ISRN KTH/OPT SYST/FR 02/07 SE
Coden: TRITA/MAT-02-OS7
ISSN 1401-2294

Relations between divergence of multipliers and convergence to infeasible points in primal-dual interior methods for nonconvex nonlinear programming

by

Göran Sporre and Anders Forsgren

Optimization and Systems Theory

DEPARTMENT OF MATHEMATICS
ROYAL INSTITUTE OF TECHNOLOGY
SE-100 44 STOCKHOLM, SWEDEN

ISSN 1401-2294

RELATIONS BETWEEN DIVERGENCE OF MULTIPLIERS AND CONVERGENCE TO INFEASIBLE POINTS IN PRIMAL-DUAL INTERIOR METHODS FOR NONCONVEX NONLINEAR PROGRAMMING

Göran SPORRE* and Anders FORSGREN†

Technical Report TRITA-MAT-2002-OS7
Department of Mathematics
Royal Institute of Technology
April 2002

Abstract

Recently, infeasibility issues in interior methods for nonconvex nonlinear programming have been studied. In particular, it has been shown how many line-search interior methods may converge to an infeasible point which is on the boundary of the feasible region with respect to the inequality constraints. The convergence is such that the search direction does not tend to zero, but the step length does. Such "false" convergence may occur even for a fixed value of the barrier parameter. In this paper, two commonly used reformulations for handling infeasibility are studied. It is shown that when the above-mentioned convergence difficulty occurs, the multiplier search directions diverge along a direction that can be characterized in terms of the gradients of the equality constraints and the asymptotically active inequality constraints. Further, for a penalty-barrier interior method, similar convergence difficulties along with diverging multiplier search directions can occur when the penalty and barrier parameters are driven to zero. The divergence is motivated through an analysis of the penalty-barrier trajectory. Further, the direction of divergence is characterized.

Key words. Nonlinear programming, interior method, primal-dual interior method, infeasibility.

AMS subject classifications. 65K05, 90C30, 90C51, 90C26.

1. Introduction

We study properties of so-called interior methods for solving problems of the form

$$(P) \qquad \begin{array}{ll} \underset{x \in I\!R^n}{\text{minimize}} & f(x) \\ \text{subject to} & c_i(x) \geq 0, \quad i = 1, \ldots, m, \end{array}$$

*Optimization and Systems Theory, Department of Mathematics, Royal Institute of Technology, SE-100 44 Stockholm, Sweden (goran.sporre@math.kth.se). Research supported by the Swedish Research Council (VR).

†Optimization and Systems Theory, Department of Mathematics, Royal Institute of Technology, SE-100 44 Stockholm, Sweden (anders.forsgren@math.kth.se). Research supported by the Swedish Research Council (VR).

or certain reformulations of such problems. The methods are interior in the sense that all iterates strictly satisfy the inequality constraints of the problem to which the method is applied. Throughout, f and c, with components c_i for $i = 1, \ldots, m$, are assumed to be twice continuously differentiable functions on $\mathbb{R}^n$. The analysis in this paper can without theoretical difficulties be generalized to cover equality constraints as well. However, to avoid a too cumbersome notation we assume that there only are inequality constraints in (P).

Since all iterates must satisfy the inequality constraints strictly, the issue of finding an initial iterate $x^{(0)}$ such that $c(x^{(0)}) > 0$ must be addressed. A common approach is to circumvent this potential difficulty by reformulating (P) so that there are easy-to-find points which strictly satisfy the inequality constraints of the reformulated problem. A reformulation typically includes introduction of new variables and additional constraints. The extra variables can normally, through careful use of linear algebra, be handled without any significant increase in computational effort.

The most common reformulation is based on subtracting a nonnegative slack variable, s_i, from $c_i(x)$ for $i = 1, \ldots, m$. This gives, what we call, the slack variable reformulation

$$(P_{sl}) \qquad \begin{array}{lll} \underset{x \in \mathbb{R}^n}{\text{minimize}} & f(x) & \\ \text{subject to} & c_i(x) - s_i = 0, & i = 1, \ldots, m, \\ & s_i \geq 0, & i = 1, \ldots, m. \end{array}$$

For (P_{sl}), each $(x^{(0)}, s^{(0)})$ such that $s^{(0)} > 0$ is acceptable as initial iterate. An alternative is to shift the constraints $c_i(x) \geq 0$, $i = 1, \ldots, m$, by s_i and require all shifts to be zero. This gives the shift reformulation

$$(P_{sh}) \qquad \begin{array}{lll} \underset{x \in \mathbb{R}^n}{\text{minimize}} & f(x) & \\ \text{subject to} & c_i(x) + s_i \geq 0, & i = 1, \ldots, m, \\ & s_i = 0, & i = 1, \ldots, m. \end{array}$$

For discussions on the use of shift variables in interior methods, see, e.g., Gill *et al.* [GMSW88], Gay, Overton and Wright [GOW98, Section 7], Jarre and Saunders [JS95], Todd [Tod94], Polyak [Pol92], Freund [Fre91], and Conn, Gould and Toint [CGT94]. See also Powell [Pow69] for a discussion on the use of shift variables in a penalty method. An acceptable initial iterate can be obtained by choosing $x^{(0)}$ arbitrary and letting $s_i^{(0)} = \max\{-c_i(x^{(0)}) + \varepsilon, 0\}$, $i = 1, \ldots, m$, for some positive ε. For ease of notation, we introduce one shift variable for each constraint, although it is only necessary to do so for constraints with $c_i(x^{(0)}) \leq 0$.

Wächter and Biegler [WB00] have recently given an example of the form (P_{sl}), where they show that a wide class of interior methods would fail to find a feasible point. In this paper we extend the discussion of that paper by also considering Lagrange multiplier estimates arising in primal-dual interior methods. In particular, we derive connections between asymptotic lack of feasibility and diverging multiplier search directions. The asymptotic lack of feasibility is assumed to be characterized by the step length being forced to zero due to the requirement of strict feasibility with respect to inequality constraints. We do this in a general framework that includes primal-dual barrier methods, and subsequently specialize to the two forms (P_{sl}) and (P_{sh}). In Section 2 we describe the class of methods that we consider, and also discuss related work. In Section 3 we give results on divergence of the multiplier search directions. We also characterize the direction of divergence in terms of constraint gradients. In Section 4, the results are specialized to barrier methods for the two reformulations (P_{sl}) and (P_{sh}). Finally, related results for penalty-barrier methods are given in Section 5.

1.1. Notation

The gradient of $f(x)$ is denoted by $g(x)$ and the Jacobian of $c(x)$ by $J(x)$. For an index set $\mathcal{J}$, $c_{\mathcal{J}}(x)$ denotes a vector with components $c_i(x)$ for $i \in \mathcal{J}$. The associated Jacobian is denoted by $J_{\mathcal{J}}(x)$. For brevity, we sometimes suppress the dependency on the variables when there is no risk of ambiguity, e.g., $J_{\mathcal{J}}(x)$ is written $J_{\mathcal{J}}$. A capital version of a letter denoting a vector means a diagonal matrix with diagonal components equal to the components of the vector. A vector of all ones is denoted by e. Its dimension is given by the context. The vector norm used can be any norm satisfying $\|(x^T \ 0)^T\| = \|x\|$ for any vector x. By $\mu \to 0^+$ we mean a sequence of positive values of μ converging to zero.

2. Background

2.1. Class of interior methods considered

The methods studied here are line-search primal-dual interior methods for nonlinear nonconvex problems. For examples of such methods, see, e.g., Forsgren and Gill [FG98], Gay, Overton and Wright [GOW98], Shanno and Vanderbei [SV99], El-Bakry *et al.* [EBTTZ96], and Conn, Gould and Toint [CGT96]. Other interior methods use a trust-region technique for safeguarding; see, e.g., Byrd, Hribar and Nocedal [BHN99], Byrd, Gilbert and Nocedal [BGN00], Conn, Gould and Toint [CGT00, Chapter 13], and references therein. Recently, interior methods with filters have been studied as well, see Ulbrich, Ulbrich and Vicente [UUV00] and Wächter and Biegler [WB01]. However, trust-region and filter methods do generally not fit into the framework of this paper and are therefore not discussed further.

To avoid repeating descriptions of the method as it is applied to each of the two above-mentioned reformulations, it is described for the general problem

$$\begin{array}{ll} \underset{w \in \mathbb{R}^{\tilde{n}}}{\text{minimize}} & \widetilde{f}(w) \\ \text{subject to} & \widetilde{c}_i(w) \geq 0, \quad i \in \mathcal{I}, \\ & \widetilde{c}_i(w) = 0, \quad i \in \mathcal{E}. \end{array} \tag{2.1}$$

The variables are denoted by w instead of x to distinguish between these variables and the x variables of the reformulated problems. A tilde on the functions is used to avoid confusion with the functions of problem (P).

