

# ON WARM STARTS FOR INTERIOR METHODS

Anders FORSGREN\*

Technical Report TRITA-MAT-2005-OS4  
Department of Mathematics  
Royal Institute of Technology  
October 2005

## Abstract

An appealing feature of interior methods for linear programming is that the number of iterations required to solve a problem tends to be relatively insensitive to the choice of initial point. This feature has the drawback that it is difficult to design interior methods that efficiently utilize information from an optimal solution to a “nearby” problem. We discuss this feature in the context of general nonlinear programming and specialize to linear programming. We demonstrate that warm start for a particular nonlinear programming problem, given a near-optimal solution for a “nearby” problem, is closely related to an SQP method applied to an equality-constrained problem. These results are further refined for the case of linear programming.

**Key words.** nonlinear programming, linear programming, interior method, warm start.

**AMS subject classifications.** 49M37, 65K05, 90C30

## 1. Introduction

This paper concerns the solution of a nonlinear program in the form

$$\begin{aligned} & \underset{x \in \mathbb{R}^n}{\text{minimize}} && f(x) \\ & \text{subject to} && c(x) \geq 0, \end{aligned} \tag{1.1}$$

where  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  and  $c : \mathbb{R}^n \rightarrow \mathbb{R}^m$  are twice-continuously differentiable. Our interest is the situation where we want to solve (1.1) given the solution to a “nearby” problem. This situation is commonly referred to as *warm start*. It may for example be the case that one is interested in resolving the problem when only some constraints have been changed. Our discussion specifically concerns *interior methods*. We will study properties of the search directions generated for the nonlinear programming case, and then specialize further to linear programming. For related discussions concerning linear programming, see, e.g. Jansen et al. [11], Kim, Park and Park [12], Gondzio and Grothey [9], Yildirim and Todd [18, 19], Yildirim and Wright [20], Gonzalez-Lima, Wei and Wolkowicz [10]. For extensions to linear semidefinite programming, see Yildirim [17].

---

\*Optimization and Systems Theory, Department of Mathematics, Royal Institute of Technology, SE-100 44 Stockholm, Sweden ([andersf@kth.se](mailto:andersf@kth.se)). Research supported by the Swedish Research Council (VR).

## 2. Background

Methods for solving (1.1) all have to solve a combinatorial problem of identifying the constraints that are active at the solution. This can roughly be done in two ways: (i) by a “hard” estimate of the active constraints at each iteration or by (ii) a “soft” estimate. Our focus is interior methods, which belong to the latter class. However, in the discussion, sequential quadratic programming methods, which belong to the former type of methods, arise too. In this section, we review basic properties of these methods, and also give a brief review of optimality conditions.

### 2.1. Optimality conditions

Given a suitable *constraint qualification*, an optimal solution  $x$  to (1.1), together with a *Lagrange multiplier vector*  $\lambda \in \mathbb{R}^m$ , has to satisfy the *first-order necessary optimality conditions* associated with (1.1). These conditions may be written in the form

$$g(x) - A(x)^T y = 0, \quad (2.1a)$$

$$c(x) - s = 0, \quad (2.1b)$$

$$Y S e = 0, \quad y \geq 0, \quad s \geq 0, \quad (2.1c)$$

where  $g(x) = \nabla f(x)$ ,  $A(x) = c'(x)$  and  $e$  is an  $m$ -dimensional vector of ones. Here, and throughout the paper, we denote by upper-case letters  $Y$  and  $S$ , the diagonal matrices formed by  $y$  and  $s$  respectively. In (2.1) we have introduced the slack variables  $s$  associated with the constraints  $c(x) \geq 0$ . They need not be present, since  $s$  may be eliminated from (2.1b). They do not affect the discussion, but simplify the notation. The analogous discussion could be made without introducing  $s$ . A constraint qualification ensures that a linearization of the constraints around a point of interest gives a suitable approximation to the constraints. We will throughout the paper assume that a constraint qualification holds at all points which we consider. For a more detailed discussion of constraint qualifications in the context of interior methods, see, e.g., Forsgren, Gill and Wright [6, Section 2.2].

Second-order optimality conditions typically involve the Hessian of the *Lagrangian*  $L(x, y)$  with respect to  $x$ , where  $L(x, y) = f(x) - y^T c(x)$ . We denote this Hessian by  $H(x, y)$ , i.e.,

$$H(x, y) = \nabla_{xx}^2 L(x, y) = \nabla^2 f(x) - \sum_{i=1}^m y_i \nabla^2 c_i(x).$$

Given a suitable constraint qualification, the curvature of the objective function on the surface of the active constraints is captured by the curvature of  $H(x, y)$  on the tangent surface of the active constraints. For further discussion, see, e.g., Forsgren, Gill and Wright [6, Section 2.2].

### 2.2. Interior methods

The interior methods of interest to this paper are based on approximately following the *barrier trajectory*. This trajectory is defined as the set of solutions to the

perturbed optimality conditions

$$g(x) - A(x)^T y = 0, \quad (2.2a)$$

$$c(x) - s = 0, \quad (2.2b)$$

$$Y S e = \mu e, \quad (2.2c)$$

where  $y > 0$  and  $s > 0$  are held implicitly. Here,  $\mu$  is a positive parameter, referred to as the *barrier parameter*. As in (2.1), we have introduced slack variables  $s$ . This slack reformulation is not of importance for the discussion, but convenient for the notation.

A primal-dual interior method computes approximate solution to (2.2) for decreasing values of  $\mu$  by Newton's method. This means that the steps  $\Delta x$ ,  $\Delta s$  and  $\Delta y$  are computed from the linear equation

$$\begin{pmatrix} H(x, y) & & -A(x)^T \\ A(x) & -I & \\ & Y & S \end{pmatrix} \begin{pmatrix} \Delta x \\ \Delta s \\ \Delta y \end{pmatrix} = - \begin{pmatrix} g(x) - A(x)^T y \\ c(x) - s \\ Y S e - \mu e \end{pmatrix}. \quad (2.3)$$

Equivalently, we may eliminate  $\Delta s$  and solve

$$\begin{pmatrix} H(x, y) & A(x)^T \\ A(x) & -Y^{-1} S \end{pmatrix} \begin{pmatrix} \Delta x \\ -\Delta y \end{pmatrix} = - \begin{pmatrix} g(x) - A(x)^T y \\ c(x) - \mu S^{-1} e \end{pmatrix}. \quad (2.4)$$

Note that there is no loss in sparsity when forming (2.4) from (2.3), since  $Y$  is diagonal. Local convexity is typically deduced by the inertia of the matrix of (2.4). If the inertia is such that the matrix has  $n$  positive eigenvalues and  $m$  negative eigenvalues, the equations are solved. Otherwise, some modification is made. The solution of this equation can either be done by factorization methods or by iterative methods. See, e.g., Forsgren [4] and Forsgren, Gill and Griffin [5] for a discussion of these issues.

