Methods for Convex and General Quadratic Programming*

Philip E. Gill† Elizabeth Wong†

UCSD Department of Mathematics
Technical Report NA-10-01
September 2010

Abstract

Computational methods are considered for finding a point that satisfies the second-order necessary conditions for a general (possibly nonconvex) quadratic program (QP). The first part of the paper defines a framework for the formulation and analysis of feasible-point active-set methods for QP. This framework defines a class of methods in which a primal-dual search pair is the solution of an equality-constrained subproblem involving a "working set" of linearly independent constraints. This framework is discussed in the context of two broad classes of active-set method for quadratic programming: *binding-direction methods* and *nonbinding-direction methods*. We recast a binding-direction method for general QP first proposed by Fletcher, and subsequently modified by Gould, as a nonbinding-direction method. This reformulation gives the primal-dual search pair as the solution of a KKT-system formed from the QP Hessian and the working-set constraint gradients. It is shown that, under certain circumstances, the solution of this KKT-system may be updated using a simple recurrence relation, thereby giving a significant reduction in the number of KKT systems that need to be solved. Furthermore, the nonbinding-direction framework is applied to QP problems with constraints in standard form, and to the dual of a convex QP.

The second part of the paper focuses on implementation issues. First, two approaches are considered for solving the constituent KKT systems. The first approach uses a variable-reduction technique requiring the calculation of the Cholesky factor of the reduced Hessian. The second approach uses a symmetric indefinite factorization of a fixed KKT matrix in conjunction with the factorization of a smaller matrix that is updated at each iteration. Finally, algorithms for finding an initial point for the method are proposed. In particular, single-phase methods involving a linearly constrained augmented Lagrangian are proposed that obviate the need for a separate procedure for finding a feasible point.

Key words. Large-scale quadratic programming, active-set methods, convex quadratic programming, dual quadratic program, nonconvex quadratic programming, KKT systems.

*Research supported in part by National Science Foundation grants DMS-0511766 and DMS-0915220, and by Department of Energy grant DE-SC0002349.

†Department of Mathematics, University of California, San Diego, La Jolla, CA 92093-0112 (pgill@ucsd.edu, elwong@ucsd.edu).

1. Introduction

The *quadratic programming* (QP) problem is to minimize a quadratic objective function subject to linear constraints on the variables. Quadratic programs arise in many areas, including economics, applied science and engineering. Important applications of quadratic programming include portfolio analysis, support vector machines, structural analysis and optimal control. Quadratic programming also forms a principal computational component of many sequential quadratic programming (SQP) methods for nonlinear programming (for a recent survey, see Gill and Wong [17]).

In the first part of the paper (comprising Sections 2 and 3), we review the optimality conditions for QP and defines a framework for the formulation and analysis of feasible-point active-set methods for QP. This framework defines a class of methods in which a primal-dual search pair is the solution of an equality-constrained subproblem involving a "working set" of linearly independent constraints. This framework is discussed in the context of two broad classes of active-set method for quadratic programming: *binding-direction methods* and *nonbinding-direction methods*. Broadly speaking, the working set for a binding direction method consists of a subset of the active constraints, whereas the working set for a nonbinding direction method may involve constraints that need not be active (nor even feasible). We recast a binding-direction method for general QP first proposed by Fletcher, and subsequently modified by Gould, as a nonbinding-direction method. This reformulation gives the primal-dual search pair as the solution of a KKT-system formed from the QP Hessian and the working-set constraint gradients. It is shown that, under certain circumstances, the solution of this KKT-system may be updated using a simple recurrence relation, thereby giving a significant reduction in the number of KKT systems that need to be solved.

The linear constraints of a QP may include an arbitrary mixture of equality and inequality constraints, where the inequality constraints may be subject to lower and/or upper bounds. Many mathematically equivalent formulations of the constraints are possible, and the choice of formulation often depends on the context. We consider the generic quadratic program

$$\begin{array}{ll} \underset{x\in\mathbb{R}^n}{\text{minimize}} & \varphi(x) = c^Tx + \tfrac{1}{2}x^THx \\ \text{subject to} & Ax = b, \qquad Dx \ge f, \end{array} \tag{1.1}$$

where A, b, c, D, f and H are constant, H is symmetric, A is $m \times n$, and D is $m_D \times n$. (In order to simplify the notation, it is assumed that the inequalities involve only lower bounds.) However, the methods to be described can be generalized to treat all forms of linear constraints. No assumptions are made about H (other than symmetry), which implies that the objective $\varphi(x)$ need not be convex. In the nonconvex case, however, convergence will be to local minimizers only.

In Section 4, the nonbinding direction method is extended to problems with constraints in standard form, which is an example of the generic form (1.1) where the inequalities are the nonnegativity constraints $x \ge 0$. It is shown that if $H = 0$, the method is equivalent to a variant of the primal simplex method in which the π-values and reduced costs are updated at each iteration. Section 5 focuses on the convex case and considers the application of the nonbinding direction method to the QP dual. The resulting method does not require the assumption of strict convexity and gives a method equivalent to the dual simplex method when $H = 0$.

Section 6 considers two alternative approaches for solving the KKT systems. The first involves the symmetric transformation of the KKT system into three smaller systems, one of which involves the explicit reduced Hessian matrix. The second approach uses a symmetric indefinite factorization of a fixed KKT matrix in conjunction with the factorization of a smaller matrix that is updated at each iteration. The use of a fixed factorization allows an "off-the shelf" sparse equation solver to be used repeatedly. This feature is ideally suited to problems with structure that can be exploited by a specialized factorization. Moreover, improvements in efficiency derived from exploiting new parallel and vector computer architectures are immediately applicable.

Finally, Sections 7 and 8 consider algorithms for finding an initial point for the nonbinding-direction method. Two single-phase methods are proposed that use the active-set framework of Section 2.

1.1. Notation

The vector $g(x)$ denotes $c + Hx$, the gradient of the objective φ evaluated at x. The vector d_i^T refers to the i-th row of the constraint matrix D, so that the i-th inequality constraint is $d_i^T x \geq f_i$. The i-th component of a vector labeled with a subscript will be denoted by $[\,\cdot\,]_i$, e.g., $[v_N]_i$ is the i-th component of the vector v_N. Similarly, a subvector of components with indices in the index set $\mathcal{S}$ is denoted by $(\,\cdot\,)_{\mathcal{S}}$, e.g., $(v_N)_{\mathcal{S}}$ is the vector with components $[v_N]_i$ for $i \in \mathcal{S}$. The symbol I is used to denote an identity matrix with dimension determined by the context. The j-th column of I is denoted by e_j. 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 i-th component $a_i b_i$ is denoted by $a \cdot b$. Given symmetric $K = \begin{pmatrix} M & N^T \\ N & G \end{pmatrix}$, with M nonsingular, the matrix $G - NM^{-1}N^T$, the Schur complement of M in K, will be denoted by K/M. We sometimes refer simply to "the" Schur complement when the relevant matrices are clear.

2. Background

In this section, we review the optimality conditions for the generic QP (1.1), and describe a framework for the formulation of feasible-point active-set QP methods. No assumptions are made about H (other than symmetry), which implies that the objective $\varphi(x)$ need not be convex. Throughout, it is assumed that the matrix A has full row-rank m. This condition is easily satisfied for the class of active-set methods considered in this paper. Given an arbitrary matrix G, equality constraints $Gu = b$ are equivalent to the full rank constraints $Gu + v = b$, if we impose $v = 0$. In this formulation, the v-variables are artificial variables that are fixed at zero.

2.1. Optimality conditions

The necessary and sufficient conditions for a local solution of the QP (1.1) involve the existence of vectors z and π of Lagrange multipliers associated with the constraints $Dx \geq f$ and $Ax = b$, respectively. The conditions are summarized by the following result, which is stated without proof (see, e.g., Borwein [4], Contesse [6] and Majthay [22]).

Result 2.1. (QP optimality conditions) *The point x is a local minimizer of the quadratic program (1.1) if and only if*

(a) *$Ax = b$, $Dx \geq f$, and there exists at least one pair of vectors π and z such that $g(x) = A^T\pi + D^Tz$, with $z \geq 0$, and $z \cdot (Dx - f) = 0$;*

(b) *$p^T Hp \geq 0$ for all nonzero p satisfying $g(x)^T p = 0$, $Ap = 0$, and $d_i^T p \geq 0$ for every i such that $d_i^T x = f_i$.* ∎

We follow the convention of refering to any x that satisfies condition (a) as a first-order KKT point.

If H has at least one negative eigenvalue and (x, π, z) satisfies condition (a) with an index i such that $z_i = 0$ and $d_i^T x = f_i$, then x is known as a dead point. Verifying condition (b) at a dead point requires finding the global minimizer of an indefinite quadratic form over a cone, which is an NP-hard problem (see, e.g., Cottle, Habetler and Lemke [7], Pardalos and Schnitger [24], and Pardalos and Vavasis [25]). This implies that the optimality of a candidate solution of a general quadratic program can be verified only if more restrictive (but computationally tractable) sufficient conditions are satisfied. A dead point is a point at which the sufficient conditions are not satisfied, but certain necessary conditions for optimality hold. Computationally tractable necessary conditions are based on the following result.

Result 2.2. (Necessary conditions for optimality) *The point x is a local minimizer of the QP (1.1) only if*

(a) *$Ax = b$, $Dx \ge f$, and there exists at least one pair of vectors π and z such that $g(x) = A^T\pi + D^T z$, with $z \ge 0$, and $z \cdot (Dx - f) = 0$;*

(b) *it holds that $p^T H p \ge 0$ for all nonzero p satisfying $Ap = 0$, and $d_i^T p = 0$ for each i such that $d_i^T x = f_i$.* ∎

Suitable sufficient conditions for optimality are given by (a)–(b) with (b) replaced by the condition that $p^T H p \ge 0$ for all p such that $Ap = 0$, and $d_i^T p = 0$ for every $i \in \mathcal{A}_+(x)$, where $\mathcal{A}_+(x)$ is the index set $\mathcal{A}_+(x) = \{i : d_i^T x = f_i \text{ and } z_i > 0\}$.

These conditions may be expressed in terms of the constraints that are satisfied with equality at x. Let x be any point satisfying the equality constraints $Ax = b$. (The assumption that A has rank m implies that there must exist at least one such x.) An inequality constraint is active at x if it is satisfied with equality. The indices associated with the active constraints comprise the active set, denoted by $\mathcal{A}(x)$. An active-constraint matrix $A_\mathfrak{a}(x)$ is a matrix with rows consisting of the rows of A and the gradients of the active constraints. By convention, the rows of A are listed first, giving the active-constraint matrix

$$A_\mathfrak{a}(x) = \begin{pmatrix} A \\ D_\mathfrak{a}(x) \end{pmatrix},$$

where $D_\mathfrak{a}(x)$ comprises the rows of D with indices in $\mathcal{A}(x)$. Note that the active-constraint matrix includes A in addition to the gradients of the active constraints. The argument x is generally omitted if it is clear where $D_\mathfrak{a}$ is defined.

With this definition of the active set, we give an equivalent statement of Result 2.2.

Result 2.3. (Necessary conditions in active-set form) *Let the columns of the matrix $Z_\mathfrak{a}$ form a basis for the null-space of $A_\mathfrak{a}$. The point x is a local minimizer of the QP (1.1) only if*

(a) *x is a first-order KKT point, i.e.,* (i) *$Ax = b$, $Dx \ge f$;* (ii) *$g(x)$ lies in* range$(A_\mathfrak{a}^T)$, *or equivalently, there exist vectors π and $z_\mathfrak{a}$ such that $g(x) = A^T\pi + D_\mathfrak{a}^T z_\mathfrak{a}$; and* (iii) *$z_\mathfrak{a} \ge 0$,*

(b) *the reduced Hessian $Z_\mathfrak{a}^T H Z_\mathfrak{a}$ is positive semidefinite.* ∎

Typically, software for general quadratic programming will terminate the iterations at a dead point. Nevertheless, it is possible to define procedures that check for optimality at a dead point, even though the chance of success in a reasonable amount of computation time will depend on the size of the problem (see Forsgren, Gill and Murray [11]).

2.2. Active-set methods

The method to be considered is a two-phase active-set method. In the first phase (the feasibility phase or phase 1), the objective is ignored while a feasible point is found for the constraints $Ax = b$ and $Dx \geq f$. In the second phase (the optimality phase or phase 2), the objective is minimized while feasibility is maintained. Given a feasible x_0, active-set methods compute a sequence of feasible iterates $\{x_k\}$ such that $x_{k+1} = x_k + \alpha_k p_k$ and $\varphi(x_{k+1}) \leq \varphi(x_k)$, where p_k is a nonzero search direction and α_k is a nonnegative step length. Active-set methods are motivated by the main result of Farkas' Lemma, which states that a feasible x must either satisfy the first-order optimality conditions or be the starting point of a feasible descent direction, i.e., a direction p such that

$$A_{\mathfrak{a}} p \geq 0 \quad \text{and} \quad g(x)^T p < 0. \tag{2.1}$$

The methods considered in this paper approximate the active set by a working set $\mathcal{W}$ of row indices of D. The working set has the form $\mathcal{W} = \{\nu_1, \nu_2, \ldots, \nu_{m_w}\}$, where m_w is the number of indices in $\mathcal{W}$. Analogous to the active-constraint matrix $A_{\mathfrak{a}}$, the $(m+m_w) \times n$ working-set matrix A_w contains the gradients of the equality constraints and inequality constraints in $\mathcal{W}$. The structure of the working-set matrix is similar to that of the active-set matrix, i.e.,

$$A_w = \begin{pmatrix} A \\ D_w \end{pmatrix},$$

where D_w is a matrix formed from the m_w rows of D with indices in $\mathcal{W}$. The vector f_w denotes the components of f with indices in $\mathcal{W}$.

There are two important distinctions between the definitions of $\mathcal{A}$ and $\mathcal{W}$.

(i) The indices of $\mathcal{W}$ define a subset of the rows of D that are linearly independent of the rows of A, i.e., the working set matrix A_w has full row rank. It follows that m_w must satisfy $0 \leq m_w \leq \min\{n - m, m_D\}$.

(ii) The active set $\mathcal{A}$ is uniquely defined at any feasible x, whereas there may be many choices for $\mathcal{W}$. The set $\mathcal{W}$ is determined by the properties of a particular active-set method.

Conventional active-set methods define the working set as a subset of the active set (see, e.g., Gill, Murray and Wright [16], and Nocedal and Wright [23]). In this paper we relax this requirement—in particular, a working-set constraint may be strictly satisfied or violated at x.

Given a working set $\mathcal{W}$ and an associated working-set matrix A_w at x, we introduce the notions of stationarity and optimality with respect to a working set. We emphasize that the definitions below do not require that the working-set constraints are active (or even feasible) at x.

Definition 2.1. (Subspace stationary point) *Let $\mathcal{W}$ be a working set defined at an x such that $Ax = b$. Then x is a* subspace stationary point with respect to $\mathcal{W}$ (or, equivalently, with respect to A_w) *if $g(x) \in \text{range}(A_w^T)$, i.e., there exists a vector y such that $g(x) = A_w^T y$. Equivalently, x is a subspace stationary point with respect to the working set $\mathcal{W}$ if the reduced gradient $Z_w^T g$ is zero, where the columns of Z_w form a basis for the null-space of A_w.* ∎

At a subspace stationary point, the components of y are the Lagrange multipliers associated with a QP with equality constraints $Ax = b$ and $D_w x = f_w$. To be consistent with the optimality conditions of Result 2.3, we denote the first m components of y as π (the multipliers associated with $Ax = b$) and the last m_w components of y as z_w (the multipliers associated with the constraints in $\mathcal{W}$). With this notation, the identity $g(x) = A_w^T y = A^T\pi + D_w^T z_w$ holds at a subspace stationary point.