In a classical penalty-barrier method, see Fiacco and McCormick [FM68], the problem (2.1) is solved by approximately minimizing the function

$$\widetilde{f}(w) - \mu_{\mathcal{I}} \sum_{i\in\mathcal{I}} \ln(\widetilde{c}_i(w)) + \frac{1}{2\mu_{\mathcal{E}}} \widetilde{c}_{\mathcal{E}}(w)^T \widetilde{c}_{\mathcal{E}}(w), \tag{2.2}$$

for sequences of decreasing and positive values of the penalty parameter, $\mu_{\mathcal{E}}$, and the barrier parameter, $\mu_{\mathcal{I}}$. For convenience in notation, we let $\mu = \mu_{\mathcal{E}} = \mu_{\mathcal{I}} > 0$. The penalty-barrier trajectory, $x(\mu)$, is then defined as the set of these minimizers of (2.2) as a function of μ. Under suitable assumptions, the trajectory converges to a solution of (2.1) as μ is driven to zero; see Fiacco and McCormick [FM68, Thm. 17]. In (2.2), we allow different values of the barrier and penalty parameters to facilitate a comparison between penalty-barrier methods and barrier methods. In a barrier method, the first two terms of (2.2) are minimized subject to $c_{\mathcal{E}}(w) = 0$. A barrier trajectory is defined analogously to the penalty-barrier trajectory.

The first-order optimality conditions for (2.2), when $\mu = \mu_{\mathcal{E}} = \mu_{\mathcal{I}}$, imply the existence of a vector $\lambda(\mu)$ such that

$$\widetilde{g}(w(\mu)) - \widetilde{J}(w(\mu))^T \lambda(\mu) = 0, \quad \text{and} \quad \lambda_i(\mu) = \begin{cases} -\widetilde{c}_i(w(\mu))/\mu, & i \in \mathcal{E}; \\ \mu/\widetilde{c}_i(w(\mu)), & i \in \mathcal{I}. \end{cases} \tag{2.3}$$

Then, $\lambda(\mu)$ can be viewed as an estimate of the multipliers to (2.1). A primal-dual interior method tries to approximately follow the trajectory $(w(\mu), \lambda(\mu))$. The search directions are computed by applying Newton's method to the primal-dual penalty-barrier equations

$$\widetilde{g}(w) - \widetilde{J}(w)^T \lambda = 0, \tag{2.4a}$$

$$\widetilde{c}_i(w)\lambda_i - \mu_{\mathcal{I}} = 0, \quad i \in \mathcal{I}, \tag{2.4b}$$

$$\widetilde{c}_i(w) + \mu_{\mathcal{E}}\lambda_i = 0, \quad i \in \mathcal{E}, \tag{2.4c}$$

which are equivalent to (2.3). Here, we have allowed the slightly more general form where μ differs for the inequality and equality constraints. Positivity of $\widetilde{c}_i(w)$ for $i \in \mathcal{I}$ is held implicitly to guarantee that (2.2) is well defined. Considering (2.3), it is natural to also implicitly hold $\lambda_i > 0$ for $i \in \mathcal{I}$.

Setting $\mu_{\mathcal{E}}$ equal to zero in (2.4) gives the primal-dual *barrier* equations. Note that these equations correspond to the KKT-conditions for the barrier problem along with the definition of $\lambda_i(\mu)$ for $i \in \mathcal{I}$ in (2.3). The Newton equations, associated with (2.4), for the search directions Δw, $\Delta\lambda_{\mathcal{I}}$ and $\Delta\lambda_{\mathcal{E}}$, are

$$\begin{pmatrix} \widetilde{H} & -\widetilde{J}_{\mathcal{I}}^T & -\widetilde{J}_{\mathcal{E}}^T \\ \Lambda_{\mathcal{I}}\widetilde{J}_{\mathcal{I}} & \widetilde{C}_{\mathcal{I}} & 0 \\ \widetilde{J}_{\mathcal{E}} & 0 & \mu_{\mathcal{E}} I \end{pmatrix} \begin{pmatrix} \Delta w \\ \Delta\lambda_{\mathcal{I}} \\ \Delta\lambda_{\mathcal{E}} \end{pmatrix} = - \begin{pmatrix} \widetilde{g} - \widetilde{J}_{\mathcal{I}}^T\lambda_{\mathcal{I}} - \widetilde{J}_{\mathcal{E}}^T\lambda_{\mathcal{E}} \\ \widetilde{C}_{\mathcal{I}}\lambda_{\mathcal{I}} - \mu_{\mathcal{I}} e \\ \widetilde{c}_{\mathcal{E}} + \mu_{\mathcal{E}}\lambda_{\mathcal{E}} \end{pmatrix}, \tag{2.5}$$

where $\widetilde{H} = \widetilde{H}(w, \lambda)$ denotes the Hessian of the Lagrangian $\widetilde{f}(w) - \lambda^T\widetilde{c}(w)$ with respect to w. To get a practically sound and well-defined algorithm, many details need to be considered. However, the results of this paper only rely on the algorithm fitting into the framework of Algorithm 2.1.

Algorithm 2.1. *Framework for line-search interior method.*

Choose an initial iterate $(w^{(0)}, \lambda^{(0)})$ such that $\widetilde{c}_{\mathcal{I}}(w^{(0)}) > 0$ and $\lambda_{\mathcal{I}}^{(0)} > 0$;
Choose a positive $\mu_{\mathcal{I}}^{(0)}$ and a nonnegative $\mu_{\mathcal{E}}^{(0)}$;
$k \leftarrow 0$;
repeat
  Compute $\Delta w^{(k)}$, $\Delta\lambda^{(k)}$ from (2.5) with $w = w^{(k)}$ and $\lambda = \lambda^{(k)}$;
  Choose $\alpha^{(k)} \in (0, 1]$ such that $\widetilde{c}_{\mathcal{I}}(w^{(k)}+\alpha^{(k)}\Delta w^{(k)}) > 0$ and $\lambda_{\mathcal{I}}^{(k)}+\alpha^{(k)}\Delta\lambda_{\mathcal{I}}^{(k)} > 0$;
  $w^{(k+1)} \leftarrow w^{(k)} + \alpha^{(k)}\Delta w^{(k)}$;
  $\lambda^{(k+1)} \leftarrow \lambda^{(k)} + \alpha^{(k)}\Delta\lambda^{(k)}$;
  Choose a positive $\mu_{\mathcal{I}}^{(k+1)}$ and a nonnegative $\mu_{\mathcal{E}}^{(k+1)}$;
  $k \leftarrow k + 1$;
until converged;

We assume throughout that the matrix of (2.5) is nonsingular at the iterates. This means that we assume that the matrix $(\widetilde{J}_{\mathcal{E}}\ \ \mu_{\mathcal{E}} I)$ has full row rank. In this situation, nonsingularity may be accomplished by modifying $\widetilde{H}$ if necessary, see, e.g., Forsgren and Gill [FG98]. To simplify the discussion, we assume that the matrix is nonsingular without modifications. In fact, the results of this paper hold even if the (1,1) block differs from the one in (2.5), as long as the sequence of differences is bounded, see Section 3.2 for details. Further, satisfying the last block row of (2.5) is not of critical importance in this paper.

The choice of step length α in Algorithm 2.1 only guarantees strict feasibility with respect to inequality constraints for the next iterate and positivity of the corresponding multipliers. The choice of step length is closely connected with the choice of line-search strategy and merit function. See, e.g., Moré and Sorensen [MS79] and Nocedal and Wright [NW99], for general discussions on these topics, Murray and Wright [MW94] and Melman [Mel96] for discussions on line-search in barrier methods, and Forsgren and Gill [FG98] for a merit function for a primal-dual penalty-barrier method. For nonconvex problems, the set $\{w \in I\!R^{\tilde{n}} : \widetilde{c}_{\mathcal{I}}(w) \geq 0\}$ may be nonconvex. Hence, there may exist $\alpha \in (0, \alpha^{(k)})$ such that $\widetilde{c}_{\mathcal{I}}(w^{(k)} + \alpha\Delta w^{(k)}) < 0$, i.e., the step crosses the infeasible region with respect to the inequality constraints. We refer to this phenomenon as *tunneling*. The occurrence of tunneling may make the line-search more complicated. The region defined by the inequality constraints in the slack variable reformulation (P_{sl}) is convex, and hence tunneling will not occur in this situation. Such convexity does not hold for the shift variable reformulation (P_{sh}), and hence tunneling may occur. However, the important issues related to the line-search are not the focus of the present paper and, hence, are not discussed further.