In order to enforce convergence of the method, typically a linesearch strategy, a trust-region strategy or a filter strategy may be used. We shall not be concerned with the precise method, but focus on the linear equations to be solved. For more detailed descriptions on interior methods, see e.g., Forsgren, Gill and Wright [6] or Wright [15]. Note that for a solution of (2.2), no constraints are active, since  $\mu > 0$ . Hence, the active constraints at the solution are determined implicitly as  $\mu$  tends to zero.

### 2.3. Sequential quadratic programming methods

In contrast to an interior method, where one system of linear equations is solved at each iteration, a sequential programming method has a subproblem which is an inequality-constrained quadratic program on the form

$$\begin{aligned} & \underset{p \in \mathbb{R}^n}{\text{minimize}} && \frac{1}{2} p^T H(x, y) p + g(x)^T p \\ & \text{subject to} && A(x) p \geq -c(x). \end{aligned} \quad (2.5)$$

If the problem is locally convex, this subproblem is well defined. Otherwise, some modification is made. We will assume that local convexity holds in our discussion, and not consider the modifications. If  $p^*$  denotes the optimal solution of (2.5) and  $y^*$  denotes the corresponding Lagrange multiplier vector, then the next iterate for solving (1.1) is given by  $x + p^*$ , and the next Lagrange multiplier estimate is given by  $y^*$ . Again, some strategy is required to ensure convergence, but the basis of the subproblem is the solution of a quadratic program on the form (2.5). Note that the prediction of the constraints that are active at the solution of (1.1) are given by the constraints active at the solution of (2.5). For a thorough discussion on sequential quadratic programming methods, see, e.g., Nocedal and Wright [13, Chapter 18].

### 3. Warm starts of interior methods for nonlinear programming

We now return to the issue of solving (1.1) by a primal-dual interior method. Specifically, we consider the warm-start situation. Assume that the initial point is given as a near-optimal solution on the trajectory to a different problem

$$\begin{aligned} & \underset{x \in \mathbb{R}^n}{\text{minimize}} && \tilde{f}(x) \\ & \text{subject to} && \tilde{c}(x) \geq 0, \end{aligned} \quad (3.1)$$

where  $\tilde{f}$  and  $\tilde{c}$  have the same properties as  $f$  and  $c$  of (1.1). We denote this point by  $\tilde{x}$ . This means that for a small barrier parameter  $\tilde{\mu}$ , we assume that  $\tilde{x}$ ,  $\tilde{s}$  and  $\tilde{y}$  solve

$$\tilde{g}(\tilde{x}) - \tilde{A}(\tilde{x})^T \tilde{y} = 0, \quad (3.2a)$$

$$\tilde{c}(\tilde{x}) - \tilde{s} = 0, \quad (3.2b)$$

$$\tilde{Y} \tilde{S} e = \tilde{\mu} e, \quad \tilde{y} > 0, \quad \tilde{s} > 0, \quad (3.2c)$$

where  $\tilde{g}(\tilde{x}) = \nabla \tilde{f}(\tilde{x})$  and  $\tilde{A}(\tilde{x}) = \tilde{c}'(\tilde{x})$ . Throughout the paper, we will consider quantities related to matrices that are implicitly dependent on  $\tilde{\mu}$ . The notation  $O(\tilde{\mu})$  will be used to denote a quantity that converges to zero at least as fast as  $\tilde{\mu}$ . Analogously,  $\Theta(\tilde{\mu})$  denotes a quantity that converges to zero at the same rate as  $\tilde{\mu}$ ,  $\Theta(1)$  denotes a bounded quantity that is bounded away from zero as  $\tilde{\mu} \rightarrow 0$  and  $\Omega(1/\tilde{\mu})$  denotes a quantity whose inverse converges to zero at least as fast as  $\tilde{\mu}$ .

From the point given by (3.1), we want to take a primal-dual interior step towards solving (1.1) for a given barrier parameter  $\mu$ . By (2.3) and (3.2), the Newton equations take the form

$$\begin{pmatrix} H(\tilde{x}, \tilde{y}) & & -A(\tilde{x})^T \\ A(\tilde{x}) & -I & \\ & \tilde{Y} & \tilde{S} \end{pmatrix} \begin{pmatrix} \Delta x \\ \Delta s \\ \Delta y \end{pmatrix} = \begin{pmatrix} -g(\tilde{x}) + A(\tilde{x})^T \tilde{y} \\ \tilde{c}(\tilde{x}) - c(\tilde{x}) \\ (\mu - \tilde{\mu})e \end{pmatrix}. \quad (3.3)$$

In order to make a more detailed analysis, we assume that the set  $\{1, 2, \dots, m\}$  can be partitioned into two sets,  $A$  and  $I$ , according to which constraints of (3.1) that are "almost active" at  $\tilde{x}$ , and which constraints that are "not almost active" at  $\tilde{x}$ .

This means that  $A \cup I = \{1, \dots, m\}$ , where  $A = \{i \in \{1, 2, \dots, m\} : \tilde{c}_i(\tilde{x}) = \Theta(\tilde{\mu})\}$  and  $I = \{i \in \{1, 2, \dots, m\} : \tilde{c}_i(\tilde{x}) = \Theta(1)\}$ . This would typically be the case in the neighborhood of a local minimizer of (1.1) where strict complementarity holds. We will throughout this section let subscript “ $A$ ” denote quantities associated with  $A$ , and similarly for  $I$ . For example, the matrix  $A(\tilde{x})$  is partitioned into  $A_A(\tilde{x})$  and  $A_I(\tilde{x})$ . By the above assumption, it follows that  $\tilde{y}_I = \Theta(\tilde{\mu})$ , and that  $\tilde{s}_A = \Theta(\tilde{\mu})$ . We will also make the assumption that  $A_A(\tilde{x})$  has full row rank, and that  $H(\tilde{x}, \tilde{y})$  is positive definite on the nullspace of  $A_A(\tilde{x})$ . With this partition, (3.3) may be written as

$$\begin{pmatrix} H(\tilde{x}, \tilde{y}) & & -A_A(\tilde{x})^T & -A_I(\tilde{x})^T \\ A_A(\tilde{x}) & -I & & \\ A_I(\tilde{x}) & & -I & \\ & \tilde{Y}_A & & \tilde{S}_A \\ & & \tilde{Y}_I & \tilde{S}_I \end{pmatrix} \begin{pmatrix} \Delta x \\ \Delta s_A \\ \Delta s_I \\ \Delta y_A \\ \Delta y_I \end{pmatrix} = \begin{pmatrix} -g(\tilde{x}) + A(\tilde{x})^T \tilde{y} \\ -c_A(\tilde{x}) \\ \tilde{c}_I(\tilde{x}) - c_I(\tilde{x}) \\ (\mu - \tilde{\mu})e \\ (\mu - \tilde{\mu})e \end{pmatrix}.$$

