The Newton modified barrier method for QP problems

A. Melman
Department of Industrial Engineering and Management, Ben-Gurion University, Beer-Sheva 84105, Israel
E-mail: melman@bgumail.bgu.ac.il

R. Polyak*
Department of Operations Research, George Mason University, Fairfax, VA 22030, USA
E-mail: rpolyak@gmuvax.gmu.edu

The Modified Barrier Functions (MBF) have elements of both Classical Lagrangians (CL) and Classical Barrier Functions (CBF). The MBF methods find an unconstrained minimizer of some smooth barrier function in primal space and then update the Lagrange multipliers, while the barrier parameter either remains fixed or can be updated at each step. The numerical realization of the MBF method leads to the Newton MBF method, where the primal minimizer is found by using Newton's method. This minimizer is then used to update the Lagrange multipliers. In this paper, we examine the Newton MBF method for the Quadratic Programming (QP) problem. It will be shown that under standard second-order optimality conditions, there is a ball around the primal solution and a cut cone in the dual space such that for a set of Lagrange multipliers in this cut cone, the method converges quadratically to the primal minimizer from any point in the aforementioned ball, and continues to do so after each Lagrange multiplier update. The Lagrange multipliers remain within the cut cone and converge linearly to their optimal values. Any point in this ball will be called a "hot start". Starting at such a "hot start", at most $\mathcal{O}(\ln \ln \varepsilon^{-1})$ Newton steps are sufficient to perform the primal minimization which is necessary for the Lagrange multiplier update. Here, $\varepsilon > 0$ is the desired accuracy. Because of the linear convergence of the Lagrange multipliers, this means that only $\mathcal{O}(\ln \varepsilon^{-1})\mathcal{O}(\ln \ln \varepsilon^{-1})$ Newton steps are required to reach an ε-approximation to the solution from any "hot start". In order to reach the "hot start", one has to perform $\mathcal{O}(\sqrt{m} \ln C)$ Newton steps, where m characterizes the size of the problem and $C > 0$ is the condition number of the QP problem. This condition number will be characterized explicitly in terms of key parameters of the QP problem, which in turn depend on the input data and the size of the problem.

Keywords: Modified Barrier Function, Lagrange multipliers, relaxation operator, condition number, "hot start", complexity.

* Partially supported by NASA Grant NAG3-1397 and National Science Foundation Grant DMS-9403218.

1. Introduction

The most remarkable property of the Classical Barrier Functions (CBFs) is their self-concordance (see [19]). This guarantees for each Newton step, starting from a "warm start", i.e., a point "close" to the central path (see [11]), the decrease of the duality gap by a factor which depends only on the size of the problem (see [11, 19] and bibliography in it). Every new approximation is again "well defined" (see [31]) for the CBF when the barrier parameter is decreased by a factor, which depends only on the size of the problem.

In other words, a "careful" barrier parameter update allows us to stay in the Newton area, starting from the first "warm start" up to the end of the process. It guarantees that every Newton step reduces the duality gap by a constant factor. In the case of linear programming (LP) (see [28]), this factor is given by $(1-(41\sqrt{m})^{-1})$ (recently Smale and Shub [32] proved that this factor can be improved to $(1-(13\sqrt{m})^{-1})$). This means that $O(\sqrt{m})$ Newton steps are required to reduce the duality gap by a factor of two. A more drastic barrier parameter update does not lead to any improvements in the complexity bound. The so-called long step interior point methods (see [11, 30]) do not have better complexity bounds, although their practical performance is better. In this paper, we will estimate the complexity of the Newton Modified Barrier Function (MBF) method for quadratic programming (QP) problems.

In contrast to CBF methods (see [7, 9]), the MBF method for QP converges for a fixed positive barrier parameter, whether the constrained optimization problem is degenerate or not (see [15, 26]). Note that in the LP case, the convergence of the MBF method is R-linear (see [27]).

In the case of nondegenerate constrained optimization, the primal and dual MBF sequences converge at least linearly (see [24]), at a rate which can be made as small as one wants by choosing a fixed but sufficiently small barrier parameter.

We will consider the MBF method with a fixed barrier parameter, determined by the condition of the QP problem. We will show that from a certain point on, the Newton MBF method converges such that, after every Lagrange multiplier update, the current primal iterate is "well defined" (see [31]) for the next MBF, while the barrier parameter is fixed.

This means that from this point on, the "warm start" turns into a "hot start", i.e., with $\varepsilon > 0$ the desired accuracy, only $O(\ln \ln \varepsilon^{-1})$ Newton steps are necessary to compute the current primal minimizer and to update the dual variables, which leads to an improvement of the primal and dual approximation by a given factor $0 < \gamma < 1$. Moreover, the number of Newton steps is decreasing from one Lagrange multiplier update to another until finally only one Newton step suffices for the multiplier update.

We will characterize the "hot start" through the condition number $C > 0$ of the constrained optimization problem, which was introduced in [24] (see also [25] for LP). The condition number is in turn characterized through the basic parameters of the QP problem, which depend on the input data and the size of the problem.

In order to reach the " hot start", one can use any Interior Point Method. Numerical results with MBF methods for both linear and nonlinear programming problems (see [2,3,14,18]) strongly corroborate the theory of the "hot start" phenomenon.

The fundamental difference between the MBF approach from the interior point methods based on CBF is the convergence of the MBF method for fixed positive barrier parameter. It contributes to the stability of both the condition number of the MBF Hessian and the area where Newton's method is "well defined", i.e., where the convergence is of quadratic order. It makes the Newton MBF method numerically stable and causes the aforementioned "hot start" phenomenon.

The paper is organized as follows. After the statement of the problem, we introduce some basic facts about the MBF theory for QP problems. We then describe the Newton MBF method for QP in more detail and prove some basic properties concerning this method. Based on these properties, we will prove the existence of the "hot start" and estimate the complexity of the Newton MBF method for non-degenerate QP.

2. Statement of the problem and the MBF for QP

We assume that $Q \in \mathbb{R}^{(n,n)}$ is a symmetric positive semidefinite matrix, that $a \in \mathbb{R}^n$, that the feasible set $\Omega = \{x \mid r_i(x) \geq 0,\ i = 1,\dots,m\}$ is bounded and that

$$int\,\Omega \neq \varnothing, \tag{1}$$

where $r_i(x) = a_i^T x - b_i$; $x, a_i \in \mathbb{R}^n$; $b_i \in \mathbb{R}$ and $\|a_i\| = 1$.

We consider the following QP problem:

$$x^* = \arg\min \left\{ f_0(x) = \tfrac{1}{2} x^T Q x - a^T x \mid x \in \Omega \right\}. \tag{2}$$

Let $k > 0$. We then consider the extended feasible set

$$\Omega_k = \{x \mid r_1(x) \geq -k^{-1},\ i = 1.\dots, m\}. \tag{3}$$

The MBF for problem (2) is given by the following expression (see [24]):

$$F(x,u,k) = f_0(x) - k^{-1} \sum_{i=1}^{m} u_i \ln(kr_i(x) + 1), \tag{4}$$

and assuming that for $t \leq 0$, $\ln t = -\infty$. Taking into account that

$$\Omega = \{x \mid r_i(x) \geq 0,\ i = 1,\dots,m\} = \{x \mid k^{-1} \ln(kr_i(x) + 1) \geq 0,\ i = 1,\dots,m\},$$

we see that the MBF $F(x, u, k)$ is a classical Lagrangian for an equivalent problem. In view of (1), there exists a vector $u^* = (u_1^*, \dots, u_m^*) \in \mathbb{R}_+^m$, such that

$$Qx^* - a = Au^*, \tag{5}$$

and

$$u_i^* r_i(x^*) = 0 \quad i = 1, \dots, m. \tag{6}$$

Here, A is an $n \times m$ matrix whose columns are given by a_i.

Let $I = \{1, \dots, r\} = \{i : r_i(x^*) = 0\}$, with $r < n < m$, be the active constraints set and let $J = \{r+1, \dots, m\} = \{i : r_i(x^*) > 0\}$ be the passive constraints set. We define the matrices $A_r \in \mathbb{R}^{n,r}$, the columns of which are given by the a_i's, $i \in I$, and the matrix $A_{m-r} \in \mathbb{R}^{n,m-r}$, the columns of which are given by the vectors a_i, $i \in J$. We also define the vectors $u_{(r)}^T = (u_1, \dots, u_r)$, $u_{(m-r)}^T = (u_{r+1}, \dots, u_m)$ and a diagonal matrix $U_r = [\text{diag}(u_i)]_{i=1}^r$ with entries u_i. It will be assumed that

$$\text{rank } A_r = r \quad \text{and} \quad u_{(r)}^* \in \mathbb{R}_{++}^r. \tag{7}$$

Let $L(x, u) = f_0(x) - \sum_{i=1}^m u_i r_i(x)$ be the Classical Lagrangian for the original problem, then $\nabla_{xx}^2 L(x, u) = Q$. We will also assume that there exists $\tau > 0$ such that

$$(Qy, y) \geq \tau(y, y) \quad \forall y : A_r^T y = 0. \tag{8}$$

Expressions (7) and (8) comprise the second-order optimality conditions and the primal and dual nondegeneracy for the QP problem (2).

From the definition of $F(x, u, k)$ and in view of (5)–(8), we obtain for $k > 0$ the following MBF properties:

(P1) $F(x^*, u^*, k) = f_0(x^*)$.

(P2) $\nabla_x F(x^*, u^*, k) = Qx^* - a - Au^* = Qx^* - a - A_r u_r^* = 0$.

(P3) $\nabla_{xx}^2 F(x^*, u^*, k) = Q + kA_r U_r^* A_r^T$, where $U_r^* = [\text{diag}(u_i^*)]_{i=1}^r$.

(P4) There exists a $k_0 > 0$, such that for any fixed $k \geq k_0$, one has:

mineigenval $\nabla_{xx}^2 F(x^*, u^*, k) = $ mineigenval $\{Q + kA_r U_r^* A_r^T\} \geq \lambda > 0$;
maxeigenval $\nabla_{xx}^2 F(x^*, u^*, k) = $ maxeigenval $\{Q + kA_r U_r^* A_r^T\} \leq \Lambda < \infty$.

(P5) There exists a $k_0 > 0$, such that for any fixed $k \geq k_0$, $x^* = \arg\min\{F(x, u^*, k) \mid x \in \mathbb{R}^n\}$.

Here, $[\text{diag}(w_i)]_{i=1}^p$ stands for the $p \times p$ diagonal matrix with the (i, i)th element equal to w_i. Due to (7), we have that $\theta^* = \min\{u_i^* \mid i \in I\} > 0$ and, due to (1), $\rho^* = \max\{u_i^* \mid i \in I\} < \infty$. We will assume that $\sigma = \min\{r_i(x^*) \mid i \in J\} > 0$.

3. Basic MBF theorem for the QP problem

We will reformulate the basic theorem from [24] (see also [22,23]) for the QP problem.

Given ε, δ, $k_0 > 0$, $u^* \in \mathbb{R}^m_{++}$ and $0 \le \varepsilon \le \min\{u_i^* \mid i \in I\}$, we define for $i \in I$:

$$D_i(\cdot) = D_i(u^*, k_0, \delta, \varepsilon) = \{(u_i, k) \in \mathbb{R}^2 \mid u_i \ge \varepsilon, |u_i - u_i^*| \le \delta k,\ k \ge k_0\},$$

and for $i \in J$:

$$D_i(\cdot) = D_i(u^*, k_0, \delta, \varepsilon) = \{(u_i, k) \in \mathbb{R}^2 \mid 0 \le u_i \le \delta k,\ k \ge k_0\}.$$

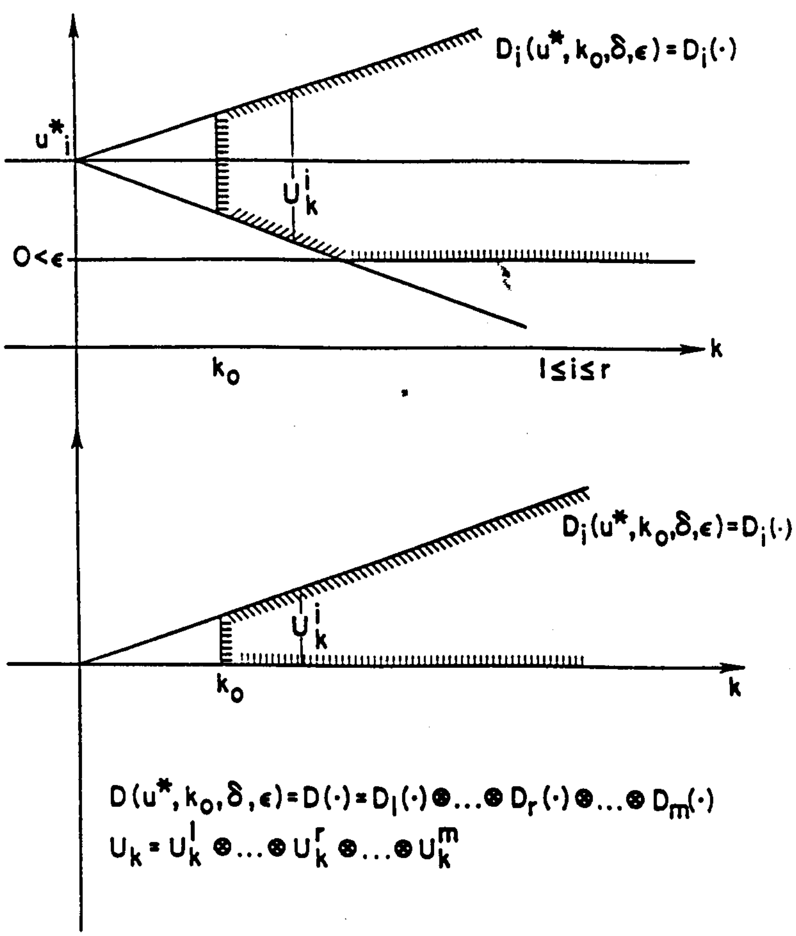

Figure 1. The sets $D_i(u^*, k_0, \delta, \varepsilon)$ and U_k^i.