The strategy for updating $\mu_{\mathcal{I}}$ and $\mu_{\mathcal{E}}$ is intentionally left unspecified to allow for different strategies. In a penalty-barrier method, they are typically left unchanged until an approximate solution of the corresponding penalty-barrier problem has been obtained. After that, they are reduced. For a barrier method, $\mu_{\mathcal{E}}$ is kept fixed at zero, while $\mu_{\mathcal{I}}$ is reduced when an approximate solution to the corresponding barrier problem is found.

2.2. Related work

This research was initiated by some recent results of Wächter and Biegler [WB00]. They study interior methods that possess two properties. Firstly, strict feasibility with respect to the inequality constraints is maintained through step length reduction. Secondly, the search direction satisfies a linearization of the equality constraints. They prove that no method with these properties can be guaranteed to converge to a feasible point for a specific family of problems. This family of problems can be viewed as slack variable reformulations of

$$\begin{array}{ll} \underset{x \in I\!R}{\text{minimize}} & f(x) \\ \text{subject to} & x^2 + a \geq 0, \\ & x - b \geq 0, \end{array} \tag{2.6}$$

where a and b are scalars and b is nonnegative. The objective can be any function since it does not enter the analysis. They prove that if $b \geq 0$ and the initial x and s satisfy $x - b - s_2 < 0$, $x < 0$, and $a + b(x^2 + a - s_1)/(x - b - s_2) \leq \min\{0, -a/2\}$, a feasible point can never be reached by any method that possesses the two properties mentioned above; see Wächter and Biegler [WB00, Thm. 1].

To illustrate, let $a = -1$ and $b = 0.5$. Introduction of slack variables and linearization of the equality constraints give

$$x^2 + 2x\Delta x - s_1 - \Delta s_1 - 1 = 0, \tag{2.7a}$$

$$x + \Delta x - s_2 - \Delta s_2 - 0.5 = 0. \tag{2.7b}$$

If the initial x is chosen to be negative and the initial s_1 satisfies $s_1 \leq x^2 - 1$, the unit step is not feasible. To see this, assume that $\alpha = 1$ is feasible. Equation (2.7a) along with the implicit positivity requirement on $s_1 + \Delta s_1$ implies that $\Delta x < (1 - x^2)/2x$ while equation (2.7b) along with the positivity of $s_2 + \Delta s_2$ gives $\Delta x > -x$. These inequalities are incompatible for negative x. Further, the convexity of $x^2 - 1$ implies that a feasible point can never be reached.

What happens in practice when a barrier method is applied to the slack variable reformulation of (2.6) is that the slack variables rapidly converge to zero and x converges to some negative number, x^*, which depends on the initial iterate. The positivity requirement on the slack variables forces the step length to converge to zero. Furthermore, $s + \Delta s < 0$ typically holds in the limit. Subsequently, this kind of behavior has been discussed in other papers as well; see Benson, Shanno, and Vanderbei [BSV00], Byrd, Marazzi, and Nocedal [BMN01], and Marazzi and Nocedal [MN00].

3. General analysis

Our concern is with primal-dual interior methods, and it is therefore interesting to notice that the multiplier search directions diverge as the “false” convergence in the primal variables occurs. The main focus of this paper is an analysis of this divergence. Convergence to zero of the step length is assumed to be characterized by the existence of $i \in \mathcal{I}$ for which $\widetilde{c}_i(w) + \nabla \widetilde{c}_i(w)^T \Delta w < 0$ holds in the limit. Note that this is exactly the generalization to arbitrary inequality constraints of the empirically observed behavior that $s + \Delta s < 0$ in the limit for the slack variable reformulation of the example (2.6) by Wächter and Biegler [WB00]. In this section we give results relating such convergence to zero of the step length to divergence of the multiplier search directions. We also characterize the direction of divergence in terms of constraint gradients.

The main theorem of this paper describes the behavior of the multiplier search directions as the value of some inequality constraint functions converge to zero in such a way that the step length is forced to converge to zero. In Section 3.1 we give results that are consequences of the part of (2.5) associated with the perturbed complementarity constraints (2.4b). These results are used to prove the main theorem in Section 3.2.

3.1. Consequences of perturbed complementarity

Although the aim of this section is to derive results that can be applied to Algorithm 2.1, the results only rely on part of the properties of that algorithmic framework. The relevant part is the one that concerns the treatment of the inequality constraints and the corresponding multipliers. For simplicity, we use the same notation as previously in this paper.

Assume that there are positive sequences $\{\lambda_i^{(k)}\}$ for $i \in \mathcal{I}$ and a sequence $\{w^{(k)}\}$ such that $\widetilde{c}(w^{(k)}) > 0$ for all k. Further, assume that there are sequences $\{\Delta\lambda_{\mathcal{I}}^{(k)}\}$ and $\{\Delta w^{(k)}\}$ such that the Newton equations corresponding to (2.4b) are satisfied, i.e.,

$$\lambda_i^{(k)} \nabla \widetilde{c}_i(w^{(k)})^T \Delta w^{(k)} + \widetilde{c}_i(w^{(k)}) \Delta\lambda_i^{(k)} = -(\widetilde{c}_i(w^{(k)})\lambda_i^{(k)} - \mu_{\mathcal{I}}^{(k)}), \quad \forall i \in \mathcal{I}, \tag{3.1}$$

holds. Further, assume that, for all $i \in \mathcal{I}$, subsequent multiplier estimates $\lambda_i^{(k)}$ and $\lambda_i^{(k+1)}$ are such that $\lambda_i^{(k+1)} - \lambda_i^{(k)} = \alpha^{(k)} \Delta\lambda_i^{(k)}$ for some positive $\alpha^{(k)}$.

We are interested in inequality constraints which asymptotically become active. However, we restrict the study to constraints that force the step length to converge to zero. This motivates the following definition.

Definition 3.1. *Let I be the set of indices $i \in \mathcal{I}$ such that $\lim_{k\to\infty} \widetilde{c}_i(w^{(k)}) = 0$ and $\limsup_{k\to\infty} \{\widetilde{c}_i(w^{(k)}) + \nabla \widetilde{c}_i(w^{(k)})^T \Delta w^{(k)}\} < 0$.*

Note that the set I depends on $\{w^{(k)}\}$, $\{\Delta w^{(k)}\}$, and $\widetilde{c}$ but this dependence is suppressed in order not to make the notation too cumbersome. The following lemma relates the asymptotic behavior of the multipliers and their search directions to the behavior of the corresponding constraint values.

Lemma 3.1. *Let $\{w^{(k)}\}$, $\{\lambda_{\mathcal{I}}^{(k)}\}$, $\{\Delta w^{(k)}\}$, $\{\Delta\lambda_{\mathcal{I}}^{(k)}\}$, and $\{\mu_{\mathcal{I}}^{(k)}\}$ be sequences such that (3.1) is satisfied for all iterates k and all indices $i \in \mathcal{I}$. Assume that $\{\lambda_{\mathcal{I}}^{(k)}\}$ and*

$\{\mu_{\mathcal{I}}^{(k)}\}$ *are positive sequences, and let* I *be given by Definition 3.1. Then, for* $i \in I$, $\lim_{k\to\infty} \lambda_i^{(k)}/\Delta\lambda_i^{(k)} = 0$ *and* $\lim_{k\to\infty} \Delta\lambda_i^{(k)} = \infty$.

Proof. Since $\lambda_{\mathcal{I}}^{(k)} > 0$ for all k, we can rewrite (3.1) as

$$\widetilde{c}_i(w^{(k)}) + \nabla\widetilde{c}_i(w^{(k)})^T \Delta w^{(k)} + \widetilde{c}_i(w^{(k)})\frac{\Delta\lambda_i^{(k)}}{\lambda_i^{(k)}} = \frac{\mu_{\mathcal{I}}^{(k)}}{\lambda_i^{(k)}}, \quad \forall i \in \mathcal{I}. \tag{3.2}$$

In the limit, the right-hand side of (3.2) is nonnegative and, for $i \in I$, the sum of the first two terms of the left-hand side is negative. This forces the third term to be positive. Since $\widetilde{c}_i(w^{(k)}) \to 0$ for $i \in I$, it follows that $\lim_{k\to\infty} \Delta\lambda_i^{(k)}/\lambda_i^{(k)} = \infty$. Inversion then gives the first result. Equation (3.1) can alternatively be written as

$$\widetilde{c}_i(w^{(k)})\Delta\lambda_i^{(k)} = \mu_{\mathcal{I}}^{(k)} - \lambda_i^{(k)}(\widetilde{c}_i(w^{(k)}) + \nabla\widetilde{c}_i(w^{(k)})^T \Delta w^{(k)}), \quad \forall i \in \mathcal{I}. \tag{3.3}$$

For $i \in I$, the right-hand side of (3.3) is positive for all k large enough. Hence, $\Delta\lambda_i^{(k)}$ is positive for all k large enough and $\{\lambda_i^{(k)}\}$ is bounded away from zero. Since $\limsup_{k\to\infty}\{\widetilde{c}_i(w^{(k)}) + \nabla\widetilde{c}_i(w^{(k)})^T \Delta w^{(k)}\} < 0$, the right-hand side is asymptotically bounded away from zero and the second result follows since $\widetilde{c}_i(w^{(k)}) \to 0$ for $i \in I$. ∎

Lemma 3.1 can be strengthened by assuming that the sequence $\{\widetilde{C}(w^{(k)})\lambda^{(k)}\}$ is bounded, that the sequence $\{w^{(k)}\}$ converges and that every constraint converging to zero does so in a way that enforces the step length to converge to zero. Thus, if $i \notin I$, then $\widetilde{c}_i(w^{(k)})$ is positive and asymptotically bounded away from zero. The following lemma makes this precise.