Subspace stationary points may be classified based on the curvature of the objective on the working set.

Definition 2.2. (Subspace minimizer) *Let x be a subspace stationary point with respect to the working set $\mathcal{W}$. Let the columns of Z_w form a basis for the null-space of A_w. Then x is a* subspace minimizer with respect to the working set $\mathcal{W}$ *if the reduced Hessian $Z_w^T H Z_w$ is positive definite. If every constraint in the working set is active, then x is called a* standard subspace minimizer*; otherwise x is called a* nonstandard subspace minimizer*.* ∎

The inertia of the reduced Hessian is related to the inertia of the $(n+m_w)\times(n+m_w)$ KKT matrix $K = \begin{pmatrix} H & A_w^T \\ A_w & \end{pmatrix}$ through the identity $\text{In}(K) = \text{In}(Z_w^T H Z_w) + (m+m_w, m+m_w, 0)$ (see Gould [19]). It follows that an equivalent characterization of a subspace minimizer is that $g(x) \in \text{range}(A_w^T)$ and K has inertia $(n, m+m_w, 0)$.

A feasible x is said to be a *degenerate point* if $g(x) \in \text{range}(A_{\mathfrak{a}}^T)$ and the rows of $A_{\mathfrak{a}}$ are linearly dependent, i.e., $\text{rank}(A_{\mathfrak{a}}) < m_{\mathfrak{a}}$. A *degenerate vertex* is a degenerate point at which the rank of $A_{\mathfrak{a}}$ is n, and more than $n-m$ of the constraints $Dx \geq f$ are active. At a degenerate point there are infinitely many vectors y such that $g(x) = A_{\mathfrak{a}}^T y$, at least one of which will have one or more zero components.

3. Quadratic Programs with Mixed Constraints

An active-set method for quadratic programming is an iterative process involving the solution of a KKT system to compute a search direction p. This section describes a *nonbinding direction method*, an active-set method based on inertia control. The method limits the number of nonpositive eigenvalues in the KKT matrix (and hence limits the number of nonpositive eigenvalues in the reduced Hessian) allowing for the efficient calculation of search directions. Inertia-controlling methods are based on the simple rule that a constraint is removed from the working set only at a *subspace minimizer*.

At a subspace minimizer x, $g(x) = A_w^T y = A^T\pi + D_w^T z_w$. If x is standard and $z_w \geq 0$, then x is optimal for the QP. Otherwise, there exists an index $\nu_s \in \mathcal{W}$ such that $[z_w]_s < 0$. To proceed, we define a descent direction that is feasible for the equality constraints and the constraints in the working set. Analogous to (2.1), p is defined so that

$$A_w p = e_{m+s} \quad \text{and} \quad g(x)^T p < 0.$$

Any vector satisfying this condition is called a *nonbinding direction* because any nonzero step along it will increase the residual of the ν_s-th inequality constraint (and hence make it inactive or nonbinding). Here we define p as the solution of the equality-constrained subproblem

$$\underset{p}{\text{minimize}} \;\; \varphi(x+p) \quad \text{subject to} \quad A_w p = e_{m+s}. \tag{3.1}$$

The optimality conditions for this subproblem imply the existence of a vector q such that $g(x+p) = A_w^T(y+q)$; i.e., q is the step to the multipliers associated with the optimal solution $x+p$. This condition, along with the feasibility condition, implies that p and q satisfy the equations

$$\begin{pmatrix} H & A_w^T \\ A_w & 0 \end{pmatrix} \begin{pmatrix} p \\ -q \end{pmatrix} = \begin{pmatrix} -(g(x) - A_w^T y) \\ e_{m+s} \end{pmatrix}. \tag{3.2}$$

The primal and dual vectors have a number of important properties that are summarized in the next result.

Result 3.1. (Properties of a nonbinding search direction) *Let x be a subspace minimizer such that $g = A_w^T y = A^T\pi + D_w^T z_w$, with $[z_w]_s < 0$. Then the vectors p and q satisfying the equations*

$$\begin{pmatrix} H & A_w^T \\ A_w & 0 \end{pmatrix} \begin{pmatrix} p \\ -q \end{pmatrix} = \begin{pmatrix} -(g(x) - A_w^T y) \\ e_{m+s} \end{pmatrix} = \begin{pmatrix} 0 \\ e_{m+s} \end{pmatrix} \tag{3.3}$$

constitute the unique primal and dual solutions of the equality constrained problem defined by minimizing $\varphi(x+p)$ subject to $A_w p = e_{m+s}$. Moreover, p and q satisfy the identities

$$g^T p = y_{m+s} = [z_w]_s \quad \text{and} \quad p^T H p = q_{m+s} = [q_w]_s, \tag{3.4}$$

where q_w denotes the vector of last m_w components of q.

Proof. The assumption that x is a subspace minimizer implies that the subproblem has a unique bounded minimizer. The optimality of p and q follow from the equations in (3.2), which represent the feasibility and optimality conditions for the minimization of $\varphi(x+p)$ on the set $\{p : A_w p = e_{m+s}\}$. The equation $g = A_w^T y$ and the definition of p from (3.3) give

$$g^T p = p^T(A_w^T y) = y^T A_w p = y^T e_{m+s} = y_{m+s} = [z_w]_s$$

Similarly, $p^T H p = p^T(A_w^T q) = e_{m+s}^T q = q_{m+s} = [q_w]_s$. ∎

Once p and q are known, a nonnegative step α is computed so that $x + \alpha p$ is feasible and $\varphi(x+\alpha p) \le \varphi(x)$. If $p^T H p > 0$, the step that minimizes $\varphi(x+\alpha p)$ as a function of α is given by $\alpha_* = -g^T p / p^T H p$. The identities (3.4) give

$$\alpha_* = -g^T p / p^T H p = -[z_w]_s / [q_w]_s.$$

Since $[z_w]_s < 0$, if $[q_w]_s = p^T H p > 0$, the optimal step α_* is positive. Otherwise $[q_w]_s = p^T H p \le 0$ and φ has no bounded minimizer along p and $\alpha_* = +\infty$.

If $x + \alpha_* p$ is unbounded or infeasible, then α must be limited by α_F, the *maximum feasible step* from x along p. The feasible step is defined as $\alpha_F = \gamma_r$, where

$$\gamma_r = \min \gamma_i, \quad \text{with} \quad \gamma_i = \begin{cases} \dfrac{d_i^T x - f_i}{-\, d_i^T p} & \text{if } d_i^T p < 0; \\ +\infty & \text{otherwise.} \end{cases}$$

The step α is then $\min\{\alpha_*, \alpha_F\}$. If $\alpha = +\infty$, the QP has no bounded solution and the algorithm terminates. In the discussion below, we assume that α is a bounded step.

The primal and dual directions p and q defined by (3.3) have the property that $x + \alpha p$ remains a subspace minimizer with respect to A_w *for any step* α. This follows from the definitions (3.3), which imply that

$$g(x+\alpha p) = g(x) + \alpha H p = A_w^T y + \alpha A_w^T q = A_w^T (y + \alpha q), \tag{3.5}$$

so that the gradient at $x+\alpha p$ is a linear combination of the columns of A_w^T. The step $x+\alpha p$ does not change the KKT matrix K associated with the subspace minimizer x, which implies that $x + \alpha p$ is also a subspace minimizer with respect to A_w. This means that $x + \alpha p$ may be interpreted as the solution of a problem in which the working-set constraint $d_{\nu_s}^T x \ge f_{\nu_s}$ is *shifted* to pass through $x + \alpha p$. The component $[y + \alpha q]_{m+s} = [z_w + \alpha q_w]_s$ is the Lagrange multiplier associated with the shifted version of $d_{\nu_s}^T x \ge f_{\nu_s}$. This property is known as the *parallel subspace property* of quadratic programming. It shows that if x is stationary with respect to a nonbinding constraint, then it remains so for all subsequent iterates for which that constraint remains in the working set.

Once α has been defined, the new iterate is $\bar{x} = x + \alpha p$. The composition of the new working set and multipliers depends on the definition of α.

Case 1: $\alpha = \alpha_*$ In this case, $\alpha = \alpha_* = -[z_w]_s/[q_w]_s$ minimizes $\varphi(x + \alpha p)$ with respect to α, giving the s-th element of $z_w + \alpha q_w$ as

$$[z_w + \alpha q_w]_s = [z_w]_s + \alpha_* [q_w]_s = 0,$$

which implies that the Lagrange multiplier associated with the shifted constraint is *zero* at $\bar{x}$. The nature of the stationarity may be determined using the next result.

Result 3.2. (Constraint deletion) *Let x be a subspace minimizer with respect to $\mathcal{W}$. Assume that $[z_w]_s < 0$. Let $\bar{x}$ denote the point $x + \alpha p$, where p is defined by (3.3) and $\alpha = \alpha_*$ is bounded. Then $\bar{x}$ is a subspace minimizer with respect to $\bar{\mathcal{W}} = \mathcal{W} - \{\nu_s\}$.*

Proof. Let K and $\bar{K}$ denote the matrices

$$K = \begin{pmatrix} H & A_w^T \\ A_w & \end{pmatrix} \quad \text{and} \quad \bar{K} = \begin{pmatrix} H & \bar{A}_w^T \\ \bar{A}_w & \end{pmatrix},$$

where A_w and $\bar{A}_w$ are the working-set matrices associated with $\mathcal{W}$ and $\bar{\mathcal{W}}$. It suffices to show that $\bar{K}$ has the correct inertia, i.e., $\text{In}(\bar{K}) = (n, m + m_w - 1, 0)$.

Consider the matrix M such that

$$M \triangleq \begin{pmatrix} K & e_{m+n+s} \\ e_{m+n+s}^T & \end{pmatrix}.$$

By assumption, x is a subspace minimizer with $\text{In}(K) = (n, m + m_w, 0)$. In particular, K is nonsingular and the Schur complement of K in M exists with

$$M/K = -e_{n+m+s}^T K^{-1} e_{n+m+s} = -e_{n+m+s}^T \begin{pmatrix} p \\ -q \end{pmatrix} = [q_w]_s.$$

It follows that

$$\text{In}(M) = \text{In}(M/K) + \text{In}(K) = \text{In}([q_w]_s) + (n, m + m_w, 0). \tag{3.6}$$

Now consider a symmetrically permuted version of M:

$$\widetilde{M} = \begin{pmatrix} 0 & 1 & & \\ 1 & 0 & d_{\nu_s}^T & \\ & d_{\nu_s} & H & \bar{A}_w^T \\ & & \bar{A}_w & \end{pmatrix}.$$

Inertia is unchanged by symmetric permutations, so $\text{In}(M) = \text{In}(\widetilde{M})$. The 2×2 block in the upper-left corner of $\widetilde{M}$, denoted by E, has eigenvalues ± 1, so that

$$\text{In}(E) = (1, 1, 0) \quad \text{with} \quad E^{-1} = E.$$

The Schur complement of E in $\widetilde{M}$ is

$$\widetilde{M}/E = \bar{K} - \begin{pmatrix} 0 & d_{\nu_s} \\ 0 & 0 \end{pmatrix} \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 0 & 0 \\ d_{\nu_s}^T & 0 \end{pmatrix} = \bar{K},$$

which implies that $\mathrm{In}(\widetilde{M}) = \mathrm{In}(\widetilde{M}/E) + \mathrm{In}(E) = \mathrm{In}(\bar{K}) + (1,1,0)$. Combining this with (3.6) yields

$$\begin{aligned}\mathrm{In}(\bar{K}) &= \mathrm{In}([q_w]_s) + (n, m+m_w, 0) - (1,1,0)\\ &= \mathrm{In}([q_w]_s) + (n-1, m+m_w-1, 0).\end{aligned}$$

Since $\alpha = \alpha_*$, $[q_w]_s$ must be positive. It follows that

$$\mathrm{In}(\bar{K}) = (1,0,0) + (n-1, m+m_w-1, 0) = (n, m+m_w-1, 0)$$

and the subspace stationary point $\bar{x}$ is a (standard) subspace minimizer with respect to the new working set $\bar{\mathcal{W}} = \mathcal{W} - \{\nu_s\}$. ∎

Case 2: $\alpha = \alpha_F$ In this case, α is the step to the blocking constraint $d_r^T x \ge f_r$, which is eligible to be added to the working set at $x + \alpha p$. However, the definition of the new working set depends on whether or not the blocking constraint is dependent on the constraints already in $\mathcal{W}$. If d_r is linearly independent of the columns of A_w^T, then the index r is added to the working set. Otherwise, we show in Result 3.4 below that a suitable working set is defined by exchanging rows d_{ν_s} and d_r in A_w. The following result provides a computable test for the independence of d_r and the columns of A_w^T.

Result 3.3. (Test for constraint dependency) *Let x be a subspace minimizer with respect to A_w. Assume that $d_r^T x \ge f_r$ is a blocking constraint at $\bar{x} = x + \alpha p$, where p satisfies (3.3). Define vectors u and v such that*

$$\begin{pmatrix} H & A_w^T \\ A_w & \end{pmatrix} \begin{pmatrix} u \\ v \end{pmatrix} = \begin{pmatrix} d_r \\ 0 \end{pmatrix}, \tag{3.7}$$

then

(a) *d_r and the columns of A_w^T are linearly independent if and only if $u \ne 0$;*

(b) *$v_{m+s} = d_r^T p < 0$, and if $u \ne 0$, then $u^T d_r > 0$.*

Proof. For part (a), equations (3.7) give $Hu + A_w^T v = d_r$ and $A_w u = 0$. If $u = 0$ then $A_w^T v = d_r$, and d_r must be dependent on the columns of A_w^T. Conversely, if $A_w^T v = d_r$, then the definition of u gives $u^T A_w^T v = u^T d_r = 0$, which implies that $u^T H u = u^T(Hu + A_w^T v) = u^T d_r = 0$. By assumption, x is a subspace minimizer with respect to A_w, which is equivalent to the assumption that H is positive definite for all u such that $A_w u = 0$. Hence $u^T H u = 0$ can hold only if u is zero.

For part (b), we use equations (3.3) and (3.7) to show that

$$v_{m+s} = e_{m+s}^T v = p^T A_w^T v = p^T(d_r - Hu) = p^T d_r - q^T A_w u = d_r^T p < 0,$$

where the final inequality follows from the fact that $d_r^T p$ must be negative if $d_r^T x \ge f_r$ is a blocking constraint. If $u \ne 0$, equations (3.7) imply $Hu + A_w^T v = d_r$ and $A_w u = 0$. Multiplying the first equation by u^T and applying the second equation gives $u^T H u = u^T d_r$. As $u \in \mathrm{null}(A_w)$ and x is a subspace minimizer, it must hold that $u^T H u = u^T d_r > 0$, as required. ∎

The next result provides expressions for the updated multipliers after a constraint is added to the working set.

Result 3.4. (Constraint addition) *Assume that x is a subspace minimizer with respect to A_w. Assume that $d_r^T x \ge f_r$ is a blocking constraint at the next iterate $\bar{x} = x + \alpha p$, where the direction p satisfies (3.3). Let u and v satisfy (3.7).*

(a) *If d_r and the columns of A_w^T are linearly independent, then the vector $\bar{y}$ formed by appending a zero component to the vector $y + \alpha q$ satisfies $g(\bar{x}) = \bar{A}_w^T \bar{y}$, where $\bar{A}_w$ denotes the matrix A_w with row d_r^T added in the last position.*

(b) *If d_r and the columns of A_w^T are linearly dependent, then the vector $\bar{y}$ such that*

$$\bar{y} = y + \alpha q - \sigma v, \quad \textit{with} \quad \sigma = [y + \alpha q]_{m+s} / v_{m+s}, \tag{3.8}$$

satisfies $g(\bar{x}) = A_w^T \bar{y} + \sigma d_r$ with $\bar{y}_{m+s} = 0$ and $\sigma > 0$.