The sets $D_i(\cdot)$ are in fact cut cones and are represented in figure 1. The set $D(\cdot)$ is defined as the direct product of the sets $D_i(\cdot)$:

$$D(\cdot) = D_1(\cdot) \otimes D_2(\cdot) \otimes \ldots \otimes D_m(\cdot).$$

We also consider the set $U_k = U_k^1 \otimes U_k^2 \otimes \ldots \otimes U_k^m$ (see figure 1), where, for $i = 1,\ldots,r$:

$$U_k^i = \{u_i : u_i^* - \delta k \leq u_i \leq u_i^* + \delta k,\ k \geq k_0 \text{ is fixed}\}$$

and for $i = r+1,\ldots,m$:

$$U_k^i = \{u_i : 0 \leq u_i \leq \delta k,\ k \geq k_0 \text{ is fixed}\}.$$

The set U_k will also be denoted later on by U_μ, with $m = 1/k$.

We define, whenever they exist:

$$\hat{x}(u,k) = \arg\min\{F(x,u,k) \mid x \in \mathbb{R}^n\}, \tag{9}$$

$$\hat{u}(u,k) = [\operatorname{diag}(kr_i(\hat{x}) + 1^{-1}]_{i=1}^m u. \tag{10}$$

We are now ready for the "Basic Theorem", which is a restatement of a part of the Basic Theorem in [24].

THEOREM 3.1

(1) If $\Omega^* = \{x \in \Omega \mid f(x) = f(x^*)\}$ is compact, then for any $(u,k) \in \mathbb{R}_{++}^m \times \mathbb{R}_{++}$, there exist $\hat{x} = \hat{x}(u,k)$ and $\hat{u} = \hat{u}(u,k)$ such that $\nabla_x F(\hat{x},u,k) = 0$, that is,

$$Q\hat{x} - a = \sum_{i=1}^{m} \frac{u_i}{kr_i(\hat{x}) + 1} a_i = A\hat{u}.$$

(2) $\hat{x}(u^*,k) = x^*$ and $\hat{u}(u^*,k) = u^*$ for any $k > 0$, i.e. u^* is a fixed point of the map $u \to \hat{u}(u,k)$.

(3) If (5)–(8) are satisfied, then there exist $\delta > 0$, $\varepsilon > 0$ and $k_0 > 0$ such that for any $(u,k) \in D(\cdot)$, there exist vectors $\hat{x} = \hat{x}(u,k) \in \Omega_k$ and $\hat{u} = \hat{u}(u,k)$ such that $\nabla_x L(\hat{x},\hat{u}) = 0$ and

$$\max\{\|\hat{x} - x^*\|_\infty, \|\hat{u} - u^*\|_\infty\} \leq \frac{C}{k}\|u - u^*\|_\infty \leq \gamma\|u - u^*\|_\infty, \tag{11}$$

with the constant C independent of $(u,k) \in D(\cdot)$.

(4) For any fixed $k \geq k_0$ and any pair $(\hat{x},u) \in \Omega_k \times U_k$, there exist $\tilde{\lambda}$ and $\tilde{\Lambda}$ such that:

$$\text{mineigenval}(\nabla_{xx}^2 F(\hat{x},u,k)) \geq \tilde{\lambda} > 0, \tag{12}$$

$$\text{maxeigenval}(\nabla_{xx}^2 F(\hat{x},u,k)) \leq \tilde{\Lambda} < \infty. \tag{13}$$

□

The important result, which will be used throughout this paper, is the expression for the rate of convergence (11). It states that for any $(u, k) \in D(\cdot)$, the new minimizer $\hat{x} = \hat{x}(u,k)$, as well as the new Lagrange multipliers $\hat{u} = \hat{u}(u,k)$, are closer to the solution by a factor of $\gamma = C/k$. Therefore, if $\gamma < 1$, then $u \in U_k \Rightarrow \hat{u} \in U_k$. In order to characterize the value of C, we first introduce two parameters, which depend on $k > 0$:

$$\alpha_k = \max\{(r_i(x^*) + k^{-1})^{-1} | i = r+1, \dots, m\} \quad \text{and} \quad \beta_k = \|\Phi_k^{-1} R_k\|_\infty,$$

where the matrices Φ_k and R_k (see [24]) are given by:

$$\Phi_k = \begin{pmatrix} Q & -A_r \\ -U_r^* A_r^T & -k^{-1} I_r \end{pmatrix}, \quad R_k = \begin{pmatrix} 0^{(n,r)} & -A_{m-r}[\mathrm{diag}(r_i(x^*) + k^{-1})^{-1}]_{r+1}^m \\ -I_r & 0^{(r,m-r)} \end{pmatrix},$$

and $Q = \nabla_{xx}^2 L(x^*, u^*)$, $U_r^* = [\mathrm{diag}(u_i^*)]_{i=1}^r$. Furthermore, I_p is the $p \times p$ identity matrix and $0^{(p,s)}$ is the $p \times s$ null matrix. We shall use this notation throughout the paper.

The following assertion is a consequence of the Basic Theorem in [24].

ASSERTION 3.1

If the second-order optimality conditions (7)–(8) are satisfied, then there exists a $k_0 > 0$ such that for any $k \geq k_0$ the following inequalities hold:

$$\alpha_k \leq \sigma^{-1} \quad \text{and} \quad \beta_k \leq c_0,$$

where σ and c_0 are independent of $k \geq k_0$.

An upper bound on c_0 is computed in appendix A.

DEFINITION 3.1

The condition of the QP problem (2) is defined as

$$C = \max\{\sigma^{-1}, c_0\}.$$

We would like to emphasize that the condition number C not only does not depend on $k \geq k_0$, it is also invariant to scaling of the input data ($\|a_i\| = 1$ for all $i = 1, \dots, m$). It can also be made independent of the monotone transformation, which is used to transform the initial problem to an equivalent one (in our case, the transformation is given by $\ln(1 + t)$). The condition number depends on the input data and the size of the problem. It is an important component in our further analysis. In particular, the value C is critical for the complexity bound of the Newton MBF methods. The larger it is, the larger k has to be to achieve a given rate of convergence, and the longer one has to follow the shifted barrier trajectory before reaching the "hot start".

An upper bound on C in terms of the QP problem parameters can be found in appendix A.

We conclude this section with the following optimality criterion by considering a "merit function" $\nu(x, u, k)$:

$$\nu(x,u,k) = \max\left\{ -\min_{1\le i\le m} r_i(x), \|\nabla_x F(x,u,k\|, \sum_{i=1}^m u_i |r_i(x)| \right\} \ge 0.$$

It is easy to see that for any $k > 0$:

$$\nu(x,u,k) = 0 \text{ iff } x = x^* \text{ and } u = u^*. \tag{14}$$

Also, there exists an $L > 0$ such that for any $k > 0$,

$$\nu(w,k) - \nu(w^*,k) \le L\|w - w^*\|, \tag{15}$$

with $\nu(w, k) \equiv \nu(x, u, k)$. From (15) we have that for any sequence $\{w^s\}$, converging to w^*, the rate of convergence of the merit function to zero can be estimated by the rate of convergence of $w^s = (x^s, u^s)$ to $w^* = (x^*, u^*)$.

4. The modified barrier method for QP

First, we will rewrite the MBF function

$$F(x,u,k) = \tfrac{1}{2} x^T Q x - a^T x - \frac{1}{k} \sum_{i=1}^m u_i \ln(k r_i(x) + 1)$$

as follows:

$$\tfrac{1}{2} x^T Q x - a^T x - \frac{1}{k} \sum_{i=1}^m u_i \left(\ln\left(r_i(x) + \frac{1}{k}\right) + \ln k \right). \tag{16}$$

Setting $\mu = 1/k$, we obtain $\Omega_k = \Omega_m = \{x \mid r_i(x) \ge -\mu\}$, and (16) becomes:

$$\tfrac{1}{2} x^T Q x - a^T x - \mu \sum_{i=1}^m u_i \ln(r_i(x) + \mu) + \mu \ln \mu \sum_{i=1}^m u_i .$$

Dividing this last expression by μ and dropping the last term, we formulate the following definition:

DEFINITION 4.1

The Modified Barrier Function for QP is defined as

$$\phi(x,u,\mu) \triangleq \frac{1}{\mu}\left(\tfrac{1}{2}x^T Qx - a^T x\right) - \sum_{i=1}^{m} u_i \ln(r_i(x) + \mu).$$

The reason for this definition is one of convenience only. It makes no difference whether we minimize the original MBF or $\phi(x, u, \mu)$ with respect to x, since they differ only by a constant and a factor, both independent of x. The update formula for the Lagrange multipliers remains of course the same, and we shall therefore obtain the same iterates as with the original MBF. Therefore, all previous results continue to hold for $\phi(x, u, \mu)$.

The first and second order derivatives of $\phi(x, u, \mu)$ with respect to x are given by

$$\nabla_x \phi(x,u,\mu) = g(x,u,\mu) = g(\cdot) = \frac{Qx-a}{\mu} - \sum_{i=1}^{m} u_i \frac{a_i}{r_i(x)+\mu},$$

$$\nabla_{xx}^2 \phi(x,u,\mu) = H(x,u,\mu) = H(\cdot) = \frac{Q}{\mu} + \sum_{i=1}^{m} u_i \frac{a_i a_i^T}{(r_i(x)+\mu)^2}.$$

The following assertion holds:

ASSERTION 4.1

(1) The MBF Hessian $H(x, u, \mu)$ is positive definite in x for any $u \in \mathbb{R}^m_{++}$ and $\mu > 0$ if:

(a) Q is positive definite

or

(b) Q positive semidefinite and Ω is bounded.

(2) If the second-order optimality conditions are fulfilled, then the MBF Hessian $H(x, u, \mu)$ is positive definite in a neighborhood of $\hat{x} = \hat{x}(u, \mu)$ for any $(u, \mu) \in D(\cdot)$.

Proof

Part (1a) of the assertion is obvious, part (1b) is proved in [13], and part (2) is a consequence of the Debreu theorem (see [20]) and the basic theorem in [24]. □

Let us now consider the dual problem for (2) (Wolfe duality, see [33]):

$$w^* \equiv (x^*, u^*) = \arg\max\{L(w) \mid w \in W\}, \tag{17}$$

where

$$W = \{w = (x,u) \mid \nabla_x L(x,u) = Qx - a - Au = 0, u \in \mathbb{R}^m_+\}.$$

Let $x^0 \in int\,\Omega$ and $\Omega^0 = \{x \in \Omega \mid f_0(x) \le f_0(x^0)\}$. In view of (1), the existence of x^* leads to the existence of $u^* \in \mathbb{R}^m_+$ such that $f_0(x^*) = L(w^*)$. We will now describe the MBF method for the simultaneous solution of problems (2) and (17).

To start, we choose $\mu > 0$, $x^0 \in \Omega_\mu$ and $u^0 = (1,\dots,1) \in \mathbb{R}^m$. Let $x^s \in \Omega_\mu$ and $u^s \in \mathbb{R}^m_{++}$ have been found already. The next approximation x^{s+1} and u^{s+1} are computed by:

$$x^{s+1} = \arg\min\{\phi(x, u^s, \mu) \mid x \in \mathbb{R}^n\}, \tag{18}$$

$$u^{s+1} = \mu\Big[\operatorname{diag}\big(r_i(x^{s+1}) + \mu\big)^{-1}\Big]_{i=1}^m u^s. \tag{19}$$

The following theorem holds [15]:

THEOREM 4.1

If Ω^0 and W are nonempty and bounded, then for any $\mu > 0$, we have:

(1) $f_0(x^*) > \dots > L(w^{s+1}) > L(w^s) > \dots > L(w^0)$;

(2) $\lim_{s\to\infty} \phi(x^s, u^s, \mu) = \lim_{s\to\infty} f_0(x^s) = \lim_{s\to\infty} L(w^s) = f_0(x^*)$;

(3) $\max\{u_i^*, r_i(x^*)\} > 0$, $i = 1,\dots,m$, i.e., strict complementarity holds;

(4) $\sum_{i=1}^m u_i^s r_i(x^s) = O((\mu\alpha_s)^{1/2})$, where $\lim_{s\to\infty} \alpha_s = 0$. □

In other words, the MBF method generates primal and dual sequences, which converge, respectively, to the primal and dual solutions in value (recently, Polyak and Teboulle (see [26]) proved that the dual sequence $\{u^s\}$ itself converges to u^*), for any positive barrier parameter μ, whether the primal and dual problems have unique solutions or not.