Lemma 3.2. *Let* $\{w^{(k)}\}$, $\{\lambda_{\mathcal{I}}^{(k)}\}$, $\{\Delta w^{(k)}\}$, $\{\Delta\lambda_{\mathcal{I}}^{(k)}\}$, *and* $\{\mu_{\mathcal{I}}^{(k)}\}$ *be sequences such that (3.1) is satisfied for all iterates* k *and all indices* $i \in \mathcal{I}$. *Assume that* $\{\lambda_{\mathcal{I}}^{(k)}\}$, $\{\widetilde{c}_{\mathcal{I}}(w^{(k)})\}$, *and* $\{\mu_{\mathcal{I}}^{(k)}\}$ *are positive sequences, that* $\{\widetilde{C}_{\mathcal{I}}(w^{(k)})\lambda_{\mathcal{I}}^{(k)}\}$ *is bounded, and that* $\{w^{(k)}\}$ *converges. Further, assume that if* $\lim_{k\to\infty} \widetilde{c}_i(x^{(k)}) = 0$ *then it holds that* $\limsup_{k\to\infty}\{\widetilde{c}_i(w^{(k)}) + \nabla\widetilde{c}_i(w^{(k)})^T \Delta w^{(k)}\} < 0$. *Finally, let* I *be given by Definition 3.1 and assume that* $I \neq \emptyset$. *Then,* $\lambda_{\mathcal{I}}^{(k)}/\|\Delta\lambda_{\mathcal{I}}^{(k)}\| \to 0$ *as* $k \to \infty$.

Proof. First, note that $\lim_{k\to\infty} \widetilde{c}_{\mathcal{I}}(w^{(k)})$ is well defined since $\widetilde{c}_{\mathcal{I}}(w)$ is continuous and $\{w^{(k)}\}$ is assumed to converge. Since $I \neq \emptyset$, it follows from Lemma 3.1 that $\lim_{k\to\infty} \|\Delta\lambda_{\mathcal{I}}^{(k)}\| = \infty$. We prove the result by showing that $\lambda_i^{(k)}/\|\Delta\lambda^{(k)}\| \to 0$ for all $i \in \mathcal{I}$. For i such that $\lim_{k\to\infty} \widetilde{c}_i(w^{(k)}) > 0$, the result follows upon observing that boundedness of the sequence of $\{\widetilde{c}_i(w^{(k)})\lambda_i^{(k)}\}$ implies that $\limsup_{k\to\infty} \lambda_i^{(k)} < \infty$. If $\liminf_{k\to\infty} \widetilde{c}_i(w^{(k)})$ is not positive it must be zero. By assumption it follows that $\limsup_{k\to\infty}\{\widetilde{c}_i(w^{(k)}) + \nabla\widetilde{c}_i(w^{(k)})^T \Delta w^{(k)}\} < 0$. The result then follows from Lemma 3.1. ∎

Let us return to the framework of Algorithm 2.1 and denote the right-hand side of (2.5) by $r(w, \lambda_{\mathcal{E}}, \lambda_{\mathcal{I}})$. A crude way to ensure boundedness of $\{\widetilde{C}_{\mathcal{I}}(w^{(k)})\lambda_{\mathcal{I}}^{(k)}\}$ is to make sure that $\{\|r(w^{(k)}, \lambda_{\mathcal{E}}^{(k)}, \lambda_{\mathcal{I}}^{(k)})\|\}$ is nonincreasing. This can be done by step length reduction since the search direction given by (2.5) is a descent direction with respect to $\|r(w, \lambda_{\mathcal{E}}, \lambda_{\mathcal{I}})\|^2$ as long as $r \neq 0$ and the matrix of (2.5) is nonsingular.

3.2. Direction of divergence for multiplier search directions

Above, we have shown that if $c_i(w^{(k)}) \to 0$ for some $i \in \mathcal{I}$, the corresponding multiplier search direction diverges. The following theorem shows that, under suitable assumptions, the divergence must take place in a direction that lies in the null space of the transpose of the part of the Jacobian that corresponds to the equality constraints and the inequality constraints which asymptotically become active.

Theorem 3.1. *Let $\{w^{(k)}\}$, $\{\lambda^{(k)}\}$, $\{\Delta w^{(k)}\}$, $\{\Delta\lambda^{(k)}\}$, $\{\mu_{\mathcal{E}}^{(k)}\}$, and $\{\mu_{\mathcal{I}}^{(k)}\}$ be sequences fitting the framework of Algorithm 2.1. Assume that the sequences $\{\mu_{\mathcal{E}}^{(k)}\}$, $\{\mu_{\mathcal{I}}^{(k)}\}$, $\{\Delta w^{(k)}\}$, $\{\widetilde{g}(w^{(k)}) - \widetilde{J}(w^{(k)})^T\lambda^{(k)}\}$, and $\{\widetilde{C}_{\mathcal{I}}(w^{(k)})\lambda_{\mathcal{I}}^{(k)}\}$ are bounded, that $\lim_{k\to\infty} \lambda_{\mathcal{E}}^{(k)}/\|\Delta\lambda^{(k)}\| = 0$, and that $\lim_{k\to\infty}\{w^{(k)}\}$ exists and is bounded. Further, assume that $\limsup_{k\to\infty}\{\widetilde{c}_i(w^{(k)}) + \nabla\widetilde{c}_i(w^{(k)})^T\Delta w^{(k)}\} < 0$ if $\lim_{k\to\infty}\widetilde{c}_i(w^{(k)}) = 0$. Finally, let I be given by Definition 3.1 and assume that $I \neq \emptyset$. Then, for $i \in I$, $\lim_{k\to\infty}\Delta\lambda_i^{(k)} = \infty$ and*

$$\lim_{k\to\infty} \sum_{i\in I\cup\mathcal{E}} \nabla\widetilde{c}_i(w^{(k)})\Delta\lambda_i^{(k)}/\|\Delta\lambda^{(k)}\| = 0. \tag{3.4}$$

Proof. First, note that $\lim_{k\to\infty}\widetilde{c}(w^{(k)})$ is well defined since $\widetilde{c}(w)$ is continuous and $\{w^{(k)}\}$ is assumed to converge. Lemma 3.1 implies that $\Delta\lambda_i^{(k)} \to \infty$ for $i \in I$. Since I is nonempty, $\lim_{k\to\infty}\|\Delta\lambda_I^{(k)}\| \to \infty$. Divide both sides of the first block row of (2.5) by $\|\Delta\lambda^{(k)}\|$ and let $k \to \infty$. Since $\{\widetilde{g}(w^{(k)}) - \widetilde{J}(w^{(k)})^T\lambda^{(k)}\}$ by assumption is bounded, this gives

$$\lim_{k\to\infty}\left(\frac{\nabla^2\widetilde{f}(w^{(k)})\Delta w^{(k)}}{\|\Delta\lambda^{(k)}\|} - \frac{\sum_i \lambda_i^{(k)}\nabla^2\widetilde{c}_i(w^{(k)})\Delta w^{(k)}}{\|\Delta\lambda^{(k)}\|} - \frac{\widetilde{J}(w^{(k)})^T\Delta\lambda^{(k)}}{\|\Delta\lambda^{(k)}\|}\right) = 0. \tag{3.5}$$

