R when "close" to optimality (ignoring locally infeasible problems), we use the definition

$$\mu_{k+1}^R = \begin{cases} \min\left(\frac{1}{2}\mu_k^R, \|r_{k+1}\|^{3/2}\right), & \text{if (2.17) is satisfied;} \\ \min\left(\ \mu_k^R, \|r_{k+1}\|^{3/2}\right), & \text{otherwise,} \end{cases} \tag{2.18}$$

where

$$r_{k+1} \equiv r_{\mathrm{opt}}(v_{k+1}) \triangleq \begin{pmatrix} c(x_{k+1}) \\ \min\left(x_{k+1}, g(x_{k+1}) - J(x_{k+1})^T y_{k+1}\right) \end{pmatrix}. \tag{2.19}$$

The update to μ_k is motivated by a different goal. Namely, we wish to decrease μ_k only when the trial step indicates that the merit function with penalty parameter μ_k *increases*. Thus, we use the definition

$$\mu_{k+1} = \begin{cases} \mu_k, & \mathcal{M}^\nu(v_{k+1}\,; y_k^E, \mu_k) \le \mathcal{M}^\nu(v_k\,; y_k^E, \mu_k) + \widehat{\alpha}_k \eta_S N_k \\ \max\left(\frac{1}{2}\mu_k, \mu_{k+1}^R\right), & \text{otherwise,} \end{cases} \tag{2.20}$$

where $\widehat{\alpha}_k = \min(\alpha_{\min}, \alpha_k)$ for some positive $\alpha_{\min}$. The use of the scalar $\alpha_{\min}$ increases the likelihood that μ_k will not be decreased.

2.5. Formal statement of the algorithm. In this section we formally state the proposed method as Algorithm 2.1 and include some additional details. During each iteration, the trial step is computed as described in Section 2.1, the solution estimate is updated as in Section 2.2, y_k^E is updated as in Section 2.3, and the penalty parameters are updated as in Section 2.4. The value of y_k^E is crucial for both global and local convergence. To this end, there are three possibilities. First, y_k^E is set to y_{k+1} if (x_{k+1}, y_{k+1}) is acceptable to either of the merit functions ϕ_S or ϕ_L given by (2.14). These iterates are labeled as S- and L-iterates, respectively. It is to be expected that y_k^E will be updated in this way most of the time. Second, if (x_{k+1}, y_{k+1}) is not acceptable to either of the merit functions ϕ_S or ϕ_L, we check whether we have computed an approximate first-order solution to problem (2.16) by verifying conditions (2.17) for the current value of τ_k. If these conditions are satisfied, the iterate is called an M-iterate. In this case, the regularization parameter μ_k^R and subproblem tolerance τ_k are decreased and y_k^E is updated as in (2.3). Finally, an iterate at which neither of the first two cases occur is called an F-iterate. The multiplier estimate y_k^E is not changed in an F-iterate.

Algorithm 2.1. Regularized primal-dual SQP algorithm (pdSQP)
Input (x_0, y_0);
Set algorithm parameters $\alpha_{\min} > 0$, $\eta_S \in (0,1)$, $\tau_{\text{stop}} > 0$, $n_{\max} > 0$, and $\nu > 0$;
Initialize $y_0^E = y_0$, $\tau_0 > 0$, $\mu_0^R > 0$, $\mu_0 \in [\mu_0^R, \infty)$, and $k = 0$;
Compute $f(x_0)$, $c(x_0)$, $g(x_0)$, $J(x_0)$, and $H(x_0, y_0)$;
for $k = 0, 1, 2, \ldots n_{\max}$ **do**
  Define $\bar{H}_k \approx H(x_k, y_k)$ such that $\bar{H}_k + (1/\mu_k^R) J_k^T J_k$ is positive definite;
  Solve the QP (2.9) for the search direction $\Delta v_k = (p_k, q_k)$;
  Find an α_k satisfying (2.12) and (2.13);
  Update the primal-dual estimate $x_{k+1} = x_k + \alpha_k p_k$, $\quad y_{k+1} = y_k + \alpha_k q_k$;
  Compute $f(x_{k+1})$, $c(x_{k+1})$, $g(x_{k+1})$, $J(x_{k+1})$, and $H(x_{k+1}, y_{k+1})$;
  if $\phi_S(x_{k+1}, y_{k+1}) \le \frac{1}{2}\phi_S^{\max}$ **then** [S-iterate]

$\phi_S^{\max} = \frac{1}{2}\phi_S^{\max}$;
$y_{k+1}^E = y_{k+1}$;
$\tau_{k+1} = \tau_k$;
else if $\phi_L(x_{k+1}, y_{k+1}) \le \frac{1}{2}\phi_L^{\max}$ **then** [L-iterate]
$\phi_L^{\max} = \frac{1}{2}\phi_L^{\max}$;
$y_{k+1}^E = y_{k+1}$;
$\tau_{k+1} = \tau_k$;
else if $v_{k+1} = (x_{k+1}, y_{k+1})$ satisfies (2.17) [M-iterate]
$y_{k+1}^E = \text{mid}(-10^6, y_{k+1}, 10^6)$;
$\tau_{k+1} = \frac{1}{2}\tau_k$;
else [F-iterate]
$y_{k+1}^E = y_k^E$;
$\tau_{k+1} = \tau_k$;
end if
Update μ_{k+1}^R and μ_{k+1} according to (2.18) and (2.20), respectively;
if $\|r_{k+1}\| \le \tau_{\text{stop}}$ **then** *exit*;
end (for)

3. Solution of the QP Subproblem. In this section we consider various theoretical and computational issues associated with the bound-constraint QP subproblem (2.9). In particular, it is shown that the search direction computed using subproblem (2.9) is the unique solution of the stabilized SQP subproblem (2.11), and *independent* of the value of ν. Moreover, a conventional active-set method applied to problems (2.9) and (2.11) generates identical iterates, provided a common (feasible) starting point is used.

3.1. Equivalence with stabilized SQP. First, we show that under certain conditions, the regularized QP subproblem (2.9) is equivalent to the stabilized SQP subproblem (2.11). Equivalent problems are considered in which the unknowns are written in terms of the increment (p, q) for given variables (x, y).

THEOREM 3.1. *Consider the bound constrained QP*

$$\underset{\Delta v=(p,q)}{\text{minimize}} \;\; g_M^T \Delta v + \tfrac{1}{2}\Delta v^T H_M \Delta v \quad \text{subject to} \quad x + p \ge 0, \tag{3.1}$$

where x and y are constant,

$$g_M = \begin{pmatrix} g - J^T\big(\pi + \nu(\pi - y)\big) \\ \nu\big(c + \mu(y - y^E)\big) \end{pmatrix}, \quad \text{and} \quad H_M = \begin{pmatrix} H + \frac{1}{\mu}(1+\nu)J^TJ & \nu J^T \\ \nu J & \nu\mu I \end{pmatrix},$$

with $H + \frac{1}{\mu}J^TJ$ positive definite and $\nu \ge 0$. For the same quantities c, g, J and H, consider the stabilized QP problem

$$\begin{aligned} \underset{p,q}{\text{minimize}} \quad & g^Tp + \tfrac{1}{2}p^THp + \tfrac{1}{2}\mu\|y+q\|^2 \\ \text{subject to} \quad & c + Jp + \mu(y + q - y^E) = 0, \quad x + p \ge 0. \end{aligned} \tag{3.2}$$

The following results hold.

(i) *The stabilized QP* (3.2) *has a unique bounded primal-dual solution* (p, q).

(ii) *The unique solution* $\Delta v = (p, q)$ *of the stabilized QP* (3.2) *is a solution of the bound constrained QP* (3.1) *for all* $\nu \geq 0$. *If* $\nu > 0$, *then the stabilized solution* $\Delta v = (p, q)$ *is the* unique *solution of* (3.1).