Proof. For part (a), the identity (3.5) implies that $g(x + \alpha p) = g(\bar{x}) = A_w^T (y + \alpha q)$. As d_r and the columns of A_w^T are linearly independent, we may add the index r to $\mathcal{W}$ and define the new working-set matrix $\bar{A}_w^T = \begin{pmatrix} A_w^T & d_r \end{pmatrix}$. This allows us to write $g(\bar{x}) = \bar{A}_w^T \bar{y}$, with $\bar{y}$ given by $y + \alpha q$ with an appended zero component.

Now assume that A_w^T and d_r are linearly dependent. From Result 3.3 it must hold that $u = 0$ and there exists a unique v such that $d_r = A_w^T v$. For any value of σ, it holds that

$$g(\bar{x}) = A_w^T (y + \alpha q) = A_w^T (y + \alpha q - \sigma v) + \sigma d_r.$$

If we choose $\sigma = [y + \alpha q]_{m+s} / v_{m+s}$ and define the vector $\bar{y} = y + \alpha q - \sigma v$, then

$$g(\bar{x}) = A_w^T \bar{y} + \sigma d_r, \quad \text{with} \quad \bar{y}_{m+s} = [y + \alpha q - \sigma v]_{m+s} = 0.$$

It follows that $g(\bar{x})$ is a linear combination of d_r and every column of A_w^T except d_s.

In order to show that $\sigma = [y + \alpha q]_{m+s} / v_{m+s}$ is positive, we consider the linear function $\eta(\alpha) = [y + \alpha q]_{m+s}$, which satisfies $\eta(0) = y_{m+s} < 0$. If $q_{m+s} = p^T H p > 0$, then $\alpha_* < \infty$ and $\eta(\alpha)$ is an increasing linear function of positive α with $\eta(\alpha_*) = 0$. This implies that $\eta(\alpha) < 0$ for any $\alpha < \alpha_*$ and $\eta(\alpha_k) < 0$. If $q_{m+s} \le 0$, then $\eta(\alpha)$ is a nonincreasing linear function of α so that $\eta(\alpha) < 0$ for any positive α. Thus, $[y + \alpha q]_{m+s} < 0$ for any $\alpha < \alpha_*$, and $\sigma = [y + \alpha q]_{m+s} / v_{m+s} > 0$ from part (b) of Result 3.3. ∎

Result 3.5. *Let x be a subspace minimizer with respect to the working set $\mathcal{W}$. Assume that $d_r^T x \ge f_r$ is a blocking constraint at $\bar{x} = x + \alpha p$, where p is defined by (3.3).*

(a) *If d_r is linearly independent of the columns of A_w^T, then $\bar{x}$ is a subspace minimizer with respect to the working set $\bar{\mathcal{W}} = \mathcal{W} + \{r\}$.*

(b) *If d_r is dependent on the columns of A_w^T, then $\bar{x}$ is a subspace minimizer with respect to the working set $\bar{\mathcal{W}} = \mathcal{W} + \{r\} - \{\nu_s\}$.*

Proof. Parts (a) and (b) of Result 3.4 imply that $\bar{x}$ is a subspace stationary point with respect to $\bar{\mathcal{W}}$. It remains to show that in each case, the KKT matrix for the new working set has correct inertia.

For part (a), it suffices to show that the KKT matrix for the new working set $\bar{\mathcal{W}} = \mathcal{W} + \{r\}$ has inertia $(n, m + m_w + 1, 0)$. Assume that d_r and the columns of A_w^T are linearly independent, so that the vector u of (3.7) is nonzero. Let K and $\bar{K}$ denote the KKT matrices associated with the working sets $\mathcal{W}$ and $\bar{\mathcal{W}}$, i.e.,

$$K = \begin{pmatrix} H & A_w^T \\ A_w & \end{pmatrix} \quad \text{and} \quad \bar{K} = \begin{pmatrix} H & \bar{A}_w^T \\ \bar{A}_w & \end{pmatrix},$$

where $\bar{A}_w$ is the matrix A_w with the row d_r^T added in the last position.

By assumption, x is a subspace minimizer and $\text{In}(K) = (n, m + m_w, 0)$. It follows that K is nonsingular and the Schur complement of K in $\bar{K}$ exists with

$$\bar{K}/K = -\begin{pmatrix} d_r \\ 0 \end{pmatrix}^T K^{-1} \begin{pmatrix} d_r \\ 0 \end{pmatrix} = -\begin{pmatrix} d_r^T & 0 \end{pmatrix} \begin{pmatrix} u \\ v \end{pmatrix} = -d_r^T u < 0,$$

where the last inequality follows from part (b) of Result 3.3. Then,

$$\begin{aligned}\text{In}(\bar{K}) = \text{In}(\bar{K}/K) + \text{In}(K) &= \text{In}(-u^T d_r) + (n, m + m_w, 0) \\ &= (0, 1, 0) + (n, m + m_w, 0) = (n, m + m_w + 1, 0).\end{aligned}$$

For part (b), assume that d_r and the columns of A_w^T are linearly dependent and that $\bar{\mathcal{W}} = \mathcal{W} + \{r\} - \{\nu_s\}$. By Result 3.4 and equation (3.7), it must hold that $u = 0$ and $A_w^T v = d_r$. Let A_w and $\bar{A}_w$ be the working-set matrices associated with $\mathcal{W}$ and $\bar{\mathcal{W}}$. The change in the working set replaces row s of D_w by d_r^T, so that

$$\begin{aligned}\bar{A}_w = A_w + e_{m+s}(d_r^T - d_s^T) &= A_w + e_{m+s}(v^T A_w - e_{m+s}^T A_w) \\ &= \big(I + e_{m+s}(v - e_{m+s})^T\big) A_w \\ &= M A_w,\end{aligned}$$

where $M = I + e_{m+s}(v - e_{m+s})^T$. The matrix M has $m + m_w - 1$ unit eigenvalues and one eigenvalue equal to v_{m+s}. From part (b) of Result 3.3, it holds that $v_{m+s} < 0$ and hence M is nonsingular. The new KKT matrix for $\bar{\mathcal{W}}$ can be written as

$$\begin{pmatrix} H & \bar{A}_w^T \\ \bar{A}_w & \end{pmatrix} = \begin{pmatrix} I & \\ & M \end{pmatrix} \begin{pmatrix} H & A_w^T \\ A_w & \end{pmatrix} \begin{pmatrix} I & \\ & M^T \end{pmatrix}.$$

By Sylvester's Law of Inertia, the old and new KKT matrices have the same inertia, which implies that $\bar{x}$ is a subspace minimizer with respect to $\bar{\mathcal{W}}$. ∎

The first part of this result shows that $\bar{x}$ is a subspace minimizer both before and after an independent constraint is added to the working set. This is crucial because it means that the directions p and q for the next iteration satisfy the KKT equations (3.3) with $\bar{A}_w$ in place of A_w. The second part shows that the working-set constraints can be linearly dependent only at a *standard* subspace minimizer associated with a working set that does not include constraint ν_s. This implies that it is appropriate to remove ν_s from the working set. The constraint $d_{\nu_s}^T x \ge f_{\nu_s}$ plays a significant (and explicit) role in the definition of the search direction and is called the *nonbinding working-set constraint*. The method generates sets of consecutive iterates that begin and end with a standard subspace minimizer. The nonbinding working-set constraint $d_{\nu_s}^T x \ge f_{\nu_s}$ identified at the first point of the sequence is deleted from the working set at the last point (either by deletion or replacement).

The proposed method is the basis for Algorithm 3.1 given below. Each iteration requires the solution of two KKT systems:

$$\text{Full System 1} \qquad \begin{pmatrix} H & A_w^T \\ A_w & 0 \end{pmatrix} \begin{pmatrix} p \\ -q \end{pmatrix} = \begin{pmatrix} 0 \\ e_{m+s} \end{pmatrix} \tag{3.9a}$$

$$\text{Full System 2} \qquad \begin{pmatrix} H & A_w^T \\ A_w & 0 \end{pmatrix} \begin{pmatrix} u \\ v \end{pmatrix} = \begin{pmatrix} d_r \\ 0 \end{pmatrix}. \tag{3.9b}$$

However, for those iterations for which the number of constraints in the working set increases, it is possible to *update* the vectors p and q, making it unnecessary to solve (3.9a).

Algorithm 3.1.[Nonbinding-direction method for general QP]

Find x such that $Ax = b$, $Dx \ge f$; $k = 0$;
Choose $\mathcal{W}$, any full-rank subset of $\mathcal{A}(x)$; Choose π and z_w;
$\big[x, \pi, z_w, \mathcal{W}\big] = \texttt{subspaceMin}(x, \pi, z_w, \mathcal{W})$; $m_w = |\mathcal{W}|$;
$g = c + Hx$; $s = \text{argmin}_i\ [z_w]_i$;
while $[z_w]_s < 0$ **do**
  Solve $\begin{pmatrix} H & A^T & D_w^T \\ A & 0 & 0 \\ D_w & 0 & 0 \end{pmatrix}\begin{pmatrix} p \\ -q_\pi \\ -q_w \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ e_s \end{pmatrix}$;
  $\alpha_F = \texttt{maxStep}(x, p, D, f)$;
  if $[q_w]_s > 0$ **then** $\alpha_* = -[z_w]_s/[q_w]_s$ **else** $\alpha_* = +\infty$;
  $\alpha = \min\{\alpha_*, \alpha_F\}$;
  if $\alpha = +\infty$ **then stop**; [the solution is unbounded]
  $x \leftarrow x + \alpha p$; $\pi \leftarrow \pi + \alpha q_\pi$; $z_w \leftarrow z_w + \alpha q_w$; $g \leftarrow g + \alpha H p$;
  if $\alpha_F < \alpha_*$ **then** [add constraint r to the working set]
    Choose a blocking constraint index r;
    Solve $\begin{pmatrix} H & A^T & D_w^T \\ A & 0 & 0 \\ D_w & 0 & 0 \end{pmatrix}\begin{pmatrix} u \\ v_\pi \\ v_w \end{pmatrix} = \begin{pmatrix} d_r \\ 0 \\ 0 \end{pmatrix}$;
    if $u = 0$ **then** $\sigma = [z_w]_s/[v_w]_s$ **else** $\sigma = 0$;
    $\pi \leftarrow \pi - \sigma v_\pi$; $z_w \leftarrow \begin{pmatrix} z_w - \sigma v_w \\ \sigma \end{pmatrix}$;
    $\mathcal{W} \leftarrow \mathcal{W} + \{r\}$; $m_w \leftarrow m_w + 1$;
  end;
  if $[z_w]_s = 0$ **then** [delete constraint ν_s from the working set]
    $\mathcal{W} \leftarrow \mathcal{W} - \{\nu_s\}$; $m_w \leftarrow m_w - 1$;
    for $i = s : m_w$ **do** $[z_w]_i \leftarrow [z_w]_{i+1}$;
    $s = \text{argmin}_i\ [z_w]_i$;
  end;
  $k \leftarrow k + 1$;
end do

Result 3.6. *Let x be a subspace minimizer with respect to A_w. Assume the vectors p, q, u and v are defined by (3.9). Let d_r be the gradient of a blocking constraint at $\bar{x} = x + \alpha p$ such that d_r is independent of the columns of A_w^T. If $\rho = -d_r^T p / d_r^T u$, then the vectors*

$$\bar{p} = p + \rho u \quad \textit{and} \quad \bar{q} = \begin{pmatrix} q - \rho v \\ \rho \end{pmatrix}$$

are well-defined and satisfy

$$\begin{pmatrix} H & \bar{A}_w^T \\ \bar{A}_w & \end{pmatrix}\begin{pmatrix} \bar{p} \\ -\bar{q} \end{pmatrix} = \begin{pmatrix} 0 \\ e_{m+s} \end{pmatrix}, \quad \textit{where} \quad \bar{A}_w = \begin{pmatrix} A_w \\ d_r^T \end{pmatrix}. \tag{3.10}$$

Proof. Result 3.3 implies that u is nonzero and that $u^T d_r > 0$ so that ρ is well defined (and strictly positive).

For any scalar ρ, (3.9a) and (3.9b) imply that

$$\begin{pmatrix} H & A_w^T & d_r \\ A_w & & \\ d_r^T & & \end{pmatrix}\begin{pmatrix} p + \rho u \\ -(q - \rho v) \\ -\rho \end{pmatrix} = \begin{pmatrix} 0 \\ e_{m+s} \\ d_r^T p + \rho d_r^T u \end{pmatrix}.$$

If ρ is chosen so that $d_r^T p + \rho d_r^T u = 0$, the last component of the right-hand side vanishes, and $\bar{p}$ and $\bar{q}$ satisfy (3.10) as required. ∎

4. Quadratic Programs in Standard Form

The inequality constraints of a quadratic program in standard form consist of only simple upper and lower bounds on the variables. Without loss of generality, we consider methods for the standard-form quadratic program:

$$\underset{x\in\mathbb{R}^n}{\text{minimize}} \;\; \varphi(x) = c^T x + \tfrac{1}{2}x^T H x \quad \text{subject to} \quad Ax = b, \quad x \geq 0. \tag{4.1}$$

This is an example of a mixed-constraint problem (1.1) with $D = I$ and $f = 0$. The working-set matrix has the form:

$$A_w = \begin{pmatrix} A \\ E_w \end{pmatrix},$$

where the rows of E_w are the rows of the identity corresponding to the m_w constraints in the working set. This form of the working-set matrix implies that there is a permutation P such that $E_w P = \begin{pmatrix} 0 & I_w \end{pmatrix}$, where I_w is the identity matrix of order m_w. If P is applied to A, we obtain $AP = \begin{pmatrix} A_B & A_N \end{pmatrix}$, where A_N is the matrix of columns of A corresponding to variables that are implicitly fixed on their bounds by the constraints in the working set, and A_B contains the columns associated with the complementary set of "free" variables. Following standard terminology associated with constraints in standard form, we use the term *nonbasic set* to refer to the working set. The nonbasic set is denoted by $\mathcal{N} = \{\nu_1, \nu_2, \ldots, \nu_{n_N}\}$, where $n_N = m_w$. Similarly, we define the complementary set of $n_B = n - n_N$ indices that are not in $\mathcal{N}$ as the *basic set*, with $\mathcal{B} = \{\beta_1, \beta_2, \ldots, \beta_{n_B}\}$. With these definitions, the matrices A_B and A_N have columns $\{\, a_{\beta_j} \,\}$ and $\{\, a_{\nu_j} \,\}$ respectively. The effect of P on the Hessian and working-set matrix A_w may be written as

$$P^T H P = \begin{pmatrix} H_B & H_D \\ H_D^T & H_N \end{pmatrix}, \quad \text{and} \quad A_w P = \begin{pmatrix} A_B & A_N \\ 0 & I_N \end{pmatrix}, \tag{4.2}$$

where I_N denotes the identity matrix of order n_N. As in the mixed-constraint formulation, A_w must have full row-rank. This is equivalent to requiring that A_B has full row-rank since $\text{rank}(A_w) = n_N + \text{rank}(A_B)$. If y is an n-vector, y_B (the *basic components of* y) denotes the n_B-vector whose j-th component is component β_j of y, and y_N (the *nonbasic components of* y) denotes the n_N-vector whose j-th component is component ν_j of y.