By assumption, $\{w^{(k)}\}$ is bounded. Since $\nabla^2\widetilde{f}(w)$ is continuous, $\{\nabla^2\widetilde{f}(w^{(k)})\}$ is bounded. Similarly, $\{\nabla^2\widetilde{c}_i(w^{(k)})\}$ is bounded for all i. Boundedness of $\{\nabla^2\widetilde{f}(w^{(k)})\}$ and $\{\Delta w^{(k)}\}$ implies that the first term of (3.5) converges to zero. Lemma 3.2 is applicable and along with the assumption $\lim_{k\to\infty}\lambda_{\mathcal{E}}^{(k)}/\|\Delta\lambda^{(k)}\| = 0$ and the boundedness of $\{\Delta w^{(k)}\}$ and $\{\nabla^2\widetilde{c}_i(w^{(k)})\}$, for $i \in \mathcal{E}\cup\mathcal{I}$, it implies that the second term also converges to zero. Hence, the third term of (3.5) converges to zero. The boundedness assumption on $\{\widetilde{C}_{\mathcal{I}}(w^{(k)})\lambda_{\mathcal{I}}^{(k)}\}$ implies that the right-hand side of (3.1) is bounded. For $i \in \mathcal{I}\setminus I$ it also implies that $\{\lambda_i^{(k)}\}$ is bounded since we assume that $\liminf \widetilde{c}_i(w^{(k)}) > 0$ for those indices. Since $\{w^{(k)}\}$ is convergent and $\widetilde{J}(w)$ is continuous, the first term of the left-hand side of (3.1) is also bounded, which implies that the second term is also bounded. Since $\lim_{k\to\infty}\widetilde{c}_i(w^{(k)}) > 0$ for $i \in \mathcal{I}\setminus I$, the sequence $\{\Delta\lambda_i^{(k)}\}$ must be bounded for $i \in \mathcal{I}\setminus I$. Thus, (3.4) follows. ∎

The assumption that the search direction satisfies (2.5) may seem restrictive since some methods make modifications to this system. However, Theorem 3.1 still holds if $H(w^{(k)}, \lambda^{(k)})$ is modified to $H(w^{(k)}, \lambda^{(k)}) + \delta H^{(k)}$, for $k = 0, \ldots$, as long as the sequence $\{\delta H^{(k)}\}$ is bounded. Further, the assumption that $\lim_{k\to\infty}\lambda_{\mathcal{E}}^{(k)}/\|\Delta\lambda^{(k)}\|$ is zero is not always necessary as will be seen in Section 4.

4. Specialization of results to barrier methods

As noted in Section 2.1, a barrier method is obtained by setting $\mu_{\mathcal{E}}$ to zero in (2.4) and (2.5). This allows us to specialize Theorem 3.1 to a barrier method. In this section we do this for both the slack variable reformulation, (P_{sl}), and the shift reformulation, (P_{sh}). The behavior of a barrier method applied to each of the reformulations of the problem (2.6) is also discussed. For simplicity, the subscript of $\mu_{\mathcal{I}}$ is dropped in this section since there is no risk of ambiguity.

4.1. Slack variable reformulation

The slack variable reformulation (P_{sl}) is on the form of problem (2.1) with $w^T = (x^T\ s^T)$, $\widetilde{f}(w) = f(x)$, $\widetilde{c}_{\mathcal{I}}(w) = s$ and $\widetilde{c}_{\mathcal{E}}(w) = c(x) - s$. Equation (2.5), with $\mu_{\mathcal{E}} = 0$ and $\mu_{\mathcal{I}} = \mu$, then gives

$$\begin{pmatrix} H(x,\lambda_{\mathcal{E}}) & 0 & 0 & -J(x)^T \\ 0 & 0 & -I & I \\ 0 & \Lambda_{\mathcal{I}} & S & 0 \\ J(x) & -I & 0 & 0 \end{pmatrix} \begin{pmatrix} \Delta x \\ \Delta s \\ \Delta\lambda_{\mathcal{I}} \\ \Delta\lambda_{\mathcal{E}} \end{pmatrix} = - \begin{pmatrix} g(x) - J(x)^T\lambda_{\mathcal{E}} \\ \lambda_{\mathcal{E}} - \lambda_{\mathcal{I}} \\ S\lambda_{\mathcal{I}} - \mu e \\ c(x) - s \end{pmatrix}. \tag{4.1}$$

Here, $H(x,\lambda_{\mathcal{E}})$ denotes the Hessian of $f(x) - \lambda_{\mathcal{E}}^T c(x)$ with respect to x. Note that there are no tildes since the functions involved are those of the original problem and not the functions of (P_{sl}). From the second block row of (4.1) it is clear that if the initial iterate satisfy $\lambda_{\mathcal{E}} = \lambda_{\mathcal{I}}$ then equality will hold throughout and one of the multiplier vectors can be eliminated. For the rest of this subsection, we assume that this elimination has been performed. Further, the subscript is dropped from $\lambda_{\mathcal{I}}$. Equation (4.1) can then be written as

$$\begin{pmatrix} H(x,\lambda) & 0 & -J(x)^T \\ 0 & \Lambda & S \\ J(x) & -I & 0 \end{pmatrix} \begin{pmatrix} \Delta x \\ \Delta s \\ \Delta\lambda \end{pmatrix} = - \begin{pmatrix} g(x) - J(x)^T\lambda \\ S\lambda - \mu e \\ c(x) - s \end{pmatrix}. \tag{4.2}$$

Given this system of equations for the search directions, it is straightforward to establish the following corollary to Theorem 3.1.

Corollary 4.1. *Let $\{x^{(k)}\}$, $\{s^{(k)}\}$, $\{\lambda^{(k)}\}$, $\{\Delta x^{(k)}\}$, $\{\Delta s^{(k)}\}$, $\{\Delta\lambda^{(k)}\}$, $\{\mu_{\mathcal{E}}^{(k)}\}$, and $\{\mu_{\mathcal{I}}^{(k)}\}$ be sequences fitting the framework of Algorithm 2.1 with $w^T = (x^T\ s^T)$, $\widetilde{f}(w) = f(x)$, $\widetilde{c}_{\mathcal{I}}(w) = s$, $\widetilde{c}_{\mathcal{E}}(w) = c(x) - s$, $\lambda = \lambda_{\mathcal{E}} = \lambda_{\mathcal{I}}$, and (2.5) replaced by (4.2). Assume that the sequences $\{\mu_{\mathcal{I}}^{(k)}\}$, $\{\Delta s^{(k)}\}$, $\{\Delta x^{(k)}\}$, $\{g(x^{(k)}) - J(x^{(k)})^T\lambda^{(k)}\}$, and $\{S^{(k)}\lambda^{(k)}\}$ are bounded, that $\lim_{k\to\infty}\{x^{(k)}\}$ exists and is finite, and that $\mu_{\mathcal{E}}^{(k)} = 0$ for all k. Also, assume that if $\liminf_{k\to\infty} s_i^{(k)} = 0$ then $\limsup_{k\to\infty}\{s_i^{(k)} + \Delta s_i^{(k)}\} < 0$. Finally, let I_{sl} be the set of indices i such that $\lim_{k\to\infty} s_i^{(k)} = 0$, and assume that $I_{sl} \neq \emptyset$. Then, $\lim_{k\to\infty} \Delta\lambda_i^{(k)} = \infty$ for $i \in I_{sl}$ and*

$$\lim_{k\to\infty} J_{I_{sl}}(x^{(k)})^T \Delta\lambda_{I_{sl}}^{(k)} / \|\Delta\lambda^{(k)}\| = 0. \tag{4.3}$$

Proof. First, note that $\liminf_{k\to\infty} s_i^{(k)} = 0$ means that the limit actually exists and is zero because $s_i^{(k)} + \Delta s_i^{(k)}$ is negative for all k large enough. Therefore, $\liminf_{k\to\infty} s_i^{(k)} > 0$ for all $i \notin I_{sl}$. For Theorem 3.1 to apply $\{s_i^{(k)}\}$ should be assumed to be convergent. However, it is straightforward to verify that when $\widetilde{c}_{\mathcal{I}}(w) = s$, the proofs of Theorem 3.1 and the lemmas it depends on do not require this assumption. The assumption that $\lambda_{\mathcal{E}}^{(k)}/\|\Delta\lambda^{(k)}\|$ converges to zero is not necessary since $\lambda_{\mathcal{I}} = \lambda_{\mathcal{E}}$ and Lemma 3.2 implies that $\lambda_{\mathcal{I}}^{(k)}/\|\Delta\lambda^{(k)}\|$ converges to zero. Therefore Theorem 3.1 implies that

$$\lim_{k\to\infty} \left\{ \begin{pmatrix} J(x^{(k)})^T \\ -I \end{pmatrix} \frac{\Delta\lambda^{(k)}}{\|\Delta\lambda^{(k)}\|} + \sum_{i\in I_{sl}} \frac{\Delta\lambda_i^{(k)}}{\|\Delta\lambda^{(k)}\|} \begin{pmatrix} 0 \\ e_i \end{pmatrix} \right\} = 0,$$

where e_i is the ith unit vector. The second block states that $\{\Delta\lambda_i^{(k)}/\|\Delta\lambda^{(k)}\|\}$ converges to zero for all $i \notin I_{sl}$. Inserted in the first block, this gives (4.3). ∎

The corollary could have been stated in a form that is a direct application of Theorem 3.1 to the slack variable reformulation. However, (4.2) is more frequently used than (4.1) and the corollary is therefore given so that it matches the former system of equations.