We may now approximate these equations by ignoring the  $\Theta(\tilde{\mu})$  terms in the matrix, which gives a related system of equations

$$\begin{pmatrix} H(\tilde{x}, \tilde{y}) & & -A_A(\tilde{x})^T & -A_I(\tilde{x})^T \\ A_A(\tilde{x}) & -I & & \\ A_I(\tilde{x}) & & -I & \\ & \tilde{Y}_A & & \\ & & & \tilde{S}_I \end{pmatrix} \begin{pmatrix} u \\ v_A \\ v_I \\ z_A \\ z_I \end{pmatrix} = \begin{pmatrix} -g(\tilde{x}) + A(\tilde{x})^T \tilde{y} \\ -c_A(\tilde{x}) \\ \tilde{c}_I(\tilde{x}) - c_I(\tilde{x}) \\ (\mu - \tilde{\mu})e \\ (\mu - \tilde{\mu})e \end{pmatrix}.$$

where  $v_A$  and  $\Delta y_I$  may be eliminated so as to give the equivalent equations

$$v_A = (\mu - \tilde{\mu}) \tilde{Y}_A^{-1} e, \quad (3.4a)$$

$$z_I = (\mu - \tilde{\mu}) \tilde{S}_I^{-1} e, \quad (3.4b)$$

$$\begin{pmatrix} H(\tilde{x}, \tilde{y}) & A_A(\tilde{x})^T \\ A_A(\tilde{x}) & \end{pmatrix} \begin{pmatrix} u \\ -z_A \end{pmatrix} = \begin{pmatrix} -g(\tilde{x}) + A(\tilde{x})^T \tilde{y} + A_I(\tilde{x})^T z_I \\ -c_A(\tilde{x}) + v_A \end{pmatrix}, \quad (3.4c)$$

$$v_I = c_I(\tilde{x}) - \tilde{c}_I(\tilde{x}) + A_I(\tilde{x})u. \quad (3.4d)$$

From (3.4), we may identify  $u$  and  $\tilde{y}_A + z_A$  as the solution and Lagrange multipliers of an equality-constrained quadratic program. By our assumptions, the difference between  $\Delta x$ ,  $\Delta s$ ,  $\Delta y$ , and  $u$ ,  $v$  and  $z$ , respectively, is  $O(\tilde{\mu})$ , as the following lemma states.

**Lemma 3.1.** *Let  $\tilde{x}$ ,  $\tilde{y}$ , and  $\tilde{s}$  satisfy (3.2). Assume that (i)  $A_A(\tilde{x})$  has full row rank, (ii) that  $H(\tilde{x}, \tilde{y})$  is positive definite on the nullspace of  $A_A(\tilde{x})$ , (iii) that  $\tilde{y}_I = \Theta(\tilde{\mu})$ , and (iv) that  $\tilde{s}_A = \Theta(\tilde{\mu})$ . Further, let  $\Delta x$ ,  $\Delta s$  and  $\Delta y$  satisfy (3.3), and let  $u$ ,  $v$  and  $z$  satisfy (3.4). Then,  $\Delta x = u + O(\tilde{\mu})$ ,  $\Delta s = v + O(\tilde{\mu})$  and  $\Delta y = z + O(\tilde{\mu})$ .*

**Proof.** The quantities  $u$ ,  $v$  and  $z$  are solutions of a system of linear equations whose matrix is bounded and nonsingular as  $\tilde{\mu} \rightarrow 0$ , by our assumptions. Hence, since  $\Delta x$ ,

$\Delta y$  and  $\Delta s$  satisfy (3.2), where the only difference is that some  $O(\tilde{\mu})$  elements have been added to the matrix, the result follows. ■

This means that we may use properties of  $u$ ,  $v$  and  $w$  to deduce properties of our desired quantities  $\Delta x$ ,  $\Delta s$  and  $\Delta y$ , as stated in the following theorem.

**Theorem 3.1.** *Let  $\tilde{x}$ ,  $\tilde{y}$ , and  $\tilde{s}$  satisfy (3.2). Assume that (i)  $A_A(\tilde{x})$  has full row rank, (ii) that  $H(\tilde{x}, \tilde{y})$  is positive definite on the nullspace of  $A_A(\tilde{x})$ , (iii) that  $\tilde{y}_I = \Theta(\tilde{\mu})$ , and (iv) that  $\tilde{s}_A = \Theta(\tilde{\mu})$ . Further, let  $\Delta x$ ,  $\Delta s$  and  $\Delta y$  satisfy (3.3). Then,  $\Delta x$  differs by  $O(\tilde{\mu})$  from the optimal solution to the equality-constrained quadratic program*

$$\begin{aligned} & \underset{p \in \mathbb{R}^n}{\text{minimize}} && \frac{1}{2} p^T H(\tilde{x}, \tilde{y}) p + (g(\tilde{x}) - (\mu - \tilde{\mu}) A_I(\tilde{x})^T \tilde{S}_I^{-1} e)^T p \\ & \text{subject to} && A_A(\tilde{x}) p = -c_A(\tilde{x}) + (\mu - \tilde{\mu}) \tilde{Y}_A^{-1} e, \end{aligned} \quad (3.5)$$

and  $\tilde{y}_A + \Delta y_A$  differs by  $O(\tilde{\mu})$  from the associated Lagrange multipliers.

**Proof.** The optimality conditions for (3.5) are given by

$$H(\tilde{x}, \tilde{y}) p^* + g(\tilde{x}) - (\mu - \tilde{\mu}) A_I(\tilde{x})^T \tilde{S}_I^{-1} e = A_A(\tilde{x})^T \lambda_A^*, \quad (3.6a)$$

$$A_A(\tilde{x}) p^* = -c_A(\tilde{x}) + (\mu - \tilde{\mu}) \tilde{Y}_A^{-1} e, \quad (3.6b)$$

for an optimal solution  $p^*$  together with a Lagrange multiplier vector  $\lambda_A^*$ . Rearrangement of (3.6), taking into account that  $\tilde{x}$ ,  $\tilde{s}$  and  $\tilde{y}$  satisfy (3.2), gives

$$H(\tilde{x}, \tilde{y}) p^* - A_A(\tilde{x})^T (\lambda_A^* - \tilde{y}_A) = -g(\tilde{x}) + A_A(\tilde{x})^T \tilde{y}_A + (\mu - \tilde{\mu}) A_I(\tilde{x})^T \tilde{S}_I^{-1} e \quad (3.7a)$$

$$A_A(\tilde{x}) p^* = -c_A(\tilde{x}) + (\mu - \tilde{\mu}) \tilde{Y}_A^{-1} e. \quad (3.7b)$$

By comparing (3.6c) with (3.7) it follows that  $p^* = u$  and  $\lambda_A^* - \tilde{y}_A = z_A$ . Lemma 3.1 now gives  $\Delta x = p^* + O(\tilde{\mu})$  and  $\lambda_A^* - \tilde{y}_A = \Delta y_A + O(\tilde{\mu})$ , as required. ■

A consequence of Theorem 3.1 is that if  $\mu = \tilde{\mu}$ , then the step is near-optimal to the equality-constrained problem where the active constraint are set as equalities, as summarized in the following corollary.