The method (18)–(19) requires an infinite number of operations at each step. To make this method executable, we have to change the infinite procedure of finding the primal minimizer by a finite procedure, while keeping estimation (11). Such a modification can be carried out in the following way. Let us consider a positive number τ and a pair $(\bar{x}, \bar{u}) \in \mathbb{R}^n \times \mathbb{R}^m_{++}$, with

$$\|\nabla_x \phi(\bar{x}, u, \mu)\| \le \tau\mu \|\mu[\operatorname{diag}(r_i(\bar{x}) + \mu)^{-1}]u - u\|, \tag{20}$$

$$\bar{u} = \mu[\operatorname{diag}(r_i(\bar{x}) + \mu)^{-1}]u. \tag{21}$$

If conditions (7)–(8) are satisfied, then for any $\tau > 0$ there exists $\mu > 0$ such that for any $(u, \mu) \in D(\cdot)$ the following estimate holds:

$$\max\{\|\bar{x} - x^*\|, \|\bar{u} - u\|\} \le C(1 + \tau)\mu\|u - u^*\|. \tag{22}$$

For a proof of this result, see [24, p. 206]. To find $\bar{x}$ requires a finite number of Newton steps.

Throughout the paper, we will assume that (x, u) belongs to the cut cone $D(\cdot)$ as is stated in the Basic Theorem. We also assume that μ is small enough so that $0 < \gamma = C\mu < 1$. This ratio can be made as small as one wants by decreasing μ, and, as was explained at the end of the previous section, can be estimated by checking the merit function $\nu(x, u, \mu)$. We have the following definition:

DEFINITION 4.2

The barrier parameter μ is consistent with a given ratio $0 < \gamma < 1$ if $\{\gamma^s\}$ is a majorant for the sequence $\{\nu(w^s, \mu)\}$, i.e., $\{\nu(w^s, \mu)\} \leq \gamma^s$.

This latter inequality can be verified explicitly after every Lagrange multiplier update. Therefore, the merit function ν establishes a feedback between the observed rate of convergence and the barrier parameter value, allowing its proper choice.

We conclude this section with some inequalities for MBF parameters. We shall prove a lemma, which will provide a lower bound on the Lagrange multipliers in terms of certain parameters of the MBF method. The lemma will then be used to derive two inequalities with those same parameters, one involving θ^* and the other involving σ. These inequalities will be needed further on.

Since $\|a_i\| = 1$ for all i, we have:

$$\forall x, y \in \Omega_\mu, \quad \forall i : |r_i(x) - r_i(y)| \leq \|x - y\|.$$

Recalling that Ω_μ denotes the extended feasible region and rewriting the update formula for the Lagrange multipliers as

$$\hat{u}_i = \frac{\mu u_i}{r_i(\hat{x}) + \mu},$$

where $\mu = 1/\mathrm{k}$, the following lemma holds:

LEMMA 4.1

Let $\|u^{(0)} - u^*\|_\infty \leq \omega$ for some $\omega > 0$ and let $0 < \gamma < 1$ be the rate of convergence, associated with $\mu > 0$, as in the basic convergence result (11).

Then the following inequalities hold:

(1) For the active constraints ($i \in I$):

$$u_i^{(s)} \geq \exp\left(-\left(\frac{1-\gamma^s}{1-\gamma}\right) \frac{\gamma\omega\sqrt{n}}{\mu} \right) u_i^{(0)}. \tag{23}$$

(2) For the passive constraints ($i \in J$):

$$u_i^{(s)} \geq \exp\left(- \frac{s\sigma_i}{\mu}\right)\exp\left(- \left(\frac{1-\gamma^s}{1-\gamma}\right) \frac{\gamma\omega\sqrt{n}}{\mu} \right)u_i^{(0)}, \tag{24}$$

with $\sigma_i = r_i(x^*)$.

Proof

The proof of this lemma is given in appendix A. □

We now use this lemma to prove the following theorem.

THEOREM 4.2

Let $0 < \gamma = C\mu < 1$ be the ratio of convergence associated with μ in estimation (11), then the following two inequalities hold:

$$\mu \leq \left(\frac{\gamma\sqrt{n}}{1-\gamma}\right)\theta^*, \tag{25}$$

$$\mu \leq \frac{\sigma}{|\ln \gamma|}. \tag{26}$$

Proof

(1) First we take a look at the active constraints. In order to apply the previous lemma for given ω, we choose $u^{(0)}$ as follows:

$$u_i^{(0)} = u_i^* + \omega \quad i = 1,\dots,m. \tag{27}$$

Then, taking the limit for $s \to \infty$ in (23), we obtain:

$$\lim_{s\to\infty} u_i^{(s)} \geq \lim_{s\to\infty} \exp\left(- \left(\frac{1-\gamma^s}{1-\gamma}\right) \frac{\gamma\omega\sqrt{n}}{\mu}\right)u_i^{(0)}.$$

Since we assumed that the $u_i^{(s)}$'s converge to the optimal Lagrange multipliers, this means that

$$u_i^* \geq \exp\left(- \frac{\gamma\omega\sqrt{n}}{\mu(1-\gamma)}\right)u_i^{(0)}. \tag{28}$$

We now choose an index $\bar{i}$ for which $u_{\bar{i}}^{(s)}$ converges to θ^*. For this index, (28) becomes:

$$\theta^* \geq \exp\left(- \frac{\gamma\omega\sqrt{n}}{\mu(1-\gamma)}\right)(\theta^* + \omega),$$

and therefore

$$\theta^* \geq \frac{\omega \exp\left(-\frac{\gamma\omega\sqrt{n}}{\mu(1-\gamma)}\right)}{1-\exp\left(-\frac{\gamma\omega\sqrt{n}}{\mu(1-\gamma)}\right)}$$

or

$$\theta^* \geq \frac{\omega}{\exp\left(\frac{\gamma\omega\sqrt{n}}{\mu(1-\gamma)}\right)-1}.$$

Taking the limit for $\omega \to 0$ on both sides of the previous inequality yields:

$$\theta^* \geq \lim_{\omega\to 0} \frac{\omega}{\exp\left(\frac{\gamma\omega\sqrt{n}}{\mu(1-\gamma)}\right)-1} = \frac{\mu(1-\gamma)}{\gamma\sqrt{n}}.$$

The first part of the theorem follows immediately.

(2) Turning to the passive constraints, our choice for $u^{(0)}$ is

$$u_i^{(0)} = \omega \quad i \in J.$$

We now choose an index $\bar{J}$ for which $\sigma_{\bar{J}} = \sigma = \min\{\sigma_i | i \in J\}$. The basic convergence result in theorem 3.1 then gives

$$u_{\bar{J}}^{(s)} \leq \gamma^s u_{\bar{J}}^{(0)} = \gamma^s \omega.$$

With the previous lemma, we can therefore write:

$$\gamma^s \omega \geq u_{\bar{J}}^{(s)} \geq \exp\left(-\frac{s\sigma_{\bar{J}}}{\mu}\right)\exp\left(-\left(\frac{1-\gamma^s}{1-\gamma}\right)\frac{\gamma\omega\sqrt{n}}{\mu}\right)\omega.$$

This gives

$$\gamma \geq \exp\left(-\frac{\sigma}{\mu}\right)\exp\left(-\frac{1}{s}\left(\frac{1-\gamma^s}{1-\gamma}\right)\frac{\gamma\omega\sqrt{n}}{\mu}\right).$$

Taking the limit on both sides of this latest inequality for $s \to \infty$ yields: $\gamma \geq \exp(-\sigma/\mu)$. This completes the proof. □

5. The Newton MBF method

The Newton MBF method has a preliminary and a general phase. The preliminary phase is similar to that in interior point methods (see [8, 12]). We find a "warm start" for the shifted barrier function $\phi(x, e, \mu_0)$, $e \in \mathbb{R}^m$ with a fixed μ_0, i.e., we find an approximation x^0, from which the Newton method for the system $\nabla_x \phi(x, e, \mu_0) = 0$ converges quadratically.

The general phase has two parts, which might alternate up to some point. In the first part, we will follow the shifted barrier trajectory by changing "carefully" the barrier parameter followed by a Newton step. In the second part, we will use the relaxation operator

$$R : \Omega_\mu \to \Omega_\mu : Rx = \bar{x},$$

where $\bar{x}$ is defined by (20), followed by the Lagrange multipliers update (21), while the barrier parameter is fixed. The efficient use of the relaxation parameter R is possible only in the case where the current value of the barrier parameter μ is *consistent* with the chosen ratio $0 < \gamma < 1$. In such a case, due to (22), the primal sequence $\{x^{(s)}\}$:

$$x^{(s)} = Rx^{(s-1)}$$

and the corresponding sequence of the Lagrange multipliers $\{u^{(s)}\}$:

$$u^{(s)} = \mu[\text{diag}(r_i(x^{(s)}) + \mu)^{-1}]u^{(s-1)}$$

converges to the K–K–T pair (x^*, u^*) linearly with the given ratio $0 < \gamma < 1$, i.e., the process is following the MBF trajectory fast enough. If μ is *inconsistent* with $0 < \gamma < 1$, then estimate (22) is not true and one has to find a mechanism that decreases μ without leaving the Newton area.

Before describing such a method, we estimate the computational cost for the numerical realization of the relaxation operator R. Since R is always applied to x, which is "well defined" (see [31]), it takes $\mathcal{O}(\ln \ln \varepsilon^{-1})$ Newton steps to solve the system in x: $\nabla_x \phi(x, u, \mu) = 0$ with accuracy $\varepsilon > 0$. We assume that:

(1) the accuracy $\varepsilon = 2^{-L}$ (L is the input length of the problem) is enough to find $\bar{x}$, which satisfies (20), starting from any "warm start" for the system $\nabla_x \phi(x, u, \mu) = 0$, when $u \in U_\mu$ and $\mu > 0$ is consistent with $0 < \gamma < 1$.

(2) m ($m > n$) is an upper bound on the number of bits needed to represent any number appearing in the input data.

Under these assumptions, we have $L = \mathcal{O}(m^3)$ and the numerical realization of the operator R requires at most $\mathcal{O}(\ln L) = \mathcal{O}(\ln m)$ Newton steps.

The Newton MBF method will produce a combination of the path-following trajectory for the shifted barrier function (μ is decreasing while the Lagrange multipliers are fixed: $u^0 = e$) and the MBF trajectory (μ is fixed while the Lagrange multipliers are updated, followed by an application of the R-operator to the primal vector). A key element in the method is the merit function $\nu(w, \mu)$. As soon as the barrier parameter μ becomes consistent with the rate of convergence γ, the relaxation inequality $\nu(w^{(s)}, \mu) \le \gamma^s$ holds for any $s \ge 1$, which means that the process is now on the MBF part of the trajectory and will stay on it from now on. If the latter inequality does not hold for some $s > 1$, it indicates that the current μ is inconsistent with the chosen γ and we have to decrease μ. To achieve this without leaving the Newton area, we will turn again to the shifted barrier trajectory and will follow it by "carefully" decreasing the barrier parameter μ. As soon as μ is decreased by a given factor $0 < \kappa < 1$, we will try again to turn to the MBF trajectory. If the new barrier parameter value is consistent with γ, our attempt to turn to the MBF trajectory will be successful, otherwise we will continue to decrease μ by following the shifted barrier trajectory.

Therefore, the Newton MBF method might alternate between the shifted barrier and the MBF trajectories, while always being in the Newton area. As soon as μ becomes consistent with γ, there will be no more alternation between the trajectories. From this point on, only the R-operator will be used, the sequence $\{w^{(s)} = (x^{(s)}, u^{(s)})\}$ will converge to (x^*, u^*) linearly with the rate of convergence γ and the number of Newton steps required for the numerical realization of R will decrease after every Lagrange multiplier update so that from some point on, *one* Newton step is enough for the multiplier update.

The algorithm presented in the flow-chart of figure 2 is theoretical in the sense that even though it correponds to the complexity analysis which is to follow, in practice one might change a few things. In particular, instead of going back to the shifted barrier trajectory when the inequality $\nu(\bar{w}, \mu) \leq \gamma^s$ is not satisfied, one can continue the calculation from the point $\bar{x}$ while changing μ for $\kappa\mu$.

The Newton direction $p(x, u, \mu)$ at a given point x for fixed u and μ is given by:

$$p(x, u, \mu) = -\left(\nabla^2_{xx}\phi(x, u, \mu)\right)^{-1} \nabla_x \phi(x, u, \mu).$$

Any of the three conditions of assertion 4.1 guarantee the existence of $(\nabla^2_{xx}\phi(x, u, \mu))^{-1}$. Whenever there can be no confusion, we shall write p, g, H instead of $p(x, u, \mu)$, $\nabla_x\phi(x, u, \mu)$, $\nabla^2_{xx}\phi(x, u, \mu)$, respectively. We will frequently use the H-norm $\|\cdot\|_H$, defined as

$$\|x\|_H = \sqrt{x^T H x}.$$

The flow-chart on the following page describes the Newton MBF method.