Result 4.1. (Subspace minimizer for standard form) *Let x be a feasible point with working set A_w and $g = g(x)$.*

(a) *If x is a subspace stationary point with respect to A_w, then there exists a vector π such that $g_B = A_B^T \pi$, or equivalently, $Z_B^T g_B = 0$, where the columns of Z_B form a basis for the null-space of A_B.*

(b) *If x is a subspace minimizer with respect to A_w, then $Z_B^T H_B Z_B$ is positive definite, where the columns of Z_B form a basis for the null-space of A_B. Equivalently, if x is a subspace minimizer, then the KKT matrix $K_B = \begin{pmatrix} H_B & A_B^T \\ A_B & \end{pmatrix}$ has inertia $(n_B, m, 0)$.*

∎

For constraints in standard form, we say that x is a subspace minimizer with respect to the basic set $\mathcal{B}$ (or, equivalently, with respect to A_B).

As in linear programming, the components of the vector $z = g(x) - A^T\pi$ are called the *reduced costs*. For constraints in standard form, the multipliers z_w associated inequality constraints in the working set are denoted by z_N. The components of z_N are the nonbasic components of the reduced-cost vector, i.e.,

$$z_N = (g(x) - A^T\pi)_{\mathcal{N}} = g_N - A_N^T\pi.$$

At a subspace stationary point, it holds that $g_B - A_B^T\pi = 0$, which implies that the basic components of the reduced costs are zero.

The fundamental property of constraints in standard form is that the mixed-constraint method may be formulated so that the number of variables involved in the equality-constraint QP subproblem is reduced from n to n_B. For example, the KKT equations (3.9a) may be written in terms of the basic and nonbasic variables by applying the permutation matrix P appropriately; i.e.,

$$\left(\begin{array}{cc|cc} H_B & H_D & A_B^T & \\ H_D^T & H_N & A_N^T & I_N \\ \hline A_B & A_N & & \\ & I_N & & \end{array}\right)\begin{pmatrix} p_B \\ p_N \\ -q_\pi \\ -q_N \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ e_s \end{pmatrix}, \quad \text{where } p = P\begin{pmatrix} p_B \\ p_N \end{pmatrix} \text{ and } q = \begin{pmatrix} q_\pi \\ q_N \end{pmatrix}.$$

These equations imply that p_B and q_π satisfy the smaller KKT system

$$\begin{pmatrix} H_B & A_B^T \\ A_B & 0 \end{pmatrix}\begin{pmatrix} p_B \\ -q_\pi \end{pmatrix} = \begin{pmatrix} -H_D p_N \\ -A_N p_N \end{pmatrix} = -\begin{pmatrix} (h_{\nu_s})_{\mathcal{B}} \\ a_{\nu_s} \end{pmatrix}. \tag{4.3}$$

Once p_B and q_π are known, the increment q_N for multipliers z_N associated with the constraints $p_N = e_s$ are given by $q_N = (Hp - A^T q_\pi)_{\mathcal{N}}$.

Similarly, the solution of the second KKT system (3.9b) can be computed from the KKT equation

$$\begin{pmatrix} H_B & A_B^T \\ A_B & \end{pmatrix}\begin{pmatrix} u_B \\ v_\pi \end{pmatrix} = \begin{pmatrix} e_r \\ 0 \end{pmatrix}, \tag{4.4}$$

with $u_N = 0$ and $v_N = -(Hu + A^T v_\pi)_{\mathcal{N}}$, where $u = P\begin{pmatrix} u_B \\ u_N \end{pmatrix}$ and $v = \begin{pmatrix} v_\pi \\ v_N \end{pmatrix}$.

The KKT equations (4.3) and (4.4) allow the mixed constraint algorithm to be formulated in terms of the basic variables only, which implies that the algorithm is driven by variables entering or leaving the basic set rather than constraints entering or leaving the working set. With this interpretation, changes to the KKT matrix are based on column-changes to A_B instead of row-changes to D_w. In practice, p_N is defined implicitly and only the components of p_B need be computed explicity. As in linear programming, the largest feasible step is defined using the minimum ratio test:

$$\alpha_F = \min \gamma_i, \quad \text{where} \quad \gamma_i = \begin{cases} \dfrac{[x_B]_i}{-[p_B]_i} & \text{if } [p_B]_i < 0, \\ +\infty & \text{otherwise.} \end{cases}$$

For completeness we summarize Results 3.2—3.5 in terms of the quantities associated with constraints in standard form.

Result 4.2. *Let x be a subspace minimizer with respect to the basic set $\mathcal{B}$, with $[z_N]_s < 0$. Let $\bar{x}$ be the point such that $\bar{x}_N = x_N + \alpha e_s$ and $\bar{x}_B = x_B + \alpha p_B$, where p_B is defined as in (4.3).*

(1) *The step to the minimizer of* $\varphi(x + \alpha p)$ *is* $\alpha_* = -z_{\nu_s}/[q_N]_s$. *If* α_* *is bounded and* $\alpha = \alpha_*$, *then* $\bar{x}$ *is a subspace minimizer with respect to the basic set* $\bar{\mathcal{B}} = \mathcal{B} + \{\nu_s\}$.

(2) *If* $[x_B + \alpha p_B]_{\beta_r} = 0$, *let* u_B *and* v_π *be defined by* (4.4).

   (a) e_r *and the columns of* A_B^T *are linearly independent if and only if* $u_B \neq 0$.

   (b) $[v_N]_s = [p_B]_r < 0$, *and if* $u_B \neq 0$, *then* $[u_B]_r > 0$.

   (c) *If* e_r *and the columns of* A_B^T *are linearly independent, then* $\bar{x}$ *is a subspace minimizer with respect to* $\bar{\mathcal{B}} = \mathcal{B} - \{\beta_r\}$. *Moreover,* $g_{\bar{B}}(\bar{x}) = A_{\bar{B}}^T\bar{\pi}$ *and* $g_{\bar{N}}(\bar{x}) = A_{\bar{N}}^T\bar{\pi} + \bar{z}_N$, *where* $\bar{\pi} = \pi + \alpha q_\pi$ *and* $\bar{z}_N$ *is formed by appending a zero component to the vector* $z_N + \alpha q_N$.

   (d) *If* e_r *and the columns of* A_B^T *are linearly dependent, define* $\sigma = [z_N + \alpha q_N]_s/[v_N]_s$. *Then* $\bar{x}$ *is a subspace minimizer with respect to* $\bar{\mathcal{B}} = \mathcal{B} - \{\beta_r\} + \{\nu_s\}$ *with* $g_{\bar{B}}(\bar{x}) = A_{\bar{B}}^T\bar{\pi}$ *and* $g_{\bar{N}}(\bar{x}) = A_{\bar{N}}^T\bar{\pi} + \bar{z}_N$, *where* $\bar{\pi} = \pi + \alpha q_\pi - \sigma v_\pi$ *with* $\sigma > 0$, *and* $\bar{z}_N$ *is formed by appending* σ *to* $z_N + \alpha q_N - \sigma v_N$. ∎

As in the mixed-constraint method, the direction p_B and multiplier q_π may be updated in the linearly independent case.

Result 4.3. *Let* x *be a subspace minimizer with respect to* $\mathcal{B}$. *Assume the vectors* p_B, q_π, u_B *and* v_π *are defined by* (4.3) *and* (4.4). *Let* β_r *be the index of a linearly independent blocking constraint at* $\bar{x}$, *where* $\bar{x}_N = x_N + \alpha e_s$ *and* $\bar{x}_B = x_B + \alpha p_B$. *Let* $\rho = -[p_B]_r/[u_B]_r$, *and consider the vectors* $\bar{p}_B$ *and* $\bar{q}_\pi$, *where* $\bar{p}_B$ *is the vector* $p_B + \rho u_B$ *with the* r*-th component omitted, and* $\bar{q}_\pi = q_\pi - \rho v_\pi$. *Then* $\bar{p}_B$ *and* $\bar{q}_\pi$ *are well-defined and satisfy the KKT equations for the basic set* $\mathcal{B} - \{\beta_r\}$. ∎

Algorithm 4.1 summarizes the nonbinding direction method for general QPs in standard form. For simplicity, algorithm recomputes z from π as $z = g - A^T\pi$, instead of using q_N and v_N to update z as in Result 4.3. In addition, the relation in part 2(b) of Result 4.2 is used to simplify the computation of $[v_N]_s$.

4.1. Linear programs in standard form

If the problem is a linear program (i.e., $H = 0$), then the basic set $\mathcal{B}$ must be chosen so that A_B is nonsingular (i.e., it is square with rank m). In this case, we show that Algorithm 4.1 simplifies to a variant of the primal simplex method in which the π-values and reduced costs are updated by a simple recurrence relation.

When $H = 0$, the equations (4.3) reduce to $A_B p_B = -a_{\nu_s}$ and $A_B^T q_\pi = 0$, with $p_N = e_s$ and $q_N = -A_N^T q_\pi$. As A_B is nonsingular, both q_π and q_N are zero, and the directions p_B and p_N are equivalent to those defined by the simplex method. For the singularity check (4.4), the basic and nonbasic components of u satisfy $A_B u_B = 0$ and $u_N = 0$. Similarly, $v_N = -A_N^T v_\pi$, where $A_B^T v_\pi = e_r$, As A_B is nonsingular, $u_B = 0$ and the linearly dependent case always applies. This implies that the r-th basic and the s-th nonbasic variables are always swapped, as in the primal simplex method.

As q is zero, the updates to the multiplier vectors π and z_N defined by part 2(d) of Result 4.2 depend only on the vectors v_π and v_N, and the scalar $\sigma = -[z_N]_s/[p_B]_r$. The resulting updates to the multipliers are:

$$\pi \leftarrow \pi - \sigma v_\pi, \quad \text{and} \quad z_N \leftarrow \begin{pmatrix} z_N - \sigma v_N \\ \sigma \end{pmatrix},$$

Algorithm 4.1.[Nonbinding direction method for a general QP in standard form]

Find x_0 such that $Ax_0 = b$ and $x_0 \geq 0$;
$[x, \pi, \mathcal{B}, \mathcal{N}] = \texttt{subspaceMin}(x_0)$;
$g = c + Hx$; $z = g - A^T\pi$;
$\nu_s = \operatorname{argmin}_i\{z_i\}$;
while $z_{\nu_s} < 0$ **do**
  Solve $\begin{pmatrix} H_B & A_B^T \\ A_B & \end{pmatrix} \begin{pmatrix} p_B \\ -q_\pi \end{pmatrix} = -\begin{pmatrix} (h_{\nu_s})_{\mathcal{B}} \\ a_{\nu_s} \end{pmatrix}$; $p_N = e_s$; $p = P\begin{pmatrix} p_B \\ p_N \end{pmatrix}$;
  $\alpha_F = \texttt{minRatioTest}(x_B, p_B)$;
  if $[q_N]_s > 0$ **then** $\alpha_* = -z_{\nu_s}/[q_N]_s$ **else** $\alpha_* = +\infty$;
  $\alpha = \min\{\alpha_*, \alpha_F\}$;
  if $\alpha = +\infty$ **then stop**; [the solution is unbounded]
  $x \leftarrow x + \alpha p$; $g \leftarrow g + \alpha Hp$;
  $\pi \leftarrow \pi + \alpha q_\pi$; $z = g - A^T\pi$;
  if $\alpha_F < \alpha_*$ **then** [remove the r-th basic variable]
    Find the blocking constraint index r;
    Solve $\begin{pmatrix} H_B & A_B^T \\ A_B & \end{pmatrix} \begin{pmatrix} u_B \\ v_\pi \end{pmatrix} = \begin{pmatrix} e_r \\ 0 \end{pmatrix}$;
    if $u_B = 0$ **then** $\sigma = z_{\nu_s}/[p_B]_r$ **else** $\sigma = 0$;
    $\mathcal{B} \leftarrow \mathcal{B} - \{\beta_r\}$; $\mathcal{N} \leftarrow \mathcal{N} + \{\beta_r\}$;
    $\pi \leftarrow \pi - \sigma v_\pi$; $z = g - A^T\pi$;
  end;
  if $z_{\nu_s} = 0$ **then** [add the s-th nonbasic variable]
    $\mathcal{B} \leftarrow \mathcal{B} + \{\nu_s\}$; $\mathcal{N} \leftarrow \mathcal{N} - \{\nu_s\}$;
    $\nu_s = \operatorname{argmin}_i\{z_i\}$;
  end;
  $k \leftarrow k + 1$;
end do

which are the established multiplier updates associated with the simplex method (see Gill [12] and Tomlin [27]). It follows that the simplex method is a nonbinding direction method for which *every subspace minimizer is standard.* (i.e., the nonbasic set is always a subset of the active set).

5. Dual Active-Set Methods

5.1. Optimality conditions for the dual of a convex QP

The dual of the standard-form QP (4.1) is

$$\underset{w\in\mathbb{R}^n,\,\pi\in\mathbb{R}^m}{\text{minimize}} \;\; \varphi_D(w,\pi) = \tfrac{1}{2}w^T Hw - b^T\pi \quad \text{subject to} \quad Hw - A^T\pi \geq -c. \tag{5.1}$$

The optimality conditions for the dual were first established by Dorn [8]. If the dual is unbounded, the primal constraints are infeasible. A bounded solution (w, π) may be used to define the solution and Lagrange multipliers for the primal problem (4.1). The relationship between the primal and dual solution is characterized by the following result.

Result 5.1. (Optimality conditions for the dual QP) *The point* (w, π) *solves the dual QP* (5.1) *if and only if*

(a) (w, π) *satisfies* $Hw - A^T\pi \ge -c$;

(b) *there exists an* n*-vector* y *such that* (i) $Hw = Hy$, (ii) $Ay = b$, (iii) $y \ge 0$, *and* (iii) $y \cdot (Hw - A^T\pi + c) = 0$. ∎

If the dual has a bounded solution, then part (b) implies that the vector y of multipliers for the dual is a KKT point of the primal, and hence constitutes a primal solution. Moreover, if the dual has a bounded solution and H is nonsingular, then $w = y$.

5.2. A dual nonbinding direction method for convex QP

A dual active-set method for strictly convex problems was proposed by Goldfarb and Idnani [18]. This method was extended by Powell [26] to deal with ill-conditioned problems and reformulated by Boland [3] to handle the convex case. These methods require the factorization of a matrix defined in terms of the inverse of H, and as such, they are unsuitable for large-scale QP. This difficulty was addressed by Bartlett and Biegler [1], who reformulated the Goldfarb-Idnani method so that a Schur-complement method could be used to solve the linear systems (see Section 6 for a discussion of the Schur-complement method).

In this section we formulate a dual active-set method based on applying the nonbinding-direction method of Section 3 to the dual problem (5.1). The method is suitable for QPs that are not strictly convex (as in the primal case) and, as in the Bartlett-Biegler approach, the method may be implemented without the need for customized linear algebra software.

However, the method of Section 3 cannot be applied to the dual problem (5.1) directly. When H is singular, a KKT matrix defined in terms of any subset of the dual constraint gradients is singular, and the dual QP has no subspace minimizers—i.e., the reduced Hessian is positive semidefinite and singular at every subspace stationary point. This difficulty may be overcome by including additional artificial equality constraints in the dual. Let Z be a matrix with columns that form a basis for the null space of H. Consider the regularized problem

$$\underset{w\in\mathbb{R}^n,\, \pi\in\mathbb{R}^m}{\text{minimize}} \;\; \tfrac{1}{2}w^THw - b^T\pi \quad \text{subject to} \quad Z^Tw = 0, \;\; Hw - A^T\pi \ge -c. \tag{5.2}$$

The constraints $Z^Tw = 0$ force w to lie in the column space of H, i.e., $w \in \text{range}(H)$. The next result shows that any solution of the regularized dual QP (5.2) is a solution of the dual QP (5.1).

Result 5.2. (Optimality of the regularized dual QP) *A bounded solution of the regularized dual QP* (5.2) *is a solution of the dual QP* (5.1).