**Corollary 3.1.** *Let  $\tilde{x}$ ,  $\tilde{y}$ , and  $\tilde{s}$  satisfy (3.2). Assume that (i)  $A_A(\tilde{x})$  has full row rank, (ii) that  $H(\tilde{x}, \tilde{y})$  is positive definite on the nullspace of  $A_A(\tilde{x})$ , (iii) that  $\tilde{y}_I = \Theta(\tilde{\mu})$ , and (iv) that  $\tilde{s}_A = \Theta(\tilde{\mu})$ . Further, let  $\Delta x$ ,  $\Delta s$  and  $\Delta y$  satisfy (3.3) for  $\mu = \tilde{\mu}$ . Then,  $\Delta x$  differs by  $O(\tilde{\mu})$  from the optimal solution to the equality-constrained quadratic program*

$$\begin{aligned} & \underset{p \in \mathbb{R}^n}{\text{minimize}} && \frac{1}{2} p^T H(\tilde{x}, \tilde{y}) p + g(\tilde{x})^T p \\ & \text{subject to} && A_A(\tilde{x}) p = -c_A(\tilde{x}), \end{aligned}$$

and  $\tilde{y}_A + \Delta y_A$  differs by  $O(\tilde{\mu})$  from the associated Lagrange multipliers.

Another consequence is that for  $\mu = \tilde{\mu}$ , the step is near-optimal to the ‘‘appropriate’’ inequality-constrained quadratic programming problem.

**Corollary 3.2.** *Let  $\tilde{x}$ ,  $\tilde{y}$ , and  $\tilde{s}$  satisfy (3.2). Assume that (i)  $A_A(\tilde{x})$  has full row rank, (ii) that  $H(\tilde{x}, \tilde{y})$  is positive definite on the nullspace of  $A_A(\tilde{x})$ , (iii) that  $\tilde{y}_I = \Theta(\tilde{\mu})$ , and (iv) that  $\tilde{s}_A = \Theta(\tilde{\mu})$ . Further, let  $\Delta x$ ,  $\Delta s$  and  $\Delta y$  satisfy (3.3) for  $\mu = \tilde{\mu}$ . Then,  $\Delta x$  differs by  $O(\tilde{\mu})$  from the optimal solution to the equality-constrained quadratic program*

$$\begin{aligned} & \underset{p \in \mathbb{R}^n}{\text{minimize}} && \frac{1}{2} p^T H(\tilde{x}, \tilde{y}) p + g(\tilde{x})^T p \\ & \text{subject to} && a_i(\tilde{x})^T p \geq -c_i(\tilde{x}), \quad i \in I_A^+, \\ & && a_i(\tilde{x})^T p \leq -c_i(\tilde{x}), \quad i \in I_A \setminus I_A^+, \end{aligned} \quad (3.8)$$

and  $\tilde{y} + \Delta y$  differ by  $O(\tilde{\mu})$  from the associated Lagrange multipliers, where  $I_A^+ = \{i \in A : \tilde{y}_i + z_i \geq 0\}$ , where  $z$  is given by (3.4).

The conclusion is that if  $\mu$  is small, of the order of  $\tilde{\mu}$ , the primal-dual step behaves like the sequential quadratic programming step, i.e., for small perturbations we may expect the step to give a near-optimal solution, but for larger perturbations, the step is likely to violate both inactive constraints (that are ignored) and nonnegativity of the multipliers. In addition, we have not taken into account the implicit requirement that  $y$  and  $s$  have to be maintained positive.

#### 4. Warm starts of interior methods for linear programming

The above analysis applies also to linear programming. However, linear programming is special in the sense that there always exists a strictly complementary optimal solution, if an optimal solution exists, and the analysis may be specialized further. To be consistent with the discussion in Section 3, we will consider the linear program

$$\begin{aligned} & \underset{x \in \mathbb{R}^n}{\text{minimize}} && d^T x \\ & \text{subject to} && Ax \geq b, \end{aligned} \quad (4.1)$$

and a near-optimal solution to the related linear program

$$\begin{aligned} & \underset{x \in \mathbb{R}^n}{\text{minimize}} && \tilde{d}^T x \\ & \text{subject to} && Ax \geq \tilde{b}. \end{aligned} \quad (4.2)$$

An underlying assumption is that the constraint matrix  $A$  has full column rank. Our analysis is “classical” sensitivity analysis in the sense that the constraint matrix  $A$  is assumed fixed, whereas the cost coefficients and the right hand sides may differ.

The analysis of the previous section applies. However, here we need not make any assumption about nonsingularity of the resulting limiting Newton system. Again, we assume that a “small” barrier  $\tilde{\mu}$  is given, and that  $\tilde{x}$ ,  $\tilde{s}$  and  $\tilde{y}$  solve

$$\tilde{d} - A^T \tilde{y} = 0, \quad (4.3a)$$

$$A \tilde{x} - \tilde{b} - \tilde{s} = 0, \quad (4.3b)$$

$$\tilde{Y} \tilde{S} e = \tilde{\mu}, \quad \tilde{y} > 0, \quad \tilde{s} > 0. \quad (4.3c)$$

In addition to the assumption on full column rank of  $A$ , we make the assumption that the barrier trajectory for the perturbed problem (4.2) is well defined, i.e.,  $\{(x, s) : Ax - s = \tilde{b}, s > 0\} \neq \emptyset$  and  $\{y : A^T y = \tilde{c}, y > 0\} \neq \emptyset$ .

