REGULARIZED SEQUENTIAL QUADRATIC PROGRAMMING METHODS

Philip E. Gill* Daniel P. Robinson†

UCSD Department of Mathematics
Technical Report NA-11-02
October 2011

Abstract

We present the formulation and analysis of a new sequential quadratic programming (SQP) method for general nonlinearly constrained optimization. The method pairs a primal-dual generalized augmented Lagrangian merit function with a *flexible* line search to obtain a sequence of improving estimates of the solution. This function is a primal-dual variant of the augmented Lagrangian proposed by Hestenes and Powell in the early 1970s. A crucial feature of the method is that the QP subproblems are convex, but formed from the exact second derivatives of the original problem. This is in contrast to methods that use a less accurate quasi-Newton approximation. Additional benefits of this approach include the following: (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, regularized methods, primal-dual methods.

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

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

1. Introduction

We present a sequential quadratic programming (SQP) method for optimization problems involving general linear and nonlinear constraints. The method is described in terms of the problem format:

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

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) easily carry over to the more general setting with $l \leq x \leq u$. The vector-pair (x^*, y^*) is called a first-order solution to problem (NP) if it satisfies

$$c(x^*) = 0 \;\text{ and }\; \min\left(x^*, z^*\right) = 0, \tag{1.1}$$

where y^* are the Lagrange multipliers associated with the constraints $c(x) = 0$, and z^* are the reduced costs at (x^*, y^*), i.e. $z^* = g(x^*) - J(x^*)^T y^*$

Sequential quadratic programming methods and interior methods are two alternative approaches to handling the inequality constraints in problem (NP). Sequential quadratic programming (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. Interior methods approximate a continuous path that passes through a solution of (NP). In the simplest case, the path is parameterized by a positive scalar parameter μ that may be interpreted as a perturbation for the optimality conditions for the problem (NP). Both interior methods and SQP methods have an inner/outer iteration structure, with the work for an inner iteration being dominated by the cost of solving a large sparse system of symmetric indefinite linear equations. In the case of SQP methods, these equations involve a subset of the variables and constraints; for interior methods, the equations involve all the constraints and variables.

SQP methods provide a relatively reliable "certificate of infeasibility" and they have the potential of being able to capitalize on a good initial starting point. Sophisticated matrix factorization updating techniques are used to 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 SQP methods so that exact second derivatives can 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 [11, 25]. The complexity of the QP subproblem has been a major impediment to the formulation of second-derivative SQP methods (although methods based on indefinite QP have been proposed [19, 20]). Over the years, algorithm developers have avoided

this difficulty by eschewing second derivatives and by solving a convex QP subproblem defined with a positive semidefinite quasi-Newton approximate Hessian (see, e.g., [28]); some authors enhance these basic methods with an additional subspace phase that incorporates exact second derivatives [33,34,40]. A difficulty with active-set methods is that they may require a substantial number of QP iterations when the outer iterates are far from the solution. The use of a QP subproblem is motivated by the assumption that the QP objective and constraints provide good "models" of the objective and constraints of problem (NP). This should make it unnecessary (and inefficient) to solve the QP to high accuracy during the preliminary iterations. Unfortunately, the simple expedient of limiting the number of inner iterations may have a detrimental effect upon reliability. An approximate QP solution may not predict a sufficient improvement in a merit function. Moreover, some of the QP multipliers will have the wrong sign if an active-set method is terminated before a solution is found. This may cause difficulties if the QP multipliers are used to estimate the multipliers for the nonlinear problem. These issues would largely disappear if a primal-dual *interior* method were to be used to solve the QP subproblem. These methods have the benefit of providing a sequence of feasible (i.e., correctly signed) dual iterates. Nevertheless, QP solvers based on conventional interior methods have had limited success within SQP methods because they are difficult to "warm start" from a near-optimal point (see the discussion below). This makes it difficult to capitalize on the property that, as the outer iterates converge, the solution of one QP subproblem is a very good estimate of the solution of the next.

