CONVEXIFICATION SCHEMES FOR SQP METHODS

Philip E. Gill* Elizabeth Wong*

UCSD Center for Computational Mathematics
Technical Report CCoM-14-06
July 18, 2014

Abstract

Sequential quadratic programming (SQP) methods solve nonlinear optimization problems by finding an approximate solution of a sequence of quadratic programming (QP) subproblems. Each subproblem involves the minimization of a quadratic model of the objective function subject to the linearized constraints. Depending on the definition of the quadratic model, the QP subproblem may be nonconvex, leading to difficulties in the formulation and analysis of a conventional SQP method.

Convexification is a process for defining a local convex approximation of a nonconvex problem. We describe three forms of convexification: preconvexification, concurrent convexification, and post-convexification. The methods require only minor changes to the algorithms used to solve the QP subproblem, and are designed so that modifications to the original problem are minimized and applied only when necessary.

Key words. nonlinear programming, nonlinear inequality constraints, sequential quadratic programming, second-derivative 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, elwong@ucsd.edu). Research supported in part by Northrop Grumman Aerospace Systems, and National Science Foundation grant DMS-1318480.

1. Introduction

This paper concerns the formulation of a sequential quadratic programming (SQP) method for the solution of the nonlinear optimization problem:

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

where $c\colon \mathbb{R}^n \mapsto \mathbb{R}^m$ and $f\colon \mathbb{R}^n \mapsto \mathbb{R}$ are twice-continuously differentiable. In order to simplify the notation, it is assumed that the constraints are all inequalities with zero lower bounds. However, the method to be described can be generalized to treat all forms of linear and nonlinear constraints. No assumptions are made about f and c (other than twice differentiability), which implies that the problem need not be convex. The vector $g(x)$ denotes the gradient of $f(x)$, and $J(x)$ denotes the $m \times n$ constraint Jacobian, which has ith row $\nabla c_i(x)^T$, the gradient of the ith constraint function $c_i(x)$. The Lagrangian function associated with (NP) is $L(x,y) = f(x) - c(x)^T y$, where y is the m-vector of dual variables associated with the inequality constraints $c(x) \ge 0$. The Hessian of the Lagrangian with respect to x is denoted by $H(x,y) = \nabla^2 f(x) - \sum_{i=1}^m y_i \nabla^2 c_i(x)$.

Sequential quadratic programming 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. In a line-search SQP method, the QP solution provides a direction of improvement for a merit function that represents a compromise between the (usually conflicting) aims of minimizing the objective function and minimizing the constraint violations.

Given the kth estimate (x_k, y_k) of the primal and dual solution, a conventional SQP method defines a direction $p_k = \widehat{x}_k - x_k$, where $\widehat{x}_k$ is a solution of the QP subproblem

$$\begin{aligned} \underset{x\in\mathbb{R}^n}{\text{minimize}} \quad & g(x_k)^T(x - x_k) + \tfrac{1}{2}(x - x_k)^T H(x_k, y_k)(x - x_k) \\ \text{subject to} \quad & J(x_k)(x - x_k) \ge -c(x_k), \end{aligned} \tag{1.1}$$

where the QP objective function is a quadratic approximation of $f(x)$ restricted to the surface $c(x) = 0$. If the QP subproblem (1.1) has a bounded solution, then the first-order optimality conditions imply the existence of a primal-dual pair $(\widehat{x}_k, \widehat{y}_k)$ such that

$$g(x_k) + H(x_k, y_k)(\widehat{x}_k - x_k) = J(x_k)^T \widehat{y}_k, \qquad \widehat{y}_k \ge 0, \tag{1.2}$$

$$r(\widehat{x}_k) \cdot \widehat{y}_k = 0, \qquad r(\widehat{x}_k) \ge 0, \tag{1.3}$$

where $r(x)$ is the vector of constraint residuals $r(x) = c(x_k) + J(x_k)(x - x_k)$, and $a \cdot b$ denotes the vector with ith component $a_i b_i$. At any feasible point x, the active set associated with the QP subproblem is given by

$$\mathcal{A}(x) = \{\, i : r_i(x) = [c(x_k) + J(x_k)(x - x_k)]_i = 0 \,\}.$$

The optimality conditions (1.1) may be characterized in terms of an index set $\mathcal{W} \subseteq \mathcal{A}(\widehat{x}_k)$ such that the rows of $J(x_k)$ with indices in $\mathcal{W}$ are linearly independent. It can be shown that the conditions (1.2)–(1.3) may be written in the form

$$\begin{aligned} g(x_k) + H(x_k, y_k)(\widehat{x}_k - x_k) &= J_w(x_k)^T \widehat{y}_w, & \widehat{y}_w &\ge 0, \\ c_w(x_k) + J_w(x_k)(\widehat{x}_k - x_k) &= 0, & r(\widehat{x}_k) &\ge 0, \end{aligned}$$

where $c_w(x_k)$ and $J_w(x_k)$ denote the rows of $c(x_k)$ and $J(x_k)$ associated with indices in $\mathcal{W}$. The vector $\widehat{y}_w$ is the subvector of $\widehat{y}_k$ such that $[\widehat{y}_k]_i = 0$ for $i \notin \mathcal{W}$, and $[\widehat{y}_k]_w = \widehat{y}_w$. These conditions may be written in matrix form

$$\begin{pmatrix} H(x_k, y_k) & J_w(x_k)^T \\ J_w(x_k) & 0 \end{pmatrix} \begin{pmatrix} p_k \\ -\widehat{y}_w \end{pmatrix} = -\begin{pmatrix} g(x_k) \\ c_w(x_k) \end{pmatrix}, \tag{1.4}$$