From the initial point given by  $\tilde{x}$ ,  $\tilde{s}$  and  $\tilde{y}$ , we want to take a primal-dual interior step towards solving (1.1) for a given barrier parameter  $\mu$ . The Newton equations may be written as

$$\begin{pmatrix} & A^T & & \\ A & -I & & \\ & \tilde{Y} & \tilde{S} & \end{pmatrix} \begin{pmatrix} \Delta x \\ \Delta s \\ \Delta y \end{pmatrix} = \begin{pmatrix} d - \tilde{d} \\ b - \tilde{b} \\ 0 \end{pmatrix} + (\mu - \tilde{\mu}) \begin{pmatrix} 0 \\ 0 \\ e \end{pmatrix}. \quad (4.4)$$

If  $\Delta s$  is eliminated from (4.4), the resulting equivalent system of equations becomes

$$\begin{pmatrix} & A^T \\ A & \tilde{W}^{-1} \end{pmatrix} \begin{pmatrix} \Delta x \\ \Delta y \end{pmatrix} = \begin{pmatrix} d - \tilde{d} \\ b - \tilde{b} \end{pmatrix} + (\mu - \tilde{\mu}) \begin{pmatrix} 0 \\ \tilde{Y}^{-1}e \end{pmatrix}, \quad (4.5a)$$

$$\Delta s = A\Delta x + \tilde{b} - b, \quad (4.5b)$$

where  $\tilde{W} = \tilde{S}^{-1}\tilde{Y}$ . By further eliminating  $\Delta y$ , we obtain

$$A^T\tilde{W}A\Delta x = \tilde{d} - d + A^T\tilde{W}(b - \tilde{b}) + (\mu - \tilde{\mu})A^T\tilde{S}^{-1}e, \quad (4.6a)$$

$$\tilde{W}^{-1}\Delta y = b - \tilde{b} - A\Delta x + (\mu - \tilde{\mu})\tilde{Y}^{-1}e, \quad (4.6b)$$

$$\Delta s = A\Delta x + \tilde{b} - b. \quad (4.6c)$$

From this latter system, we obtain explicit expressions for  $\Delta x$ ,  $\Delta y$  and  $\Delta s$  as

$$\Delta x = (A^T\tilde{W}A)^{-1}(\tilde{d} - d) + (A^T\tilde{W}A)^{-1}A^T\tilde{W}(b - \tilde{b} + (\mu - \tilde{\mu})\tilde{Y}^{-1}e), \quad (4.7a)$$

$$\begin{aligned} \Delta y &= \tilde{W}A(A^T\tilde{W}A)^{-1}(d - \tilde{d}) \\ &\quad + \tilde{W}(I - A(A^T\tilde{W}A)^{-1}A^T\tilde{W})(b - \tilde{b} + (\mu - \tilde{\mu})\tilde{Y}^{-1}e) \end{aligned} \quad (4.7b)$$

$$= \tilde{W}A(A^T\tilde{W}A)^{-1}(d - \tilde{d}) + Z(Z^T\tilde{W}^{-1}Z)^{-1}Z^T(b - \tilde{b} + (\mu - \tilde{\mu})\tilde{Y}^{-1}e), \quad (4.7c)$$

$$\Delta s = A\Delta x + \tilde{b} - b, \quad (4.7d)$$

where  $Z$  is a matrix whose columns form a basis for  $\text{null}(A^T)$ . Note that (4.7c) is obtained from (4.7b) from the relations

$$\begin{aligned} \tilde{W}(I - A(A^T\tilde{W}A)^{-1}A^T\tilde{W}) &= \tilde{W}^{1/2}(I - \tilde{W}^{1/2}A(A^T\tilde{W}A)^{-1}A^T\tilde{W}^{1/2})\tilde{W}^{1/2} \\ &= \tilde{W}^{1/2}(\tilde{W}^{-1/2}Z(Z^T\tilde{W}^{-1}Z)^{-1}Z^T\tilde{W}^{-1/2})\tilde{W}^{1/2} \\ &= Z(Z^T\tilde{W}^{-1}Z)^{-1}Z^T. \end{aligned} \quad (4.8)$$

A special property of linear programming is that the elements of  $\tilde{w}$  can be split into “large” and “small” elements, denoted by  $A$  and  $I$  in consistency with the analysis of Section 3, as stated in the following lemma.

**Lemma 4.1.** *Assume that  $\{(x, s) : Ax - s = \tilde{b}, s > 0\} \neq \emptyset$  and  $\{y : A^T y = \tilde{c}, y > 0\} \neq \emptyset$ . Let  $\tilde{x}$ ,  $\tilde{y}$  and  $\tilde{s}$  satisfy (4.3). Then, we may partition  $\{1, \dots, m\} = A \cup I$ , with  $A \cap I = \emptyset$ , such that*

$$\begin{aligned} \tilde{s}_i &= \Theta(1), \quad \tilde{y}_i = \Theta(\tilde{\mu}), \quad \tilde{w}_i = \Theta(\tilde{\mu}), \quad i \in I, \\ \tilde{s}_i &= \Theta(\tilde{\mu}), \quad \tilde{y}_i = \Theta(1), \quad \tilde{w}_i = \Theta(1/\tilde{\mu}), \quad i \in A, \end{aligned}$$

where  $\tilde{w}_i = \tilde{y}_i/\tilde{s}_i$ ,  $i = 1, \dots, m$ .

**Proof.** See, e.g., Wright [16, Lemma 5.13]. ■

We may get a further description of  $\Delta x$ ,  $\Delta s$  and  $\Delta y$  by utilizing the following result, which essentially states that we may obtain  $(A^T\tilde{W}A)^{-1}A^T\tilde{W}$  as a convex combination of solutions obtained from nonsingular  $n \times n$  submatrices of  $A$ .

**Theorem 4.1. (Dikin [2])** *Let  $A$  be an  $m \times n$  matrix of full column rank, let  $g$  be a vector of dimension  $m$ , and let  $D$  be a positive definite diagonal  $m \times m$  matrix. Then,*

$$(A^TDA)^{-1}A^TDg = \sum_{J \in \mathcal{J}(A)} \left( \frac{\det(D_J) \det(A_J)^2}{\sum_{K \in \mathcal{J}(A)} \det(D_K) \det(A_K)^2} \right) A_J^{-1}g_J,$$

where  $\mathcal{J}(A)$  is the collection of sets of row indices associated with nonsingular  $n \times n$  submatrices of  $A$ .

**Proof.** See, e.g., Ben-Tal and Teboulle [1, Corollary 2.1]. ■

Note that an implication of Theorem 4.1 is that  $\|(A^T\tilde{W}A)^{-1}A^T\tilde{W}\|$  is bounded when  $\tilde{W}$  varies over the set of positive definite and diagonal matrices. For further discussions on this issue, see, e.g., Forsgren [3] and Forsgren and Sporre [7].

In order to give the results on the search directions, we first need a result on the behavior of  $(A^T\tilde{W}A)^{-1}(\tilde{d} - d)$ . As in the previous section, we will let subscript “ $A$ ” denote quantities associated with  $A$ , and similarly for  $I$ .