Broadly speaking, the advantages and disadvantages of SQP methods and interior methods complement each other. Interior methods are most efficient when implemented with exact second derivatives. Moreover, they can converge in few inner iterations—even for very large problems. The inner iterates are the iterates of Newton's method for finding an approximate solution of the perturbed optimality conditions for a given μ. As the dimension and zero/nonzero structure of the Newton equations remains *fixed*, these Newton equations may be solved efficiently using either iterative or direct methods available in the form of advanced "off-the-shelf" linear algebra software. In particular, any new software for multicore and parallel architectures is immediately applicable. Moreover, the perturbation parameter μ plays an auxiliary role as an implicit regularization parameter of the linear equations. This implicit regularization plays a crucial role in the robustness of interior methods on ill-conditioned and ill-posed problems.

On the negative side, although interior methods are very effective for solving "one-off" problems, they are difficult to adapt to solving a sequence of related nonlinear problems. This difficulty may be explained in terms of the "path-following" interpretation of interior methods. In the neighborhood of an optimal solution, a step *along* the path $x(\mu)$ of perturbed solutions is well-defined, whereas a step *onto* the path from a neighboring point will be extremely sensitive to perturbations in the problem functions (and hence difficult to compute). Another difficulty with conventional interior methods is that a substantial number of iterations may be needed when the constraints are infeasible.

The idea of replacing a constrained optimization problem by a sequence of un-

constrained problems parameterized by a scalar μ has played a fundamental role in the formulation of algorithms since the early 1960s (for a seminal reference, see Fiacco and McCormick [16, 17]). One of the best-known methods for solving the equality-constrained problem (NEP) uses an unconstrained function based on the quadratic penalty function, which combines f with a term of order $1/\mu$ that "penalizes" the sum of the squares of the constraint violations. Under certain conditions (see, e.g., [17, 26, 49, 51]), the minimizers of the penalty function define a differentiable *trajectory* or *central path* that approaches the solution as $\mu \to 0$. Penalty methods approximate this path by minimizing the penalty function for a finite sequence of decreasing values of μ. In this form, the methods have a two-level structure of inner and outer iterations: the inner iterations are those of the method used to minimize the penalty function, and the outer iterations test for convergence and adjust the value of μ. As $\mu \to 0$, the Newton equations for minimizing the penalty function are increasingly ill-conditioned, and this ill-conditioning was perceived to be the reason for the poor numerical performance on some problems. In separate papers, Hestenes [36] and Powell [42] proposed the augmented Lagrangian function for (NEP), which is an unconstrained function based on augmenting the Lagrangian function with a quadratic penalty term that does not require μ to go to zero for convergence. The price that must be paid for keeping $1/\mu$ finite is the need to update estimates of the Lagrange multipliers in each outer iteration.

Since the first appearance of the Hestenes-Powell function, many algorithms have been proposed based on using the augmented Lagrangian as an objective function for sequential unconstrained minimization. Augmented Lagrangian functions have also been proposed that treat the multiplier vector as a continuous function of x; some of these ensure global convergence and permit local superlinear convergence (see, e.g., Fletcher [18]; DiPillo and Grippo [13]; Bertsekas [1, 2]; Boggs and Tolle [4]).

As methods for treating linear inequality constraints and bounds became more sophisticated, the emphasis of algorithms shifted from sequential unconstrained minimization to sequential linearly constrained minimization. In this context, the augmented Lagrangian has been used successfully within a number of different algorithmic frameworks for problem (NP). The method used in the software package LANCELOT [9] finds the approximate solution of a sequence of bound constrained problems with an augmented Lagrangian objective function. Similarly, the software package MINOS of Murtagh and Saunders [41] employs a variant of Robinson's linearly constrained Lagrangian (LCL) method [44] in which an augmented Lagrangian is minimized subject to the linearized nonlinear constraints. Friedlander and Saunders [27] define a globally convergent version of the LCL method that can treat infeasible constraints and infeasible subproblems. Augmented Lagrangian functions have also been used extensively as a merit function for sequential quadratic programming (SQP) methods (see, e.g., [3, 5, 7, 21, 28, 30, 45–48]).

The development of path-following interior methods for linear programming in the mid-1980s stimulated renewed interest in the treatment of constraints by sequential unconstrained optimization. This new attention not only resulted in a new understanding of the computational complexity of existing methods but also provided the impetus for the development of new approaches. A notable development was the

derivation of efficient path-following methods for linear programming based on applying Newton's method with respect to both the primal and dual variables. These new approaches also refocused attention on two computational aspects of penalty- and barrier-function methods for nonlinear optimization. First, the recognition of the formal equivalence between some primal-dual methods and conventional penalty methods indicated that the inherent ill-conditioning of penalty and barrier functions is not necessarily the reason for poor numerical performance. Second, the crucial role of penalty and barrier functions in problem regularization was recognized and better understood.