In the rest of the paper, we will prove that for nondegenerate QP, there comes a point where the barrier parameter μ becomes consistent with the chosen ratio γ. As soon as such a point on the shifted barrier trajectory has been reached, the "warm start" for this barrier parameter turns into a "hot start" (see definition 6.1). From this point on, every multiplier update requires at most $O(\ln m)$ Newton steps (see our previous assumptions (1) and (2)) and decreases the distance from the current primal-dual solution to (x^*, u^*) by a factor γ.

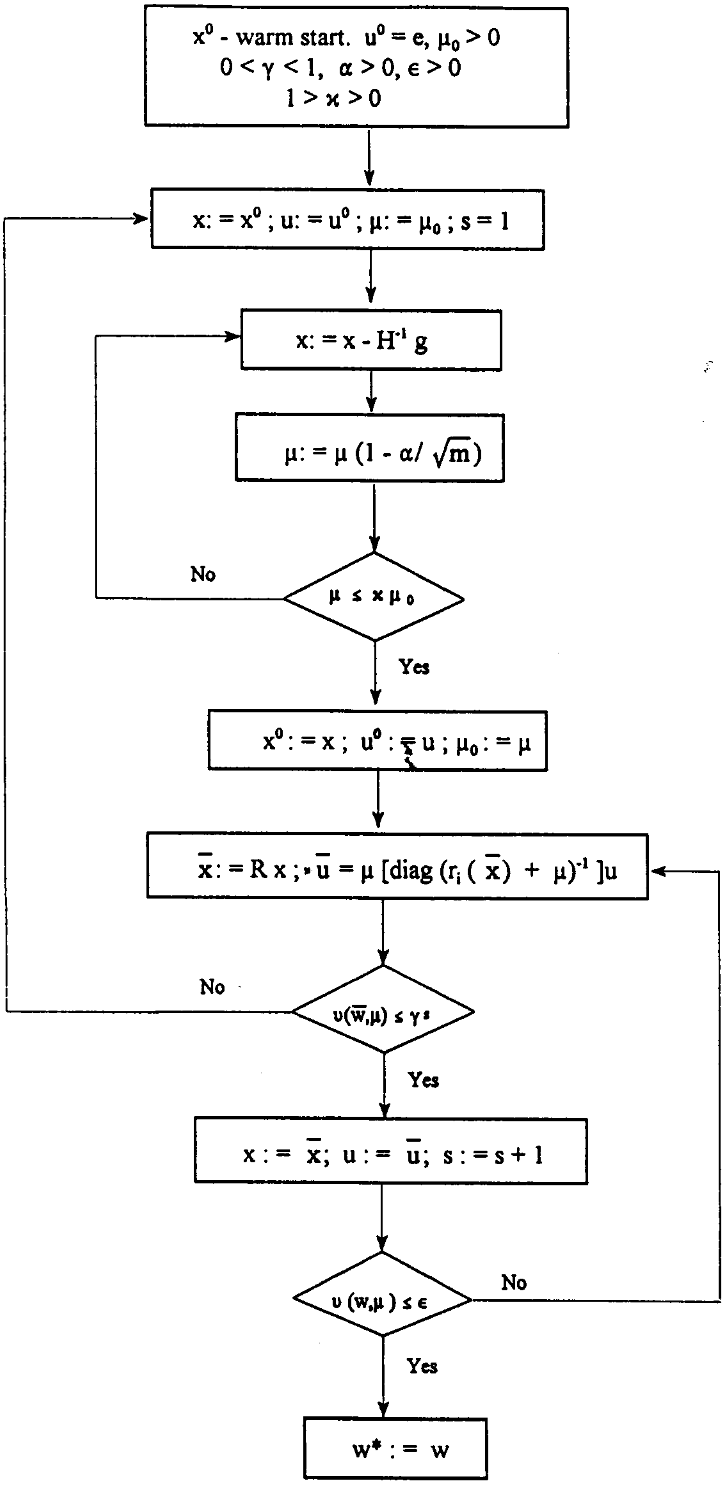

Figure 2. Flowchart of the Newton MBF method.

5.1. BASIC LEMMAS FOR THE INNER ITERATION

In this section, we shall state some basic lemmas that will be needed to determine $\beta(x)$, which was mentioned above. The proof of most of these lemmas is technical and is deferred to appendix A. We start with some notation:

$I \quad = \{i \mid r_i(x) \text{ is active}\}$,
$J \quad = \{i \mid r_i(x) \text{ is passive}\}$,
$I_\mu \quad = \{i \mid u_i \geq \mu/2\}$,
$J_\mu \quad = \{i \mid u_i < \mu/2\}$,
$|J_\mu| = $ the number of elements in J_μ,
$\theta \quad = \min\{u_i \mid i \in I_\mu\}$,
$\rho \quad = \max\{u_i\}$,
$\eta \quad = |J_\mu| \max\{u_i \mid i \in J_\mu\}$,
$\eta_1 \quad = \max\{u_i \mid i \in J_\mu\}$,
$\lambda_H \quad = $ the smallest eigenvalue of H,
$\Omega_\mu \quad = \{x \mid \forall i : r_i(x) + \mu \geq 0\}$.

In the rest of this paper, we will also use the quadratic approximation $q_x(d, u, \mu)$ for $\phi(x + d, u, \mu)$ at the point x, defined as

$$q_x(d, u, \mu) = \phi(x, u, \mu) + g^T d + \tfrac{1}{2} d^T H d.$$

As was mentioned before, we want to determine the region in which full Newton steps can be taken. In order to do this, we have to determine under what conditions, for $x \in int\, \Omega_\mu$ and $d \in \mathbb{R}^n$, $x + d$ still lies in $int\, \Omega_\mu$.

The following two lemmas deal with these issues. In the first one, we determine a condition on d for $x + d$ to lie in $int\, \Omega_\mu$ whenever x does. In this same lemma, we also compute a bound on the error in the quadratic approximation to the MBF, which will be needed for the proof of the next lemma.

LEMMA 5.1

If $x \in int\, \Omega_\mu$ and $\|d\|_H < \beta_1(x)$, then $x + d \in int\, \Omega_\mu$. Moreover, if $\|d\|_H < \beta_1(x)/2$, then

$$|\phi(x + d, u, \mu) - q_x(d, u, \mu)| < \frac{1}{3\xi_1(x)} \|d\|_H^3,$$

where

$$R_1(x) = \min\left\{\sqrt{\lambda_H}\,(r_j(x) + \mu) \mid j \in J_\mu\right\}$$

$$\beta_1(x) = \min\left\{\sqrt{\theta}, R_1(x)\right\}$$

$$\xi_1(x) = \frac{1}{2}\left(\frac{R_1(x)\sqrt{\theta}}{R_1(x) + \frac{\eta}{R_1^2(x)}\sqrt{\theta}}\right).$$

□

The following lemma gives a measure for determining the distance to the minimum for the inner iteration, when we are close to this minimum. See also lemma 2.16 in [13]. We recall that p denotes the Newton direction.

LEMMA 5.2

If

$$\|p\|_H < \beta_2(x) = \min\left\{\frac{R_1(x)}{2}, \frac{\xi_1(x)}{5}\right\},$$

then

$$\|x - \hat{x}(u, \mu)\|_H \leq \tfrac{5}{2}\|p\|_H. \qquad \square$$

As a complement to the previous lemma, we also prove the following result, similar to lemma 5 in [5]. It gives a lower bound on the reduction of the MBF that can be achieved after a linesearch along the Newton direction, when we are far from the minimum.

LEMMA 5.3

If

$$\|p\|_H > \beta_2(x) = \min\left\{\frac{R_1(x)}{2}, \frac{\xi_1(x)}{5}\right\},$$

then the reduction $\Delta\phi$ in the MBF after a linesearch along the Newton direction p, satisfies

$$\Delta\phi > \tfrac{2}{5}\beta_2^2(x).$$

This remains true if $\beta_2(x)$ is replaced by any β, such that $\beta < \beta_2(x)$. $\square$

The next step is to determine under what conditions taking a full Newton step will actually bring us closer to the minimum, and at what rate. In order to do this, we will investigate how successive Newton directions relate to each other. The main result in this respect will be lemma 5.7.

For the proof of this result, we will need the following three lemmas which give bounds on the change in various quantities depending on x when evaluated at different points.

We first define the following quantity:

DEFINITION 5.1

$$B(x) \triangleq \min_i\{r_i(x) + \mu\}.$$

LEMMA 5.4

If

$$\|x - y\| \leq \frac{B(x)}{\alpha} \quad (\alpha > 1), \tag{29}$$

then

$$\left(\frac{\alpha}{\alpha+1}\right)\frac{1}{r_i(x)+\mu} \le \frac{1}{r_i(y)+\mu} \le \left(\frac{\alpha}{\alpha-1}\right)\frac{1}{r_i(x)+\mu}. \qquad \square$$

LEMMA 5.5

If

$$\|x-y\| \le \frac{B(x)}{\alpha} \quad (\alpha > 1), \tag{30}$$

then

$$\left(\frac{\alpha}{\alpha+1}\right)^2 d^T H(x)d \le d^T H(y)d \le \left(\frac{\alpha}{\alpha-1}\right)^2 d^T H(x)d. \tag{31}$$

$\square$

LEMMA 5.6

For $x, x+p \in int\, \Omega_\mu$ with $\|p\| \le B(x)/\alpha$ $(\alpha > 1)$, the following inequality holds:

$$\|\nabla\phi(x+p)\| \le \left(\frac{\alpha^3\sqrt{n}}{(\alpha-1)^3 B(x)}\right)\|p\|_H^2. \qquad \square$$

The next lemma will determine the rate of convergence for the norms of the Newton directions in the inner iteration. The notation "q", which will be used for a Newton direction, should not be confused with the same "q" appearing in the objective function in section 2, as the explicit form of the objective function is not used here at all.

LEMMA 5.7

Let p, q and $H, \overline{H}$ be the Newton directions and Hessians at x and $x+p$, respectively, with $x, x+p \in int\, \Omega_\mu$ with $\|p\| \le B(x)/\alpha$ $(\alpha > 1)$.

Then:

$$\frac{\|q\|_{\overline{H}}}{\|p\|_H^2} \le \frac{\alpha^3\sqrt{n}}{(\alpha-1)^3 B(x)\sqrt{\lambda_{\overline{H}}}}.$$

Proof

We have

$$Hp = -\nabla\phi(x) \quad \text{and} \quad \overline{H}q = -\nabla\phi(x+p).$$

Therefore, with the previous lemma:

$$q^T\overline{H}^2 q = \|\overline{H}q\|^2 = \|\nabla\phi(x+p)\|^2 \le n\left(\frac{\alpha^3}{(\alpha-1)^3 B(x)}\right)^2 \|p\|_H^4.$$

On the other hand, one has

$$q^T \overline{H}^2 q = (\overline{H}^{1/2} q)^T \overline{H} (\overline{H}^{1/2} q) \ge \lambda_{\overline{H}} \| \overline{H}^{1/2} q \|^2 = \lambda_{\overline{H}} \| q \|^2_{\overline{H}}.$$

Therefore, we can write

$$\lambda_{\overline{H}} \| q \|^2_{\overline{H}} \le n \left(\frac{\alpha^3}{(\alpha - 1)^3 B(x)} \right)^2 \| p \|^4_H,$$

which gives

$$\| q \|^2_{\overline{H}} \le \left(\frac{n}{\lambda_{\overline{H}}} \right) \left(\frac{\alpha^3}{(\alpha - 1)^3 B(x)} \right)^2 \| p \|^4_H.$$

Dividing both sides by $\| p \|^4_H$ and taking the square root completes the proof. □

We now define a quantity $\beta_3(x, \alpha, \overline{\alpha})$, depending on x and two positive parameters, α and $\overline{\alpha}$.

DEFINITION 5.2

$$\beta_3(x, \alpha, \overline{\alpha}) \triangleq \min \left\{ \frac{R_1(x)}{2}, \frac{\xi_1(x)}{5}, \frac{1}{2}, \frac{B(x)\sqrt{\lambda_H}}{\alpha}, \left(\frac{\alpha - 1}{\alpha} \right)^3 \frac{B(x)\sqrt{\lambda_H}}{\overline{\alpha}\sqrt{n}} \right\}.$$

We will show that for certain values of α and $\overline{\alpha}$, the following property holds:

If, during the inner iteration (fixed u and μ)

$$\| p(x, u, \mu) \|_{H(x,u,\mu)} \le \beta_3(x, \alpha, \overline{\alpha}), \tag{32}$$

then, if full Newton steps are taken from this point on, the algorithm converges and each new iterate y generated in this way will satisfy:

(1) $\quad y \in int\, \Omega_\mu,$

(2) $\quad \| p(y, u, \mu) \|_{H(y,u,\mu)} \le \beta_3(y, \alpha, \overline{\alpha}).$

Before we can do this, we must find a bound on the change, after one full Newton step, in the quantities $B(x)$, λ_H, $R_1(x)$, and $\xi_1(x)$, which determine $\beta_3(x, \alpha, \overline{\alpha})$. The next lemma will provide this.