Proof. The regularized dual QP (5.2) is convex with a mixture of equality and inequality constraints. The optimality conditions follow from part (a) of Result 2.1. In particular, if (w, π) is a bounded solution of the regularized dual, then $Z^Tw = 0$, $Hw - A^T\pi \ge -c$, and there exist vectors v and y such that

$$\begin{pmatrix} Hw \\ -b \end{pmatrix} = \begin{pmatrix} Z & H \\ 0 & -A \end{pmatrix} \begin{pmatrix} v \\ y \end{pmatrix}, \tag{5.3}$$

with $y \ge 0$, and $y \cdot (Hw - A^T\pi + c) = 0$. The first set of equations in (5.3) gives $H(w - y) = Zv$, which implies that $Zv = 0$. As the columns of Z are linearly independent, it must hold that $v = 0$. Hence (w, π) satisfies $Hw - A^T\pi \ge -c$, and y is such that $Hw = Hy$, $Ay = b$, with $y \cdot (Hw - A^T\pi + c) = 0$, and $y \ge 0$. It follows that (w, π) satisfies the optimality conditions for the dual QP (5.1). ∎

Consider a feasible point (w,π) for the dual QP (5.2), i.e., $Hw - A^T\pi \geq -c$ with $w \in \text{range}(H)$. Our intention is to make the notation associated with the dual algorithm consistent with the notation for the primal. To do this, we break with the notation of Section 3 and use $\mathcal{B}$ to denote the working set of inequality constraints for the dual QP. This set consists of the indices of n_B linearly independent rows from $(\, H \quad -A^T\,)$. With this notation, the set $\mathcal{N} = \{\nu_1, \nu_2, \ldots, \nu_{n_N}\}$ defines the gradients of the dual inequalities that are not in the working set. The rows of $-A^T$ that appear in the working set form a matrix denoted by $-A_B^T$. Similarly, the columns of H and Z^T in the working-set matrix A_w are permuted so that

$$A_w Q = \begin{pmatrix} Z_B^T & Z_N^T & 0 \\ H_B & H_D & -A_B^T \end{pmatrix}, \tag{5.4}$$

where $Q = \text{diag}(P, I_m)$, with P a column permutation chosen to make H_B symmetric. As in Section 4, given any vector w of dimension n, w_B and w_N denote the vectors such that $[w_B]_j = w_{\beta_j}$ and $[w_N]_j = w_{\nu_j}$.

The linear independence of the rows of a dual working-set matrix A_w does not guarantee that the columns of the associated submatrix A_B form a basis for the primal QP—i.e., that A_B has rank m. For example, if H is positive definite, then Z is empty (a matrix with rank 0) and A_w has full rank regardless of the rank of A_B. These considerations lead us to impose the additional condition on the dual working set that the matrix

$$K_B = \begin{pmatrix} H_B & A_B^T \\ A_B & 0 \end{pmatrix} \tag{5.5}$$

is nonsingular. This condition ensures that A_B has rank m. To distinguish K_B from the full KKT matrix for the dual, we refer to K_B as the *reduced KKT matrix*. The next result concerns the properties of a subspace minimizer for the regularized dual QP (5.2).

Result 5.3. (Subspace minimizer for the regularized dual) *Consider the regularized dual problem* (5.2).

(a) *If (w,π) is a subspace stationary point with respect to the working-set matrix A_w of* (5.4)*, then there exists a vector y such that*

$$Hw = Hy, \quad \textit{where} \quad y = P\begin{pmatrix} y_B \\ y_N \end{pmatrix}, \;\; \textit{with} \;\; A_B y_B = b, \;\; \textit{and} \;\; y_N = 0.$$

(b) *A subspace stationary point at which the reduced KKT matrix* (5.5) *is nonsingular is a subspace minimizer for the dual.*

(c) *If (w,π) is a standard subspace minimizer, then $z_B = 0$ and $z_N \geq 0$, where $z = Hw - A^T\pi + c$ is the vector of residuals of the inequality constraints of the dual.*

Proof. Given a working-set matrix for the dual QP (5.2), a subspace stationary point satisfies

$$\begin{pmatrix} 0 \\ 0 \\ -b \end{pmatrix} + \begin{pmatrix} H_B & H_D & 0 \\ H_D^T & H_N & 0 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} w_B \\ w_N \\ \pi \end{pmatrix} = \begin{pmatrix} Z_B & H_B \\ Z_N & H_D^T \\ 0 & -A_B \end{pmatrix} \begin{pmatrix} y_Z \\ y_B \end{pmatrix},$$

where y_Z and y_B are, respectively, the multipliers for $Z^Tw = 0$ and the working-set constraints. An argument similar to that used in Result 5.2 yields $y_Z = 0$. Part (a) follows directly.

As the dual QP is convex, the dual subspace stationary point (w, π) is a subspace minimizer for the dual if the symmetrically permuted KKT matrix is nonsingular, i.e., if $Ku = 0$ implies $u = 0$, where

$$K = \left(\begin{array}{ccc|cc} H_B & H_D & 0 & Z_B & H_B \\ H_D^T & H_N & 0 & Z_N & H_D^T \\ 0 & 0 & 0 & 0 & -A_B \\ \hline Z_B^T & Z_N^T & 0 & 0 & 0 \\ H_B & H_D & -A_B^T & 0 & 0 \end{array}\right) \quad \text{and} \quad u = \begin{pmatrix} u_1 \\ u_2 \\ u_3 \\ u_4 \\ u_5 \end{pmatrix}.$$

As above, $u_4 = 0$, and the first and third sets of equations require that u_3 and u_5 satisfy

$$\begin{pmatrix} H_B & A_B^T \\ A_B & 0 \end{pmatrix} \begin{pmatrix} u_5 \\ u_3 \end{pmatrix} = 0,$$

which implies that u_5 and u_3 are zero. Finally, the fourth and fifth set of equations together with the identity $u_3 = 0$ yield $u_1 = 0$ and $u_2 = 0$. It follows that $u = 0$ and the KKT matrix is nonsingular.

Part (c) follows from the definition of a standard subspace minimizer for the dual problem (see (5.4)). ∎

At a subspace stationary point, the variables y (the dual variables of the dual problem) define a basic solution of the primal equality constraints. Moreover, the residuals of the dual inequality constraints are $z = Hw - A^T\pi + c = g(w) - A^T\pi = g(y) - A^T\pi$, which are the primal reduced-costs corresponding to both w and y. These considerations lead us to redefine the notation for the multipliers y_B of the dual working-set constraints so that $x \triangleq y_B$.

Let (w, π) be a nonoptimal dual subspace minimizer for the dual QP (5.2). (It will be shown below that if the QP gradient $g(w) = c + Hw$ is known, the vector w need not be computed explicitly.) As (w, π) is not optimal, there is at least one negative component of the dual multiplier vector x_B, say x_{β_r}. The application of the nonbinding-direction method of Section 3 to the dual gives a search direction $(\Delta w, q_\pi)$ that is feasible for the dual working-set constraints and increases a designated constraint with a negative multiplier. The constraints of the equality-constraint QP subproblem analogous to (3.1) are

$$Z_B^T \Delta w_B + Z_N^T \Delta w_N = 0, \quad \text{and} \quad H_B \Delta w_B + H_D \Delta w_N - A_B^T q_\pi = e_r.$$

The permuted equations analogous to (3.9a) for the dual direction $(\Delta w, q_\pi)$ are

$$\left(\begin{array}{ccc|cc} H_B & H_D & 0 & Z_B & H_B \\ H_D^T & H_N & 0 & Z_N & H_D^T \\ 0 & 0 & 0 & 0 & -A_B \\ \hline Z_B^T & Z_N^T & 0 & 0 & 0 \\ H_B & H_D & -A_B^T & 0 & 0 \end{array}\right) \begin{pmatrix} \Delta w_B \\ \Delta w_N \\ q_\pi \\ -p_Z \\ -p_B \end{pmatrix} = \left(\begin{array}{c} 0 \\ 0 \\ 0 \\ \hline 0 \\ e_r \end{array}\right), \tag{5.6}$$

where p_B and p_Z denote the increments for the multipliers of the dual working set. These equations give

$$H\Delta w = Zp_Z + Hp, \quad \text{where} \quad p = P\begin{pmatrix} p_B \\ p_N \end{pmatrix}, \text{ with } p_N = 0.$$

It follows that $p_Z = 0$, with p_B and q_π defined by the equations

$$\begin{pmatrix} H_B & A_B^T \\ A_B & 0 \end{pmatrix} \begin{pmatrix} p_B \\ -q_\pi \end{pmatrix} = \begin{pmatrix} e_r \\ 0 \end{pmatrix}. \tag{5.7}$$

As $p_Z = 0$, it holds that $H\Delta w = Hp$, which allows the vector $\Delta z = H\Delta w - A^T q_\pi$, the change in the residuals for the dual inequalities, to be computed as $\Delta z = Hp - A^T q_\pi$.

If the curvature $\Delta w^T H \Delta w = [p_B]_r$ is nonzero, the step $\alpha_* = -[x_B]_r/[p_B]_r$ minimizes the dual objective $\varphi_D(w + \alpha\Delta w, \pi + \alpha q_\pi)$ with respect to α, and the r-th element of $x_B + \alpha_* p_B$ is zero. If the x_B are interpreted as estimates of the primal variables, the step from x_B to $x_B + \alpha_* p_B$ increases the negative (and hence infeasible) primal variable $[p_B]_r$ until it reaches its bound of zero. If $\alpha = \alpha_*$ gives a feasible point for the dual inequalities, i.e., if the residuals $z + \alpha_*\Delta z$ of the dual inequalities are nonnegative, then the new iterate is $(w + \alpha_*\Delta w, \pi + \alpha_* q_\pi)$. In this case, the nonbinding working-set constraint is removed from the dual working set, which means that the index β_r is moved to $\mathcal{N}$ and the associated entries of H and A are removed from H_B and A_B.

If $\alpha = \alpha_*$ is unbounded, or $(w + \alpha_*\Delta w, \pi + \alpha_* q_\pi)$ is not feasible for the dual, the step is the largest α such that $g(w + \alpha\Delta w) - A^T(\pi + \alpha q_\pi)$ is nonnegative. The required value is

$$\alpha_F = \min_{1 \le i \le n_N} \{\gamma_i\}, \quad \text{where} \quad \gamma_i = \begin{cases} \dfrac{[z_N]_i}{-[\Delta z_N]_i} & \text{if } [\Delta z_N]_i < 0 \\ +\infty & \text{otherwise.} \end{cases} \tag{5.8}$$

If $\alpha_F < \alpha_*$ then at least one of the dual residuals is zero at $(w + \alpha_F\Delta w, \pi + \alpha_F q_\pi)$, and the index of one of these, ν_s say, is moved to $\mathcal{B}$. The composition of the new working set is determined by the following result, adapted to the dual QP from Result 3.3.

Result 5.4. (Test for dual constraint dependency) *Assume that (w, π) is a subspace minimizer for the dual. Assume that the ν_s-th dual inequality constraint is blocking at $(\bar{w}, \bar{\pi}) = (w, \pi) + \alpha(\Delta w, q_\pi)$, where $(\Delta w, q_\pi)$ satisfies (5.6). Define vectors u, u_π and v such that $Hu = h_{\nu_s} - Hv$, where*

$$v = P\begin{pmatrix} v_B \\ v_N \end{pmatrix}, \quad \textit{with} \quad v_N = 0, \quad \textit{and} \quad \begin{pmatrix} H_B & A_B^T \\ A_B & 0 \end{pmatrix}\begin{pmatrix} v_B \\ u_\pi \end{pmatrix} = \begin{pmatrix} (h_{\nu_s})_B \\ a_{\nu_s} \end{pmatrix}, \tag{5.9}$$

then

(a) *the gradient of the ν_s-th dual constraint is linearly independent of the gradients of the working-set constraints if and only if $u \neq 0$;*

(b) *$[v_B]_r = [\Delta z]_{\nu_s} < 0$; and if $u = 0$, then $u_\pi = 0$, otherwise $[Hu - A^T u_\pi]_{\nu_s} > 0$.*

Proof. Parts (a) and (b) are restatements of Result 3.3 in terms of the dual constraints. The additional result of part (b) that $u = 0$ implies $u_\pi = 0$ is shown as follows. If $u = 0$, then $Hv = h_{\nu_s}$, which, combined with the first set of equations of the reduced KKT system and the identity $v_N = 0$, implies $A_B^T u_\pi = 0$. As A_B^T has linearly independent columns, $u_\pi = 0$. ∎

If the vector u of Result 5.4 is zero, then $(w + \alpha_F q_\pi, \pi + \alpha_F q_\pi)$ is a subspace minimizer with respect to the working set defined with constraint β_r replaced by constraint ν_s. Otherwise, ν_s is moved to $\mathcal{B}$, which has the effect of adding the column a_{ν_s} to A_B, and adding a row and column to H_B.

The next result summarizes Results 3.2—3.5 in terms of the quantities associated with the dual QP (5.2).

Result 5.5. *Let (w, π) be a subspace minimizer with respect to the dual working set $\mathcal{B}$, with $[x_B]_r < 0$. Let $(\bar{w}, \bar{\pi})$ be the point $(w + \alpha\Delta w, \pi + \alpha q_\pi)$, where Δw and q_π are defined as in (5.6).*

(1) *If $\alpha = \alpha_*$ is bounded, then $(\bar{w}, \bar{\pi})$ is a subspace minimizer with respect to the dual working set $\bar{\mathcal{B}} = \mathcal{B} - \{\beta_r\}$.*

(2) *Alternatively, if the ν_s-th inequality is added to the dual working set at $(\bar{w}, \bar{\pi})$, let u and v_π be defined by (5.9).*

(a) *The gradient of the ν_s-th constraint is linearly independent of the gradients of the dual working-set constraints if and only if $u \neq 0$.*

(b) *If the gradient of the ν_s-th constraint is linearly independent of the gradients of the dual working-set constraints, then $(\bar{w}, \bar{\pi})$ is a subspace minimizer with respect to the dual working set $\bar{\mathcal{B}} = \mathcal{B} + \{\nu_s\}$. Moreover, the vector $\bar{x}_B$ at $(\bar{w}, \bar{\pi})$ is formed by adding a zero component to the vector $x_B + \alpha p_B$.*

(c) *If the gradient of the ν_s-th constraint is linearly dependent on the gradients of the dual working-set constraints, define $\sigma = [x_B + \alpha p_B]_r / [v_B]_r$. Then $(\bar{w}, \bar{\pi})$ is a subspace minimizer with respect to $\bar{\mathcal{B}} = \mathcal{B} - \{\beta_r\} + \{\nu_s\}$ with associated multipliers $\bar{x}_B$ given by the vector $x_B + \alpha p_B - \sigma v_B$ with its r-th component replaced by σ.* ∎

As in the primal methods of Sections 3 and 4, the search directions may be updated if no column swap is made. For the dual QP, the updated directions involve the scalar $\rho = -\Delta z_{\nu_s} / [Hu - A^T u_\pi]_{\nu_s}$, which is positive from Result 5.4. Given ρ, the updated directions are $\bar{p}_B$ and $\bar{q}_\pi$, where $\bar{p}_B$ is the vector $p_B - \rho v_B$ augmented by ρ, and $\bar{q}_\pi = q_\pi + \rho u_\pi$.

In practice, it is not necessary to compute the vector u of Result 5.4 explicitly. If the vector $h_{\nu_s} - Hv$ is zero then Hu must be zero and the restriction $Z^T u = 0$ imposed by the full dual KKT equations analogous to (3.9b) implies that $u = 0$.

It remains to show that K_B remains nonsingular throughout the computation.