In this paper we formulate and analyze a new sequential quadratic programming (SQP) method for nonlinearly constrained optimization. The method pairs a primal-dual generalized augmented Lagrangian merit function with a *flexible* line search to obtain a sequence of improving estimates of the solution. This function is a primal-dual variant of the augmented Lagrangian proposed by Hestenes and Powell in the early 1970s. A crucial feature of the method is that the QP subproblems are convex, but formed from the exact second derivatives of the original problem. This is in contrast to methods that use a less accurate quasi-Newton approximation. Additional benefits of this approach include the following: (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. Preliminary numerical experiments on a subset of problems from the `CUTEr` test collection indicate that the proposed SQP method is significantly more efficient than our current SQP package `SNOPT`.

The paper is organized in five sections. Section 1 is a review of some of the basic properties of SQP methods. In Section 2, the steps of the 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 the stabilized SQP method [15, 35, 38, 50]. A rather general global convergence result is established in Section 4 that does not make any constraint qualification or non-degeneracy assumption.

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 $a \cdot b$. 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, repectively. The Hessian of the Lagrangian with respect to x is denoted by $H(x, y) = H(x) - \sum_{i=1}^{m} y_i H_i(x)$.

Background

Some of the most efficient algorithms for nonlinear optimization are sequential quadratic programming (SQP) methods. 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. Given a current 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

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

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

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

Given any $x \geq 0$, let $\mathcal{A}_0$ and $\mathcal{F}_0$ denote the index sets

$$\mathcal{A}_0(x) = \{i : x_i = 0\} \;\text{ and }\; \mathcal{F}_0(x) = \{1, 2, \ldots, n\}/\mathcal{A}_0(x), \tag{1.3}$$

If x is feasible for the constraints $c_k + J_k(x - x_k) = 0$, then $\mathcal{A}_0(x)$ is the *active set* at x. If the set $\mathcal{A}_0$ 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 $x \geq 0$ replaced by $x_i = 0$ for $i \in \mathcal{A}_0$. In general, the optimal $\mathcal{A}_0$ 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

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

where $\bar{H}_{\scriptscriptstyle F}$ is the matrix of free rows and columns of $\bar{H}_k$, $J_{\scriptscriptstyle F}$ 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 $\mathcal{A}_0$.

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 $(1, 2, \ldots, n)$ of a "working set" of indices that estimates the set $\mathcal{A}_0$ at a solution of (1.2). If $\mathcal{N}$ is a subset of $\mathcal{A}_0$, 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* add a positive-definite regularization term in the $(2, 2)$ block of (1.4). The magnitude of the regularization is generally based an heuristic arguments, which gives 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 merit 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, y^E is an estimate of an optimal Lagrange multiplier vector y^*. This function, proposed by Robinson [43], and Gill and Robinson [31], may be derived by applying the primal-dual penalty function of Forsgren and Gill [23] to a problem in which the constraints are shifted by a constant vector (see Powell [42]). 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

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

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

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

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 $\nu\geq 0$, we define

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

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 $H_M^\nu(x,y\,;\mu)$ 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_M^ν at the kth primal-dual iterate $v_k=(x_k,y_k)$, consider the convex QP subproblem

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

where $\mathcal{M}^\nu(v)$ denotes the merit function evaluated at v. 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 $\mathcal{F}_0(x_k+p_k)$ may be written in matrix form as:

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

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

$$s_i = \begin{cases} [\,x_k\,]_i & \text{if } i\in\mathcal{A}_0(x_k+p_k); \\ 0 & \text{if } i\in\mathcal{F}_0(x_k+p_k). \end{cases}$$

(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 $\mathcal{A}_0(x_k + p_k) = \mathcal{A}_0(x_k)$, 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 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 [8, 28].