Proof. For part (i), let $\Delta v = (p, q)$ denote an arbitrary feasible point for the constraints of the stabilized QP (3.2). Given the particular feasible point $\Delta v_0 = (0, \pi - y)$, consider an n-vector of variables w defined by the linear transformation

$$\Delta v = \Delta v_0 + Mw, \quad \text{where} \quad M = \begin{pmatrix} \mu I \\ -J \end{pmatrix}.$$

The matrix M is $(n + m) \times n$ with rank n, and its columns form a basis for the null-space of the constraint matrix $\begin{pmatrix} J & \mu I \end{pmatrix}$. Using this transformation gives rise to the equivalent problem

$$\underset{w \in \mathbb{R}^n}{\text{minimize}} \;\; \frac{\mu}{2} w^T \big(H + \frac{1}{\mu} J^T J\big) w + w^T \big(g - J^T \pi\big) \;\; \text{subject to} \;\; x + \mu w \geq 0.$$

The matrix $H + \frac{1}{\mu} J^T J$ is positive definite by assumption, and it follows that the stabilized QP (3.2) is equivalent to a convex program with a strictly convex objective. The existence of a bounded unique solution follows directly.

For part (ii), we begin by stating the first-order conditions for (p, q) to be a solution of the stabilized QP (3.2):

$$\begin{aligned} c + Jp + \mu(y + q - y^{E}) &= 0, & \mu(y + q) &= \mu w, \\ g + Hp - J^T w - z &= 0, & z &\geq 0, \\ z \cdot (x + p) &= 0, & x + p &\geq 0, \end{aligned}$$

where w and z denote the dual variables for the equality and inequality constraints of problem (3.2), respectively. Eliminating w using the equation $w = y + q$ gives

$$c + Jp + \mu(y + q - y^{E}) = 0, \tag{3.3a}$$

$$g + Hp - J^T(y + q) - z = 0, \qquad z \geq 0, \tag{3.3b}$$

$$z \cdot (x + p) = 0, \qquad x + p \geq 0. \tag{3.3c}$$

First, we prove part (ii) for the case $\nu > 0$. The optimality conditions for (3.1) are

$$g_{M} + H_{M} \Delta v = \begin{pmatrix} z \\ 0 \end{pmatrix}, \qquad z \geq 0, \tag{3.4}$$

$$z \cdot (x + p) = 0, \qquad x + p \geq 0.$$

Pre-multiplying the equality of (3.4) by the nonsingular matrix T such that

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

and using the definition (2.2a) yields the equivalent conditions

$$g + Hp - J^T(y + q) - z = 0 \quad \text{and} \quad c + Jp + \mu(y + q - y^{E}) = 0,$$

which are identical to the relevant equalities in (3.3). Thus, the solutions of (3.2) and (3.1) are identical in the case $\nu > 0$.

It remains to consider the case $\nu = 0$. In this situation, the objective function of the QP (3.1) includes only the primal variables p, which implies that the problem may be written as

$$\underset{p}{\text{minimize}} \;\; (g - J^T\pi)^T p + \tfrac{1}{2} p^T\big(H + \frac{1}{\mu} J^T J\big)p \quad \text{subject to} \;\; x + p \ge 0, \tag{3.5}$$

with q an arbitrary vector. Although there are infinitely many solutions of (3.1), the vector p associated with a particular solution (p, q) is unique because it is the solution of problem (3.5) for a positive-definite matrix $H + \frac{1}{\mu} J^T J$. The optimality conditions for (3.5) are

$$\begin{aligned} g - J^T\pi + \big(H + \frac{1}{\mu} J^T J\big)p &= z, & z &\ge 0, \\ z \cdot (x + p) &= 0, & x + p &\ge 0. \end{aligned} \tag{3.6}$$

For the given y and optimal p, define the m-vector q such that

$$q = -\frac{1}{\mu}\big(Jp + c + \mu(y - y_e)\big) = -\frac{1}{\mu}\big(Jp + \mu(y - \pi)\big). \tag{3.7}$$

Equation (3.7) and the equality of (3.6) may be combined to give the matrix equation

$$\begin{pmatrix} g - J^T y + 2J^T(y - \pi) \\ \mu(y - \pi) \end{pmatrix} + \begin{pmatrix} H + \frac{2}{\mu} J^T J & J^T \\ J & \mu I \end{pmatrix} \begin{pmatrix} p \\ q \end{pmatrix} = \begin{pmatrix} z \\ 0 \end{pmatrix}.$$

Applying the nonsingular matrix $\begin{pmatrix} I & -\frac{2}{\mu} J^T \\ 0 & I \end{pmatrix}$ to both sides of this equation yields

$$\begin{pmatrix} g - J^T y \\ c + \mu(y - y_e) \end{pmatrix} + \begin{pmatrix} H & -J^T \\ J & \mu I \end{pmatrix} \begin{pmatrix} p \\ q \end{pmatrix} = \begin{pmatrix} z \\ 0 \end{pmatrix}.$$

It follows that if $\nu = 0$, then the unique solution of (3.2) is a solution of (3.1), which is what we wanted to show.

When $\nu > 0$, the uniqueness of the solution $\Delta v = (p, q)$ follows from the observation that QP (3.1) is then convex with a strictly convex objective. □

Theorem 3.1 shows that the direction defined by the bound-constrained QP is *independent of the parameter* ν. Moreover, this direction may be defined as the solution of an equivalent stabilized SQP subproblem (2.11) that does not include ν at all. However, the parameter ν does appear explicitly in the definition of the merit function $\mathcal{M}^\nu$ (2.1), and therefore plays an important role in influencing the length of the step during the flexible line search. The value of ν determines the proximity of the primal-dual iterates to the so-called "primal-dual trajectory", which is the one-parameter family of points $\big(x(\mu), y(\mu)\big)$, such that $x(\mu)$ is a minimizer of the conventional augmented Lagrangian for fixed y^E. The definition of $\mathcal{M}^\nu$ implies that larger values of ν tend to force the iterates to be close to the primal-dual trajectory. If $\nu = 0$ then the method reverts to a regularized SQP method based on the (primal) conventional augmented Lagrangian (for which *no* emphasis is placed on staying close to the *primal-dual* trajectory). The algorithm may be modified to allow for the choice

$\nu = 0$ by always setting y^E_{k+1} to be $\pi(x_{k+1})$; this does emphasize the primal-dual trajectory, but only *after* the major iteration has been completed. The use of the *primal-dual* augmented Lagrangian function allows the emphasis on the dual variables *during* the line search.

3.2. Equivalent iterates of an active-set method. In Section 3.1 it is shown that, if $\nu > 0$ then the *unique* solutions of subproblems (2.9) and (2.11) are identical. If $\nu = 0$ then the solution of (2.9) is no longer unique, but there is a particular solution that is identical to the unique solution of (2.11). In this section we extend this analysis to consider the relationship between the iterates of an active-set method applied to each problem.

3.2.1. An active-set method. For the remainder of this section, the indices associated with the SQP iteration are omitted and it will be assumed that the constraints of the QP involve the constraints linearized at the point $\bar{x}$. In all cases, the suffix j will be reserved for the iteration index of the QP algorithm.

We start by defining a "conventional" active-set method on a generic convex QP with constraints written in standard form. The problem format is

$$\begin{array}{ll} \underset{x}{\text{minimize}} & \mathcal{Q}(x) = g^T(x-\bar{x}) + \tfrac{1}{2}(x-\bar{x})^T H (x-\bar{x}) \\ \text{subject to} & c + A(x-\bar{x}) = 0, \quad x \ge 0, \end{array} \tag{3.8}$$

where $\bar{x}$, c, A, g and H are constant, with H positive-definite. Throughout, we assume that the constraints are feasible, i.e., there exists at least one nonnegative x such that $c + A(x-\bar{x}) = 0$.

Given a feasible x_0, active-set methods generate a feasible sequence $\{x_j\}$ such that $\mathcal{Q}(x_{j+1}) \le \mathcal{Q}(x_j)$ with $x_{j+1} = x_j + \alpha_j p_j$. Let the index sets $\mathcal{A}$ and $\mathcal{F}$ be defined as in (1.3). At the start of the jth QP iteration, given primal-dual iterates (x_j, w_j), new estimates $(x_j + p_j, w_j + q_j)$ are defined by solving a QP formed by fixing the variables with indices in $\mathcal{A}(x_j)$ and defining p_j such that $x_j + p_j$ minimizes $\mathcal{Q}(x)$ with respect to the free variables, subject to the equality constraints. With this definition, the quantities $w_j + q_j$ are the Lagrange multipliers at the minimizer $x_j + p_j$. The components of p_j with indices in $\mathcal{A}(x_j)$ are zero, and the free components $p_F = [\,p_j\,]_F$ are determined from the equations

$$\begin{pmatrix} H_F & -A_F^T \\ A_F & 0 \end{pmatrix} \begin{pmatrix} p_F \\ q_j \end{pmatrix} = - \begin{pmatrix} [\,g + H(x_j-\bar{x}) - A^T w_j\,]_F \\ c + A(x_j - \bar{x}) \end{pmatrix}, \tag{3.9}$$

where $[\,\cdot\,]_F$ denotes the subvector of components with indices in $\mathcal{F}(x_j)$. The choice of step length α_j is based on remaining feasible with respect to the satisfied bounds. If $x_j + p_j$ is feasible, i.e., $x_j + p_j \ge 0$, then α_j will be taken as unity. Otherwise, α is set to α_M, the largest feasible step along p_j. Finally, the iteration index j is incremented by one and the iteration is repeated.