**Lemma 4.2.** *It holds that  $\|(A^T\tilde{W}A)^{-1}(\tilde{d} - d)\| = O(\tilde{\mu})$  if  $\tilde{d} - d \in \text{range}(A_A^T)$  and  $\|(A^T\tilde{W}A)^{-1}(\tilde{d} - d)\| = \Omega(1/\tilde{\mu})$  otherwise.*

**Proof.** First assume that  $\tilde{d} - d \in \text{range}(A_A^T)$ . Then,  $\tilde{d} - d = A_A^T u$  for some  $u$ . We then get

$$(A^T\tilde{W}A)^{-1}(\tilde{d} - d) = (A^T\tilde{W}A)^{-1}A_A^T u = (A^T\tilde{W}A)^{-1}A^T\tilde{W}\tilde{W}^{-1}I_A u, \quad (4.9)$$

where  $I_A$  is the diagonal matrix with ones in diagonals corresponding to  $A$  and zeros in diagonals corresponding to  $I$ . Taking norms in (4.9), taking into account that  $\|\tilde{W}^{-1}I_A\| = \|\tilde{W}_A^{-1}\|$ , gives

$$\|(A^T\tilde{W}A)^{-1}(\tilde{d} - d)\| \leq \|(A^T\tilde{W}A)^{-1}A^T\tilde{W}\| \|\tilde{W}_A^{-1}\| \|u\|. \quad (4.10)$$

Theorem 4.1 gives  $\|(A^T\tilde{W}A)^{-1}A^T\tilde{W}\| = O(1)$ , Lemma 4.1 shows that  $\|\tilde{W}_A^{-1}\| = O(\tilde{\mu})$  and  $u$  is a fixed vector. Consequently, (4.10) gives  $\|(A^T\tilde{W}A)^{-1}(\tilde{d} - d)\| = O(\tilde{\mu})$ , as required.

Now assume that  $\tilde{d} - d \notin \text{range}(A_A^T)$ . Then,  $Z_A^T(\tilde{d} - d) \neq 0$ , where  $Z_A$  is a matrix whose columns form a basis for the null space of  $A_A$ . Let  $\tilde{v} = (A^T\tilde{W}A)^{-1}(\tilde{d} - d)$ , i.e.,  $\tilde{v}$  solves

$$(A_A^T\tilde{W}_A A_A + A_I^T\tilde{W}_I A_I)\tilde{v} = \tilde{d} - d. \quad (4.11)$$

Premultiplication of (4.11) by  $Z_A^T$  gives

$$Z_A^T A_I^T \widetilde{W}_I A_I \tilde{v} = Z_A^T (\tilde{d} - d),$$

and consequently

$$\|\tilde{v}\| \geq \frac{\|Z_A^T (\tilde{d} - d)\|}{\|Z_A^T A_I^T \widetilde{W}_I A_I\|}. \quad (4.12)$$

Lemma 4.1 shows that  $\|Z_A^T A_I^T \widetilde{W}_I A_I\| = O(\tilde{\mu})$ . Since  $Z_A^T (\tilde{d} - d) \neq 0$ , it follows from (4.12) that  $\|\tilde{v}\| = \Omega(1/\tilde{\mu})$ , as required. ■

Analogously, a result on the behavior of  $\widetilde{W}(I - A(A^T \widetilde{W} A)^{-1} A^T \widetilde{W})(b - \tilde{b})$  is required.

**Lemma 4.3.** *It holds that  $\|\widetilde{W}(I - A(A^T \widetilde{W} A)^{-1} A^T \widetilde{W})(b - \tilde{b})\| = O(\tilde{\mu})$  if  $\tilde{b}_A - b_A \in \text{range}(A_A)$  and  $\|\widetilde{W}(I - A(A^T \widetilde{W} A)^{-1} A^T \widetilde{W})(b - \tilde{b})\| = \Omega(1/\tilde{\mu})$  otherwise.*

**Proof.** First assume that  $b_A - \tilde{b}_A \in \text{range}(A_A)$ . Then,  $b - \tilde{b} = Au + r$  for some  $u$  and  $r$  with  $r_A = 0$ . We then get

$$\begin{aligned} \widetilde{W}(I - A(A^T \widetilde{W} A)^{-1} A^T \widetilde{W})(b - \tilde{b}) &= (I - \widetilde{W} A(A^T \widetilde{W} A)^{-1} A^T) \widetilde{W}(Au + r) \\ &= (I - \widetilde{W} A(A^T \widetilde{W} A)^{-1} A^T) \widetilde{W}r. \end{aligned} \quad (4.13)$$

Taking norms in (4.13) gives

$$\|\widetilde{W}(I - A(A^T \widetilde{W} A)^{-1} A^T \widetilde{W})(b - \tilde{b})\| \leq \|I - \widetilde{W} A(A^T \widetilde{W} A)^{-1} A^T\| \|\widetilde{W}r\|. \quad (4.14)$$

Theorem 4.1 gives  $\|I - \widetilde{W} A(A^T \widetilde{W} A)^{-1} A^T\| = O(1)$ . Note that since  $r_A = 0$ , we obtain  $\|\widetilde{W}r\| = \|\widetilde{W}_I r_I\|$ . Lemma 4.1 shows that  $\|\widetilde{W}_I\| = O(\tilde{\mu})$  and  $r$  is a fixed vector. Consequently, (4.14) gives  $\|\widetilde{W}(I - A(A^T \widetilde{W} A)^{-1} A^T \widetilde{W})(b - \tilde{b})\| = O(\tilde{\mu})$ , as required.

Now assume that  $b_A - \tilde{b}_A \notin \text{range}(A_A)$ . Then,  $\text{null}(A_A^T) \neq \emptyset$ , and there is an orthonormal matrix  $Z_A$  whose columns form a basis for  $\text{null}(A_A^T)$ . Since  $b_A - \tilde{b}_A \notin \text{range}(A_A)$ , it holds that  $Z_A^T (b_A - \tilde{b}_A) \neq 0$ . Moreover, there is an orthonormal matrix  $Z$  whose columns form a basis for  $\text{null}(A^T)$  of the form

$$Z = \begin{pmatrix} Z_A & Z_1 \\ 0 & Z_2 \end{pmatrix},$$

where  $Z_1$  and  $Z_2$  are suitably dimensioned, possibly empty. It follows from (4.8) that

$$\widetilde{W}(I - A(A^T \widetilde{W} A)^{-1} A^T \widetilde{W})(b - \tilde{b}) = Z(Z^T \widetilde{W}^{-1} Z)^{-1} Z^T (b - \tilde{b}). \quad (4.15)$$

Consequently,

$$\|\widetilde{W}(I - A(A^T \widetilde{W} A)^{-1} A^T \widetilde{W})(b - \tilde{b})\| = \|(Z^T \widetilde{W}^{-1} Z)^{-1} Z^T (b - \tilde{b})\|.$$

Let  $y = (Z^T \widetilde{W}^{-1} Z)^{-1} Z^T (b - \tilde{b})$ . Then  $Z^T \widetilde{W}^{-1} Z y = Z^T (b - \tilde{b})$ , or equivalently

$$\begin{pmatrix} Z_A^T & 0 \\ Z_1^T & Z_2^T \end{pmatrix} \begin{pmatrix} \widetilde{W}_A^{-1} & 0 \\ 0 & \widetilde{W}_I^{-1} \end{pmatrix} \begin{pmatrix} Z_A & Z_1 \\ 0 & Z_2 \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = \begin{pmatrix} Z_A^T & 0 \\ Z_1^T & Z_2^T \end{pmatrix} \begin{pmatrix} b_A - \tilde{b}_A \\ b_I - \tilde{b}_I \end{pmatrix}.$$