where $p_k = \widehat{x}_k - x_k$. The working set $\mathcal{W}$ is said to be *second-order consistent* with respect to $H(x_k, y_k)$ if the reduced Hessian $Z_w^T H(x_k, y_k) Z_w$ is positive definite, where the columns of Z_w form a basis for the null-space of $J_w(x_k)$. If $\mathcal{W}$ is second-order consistent with respect to $H(x_k, y_k)$, then the equations (1.4) are nonsingular and define unique vectors p_k and $\widehat{y}_w$ that satisfy

$$p_k^T H(x_k, y_k) p_k = -\big(g(x_k) - J_w(x_k)^T \widehat{y}_w\big)^T p_k = -g_L(x_k, \widehat{y}_k)^T p_k,$$

where $g_L(x, y)$ denotes the gradient of the Lagrangian function with respect to x, i.e., $g_L(x, y) = g(x) - J(x)^T y$. If $H(x_k, y_k)$ is positive definite, then $g_L(x_k, \widehat{y}_k)^T p_k < 0$ and the QP search direction is a descent direction for the Lagrangian function defined with multipliers $y = \widehat{y}_k$. The condition $p_k^T H(x_k, y_k) p_k > 0$ is sufficient for there to exist a step length that provides a sufficient decrease for several merit functions that have been proposed in the literature; e.g., the ℓ_1 penalty function (Han [20] and Powell [22]), and various forms of the augmented Lagrangian merit function (Han [20], Schittkowski [23], and Gill, Murray, Saunders and Wright [11]).

If the problem (NP) is not convex, the Hessian of the Lagrangian may be indefinite, even in the neighborhood of a solution. This situation creates a number of difficulties in the formulation and analysis of a conventional SQP method.

(i) In the nonconvex case, the QP subproblem (1.1) may be nonconvex, which implies that the objective may be unbounded below in the feasible region, and that there may be many local solutions. In addition, nonconvex QP is NP-hard—even for the calculation of a local minimizer [1, 6]. 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 [2, 3]).

(ii) If $H(x_k, y_k)$ is not positive definite, then p_k may not be a descent direction for the merit function. This implies that an alternative direction must be found or the line search must allow the merit function to increase on some iterations, see, e.g., Grippo, Lampariello and Lucidi [17–19], Toint [27], and Zhang and Hager [29]).

Over the years, algorithm developers have avoided these difficulties by solving a convex QP subproblem defined with a positive semidefinite quasi-Newton approximate Hessian. In this form, SQP methods have proved reliable and efficient for many such problems. For example, under mild conditions the general-purpose solvers NLPQL [24], NPSOL [10, 11], DONLP [26], and SNOPT [9] typically find a (local) optimum from an arbitrary starting point, and they require relatively few evaluations of the problem functions and gradients.

2. Convexification

Convexification is a process for defining a local convex approximation of a nonconvex problem. This approximation may be defined on the full space of variables or on just some subset. Many model-based optimization methods use some form of convexification. For example, line-search methods for unconstrained and linearly-constrained optimization define a convex local quadratic model in which the Hessian $H(x_k)$ is replaced by a positive-definite matrix $H(x_k) + E_k$ (see, e.g., Greenstadt [16], Gill and Murray [8], Schnabel and Eskow [25], and Forsgren and Murray [7]). All of these methods are based on convexifying an unconstrained or equality-constrained local model. In this paper we consider a method that convexifies the inequality-constrained subproblem directly. The method extends some approaches proposed by Gill and Robinson [12, Section 4] and Kungurtsev [21].

In the context of SQP methods, the purpose of the convexification is to find a primal-dual pair $(x_k, \widehat{y}_k)$ and matrix $\mathit{\Delta} H_k$ such that

$$p_k^T\big(H(x_k, y_k) + \mathit{\Delta} H_k\big)p_k \ge \bar{\gamma} p_k^T p_k,$$

where $\bar{\gamma}$ is a fixed positive scalar that defines a minimum acceptable value of the curvature of the Lagrangian. Ideally, any algorithm for computing $\mathit{\Delta} H_k$ should satisfy two requirements. First, the convexification should be "minimally invasive", i.e., if $H(x_k, y_k)$ is positive definite or $p_k^T H(x_k, y_k)p_k \ge \bar{\gamma} p_k^T p_k$, then $\mathit{\Delta} H_k$ should be zero. Second, it must be possible to store the modification $\mathit{\Delta} H_k$ implicitly, without the need to modify the elements of $H(x_k, y_k)$.

The convexification discussed here can take three forms: preconvexification, concurrent convexification, and post-convexification. Not all of these forms are needed at a given iteration.

2.1. Concurrent QP Convexification

The concurrent convexification algorithm is defined in terms of a generic QP of the form

$$\begin{aligned} &\underset{x\in\mathbb{R}^n}{\text{minimize}} && \varphi(x) = g^T(x - x_I) + \tfrac{1}{2}(x - x_I)^T H(x - x_I) \\ &\text{subject to} && Ax \ge Ax_I - b, \end{aligned} \tag{2.1}$$

where x_I, b, A, g, and H are constant. In the SQP context, $x_I = x_k$, $g = g(x_k)$, $b = c(x_k)$, $A = J(x_k)$, and H is the Hessian of the Lagrangian or as an approximation of it. Thus, the objective is not necessarily convex and the QP subproblem may be indefinite. To avoid indefinite subproblems, we apply a concurrent convexification method that is designed to minimize the modifications to the Hessian.