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., [35, 50]). 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 [39] 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 [12] and define $\alpha_k = 2^{-j}$, where j is the smallest nonnegative integer such that

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

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

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

is a sufficiently negative real number that will allow us to prove global convergence of our proposed method. 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 is well-defined since 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 [31] 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

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

are feasibility and stationarity measures at the point (x, y), respectively. These optimality measures are based on the optimality conditions for problem (NP) rather than for minimizing the merit function $\mathcal{M}^\nu$. 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 $\phi_S(v_k) \leq \frac{1}{2}\phi_S^{\max}$ or $\phi_L(v_k) \leq \frac{1}{2}\phi_L^{\max}$, 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

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

has been found. The test for optimality is

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

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

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

where

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

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

$$\mu_{k+1} = \begin{cases} \mu_k, & \mathcal{M}^\nu(v_{k+1}\,; y_k^E, \mu_k) \leq \mathcal{M}^\nu(v_k\,; y_k^E, \mu_k) + \min(\alpha_{\min}, \alpha_k)\eta_S N_k \\ \max\big(\tfrac{1}{2}\mu_k, \mu_{k+1}^R\big), & \text{otherwise,} \end{cases} \tag{2.20}$$

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 $\alpha_{\min} > 0$, $\eta_S \in (0,1)$, $\tau_{\text{stop}} > 0$, 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 $k = 0, 1, 2, \ldots$ **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 $\|r_k\| \le \tau_{\text{stop}}$ **then** *exit*;
end (for)

---

3. Solution of the QP Subproblem

In this section we consider various theoretical and computational issues associated with the 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, an 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

In this section it is shown 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 steps (p, q) for given variables (x, y).

Theorem 3.1. *Consider the bound constrained QP*

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

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 $\nu \geq 0$. For the same quantities c, g, J and H, consider the stabilized QP problem

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

The following results hold.

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

(b) *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 (a), 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

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

The matrix $H + \frac{1}{\mu}J^TJ$ 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 (b), 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 (b) for the case $\nu > 0$. The optimality conditions for (3.1) are

$$\begin{aligned} g_M + H_M \Delta v &= \begin{pmatrix} z \\ 0 \end{pmatrix}, & z &\geq 0, \\ z \cdot (x + p) &= 0, & x + p &\geq 0. \end{aligned} \tag{3.4}$$

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

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

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

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

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 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, and 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 continue our study of these subproblems by considering the iterates that result when solved with an active-set method.

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

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

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 $\mathcal{Q}(x_{j+1}) \leq \mathcal{Q}(x_j)$ with $x_{j+1} = x_j + \alpha_j p_j$. Let the index sets $\mathcal{A}_0$ and $\mathcal{F}_0$ 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 $\mathcal{A}_0(x_j)$ 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 $\mathcal{A}_0(x_j)$ 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 $\mathcal{F}_0(x_j)$. 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., $x_j + p_j \geq 0$, 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

In this section we apply the active-set method to a QP of the form

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

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

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

with $H + \frac{1}{\mu}J^TJ$ positive definite, $\nu \geq 0$, 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

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

and

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

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 $\mathcal{F}_0(x_j)$ 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 QP (3.10). Then, for every $\nu \geq 0$, 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. First, we 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

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

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

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

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 $\mathcal{F}_0(x_k+p_k)$, then (p_F, q_k) satisfies the equations*

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

where $\mathcal{F}$ is defined in terms of the set $\mathcal{F}_0(x_k + p_k)$ and s is a nonnegative vector such that

$$s_i = \begin{cases} [\, x_k \,]_i & \text{if } i \in \mathcal{A}_0(x_k + p_k); \\ 0 & \text{if } i \in \mathcal{F}_0(x_k + p_k). \end{cases}$$

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:

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

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 "$\bar{x}$" $= (\bar{x}, \bar{y})$,

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

(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

$$\begin{pmatrix} H_F & 0 & -J_F^T \\ 0 & \mu I & -\mu I \\ J_F & \mu I & 0 \end{pmatrix} \begin{pmatrix} p_F \\ \bar{p}_F \\ q_j \end{pmatrix} = - \begin{pmatrix} [\, g + H(x_j - \bar{x}) - J^T w_j \,]_F \\ \mu y_j - \mu w_j \\ c + \mu(y_j - y^E) + J(x_j - \bar{x}) \end{pmatrix}, \tag{3.16}$$

where p_F and $\bar{p}_F$ 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 $\bar{p}_F = q_j - y_j + w_j$, which implies that