It must be emphasized that this active-set method is not well defined unless the equations (3.9) have a solution at every (x_j, w_j).

3.2.2. Solution of the bound-constrained subproblem. Next we apply the active-set method to a bound constrained QP of the form

$$\underset{v=(x,y)}{\text{minimize}} \;\; g_M^T(v-\bar{v}) + \tfrac{1}{2}(v-\bar{v})^T H_M (v-\bar{v}) \quad \text{subject to} \;\; x \ge 0, \tag{3.10}$$

where $\bar{v} = (\bar{x}, \bar{y})$, and

$$g_M = \begin{pmatrix} g - J^T\big(\pi + \nu(\pi - \bar{y})\big) \\ \nu\big(c + \mu(\bar{y} - y^E)\big) \end{pmatrix}, \quad H_M = \begin{pmatrix} H + \frac{1}{\mu}(1+\nu)J^TJ & \nu J^T \\ \nu J & \nu\mu I \end{pmatrix},$$

with $H + \frac{1}{\mu}J^TJ$ positive definite, $\nu \ge 0$, and $\pi = y^E - c/\mu$ (see (2.3)). The matrix H_M is positive semidefinite under the given assumptions. This follows from the identity

$$L^T H_M L = \begin{pmatrix} H + \frac{1}{\mu}J^TJ & 0 \\ 0 & \nu\mu I_m \end{pmatrix}, \quad \text{where} \quad L = \begin{pmatrix} I_n & 0 \\ -\frac{1}{\mu}J & I_m \end{pmatrix}.$$

The matrix L is nonsingular, and Sylvester's Law of Inertia gives

$$\text{In}(H_M) = \text{In}(L^T H_M L) = \text{In}\Big(H + \frac{1}{\mu}J^TJ\Big) + (m,0,0) = (n+m,0,0) \ \text{ for } \nu > 0,$$

and

$$\text{In}(H_M) = \text{In}\Big(H + \frac{1}{\mu}J^TJ\Big) + (0,0,m) = (n,0,m) \ \text{ for } \nu = 0.$$

It follows that problem (3.10) is a convex QP, and we may apply the active-set method of Section 3.2.1.

Given the jth QP iterate $v_j = (x_j, y_j)$, the generic active-set method applied to (3.10) defines the next iterate as $v_{j+1} = v_j + \alpha_j \Delta v_j$, where the free components of the vector $\Delta v_j = (p_j, q_j)$ satisfy the equations

$$[\, H_M \,]_F \Delta v_F = -[\, g_M + H_M(v_j - \bar{v}) \,]_F, \tag{3.11}$$

where $\Delta v_F = (p_F, q_j)$ and the index set $\mathcal{F}(x_j)$ is defined as in (1.3). The equations (3.11) appear to be ill-conditioned for small μ because of the $O(1/\mu)$ term in the $(1,1)$ block of the matrix H_M. However, this ill-conditioning is superficial. The next result shows that Δv_F may be determined by solving an equivalent nonsingular primal-dual system with conditioning dependent on that of the original problem.

THEOREM 3.2. *Consider the application of the active-set method to the bound constrained QP* (3.10)*. Then, for every* $\nu \ge 0$*, there exists a positive* $\bar{\mu}$ *such that, for all* $0 < \mu < \bar{\mu}$*, the free components of the QP search direction* (p_j, q_j) *satisfy the nonsingular primal-dual system*

$$\begin{pmatrix} H_F & -J_F^T \\ J_F & \mu I \end{pmatrix} \begin{pmatrix} p_F \\ q_j \end{pmatrix} = -\begin{pmatrix} [\, g + H(x_j - \bar{x}) - J^T y_j \,]_F \\ c + \mu(y_j - y^E) + J(x_j - \bar{x}) \end{pmatrix}. \tag{3.12}$$

Proof. Consider the definition of the search direction when $\nu > 0$. In this case it suffices to show that the linear systems (3.11) and (3.12) are equivalent. For any positive ν, we may define the matrix

$$T = \begin{pmatrix} I & -\frac{1+\nu}{\nu\mu} J_F^T \\ 0 & \frac{1}{\nu} I_m \end{pmatrix},$$

where the identity matrix I has dimension n_F, the column dimension of J_F. The matrix T is nonsingular with $n_F + m$ rows and columns. It follows that the equations

$$T[\, H_M \,]_F \Delta v_F = -T[\, g_M + H_M(v_j - \bar{v}) \,]_F$$

have the same solution as those of (3.11). The primal-dual equations (3.12) follow by direct multiplication. The nonsingularity of the equations (3.12) follows from the nonsingularity of T, and the fact that H_M (and all symmetric submatrices formed from its rows and columns) is nonsingular.

The resulting equations (3.12) are independent of ν, but the simple proof above is not applicable when $\nu = 0$ because T is undefined in this case. For $\nu = 0$, the QP objective includes only the primal variables x, which implies that problem (3.10) may be written as

$$\underset{x\geq 0}{\text{minimize}}\;\; \big(g - J^T\pi\big)^T(x-\bar{x}) + \tfrac{1}{2}(x-\bar{x})^T\Big(H + \frac{1}{\mu}J^TJ\Big)(x-\bar{x}),$$

with y arbitrary. The active-set equations analogous to (3.11) are then

$$\Big(H_F + \frac{1}{\mu}J_F^TJ_F\Big)p_F = -\Big[g + \Big(H + \frac{1}{\mu}J^TJ\Big)(x_j - \bar{x}) - J^T\pi\Big]_F. \tag{3.13}$$

For *any* choice of y_j, define the m-vector q_j such that

$$q_j = -\frac{1}{\mu}\Big(J_Fp_F + \mu(y_j - \pi) + J(x_j - \bar{x})\Big), \tag{3.14}$$

where $\pi = y^E - c/\mu$ (see (2.3)). Equations (3.13) and (3.14) may be combined to give equations $K\Delta v_F = -r$, where $\Delta v_F = (p_F, q_j)$,

$$K = \begin{pmatrix} H_F + \frac{2}{\mu}J_F^TJ_F & J_F^T \\ J_F & \mu I \end{pmatrix}$$

and right-hand side

$$r = \begin{pmatrix} [\,g + H(x_j - \bar{x})\,]_F + \frac{2}{\mu}J_F^TJ(x_j-\bar{x}) - J_F^Ty_j + 2J_F^T(y_j - \pi) \\ \mu(y_j - \pi) + J(x_j - \bar{x}) \end{pmatrix}.$$

Forming the equations $TK\Delta v_F = -Tr$, where T is the nonsingular matrix

$$T = \begin{pmatrix} I & -\frac{2}{\mu}J_F^T \\ 0 & I_m \end{pmatrix},$$

gives the equivalent system

$$\begin{pmatrix} H_F & -J_F^T \\ J_F & \mu I \end{pmatrix}\begin{pmatrix} p_F \\ q_j \end{pmatrix} = -\begin{pmatrix} [\,g + H(x_j - \bar{x}) - J^Ty_j\,]_F \\ c + \mu(y_j - y^E) + J(x_j - \bar{x}) \end{pmatrix},$$

which is identical to the system (3.12). □

THEOREM 3.3. *Let* (p_k, q_k) *be the solution of the QP subproblem* (2.7). *If* p_F *denotes the components of* p_k *with indices in* $\mathcal{F}(x_k + p_k)$, *then* (p_F, q_k) *satisfies the equations*

$$\begin{pmatrix} \bar{H}_F & -J_F^T \\ J_F & \mu I \end{pmatrix}\begin{pmatrix} p_F \\ q_k \end{pmatrix} = -\begin{pmatrix} [\,g_k - J_k^Ty_k - \bar{H}_ks\,]_F \\ c_k + \mu(y_k - y^E) - J_ks \end{pmatrix},$$

where $\mathcal{F}$ *is defined in terms of the set* $\mathcal{F}(x_k + p_k)$ *and* s *is a nonnegative vector such that*

$$s_i = \begin{cases} [\,x_k\,]_i & \text{if } i \in \mathcal{A}(x_k + p_k); \\ 0 & \text{if } i \in \mathcal{F}(x_k + p_k). \end{cases}$$

Proof. The proof is analogous to that of Theorem 3.2. □

3.2.3. Solution of the stabilized SQP subproblem. In this section we show that under certain conditions, the conventional active-set method applied to the stabilized SQP subproblem (3.2) and the bound-constrained QP (3.1) will generate identical iterates.