The first block of equations implies that

$$\begin{pmatrix} Z_A^T \widetilde{W}_A^{-1} & 0 \end{pmatrix} \begin{pmatrix} Z_A & Z_1 \\ 0 & Z_2 \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = Z_A^T (b_A - \tilde{b}_A).$$

Consequently,

$$\| \begin{pmatrix} Z_A^T \widetilde{W}_A^{-1} & 0 \end{pmatrix} \| \| Z \| \| y \| \geq \| Z_A^T (b_A - \tilde{b}_A) \|.$$

By assumption,  $\| Z_A^T (b_A - \tilde{b}_A) \| \neq 0$ . Lemma 4.1 shows that  $\| \widetilde{W}_A^{-1} \| = O(\tilde{\mu})$ , and  $Z$  is a fixed matrix. Consequently, (4.15) gives  $\| \widetilde{W} (I - A(A^T \widetilde{W} A)^{-1} A^T \widetilde{W}) (b - \tilde{b}) \| = \Omega(1/\tilde{\mu})$ , as required. ■

The following theorem now gives a characterization of  $\Delta x$ . Note that  $\Delta x$  is unbounded if  $\tilde{d} - d \notin \text{range}(A_A^T)$ .

**Theorem 4.2.** *If  $\tilde{d} - d \in \text{range}(A_A^T)$ , then*

$$\Delta x = (A^T \widetilde{W} A)^{-1} A^T \widetilde{W} (b - \tilde{b} + (\mu - \tilde{\mu}) \tilde{Y}^{-1} e) + O(\tilde{\mu}).$$

*Otherwise,  $\Delta x = \Omega(1/\tilde{\mu})$ .*

**Proof.** This is a consequence of Lemma 4.2 in conjunction with (4.7a). ■

Analogously, the following theorem now gives a characterization of  $\Delta y$ . Note that that  $\Delta y$  is unbounded if  $b_A - \tilde{b}_A \notin \text{range}(A_A)$ .

**Theorem 4.3.** *If  $\tilde{b}_A - b_A \in \text{range}(A_A)$ , then*

$$\begin{aligned} \Delta y &= \widetilde{W} A (A^T \widetilde{W} A)^{-1} (d - \tilde{d}) \\ &\quad + (\mu - \tilde{\mu}) \widetilde{W} (I - A(A^T \widetilde{W} A)^{-1} A^T \widetilde{W}) \tilde{Y}^{-1} e + O(\tilde{\mu}) \\ &= \widetilde{W} A (A^T \widetilde{W} A)^{-1} (d - \tilde{d}) \\ &\quad + (\mu - \tilde{\mu}) Z (Z^T \widetilde{W}^{-1} Z)^{-1} Z^T \tilde{Y}^{-1} e + O(\tilde{\mu}). \end{aligned}$$

*Otherwise,  $\Delta y = \Omega(1/\tilde{\mu})$ .*

**Proof.** This is a consequence of Lemma 4.3 in conjunction with (4.7c). ■

Finally, for the case of primal and dual nondegeneracy, it follows that both primal and dual steps are bounded.

**Corollary 4.1.** *If  $A_A$  is square and nonsingular, then*

$$\begin{aligned}\Delta x &= A_A^{-1}(b_A - \tilde{b}_A + (\mu - \tilde{\mu})\tilde{Y}_A^{-1}e) + O(\tilde{\mu}), \\ \Delta s_A &= (\mu - \tilde{\mu})\tilde{Y}_I^{-1}e + O(\tilde{\mu}), \\ \Delta s_I &= A_I A_A^{-1}(b_A - \tilde{b}_A + (\mu - \tilde{\mu})\tilde{Y}_A^{-1}e) - b_I + \tilde{b}_I + O(\tilde{\mu}), \\ \Delta y_A &= A_A^{-T}(d - \tilde{d} - (\mu - \tilde{\mu})A_I^T \tilde{S}_I^{-1}e) + O(\tilde{\mu}), \\ \Delta y_I &= (\mu - \tilde{\mu})\tilde{S}_I^{-1}e + O(\tilde{\mu}).\end{aligned}$$

**Proof.** If  $A_A$  is square and nonsingular, Theorem 4.1 in conjunction with Lemma 4.1 gives

$$(A^T \tilde{W} A)^{-1} A_A^T \tilde{W}_A = A_A^{-1} + O(\tilde{\mu}) \quad \text{and} \quad \|(A^T \tilde{W} A)^{-1}\| = O(\tilde{\mu}). \quad (4.16)$$

Consequently,  $\|(A^T \tilde{W} A)^{-1} A_I^T \tilde{W}_I Y_I^{-1}\| = O(\tilde{\mu})$ . The result for  $\Delta x$  now follows by using (4.16) in Theorem 4.2. The result for  $\Delta s$  follows from  $\Delta x$ .

Analogously, we may let

$$Z = \begin{pmatrix} Z_A \\ Z_I \end{pmatrix} = \begin{pmatrix} -A_A^{-T} A_I \\ I \end{pmatrix},$$

for which

$$(Z^T \tilde{W}^{-1} Z)^{-1} Z_I^T \tilde{W}_I^{-1} = I + O(\tilde{\mu}) \quad \text{and} \quad \|(Z^T \tilde{W}^{-1} Z)^{-1}\| = O(\tilde{\mu}). \quad (4.17)$$

Consequently,  $\|(Z^T \tilde{W}^{-1} Z)^{-1} Z_A^T \tilde{Y}_A^{-1}\| = O(\tilde{\mu})$ . The result for  $\Delta y$  now follows by using (4.16) and (4.17) in Theorem 4.3. ■

Corollary 4.1 gives the result from the nonlinear programming case, specialized to linear programming, since the limiting Newton equations are nonsingular in this situation.