Result 5.6. (Nonsingularity of the reduced KKT matrix) *Consider an iteration of the dual nonbinding direction method that starts at a subspace minimizer (w, π) at which the reduced KKT matrix (5.5) is nonsingular. Then the next iterate $(\bar{w}, \bar{\pi})$ is also a subspace minimizer with a nonsingular reduced KKT matrix.*

Proof. Let $K_{\bar{B}}$ denote the reduced KKT matrix for $\bar{\mathcal{B}}$, the basic set at the next iterate. The three cases to consider are the addition, deletion and replacement of a working-set constraint.

First, consider the case of a constraint deleted from the working set, so that $\alpha = \alpha_*$ and $\bar{\mathcal{B}} = \mathcal{B} - \{\beta_r\}$. Since α is finite, it follows that $[p_B]_r > 0$. Define the matrix

$$M = \begin{pmatrix} K_B & e_r \\ e_r^T & \end{pmatrix}.$$

Then $M/K_B = -e_r^T K_B^{-1} e_r = -[p_B]_r < 0$ so that $\operatorname{In}(M) = (0,1,0) + \operatorname{In}(K_B)$.

Since inertia is unchanged by symmetric permutations, we consider a permuted version of M,

$$\widetilde{M} = \begin{pmatrix} h_{\beta_r,\beta_r} & 1 & (h_{\beta_r})_{\bar{\mathcal{B}}}^T & a_{\beta_r}^T \\ 1 & 0 & 0 & 0 \\ (h_{\beta_r})_{\bar{\mathcal{B}}} & 0 & H_{\bar{B}} & A_{\bar{B}}^T \\ a_{\beta_r} & 0 & A_{\bar{B}}^T & \end{pmatrix}.$$

Let E denote the nonsingular 1×1 block of $\widetilde{M}$, with $\operatorname{In}(E) = (1,1,0)$. Then

$$\widetilde{M}/E = K_{\bar{B}} - \begin{pmatrix} (h_{\beta_r})_{\mathcal{B}} & 0 \\ a_{\beta_r} & 0 \end{pmatrix} \begin{pmatrix} 0 & 1 \\ 1 & -h_{\beta_r,\beta_r} \end{pmatrix} \begin{pmatrix} (h_{\beta_r})_{\mathcal{B}}^T & a_{\beta_r}^T \\ 0 & 0 \end{pmatrix} = K_{\bar{B}},$$

so that $\mathrm{In}(\widetilde{M}) = \mathrm{In}(M) = \mathrm{In}(K_{\bar{B}}) + (1,1,0)$. Combining this with the above relation for $\mathrm{In}(M)$ implies $\mathrm{In}(K_{\bar{B}}) = \mathrm{In}(K_B) - (1,0,0)$.

Now assume that a constraint is added to the working set, with $\bar{\mathcal{B}} = \mathcal{B} + \{\nu_s\}$. This occurs when a feasible step is taken and $u \neq 0$. We show that the matrix

$$M = \begin{pmatrix} H_B & A_B^T & (h_{\nu_s})_{\mathcal{B}} \\ A_B & & a_{\nu_s} \\ (h_{\nu_s})_{\mathcal{B}}^T & a_{\nu_s}^T & h_{\nu_s,\nu_s} \end{pmatrix},$$

(which has the same inertia as $K_{\bar{B}}$) is nonsingular. Since K_B is nonsingular, the Schur complement M/K_B exists such that

$$\begin{aligned} M/K_B &= h_{\nu_s,\nu_s} - \big((h_{\nu_s})_{\mathcal{B}}^T \quad a_{\nu_s}^T\big) K_B^{-1} \begin{pmatrix} (h_{\nu_s})_{\mathcal{B}} \\ a_{\nu_s} \end{pmatrix} \\ &= h_{\nu_s,\nu_s} - \big((h_{\nu_s})_{\mathcal{B}}^T \quad a_{\nu_s}^T\big) \begin{pmatrix} v_B \\ u_\pi \end{pmatrix} \\ &= e_s^T (h_{\nu_s})_{\mathcal{N}} - e_s^T H_D^T v_B - e_s^T A_N^T u_\pi \\ &= [Hu - A^T u_\pi]_{\nu_s}, \end{aligned}$$

where the last equality holds from the equations (5.9) of Result 5.4. Part (b) of the same result implies $M/K_B > 0$. Then $\mathrm{In}(K_{\bar{B}}) = (1,0,0) + \mathrm{In}(K_B)$ and $K_{\bar{B}}$ is nonsingular.

Finally, assume that a constraint is replaced in the working-set, with $\bar{\mathcal{B}} = \mathcal{B} + \{\nu_s\} - \{\beta_r\}$. In this case it must hold that $u = 0$ and $u_\pi = 0$. If v denotes the vector $v = (v_B, u_\pi) = (v_B, 0)$, then

$$K_B v = K_B \begin{pmatrix} v_B \\ u_\pi \end{pmatrix} = \begin{pmatrix} H_B \\ A_B \end{pmatrix} v_B = \begin{pmatrix} (h_{\nu_s})_{\mathcal{B}} \\ a_{\nu_s} \end{pmatrix}.$$

The updated reduced KKT matrix can be written in terms of the symmetric rank-one modification to K_B:

$$\begin{aligned} K_{\bar{B}} &= K_B + (K_B v - K_B e_r) e_r^T + e_r (K_B v - K_B e_r)^T + e_r \big((v - e_r)^T K_B (v - e_r)\big) e_r^T \\ &= \big(I + e_r (v - e_r)^T\big) K_B \big(I + (v - e_r) e_r^T\big). \end{aligned}$$

Since $[v_B]_r \neq 0$ by part (b) of Result 5.4, the matrix $I + e_r(v - e_r)^T$ and its transpose are nonsingular. Therefore, $\mathrm{In}(K_{\bar{B}}) = \mathrm{In}(K_B)$.

∎

Algorithm 5.1 summarizes the dual nonbinding direction method for solving a convex quadratic programming problem in standard form.

5.3. Dual linear programs

If the primal QP is a linear program, then $H = 0$ and we may choose Z as the identity matrix for the regularized problem (5.2). It follows from Result 5.3 that (w, π) is a subspace minimizer if A_B is nonsingular (i.e., it is square with rank m). In this case, equations (5.7) and (5.9) give $p_B = 0$, $u = 0$ and $u_\pi = 0$, with q_π and v_B determined from the nonsingular systems $A_B^T q_\pi = -e_r$ and $A_B v_B = a_{\nu_s}$. As $u = 0$, the linearly dependent case always applies and indices r and s are always swapped in $\mathcal{B}$, as in the dual simplex method. The update defined by part (b) of Result 3.4 for the dual multiplier vector x_B is given by $\bar{x}_B = x_B - \sigma v_B$, with $\sigma = [x_B]_r/[v_B]_r$.

Algorithm 5.1.[Dual nonbinding direction method for a convex QP in standard form]

Find (x_0, π_0) such that $Ax_0 = b$ and $c + Hx_0 - A^T\pi_0 \geq 0$;
$[x, \pi, \mathcal{B}, \mathcal{N}] =$ `subspaceMin`(x_0, π_0);
$g = c + Hx$; $z = g - A^T\pi$;
$\beta_r = \text{argmin}_i\{[x_B]_i\}$;
while $x_{\beta_r} < 0$ **do**
  Solve $\begin{pmatrix} H_B & A_B^T \\ A_B & \end{pmatrix}\begin{pmatrix} p_B \\ -q_\pi \end{pmatrix} = \begin{pmatrix} e_r \\ 0 \end{pmatrix}$; $p_N = 0$; $p = P\begin{pmatrix} p_B \\ p_N \end{pmatrix}$; $\Delta z = Hp - A^T q_\pi$;
  $\alpha_F =$ `minRatioTest`$(z_N, \Delta z_N)$;
  if $[p_B]_r > 0$ **then** $\alpha_* = -[x_B]_r/[p_B]_r$ **else** $\alpha_* = +\infty$;
  $\alpha = \min\{\alpha_*, \alpha_F\}$;
  if $\alpha = +\infty$ **then stop**; [the primal is infeasible]
  $x \leftarrow x + \alpha p$; $g \leftarrow g + \alpha Hp$;
  $\pi \leftarrow \pi + \alpha q_\pi$; $z = g - A^T\pi$;
  if $\alpha_F < \alpha_*$ **then** [add the dual working-set constraint ν_s]
    Find the blocking constraint index ν_s;
    Solve $\begin{pmatrix} H_B & A_B^T \\ A_B & \end{pmatrix}\begin{pmatrix} v_B \\ u_\pi \end{pmatrix} = \begin{pmatrix} (h_{\nu_s})_{\mathcal{B}} \\ a_{\nu_s} \end{pmatrix}$; $v_N = 0$; $v = P\begin{pmatrix} v_B \\ v_N \end{pmatrix}$;
    if $h_{\nu_s} - Hv = 0$ **then** $\sigma = [x_B]_r/[v_B]_r$ **else** $\sigma = 0$;
    $\mathcal{B} \leftarrow \mathcal{B} + \{\nu_s\}$; $\mathcal{N} \leftarrow \mathcal{N} - \{\nu_s\}$;
    $\pi \leftarrow \pi - \sigma u_\pi$; $z = g - A^T\pi$;
  end;
  if $x_{\beta_r} = 0$ **then** [delete the dual working-set constraint β_r]
    $\mathcal{B} \leftarrow \mathcal{B} - \{\beta_r\}$; $\mathcal{N} \leftarrow \mathcal{N} + \{\beta_r\}$;
    $\beta_r = \text{argmin}_i\{[x_B]_i\}$;
  end;
  $k \leftarrow k + 1$;
end do

5.4. Finding an initial dual-feasible point

An initial dual-feasible point may be defined by applying a conventional phase-one method to the dual constraints, e.g., by minimizing the sum of infeasibilities for the dual constraints $Hw - A^T\pi \geq -c$. If H is nonsingular and A has full rank, we may define $\mathcal{N} = \emptyset$ and compute (x_0, π_0) from the equations

$$\begin{pmatrix} H & A^T \\ A & 0 \end{pmatrix}\begin{pmatrix} x_0 \\ -\pi_0 \end{pmatrix} = -\begin{pmatrix} c \\ -b \end{pmatrix}.$$

This choice of $\mathcal{B}$ gives $Ax_0 = b$, with (w_0, π_0) a dual subspace minimizer.

5.5. Degeneracy

The dual problem (5.1) has fewer inequality constraints than variables, which implies that if H and A have no common nontrivial null vector, then the dual constraint gradients, the rows of $\begin{pmatrix} H & -A^T \end{pmatrix}$, are linearly independent, and the dual feasible region has no degenerate points. In this situation, Algorithm 5.1 cannot cycle, and will either terminate with an optimal solution or declare the dual problem to be unbounded. This nondegeneracy property does not hold for a dual linear program, but it does hold for strictly convex problems, and for any QP with H and A of the form

$$H = \begin{pmatrix} \bar{H} & 0 \\ 0 & 0 \end{pmatrix} \quad \text{and} \quad A = \begin{pmatrix} \bar{A} & -I_m \end{pmatrix},$$

where $\bar{H}$ is an $(n-m)\times(n-m)$ positive-definite matrix.

6. Two Implementations

At each iteration of the primal and dual methods discussed in Sections 4 and 5, it is necessary to solve one or two systems of the form

$$\begin{pmatrix} H_B & A_B^T \\ A_B & 0 \end{pmatrix} \begin{pmatrix} y \\ w \end{pmatrix} = \begin{pmatrix} h \\ f \end{pmatrix}, \tag{6.1}$$

where h and f are given by right-hand sides of the equations (4.3) or (4.4). This section discusses two alternative approaches for solving (6.1). The first involves the symmetric transformation of the KKT system into three smaller systems, one of which involves the explicit reduced Hessian matrix. The second approach uses a symmetric indefinite factorization of a fixed KKT matrix in conjunction with the factorization of a smaller matrix that is updated at each iteration.

6.1. Variable reduction

The variable-reduction method involves transforming the equations (6.1) to block-triangular form using the nonsingular block-diagonal matrix $\operatorname{diag}(Q, I_m)$. Consider a column permutation P such that

$$AP = \begin{pmatrix} B & S & N \end{pmatrix},$$

with B an $m\times m$ nonsingular matrix and S an $m\times n_S$ matrix with $n_S = n_B - m$ (the matrix P represents the particular permutation of (4.2) such that $A_B = \begin{pmatrix} B & S \end{pmatrix}$ and $A_N = N$). The n_S variables associated with S are called the *superbasic* variables. Given P, consider the nonsingular $n\times n$ matrix Q such that

$$Q = P \begin{pmatrix} -B^{-1}S & I_m & 0 \\ I_{n_S} & 0 & 0 \\ 0 & 0 & I_N \end{pmatrix}.$$

The columns of Q may be partitioned so that $Q = \begin{pmatrix} Z & Y & W \end{pmatrix}$, where

$$Z = P \begin{pmatrix} -B^{-1}S \\ I_{n_S} \\ 0 \end{pmatrix}, \quad Y = P \begin{pmatrix} I_m \\ 0 \\ 0 \end{pmatrix} \quad \text{and} \quad W = P \begin{pmatrix} 0 \\ 0 \\ I_N \end{pmatrix}.$$

The columns of the $n \times n_S$ matrix Z form a basis for the null-space of A_w, with

$$A_w Q = \begin{pmatrix} A \\ E_w \end{pmatrix} Q = \begin{pmatrix} 0 & B & N \\ 0 & 0 & I_N \end{pmatrix}.$$

Suppose that we wish to solve a generic KKT system

$$\begin{pmatrix} H & A^T & E_w^T \\ A & & \\ E_w & & \end{pmatrix} \begin{pmatrix} y \\ w_1 \\ w_2 \end{pmatrix} = \begin{pmatrix} h \\ f_1 \\ f_2 \end{pmatrix}.$$

Then the vector y may be computed as $y = Yy_Y + Zy_Z + Wy_W$, where y_Y, y_Z, y_W and w are defined using the equations

$$\begin{pmatrix} Z^THZ & Z^THY & Z^THW & & \\ Y^THZ & Y^THY & Y^THW & B^T & \\ W^THZ & W^THY & W^THW & N^T & I_N \\ & B & N & & \\ & & I_N & & \end{pmatrix} \begin{pmatrix} y_Z \\ y_Y \\ y_W \\ w_1 \\ w_2 \end{pmatrix} = \begin{pmatrix} h_Z \\ h_Y \\ h_W \\ f_1 \\ f_2 \end{pmatrix}, \tag{6.2}$$

with $h_Z = Z^Th$, $h_Y = Y^Th$, and $h_W = W^Th$. This leads to

$$\begin{aligned} y_W &= f_2, & & \\ By_Y &= f_1 - Nf_2, & y_R &= Yy_Y + Wy_W, & & \\ Z^THZy_Z &= Z^T(h - Hy_R), & y_T &= Zy_Z, & y &= y_R + y_T, \\ B^Tw_1 &= Y^T(h - Hy), & w_2 &= W^T(h - Hy) - N^Tw_1. & & \end{aligned}$$

These equations may be solved using a Cholesky factorization of Z^THZ and an LU factorization of B. The factors of B allow efficient calculation of matrix-vector products Z^Tv or Zv without the need to form the inverse of B.