Consider the application of the "generic" active-set method of Section 3.2.1 to the stabilized QP:

$$\begin{array}{ll} \underset{x,y}{\text{minimize}} & g^T(x-\bar{x}) + \tfrac{1}{2}(x-\bar{x})^T H (x-\bar{x}) + \tfrac{1}{2}\mu\|y\|^2 \\ \text{subject to} & c + J(x-\bar{x}) + \mu(y - y^{\scriptscriptstyle E}) = 0, \quad x \ge 0. \end{array} \tag{3.15}$$

In terms of the data "$(x, \bar{x}, H, g, A, c)$" for the generic QP (3.8), we have variables "x" $= (x, y)$, with "$\bar{x}$" $= (\bar{x}, \bar{y})$,

$$\text{“}H\text{”} = \begin{pmatrix} H & 0 \\ 0 & \mu I \end{pmatrix}, \quad \text{“}g\text{”} = \begin{pmatrix} g \\ \mu\bar{y} \end{pmatrix}, \quad \text{“}A\text{”} = \begin{pmatrix} J & \mu I \end{pmatrix}, \quad \text{and} \quad \text{“}c\text{”} = c + \mu(\bar{y} - y^{\scriptscriptstyle E}).$$

(The discussion of the properties of the stabilized QP relative to the generic form (3.8) is not affected by the nonnegativity constraints being applied to only a subset of the variables in (3.15).) After some simplification, the equations analogous to (3.9) may be written as

$$\begin{pmatrix} H_{\scriptscriptstyle F} & 0 & -J_{\scriptscriptstyle F}^T \\ 0 & \mu I & -\mu I \\ J_{\scriptscriptstyle F} & \mu I & 0 \end{pmatrix} \begin{pmatrix} p_{\scriptscriptstyle F} \\ \bar{p}_{\scriptscriptstyle F} \\ q_j \end{pmatrix} = - \begin{pmatrix} [\, g + H(x_j - \bar{x}) - J^T w_j \,]_{\scriptscriptstyle F} \\ \mu y_j - \mu w_j \\ c + \mu(y_j - y^{\scriptscriptstyle E}) + J(x_j - \bar{x}) \end{pmatrix}, \tag{3.16}$$

where $p_{\scriptscriptstyle F}$ and $\bar{p}_{\scriptscriptstyle F}$ denote the free components of the search directions for the x and y variables respectively. (Observe that the right-hand side of (3.16) is independent of $\bar{y}$.) The second block of equations gives $\bar{p}_{\scriptscriptstyle F} = q_j - y_j + w_j$, which implies that

$$y_{j+1} = y_j + \bar{p}_{\scriptscriptstyle F} = y_j + q_j - y_j + w_j = w_j + q_j = w_{j+1},$$

so that the primal y-variables and dual variables of the stabilized QP are identical.

Similarly, substituting for $\bar{p}_{\scriptscriptstyle F}$ in the third block of equations in (3.17), and using the primal-dual equivalence $w_j = y_j$ gives

$$\begin{pmatrix} H_{\scriptscriptstyle F} & -J_{\scriptscriptstyle F}^T \\ J_{\scriptscriptstyle F} & \mu I \end{pmatrix} \begin{pmatrix} p_{\scriptscriptstyle F} \\ q_j \end{pmatrix} = - \begin{pmatrix} [\, g + H(x_j - \bar{x}) - J^T y_j \,]_{\scriptscriptstyle F} \\ c + \mu(y_j - y^{\scriptscriptstyle E}) + J(x_j - \bar{x}) \end{pmatrix}, \tag{3.17}$$

which are identical to the equations associated with those for the QP subproblem (3.10).

The preceding discussion constitutes a proof of the following result.

THEOREM 3.4. *Consider the application of the active-set method to the bound constrained QP* (3.10) *and stabilized QP* (3.15) *defined with the same quantities c, g, J and H. Consider any x_0 and y_0 such that (x_0, y_0) is feasible for the stabilized QP* (3.15)*. Then, for every $\nu \ge 0$, there exists a positive $\bar{\mu}$ such that, for all $0 < \mu < \bar{\mu}$, the active-set method generates identical primal-dual iterates $\{(x_j, y_j)\}_{j\ge 0}$.* ∎

4. Convergence. The convergence of Algorithm 2.1 is discussed under the following assumptions.

ASSUMPTION 4.1. *Each $\bar{H}(x_k, y_k)$ is chosen so that the sequence $\{\bar{H}(x_k, y_k)\}_{k\ge 0}$ is bounded, with $\{\bar{H}(x_k, y_k) + (1/\mu_k^{\scriptscriptstyle R}) J(x_k)^T J(x_k)\}_{k\ge 0}$ uniformly positive definite.*

ASSUMPTION 4.2. *The functions f and c are twice continuously differentiable.*

Assumption 4.3. *The sequence $\{x_k\}_{k\ge 0}$ is contained in a compact set.*

In the "worst" case, i.e., when all iterates are eventually M-iterates or F-iterates, Algorithm 2.1 emulates a *primal-dual* augmented Lagrangian method [5, 6, 38]. Consequently, it is possible that $y^{\scriptscriptstyle E}_k$ and $\mu^{\scriptscriptstyle R}_k$ will remain fixed over a sequence of iterations, although this is unusual in practice. The following result concerns this situation.

Theorem 4.1. *Let Assumptions 4.1–4.3 hold. If there exists an integer $\widehat{k}$ such that $\mu^{\scriptscriptstyle R}_k \equiv \mu^{\scriptscriptstyle R} > 0$ and k is an $\mathcal{F}$-iterate for all $k \ge \widehat{k}$, then the following hold:*

(i) *solutions $\{\Delta v_k\}_{k\ge\widehat{k}}$ to subproblem (2.9) are bounded above;*

(ii) *solutions $\{\Delta v_k\}_{k\ge\widehat{k}}$ to subproblem (2.9) are bounded away from zero; and*

(iii) *there exists a constant $\epsilon > 0$ such that*

$$\nabla\mathcal{M}^\nu(v_k\,; y^{\scriptscriptstyle E}_k, \mu^{\scriptscriptstyle R}_k)^T\Delta v_k \le -\epsilon \ \text{ for all } k \ge \widehat{k}.$$

Proof. The assumptions of this theorem imply that

$$\tau_k \equiv \tau > 0, \ \ \mu^{\scriptscriptstyle R}_k = \mu^{\scriptscriptstyle R}, \ \text{ and } \ y^{\scriptscriptstyle E}_k = y^{\scriptscriptstyle E} \ \text{ for all } k \ge \widehat{k}. \tag{4.1}$$