We will use the following notation:

$p \quad = p(x, u, \mu),$
$\overline{x} \quad = x + p,$
$q \quad = p(\overline{x}, u, \mu),$
λ_H = the smallest eigenvalue of $H(x, u, \mu)$,
$\lambda_{\overline{H}}$ = the smallest eigenvalue of $H(\overline{x}, u, \mu)$.

LEMMA 5.8

If

$$\|p\|_H \le \frac{B(x)\sqrt{\lambda_H}}{\alpha},$$

then

(1) $$\left(\frac{\alpha-1}{\alpha}\right)B(x) \le B(\bar{x}) \le \left(\frac{\alpha+1}{\alpha}\right)B(x),$$

(2) $$\left(\frac{\alpha}{\alpha+1}\right)^2 \le \frac{\lambda_{\bar{H}}}{\lambda_H} \le \left(\frac{\alpha}{\alpha-1}\right)^2,$$

(3) $$\left(\frac{\alpha-1}{\alpha+1}\right)R_1(x) \le R_1(\bar{x}) \le \left(\frac{\alpha+1}{\alpha-1}\right)R_1(x),$$

(4) $$\left(\frac{\alpha-1}{\alpha+1}\right)^3 \xi_1(x) \le \xi_1(\bar{x}) \le \left(\frac{\alpha+1}{\alpha-1}\right)^3 \xi_1(x).$$

□

We are now ready for the main results of this section. In the following lemma, we will use the previous lemma and lemma 5.4 to find a relation between α and $\bar{\alpha}$ so that (32) will be satisfied at the point $\bar{x}$ also.

LEMMA 5.9

If $\|p\|_H \le \beta_3(x, \alpha, \bar{\alpha})$ with

$$\alpha > 1,$$

$$\bar{\alpha} > \left(\frac{\alpha+1}{\alpha}\right)\left(\frac{\alpha+1}{\alpha-1}\right)^3,$$

then $\|q\|_{\bar{H}} \le \beta_3(\bar{x}, \alpha, \bar{\alpha})$.

Proof

From lemma 5.7, we have that:

$$\begin{aligned}
\|q\|_{\bar{H}} &\le \left(\frac{\alpha}{\alpha-1}\right)^3 \frac{\sqrt{n}}{B(x)\sqrt{\lambda_{\bar{H}}}}\|p\|_H^2 \\
&\le \left(\frac{\alpha}{\alpha-1}\right)^3 \frac{\sqrt{n}}{B(x)\sqrt{\lambda_{\bar{H}}}}\left(\frac{\alpha-1}{\alpha}\right)^3 \frac{B(x)\sqrt{\lambda_H}}{\bar{\alpha}\sqrt{n}}\|p\|_H \\
&\le \frac{1}{\bar{\alpha}}\left(\frac{\lambda_H}{\lambda_{\bar{H}}}\right)^{1/2}\|p\|_H \le \frac{1}{\bar{\alpha}}\left(\frac{\alpha+1}{\alpha}\right)\|p\|_H.
\end{aligned}$$

Now, we assumed that

$$\|p\|_H \leq \beta_3(x, \alpha, \overline{\alpha}) = \min\left\{\frac{R_1(x)}{2}, \frac{\xi_1(x)}{5}, \frac{1}{2}, \frac{B(x)\sqrt{\lambda_H}}{\alpha}, \left(\frac{\alpha-1}{\alpha}\right)^3 \frac{B(x)\sqrt{\lambda_H}}{\overline{\alpha}\sqrt{n}}\right\},$$

so from lemma 5.8 we have

$$R_1(x) \leq \left(\frac{\alpha+1}{\alpha-1}\right) R_1(\overline{x}),$$

$$\xi_1(x) \leq \left(\frac{\alpha+1}{\alpha-1}\right)^3 \xi_1(\overline{x}),$$

$$B(x)\sqrt{\lambda_H} \leq \left(\frac{\alpha+1}{\alpha-1}\right) B(\overline{x})\sqrt{\lambda_{\overline{H}}},$$

and therefore

$$\|p\|_H \leq \left(\frac{\alpha+1}{\alpha-1}\right)^3 \beta_3(\overline{x}, \alpha, \overline{\alpha}).$$

We have obtained that

$$\|q\|_{\overline{H}} \leq \frac{1}{\overline{\alpha}}\left(\frac{\alpha+1}{\alpha}\right)\left(\frac{\alpha+1}{\alpha-1}\right)^3 \beta_3(\overline{x}, \alpha, \overline{\alpha}).$$

The assumption on $\overline{\alpha}$ concludes the proof. □

A convenient choice for α and $\overline{\alpha}$ is 6 and 125/36, respectively. This choice satisfies all conditions imposed on α and $\overline{\alpha}$ and we will use it in our next definition.

DEFINITION 5.3

$$\beta(x) \triangleq \beta_3(x, 6, \tfrac{125}{36}) = \min\left\{\frac{\xi_1(x)}{5}, \frac{1}{2}, \frac{B(x)\sqrt{\lambda_H}}{6\sqrt{n}}\right\}.$$

Note that $R_1(x) \geq \sqrt{\lambda_H} B(x)$, so that it could be left out of the definition.

Substituting those same values for α and $\overline{\alpha}$ in previous lemmas, we have proved the following theorem, which summarizes the results of this section:

THEOREM 5.1

(1) If $\|p\|_H \leq \beta(x)$, then

(i) $\|x - \hat{x}(u, \mu)\|_H \leq \frac{5}{2}\|p\|_H$,

(ii) $\dfrac{\|q\|_{\overline{H}}}{\|p\|_H^2} \le \dfrac{2\sqrt{n}}{B(x)\sqrt{\lambda_{\overline{H}}}}$,

(iii) $\|q\|_{\overline{H}} \le \frac{2}{5}\|p\|_H$.

(2) If $\|p\|_H > \beta(x)$, then $\Delta\phi > \frac{2}{5}\beta^2(x)$. □

5.2. BOUNDS ON ALGORITHM PARAMETERS IN A NEIGHBORHOOD OF THE SOLUTION

So far we have derived results for points that were not necessarily close to the solution. However, since we are ultimately interested in finding a lower bound on $\beta(x)$ in a neighborhood of the solution, we will now consider the algorithm for (x, u) lying in the set S_1, which is defined as follows:

DEFINITION 5.4

$$S_1 \triangleq \left\{(x, u) \mid \max\{\|x - x^*\|, \sqrt{n}\,\|u - u^*\|_\infty\} < \frac{\mu}{2}\right\}.$$

In this section, we will compute bounds on S_1 for previously defined quantities in terms of their values at the solution. These bounds will be used in the next section to prove the final results. The proofs of the first two lemmas are quite technical and they can be found in appendix A. The first lemma gives bounds on the Lagrange multipliers and on the quantity $r_i(x) + \mu$ in S_1.

LEMMA 5.10

Let $(x, u) \in S_1$ and let the rate of convergence of the algorithm be given by $\gamma = 1/(2\sqrt{n})$.

Then

(1) for the active constraints $(i \in I)$:

$$\tfrac{1}{2}u_i^* < u_i < \tfrac{3}{2}u_i^*, \tag{33}$$

and

$$\frac{\mu}{2} < r_i(x) + \mu < \frac{3\mu}{2},$$

(2) for the passive constraints $(i \in J)$:

$$r_i(x) + \mu > \sigma + \frac{\mu}{2}.$$

□

We can use the results from the lemma we just proved to draw the picture in figure 3. It shows that for $(x, u) \in S_1$, the sets I_μ and J_μ are identical to I and J, respectively, since all active Lagrange multipliers will lie to the right of $\theta^*/2$.

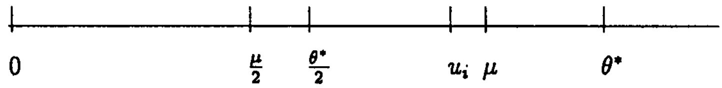

Figure 3. Location of Lagrange multipliers.

We will now have a look at the eigenvalues of the Hessian of the modified barrier function in S_1. The following lemma gives upper and lower bounds for the smallest and largest eigenvalues of the Hessian of the function $\phi(x, u, \mu)$ for fixed u and μ and for $(x, u) \in S_1$.

LEMMA 5.11

For $(x, u) \in S_1$, the smallest and largest eigenvalues of the Hessian of $\phi(x, u, \mu)$ at x for fixed u and μ are bounded as follows:

$$\tfrac{2}{9}\lambda^* \le \lambda_H \le 6\lambda^* + \frac{\mu(m-r)}{2\sqrt{n}(\sigma+\mu/2)^2},$$

$$\tfrac{2}{9}\Lambda^* \le \Lambda_H \le 6\Lambda^* + \frac{\mu(m-r)}{2\sqrt{n}(\sigma+\mu/2)^2}.$$

□

We now define the following two quantities:

DEFINITION 5.5

$$\tilde{\lambda} \triangleq \tfrac{2}{9}\lambda^*,$$

$$\tilde{\Lambda} \triangleq 6\Lambda^* + \frac{\mu(m-r)}{2\sqrt{n}(\sigma+\mu/2)^2}.$$

Here, λ^* and Λ^* are the smallest and largest eigenvalues of $\nabla^2_{xx}\phi(x^*, u^*, \mu)$, respectively (for fixed μ).

The following lemma gives a lower bound on $\beta(x)$ in S_1, but first we will define the following quantities:

DEFINITION 5.6

Definition of ξ^* and β^*:

$$\xi^* \triangleq \frac{1}{10}\left(\frac{\sqrt{2}}{\sqrt{\theta^*}} + \frac{27(m-r)\mu}{4\sqrt{2n}\lambda^{*3/2}(\sigma+\mu/2)^3}\right)^{-1},$$

$$\beta^* \triangleq \min\left\{\frac{1}{2}, \xi^*, \frac{\mu\sqrt{\lambda^*}}{18\sqrt{2}\sqrt{n}}\right\}.$$

LEMMA 5.12

For $(x, u) \in S_1$, $\beta(x) \geq \beta^*$.

Proof

From the definition of $B(x)$ and lemmas 5.10 and 5.11, we have immediately that

$$B(x) \geq \frac{\mu}{2} \quad \text{and} \quad B(x)\sqrt{\lambda_H} \geq \frac{\mu\sqrt{\lambda^*}}{3\sqrt{2}}.$$

Since $J_\mu = J$, we have for $R_1(x)$ with lemmas 5.10 and 5.11:

$$R_1(x) \geq \frac{\sqrt{2\lambda^*}}{3}\left(\sigma + \frac{\mu}{2}\right),$$

whereas for η we can write:

$$\eta \leq \frac{(m-r)\mu}{2\sqrt{n}}.$$

We now use these bounds and lemma 5.10 to compute the lower bound on $\xi_1(x)$ in S_1:

$$\xi_1(x) = \frac{1}{2}\left(\frac{1}{\sqrt{\theta}} + \frac{\eta}{R_1^3(x)}\right)^{-1} \geq \frac{1}{2}\left(\frac{\sqrt{2}}{\sqrt{\theta^*}} + \frac{27(m-r)\mu}{4\sqrt{2n}\lambda^{*3/2}(\sigma + \mu/2)^3}\right)^{-1}.$$

From these bounds and from the definition of $\beta(x)$, it follows that

$$\beta(x) \geq \beta^*.$$

□

5.3 BEHAVIOR OF THE ALGORITHM IN A NEIGHBORHOOD OF THE SOLUTION AND FINAL RESULT

In this last section, we consider the behavior of the algorithm in a subset of S_1 and present the final results. The subset $S \subset S_1$ we will look at is defined as follows:

DEFINITION 5.7

$$S \triangleq \left\{(x,u) : \max\{\|x - x^*\|_\infty, \|u - u^*\|_\infty\} < \frac{\mu}{2\sqrt{n}}\right\}.$$

Starting from a point $(x^{(0)}, u^{(0)})$ in S, let us now examine the iterates $(x^{(s)}, u^{(s)})$. We will use the following notation:

- $\phi_{s-1} = \phi(x, u^{(s-1)}, \mu)$.
- $x^{(s)}$ is the sth iterate, i.e., the approximate minimum of ϕ_{s-1}.
- $\hat{x}^{(s)}$ is the exact minimum of ϕ_{s-1}.
- $H^{(s)}(x)$ is the Hessian of ϕ_s at the point x.
- $p_s(x)$ is the Newton direction of ϕ_s at the point x.
- $p(x)$ is the Newton direction of $\phi(x, u, \mu)$ at the point x for fixed u.
- $\lambda_s(x)$ is the smallest eigenvalue of H^s at the point x.
- $\lambda(x)$ is the smallest eigenvalue of $H(x)$.

Figure 4 illustrates the labelling of the iterates.