The equations simplify considerably for the KKT systems (3.9a) and (3.9b). In the case of (3.9a), the equations are:

$$\begin{aligned} Bp_Y &= -a_{\nu_s}, & p_R &= P\begin{pmatrix} p_Y \\ 0 \\ e_s \end{pmatrix}, & & \\ Z^THZp_Z &= -Z^THp_R, & p_T &= Zp_Z, & p &= p_R + p_T, \\ B^Tq_B &= \ (Hp)_{\mathcal{B}}, & q_Z &= (Hp - A^Tq_B)_{\mathcal{N}}. & & \end{aligned} \tag{6.3}$$

Similarly for (3.9b), it holds that $u_Y = 0$, $u_R = 0$, and

$$\begin{aligned} Z^THZu_Z &= Z^Te_{\beta_r}, & u &= Zu_Z, \\ B^Tv_B &= (e_{\beta_r} - Hu)_{\mathcal{B}}, & v_Z &= -(Hu + A^Tv_B)_{\mathcal{N}}. \end{aligned} \tag{6.4}$$

These equations allow us to specialize Part 2(a) of Result 4.2, which gives the conditions for the linear independence of the new matrix A_B.

Result 6.1. *Let x be a subspace minimizer with respect to the basic set $\mathcal{B}$. Assume that p and q are defined by (4.3) and that x_{β_r} is the incoming nonbasic variable at the next iterate. Let the vectors u_B and v_π be defined by (4.4).*

- (i) *If x_{β_r} is superbasic, then e_r and the rows of A_B are linearly independent (i.e., the matrix obtained by removing the rth column of A_B has rank m).*
- (ii) *If x_{β_r} is not superbasic, then e_r and the rows of A_B are linearly independent if and only if $S^Tz \neq 0$, where z is the solution of $B^Tz = e_r$.*

Proof. From (6.4), $u = Zu_Z$, which implies that u_B is nonzero if and only if u_Z is nonzero. Similarly, the nonsingularity of Z^THZ implies that u_Z is nonzero if and only if $Z^Te_{\beta_r}$ is nonzero. Now

$$Z^Te_{\beta_r} = \begin{pmatrix} -S^TB^{-T} & I_{n_S} & 0 \end{pmatrix} e_r.$$

If $r > m$, then x_{β_r} will change from being superbasic to nonbasic, and $Z^Te_{\beta_r} = e_{r-m} \neq 0$. However, if $r \le m$, then

$$Z^Te_{\beta_r} = -S^TB^{-T}e_r = -S^Tz,$$

where z is the solution of $B^Tz = e_r$. ∎

6.2. Fixed-factorization updates

Solving a single linear system can be done very effectively using sparse matrix factorization techniques. However, within a QP algorithm, many closely related systems must be solved where the KKT matrix differs by a single row and column. Instead of reformulating the matrix at each iteration, the matrix may be "bordered" in a way that reflects the changes to the basic and nonbasic sets (see Bisschop and Meeraus [2], and Gill et al. [15]).

Let $\mathcal{B}_0$ and $\mathcal{N}_0$ denote the initial basic and nonbasic sets that define the KKT system in (6.1). There are four cases to consider:

(1) a nonbasic variable moves to the basic set and is not in $\mathcal{B}_0$,

(2) a basic variable in $\mathcal{B}_0$ becomes nonbasic,

(3) a basic variable not in $\mathcal{B}_0$ becomes nonbasic, and

(4) a nonbasic variable moves to the basic set and is in $\mathcal{B}_0$.

For case (1), let ν_s be the nonbasic variable that has become basic. The next KKT matrix can be written as

$$\left(\begin{array}{cc|c} H_B & A_B^T & (h_{\nu_s})_{\mathcal{B}_0} \\ A_B & 0 & a_{\nu_s} \\ \hline (h_{\nu_s})_{\mathcal{B}_0}^T & a_{\nu_s}^T & h_{\nu_s,\nu_s} \end{array}\right).$$

Suppose that at the next stage, another nonbasic variable ν_r becomes basic. The KKT matrix is augmented in a similar fashion, i.e.,

$$\left(\begin{array}{ccc|c} H_B & A_B^T & (h_{\nu_s})_{\mathcal{B}_0} & (h_{\nu_r})_{\mathcal{B}_0} \\ A_B & 0 & a_{\nu_s} & a_{\nu_r} \\ (h_{\nu_s})_{\mathcal{B}_0}^T & a_{\nu_s}^T & h_{\nu_s,\nu_s} & h_{\nu_s,\nu_r} \\ \hline (h_{\nu_r})_{\mathcal{B}_0}^T & a_{\nu_r}^T & h_{\nu_r,\nu_s} & h_{\nu_r,\nu_r} \end{array}\right).$$

Now consider case 2 and let $\beta_r \in \mathcal{B}_0$ become nonbasic. The change to the basic set is reflected in the new KKT matrix

$$\left(\begin{array}{cccc|c} H_B & A_B^T & (h_{\nu_s})_{\mathcal{B}_0} & (h_{\nu_r})_{\mathcal{B}_0} & e_r \\ A_B & 0 & a_{\nu_s} & a_{\nu_r} & 0 \\ (h_{\nu_s})_{\mathcal{B}_0}^T & a_{\nu_s}^T & h_{\nu_s,\nu_s} & h_{\nu_s,\nu_r} & 0 \\ (h_{\nu_r})_{\mathcal{B}_0}^T & a_{\nu_r}^T & h_{\nu_r,\nu_s} & h_{\nu_r,\nu_r} & 0 \\ \hline e_r^T & 0 & 0 & 0 & 0 \end{array}\right).$$

The unit row and column augmenting the matrix has the effect of zeroing out the components corresponding to the removed basic variable.

In case (3), the basic variable must have been added to the basic set at a previous stage as in case (1). Thus, removing it from the basic set can be done by removing the row and column in the augmented part of the KKT matrix corresponding to its addition to the basic set. For example, if ν_s is the basic to be removed, then the new KKT matrix is given by

$$\begin{pmatrix} H_B & A_B^T & (h_{\nu_r})_{\mathcal{B}_0} & e_r \\ A_B & 0 & a_{\nu_r} & 0 \\ (h_{\nu_r})_{\mathcal{B}_0}^T & a_{\nu_r}^T & h_{\nu_r,\nu_r} & 0 \\ e_r^T & 0 & 0 & 0 \end{pmatrix}.$$

For case (4), a nonbasic variable in $\mathcal{B}_0$ implies that at some previous stage, the variable was removed from $\mathcal{B}_0$ as in case (2). The new KKT matrix can be formed by removing the unit row and column in the augmented part of the KKT matrix corresponding to the removal the variable from the basic set. In this example, the new KKT matrix becomes

$$\begin{pmatrix} H_B & A_B^T & (h_{\nu_r})_{\mathcal{B}_0} \\ A_B & 0 & a_{\nu_r} \\ (h_{\nu_r})_{\mathcal{B}_0}^T & a_{\nu_r}^T & h_{\nu_r,\nu_r} \end{pmatrix}.$$

After k iterations, the KKT system is maintained as a symmetric augmented system of the form

$$\begin{pmatrix} K & V \\ V^T & D \end{pmatrix} \begin{pmatrix} r \\ \eta \end{pmatrix} = \begin{pmatrix} b \\ f \end{pmatrix} \quad \text{with} \quad K = \begin{pmatrix} H_B & A_B^T \\ A_B & \end{pmatrix}, \tag{6.5}$$

where D is of dimension at most $2k$.

6.3. Schur complement and block LU methods

Although the augmented system (in general) increases in dimension by one at every iteration, the 1×1 block K is fixed and defined by the initial set of basic variables. The *Schur complement method* assumes that factorizations for K and the *Schur complement* $C = D - V^T K^{-1} V$ exist. Then the solution of (6.5) can be determined by solving the equations

$$Kt = b, \quad C\eta = f - V^T t, \quad Kr = b - V\eta.$$

The work required is dominated by two solves with the fixed matrix K and one solve with the Schur complement C. If the number of changes to the basic set is small enough, dense factors of C may be maintained.

The Schur complement method can be extended to a *block LU method* by storing the augmented matrix in block factors

$$\begin{pmatrix} K & V \\ V^T & D \end{pmatrix} = \begin{pmatrix} L & \\ Z^T & I \end{pmatrix} \begin{pmatrix} U & Y \\ & C \end{pmatrix}, \tag{6.6}$$

where $K = LU$, $LY = V$, $U^T Z = V$, and $C = D - Z^T Y$ is the Schur-complement matrix.

The solution of (6.5) can be computed by forming the block factors and by solving the equations

$$Lt = b, \quad C\eta = f - Z^T t, \quad Ur = t - Y\eta.$$

This method requires a solve with L and U each, one multiply with Y and Z^T, and one solve with the Schur complement C. For more details, see Gill et al. [14], Eldersveld and Saunders [9], and Huynh [21].

As the iterations of the QP algorithm proceed, the size of C increases and the work required to solve with C increases. It may be necessary to restart the process by discarding the existing factors and re-forming K based on the current set of basic variables.

6.3.1. Updating the block LU factors

Suppose the current KKT matrix is bordered by the vectors v and w, and the scalar σ

$$\left(\begin{array}{cc|c} K & V & v \\ V^T & D & w \\ \hline v^T & w^T & \sigma \end{array}\right).$$

The block LU factors Y and Z, and the Schur complement C are updated every time the system is bordered, the matrices can be updated. The number of columns in matrices Y and Z and the dimension of the Schur complement increase by one. The updates y, z, c and d are defined by the equations

$$\begin{aligned} Ly &= v, & U^T z &= v, \\ c = w - Z^T y &= w - Y^T z, & d &= \sigma - z^T y, \end{aligned}$$

so that the new block LU factors satisfy

$$\left(\begin{array}{c|cc} K & V & v \\ \hline V^T & D & w \\ v^T & w^T & \sigma \end{array}\right) = \left(\begin{array}{c|cc} L & & \\ \hline Z^T & I & \\ z^T & & 1 \end{array}\right) \left(\begin{array}{c|cc} U & Y & y \\ \hline & C & c \\ & c^T & d \end{array}\right).$$

7. Finding the first subspace minimizer

The methods described in Sections 4 and 5 have the property that if the initial iterate x_0 is a subspace minimizer, then all subsequent iterates are subspace minimizers (see Result 4.2). Methods for finding an initial subspace minimizer utilize an initial estimate x_I of the solution together with matrices A_B and A_N associated with an estimate of the optimal basic and nonbasic partitions of A. These estimates are often available from the known solution of a related QP—e.g., from the solution of the previous QP subproblem in the SQP context. The initial point x_I may or may not be feasible, and the associated matrix A_B may or may not have rank m.

The methods of Sections 4 and 5 require that the matrix A_B has rank m, and it is necessary to identify a set of linearly independent basic columns of A. One algorithm for doing this has been proposed by Gill, Murray and Saunders [13], who use a sparse LU factorization of A_B^T to identify a square nonsingular subset of the columns of A_B. If necessary, a "basis repair" scheme is used to define additional unit columns that make A_B have full rank. The nonsingular matrix B obtained as a by-product of this process may be expressed in terms of A using a column permutation P such that

$$AP = \begin{pmatrix} A_B & A_N \end{pmatrix} = \begin{pmatrix} B & S & A_N \end{pmatrix}. \tag{7.1}$$

Given x_I, a point x_0 satisfying $Ax = b$ may be computed as

$$x_0 = x_I + P \begin{pmatrix} p_Y \\ 0 \\ 0 \end{pmatrix}, \quad \text{where} \quad Bp_Y = -(Ax_I - b).$$

The calculation of an initial subspace minimizer will depend on how the KKT equations during the regular iterations (see Section 6). If the matrix

$$K_B = \begin{pmatrix} H_B & A_B^T \\ A_B & 0 \end{pmatrix} \tag{7.2}$$

has m negative eigenvalues, then the inertia of K_B is correct and x_0 is used as the initial point for a sequence of Newton-type iterations in which $\varphi(x)$ is minimized with the nonbasic components of x fixed at their current values. Consider the equations

$$\begin{pmatrix} H_B & A_B^T \\ A_B & 0 \end{pmatrix} \begin{pmatrix} p_B \\ \pi \end{pmatrix} = -\begin{pmatrix} g_B \\ 0 \end{pmatrix}.$$

If p_B is zero, x is a subspace stationary point (with respect to A_B) at which K_B has correct inertia and we are done. If p_B is nonzero, two situations are possible.

If $x_B + p_B$ is infeasible, then feasibility is retained by determining the maximum nonnegative step $\alpha < 1$ such that $x_B + \alpha p_B$ is feasible. A variable on its bound at $x_B + \alpha p_B$ is then removed from the basic set and the iteration is repeated. The removal of a basic variable cannot increase the number of negative eigenvalues of K_B and a subspace minimizer must be determined in a finite number of steps.

If $x_B + p_B$ is feasible, then p_B is the step to the minimizer of $\varphi(x)$ with respect to the basic variables and it must hold that $x_B + p_B$ is a subspace minimizer.

A KKT matrix with incorrect inertia has too many negative eigenvalues. In this case, an appropriate K_B may be obtained by defining additional artificial constraints that are deleted during the course of subsequent iterations. For example, if $n - m$ variables are temporarily fixed at their current values, then A_B is a square nonsingular matrix and K_B necessarily has exactly m negative eigenvalues. The formulation of the artificial constraints depends on the method used to solve the reduced KKT equations.

7.1. Variable-reduction method

In the variable reduction method a dense Cholesky factor of the reduced Hessian $Z^T H Z$ is updated to reflect changes in the basic set. (see Section 6.1). At the initial x_0 a partial Cholesky factorization with interchanges is used to find an upper-triangular matrix R that is the factor of the largest positive-definite leading submatrix of $Z^T H Z$. The use of interchanges tends to maximize the dimension of R. Let Z_R denote the columns of Z corresponding to R, and let Z be partitioned as $Z = \begin{pmatrix} Z_R & Z_A \end{pmatrix}$. A nonbasic set for which Z_R defines an appropriate null space can be obtained by fixing the variables corresponding to the columns of Z_A at their current values. As described above, minimization of $\varphi(x)$ then proceeds within the subspace defined by Z_R. If a variable is removed from the basic set, a row and column is removed from the reduced Hessian and an appropriate update is made to the Cholesky factor.

7.2. Fixed factorization updates

If fixed factorization updates to the KKT matrix are being used, the procedure for finding an initial subspace minimizer is given as follows.

1. Factor the reduced KKT matrix (7.2) system in the form $K_B = LDL^T$, where L is unit lower-triangular and D is block diagonal with 1×1 and 2×2 blocks. If the inertia of K_B is correct, then we are done.

2. If the inertia is incorrect, factor

   $$H_A = H_B + \rho A_B^T A_B = L_A D_A L_A^T,$$

   where ρ is a modest positive penalty parameter. As the inertia of K_B is not correct, D_A will have some negative eigenvalues for all positive ρ.

   The factorization of H_A may be written in the form

   $$H_A = L_A U \Lambda U^T L_A^T = V \Lambda V^T,$$

   where $U \Lambda U^T$ is the spectral decomposition of D_A. The block diagonal structure of D_A implies that U is a block-diagonal orthonormal matrix. The inertia of Λ is the same as the inertia of H_A, and there exists a positive semidefinite diagonal matrix E such that $\Lambda + E$ is positive definite. If $\bar{H}_A$ is the positive-definite matrix $V(\Lambda + E)V^T$, then

   $$\bar{H}_A = H_A + VEV^T = H_A + \sum_{e_{jj} > 0} e_{jj} v_j v_j^T.$$

Suppose that H_A has r negative eigenvalues. Define V_B as the $r \times n_B$ matrix consisting of the columns of V associated with the positive components of E. The augmented KKT matrix

$$\begin{pmatrix} H_B & A_B^T & V_B \\ A_B & 0 & 0 \\ V_B^T & 0 & 0 \end{pmatrix}$$

has exactly $m + r$ negative eigenvalues and hence has correct inertia.

3. The minimization of $\varphi(x)$ proceeds subject to the original constraints and the (general) artificial constraints $V_B^T x_B = 0$.

The efficiency of this scheme will depend on the number of negative eigenvalues of H_A. In practice, if the number of negative eigenvalues exceeds a preassigned threshold, then an temporary vertex is defined by fixing the variables associated with the columns of S in (7.1).