Concurrent convexification is based on applying a modified active-set method to the QP problem (2.1). The method of Gill and Wong [14] is a two-phase method for general (i.e., nonconvex) QP. In the first phase, the objective function is ignored while a conventional phase-1 linear program is used to find a feasible point x_0 for the constraints $Ax \ge Ax_I - b$. On completion of the first phase, a working set $\mathcal{W}_0$ is available that contains the indices of a linearly independent subset of the constraints that are active at x_0. If A_0 denotes the $m_0 \times n$ matrix of rows of A with indices in $\mathcal{W}_0$, then

$$A_0 x_0 = A_0 x_I - b_0. \tag{2.2}$$

In the second phase, a sequence of primal-dual iterates $\{(x_j, y_j)\}_{j\ge 0}$, and linearly independent working sets $\{\mathcal{W}_j\}$ are generated such that: (i) $\{x_j\}_{j\ge 0}$ is feasible; (ii) $\varphi(x_j) \le \varphi(x_{j-1})$; and (iii) for every $j \ge 1$, (x_j, y_j) is the primal and dual solution of the equality constrained problem defined by minimizing $\varphi(x)$ on the working set $\mathcal{W}_j$. The vector x_j associated with the primal-dual pair (x_j, y_j) is known as a *subspace minimizer* with respect to $\mathcal{W}_j$. If A_j denotes the

$m_j \times n$ matrix of rows of A with indices in $\mathcal{W}_j$, then a subspace minimizer is formally defined as the point x_j such that $g(x_j) = A_j^T y_j$, and the KKT matrix

$$K_j = \begin{pmatrix} H & A_j^T \\ A_j & 0 \end{pmatrix} \tag{2.3}$$

has m_j negative eigenvalues. For any K_j satisfying this property, the working set $\mathcal{W}_j$ is said to be *second-order consistent with respect to* H.

In general, the first iterate x_0 will not minimize $\varphi(x)$ on $\mathcal{W}_0$, and one or more preliminary iterations are needed to find the first subspace minimizer x_1. An estimate of x_1, is defined by solving the equality-constrained QP subproblem:

$$\underset{x}{\text{minimize}} \;\; \varphi(x) \quad \text{subject to} \quad A_0(x - x_I) + b_0 = 0. \tag{2.4}$$

If the KKT matrix K_0 is second-order consistent, then the solution of this subproblem is given by $x_0 + p_0$, where p_0 satisfies the nonsingular equations

$$\begin{pmatrix} H & A_0^T \\ A_0 & 0 \end{pmatrix} \begin{pmatrix} p_0 \\ -\widehat{y}_0 \end{pmatrix} = - \begin{pmatrix} g(x_0) \\ b_0 + A_0(x_0 - x_I) \end{pmatrix} = - \begin{pmatrix} g(x_0) \\ 0 \end{pmatrix}, \tag{2.5}$$

If $x_0 + p_0$ is feasible, then $(x_1, y_1) = (x_0 + p_0, y_0)$, otherwise one of the constraints violated at $x_0 + p_0$ is added to the working set and the iteration is repeated. Eventually, the working set will include enough constraints to define an appropriate primal-dual pair (x_1, y_1).

If the first subspace minimizer x_1 is not optimal, then the method proceeds to find the sequence of subspace minimizers x_2, x_3, ..., described above. At any given iteration, not all the constraints in $\mathcal{W}_j$ are necessarily active at x_j. If every working-set constraint *is* active, then $\mathcal{W}_j \subseteq \mathcal{A}(x_j)$, and x_j is called a *standard* subspace minimizer; otherwise x_j is a *nonstandard* subspace minimizer. The method is formulated so that there is a subsequence of "standard" iterates intermixed with a finite number of consecutive "nonstandard" iterates. If the multipliers y_j are nonnegative at a standard iterate, then x_j is optimal for (2.1) and the algorithm is terminated. Otherwise, the working set constraint with a negative multiplier is identified and designated as the *nonbinding working-set constraint* associated with the subsequent consecutive sequence of nonstandard iterates. If the index of the nonbinding constraint corresponds to row s of A, then $[y_j]_s < 0$. There follows a sequence of "intermediate" iterations in which the constraint $a_s^T x \geq a_s^T x_I - b_s$ remains in the working set, but its multiplier is driven to zero. At each of these iterations, a search direction is defined by solving the equality-constrained subproblem

$$\underset{p \in \mathbb{R}^n}{\text{minimize}} \;\; \varphi(x_j + p) \quad \text{subject to} \quad a_i^T p = \begin{cases} 0 & \text{if } i \neq s,\; i \in \mathcal{W}_j, \\ 1 & \text{if } i = s. \end{cases} \tag{2.6}$$

In order to simplify the discussion it is assumed that H has not been modified in any previous iteration. In matrix form, the optimality conditions for the subproblem (2.6) are

$$\begin{pmatrix} H & A_j^T \\ A_j & 0 \end{pmatrix} \begin{pmatrix} p_j \\ -q_j \end{pmatrix} = \begin{pmatrix} 0 \\ e_s \end{pmatrix}, \tag{2.7}$$

where $y_j + q_j$ are the multipliers at the minimizer $x_j + p_j$, and e_s denotes sth column of the identity matrix. (In order to simplify the notation, it is assumed that the nonbinding constraint corresponds to the sth row of A, which implies that a_s^T is the sth row of both A and A_j.) Any nonzero step along p_j increases the residual of the nonbinding constraint while maintaining the residuals of the other working-set constraints at zero (i.e., the nonbinding constraint becomes inactive while the other working-set constraints remain active).