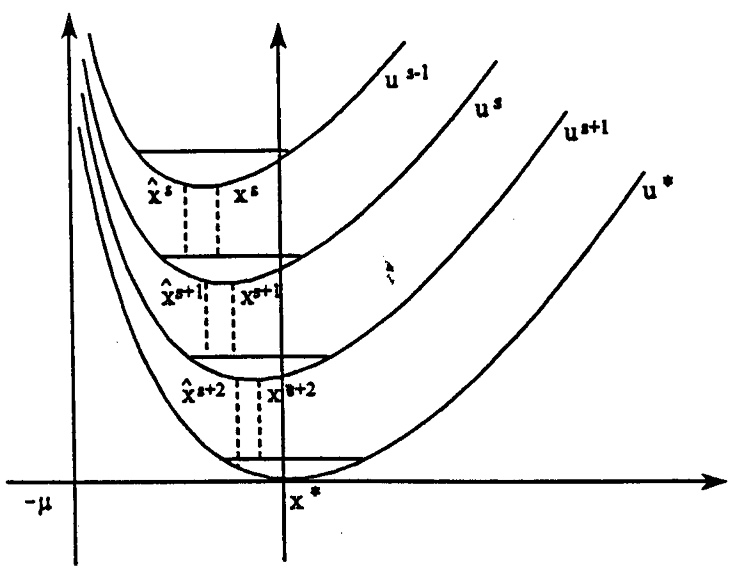

Figure 4. Labelling of the iterates.

Under the assumption that the rate of convergence of the algorithm is given by $\gamma = 1/(2\sqrt{n})$ and that $(x^{(0)}, u^{(0)}) \in S$, all iterates will lie in S and they satisfy:

$$\max\{\|x^{(s)} - x^*\|_\infty, \|u^{(s)} - u^*\|_\infty\} < \gamma^s \omega.$$

We now recall that, starting with $x^{(0)}$, the algorithm first checks whether this point is in the region of quadratic convergence for ϕ_0. If not, we perform a linesearch and continue to do so until this region is reached, from which point on full Newton steps are taken until a point close enough to the minimum is reached and accepted as the new iterate $x^{(1)}$. This point is then used to update the Lagrange multipliers and construct ϕ_1. We then, again, check if this point lies in the region of quadratic convergence for ϕ_1 and so on.

The next lemma is the last one we need to prove our main results. Its purpose is to provide a few bounds, which will be used to determine if there is an $\bar{s}$ for which $x^{(\bar{s})}$ lies in the region of quadratic convergence for Newton's method for $\phi_{\bar{s}}$ and whether this will remain so for subsequently generated iterates.

LEMMA 5.13

Let the rate of convergence of the algorithm be given by $\gamma = 1/(2\sqrt{n})$ and let $(x^{(s)}, u^{(s)}) \in S$. Assume that each inner minimization is carried out with accuracy $\varepsilon > 0$. Then we have:

(1) $$\| x^{(s)} - \hat{x}^{(s+1)} \| \le 2\gamma^s \omega \sqrt{n},$$

(2) $$\| p_s(x^{(s)}) \|_{H^s(x^{(s)})} \le \frac{2\tilde{\Lambda}}{\sqrt{\tilde{\lambda}}} \gamma^s \omega \sqrt{n}.$$

Proof

(1) We begin with the first part.

$$\begin{aligned} \| x^{(s)} - \hat{x}^{(s+1)} \| &\le \| x^{(s)} - x^* \| + \| \hat{x}^{(s+1)} - x^* \| \\ &\le \| x^{(s)} - x^* \| + \| x^{(s+1)} - x^* \| + \| x^{(s+1)} - \hat{x}^{(s+1)} \| \\ &\le \gamma^s \omega \sqrt{n} + \gamma^{s+1} \omega \sqrt{n} + \frac{\varepsilon}{4} \le 2\gamma^s \omega \sqrt{n}, \end{aligned}$$

where we have used the assumption that each minimization is carried out with accuracy $\varepsilon/4$.

(2) The second part follows almost immediately from part (1) and from a standard theorem about convex functions which is stated in appendix A as lemma 8.4. We have:

$$\begin{aligned} \| p_s(x^{(s)}) \|_{H^s(x^{(s)})} &= \left(\left(\nabla \phi_s(x^{(s)}) \right)^T \left[H^s(x^{(s)}) \right]^{-1} H^s(x^{(s)}) \left[H^s(x^{(s)}) \right]^{-1} \nabla \phi_s(x^{(s)}) \right)^{1/2} \\ &\le \left(\frac{1}{\tilde{\lambda}} \| \nabla \phi_s(x^{(s)}) \|^2 \right)^{1/2}. \end{aligned}$$

We can therefore write:

$$\| p_s(x^{(s)}) \|_{H^s(x^{(s)})} \le \frac{1}{\sqrt{\tilde{\lambda}}} \| \nabla \phi_s(x^{(s)}) \| \le \frac{\tilde{\Lambda}}{\sqrt{\tilde{\lambda}}} 2\gamma^s \omega \sqrt{n}.$$

In the first inequality we used the fact that $\phi(x, u, \mu)$ is strongly convex; the latter inequality uses lemmas 5.13 and 8.4.

This completes the proof. □

We are now ready to state the main results.

THEOREM 5.2

When the algorithm reaches a point $(x^{(\bar{s})}, u^{(\bar{s})})$ satisfying

$$\max\{\|x^{(\bar{s})} - x^*\|_\infty, \|u^{(\bar{s})} - u^*\|_\infty\} < \frac{1}{2\sqrt{n}} \min\left\{\mu, \frac{\sqrt{\tilde{\lambda}}}{\tilde{\Lambda}} \beta^*\right\}, \tag{34}$$

and $\varepsilon < (1/(2\sqrt{n}))^{\bar{s}} \mu/(2\sqrt{n})$ (otherwise we have reached the desired accuracy and there is no point in continuing the algorithm), then $x^{(\bar{s})}$ will lie in the region of quadratic convergence for $\phi_{\bar{s}}$ and the same will be true for each subsequently generated pair of primal and dual iterate $(x^{(s)}, u^{(s)})$.

Proof

Suppose we start from some initial point in S, then the iterate $x^{(\bar{s})}$ will certainly lie in the Newton area for $\phi_{\bar{s}}$ if $\|p\|_H$ for this point falls below β^*, which is a lower bound on $\beta(x)$ in S_1. With the previous lemma one sees that this will certainly be true if

$$2\frac{\tilde{\Lambda}}{\sqrt{\tilde{\lambda}}} \gamma^{\bar{s}} \omega \sqrt{n} \le \beta^*$$

or

$$\gamma^{\bar{s}} \omega \le \frac{\sqrt{\tilde{\lambda}}}{2\tilde{\Lambda}\sqrt{n}} \beta^*.$$

This completes the proof. □

Following this theorem we define the following set:

DEFINITION 5.8

$$T \triangleq \left\{(x, u) : \max\{\|x - x^*\|_\infty, \|u - u^*\|_\infty\} < \frac{1}{2\sqrt{n}} \min\left\{\mu, \frac{\sqrt{\tilde{\lambda}}}{\tilde{\Lambda}} \beta^*\right\}\right\}.$$

THEOREM 5.3

Let the algorithm have reached the point $(x^{(\bar{s})}, u^{(\bar{s})})$, satisfying the conditions of the previous theorem and let ε also be as in this theorem. Then from this point on, the convergence of $\|p\|_H$ to zero in any inner iteration (u and μ are fixed) will be quadratic with the rate of convergence given by:

$$\frac{\|q\|_{\bar{H}}}{\|p\|_H^2} \le \frac{1}{3\beta^*}.$$

Proof

When we start from a point (x, u) in T, the iterates obtained in the inner iteration will converge to $\hat{x} = \arg\min_x \phi(x, u, \mu)$, and all of them will satisfy $\|p\|_H \le \beta^*$. This inner iteration terminates with a point $\bar{x}$, satisfying

$$\|\bar{x} - \hat{x}\| \le \varepsilon/4,$$

We therefore also have

$$\|\hat{x} - x^*\| \le \|\bar{x} - \hat{x}\| + \|\bar{x} - x^*\| \le \|\bar{x} - \hat{x}\| + \sqrt{n}\,\|\bar{x} - x^*\|_\infty$$
$$\le \frac{\varepsilon}{4} + \frac{1}{2}\|u - u^*\|_\infty \le \frac{\varepsilon}{4} + \frac{\mu}{4\sqrt{n}}.$$

We now show that any iterate w, obtained during the minimization of $\phi(x, u, \mu)$ starting from $(x, u) \in T$, is such that (w, u) lies in S.

We start by looking at what happens to the first iterate z, obtained by taking a full Newton step from x. We show that (z, u) lies inside S.

$$\|z - \hat{x}\| \le \frac{1}{\sqrt{\lambda(z)}}\|z - \hat{x}\|_{H(z)} \le \frac{5}{2\sqrt{\lambda(z)}}\|p(z)\|_{H(z)}$$
$$\le \frac{5}{2\sqrt{\lambda(z)}}\,\frac{2}{5}\|p(x)\|_{H(x)} \qquad \text{(from theorem 5.1)}$$
$$\le \frac{\sqrt{\lambda(x)}}{\sqrt{\lambda(z)}}\,\frac{1}{\sqrt{\lambda(x)}}\|p(x)\|_{H(x)}$$
$$\le \frac{7}{6}\,\frac{1}{\sqrt{\lambda(x)}}\,\beta^* \qquad \text{(from lemma 5.8 with } \alpha = 6)$$
$$\le \frac{7}{6}\,\frac{1}{\sqrt{\lambda(x)}}\,\frac{\mu\sqrt{\tilde{\lambda}}}{12\sqrt{n}} \le \frac{7}{72\sqrt{n}}\mu.$$

We can therefore write

$$\|z - x^*\|_\infty \le \|z - x^*\| \le \|z - \hat{x}\| + \|\hat{x} - x^*\|$$
$$\le \frac{7}{72\sqrt{n}}\mu + \frac{1}{4\sqrt{n}}\mu + \frac{\varepsilon}{4}$$
$$< \frac{\mu}{2\sqrt{n}}.$$

This means that $(z, u) \in S$. The exact same procedure can now be carried out with z instead of x and we obtain in this way that all iterates lie in S. This can now be used together with the rate of convergence result in lemma 5.7. Since all these iterates lie in $S \subset S_1$, we can apply the bounds on various quantities which were obtained for this set.

This gives (see theorem 5.1):

$$\frac{\|q\|_{\overline{H}}}{\|p\|_H^2} \leq \frac{2\sqrt{n}}{B(x)\sqrt{\lambda_{\overline{H}}}} = \left(\frac{B(x)\sqrt{\lambda_{\overline{H}}}}{2\sqrt{n}}\right)^{-1} \leq \left(\frac{\mu\sqrt{\tilde{\lambda}}}{4\sqrt{n}}\right)^{-1} \leq \frac{1}{3\beta^*},$$

which completes the proof. □

6. Complexity estimation

Theorems 5.2 and 5.3 show that there comes a point, where in primal space only the "pure" Newton method is used, i.e., Newton's method with stepsize one, while the sequence of Lagrange multipliers remains in U_μ (as defined in section 3) with μ fixed. In other words, from some point on, the primal iterates are in the Newton area and remain there after every Lagrange multiplier update, while the Lagrange multipliers remain in U_μ and the distance to both the primal and dual solutions is reduced by a factor $0 < \gamma < 1$ after every multiplier update.

We say that $u \in U_\mu$ is well defined for the barrier parameter μ if $\gamma = C(1+\tau)\mu = (2\sqrt{n})^{-1}$. Then estimation (22) gives:

$$\max\{\|\hat{x} - x^*\|_\infty, \|\hat{u} - u^*\|_\infty\} \leq \gamma \|u - u^*\|_\infty .$$

We say that $x \in \Omega_\mu$ is well defined for a particular $u \in U_\mu$ if x lies in the area of quadratic convergence for Newton's method (see [31]) when applied to solving $\nabla_x \phi(x, u, \mu) = 0$.

DEFINITION 6.1

We call an approximation $x \in \Omega_\mu$ a "hot start" if

$$x \in \Omega_\mu \text{ is well defined for } u \in U_\mu$$

implies

$$\bar{x} \in \Omega_\mu \text{ is well defined for } \bar{u} \in U_\mu,$$

where the pair $(\bar{x}, \bar{u})$ is defined by (20)–(21).

Theorems 5.2 and 5.3 tell us that there exists a neigborhood T of (x^*, u^*) such that for any pair $(x, u) \in T$, the vector x is a "hot start". The first question is what the size of this neigborhood is and by which problem parameters it is determined. The second question is what it takes to reach T. First of all note that from theorem 4.2 with $\gamma = 1/(2\sqrt{n})$, we obtain $\theta^* \geq \mu$ and $\sigma \geq \mu \ln(2\sqrt{n})$. In view of the formula for β^* (see definition 5.6) and assuming that $\theta^* = \mu$ and $\sigma = \mu \ln(2\sqrt{n})$, we obtain a lower bound for β^*, since these assumptions reduce ξ^* and therefore also β^*. Under the assumptions of assertion 4.1, we have

$$\tilde{\lambda} \sim \mathcal{O}(\mu^{-1}), \quad \tilde{\Lambda} \sim \mathcal{O}(\mu^{-2}).$$

Therefore (see definition 5.5), $\lambda^* = \mathcal{O}(\mu^{-1})$.