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

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

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

$$\begin{pmatrix} H_F & -J_F^T \\ J_F & \mu I \end{pmatrix} \begin{pmatrix} p_F \\ q_j \end{pmatrix} = - \begin{pmatrix} [\, g + H(x_j - \bar{x}) - J^T y_j \,]_F \\ c + \mu(y_j - y^E) + J(x_j - \bar{x}) \end{pmatrix}, \tag{3.17}$$

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 $\nu \geq 0$, there exists a positive $\bar{\mu}$ such that, for all $0 < \mu < \bar{\mu}$, the active-set method generates identical primal-dual iterates $\{(x_j, y_j)\}_{j \geq 0}$.* ∎

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 $\{\bar{H}(x_k, y_k)\}_{k \geq 0}$ is bounded, with $\{\bar{H}(x_k, y_k) + (1/\mu_k^R) J(x_k)^T J(x_k)\}_{k \geq 0}$ uniformly positive definite.*

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

Assumption 4.3. *The sequence $\{x_k\}_{k \geq 0}$ 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 [9, 10, 43]. Consequently, it is possible that y_k^E and μ_k^R will remain fixed over a sequence of iterations, although this is rare in practice. Nonetheless, our convergence result must consider this situation, which we now investigate.

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

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

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

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

$$\nabla \mathcal{M}^\nu(v_k\,; y_k^E, \mu_k^R)^T \Delta v_k \leq -\epsilon \;\; \text{for all } k \geq \widehat{k}.$$

Proof. The assumptions of this theorem guarantee that

$$\tau_k \equiv \tau > 0, \;\; \mu_k^R = \mu^R, \;\; \text{and} \;\; y_k^E = y^E \;\; \text{for all } k \geq \widehat{k}. \tag{4.1}$$

We first prove part (i). As in the proof of Theorem 3.1, we know that the solution to (2.9) satisfies

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

and w^* is the unique solution of

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

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

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

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

for all $k \in \mathcal{S}_1$. We may then conclude from the definition of H_M^ν, 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 μ_k^R 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). Using the matrix

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

the fact that $\Delta v = 0$ is feasible for the convex problem (2.9), that Δv_k is the solution to problem (2.9) for $\nu > 0$ chosen in Algorithm 2.1, (4.1), and Assumption 4.1, it

follows that

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

for some $\lambda_{\min} > 0$. Combining this with (4.2) we deduce that

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

from which $\lim_{k\in\mathcal{S}_2} q_k = 0$ follows from Assumptions 4.2 and 4.3. This contradicts part (ii), which shows 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) satisfying (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.
Using the definition of an S-iterate, M-iterates, and the functions ϕ_S, and ϕ_L we may conclude that the number of S-iterates and L-iterates must be finite. We now claim that there must be an infinite number of M-iterates. To prove this, we 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 that the update to μ_k^R given by (2.18) and the assumption of this case, that μ_k^R is eventually never decreased any further, and that

the update to μ_k given by (2.20) implies that μ_k is also eventually fixed. Gathering these facts gives the existence of an integer $\widehat{k}$ such that

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

It then follows from (2.20) that

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

where N_k is defined by (2.13). Moreover, parts (ii) and (iii) of Theorem 4.1 ensures 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, we first note that parts (i) and (iii) of Theorem 4.1 and Assumption 4.2 would ensure that $\{\alpha_k\}_{k\geq\widehat{k}}$ is bounded away from zero if a *standard* Armijo line search was 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 α_k that we actually compute can be no smaller since the actual definition of N_k is less restrictive and we use a flexible line search that makes step acceptance more likely. Combining these facts with (4.3), we conclude that

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

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. A contradiction has been reached so there exists infinitely many M-iterations, and *all* iterates are M-iterates and F-iterates for all 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. Part (ii) of Theorem 4.2, i.e., generating infinitely many M-iterates, is the fall-back position of Algorithm 2.1. We believe this result is the best that can be expected since we have not assumed any constraint qualification. In fact, the assumptions we have made does no preclude the possibility that problem NP is infeasible. Also, it has recently been proved [14,15,37] that iterates generated from the stabilized SQP subproblem exhibit superlinear convergence under rather mild conditions; in particular, strict complementarity is not assumed and no constraint qualification is required.

5. Conclusions

In this paper we developed and analyzed 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.