Once the direction (p_j, q_j) has been computed, the computation of the next iterate x_{j+1} depends on the value of $p_j^T H p_j$, the curvature of φ along p_j. There are two cases to consider.

Case 1: $p_j^T H p_j > 0$**.** If the curvature is positive along p_j, then the QP iteration is completed without modification. This will always be the outcome when φ is convex. In this case, the step to the minimizer of φ along the search direction p_j is given by

$$\alpha_j^* = -g(x_j)^T p_j / p_j^T H p_j = -[y_j]_s / p_j^T H p_j. \tag{2.8}$$

The definition of α_j^* implies that the multiplier $[y_j + \alpha_j^* q_j]_s$ associated with the nonbinding constraint at $x_j + \alpha_j^* p_j$ is zero. This implies that if $x_j + \alpha_j^* p_j$ is feasible with respect to the constraints that are not in the working set, then the nonbinding constraint index can be removed from $\mathcal{W}_j$ without changing the multiplier associated with the other (active) working-set constraints. This gives a new standard iterate $x_{j+1} = x_j + \alpha_j^* p_j$, with working set $\mathcal{W}_{j+1} = \mathcal{W}_j \setminus \{s\}$. Either x_{j+1} is optimal or a new nonbinding constraint is identified and the process is repeated. If $x_j + \alpha_j^* p_j$ is not feasible, then x_{j+1} is defined as $x_j + \alpha_j p_j$, where α_j is the smallest step that gives a feasible $x_j + \alpha_j p_j$. The point x_{j+1} must have at least one constraint that is active but not in $\mathcal{W}_j$. If t is the index of this constraint, and a_t and the vectors $\{a_i\}_{i \in \mathcal{W}_j}$ are linearly independent, then t is added to the working set to give $\mathcal{W}_{j+1}$. At the next iteration, a new value of (p_j, q_j) is computed using the equations (2.7) defined with A_{j+1}. If a_t and $\{a_i\}_{i \in \mathcal{W}_j}$ are linearly dependent, then it is shown in [14] that the working set $\mathcal{W}_{j+1} = \{\mathcal{W}_j \setminus \{s\}\} \cup \{t\}$ defined by replacing the index t with index s, is linearly independent. Moreover, $x_{j+1} = x_j + \alpha_j p_j$ is a subspace minimizer with respect to $\mathcal{W}_{j+1}$.

Case 2: $p_j^T H p_j \le 0$**.** In this case H is not positive definite and the QP Hessian is modified so that it has sufficiently large positive curvature along p_j. If $p_j^T H p_j \le 0$, then $\varphi(x_j + \alpha p_j)$ is unbounded below for positive values of α. In this case, either the unmodified QP is unbounded, or there exists a constraint index t and a nonnegative step $\widehat{\alpha}_j$ such that the constraint residuals satisfy $r_t(x_j + \widehat{\alpha}_j p_j) = 0$, $r(x_j + \widehat{\alpha}_j p_j) \ge 0$, and $\widehat{\alpha}_j$ minimizes $\varphi(x_j + \alpha p_j)$ for all feasible $x_j + \alpha p_j$.

If $p_j^T H p_j < 0$, a positive semidefinite rank-one matrix $\sigma a_s a_s^T$ is added to H implicitly. This modifies the quadratic program that is being solved, but the current iterate x_j remains a subspace minimizer for the modified problem. The only computed quantities that are altered by the modification are the curvature and the multiplier y_s associated with the nonbinding working-set constraint. The modified Hessian is defined as $H(\bar{\sigma}) = H + \bar{\sigma} a_s a_s^T$ for some $\bar{\sigma} > 0$. Gill and Wong [14] show that the curvature $p_j^T H p_j$ is nondecreasing during a sequence of nonstandard iterations associated with a nonbinding index s. This implies that a modification of the Hessian will occur only at the first nonstandard iterate.

For an arbitrary σ, the gradient of the modified objective at x_j is

$$g + H(\sigma)(x_j - x_I) = g + (H + \sigma a_s a_s^T)(x_j - x_I).$$

As (x_j, y_j) is a standard subspace minimizer for the unmodified problem, the identities $g(x_j) = g + H(x_j - x_I) = A_j^T y_j$ and $a_s^T(x_j - x_I) = -b_s$ hold, and the gradient of the modified objective is given by

$$\begin{aligned} g + H(\sigma)(x_j - x_I) &= g + H(x_j - x_I) + \sigma a_s a_s^T(x_j - x_I) \\ &= g(x_j) + \sigma a_s^T(x_j - x_I) a_s \\ &= A_j^T\big(y_j - \sigma b_s e_s\big) = A_j^T y(\sigma), \quad \text{with} \quad y(\sigma) = y_j - \sigma b_s e_s. \end{aligned}$$

This implies that x_j is a subspace minimizer of the modified problem for all $\sigma \ge 0$. Moreover, the multipliers of the modified problem are the same as those of the unmodified problem except for the multiplier y_s associated with the nonbinding constraint, which is shifted by $-\sigma b_s$.

Once the Hessian is modified, the equations (2.7) for the primal-dual direction become

$$\begin{pmatrix} H + \bar\sigma a_s a_s^T & A_j^T \\ A_j & 0 \end{pmatrix} \begin{pmatrix} \bar p_j \\ -\bar q_j \end{pmatrix} = \begin{pmatrix} 0 \\ e_s \end{pmatrix},$$

which are equivalent to

$$\begin{pmatrix} H & A_j^T \\ A_j & 0 \end{pmatrix} \begin{pmatrix} p_j \\ -(\bar q_j - \bar\sigma e_s) \end{pmatrix} = \begin{pmatrix} 0 \\ e_s \end{pmatrix}.$$

A comparison with (2.7) yields

$$\bar p_j = p_j \quad \text{and} \quad \bar q_j = q_j + \bar\sigma e_s.$$

which implies that the QP direction is not changed by the modification.