For ξ^*, this gives

$$\xi^* \geq \frac{\mu^{1/2}}{10}\left(\sqrt{2} + \frac{27(m-r)}{4\sqrt{2n}(\ln 2\sqrt{n} + \frac{1}{2})^3(\mu\lambda^*)^{3/2}}\right)^{-1}. \tag{35}$$

This means that $\xi^* \sim \mathcal{O}(\mu^{1/2})$ or $\xi^* \sim \mathcal{O}(\frac{\sqrt{n}(\ln n)^3}{m-r}\mu^{1/2})$, depending on whether the first or the second term in the sum on the RHS of (35) is larger.

In the definition of β^*, we also have the expression $\mu\sqrt{\lambda^*}/\sqrt{n}$, which in view of our assumptions will be $\mathcal{O}(\sqrt{\mu}/\sqrt{n})$.

If we put all of this together, we obtain for the RHS in (34):

$$\frac{1}{2\sqrt{n}} \min\left\{\mu, \frac{\sqrt{\tilde{\lambda}}}{\tilde{\Lambda}}\beta^*\right\} \sim \mathcal{O}\left(\mu^2 \min\left\{\frac{1}{n}, \frac{(\ln n)^3}{m-r}\right\}\right).$$

Assuming that $m > n$, we obtain for the radius ρ of $T: \rho = \mathcal{O}(\mu^2 m^{-1})$. Keeping in mind that $\mu = 1/(2C\sqrt{n})$, we obtain

$$\rho = \mathcal{O}\left(\frac{1}{C^2 mn}\right). \tag{36}$$

We first estimate the number of Newton steps necessary to reach the set T. We recall that by applying the path-following method to the shifted barrier function $\phi(x, e, \mu)$, one can reduce the duality gap to a given level $\bar{\mu} > 0$ in $\mathcal{O}(\sqrt{m}\ln \bar{\mu}^{-1})$ Newton steps. Since after such a reduction, the primal approximation x will remain in the Newton area for the system $\phi(x, e, \bar{\mu}) = 0$, we can apply to x the operator R. Then, due to (22) with $\tau = 1$, we obtain for the pair

$$\bar{x} = Rx, \quad \bar{u} = \bar{\mu}[\mathrm{diag}(r_i(\bar{x}) + \bar{\mu})^{-1}]e: \tag{37}$$

$$\max\{\|\bar{x} - x^*\|, \|\bar{u} - u^*\|\} \leq 2C\bar{\mu}\|e - u^*\|.$$

Therefore, to guarantee that $\{\|\bar{x} - x^*\|, \|\bar{u} - u^*\|\} \leq \rho$, we must have:

$$2C\bar{\mu}\|e - u^*\| \leq \rho,$$

i.e.,

$$\bar{\mu} \le \tfrac{1}{2}\rho C^{-1}\|e-u^*\|^{-1} \le \tfrac{1}{2}\rho C^{-1}\|u^*\|^{-1}.$$

With (36), this gives:

$$\bar{\mu} = \mathcal{O}\left(\frac{1}{C^3 mn\|u^*\|}\right).$$

Hence, it will take $\mathcal{O}(\sqrt{m}\ln(C^3mn\|u^*\|\mu_0))$ Newton steps from a "warm start" x^0, that corresponds to the barrier parameter μ_0, to the "warm start" x^0, corresponding to the barrier parameter $\bar{\mu}$. Due to assumptions (1) and (2) at the beginning of section 5, it takes another $\mathcal{O}(\ln m)$ Newton steps to apply the operator R (see (37)) and to obtain the "hot start" $\bar{x} \in \Omega_{\bar{\mu}}$ and the corresponding vector $\bar{u} \in U_{\bar{\mu}}$ of the Lagrange multipliers. Therefore, it requires

$$N_1 = \mathcal{O}\left(\sqrt{m}\ln(C^3mn\|u^*\|)\right) \tag{38}$$

Newton steps (assuming $\mu_0 = 1$) to get from the "warm start" $x^0 \in \Omega_{\mu_0}$ for the system $\phi(x, e, \mu) = 0$ to the "hot start" $\bar{x} \in \Omega_{\bar{\mu}}$ and to obtain a vector $\bar{u} \in U_{\bar{\mu}}$ such that

$$\max\{\|\bar{x} - x^*\|_\infty, \|\bar{u} - u^*\|_\infty\} \le \gamma\|e - u^*\|_\infty,$$

$\gamma = (2\sqrt{n})^{-1}$. Moreover, $\bar{x}$ is again in the Newton area for the system $\nabla_x\phi(x, \bar{u}, \bar{\mu}) = 0$. Therefore, it takes at most $\mathcal{O}(\ln m)$ Newton steps to obtain

$$x^{(1)} = R\bar{x}, \quad u^{(1)} = [\operatorname{diag}(r_i(Rx^{(1)}) + \bar{\mu})^{-1}]\bar{u}$$

and again we have

$$\max\{\|x^{(1)} - x^*\|_\infty, \|u^{(1)} - u^*\|_\infty\} \le \gamma\|\bar{u} - u^*\|_\infty,$$

$\gamma = (2\sqrt{n})^{-1}$ and $x^{(1)}$ is well defined for the system $\nabla\phi(x, u^{(1)}, \bar{\mu}) = 0$. This means that in s outer iterations, the MBF method produces a sequence

$$x^{(s)} = Rx^{(s-1)} \quad \text{and} \quad u^{(s)} = [\operatorname{diag}(r_i(Rx^{(s-1)}) + \bar{\mu})^{-1}]u^{(s-1)}$$

such that

$$\max\{\|x^{(s)} - x^*\|_\infty, \|u^{(s)} - u^*\|_\infty\} \le \gamma^s\|e - u^*\|_\infty.$$

Every outer iteration requires at most $\mathcal{O}(\ln m)$ inner Newton steps. Therefore, to get an an approximation with accuracy 2^{-L} from any "hot start" (and with $\gamma = (2\sqrt{n})^{-1}$) takes

$$N_2 = \mathcal{O}\left(\left(L - \ln(C^2mn)\right)(\ln n)^{-1}\ln m\right) \tag{39}$$

Newton steps. We point out that N_2 is an upper bound. In fact, the number of Newton steps is decreasing drastically after every multiplier update until one Newton step is enough for the update (see [2,3,14,18]). Therefore, the main part of the estimate $N_1 + N_2$ for the total number of Newton steps is N_1. This means that the complexity of the Newton MBF method is a function of the condition number, the size of the problem and $\|u^*\|$. The important point here is that the complexity grows logarithmically. Assuming that $C > mn\|u^*\|$, we obtain from (38) and (39) that the total number of Newton steps is given by

$$N = \mathcal{O}(\sqrt{m}\ln C) + \mathcal{O}\Big((L - \ln C)(\ln n)^{-1}\ln m\Big). \tag{40}$$

The first term is the number of Newton steps from the "warm start" to the "hot start", while the second term is the number of Newton steps from the "hot start" to the solution.

We have to mention though that the estimate for N can be achieved if we have an estimate to the barrier parameter $\bar{\mu}$ in advance. Without knowing such an estimate, we will make a few attempts to turn from the shifted barrier to the MBF trajectory, before finding the first "hot start". Every such attempt requires $\mathcal{O}((L - \ln C)(\ln n)^{-1}\ln m)$ Newton steps. Therefore, the few attempts which might be necessary before the process turns to the MBF trajectory cannot substantially change estimate (40). Note that the upper bound for the number of attempts is $\mathcal{O}(\ln C)$. This means that even in the case of a degenerate problem, when we can assume that $C = 2^L$, the complexity of the Newton MBF method remains polynomial.

Recalling that for a linear programming problem for example, an accuracy of $\varepsilon = 2^{-L}$, with $L = \mathcal{O}(mn)$ is required to obtain an exact solution and under the assumption that $\varepsilon \ll (C^3 mn\|u^*\|)^{-1}$, this potentially represents a significant improvement over the classical barrier method where the complexity is given by $\mathcal{O}(\sqrt{m}\ln\varepsilon^{-1})$.

7. Concluding remarks

The global convergence of the MBF methods for any positive barrier parameter (see [15,26,27]) together with local properties P1–P5 not only contribute to numerical stability and a substantial improvement of the convergence rate, compared to classical interior point methods (see [7]), but also exhibit the "hot start" phenomenon, which leads to a significant improvement of the complexity bound. The reason for this is the fact that the MBF method converges linearly for a fixed barrier parameter and the fact that the condition number of the MBF Hessian is stable uniformly in $u \in U_\mu$. This implies that there comes a point where all the iterates remain well defined in the primal as well as in the dual space.

This means that after each Lagrange multiplier update, the primal minimizer is again in the area where $\|p\|_H$ converges to zero quadratically, while the new vector of Lagrange multipliers remains in U_μ, where the basic estimate (11) holds. Therefore, the procedure of finding the current minimizer is not expensive (only $\mathcal{O}(\ln\ln\varepsilon^{-1})$ Newton steps per update), while the improvement of the primal and dual approximation is by a factor $0 < \gamma < 1$ as soon as $\mu \leq \gamma / C$. In this case, one can improve the current approximation by a factor γ in $\mathcal{O}(\ln \ln \varepsilon^{-1})$ Newton steps. Moreover, the number of Newton steps decreases after every Lagrange multiplier update.

We would like to emphasize that, contrary to Interior Point methods that are based on CBFs (see [11]), or Shifted Barrier Functions (see [8,10]), the area in which $\|p\|_H$ converges to zero does not vanish when the primal approximation approaches the solution. In other words, the "Kantorovich ball" (see [17]) in which the quadratic convergence of $\|p\|_H$ to zero occurs, is determined by $\beta(x)$, which is bounded away from zero uniformly in c, whenever $u \in U_\mu$, i.e., there is a β^* such that $\beta(x) > \beta^* > 0$. The size of the neighborhood of x where $\|p\|_H$ falls below this value, and continues to be below this value for all subsequent iterations, is characterized in terms of the condition number of the QP problem, the size of the problem and $\|u^*\|$.

To reach the "hot start", one has to perform $\mathcal{O}(\sqrt{m} \ln(C^3 mn \|u^*\|))$ Newton steps, starting from any "warm start", using the path-following trajectory for the shifted barrier function with Newton's method.

It might be that for certain problems, the "hot start" might never be reached. The radius of the ball T, where the "hot start" begins, is a function of the condition number C and the size of the problem (see (36)). In order to be in this situation, we must have $(C^2 mn)^{-1} \leq 2^{-L}$. Keeping in mind assumptions (1) and (2) (from the beginning of section 5), we have $L = \mathcal{O}(m^3)$ $(m \geq n)$. Therefore, to miss the "hot start" one must have a condition number satisfying:

$$C \geq 2^{\mathcal{O}(m^3)} (mn)^{-1/2}.$$

So even for small problems, the condition number must be beyond imagination. Practically, it is unlikely to have such a condition number for a problem with $m \geq 100$. But even if we miss the "hot start" because $C \geq 2^L$, the algorithm will simply follow the shifted barrier trajectory until the end, with the classical complexity.

We have to keep in mind that the MBF method, in contrast to interior point methods based on the classical barrier functions, coverges for fixed positive μ. Practically (see [14]), most of the LP problems from the Netlib library were solved to high accuracy with a fixed $\mu = 10^{-4}$ in less than 100 Newton steps and to reach the "hot start", very few Lagrange multiplier updates are needed.