We first prove part (i). As discussed in the proof of Theorem 3.1, it is known that the solution of (2.9) satisfies

$$\Delta v_k = \begin{pmatrix} 0 \\ \pi_k - y_k \end{pmatrix} + M_k w^*, \ \text{ where } \ M_k = \begin{pmatrix} \mu^{\scriptscriptstyle R} I \\ -J_k \end{pmatrix},$$

and w^* is the unique solution of

$$\underset{w\in\mathbb{R}^n}{\text{minimize}} \ \ \tfrac{1}{2}\mu^{\scriptscriptstyle R} w^T\Big(\bar{H}_k + \frac{1}{\mu^{\scriptscriptstyle R}} J_k^T J_k\Big)w + w^T\big(g_k - J_k^T\pi_k\big) \ \text{ subject to } \ x_k + \mu^{\scriptscriptstyle R} w \ge 0,$$

for all $k \ge \widehat{k}$. It follows from Assumption 4.1 that $\{\Delta v_k\}_{k\ge\widehat{k}}$ is uniformly bounded provided that the quantities $g_k - J_k^T\pi_k$, M_k, π_k, and y_k are all uniformly bounded for $k \ge \widehat{k}$. The boundedness of $g_k - J_k^T\pi_k$, π_k and M_k follow from Assumption 4.2, Assumption 4.3, (4.1), and (2.3). It remains to prove that $\{y_k\}_{k\ge\widehat{k}}$ is bounded. To this end, we first note that as $\mu^{\scriptscriptstyle R}_k = \mu^{\scriptscriptstyle R}$ for all $k \ge \widehat{k}$, the update to μ_k given by (2.20) implies that $\mu_k \equiv \mu \ge \mu^{\scriptscriptstyle R}$ for some μ and all k sufficiently large. For all subsequent iterations the primal-dual merit function is monotonically decreasing, i.e., $\mathcal{M}^\nu(x_{k+1}, y_{k+1}\,; y^{\scriptscriptstyle E}, \mu) \le \mathcal{M}^\nu(x_k, y_k\,; y^{\scriptscriptstyle E}, \mu)$. It follows that $\{y_k\}_{k\ge\widehat{k}}$ must be bounded, since if there were a subsequence such that $\|y_k\|$ goes to infinity, then for the same subsequence $\mathcal{M}^\nu$ would also go to infinity because both $\{f_k - c_k^T y^{\scriptscriptstyle E} + \frac{1}{2\mu}\|c_k\|^2\}_{k\ge\widehat{k}}$ and $\{c_k\}_{k\ge\widehat{k}}$ are bounded from Assumptions 4.2 and 4.3. This completes the proof of part (i).

Part (ii) is established by showing that $\{\Delta v_k\}_{k\ge\widehat{k}}$ is bounded away from zero. If this were not the case, there would exist a subsequence $\mathcal{S}_1 \subseteq \{k : k \ge \widehat{k}\}$ such that $\lim_{k\in\mathcal{S}_1} \Delta v_k = 0$. It follows that the solution Δv_k to problem (2.9) satisfies

$$\begin{pmatrix} z_k \\ 0 \end{pmatrix} = H^\nu_{\scriptscriptstyle M}(v_k\,; \mu^{\scriptscriptstyle R})\Delta v_k + \nabla\mathcal{M}^\nu(v_k\,; y^{\scriptscriptstyle E}, \mu^{\scriptscriptstyle R}) \ \text{ and } \ 0 = \min(x_k + p_k, z_k)$$

for all $k \in \mathcal{S}_1$. We may then conclude from the definition of $H^\nu_{\scriptscriptstyle M}$ and Assumptions 4.1–4.3, and (4.1) that for $k \in \mathcal{S}_1$ sufficiently large, iterate v_k will satisfy condition (2.17), be an M-iterate, and $\mu^{\scriptscriptstyle R}_k$ would be decreased. This contradicts the assumption that

$\mu_k^R \equiv \mu^R$ for all $k \geq \widehat{k}$. It follows that $\{\|v_k\|\}_{k\geq\widehat{k}}$ is bounded away from zero and part (ii) holds.

The proof of part (iii) is also by contradiction. Assume that there exists a subsequence $\mathcal{S}_2$ of $\{k : k \geq \widehat{k}\}$ such that

$$\lim_{k\in\mathcal{S}_2} \nabla\mathcal{M}^\nu(v_k\,; y^E, \mu^R)^T\Delta v_k = 0, \tag{4.2}$$

where we have used (4.1). As the vector $\Delta v = 0$ is feasible for the convex problem (2.9), and Δv_k is the solution to problem (2.9) for $\nu > 0$ chosen in Algorithm 2.1, it must hold that

$$\begin{aligned}
-\nabla\mathcal{M}^\nu(v_k\,; y^E, \mu^R)^T\Delta v_k &\geq \tfrac{1}{2}\Delta v_k^T H_M^\nu(v_k\,; \mu^R)\Delta v_k \\
&= \tfrac{1}{2}\Delta v_k^T L_k^{-T} L_k^T H_M^\nu(v_k\,; \mu^R) L_k L_k^{-1}\Delta v_k \\
&= \tfrac{1}{2}\Delta v_k^T L_k^{-T} \begin{pmatrix} \bar{H}_k + \frac{1}{\mu^R}J_k^T J_k & 0 \\ 0 & \nu\mu^R \end{pmatrix} L_k^{-1}\Delta v_k,
\end{aligned}$$

where L_k denotes the nonsingular matrix

$$L_k = \begin{pmatrix} I & 0 \\ -\frac{1}{\mu^R}J_k & I \end{pmatrix}, \quad \text{with} \quad L_k^{-1}\Delta v_k = \begin{pmatrix} p_k \\ q_k + \frac{1}{\mu^R}J_k p_k \end{pmatrix}.$$

Assumption 4.1 yields

$$\begin{aligned}
-\nabla\mathcal{M}^\nu(v_k\,; y^E, \mu^R)^T\Delta v_k &\geq \tfrac{1}{2}p_k^T\Big(\bar{H}_k + \frac{1}{\mu^R}J_k^T J_k\Big)p_k + \tfrac{1}{2}\nu\mu^R\|q_k + (1/\mu^R)J_k p_k\|^2 \\
&\geq \lambda_{\min}\|p_k\|^2 + \tfrac{1}{2}\nu\mu^R\|q_k + (1/\mu^R)J_k p_k\|^2,
\end{aligned}$$

for some $\lambda_{\min} > 0$. Combining this with (4.2) gives the limit

$$\lim_{k\in\mathcal{S}_2} p_k = \lim_{k\in\mathcal{S}_2}\Big(q_k + \frac{1}{\mu^R}J_k p_k\Big) = 0,$$

in which case $\lim_{k\in\mathcal{S}_2} q_k = 0$ follows from Assumptions 4.2 and 4.3. This contradicts the result of part (ii), which implies that $\lim_{k\in\mathcal{S}_2}\Delta v_k = 0$. It follows that part (iii) must hold. ◻

We may now state our convergence result for Algorithm 2.1.

THEOREM 4.2. *Let Assumptions* 4.1–4.3 *hold. If* v_k *denotes the kth iterate generated by Algorithm* 2.1, *then either:*

(i) *Algorithm* 2.1 *terminates with an approximate primal-dual first-order solution* v_k *satisfying*

$$\|r_{\text{opt}}(v_k)\| \leq \tau_{\text{stop}},$$

where r_{opt} *is defined by* (2.19)*; or*

(ii) *there exists a subsequence* $\mathcal{S}$ *such that* $\lim_{k\in\mathcal{S}}\mu_k^R = 0$, $\{y_k^E\}_{k\in\mathcal{S}}$ *is bounded,* $\lim_{k\in\mathcal{S}}\tau_k = 0$, *and for each* $k \in \mathcal{S}$ *the vector* v_{k+1} *is an approximate minimizer of the primal-dual augmented Lagrangian function* (2.1) *that satisfies* (2.17).

Proof. There are two cases to consider.

Case 1. A subsequence of $\{\|r_{\text{opt}}(v_k)\|\}_{k\geq 0}$ converges to zero.
In this case it is clear from the definition of S-iterates, M-iterates, ϕ_S, and ϕ_L, and the fact that $\tau_{\text{stop}} > 0$ that part (i) will be satisfied for some k sufficiently large.