8. Getting feasible

The QP method formulated in Section 4 assumes that the initial point x_0 is feasible, i.e., x_0 satisfies $Ax_0 = b$ and $x_0 \geq 0$. There are two approaches to finding a feasible point. The first, common in linear programming, is to find a point x_0 that satisfies $Ax = b$ and then iterate (if necessary) to satisfy the bounds $x \geq 0$. The second method defines a nonnegative x_0 and iterates to satisfy the constraints $Ax = b$. Here we use the former approach and assume that the initial iterate x_0 satisfies $Ax_0 = b$ (such an x_0 must exist under the assumption that A has rank m). The idea is to minimize the norm of the violations of the nonnegativity constraints subject to the linear constraints and satisfied bounds.

Suppose that the bounds $x \geq 0$ are written in the equivalent form $x = u - v$, $u \geq 0$ and $v = 0$. The idea is to relax the equality constraint $v = 0$ by minimizing some norm of v. The choice of the one-norm gives the following piecewise-linear program for a feasible point:

$$\underset{x\in\mathbb{R}^n;\; u,v\in\mathbb{R}^n}{\text{minimize}} \; \|v\|_1 \quad \text{subject to} \quad Ax = b, \;\; x - u + v = 0, \;\; u \geq 0.$$

By adding the restriction $v \geq 0$, the objective $\|v\|_1$ may be replaced by $e^T v$, giving the conventional linear program

$$\underset{x\in\mathbb{R}^n;\; u,v\in\mathbb{R}^n}{\text{minimize}} \; e^T v \quad \text{subject to} \quad Ax = b, \;\; x - u + v = 0, \;\; u \geq 0, \;\; v \geq 0.$$

The optimal u and v are the magnitudes of the positive and negative parts of the vector x satisfying $Ax = b$ that is closest in one-norm to the positive orthant. If the constraints are feasible, then $v = 0$ and $x\,(= u) \geq 0$.

At an initial x_0 such that $Ax_0 = b$, the v_i corresponding to feasible components of x_0 may be fixed at zero, so that the number of infeasibilities cannot increase during subsequent iterations. In this case, if the constraints are infeasible, the optimal solution minimizes the sum of the violations of those bounds that are violated at x_0, subject to $Ax = b$. It is necessary to allow every element of v to move if the sum of the violations is to be minimized.

In practice, the variables u and v need not be stored explicitly, and the computation may be arranged in such a way that the equations to be solved have the same dimension as those of the second phase of a conventional two-phase method. During the solution of the QP, the search is restricted to pairs (u, v) with components satisfying $u_i \geq 0$, $v_i \geq 0$, and $u_i v_i = 0$. A feasible pair (u, v) is reconstructed from any x such that $Ax = b$. In particular, $(u_i, v_i) = (x_i, 0)$ if $x_i \geq 0$, and $(u_i, v_i) = (0, -x_i)$ if $x_i < 0$. This implies that at any given iteration, an infeasible x_i must be basic because it corresponds to $(u_i, v_i) = (0, -x_i)$ an (implicit) positive elastic variable v_i. In implicit form, the strategy of fixing $v_i = 0$

for components associated with feasible components of x is equivalent to imposing the original nonnegativity constraint once a variable is feasible. An important practical benefit is derived by allowing several constraints to become feasible in one step. When treating the u-v variables implicitly, this is done by allowing an infeasible variable to "pass through" a bound and become strictly positive if this reduces either the number of infeasibilities or the (piecewise) sum of infeasibilities. Explicitly, this is equivalent to performing a swap of the basic variables u_i and v_i associated with the infeasible variable $x_i = u_i - v_i$.

The minimum one-norm problem is equivalent to the standard method for minimizing the sum of infeasibilities that has been used in QP and LP packages for many years. The introduction of the u and v variables permits the simple formulation of single-phase methods based on minimizing a sequence of linearly constrained ℓ_1-penalty functions. This method is based on the following equivalent form of a QP in standard form:

$$\underset{x\in\mathbb{R}^n;\ u,v\in\mathbb{R}^n}{\text{minimize}}\ \varphi(x) \quad \text{subject to} \quad Ax = b,\ \ x - u + v = 0,\ \ u \ge 0,\ \ v = 0.$$

The first-order optimality conditions may be written in the form:

$$\begin{aligned} Ax - b &= 0, & u &\ge 0, & & \\ x - u + v &= 0, & v &= 0, & & \\ c + Hx - A^T\pi - z &= 0, & y_u &= z, & y_v &= -z, \\ y_u \cdot u &= 0, & y_u &\ge 0, & & \end{aligned}$$

where y_u and y_v are the multipliers for the constraints $u \ge 0$ and $v = 0$. These conditions imply that the dual variables z satisfy $z \ge 0$ and $z \cdot u = 0$, so that the z's are the QP reduced costs when the problem is feasible.

The composite ℓ_1 method is defined by minimizing the ℓ_1 penalty function $\varphi(x) + \rho e^T v$ subject to the linear constraints. This gives the quadratic program:

$$\underset{x\in\mathbb{R}^n;\ u,v\in\mathbb{R}^n}{\text{minimize}}\ \varphi(x) + \rho e^T v \quad \text{subject to} \quad Ax = b,\ \ x - u + v = 0,\ \ u \ge 0,\ \ v \ge 0,$$

where ρ is a positive penalty parameter. This technique is often called *elastic programming* in the linear and nonlinear programming literature (see, e.g., Brown and Graves [5], and Gill, Murray and Saunders [13]). In this context, the variables u and v are known as *elastic variables.*

The composite ℓ_1 method is ideally suited to an implementation based on the variable-reduction method for solving the KKT equations (see Section 6.1). If the fixed-factorization (i.e., Schur-complement) method of Section 6.2 is used, a composite objective method may be defined in terms of an augmented Lagrangian function. This function is based on the properties of a composite objective function defined in terms of the quadratic penalty function. Consider an increasing sequence of penalty parameters ρ and solving a corresponding sequence of QPs of the form

$$\underset{x\in\mathbb{R}^n;\ u,v\in\mathbb{R}^n}{\text{minimize}}\ \varphi(x) + \tfrac{1}{2}\rho\|v\|_2^2 \quad \text{subject to} \quad Ax = b,\ \ x - u + v = 0,\ \ u \ge 0,\ \ v \ge 0. \tag{8.1}$$

The optimality conditions of this problem are

$$\begin{aligned} Ax - b &= 0, & u &\ge 0, & & \\ x - u + v &= 0, & & & & \\ c + Hx - A^T\pi - z &= 0, & y_u &= z, & z &= \rho v, \\ y_u \cdot u &= 0, & y_u &\ge 0. & & \end{aligned} \tag{8.2}$$

These conditions imply that $z = \rho v \ge 0$, and $u \cdot v = 0$.

Consider a hypothetical method based on solving (8.1) for an increasing sequence of values of ρ. As in the ℓ_1 case, (8.1) may be solved without defining u and v explicity. Let $\mathcal{V}(x)$ denote the indices of the violated constraints at a point x. Let E_V denote a matrix with rows e_i^T with $i \in \mathcal{V}(x)$ (i.e., E_V is the matrix of normals of the violated constraints at x). A typical KKT system has the form

$$\begin{pmatrix} H_B + \rho E_V^T E_V & A_B^T \\ A_B & \end{pmatrix} \begin{pmatrix} y \\ w \end{pmatrix} = \begin{pmatrix} h \\ f \end{pmatrix}.$$

Note that all infeasible x components are necessarily basic. These equations may be solved by calculating y and w as part of the solution of the bordered equations

$$\begin{pmatrix} H_B & A_B^T & E_V^T \\ A_B & & \\ E_V & & -\frac{1}{\rho} I_V \end{pmatrix} \begin{pmatrix} y \\ w \\ w_V \end{pmatrix} = \begin{pmatrix} h \\ f \\ 0 \end{pmatrix}, \tag{8.3}$$

where I_V is the identity matrix of dimension $|\mathcal{V}(x)|$. Observe that the numerical conditioning of these equations does not deteriorate as ρ increases.

As the QP iterations proceed, in addition to the usual changes to the basic-nonbasic partition associated with an x-component moving on or off its bound, a violated (and hence basic) x_j may increase from its negative value and become feasible. This corresponds to the associated v_j and u_j being swapped in a basis for (8.1), i.e., v_j will move on to its bound and u_j will move off its bound. Thus, in the "implicit" form of (8.1), x_j remains basic, but moves inside its bound, requiring the removal of its contribution to the composite objective function. This amounts to removing one of the diagonals from the penalty term in the matrix $H_B + \rho E_V^T E_V$, which can be done by performing a Schur-complement update that removes the appropriate row and column from the border of (8.3).

A somewhat less welcome result is that even if the original QP is feasible, some of the constraints violated at x_0 may remain violated at a solution of (8.1) *for all positive values of* ρ. In order to illustrate why this is the case, assume that the original QP is feasible, with solution (x^*, π^*, z^*), and let $(\bar{x}, \bar{\pi}, \bar{z})$ denote a solution of (8.1) for a given value of ρ. The optimality conditions (8.2) imply that if x_i^* is active with a positive reduced cost z_i^*, then, for sufficiently large ρ, it holds that $\bar{x}_i < 0$, i.e., $\bar{x}_i \le 0$ for all active x_i^* and the violated constraints predict the active set. If the original QP is infeasible, the optimal u and v approximate the magnitudes of the positive and negative parts of the x satisfying $Ax = b$ closest in two-norm to the positive orthant.

This difficulty may be eliminated by including a Lagrangian term in the definition of the composite objective, giving the linearly constrained augmented Lagrangian:

$$\begin{array}{ll} \underset{x\in\mathbb{R}^n;\ u,v\in\mathbb{R}^n}{\text{minimize}} & \varphi(x) + z_E^T v + \frac{1}{2}\rho\|v\|_2^2 \\ \text{subject to} & Ax = b, \quad x - u + v = 0, \quad u \ge 0, \end{array}$$

where ρ is a positive penalty parameter and z_E is an estimate of the multipliers for the constraints $u \ge 0$ (which are the negative of multiplier estimates of the constraints $v = 0$). In this case, the penalty parameter need not go to infinity.

9. Summary

We have considered computational methods for finding a point satisfying the second-order necessary conditions for a general (possibly nonconvex) quadratic program (QP). The methods are based on a framework for the formulation and analysis of feasible-point active-set methods that defines a primal-dual search pair as the solution of an equality-constrained subproblem involving a "working set" of linearly independent constraints. This framework is discussed in the context of two broad classes of active-set method for quadratic programming: *binding-direction methods* and *nonbinding-direction methods*. A binding-direction method for general QP first proposed by Fletcher, and subsequently modified by Gould, is recast as a nonbinding-direction method. This reformulation gives the primal-dual search pair as the solution of a KKT-system formed from the QP Hessian and the working-set constraint gradients. It is shown that, under certain circumstances, the solution of this KKT-system may be updated using a simple recurrence relation, thereby giving a significant reduction in the number of KKT systems that need to be solved. The nonbinding-direction framework is applied to QP problems with constraints in standard form, and to the dual of a convex QP.

In the second part of the paper, two approaches are presented for solving the constituent KKT systems. The first approach uses a variable-reduction technique requiring the calculation of the Cholesky factor of the reduced Hessian. The second approach uses a symmetric indefinite factorization of a fixed KKT matrix in conjunction with the factorization of a smaller matrix that is updated at each iteration. Finally, a linearly constrained augmented Lagrangian is proposed that obviates the need for a separate procedure for finding a feasible point.

References

[1] R. A. Bartlett and L. T. Biegler. QPSchur: a dual, active-set, Schur-complement method for large-scale and structured convex quadratic programming. *Optim. Eng.*, 7(1):5–32, 2006.

[2] J. Bisschop and A. Meeraus. Matrix augmentation and partitioning in the updating of the basis inverse. *Math. Program.*, 13:241–254, 1977.

[3] N. L. Boland. A dual-active-set algorithm for positive semi-definite quadratic programming. *Math. Programming*, 78(1, Ser. A):1–27, 1997.

[4] J. M. Borwein. Necessary and sufficient conditions for quadratic minimality. *Numer. Funct. Anal. and Optimiz.*, 5:127–140, 1982.

[5] G. G. Brown and G. W. Graves. Elastic programming: A new approach to large-scale mixed-integer optimization, November 1975. Presented at the presented at ORSA/TIMS meeting, Las Vegas, NV.

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

[7] R. W. Cottle, G. J. Habetler, and C. E. Lemke. On classes of copositive matrices. *Linear Algebra Appl.*, 3:295–310, 1970.

[8] W. S. Dorn. Duality in quadratic programming. *Quart. Appl. Math.*, 18:155–162, 1960/1961.

[9] S. K. Eldersveld and M. A. Saunders. A block-LU update for large-scale linear programming. *SIAM J. Matrix Anal. Appl.*, 13:191–201, 1992.

[10] R. Fletcher. A general quadratic programming algorithm. *J. Inst. Math. Applics.*, 7:76–91, 1971.

[11] 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.

[12] P. E. Gill. Recent developments in numerically stable methods for linear programming. *IMA Bulletin*, 10:180–186, 1973.

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

[14] P. E. Gill, W. Murray, M. A. Saunders, and M. H. Wright. Sparse matrix methods in optimization. *SIAM J. Sci. Statist. Comput.*, 5(3):562–589, 1984.

[15] P. E. Gill, W. Murray, M. A. Saunders, and M. H. Wright. A Schur-complement method for sparse quadratic programming. In M. G. Cox and S. J. Hammarling, editors, *Reliable Numerical Computation*, pages 113–138. Oxford University Press, 1990.

[16] P. E. Gill, W. Murray, and M. H. Wright. *Numerical Linear Algebra and Optimization, volume 1.* Addison-Wesley Publishing Company, Redwood City, 1991.

[17] P. E. Gill and E. Wong. Sequential quadratic programming methods. In J. Lee and S. Leyffer, editors, *Mixed-Integer Nonlinear Optimization: Algorithmic Advances and Applications*, The IMA Volumes in Mathematics and its Applications. Springer Verlag, Berlin, Heidelberg and New York, 2010. To appear.

[18] D. Goldfarb and A. Idnani. A numerically stable dual method for solving strictly convex quadratic programs. *Math. Programming*, 27(1):1–33, 1983.

[19] N. I. M. Gould. On practical conditions for the existence and uniqueness of solutions to the general equality quadratic programming problem. *Math. Program.*, 32:90–99, 1985.

[20] N. I. M. Gould. An algorithm for large-scale quadratic programming. *IMA J. Numer. Anal.*, 11(3):299–324, 1991.

[21] H. M. Huynh. *A Large-Scale Quadratic Programming Solver Based on Block-LU Updates of the KKT System.* PhD thesis, Program in Scientific Computing and Computational Mathematics, Stanford University, Stanford, CA, 2008.

[22] A. Majthay. Optimality conditions for quadratic programming. *Math. Programming*, 1:359–365, 1971.

[23] J. Nocedal and S. J. Wright. *Numerical Optimization.* Springer-Verlag, New York, 1999.

[24] P. M. Pardalos and G. Schnitger. Checking local optimality in constrained quadratic programming is NP-hard. *Oper. Res. Lett.*, 7(1):33–35, 1988.

[25] P. M. Pardalos and S. A. Vavasis. Quadratic programming with one negative eigenvalue is NP-hard. *J. Global Optim.*, 1(1):15–22, 1991.

[26] M. J. D. Powell. On the quadratic programming algorithm of Goldfarb and Idnani. *Math. Programming Stud.*, (25):46–61, 1985.

[27] J. A. Tomlin. On pricing and backward transformation in linear programming. *Math. Programming*, 6:42–47, 1974.