For any $\sigma \ge 0$, let $\alpha_j(\sigma)$ denote the step associated with the search direction for the modified QP. The identities $a_s^T p_j = 1$ and $a_s^T(x_j - x_I) = -b_s$ imply that

$$\begin{aligned} \alpha_j(\sigma) &= -\frac{\big(g + (H + \sigma a_s a_s^T)(x_j - x_I)\big)^T p_j}{p_j^T(H + \sigma a_s a_s^T) p_j} \\ &= -\frac{g(x_j)^T p_j + \sigma a_s^T(x_j - x_I)}{p_j^T H p_j + \sigma} \\ &= -\frac{g(x_j)^T p_j - \sigma b_s}{p_j^T H p_j + \sigma} = -\frac{y_s - \sigma b_s}{p_j^T H p_j + \sigma} = -\frac{y_s(\sigma)}{p_j^T H p_j + \sigma}. \end{aligned} \tag{2.9}$$

This implies that $\bar\sigma$ must satisfy

$$\bar\sigma > \sigma_{\min} = -p_j^T H p_j.$$

The derivative of $\alpha_j(\sigma)$ with respect to σ is given by

$$\alpha_j'(\sigma) = \frac{1}{(p_j^T H p_j + \sigma)^2}\Big(y_s + b_s p_j^T H p_j\Big) = \frac{y_s(\sigma_{\min})}{(p_j^T H p_j + \sigma)^2}. \tag{2.10}$$

The choice of $\bar\sigma$ depends on two scalar parameters $y_{\mathtt{tol}}$ and $d_{\max}$. The scalar $d_{\max}$ defines the maximum change in x at each QP iteration. The scalar $y_{\mathtt{tol}}$ is the dual optimality tolerance and is used to define what is meant by a "nonoptimal" multiplier. In particular, the nonbinding multiplier must satisfy $y_s < -y_{\mathtt{tol}}$ in order to qualify as being nonoptimal.

There are two cases to consider for the choice of $\bar\sigma$.

Case (i): $b_s < 0$. In this case, $y_s(\sigma)$ is an increasing function of σ, which implies that there exists $\sigma_{\texttt{opt}} = (y_s - y_{\texttt{tol}})/b_s > 0$ such that $y_s(\sigma_{\texttt{opt}}) = y_{\texttt{tol}} > 0$. This modification changes the multiplier associated with the nonbinding constraint from nonoptimal to optimal. However, if $\sigma_{\texttt{opt}} < \sigma_{\min}$, then the curvature is not sufficiently positive and σ must be increased so that it is larger than $\sigma_{\texttt{opt}}$. The definition

$$\bar{\sigma} = \begin{cases} \sigma_{\texttt{opt}} & \text{if } \sigma_{\texttt{opt}} \geq 2\sigma_{\min}; \\ 2\sigma_{\min} & \text{if } \sigma_{\texttt{opt}} < 2\sigma_{\min}, \end{cases}$$

guarantees that the curvature along p_j is sufficiently positive with an optimal modified multiplier $y_s(\bar{\sigma})$. In either case, the QP algorithm proceeds by selecting an alternative nonbinding constraint without taking a step along p_j.

If $b_s < 0$ and $y_s(\sigma_{\min}) < 0$, then $y_s(\sigma)$ increases from the negative value of $y_s(\sigma_{\min})$ to $-y_{\texttt{tol}}$ as σ increases from $\sigma_{\min}$ to the positive value $\sigma_{\texttt{nonopt}} = (y_s + y_{\texttt{tol}})/b_s$. This implies that if σ is chosen in the range $\sigma_{\min} < \sigma \leq \sigma_{\texttt{nonopt}}$, then the multiplier for the nonbinding constraint remains nonoptimal, and it is possible to both convexify and keep the current nonbinding constraint. However, in the SQP context it is unusual for a nonbinding constraint to have a negative value of b_s when x_k is far from a solution. For an SQP subproblem, b is the vector $c(x_k)$, and a negative value of b_s implies that the sth nonlinear constraint is violated at x_k. The linearization of a violated nonlinear constraint is likely to be retained in the working set because the SQP step is designed to reduce the nonlinear constraint violations. The picture changes when x_k is close a solution and the violations of the nonlinear constraints in the QP working set are small. In this case, if strict complementarity does not hold at the solution of the nonlinear problem and x_k is converging to a point that satisfies the second-order necessary conditions, but not the second-order sufficient conditions, then both b_s and y_s may be small and negative. It is for this reason that even if $y_s(\sigma_{\min})$ is negative, $\bar{\sigma}$ is chosen large enough that the multiplier changes sign and the nonbinding constraint is retained in the QP working set.

Case (ii): $b_s \geq 0$. In this case, $y_s(\sigma_{\min}) = y_s - b_s\sigma_{\min} < 0$ and $y_s(\sigma_{\min})$ decreases monotonically for all increasing $\sigma > \sigma_{\min}$. The step-length function $\alpha_j(\sigma)$ has a pole at $\sigma = -p_j^T H p_j$ and decreases monotonically, with $\alpha_j(\sigma) \to b_s \geq 0$ as $\sigma \to +\infty$. The behavior of $x(\sigma)$ is depicted in Figure 1 for a two-variable QP with constraints $a^T(x - x_I) \geq -b$, $x_1 \geq 0$, and $x_2 \geq 0$. The next iterate of the QP algorithm lies on the ray $x(\sigma) = x_j + \alpha_j(\sigma)p_j$. As $\sigma \to \infty$, $x(\sigma)$ moves closer to the point $x_j + b_s p_j$ on the hyperplane $a^T(x - x_I) = 0$.