Let us return to the slack variable reformulation of (2.6). The practical behavior of a barrier method applied to this problem is that both slack variables converge to zero while no feasible point is reached, even asymptotically. Further, the assumptions of Corollary 4.1 are typically satisfied and the corollary implies that the vector of multiplier search directions must therefore diverge in a direction that asymptotically lies in the null space of $J(x^*)^T$, i.e., in the null space of the matrix $(2x^*\ 1)$, where x^* is the limit point. This behavior is also observed in practice.

It is worth noting that the ill-conditioning of the matrix in (4.2) increases as the limit point is approached. In the limit the matrix is singular. To see this, note that the third block column of the matrix loses rank when both s_1 and s_2 converge to zero.

4.2. Shift reformulation

In the previous section Theorem 3.1 was applied to the slack variable reformulation, (P_{sl}). In this section we turn our attention to the shift reformulation, (P_{sh}), which is a problem on the form (2.1) if $w^T = (x^T\ s^T)$, $\widetilde{f}(w) = f(x)$, $\widetilde{c}_{\mathcal{I}}(w) = c(x) + s$, and $\widetilde{c}_{\mathcal{E}}(w) = s$. With $\mu_{\mathcal{E}} = 0$ and $\mu_{\mathcal{I}} = \mu$, equation (2.5) gives

$$\begin{pmatrix} H(x,\lambda_{\mathcal{I}}) & 0 & -J(x)^T & 0 \\ 0 & 0 & -I & -I \\ \Lambda_{\mathcal{I}} J(x) & \Lambda_{\mathcal{I}} & C(x)+S & 0 \\ 0 & I & 0 & 0 \end{pmatrix} \begin{pmatrix} \Delta x \\ \Delta s \\ \Delta\lambda_{\mathcal{I}} \\ \Delta\lambda_{\mathcal{E}} \end{pmatrix} = - \begin{pmatrix} g(x) - J(x)^T\lambda_{\mathcal{I}} \\ -\lambda_{\mathcal{E}} - \lambda_{\mathcal{I}} \\ (C(x)+S)\lambda_{\mathcal{I}} - \mu e \\ s \end{pmatrix}. \quad (4.4)$$

Here, $H(x,\lambda_{\mathcal{I}})$ denotes the Hessian of $f(x) - \lambda_{\mathcal{I}}^T c(x)$ with respect to x. If initially $\lambda_{\mathcal{E}} = -\lambda_{\mathcal{I}}$, one set of multipliers can be eliminated. Eliminating $\lambda_{\mathcal{E}}$ from (4.4) and dropping the subscript from $\lambda_\mathcal{I}$ gives

$$\begin{pmatrix} H(x,\lambda) & 0 & -J(x)^T \\ 0 & I & 0 \\ \Lambda J(x) & \Lambda & C(x)+S \end{pmatrix} \begin{pmatrix} \Delta x \\ \Delta s \\ \Delta\lambda \end{pmatrix} = - \begin{pmatrix} g(x) - J(x)^T\lambda \\ s \\ (C(x)+S)\lambda - \mu e \end{pmatrix}. \tag{4.5}$$

In parallel to Section 4.1, we give a corollary which shows that the multiplier search directions corresponding to constraints for which $\lim_{k\to\infty} c_i(x^{(k)}) + s_i^{(k)} = 0$ must diverge and do so in a direction in the null space of the transpose of the Jacobian at the limit point.

Corollary 4.2. *Let* $\{x^{(k)}\}$, $\{s^{(k)}\}$, $\{\lambda^{(k)}\}$, $\{\Delta x^{(k)}\}$, $\{\Delta s^{(k)}\}$, $\{\Delta\lambda^{(k)}\}$, $\{\mu_\mathcal{E}^{(k)}\}$, *and* $\{\mu_\mathcal{I}^{(k)}\}$ *be sequences fitting the framework of the algorithm in Section 2.1 with* $w^T = (x^T\ s^T)$, $\widetilde{f}(w) = f(x)$, $\widetilde{c}_\mathcal{I}(w) = c(x) + s$, $\widetilde{c}_\mathcal{E}(w) = s$, $\lambda = -\lambda_\mathcal{E} = \lambda_\mathcal{I}$, *and (2.5) replaced by (4.5). Assume that the sequences* $\{\mu_\mathcal{I}^{(k)}\}$, $\{\Delta s^{(k)}\}$, $\{\Delta x^{(k)}\}$, $\{g(x^{(k)}) - J(x^{(k)})^T\lambda^{(k)}\}$, *and* $\{(C(x^{(k)}) + S^{(k)})\lambda^{(k)}\}$ *are bounded, that both* $\lim_{k\to\infty}\{x^{(k)}\}$ *and* $\lim_{k\to\infty}\{s^{(k)}\}$ *exist and are finite, and that* $\mu_\mathcal{E}^{(k)} = 0$ *for all* k*. Also, assume that if* $\liminf_{k\to\infty}\{c_i(x^{(k)}) + s_i^{(k)}\} = 0$ *then* $\limsup_{k\to\infty}\{c_i(x^{(k)}) + \nabla c_i(x^{(k)})\Delta x^{(k)}\} < 0$*. Finally, let* I_{sh} *be the set of indices* i *such that* $\lim_{k\to\infty}\{c_i(x^{(k)}) + s_i^{(k)}\} = 0$ *and assume that* $I_{sh} \neq \emptyset$*. Then,* $\Delta\lambda_i^{(k)} \to \infty$ *for* $i \in I_{sh}$ *and*

$$\lim_{k\to\infty} J_{I_{sh}}(x^{(k)})^T \Delta\lambda_{I_{sh}}^{(k)} / \|\Delta\lambda^{(k)}\| = 0. \tag{4.6}$$

Proof. The lim sup expression does not contain the s_i variable since $s_i^{(k)} + \Delta s_i^{(k)} = 0$. The assumption that $\lambda_\mathcal{E}^{(k)}/\|\Delta\lambda^{(k)}\|$ converges to zero is not necessary since $\lambda_\mathcal{I} = -\lambda_\mathcal{E}$ and Lemma 3.2 implies that $\lambda_\mathcal{I}^{(k)}/\|\Delta\lambda^{(k)}\|$ converges to zero. Applying Theorem 3.1 and keeping in mind that $\Delta\lambda = \Delta\lambda_\mathcal{I} = -\Delta\lambda_\mathcal{E}$ gives

$$\lim_{k\to\infty}\left\{ \sum_{i\in I_{sh}} \begin{pmatrix} \nabla c_i(x) \\ e_i \end{pmatrix} \frac{\Delta\lambda_i}{\|\Delta\lambda^{(k)}\|} + \begin{pmatrix} 0 \\ I \end{pmatrix} \frac{-\Delta\lambda^{(k)}}{\|\Delta\lambda^{(k)}\|} \right\} = 0,$$

where e_i is the ith unit vector. The second block implies that $\{\Delta\lambda_i^{(k)}/\|\Delta\lambda^{(k)}\|\}$ converges to zero for all $i \notin I_{sh}$. Inserted in the first block, this gives (4.6). ∎

Above, we have shown that if some components of $c(x) + s$ converge to zero, the corresponding multiplier search directions diverge along a direction in the null space of the transpose of the Jacobian. Numerical experiments suggest that such divergence occurs for a shift variable reformulation of (2.6), with $a < 0$, $b \geq 0$, and with suitable initial values. The objective cannot be left unspecified as it was in the slack variable case, so we let $f(x) = x$. Then, the shift reformulation of (2.6) with $a = -1$ and $b = 0.5$ is

$$\begin{array}{ll} \underset{x\in I\!R,\, s\in I\!R^2}{\text{minimize}} & x \\ \text{subject to} & x^2 - 1 + s_1 \geq 0, \\ & x - 0.5 + s_2 \geq 0, \\ & s_1 = 0, \\ & s_2 = 0. \end{array}$$

Empirically, for suitable choices of initial values, x in practice converges to $x^* = -1$ and s converges to $s^* = (0\ 1.5)^T$ which is not feasible. Unfortunately, we have not been able to prove this analytically. As x and s converge to x^* and s^*, the multiplier search directions asymptotically diverge in the direction $(1\ 2)^T$. As predicted by Corollary 4.2, this direction lies in the null space of the transpose of the Jacobian of the constraints of (2.6) at x^*. Notice that the third block column of the matrix in equation (4.5) does not have full rank in the limit, i.e. when $x = x^*$ and $s = s^*$. This is analogous to the slack variable case. It can be noted that if the initial x is negative and the initial $s_1 < 1$, s_1 will be less than one throughout. Therefore, to obtain a feasible point, tunneling, as defined in Section 2.1, must occur.