In Section 4 a global convergence result was provided without requiring any constraint qualification or non-degeneracy assumption; for the general problem class considered, we believe this is the best result one may obtain.

The new algorithm—called pdSQP—combines the favorable properties of augmented Lagrangian, stabilized SQP, and SQP methods. The strong global convergence properties (convergence with no constraint qualification or non-degeneracy assumption) are inherited from the augmented Lagrangian method. For degenerate problems, it remains to be seen whether the *theoretical* superlinear convergence of iterates derived from stabilized SQP subproblems is inherited by pdSQP and observed in *practice*.

The effective use of exact second-derivatives in SQP methods has been a significant challenge. The structure of algorithm pdSQP allows the use of inertia-controlling factorizations [22, 24] to obtain a modified second derivative matrix that is less ad-hoc than traditional schemes. Thus, we believe that pdSQP is a new SQP method that effectively and efficiently incorporates second derivatives.

One possible enhancement to pdSQP is additional regularization in the form of *explicit* bounds on the dual variables in each subproblem. For simplicity, this refinement was not considered here, but bounds on the dual variables are easily incorporated. (See [31, 43] for examples of how "artificial" bounds on the dual variables may be included.)

A carefully designed update strategy for the regularization parameter μ^R used in Algorithm pdSQP is the focus of current research. We anticipate that such strategies will allow projected gradient methods to compute an *approximate* solution of each subproblem (when far from a solution); ultimately, this may allow pdSQP to be "scaled-up" to solve problems larger than previously possible by other SQP methods. Moreover, rapid and careful decrease of μ^R near a solution should allow superlinear convergence of the iterates under standard assumption.

References

[1] D. P. Bertsekas. *Constrained optimization and Lagrange multiplier methods*. Computer Science and Applied Mathematics. Academic Press Inc. [Harcourt Brace Jovanovich Publishers], New York, 1982. 4

[2] D. P. Bertsekas. *Constrained optimization and Lagrange multiplier methods*. Athena Scientific, Belmont, Massachusetts, 1996. 4

[3] P. T. Boggs, A. J. Kearsley, and J. W. Tolle. A global convergence analysis of an algorithm for large-scale nonlinear optimization problems. *SIAM J. Optim.*, 9(4):833–862, 1999. 4

[4] P. T. Boggs and J. W. Tolle. A family of descent functions for constrained optimization. *SIAM J. Numer. Anal.*, 21:1146–1161, 1984. 4

[5] 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. 4

[6] I. Bongartz, A. R. Conn, N. I. M. Gould, and P. L. Toint. CUTE: Constrained and unconstrained testing environment. *ACM Trans. Math. Software*, 21(1):123–160, 1995.

[7] R. H. Byrd, R. A. Tapia, and Y. Zhang. An SQP augmented Lagrangian BFGS algorithm for constrained optimization. *SIAM J. Optim.*, 20:210–241, 1992. 4

[8] 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. 9

[9] 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. 4, 20

[10] 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. 20

[11] L. B. Contesse. Une caractérisation complète des minima locaux en programmation quadratique. *Numer. Math.*, 34:315–332, 1980. 2

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

[13] G. DiPillo and L. Grippo. A new class of augmented Lagrangians in nonlinear programming. *SIAM J. Control Optim.*, 17:618–628, 1979. 4

[14] D. Fernández and M. Solodov. Local convergence of exact and inexact augmented lagrangian methods under the second-order sufficiency condition. impa preprint a677, 2010. 23

[15] 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. 5, 23

[16] A. V. Fiacco and G. P. McCormick. *Nonlinear Programming: Sequential Unconstrained Minimization Techniques*. John Wiley and Sons, Inc., New York-London-Sydney, 1968. 4

[17] A. V. Fiacco and G. P. McCormick. *Nonlinear Programming*. Classics in Applied Mathematics. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, second edition, 1990. Reprint of the 1968 original. 4

[18] R. Fletcher. A class of methods for nonlinear programming with termination and convergence properties. In J. Abadie, editor, *Integer and Nonlinear Programming*, pages 157–175. North-Holland, The Netherlands, 1970. 4

[19] 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. 2

[20] R. Fletcher and S. Leyffer. User manual for filterSQP. Technical Report NA/181, Dept. of Mathematics, University of Dundee, Scotland, 1998. 2