A preliminary value of $\bar{\sigma}$ is chosen to provide a change of variables such that

$$\|x_{j+1} - x_j\|_2 \leq d_{\max},$$

where $d_{\max}$ is the preassigned maximum change in x at each QP iteration. If $\alpha_T = d_{\max}/\|p_j\|_2$, then the substitution of $\alpha_j(\bar{\sigma}) = \alpha_T$ in (2.9) gives $\bar{\sigma} = -(y_s + \alpha_T p_j^T H p_j)/(\alpha_T - b_s)$. However, the limit $\alpha_j(\sigma) \to b_s \geq 0$ as $\sigma \to +\infty$, implies that this value of $\bar{\sigma}$ may be large if $\alpha_j(\bar{\sigma})$ is close to b_s. In order to avoid this difficulty, the value of $\bar{\sigma}$ is used as long as the associated value of $\alpha_j(\bar{\sigma})$ is sufficiently larger than b_s, i.e.,

$$\alpha_j(\bar{\sigma}) = \begin{cases} \alpha_T & \text{if } \alpha_T \geq 2b_s; \\ 2b_s & \text{if } \alpha_T < 2b_s, \end{cases} \quad \text{so that} \quad \bar{\sigma} = \begin{cases} -\dfrac{y_s + \alpha_T p_j^T H p_j}{\alpha_T - b_s} & \text{if } \alpha_T \geq 2b_s, \\ -\dfrac{y_s + 2b_s p_j^T H p_j}{b_s} & \text{if } \alpha_T < 2b_s. \end{cases}$$

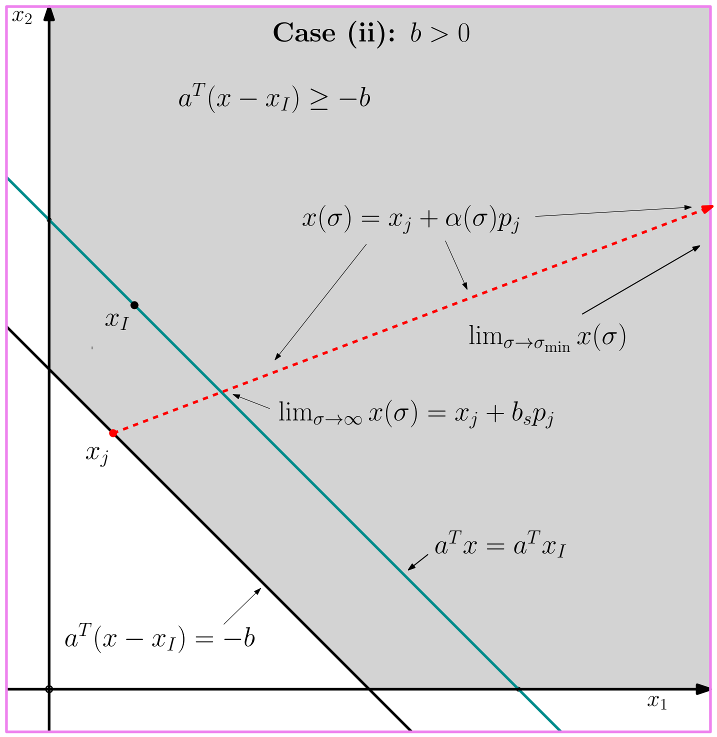

Figure 1: The figures depict a QP with constraints $a^T(x - x_I) \geq -b$, $x_1 \geq 0$, and $x_2 \geq 0$. The point x_j is a standard subspace minimizer with working-set constraint $a^T(x - x_I) \geq -b$. The surface of the hyperplane $a^T(x - x_I) = 0$ is marked in green. The QP base point x_I is feasible for $b \geq 0$. The QP search direction is the red dotted line. The next iterate of the QP algorithm lies on the ray $x(\sigma) = x_j + \alpha_j(\sigma)p_j$. As σ increases from its initial value of $\sigma_{\min}$, the new iterate $x(\sigma)$ moves closer to the point $x_j + b_s p_j$ on the hyperplane $a^T(x - x_I) = 0$.

Overall, if this algorithm is applied to a nonconvex QP of the form (2.1), then a solution is found for the convexified QP,

$$\begin{aligned} \underset{x\in\mathbb{R}^n}{\text{minimize}} \quad & \varphi(x) = g^T(x - x_I) + \tfrac{1}{2}(x - x_I)^T(H + E)(x - x_I) \\ \text{subject to} \quad & Ax \geq Ax_I - b, \end{aligned} \tag{2.11}$$

where E is a positive-semidefinite matrix of the form $E = A^T\bar{\Sigma}A$, with $\bar{\Sigma}$ a positive semidefinite diagonal matrix. In general, most of the diagonal elements of $\bar{\Sigma}$ are zero. The modification E may be reconstructed from A and a sparse representation of $\bar{\Sigma}$.

2.2. Preconvexification

The concurrent convexification method of Section 2.1 has the property that if x_0 is a subspace minimizer, then all subsequent iterates are subspace minimizers. Methods for finding an initial subspace minimizer utilize an initial estimate x_0 of the solution together with an initial working set $\mathcal{W}_0$ of linearly independent constraints. These estimates are often available from a phase-one linear program or, in the SQP context, the solution of the previous QP subproblem.