5. Similar results for penalty-barrier methods

In this section we turn our attention from barrier to penalty-barrier methods. Since the qualitative differences, in this context, between a method based on the slack variable reformulation and one based on the shift reformulation are small, we only discuss the slack variable case. It turns out that as the penalty and barrier parameters are driven to zero, a behavior similar to the one observed for barrier methods may occur. To motivate this, we give a theorem that is reminiscent of Theorem 3.1, but concerns the asymptotic behavior of the trajectory instead of the search directions.

Applying the system of equations (2.5) to the slack variable reformulation (P_{sl}) with arbitrary positive parameters $\mu_{\mathcal{E}}$ and $\mu_{\mathcal{I}}$ gives a system that resembles (4.1). As in Section 4.1, we assume that $\lambda_{\mathcal{E}}^{(0)} = \lambda_{\mathcal{I}}^{(0)}$, so that one set of multipliers can be eliminated. Further, for all k, let $\mu^{(k)} = \mu_{\mathcal{E}}^{(k)}$ and assume that $\mu_{\mathcal{E}}^{(k)} = \mu_{\mathcal{I}}^{(k)}$. This gives the system

$$\begin{pmatrix} H(x,\lambda) & 0 & -J(x)^T \\ 0 & \Lambda & S \\ J(x) & -I & \mu I \end{pmatrix} \begin{pmatrix} \Delta x \\ \Delta s \\ \Delta\lambda \end{pmatrix} = - \begin{pmatrix} g(x) - J(x)^T\lambda \\ S\lambda - \mu e \\ c(x) - s + \mu\lambda \end{pmatrix}, \tag{5.1}$$

where the subscript of λ has been dropped. In contrast to the barrier method, the last block column cannot lose rank when μ is fixed and positive. However, as μ goes to zero, there is nothing preventing ill conditioning and asymptotic loss of rank.

Let us return to the problem (2.6). Numerical experiments suggest that a penalty transformation of the equality constraints can be helpful in avoiding the convergence difficulties discussed in Section 4.1. For example, with $a = -1$, $b = 0.5$, $f(x) = x$, and a negative initial x, a penalty-barrier method converges to the true solution for reasonable initial values if the Hessian of the Lagrangian is modified whenever necessary. This result is in contrast to the one for a barrier method, as discussed in Section 4.1. The behavior is interesting since, if μ is decreased slowly, x remains negative for large values of μ and the changes between subsequent iterates are small. Then, as μ drops below a threshold value, x makes a large jump to a positive value. The reason for this lies with the penalty-barrier trajectory. For large values of μ, there are two minimizers of the penalty-barrier function. One has negative and the other positive x component. Below some threshold value of μ, the negative minimizer vanishes. This is what forces the jump.

However, there are a and b values for which x remains negative throughout. The reason is that the part of the trajectory with negative values of x exists for all positive μ. Therefore, a trajectory that converges to a point (x^*, s^*), with x^* negative and $s^* = 0$ is followed approximately. However, the limit point is not even feasible. The sequence of multiplier search directions then asymptotically diverge in the direction $(2x^* \;\; 1)^T$ which lies in the null space of the transpose of the Jacobian at x^*. Note that this does *not* follow from Theorem 3.1 since for each $\mu > 0$, the problem has a well-defined interior solution with negative x. Therefore, the condition that $\limsup_{k\to\infty}\{c_i(w^{(k)}) + \nabla c_i(w^{(k)})^T \Delta w^{(k)}\} < 0$ for all inequality constraints that are asymptotically active does not hold in general. However, the behavior is supported by the theorem below. To prove the theorem, we need a lemma that is a counterpart to Lemma 3.1 for the penalty-barrier trajectory.

Lemma 5.1. *Assume that the penalty-barrier trajectory $(w(\mu), \lambda(\mu))$ associated with problem (2.1) exists for $\mu \in (0, \bar{\mu})$ for some positive constant $\bar{\mu}$ and that $w(\mu)$ converges as $\mu \to 0^+$. Let $I_c = \{i \in \mathcal{E} \cup \mathcal{I} : \lim_{\mu\to 0^+} c_i(w(\mu)) \neq 0\}$. Further, for each $i \in I_c$, assume that $\lambda_i'(\mu)$ exists for all μ small enough and that $\gamma = \lim_{\mu\to 0^+} \lambda_i'(\mu)/\lambda_i(\mu)$ is well defined in the sense that it exists in $I\!R \cup \{-\infty\}$. Then,*

$$\lim_{\mu\to 0^+} |\lambda_i(\mu)| = \infty, \quad i \in I_c, \tag{5.2a}$$

$$\lim_{\mu\to 0^+} \lambda_i(\mu)/\lambda_i'(\mu) = 0, \quad i \in I_c. \tag{5.2b}$$

Proof. Equation (2.4c) implies that $c_i(w(\mu)) + \mu\lambda_i(\mu) = 0$ for $i \in \mathcal{E}$. Note that $\lim_{\mu\to 0^+} c_i(w(\mu))$ is well defined since $c(w)$ is continuous and $w(\mu)$ is assumed to converge. Therefore, $\lim_{k\to\infty} |\lambda_i(\mu)| = \infty$ for $i \in I_c \cap \mathcal{E}$ as $\mu \to 0^+$ since $\lim_{\mu\to 0^+} c_i(w(\mu)) \neq 0$ for such indices i. Similarly, $c_i(w(\mu))\lambda_i(\mu) = \mu e$ for $i \in \mathcal{I}$ implies that $|\lambda_i(\mu)| \to \infty$ as $\mu \to 0^+$ for $i \in \mathcal{I} \cap I_c$, proving (5.2a). We can without loss of generality assume that $\lambda_i(\mu) \to +\infty$ as $\mu \to 0^+$. Then, $\ln(\lambda_i(\mu))$ is also unbounded above. The derivative of $\ln(\lambda_i(\mu))$ is $\lambda_i'(\mu)/\lambda_i(\mu)$, i.e. the quotient in the limit defining γ. Since γ is assumed to be well defined, in the sense stated above, and since no function with bounded derivative can be unbounded on a finite interval, γ must be $-\infty$. Then, (5.2b) follows by taking the inverse. ∎

The following theorem gives a characterization of the asymptotic behavior of the multiplier part of the penalty-barrier trajectory as the w part approaches an infeasible point. The theorem can be seen as a counterpart to Theorem 3.1 in the same way as the previous lemma is a counterpart to Lemma 3.1.

Theorem 5.1. *Let the assumptions of Lemma 5.1 hold. Further, assume that I_c is nonempty, that $\lambda'(\mu)$ exists for all positive μ small enough, and that $w'(\mu)$ exists and is bounded as $\mu \to 0^+$. Also, for $i \notin I_c$, assume that if $\lambda_i(\mu)$ is unbounded as*

$\mu \to 0^+$ *then* $|\lambda_i(\mu)| \to \infty$. *Then,*

$$\lim_{\mu\to 0^+} \sum_{i:|\lambda_i'(\mu)|\to\infty} \nabla c_i(w(\mu)) \frac{\lambda_i'(\mu)}{\|\lambda'(\mu)\|} = 0. \tag{5.3}$$

Proof. Differentiating (2.4a) along the trajectory $(w(\mu), \lambda_{\mathcal{E}}(\mu), \lambda_{\mathcal{I}}(\mu))$ with respect to μ gives

$$H(w(\mu), \lambda(\mu))w'(\mu) - J(w(\mu))^T \lambda'(\mu) = 0. \tag{5.4}$$

Since I_c is nonempty, Lemma 5.1 implies that $\|\lambda'(\mu)\|$ is unbounded as $\mu \to 0^+$, and for μ small enough, we can divide (5.4) by $\|\lambda'(\mu)\|$. Letting $\mu \to 0^+$ gives

$$\lim_{\mu\to 0^+} \left(\frac{\nabla^2 f(w(\mu))w'(\mu)}{\|\lambda'(\mu)\|} - \frac{\sum_i \lambda_i(\mu)\nabla^2 c_i(w(\mu))w'(\mu)}{\|\lambda'(\mu)\|} - \frac{J(w(\mu))^T\lambda'(\mu)}{\|\lambda'(\mu)\|} \right) = 0. \tag{5.5}$$