#### 4.1. Example linear programming problem

Consider the example linear programming problem where  $A$ ,  $\tilde{b}$  and  $\tilde{d}$  are given by

$$A = \begin{pmatrix} 1 & 0 \\ -1 & 0 \\ 0 & 1 \end{pmatrix}, \quad \tilde{b} = \begin{pmatrix} -1 \\ -1 \\ 0 \end{pmatrix}, \quad \text{and} \quad \tilde{d} = \begin{pmatrix} 0 \\ 1 \end{pmatrix}.$$

Then,  $\tilde{x} = (0 \ \tilde{\mu})^T$ ,  $\tilde{y} = (\tilde{\mu} \ \tilde{\mu} \ 1)^T$  and  $\tilde{s} = (1 \ 1)^T$ . Hence,  $\tilde{w} = \tilde{S}^{-1} \tilde{Y} e = (\tilde{\mu} \ \tilde{\mu} \ 1/\tilde{\mu})^T$ . Accordingly,

$$\tilde{W} A = \begin{pmatrix} \tilde{\mu} & 0 \\ -\tilde{\mu} & 0 \\ 0 & \frac{1}{\tilde{\mu}} \end{pmatrix}, \quad A^T \tilde{W} A = \begin{pmatrix} 2\tilde{\mu} & 0 \\ 0 & \frac{1}{\tilde{\mu}} \end{pmatrix}.$$

$$Z = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}, \quad \widetilde{W}^{-1}Z = \begin{pmatrix} \frac{1}{\tilde{\mu}} \\ \frac{1}{\tilde{\mu}} \\ 0 \end{pmatrix}, \quad (Z^T\widetilde{W}^{-1}Z)^{-1} = \frac{\tilde{\mu}}{2}.$$

We may use Theorem 4.1 to express

$$(A^T\widetilde{W}A)^{-1}A^T\widetilde{W} = \frac{1}{2} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} + \frac{1}{2} \begin{pmatrix} 0 & -1 & 0 \\ 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} \frac{1}{2} & -\frac{1}{2} & 0 \\ 0 & 0 & 1 \end{pmatrix}. \quad (4.18)$$

In this example  $A_A = (0 \ 1)$ . Since  $A_A$  has full row rank, a combination of Theorem 4.3 and (4.18) gives

$$\Delta y = \begin{pmatrix} \frac{1}{2} & 0 \\ -\frac{1}{2} & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} d_1 \\ d_2 - 1 \end{pmatrix} + (\mu - \tilde{\mu}) \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} + O(\tilde{\mu}).$$

However, since  $A_A$  does not have full column rank, Theorem 4.2 shows that  $\Delta x$  is  $\Omega(1/\tilde{\mu})$  unless  $d_1 = \tilde{d}_1$ . For  $d_1 = \tilde{d}_1$ , a combination of Theorem 4.2 and (4.18) gives

$$\Delta x = \begin{pmatrix} \frac{1}{2} & -\frac{1}{2} & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} b_1 + 1 \\ b_2 + 1 \\ b_3 \end{pmatrix} + (\mu - \tilde{\mu}) \begin{pmatrix} 0 \\ 1 \end{pmatrix} + O(\tilde{\mu}).$$

For large-scale problems, the explicit representations from this small problem are naturally not available, but we have included the example to give a feeling of what Theorems 4.2 and 4.3 say.

## 5. Summary

We have characterized search directions that would arise in warm starts for interior methods, first for the general nonlinear programming case, and then more specialized results for the linear programming case. The difficulties in warm starts for interior methods are emphasized by these characterizations, since the directions are similar to directions what would arise in a sequential-quadratic-programming method applied to the active constraints only.

The results are related to “false convergence” of interior methods on nonconvex problems in that the iterates are close to the boundary of the inequality constraints, but not well centered with respect to the trajectory. See e.g., Wächter and Biegler [14] or Forsgren and Sporre [8], for further discussions on false convergence.

## References

- [1] A. Ben-Tal and M. Teboulle. A geometric property of the least squares solution of linear equations. *Linear Algebra Appl.*, 139:165–170, 1990.
- [2] I. I. Dikin. On the speed of an iterative process. *Upravlyaemye Sistemy*, 12:54–60, 1974.
- [3] A. Forsgren. On linear least-squares problems with diagonally dominant weight matrices. *SIAM J. Matrix Anal. Appl.*, 17:763–788, 1996.

- 
- [4] A. Forsgren. Inertia-controlling factorizations for optimization algorithms. *Appl. Num. Math.*, 43:91–107, 2002.
  - [5] A. Forsgren, P. E. Gill, and J. D. Griffin. Iterative solution of augmented systems arising in interior methods. Report TRITA-MAT-2005-OS3, Department of Mathematics, Royal Institute of Technology, Stockholm, Sweden, 2005.
  - [6] A. Forsgren, P. E. Gill, and M. H. Wright. Interior methods for nonlinear optimization. *SIAM Rev.*, 44(4):525–597 (electronic) (2003), 2002.
  - [7] A. Forsgren and G. Sporre. On weighted linear least-squares problems related to interior methods for convex quadratic programming. *SIAM J. Matrix Anal. Appl.*, 23:42–56, 2001.
  - [8] A. Forsgren and G. Sporre. Relations between divergence of multipliers and convergence to infeasible points in primal-dual interior methods for nonconvex nonlinear programming. Report TRITA-MAT-2002-OS7, Department of Mathematics, Royal Institute of Technology, Stockholm, Sweden, 2002.
  - [9] J. Gondzio and A. Grothey. Reoptimization with the primal-dual interior point method. *SIAM J. Optim.*, 13(3):842–864 (electronic) (2003), 2002.
  - [10] M. Gonzalez-Lima, H. Wei, and H. Wolkowicz. A stable iterative method for linear programming. Report CORR 2004-26, Department of Combinatorics and Optimization, University of Waterloo, 2004.
  - [11] B. Jansen, J. J. de Jong, C. Roos, and T. Terlaky. Sensitivity analysis in linear programming: just be careful! *European J. Oper. Res.*, 101:15–28, 1997.
  - [12] W.-J. Kim, C.-K. Park, and S. Park. An  $\epsilon$ -sensitivity analysis in the primal-dual interior point method. *European J. Oper. Res.*, 116:629–639, 1999.
  - [13] J. Nocedal and S. J. Wright. *Numerical Optimization*. Springer, New York, 1999. ISBN 0-387-98793-2.
  - [14] A. Wächter and L. T. Biegler. Failure of global convergence for a class of interior point methods for nonlinear programming. *Math. Program.*, 88:565–574, 2000.
  - [15] M. H. Wright. The interior-point revolution in optimization: history, recent developments, and lasting consequences. *Bull. Amer. Math. Soc. (N.S.)*, 42(1):39–56 (electronic), 2005.
  - [16] S. J. Wright. *Primal-Dual Interior-Point Methods*. SIAM, Society for Industrial and Applied Mathematics, Philadelphia, 1997. ISBN 0-89871-382-X.
  - [17] E. A. Yildirim. An interior-point perspective on sensitivity analysis in semidefinite programming. *Math. Oper. Res.*, 28(4):649–676, 2003.
  - [18] E. A. Yildirim and M. J. Todd. Sensitivity analysis in linear programming and semidefinite programming using interior-point methods. *Math. Program.*, 90(2, Ser. A):229–261, 2001.
  - [19] E. A. Yildirim and M. J. Todd. An interior-point approach to sensitivity analysis in degenerate linear programs. *SIAM J. Optim.*, 12(3):692–714 (electronic), 2002.
  - [20] E. A. Yildirim and S. J. Wright. Warm-start strategies in interior-point methods for linear programming. *SIAM J. Optim.*, 12(3):782–810 (electronic), 2002.