Case 2. The sequence $\{\|r_{\text{opt}}(v_k)\|\}_{k\geq 0}$ is bounded away from zero.
From the definitions of an S-iterate, M-iterate, and the functions ϕ_S, and ϕ_L, we conclude that the number of S-iterates and L-iterates must be finite. We claim that there must be an infinite number of M-iterates. To prove this, assume to the contrary that the number of M-iterates is finite, so that all iterates are F-iterates for k sufficiently large. It follows from the form of the update to μ_k^R (2.18) and the assumption made in this case, that eventually μ_k^R remains constant. In this case, the update to μ_k given by (2.20) implies that eventually, μ_k also remains constant. These arguments imply the existence of an integer $\widehat{k}$ such that

$$\mu_k^R \equiv \mu^R \leq \mu \equiv \mu_k, \;\; y_k^E \equiv y^E, \;\; \tau_k \equiv \tau > 0, \;\text{ and } k \text{ is an F-iterate for all } k \geq \widehat{k}.$$

It follows from (2.20) that

$$\mathcal{M}^\nu(v_{k+1}\,; y^E, \mu) \leq \mathcal{M}^\nu(v_k\,; y^E, \mu) + \min(\alpha_{\min}, \alpha_k)\eta_S N_k \;\text{ for all } k \geq \widehat{k}, \tag{4.3}$$

where N_k is defined by (2.13). Moreover, parts (ii) and (iii) of Theorem 4.1 ensure that $\{N_k\}_{k\geq\widehat{k}}$ is a negative sequence bounded away from zero. We also claim that $\{\alpha_k\}_{k\geq\widehat{k}}$ is bounded away from zero. To see this, note that parts (i) and (iii) of Theorem 4.1 and Assumption 4.2 ensure that $\{\alpha_k\}_{k\geq\widehat{k}}$ is bounded away from zero if a *standard* Armijo line search is used, i.e., if $\mu_k^F = \mu^R$ and $N_k = \Delta v_k^T \nabla\mathcal{M}^\nu(v_k; y^E, \mu^R)$ in (2.12). However, the computed value of α_k can be no smaller because the definition of N_k is less restrictive and the use of a flexible line search makes the acceptance of a step more likely. Combining these facts with (4.3), yields

$$\mathcal{M}^\nu(v_{k+1}\,; y^E, \mu) \leq \mathcal{M}^\nu(v_k\,; y^E, \mu) - \kappa \;\text{ for all } k \geq \widehat{k} \text{ and some } \kappa > 0,$$

so that

$$\lim_{k\to\infty} \mathcal{M}^\nu(v_k\,; y^E, \mu) = -\infty.$$

However, Assumptions 4.2 and 4.3 ensure that this is not possible. This contradiction implies that there must exist infinitely many M-iterations, and *every* iterate is an M-iterate or and F-iterate for k sufficiently large. Part (ii) now follows from (2.18) and the properties of the updates to τ_k and y_k^E used for M-iterates and F-iterates in Algorithm 2.1. □

The "ideal" scenario is that Algorithm 2.1 generates many S-iterates/L-iterates that rapidly converge to an approximate solution of NP; this corresponds to part (i) of Theorem 4.2. The scenario considered in part (ii) of Theorem 4.2, i.e., the generation of infinitely many M-iterates, is the fall-back position of Algorithm 2.1. This result would appear to be the best that can be expected because no constraint qualification has been assumed. In fact, the assumptions we have made does no preclude the possibility that problem NP is infeasible. It has been proved recently [12, 31, 11] that iterates generated from the stabilized SQP subproblem exhibit superlinear convergence under mild conditions; in particular, strict complementarity is not assumed and no constraint qualification is required.

5. Convexification of the Bound-Constrained Subproblem. An important feature of the proposed method is that the QP subproblem is convex. A conventional QP subproblem defined in terms of the Hessian of the Lagrangian is not convex, in general. To avoid solving an indefinite subproblem, most existing methods are based on solving a convex QP based on a positive semidefinite approximation to the Hessian. This convex subproblem is used to either define the search direction directly, or identify the constraints for an equality-constrained QP subproblem using the exact derivatives.

Here we take a different approach and define a *convexified* QP subproblem defined in terms of the exact Hessian of the Lagrangian. The convex problem is defined in such a way that if the inner iterations do not alter the active set, then the computed direction is a variant of the second-derivative stabilized SQP direction. The method is based on defining a matrix $\bar{H}_k$ (not necessarily positive definite) such that

$$H_{\mathcal{M}} = \begin{pmatrix} \bar{H}_k + \frac{1}{\mu}(1+\nu)J_k^TJ_k & \nu J_k^T \\ \nu J_k & \nu\mu I \end{pmatrix}$$

is positive definite. The method proposed in this paper defines $\bar{H}_k$ as a diagonal modification of the Lagrangian Hessian H_k.

For the remainder of this section we focus on the solution of a single QP subproblem and omit the suffix k.

If $\mathcal{F}(x)$ denotes the index set of the free variables at x, let J_F and J_A denote the columns of J associated with $\mathcal{F}(x)$ and its complement $\mathcal{A}(x)$, respectively. For given H and J, let K and K_F denote the matrices

$$K = \begin{pmatrix} H + D & J^T \\ J & -\mu I \end{pmatrix} \quad \text{and} \quad K_F = \begin{pmatrix} H_F + D_F & J_F^T \\ J_F & -\mu I \end{pmatrix}, \tag{5.1}$$

where D is a diagonal modification. We are particularly interested in diagonal matrices D that endow K_F with the property of *second-order consistency.*

Definition 5.1. *If $D_A = 0$, D_F is positive semidefinite, and the inertia of K_F is $(n_F, m, 0)$, then K_F is said to be* second-order consistent.

Methods for computing the diagonal modification D_F are based on some variant of the symmetric indefinite factorization of the matrix

$$\begin{pmatrix} H_F & J_F^T \\ J_F & -\mu I \end{pmatrix}.$$

Methods include: (i) the *inertia controlling LBL^T factorization* (Forsgren [15], Forsgren and Gill [16]); (ii) an LBL^T factorization with pivot modification (Gould [25]); and (iii) tile preordering in conjunction with pivot modification (Gill and Wong [23]).

Generally, in the neighborhood of a solution no modification will be required, i.e., $D = 0$.

The second-order consistent K_F is used to define a convex QP at the point x. First, we show how K_F may be used to provide a positive define matrix with respect to the free variables.

Lemma 5.2. *If the KKT matrix K_F (5.1) is second-order consistent, then the matrix*

$$B = \begin{pmatrix} \widehat{H}_F + \frac{1}{\mu}(1+\nu)J_F^TJ_F & \nu J_F^T \\ \nu J_F & \nu\mu I \end{pmatrix}, \quad \textit{where} \;\; \widehat{H}_F = H_F + D_F, \tag{5.2}$$

is positive definite for $\nu > 0$, and positive semidefinite for $\nu = 0$.

Proof. Consider the identity

$$S^T \begin{pmatrix} \widehat{H}_F & J_F^T \\ J_F & -\mu I \end{pmatrix} S = \begin{pmatrix} \widehat{H}_F + \frac{1}{\mu} J_F^T J_F & 0 \\ 0 & -\mu I_m \end{pmatrix}, \quad \text{with} \quad S = \begin{pmatrix} I_F & 0 \\ \frac{1}{\mu} J_F & I_m \end{pmatrix}.$$

Using the fact that K_F is second-order consistent, the nonsingularity of the matrix S, and Sylvester's law of inertia, it follows that

$$(n_F, m, 0) = \text{In}(K_F) = \text{In}(S^T K_F S) = \text{In}\left(\widehat{H}_F + \frac{1}{\mu} J_F^T J_F\right) + (0, m, 0),$$

which implies that $\widehat{H}_F + \frac{1}{\mu} J_F^T J_F$ is positive definite. Similarly,

$$L^T B L = \begin{pmatrix} \widehat{H}_F + \frac{1}{\mu} J_F^T J_F & 0 \\ 0 & \nu\mu I_m \end{pmatrix}, \quad \text{where} \quad L = \begin{pmatrix} I_F & 0 \\ -\frac{1}{\mu} J & I_m \end{pmatrix}.$$

The matrix L is nonsingular, and Sylvester's law of inertia gives

$$\text{In}(B) = \text{In}(L^T B L) = \text{In}\left(\widehat{H}_F + \frac{1}{\mu} J_F^T J_F\right) + (m, 0, 0) = (n_F + m, 0, 0) \ \text{ for } \nu > 0,$$

and

$$\text{In}(B) = \text{In}\left(\widehat{H}_F + \frac{1}{\mu} J_F^T J_F\right) + (0, 0, m) = (n_F, 0, m) \ \text{ for } \nu = 0.$$

It follows that (5.2) is positive definite or positive semidefinite depending on the value of ν. □

The next lemma gives the inertia of a KKT matrix that includes the gradients of the constraints with indices in $\mathcal{A}(x)$. If these gradients form the columns of the matrix P_A, then $P_A P_A^T x$ is the projection of x onto the bounds in $\mathcal{A}(x)$. The complementary projection may be defined in terms of the matrix P_F with (unit) columns orthogonal to P_A.