If a potential KKT matrix K_0 has too many negative or zero eigenvalues, then $\mathcal{W}_0$ is not a second-order consistent working set. In this case, an appropriate K_0 may be obtained by imposing temporary constraints that are deleted during the course of the subsequent QP iterations. For example, if n variables are temporarily fixed at their current values, then A_0 is the identity matrix and K_0 necessarily has exactly n negative eigenvalues regardless of the eigenvalues of $H(x_k, y_k)$. The form of the temporary constraints depends on the method used to solve the KKT equations, see, e.g., Gill and Wong [14, Section 6]. Once the temporary constraints are imposed, concurrent convexification can proceed as in Section 2.1 as the temporary constraints are removed from the working set during subsequent iterations.

A disadvantage of using temporary constraints is that it may be necessary to factor two KKT matrices if the initial working set is not second-order consistent. An alternative approach is to utilize the given working set $\mathcal{W}_0$ without modification and use *preconvexification*, which involves the definition of a positive-semidefinite E_0 such that the matrix

$$K_0 = \begin{pmatrix} H + E_0 & A_0^T \\ A_0 & 0 \end{pmatrix} \tag{2.12}$$

is second-order consistent. A suitable modification E_0 may be based on some variant of the symmetric indefinite or block-triangular factorizations of K_0. Appropriate methods include: (i) the inertia controlling LBL^T factorization (Forsgren [4], Forsgren and Gill [5]); (ii) an LBL^T factorization with pivot modification (Gould [15]); and (iii) a conventional LBL^T factorization of (2.12) with $E_0 = \sigma I$ for some nonnegative scalar σ (Wächter and Biegler [28]). In each case, the modification E_0 is zero if $\mathcal{W}_0$ is already second-order consistent.

2.3. Post-convexification

As concurrent convexification generates a sequence of second-order-consistent working sets, the SQP search direction $p_k = \widehat{x}_k - x_k$ must satisfy the second-order-consistent KKT system

$$\begin{pmatrix} H_k + E_k & J_w(x_k)^T \\ J_w(x_k) & 0 \end{pmatrix} \begin{pmatrix} p_k \\ -\widehat{y}_w \end{pmatrix} = -\begin{pmatrix} g(x_k) \\ c_w(x_k) \end{pmatrix}, \tag{2.13}$$

where $H_k = H(x_k, y_k)$, E_k is the matrix defined by the pre- and/or concurrent convexification, and $c_w(x_k)$ and $J_w(x_k)$ are the rows of $c(x_k)$ and $J(x_k)$ associated with indices in the final QP working set $\mathcal{W}$ (cf. (1.4)). In most cases, concurrent convexification is sufficient to give $p_k^T(H_k + E_k)p_k > 0$, but it may hold that $p_k^T(H_k + E_k)p_k \le 0$. In this case, p_k is not a descent direction for $g_L(x_k, \widehat{y}_k)$, and an additional *post-convexification step* is necessary. In the following discussion, there is no loss of generality in assuming that $E_k = 0$, i.e., it is assumed that H_k has not been modified during the preconvexification or concurrent convexification stages. Post-convexification is based on the following result.

Result 2.1. *If J_w is a second-order-consistent working-set matrix associated with a symmetric H, then there exists a nonnegative $\bar{\sigma}$ such that the matrix $\bar{H} = H + \bar{\sigma} J_w^T J_w$ is positive definite. In addition, the solution of the equations*

$$\begin{pmatrix} H & J_w^T \\ J_w & 0 \end{pmatrix} \begin{pmatrix} p \\ -\widehat{y}_w \end{pmatrix} = -\begin{pmatrix} g \\ c_w \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} \bar{H} & J_w^T \\ J_w & 0 \end{pmatrix} \begin{pmatrix} \bar{p} \\ -\bar{y}_w \end{pmatrix} = -\begin{pmatrix} g \\ c_w \end{pmatrix}$$

are related by the identities $\bar{p} = p$ and $\bar{y}_w = \widehat{y}_w - \bar{\sigma} c_w$. ∎

If the solution $(\widehat{x}_k, \widehat{y}_k)$ of the QP subproblem does not satisfy the descent condition, then $p_k = \widehat{x}_k - x_k$ is such that

$$p_k^T H(x_k, y_k) p_k = -g_L(x_k, \widehat{y}_k)^T p_k < \bar{\gamma} p_k^T p_k,$$

for some positive $\bar{\gamma}$. The result implies that multipliers $\bar{y}_k$ such that $[\bar{y}_k]_i = 0$, for $i \notin \mathcal{W}$, and $[\bar{y}_k]_w = \widehat{y}_w - \bar{\sigma} c_w(x_k)$, provide the required curvature

$$p_k^T \bar{H}(x_k, y_k) p_k = -g_L(x_k, \bar{y}_k)^T p_k = \gamma p_k^T p_k,$$

where $\bar{\sigma} = \big(\gamma p_k^T p_k - p_k^T H(x_k, y_k) p_k\big) / \|c_w(x_k)\|^2$ with γ chosen such that $\gamma \ge \bar{\gamma}$. The extension of this result to the situation where $(\widehat{x}_k, \widehat{y}_k)$ satisfy the modified KKT equations (2.13) is obvious.

3. Summary

Convexification algorithms are proposed for the QP subproblem in an SQP method for nonlinearly constrained optimization. Three forms of convexification are defined: preconvexification, concurrent convexification, and post-convexification. The methods require only minor changes to the algorithms used to solve the QP subproblem, and are designed so that modifications to the original problem are minimized and applied only when necessary.