When $\mu \to 0^+$, the Hessian $\nabla^2 f(w(\mu))$ is bounded since $\nabla^2 f(w)$ is a continuous functions and $w(\mu)$ is bounded when $\mu \to 0^+$. Similarly, $\nabla^2 c_i(w(\mu))$ is bounded for all i as $\mu \to 0^+$. Boundedness of $w'(\mu)$ then implies that the first term of (5.5) converges to zero. If $\lambda_i(\mu)/\|\lambda'(\mu)\|$ converges to zero for all i then the second term of (5.5) also converges to zero. We now show that this is the case. For $i \in I_c$, Lemma 5.1 gives $\lambda_i(\mu)/\|\lambda'(\mu)\| \to 0$. For $i \in \mathcal{I} \setminus I_c$, there are two cases, either $\lambda_i(\mu)$ is bounded or $\lambda_i(\mu) \to \infty$ as $\mu \to 0^+$. If it is bounded the desired result follows immediately. If it is unbounded, an argument equivalent to the one in the proof of Lemma 5.1 shows that $\lambda_i(\mu)/\lambda_i'(\mu) \to 0$ as $\mu \to 0^+$. Thus, the second term converges to zero implying that the third term also converges to zero. Since $\|\lambda'(\mu)\| \to \infty$ and $\nabla c(w(\mu))$ is bounded as $\mu \to 0^+$, only those indices for which $\lambda_i'(\mu) \to \infty$ as $\mu \to 0^+$ give any contribution. This completes the proof. ∎

Theorem 5.1 concerns the trajectory while the discussion in the beginning of this section concerns the search direction. However, these are related topics since at a point on the trajectory, the search direction given by (2.5), after a reduction of μ, is parallel to the tangent of the trajectory. To see this, evaluate (2.5) at a point on the trajectory and compare the system with the system obtained by differentiating the primal-dual penalty-barrier equations (2.2) along the trajectory with respect to μ. The two systems of equations differ only by a scaling of the right-hand side vector. Therefore, the solution vectors are parallel.

Let us return to the slack variable reformulation of (2.6). Empirically, $\lambda_i(\mu)$ is asymptotically proportional to the inverse of μ while $x(\mu)$ and $s(\mu)$ are asymptotically affine in μ as $\mu \to 0^+$. The assumptions of the theorem are therefore satisfied and we get an explanation of the empirical observation that the multiplier search directions diverge in a direction in the null space of the transpose of the Jacobian.

6. Summary and discussion

The main result of this paper concerns the behavior of the multiplier search directions as the value of some constraint are driven to zero prematurely. The divergence of the multiplier search directions relies entirely on the search directions satisfying a linearization of the perturbed complementarity constraints. To establish that the direction of divergence lies in the null space of the transpose of the part of the Jacobian which corresponds to the equality constraints and the asymptotically active inequality constraints, we use the Newton equations for the the part of the primal-dual equations relating the gradient of the objective and the gradients of the constraints.

Further, we have demonstrated that a penalty transformation of the equality constraints can have a regularizing effect. However, this is by no means a complete remedy since the penalty transformation can cause the penalty-barrier function (2.2) to be unbounded below even if the original problem as well as the barrier subproblem are bounded below. Also, convergence difficulties similar to those demonstrated by Wächter and Biegler [WB00] for barrier methods can occur as the barrier and penalty parameters are driven to zero. How frequent this kind of behavior is for practical problems is a question for future research.

An interesting topic for further research is a formal analysis of the convergence difficulties for a shift reformulation similar to the one Wächter and Biegler have done for the slack reformulation. Obviously, one has to make additional assumptions on the method used since a linearization of the equality constraints of a shift reformulated problem does not give enough information.

References

[BGN00] R. H. Byrd, J. C. Gilbert, and J. Nocedal. A trust region method based on interior point techniques for nonlinear programming. *Math. Program.*, 89, 149–185, 2000.

[BHN99] R. H. Byrd, M. E. Hribar, and J. Nocedal. An interior point algorithm for large-scale nonlinear programming. *SIAM J. Optim.*, 9, 877–900, 1999.

[BMN01] R. H. Byrd, M. Marazzi, and J. Nocedal. On the convergence of Newton iterations to non-stationary points. Report OTC 2001/01, Optimization Technology Center, Department of Electrical Engineering and Computer Science, Northwestern University, Evanston, IL 60208, 2001.

[BSV00] H. Y. Benson, D. F. Shanno, and R. J. Vanderbei. Interior-point methods for nonconvex nonlinear programming: Jamming and comparative numerical testing. Report ORFE-00-02, Department of Operations Research and Financial Engineering, Princeton University, 2000.

[CGT94] A. R. Conn, N. I. M. Gould, and Ph. L. Toint. A note on using alternative second-order models for the subproblems arising in barrier function methods for minimization. *Numer. Math.*, 68, 17–33, 1994.

[CGT96] A. R. Conn, N. I. M. Gould, and Ph. L. Toint. A primal-dual algorithm for minimizing a nonconvex function subject to bound and linear equality constraints. Report RC 20639, IBM T. J. Watson Research Center, Yorktown Heights, New York, 1996.

[CGT00] A. R. Conn, N. I. M. Gould, and P. L. Toint. *Trust-region methods.* MPS/SIAM Series on Optimization. Society for Industrial and Applied Mathematics (SIAM), Mathematical Programming Society (MPS), Philadelphia, PA, 2000. ISBN 0-89871-460-5.

[EBTTZ96] A. S. El-Bakry, R. A. Tapia, T. Tsuchiya, and Y. Zhang. On the formulation and theory of the newton interior-point method for nonlinear programming. *J. Optim. Theory Appl.*, 89, 507–541, 1996.

[FG98] A. Forsgren and P. E. Gill. Primal-dual interior methods for nonconvex nonlinear programming. *SIAM J. Optim.*, 8, 1132–1152, 1998.

[FM68] A. V. Fiacco and G. P. McCormick. *Nonlinear Programming: Sequential Unconstrained Minimization Techniques*. John Wiley and Sons, Inc., New York, 1968. Republished by Society for Industrial and Applied Mathematics, Philadelphia, 1990. ISBN 0-89871-254-8.

[Fre91] R. M. Freund. Theoretical efficiency of a shifted-barrier-function algorithm for linear programming. *Linear Algebra Appl.*, 152, 19–41, 1991.

[GMSW88] P. E. Gill, W. Murray, M. A. Saunders, and M. H. Wright. Shifted barrier methods for linear programming. Report SOL 88-9, Department of Operations Research, Stanford University, Stanford, CA, 1988.

[GOW98] D. M. Gay, M. L. Overton, and M. H. Wright. A primal-dual interior method for nonconvex nonlinear programming. In Y. Yuan, editor, *Advances in Nonlinear Programming*, pages 31–56. Kluwer Academic Publishers, Dordrecht, 1998.

[JS95] F. Jarre and M. A. Saunders. A practical interior-method for convex programming. *SIAM J. Optim.*, 5, 149–171, 1995.

[Mel96] A. Melman. A linesearch procedure in barrier methods for some convex programming problems. *SIAM J. Optim.*, 6, 283–298, 1996.

[MN00] M. Marazzi and J. Nocedal. Feasibility control in nonlinear optimization. Report OTC 2000/04, Optimization Technology Center, Department of Electrical Engineering and Computer Science, Northwestern University, Evanston, IL 60208, 2000.

[MS79] J. J. Moré and D. C. Sorensen. On the use of directions of negative curvature in a modified Newton method. *Math. Program.*, 16, 1–20, 1979.

[MW94] W. Murray and M. H. Wright. Line search procedures for the logarithmic barrier function. *SIAM J. Optim.*, 4, 229–246, 1994.

[NW99] J. Nocedal and S. J. Wright. *Numerical Optimization*. Springer, New York, 1999. ISBN 0-387-98793-2.

[Pol92] R. Polyak. Modified barrier functions (theory and methods). *Math. Program.*, 54, 177–222, 1992.

[Pow69] M. J. D. Powell. A method for nonlinear constraints in minimization problems. In R. Fletcher, editor, *Optimization*, pages 283–298, London and New York, 1969. Academic Press.

[SV99] D. F. Shanno and R. J. Vanderbei. An interior-point algortihm for nonconvex nonlinear programming. *J. Comput. Optim. Appl.*, 13, 231–253, 1999.

[Tod94] M. J. Todd. Scaling, shifting and weighting in interior-point methods. *Comput. Optim. Appl.*, 3, 305–315, 1994.

[UUV00] M. Ulbrich, S. Ulbrich, and L. N. Vicente. A globally convergent primal-dual interior-point filter method for nonconvex nonlinear programming. Technical Report TR00-12, Department of Computational and Applied Mathematics, Rice University, April 2000.

[WB00] 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.

[WB01] A. Wächter and L. T. Biegler. Global and local convergence of line search filter methods for nonlinear programming. Technical Report CAPD B-01-09, Department of Chemical Engineering, Carnegie Mellon University, Pittsburgh, PA 15213, August 2001. Revised November 2001.