Lemma 5.3. *Consider the matrix*

$$K_A = \begin{pmatrix} \widehat{H} + \frac{1}{\mu}(1+\nu) J^T J & \nu J^T & P_A \\ \nu J & \nu\mu I & 0 \\ P_A^T & 0 & 0 \end{pmatrix}, \quad \textit{where} \ \widehat{H} = H + D$$

and the n_A rows of P_A^T comprise the gradients of the bounds in $\mathcal{A}(x)$. Then

$$\text{In}(K_A) = (n_A, n_A, 0) + \text{In}\begin{pmatrix} \widehat{H}_F + \frac{1}{\mu}(1+\nu) J_F^T J_F & \nu J_F^T \\ \nu J_F & \nu\mu I \end{pmatrix}.$$

Proof. Applying the column permutation $P = \begin{pmatrix} P_F & P_A \end{pmatrix}$ yields

$$\begin{pmatrix} P^T & 0 & 0 \\ 0 & I_m & 0 \\ 0 & 0 & I_A \end{pmatrix} K_A \begin{pmatrix} P & 0 & 0 \\ 0 & I_m & 0 \\ 0 & 0 & I_A \end{pmatrix} = \begin{pmatrix} \bar{H}_F & \bar{H}_D & \nu J_F^T & 0 \\ \bar{H}_D & \bar{H}_A & \nu J_A^T & I_A \\ \nu J_F & \nu J_A & \nu\mu I & 0 \\ 0 & I_A & 0 & 0 \end{pmatrix}, \tag{5.3}$$

where $\bar{H}_F$, $\bar{H}_A$, and $\bar{H}_D$ are the diagonal and off-diagonal blocks of the partition

$$P^T\big(\widehat{H} + \frac{1}{\mu}(1+\nu)J^TJ\big)P = \begin{pmatrix} \bar{H}_F & \bar{H}_D \\ \bar{H}_D^T & \bar{H}_A \end{pmatrix}.$$

Note that $\bar{H}_F = \widehat{H}_F + \frac{1}{\mu}(1+\nu)J_F^TJ_F$. The matrix of (5.3) is similar (via symmetric permutations) to

$$\begin{pmatrix} 0 & I_A & 0 & 0 \\ I_A & \bar{H}_A & \bar{H}_D^T & \nu J_A^T \\ 0 & \bar{H}_D & \bar{H}_F & \nu J_F^T \\ 0 & \nu J_A & \nu J_F & \nu\mu I \end{pmatrix} = L \begin{pmatrix} 0 & I_A & 0 & 0 \\ I_A & \bar{H}_A & 0 & 0 \\ 0 & 0 & \bar{H}_F & \nu J_F^T \\ 0 & 0 & \nu J_F & \nu\mu I \end{pmatrix} L^T,$$

where L is the nonsingular matrix

$$L = \begin{pmatrix} I_A & 0 & 0 & 0 \\ 0 & I_A & 0 & 0 \\ \bar{H}_D & 0 & I_F & 0 \\ \nu J_A & 0 & 0 & I \end{pmatrix}.$$

Sylvester's law of inertia gives

$$\begin{aligned} \mathrm{In}(K_A) &= \mathrm{In}\begin{pmatrix} 0 & I_A \\ I_A & \bar{H}_A \end{pmatrix} + \mathrm{In}\begin{pmatrix} \bar{H}_F & \nu J_F^T \\ \nu J_F & \nu\mu I \end{pmatrix} \\ &= (n_A, n_A, 0) + \mathrm{In}\begin{pmatrix} \widehat{H}_F + \frac{1}{\mu}(1+\nu)J_F^TJ_F & \nu J_F^T \\ \nu J_F & \nu\mu I \end{pmatrix}, \end{aligned}$$

as required. □

The main result on the properties of the convexified Hessian uses the following lemma of Debreu (for one of many proofs in the literature, see Gill and Robinson [22]).

LEMMA 5.4 (Debreu). *Let B be a symmetric matrix of order r. Let C be an $m\times r$ matrix with rank m. If the matrix $\begin{pmatrix} B & C^T \\ C & 0 \end{pmatrix}$ has inertia $(r, m, 0)$, then $B + \frac{1}{\mu}C^TC$ is positive definite for all $\mu > 0$ sufficiently small.* ∎

THEOREM 5.5. *If the KKT matrix K_F (5.1) is a second-order consistent matrix, then the matrix*

$$H_{\mathcal{M}} = \begin{pmatrix} \bar{H} + \frac{1}{\mu}(1+\nu)J^TJ & \nu J^T \\ \nu J & \nu\mu I \end{pmatrix}, \quad \textit{with} \quad \bar{H} = H + D + \frac{1}{\mu^C}P_AP_A^T,$$

is positive definite for all $\mu^C > 0$ sufficiently small.

Proof. Consider the matrix

$$K_A = \begin{pmatrix} \widehat{H} + \frac{1}{\mu}(1+\nu)J^TJ & \nu J^T & P_A \\ \nu J & \nu\mu I & 0 \\ P_A^T & 0 & 0 \end{pmatrix}, \quad \text{where } \widehat{H} = H + D$$

and P_A contains the unit vectors associated with the active bounds. Then

$$\begin{aligned} \mathrm{In}(K_A) &= (n_A, n_A, 0) + \mathrm{In}\begin{pmatrix} \widehat{H}_F + \frac{1}{\mu}(1+\nu)J_F^TJ_F & \nu J_F^T \\ \nu J_F & \nu\mu I \end{pmatrix} \\ &= (n_A, n_A, 0) + (n_F + m, 0, 0) \qquad \text{(from Lemma 5.2)} \\ &= (n + m, n_A, 0). \end{aligned}$$

This identity implies that K_A satisfies the conditions of Lemma 5.4 with

$$B = \begin{pmatrix} \widehat{H} + \frac{1}{\mu}(1+\nu)J^TJ & \nu J^T \\ \nu J & \nu\mu I \end{pmatrix} \quad \text{and} \quad C^T = \begin{pmatrix} P_A \\ 0 \end{pmatrix},$$

in which case, the matrix

$$B + \frac{1}{\mu^C} C^TC = \begin{pmatrix} \widehat{H} + \frac{1}{\mu}(1+\nu)J^TJ & \nu J^T \\ \nu J & \nu\mu I \end{pmatrix} + \frac{1}{\mu^C} \begin{pmatrix} P_A \\ 0 \end{pmatrix} \begin{pmatrix} P_A^T & 0 \end{pmatrix} = H_{\mathcal{M}}$$

is positive definite for $\mu^C > 0$ sufficiently small, which completes the proof. ◻

Theorem 5.5 implies that

$$H_{\mathcal{M}} = \begin{pmatrix} \bar{H} + \frac{1}{\mu}(1+\nu)J^TJ & \nu J^T \\ \nu J & \nu\mu I \end{pmatrix} \quad \text{with} \quad \bar{H} = H + D + \frac{1}{\mu^C} P_A P_A^T,$$

is an appropriate positive-definite Hessian for the convexified QP. An important property of this definition is that

$$\begin{pmatrix} \bar{H}_F & J_F^T \\ J_F & -\mu I \end{pmatrix} = \begin{pmatrix} H_F + D_F & J_F^T \\ J_F & -\mu I \end{pmatrix}.$$

It follows that if the active set at the QP solution $x+p$ is $\mathcal{A}(x)$ (i.e., $\mathcal{A}(x) = \mathcal{A}(x+p)$ and the QP does not change the active set), and $D_F = 0$ (which will hold near a solution that satisfies certain standard second-order optimality conditions), then the QP step is defined in terms of the (unmodified) matrices H_F and J_F.

6. Summary and Future Work. This paper considers the formulation and analysis of an SQP method for solving general nonlinear optimization problems. The algorithm is based on the natural pairing of a generalized primal-dual augmented Lagrangian function with a *flexible* line search. The global convergence result of Section 4 gives convergence without requiring a constraint qualification or nondegeneracy assumption. We believe that this is the best result that can be obtained.