It should be noted that the post-convexification Result 2.1 holds even if a conventional general QP method is used to solve the QP subproblem (provided that the method gives a final working set that is second-order consistent). It follows that post-convexification will define a descent direction regardless of whether or not concurrent convexification is used. The purpose of concurrent convexification is to reduce the probability of needing post-convexification, and to avoid the difficulties associated with solving an indefinite QP problem.

The methods defined here are the basis of the second-derivative solver in the dense SQP package DNOPT of Gill, Saunders and Wong [13]. All of the methods may be extended to problems in which the constraints are written in so-called standard form: $c(x) = 0$ and $x \ge 0$ (see Gill and Wong [14, Section 4]). In this case, the inequality constraints for the QP subproblem are simple bounds $x \ge 0$, and all the modification matrices are diagonal.

References

[1] L. B. Contesse. Une caractérisation complète des minima locaux en programmation quadratique. *Numer. Math.*, 34:315–332, 1980. 3

[2] 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. SIAM. 3

[3] R. Fletcher and S. Leyffer. User manual for filterSQP. Technical Report NA/181, Dept. of Mathematics, University of Dundee, Scotland, 1998. 3

[4] A. Forsgren. Inertia-controlling factorizations for optimization algorithms. *Appl. Num. Math.*, 43:91–107, 2002. 10

[5] A. Forsgren and P. E. Gill. Primal-dual interior methods for nonconvex nonlinear programming. *SIAM J. Optim.*, 8:1132–1152, 1998. 10

[6] 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. 3

[7] A. Forsgren and W. Murray. Newton methods for large-scale linear equality-constrained minimization. *SIAM J. Matrix Anal. Appl.*, 14:560–587, 1993. 4

[8] P. E. Gill and W. Murray. Newton-type methods for unconstrained and linearly constrained optimization. *Math. Program.*, 7:311–350, 1974. 4

[9] 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

[10] P. E. Gill, W. Murray, M. A. Saunders, and M. H. Wright. User's guide for NPSOL (Version 4.0): a Fortran package for nonlinear programming. Report SOL 86-2, Department of Operations Research, Stanford University, Stanford, CA, 1986. 3

[11] 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. 3

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

[13] P. E. Gill, M. A. Saunders, and E. Wong. User's Guide for DNOPT: a Fortran package for medium-scale nonlinear programming. Center for Computational Mathematics Report CCoM 14-05, Department of Mathematics, University of California, San Diego, La Jolla, CA, 2014. 11

[14] P. E. Gill and E. Wong. Methods for convex and general quadratic programming. Center for Computational Mathematics Report CCoM 14-03, University of California, San Diego, La Jolla, CA, 2014. 4, 6, 10, 11

[15] N. I. M. Gould. On modified factorizations for large-scale linearly constrained optimization. *SIAM J. Optim.*, 9:1041–1063, 1999. 10

[16] J. Greenstadt. On the relative efficiencies of gradient methods. *Math. Comp.*, 21:360–367, 1967. 4

[17] L. Grippo, F. Lampariello, and S. Lucidi. Newton-type algorithms with nonmonotone line search for large-scale unconstrained optimization. In *System modelling and optimization (Tokyo, 1987)*, volume 113 of *Lecture Notes in Control and Inform. Sci.*, pages 187–196. Springer, Berlin, 1988. 3

[18] L. Grippo, F. Lampariello, and S. Lucidi. A truncated Newton method with nonmonotone line search for unconstrained optimization. *J. Optim. Theory Appl.*, 60(3):401–419, 1989. 3

[19] L. Grippo, F. Lampariello, and S. Lucidi. A class of nonmonotone stabilization methods in unconstrained optimization. *Numer. Math.*, 59(8):779–805, 1991. 3

[20] S. P. Han. A globally convergent method for nonlinear programming. *J. Optim. Theory Appl.*, 22:297–309, 1977. 3

[21] V. Kungurtsev. *Second-Derivative Sequential Quadratic Programming Methods for Nonlinear Optimization*. PhD thesis, Department of Mathematics, University of California San Diego, La Jolla, CA, 2013. 4

[22] M. J. D. Powell. A fast algorithm for nonlinearly constrained optimization calculations. In G. A. Watson, editor, *Numerical Analysis, Dundee 1977*, number 630 in Lecture Notes in Mathematics, pages 144–157, Heidelberg, Berlin, New York, 1978. Springer Verlag. 3

[23] 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. 3

[24] K. Schittkowski. NLPQL: a Fortran subroutine for solving constrained nonlinear programming problems. *Ann. Oper. Res.*, 11:485–500, 1985/1986. 3

[25] R. B. Schnabel and E. Eskow. A new modified Cholesky factorization. *SIAM J. Sci. and Statist. Comput.*, 11:1136–1158, 1990. 4

[26] P. Spellucci. An SQP method for general nonlinear programs using only equality constrained subproblems. *Math. Program.*, 82:413–448, 1998. 3

[27] P. L. Toint. An assessment of nonmonotone linesearch techniques for unconstrained optimization. *SIAM J. Sci. Comput.*, 17(3):725–739, 1996. 3

[28] A. Wächter and L. T. Biegler. Line search filter methods for nonlinear programming: motivation and global convergence. *SIAM J. Optim.*, 16(1):1–31 (electronic), 2005. 10

[29] H. Zhang and W. W. Hager. A nonmonotone line search technique and its application to unconstrained optimization. *SIAM J. Optim.*, 14(4):1043–1056 (electronic), 2004. 3