[21] R. Fontecilla, T. Steihaug, and R. A. Tapia. A convergence theory for a class of quasi-Newton methods for constrained optimization. *SIAM J. Numer. Anal.*, 24:1133–1151, 1987. 4

[22] A. Forsgren. Inertia-controlling factorizations for optimization algorithms. *Appl. Num. Math.*, 43:91–107, 2002. 24

[23] A. Forsgren and P. E. Gill. Primal-dual interior methods for nonconvex nonlinear programming. *SIAM J. Optim.*, 8:1132–1152, 1998. 7

[24] A. Forsgren, P. E. Gill, and W. Murray. On the identification of local minimizers in inertia-controlling methods for quadratic programming, August 29–September 2 1988. Presented at the 13th International Symposium on Mathematical Programming, Tokyo, Japan. 24

[25] 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. 2

[26] A. Forsgren, P. E. Gill, and M. H. Wright. Interior methods for nonlinear optimization. *SIAM Rev.*, 44:525–597, 2002. 4

[27] M. P. Friedlander and M. A. Saunders. A globally convergent linearly constrained Lagrangian method for nonlinear optimization. *SIAM J. Optim.*, 15(3):863–897, 2005. 4

[28] P. E. Gill, W. Murray, and M. A. Saunders. SNOPT: An SQP algorithm for large-scale constrained optimization. *SIAM Rev.*, 47:99–131, 2005. 3, 4, 9

[29] P. E. Gill, W. Murray, and M. A. Saunders. User's guide for SNOPT Version 7: Software for large-scale nonlinear programming. Numerical Analysis Report 06-2, Department of Mathematics, University of California, San Diego, La Jolla, CA, 2006.

[30] 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. 4

[31] P. E. Gill and D. P. Robinson. A primal-dual augmented Lagrangian. *Computational Optimization and Applications*, pages 1–25, 2010. 7, 10, 24

[32] N. I. M. Gould, D. Orban, and P. L. Toint. CUTEr and SifDec: A constrained and unconstrained testing environment, revisited. *ACM Trans. Math. Software*, 29(4):373–394, 2003.

[33] N. I. M. Gould and D. P. Robinson. A second derivative SQP method: Global convergence. *SIAM J. Optim.*, 20(4):2023–2048, 2010. 3

[34] 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. 3

[35] 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. 5, 9

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

[37] A. Izmailov and M. Solodov. Stabilized sqp revisited. *Mathematical Programming*, pages 1–28, 2010. 23

[38] 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. 5

[39] N. Maratos. *Exact Penalty Function Algorithms for Finite-Dimensional and Control Optimization Problems*. PhD thesis, Department of Computing and Control, University of London, 1978. 10

[40] 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. 3

[41] B. A. Murtagh and M. A. Saunders. MINOS 5.5 User's Guide. Report SOL 83-20R, Department of Operations Research, Stanford University, Stanford, CA, Revised 1998. 4

[42] 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. 4, 7

[43] D. P. Robinson. *Primal-Dual Methods for Nonlinear Optimization*. PhD thesis, Department of Mathematics, University of California, San Diego, September 2007. 7, 20, 24

[44] S. M. Robinson. A quadratically-convergent algorithm for general nonlinear programming problems. *Math. Program.*, 3:145–156, 1972. 4

[45] 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. 4

[46] 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. 4

[47] K. Schittkowski. On the convergence of a sequential quadratic programming method with an augmented Lagrangian line search function. *Math. Operationsforsch. Statist. Ser. Optim.*, 14(2):197–216, 1983. 4

[48] R. A. Tapia. Diagonalized multiplier methods and quasi-Newton methods for constrained optimization. *J. Optim. Theory Appl.*, 22:135–194, 1977. 4

[49] M. H. Wright. Interior methods for constrained optimization. In *Acta Numerica, 1992*, pages 341–407. Cambridge University Press, New York, USA, 1992. 4

[50] S. J. Wright. Superlinear convergence of a stabilized SQP method to a degenerate solution. *Comput. Optim. Appl.*, 11(3):253–275, 1998. 5, 9

[51] S. J. Wright and D. Orban. Properties of the log-barrier function on degenerate nonlinear programs. *Math. Oper. Res.*, 27(3):585–613, 2002. 4