The new algorithm, which we designate pdSQP, combines the favorable properties of augmented Lagrangian methods, stabilized SQP methods, and SQP methods. Strong global convergence is a property inherited from the augmented Lagrangian method. Fast local convergence is a property inherited from the SQP method. It remains to be seen if the superlinear convergence rate associated with stabilized SQP on degenerate problems is inherited by pdSQP. In a forthcoming paper [18], we establish quadratic convergence under the assumptions of strict complementarity, the Mangasarian-Fromovitz constraint qualification, and a certain second-order sufficient optimality condition. The local convergence of pdSQP under weaker conditions that are usually associated with stabilized SQP and augmented Lagrangian methods [41, 28, 42, 35, 12, 32] is also considered. In a separate paper [19] we will also consider issues that include modifications that allow convergence to second-order points, and the performance of pdSQP on practical problems.

Over the years, the use of exact second-derivatives in SQP methods has presented a significant challenge to the formulation of robust and efficient methods. A key result of this paper is the combination of a new convexification procedure and an inertia-controlling factorization to obtain a convex subproblem that is based on exact second derivatives. This approach provides an effective and efficient way of incorporating exact second derivatives in SQP methods.

One possible enhancement to pdSQP is the use of additional regularization in the form of *explicit* bounds on the dual variables in the QP subproblem. For reasons of brevity, this refinement is not considered here. However, explicit temporary bounds on the dual variables are easily incorporated (see, e.g., Robinson [38], and Gill and Robinson [22]).

The formulation of improved update strategies for the regularization parameter μ^R is the focus of current research. It is anticipated that such strategies will allow the use of projected gradient methods for the computation of an *approximate* solution of each QP subproblem when far from a solution. Approaches such as this should allow future implementations of pdSQP to solve larger problems than is possible by current SQP methods. In addition, reliable techniques that allow the rapid decrease of μ^R near a solution should give superlinear convergence under the standard assumptions.

REFERENCES

[1] P. T. Boggs and J. W. Tolle. Sequential quadratic programming. In *Acta Numerica, 1995*, volume 4 of *Acta Numer.*, pages 1–51. Cambridge Univ. Press, Cambridge, 1995.

[2] A. R. Conn, N. Gould, A. Sartenaer, and P. L. Toint. Convergence properties of an augmented Lagrangian algorithm for optimization with a combination of general equality and linear constraints. *SIAM J. Optim.*, 6(3):674–703, 1996.

[3] A. R. Conn, N. I. M. Gould, and P. L. Toint. Testing a class of method for solving minimization problems with simple bounds on the variables. Preprint, 1986.

[4] A. R. Conn, N. I. M. Gould, and P. L. Toint. Global convergence of a class of trust region algorithms for optimization with simple bounds. *SIAM J. Numer. Anal.*, 25:433–460, 1988.

[5] A. R. Conn, N. I. M. Gould, and P. L. Toint. A comprehensive description of LANCELOT. Technical Report 91/10, Département de Mathématique, Facultés Universitaires de Namur, 1991.

[6] A. R. Conn, N. I. M. Gould, and P. L. Toint. A globally convergent augmented Lagrangian algorithm for optimization with general constraints and simple bounds. *SIAM J. Numer. Anal.*, 28:545–572, 1991.

[7] A. R. Conn, N. I. M. Gould, and P. L. Toint. `LANCELOT`*: a Fortran package for large-scale nonlinear optimization (Release A)*. Lecture Notes in Computation Mathematics 17. Springer Verlag, Berlin, Heidelberg, New York, London, Paris and Tokyo, 1992.

[8] A. R. Conn, N. I. M. Gould, and P. L. Toint. Large-scale nonlinear constrained optimization. Technical Report 92/2, Département de Mathématique, Facultés Universitaires de Namur, 1992.

[9] L. B. Contesse. Une caractérisation complète des minima locaux en programmation quadratique. *Numer. Math.*, 34:315–332, 1980.

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

[11] D. Fernández and M. Solodov. Local convergence of exact and inexact augmented lagrangian methods under the second-order sufficiency condition. impa preprint a677, 2010.

[12] D. Fernández and M. Solodov. Stabilized sequential quadratic programming for optimization and a stabilized Newton-type method for variational problems. *Math. Program.*, 125(1, Ser. A):47–73, 2010.

[13] R. Fletcher. An ℓ_1 penalty method for nonlinear constraints. In P. T. Boggs, R. H. Byrd, and R. B. Schnabel, editors, *Numerical Optimization 1984*, pages 26–40, Philadelphia, 1985.

[14] R. Fletcher and S. Leyffer. User manual for filterSQP. Technical Report NA/181, Dept. of Mathematics, University of Dundee, Scotland, 1998.

[15] A. Forsgren. Inertia-controlling factorizations for optimization algorithms. *Appl. Num. Math.*, 43:91–107, 2002.

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

[17] A. Forsgren, P. E. Gill, and W. Murray. On the identification of local minimizers in inertia-controlling methods for quadratic programming. *SIAM J. Matrix Anal. Appl.*, 12:730–746, 1991.

[18] P. E. Gill, V. Kungurtsev, and D. P. Robinson. Convergence results for a stabilized sqp method. *In preparation*, 2012.

[19] P. E. Gill, V. Kungurtsev, and D. P. Robinson. Implementation of a stabilized sqp method. *In preparation*, 2012.

[20] P. E. Gill, W. Murray, and M. A. Saunders. SNOPT: An SQP algorithm for large-scale constrained optimization. *SIAM Rev.*, 47:99–131, 2005.

[21] P. E. Gill, W. Murray, M. A. Saunders, and M. H. Wright. Some theoretical properties of an augmented Lagrangian merit function. In P. M. Pardalos, editor, *Advances in Optimization and Parallel Computing*, pages 101–128. North Holland, North Holland, 1992.

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

[23] P. E. Gill and E. Wong. Methods for convex and general quadratic programming. Numerical Analysis Report 10-1, Department of Mathematics, University of California, San Diego, La Jolla, CA, 2010.

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

[25] N. I. M. Gould. On modified factorizations for large-scale linearly-constrained optimization. *SIAM J. Optim.*, 9:1041–1063, 1999.

[26] N. I. M. Gould and D. P. Robinson. A second derivative SQP method: Global convergence. *SIAM J. Optim.*, 20(4):2023–2048, 2010.

[27] N. I. M. Gould and D. P. Robinson. A second derivative SQP method: Local convergence and practical issues. *SIAM J. Optim.*, 20(4):2049–2079, 2010.

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

[29] S. P. Han. Superlinearly convergent variable metric algorithms for general nonlinear programming problems. *Math. Programming*, 11(3):263–282, 1976/77.

[30] M. R. Hestenes. Multiplier and gradient methods. *J. Optim. Theory Appl.*, 4:303–320, 1969.

[31] A. Izmailov and M. Solodov. Stabilized SQP revisited. *Mathematical Programming*, pages 1–28, 2010.

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

[33] N. Maratos. *Exact Penalty Function Algorithms for Finite-Dimensional and Control Optimization Problems*. PhD thesis, Department of Computing and Control, University of London, 1978.

[34] J. L. Morales, J. Nocedal, and Y. Wu. A sequential quadratic programming algorithm with an additional equality constrained phase. Technical Report OTC-05, Northwestern University, 2008.

[35] C. Oberlin and S. J. Wright. Active set identification in nonlinear programming. *SIAM J. Optim.*, 17(2):577–605, 2006.

[36] M. J. D. Powell. A method for nonlinear constraints in minimization problems. In R. Fletcher, editor, *Optimization*, pages 283–298, London and New York, 1969. Academic Press.

[37] M. J. D. Powell. A fast algorithm for nonlinearly constrained optimization calculations. Technical Report 77/NA 2, Department of Applied Mathematics and Theoretical Physics, University of Cambridge, England, 1977.

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

[39] K. Schittkowski. The nonlinear programming method of Wilson, Han, and Powell with an augmented Lagrangian type line search function. I. Convergence analysis. *Numer. Math.*, 38(1):83–114, 1981/82.

[40] K. Schittkowski. The nonlinear programming method of Wilson, Han, and Powell with an augmented Lagrangian type line search function. II. An efficient implementation with linear least squares subproblems. *Numer. Math.*, 38(1):115–127, 1981/82.

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

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