Appendix A

A.1 FROBENIUS' FORMULA

The following expression is valid if the appropriate matrix inverses exist (see [6, p. 102]):

$$\begin{pmatrix} M & N \\ R & S \end{pmatrix}^{-1} = \begin{pmatrix} (M - NS^{-1}R)^{-1} & -(M - NS^{-1}R)^{-1}NS^{-1} \\ -S^{-1}R(M - NS^{-1}R)^{-1} & S^{-1} + S^{-1}R(M - NS^{-1}R)^{-1}NS^{-1} \end{pmatrix}.$$

A.2 SOME LEMMAS ON THE NORMS OF SPECIAL MATRICES

For $A \in \mathbb{R}^{(n,n)}$, $B \in \mathbb{R}^{(r,n)}$, $x \in \mathbb{R}^n$, $y \in \mathbb{R}^r$ with:

(1) $\|B^T y\| \geq m_0 \|y\|$ $(m_0 > 0)$ $\forall y \in \mathbb{R}^r$, which implies

$$\left\| \left(BB^T\right)^{-1} \right\| \leq 1/m_0^2,$$

(2) $(Ax, x) \geq l_0 \|x\|^2$ $\forall x$ such that $Bx = 0$,

the following three lemmas hold (see [21]):

LEMMA A.1

If

$$\frac{1}{\mu} \geq \frac{2\|A\| \, \|B\|^2}{m_0^4}\left(1 + \frac{2}{l_0} + \frac{l_0}{2\|A\|}\right),$$

then

$$\left\| \left(A + \frac{1}{\mu} B^T B\right)^{-1} \right\| \leq \frac{2}{l_0}.$$

LEMMA A.2

If

$$\frac{1}{\mu} \geq \frac{2\|A\| \, \|B\|}{m_0^3}\left(1 + \frac{\|A\|^2}{l_0^2}\right)^{1/2},$$

then

$$\left\| \left(A + \frac{1}{\mu} B^T B\right)^{-1} B^T \right\| \leq \frac{2\mu \|B\|}{m_0^3}\left(1 + \frac{\|A\|^2}{l_0^2}\right)^{1/2}.$$

LEMMA A.3

If

$$\frac{1}{\mu} \geq \frac{2\|A\| \, \|B\|}{m_0^3}\left(1 + \frac{\|A\|^2}{l_0^2}\right)^{1/2},$$

then

$$\left\| I_r - \frac{1}{\mu} B\left(A + \frac{1}{\mu} B^T B\right)^{-1} B^T \right\| \leq \frac{2\mu \|A\| \, \|B\|^2}{m_0^3}\left(1 + \frac{\|A\|^2}{l_0^2}\right)^{1/2}.$$

A.3. ESTIMATION OF THE CONDITION NUMBER C

In order to find an upper bound on C, we will estimate c_0 which was defined in section 3. We start by writing down Φ_μ^{-1} using Frobenius' formula (see the first section of this appendix):

$$\Phi_\mu^{-1} = \begin{pmatrix} F^{-1} & -\frac{1}{\mu}F^{-1}A_r \\ \frac{1}{\mu}U_r^* A_r^T F^{-1} & -\frac{1}{\mu}\left(I_r - \frac{1}{\mu}U_r^* A_r^T F^{-1} A_r\right) \end{pmatrix},$$

where $F = Q + \frac{1}{\mu}A_r U_r^* A_r^T$. $\Phi_\mu^{-1} R_\mu$ is therefore given by:

$$\Phi_\mu^{-1} R_\mu = \begin{pmatrix} \frac{1}{\mu}F^{-1}A_r & F^{-1}A_{m-r}[\operatorname{diag}(\sigma_i + \mu)^{-1}]_{r+1}^m \\ \frac{1}{\mu}\left(I_r - \frac{1}{\mu}U_r^* A_r^T F^{-1} A_r\right) & \frac{1}{\mu}U_r^* A_r^T F^{-1} A_{m-r}[\operatorname{diag}(\sigma_i + \mu)^{-1}]_{r+1}^m \end{pmatrix}.$$

To estimate $\|\Phi_\mu^{-1} R_\mu\|_\infty$, we shall use the fact that for a matrix, composed of other matrices A, B, C and D:

$$\left\|\begin{pmatrix} M & N \\ R & S \end{pmatrix}\right\|_\infty \le \max\{\|M\|_\infty + \|N\|_\infty, \|R\|_\infty + \|S\|_\infty\},$$

and that for a matrix $A \in \mathbb{R}^{(p,s)}$:

$$\|A\|_\infty \le \sqrt{s}\,\|A\|.$$

In order to compute an upper bound for the 2-norm of each of the four matrices in $\Phi_\mu^{-1} R_\mu$, we shall use three lemmas from [20], which are restated in this appendix as lemmas A.1, A.2 and A.3. The result will be that for μ "small enough", the upper bound on each matrix will be independent of μ.

We rewrite $A_r U_r^* A_r^T$ as $(U_r^{*1/2} A_r^T)^T (U_r^{*1/2} A_r^T)$. With

$$\|A_r y\| \ge m_0 \|y\| \quad \text{and} \quad \|(A_r^T A_r)^{-1}\| \le \frac{1}{m_0^2}$$

implying

$$\|U_r^{*1/2} A_r y\| \ge \sqrt{\theta^*}\, m_0 \|y\| \quad \text{and} \quad \|(A_r^T U_r^* A_r)^{-1}\| \le \frac{1}{\theta^* m_0^2},$$

the aforementioned lemmas can easily be applied to our four submatrices. We look at each of these separately.

(1) For the first one, we have

$$\frac{1}{\mu}F^{-1}A_r = \frac{1}{\mu}\left(Q + \frac{1}{\mu}\left(U_r^{*1/2}A_r^T\right)^T\left(U_r^{*1/2}A_r^T\right)\right)^{-1}\left(U_r^{*1/2}A_r^T\right)^T U_r^{*-1/2}.$$

Taking norms yields:

$$\left\|\frac{1}{\mu}F^{-1}A_r\right\| \le \frac{\|U_r^{*-1/2}\|}{\mu}\left\|\left(Q+\frac{1}{\mu}\left(U_r^{*1/2}A_r^T\right)^T\left(U_r^{*1/2}A_r^T\right)\right)^{-1}\left(U_r^{*1/2}A_r^T\right)^T\right\|.$$

From lemmas A.1 and A.2, we have that for

$$\frac{1}{\mu} \ge \frac{2\|Q\|\,\|U_r^{*1/2}A_r^T\|}{\theta^{*3/2}m_0^3}\max\left\{\frac{\|U_r^{*1/2}A_r^T\|}{\sqrt{\theta^*}m_0}\left(1+\frac{2}{l_o}+\frac{l_0}{2\|Q\|}\right),\left(1+\frac{\|Q\|^2}{l_o^2}\right)^{1/2}\right\}, \tag{41}$$

the following estimate holds:

$$\left\|\left(Q+\frac{1}{\mu}\left((U_r^{*1/2}A_r^T\right)^T\left(U_r^{*1/2}A_r^T\right)\right)^{-1}\left(U_r^{*1/2}A_r^T\right)^T\right\| \le \frac{2\mu\|U_r^{*1/2}A_r^T\|}{\theta^{*3/2}m_0^3}\left(1+\frac{\|Q\|^2}{l_o^2}\right)^{1/2}.$$

This, together with A.3, gives

$$\left\|\frac{1}{\mu}F^{-1}A_r\right\| \le \frac{2\|U_r^{*1/2}A_r^T\|\,\|U_r^{*-1/2}\|}{\theta^{*3/2}m_0^3}\left(1+\frac{\|Q\|^2}{l_o^2}\right)^{1/2}.$$

Finally,

$$\left\|\frac{1}{\mu}F^{-1}A_r\right\| \le \alpha_1,$$

where we have defined

$$\alpha_1 = \frac{2\|A_r^T\|\sqrt{\rho^*}}{\theta^{*2}m_0^3}\left(1+\frac{\|Q\|^2}{l_o^2}\right)^{1/2}.$$

(2) We now look at the second submatrix:

$$\|F^{-1}A_{m-r}[\operatorname{diag}(\sigma_i+\mu)^{-1}]_{r+1}^m\| \le \|F^{-1}\|\,\|A_{m-r}[\operatorname{diag}(\sigma_i+\mu)^{-1}]_{r+1}^m\|.$$

Lemma A.1 yields:

$$\|F^{-1}A_{m-r}[\operatorname{diag}(\sigma_i+\mu)^{-1}]_{r+1}^m\| \le \frac{2}{l_0}\|F^{-1}\|(\sigma+\mu)^{-1},$$

and therefore:

$$\|F^{-1}A_{m-r}[\operatorname{diag}(\sigma_i+\mu)^{-1}]_{r+1}^m\| \le \alpha_2,$$

where

$$\alpha_2 = \frac{2}{\sigma l_0}\|A_{m-r}\|.$$

(3) The third submatrix gives:

$$\left\| \frac{1}{\mu}\left(I_r - \frac{1}{\mu} U_r^* A_r^T F^{-1} A_r \right) \right\|$$

$$= \left\| \frac{U_r^{*1/2}}{\mu}\left(I_r - \frac{1}{\mu}\left(U_r^{*1/2} A_r^T\right)^T F^{-1}\left(U_r^{*1/2} A_r^T\right)^T \right) U_r^{*-1/2} \right\|$$

$$\leq \frac{\|U_r^{*1/2}\| \, \|U_r^{*-1/2}\|}{\mu} \left\| I_r - \frac{1}{\mu}\left(U_r^{*1/2} A_r^T\right)^T F^{-1} \left(U_r^{*1/2} A_r^T\right)^T \right\|.$$

For μ satisfying (41), lemma A.3 gives:

$$\left\| I_r - \frac{1}{\mu}\left(U_r^{*1/2} A_r^T\right)^T F^{-1} \left(U_r^{*1/2} A_r^T\right)^T \right\| \leq \frac{2\mu \|Q\| \, \|U_r^{*1/2} A_r^{T2}}{(\sqrt{\theta^*} m_0)^5} \left(1 + \frac{\|Q\|^2}{l_o 2}\right)^{1/2}.$$

This, together with (42), yields:

$$\left\| \frac{1}{\mu}\left(I_r - \frac{1}{\mu} U_r^* A_r^T F^{-1} A_r \right) \right\| \leq \alpha_3,$$

with

$$\alpha_3 = \left(\frac{\rho^*}{\theta^*}\right)^{3/2} \frac{\|Q\| \, \|A_r^T\|^2}{\theta^{*3/2} m_0^5} \left(1 + \frac{\|Q\|^2}{l_o^2}\right)^{1/2}.$$

(4) Finally, we consider the fourth submatrix:

$$\left\| \frac{1}{\mu} U_r^* A_r^T F^{-1} A_{m-r} [\operatorname{diag}(\sigma_i + \mu)^{-1}]_{r+1}^m \right\|$$

$$\leq \left\| \frac{1}{\mu} U_r^* A_r^T F^{-1} \right\| \|A_{m-r}\| (\sigma + \mu)^{-1}$$

$$\leq \left\| \frac{1}{\mu} A_r^T F^{-1} \right\| \|A_{m-r}\| \, \|U_r^*\| \sigma^{-1}$$

$$\leq \left\| \frac{1}{\mu} \left((F^{-1})^T A_r\right)^T \right\| \|A_{m-r}\| \, \|U_r^*\| \sigma^{-1}$$

$$\leq \left\| \frac{1}{\mu} F^{-1} A_r \right\| \|A_{m-r}\| \, \|U_r^*\| \sigma^{-1}$$

$$\leq \frac{\rho^*}{\sigma} \|A_{m-r}\| \alpha_1$$

$$\leq \alpha_4 \alpha_1,$$

with

$$\alpha_4 = \frac{\rho^*}{\sigma} \| A_{m-r} \|.$$

Since $c_0 \le 2 \| \Phi_\mu^{-1} R_\mu \|_\infty$, we have obtained that

$$c_0 \le 2\sqrt{n} \max\{\alpha_1 + \alpha_2, \alpha_3 + \alpha_1 \alpha_4\}. \tag{43}$$

A.4. AN INEQUALITY FOR $\ln(x+1)$

We want to estimate $|\ln(1+x)|$ for $|x| \le 1/8$. From inequalities 4.1.33 in [1], we have

$$\frac{x}{1+x} \le \ln(1+x) \le x \quad (x > -1).$$

Therefore, since $|x| \le 1/8$:

$$|\ln(1+x)| \le \max\left\{1, \frac{1}{1-|x|}\right\} |x|$$

$$\le \max\left\{1, \tfrac{8}{7}\right\} |x|.$$

This means that

$$|\ln(1+x)| \le \tfrac{8}{7} |x|. \tag{44}$$

A.5. A LEMMA ABOUT CONVEX FUNCTIONS

LEMMA A.4

For a twice continuously differentiable function $\psi(x)$, defined on an open convex set $F \subset \mathbb{R}^n$, such that

$$\forall x \in F : \| \nabla^2 \psi(x) \| \le \Lambda,$$

and with

$$\hat{x} = \arg\min_x \psi(x),$$

the following inequality holds:

$$\| \nabla \psi(x) \| \le \Lambda \| x - \hat{x} \|.$$

Proof

Expanding $\nabla \psi(x)$ in a Taylor series around $\hat{x}$, we obtain, with $\tilde{x}$ a point on the segment between x and $\hat{x}$:

$$\| \nabla \psi(x) - \nabla \psi(\hat{x}) \| \le \| \nabla^2 \psi(\tilde{x}) \| \; \| x - \hat{x} \|.$$

Since $\nabla\psi(\hat{x}) = 0$, we have

$$\|\nabla\psi(x)\| \le \|\nabla^2\psi(\tilde{x})\| \, \|x - \hat{x}\| \le \Lambda\|x - \hat{x}\|. \qquad \square$$

A.6. PROOFS OF LEMMAS IN SECTIONS 4 AND 5

LEMMA A.5 (lemma 4.1)

Let $\|u^{(0)} - u^*\|_\infty \le \omega$ for some $\omega > 0$ and let $\gamma < 1$ be the rate of convergence, associated with $\mu > 0$, as in the basic convergence result (11).

Furthermore, we recall that the constraint functions satisfy the following Lipschitz condition:

$$\forall x, y \in \Omega_\mu, \quad \forall i : |r_i(x) - r_i(y)| \le \|x - y\|.$$

Then the following inequalities hold:

(1) For the active constraints ($i \in I$):

$$u_i^{(s)} \ge \exp\left(- \left(\frac{1-\gamma^s}{1-\gamma} \right) \frac{\gamma\omega\sqrt{n}}{\mu} \right) u_i^{(0)}.$$

(2) For the passive constraints ($i \in J$):

$$u_i^{(s)} \ge \exp\left(- \frac{s\sigma_i}{\mu} \right) \exp\left(- \left(\frac{1-\gamma^s}{1-\gamma} \right) \frac{\gamma\omega\sqrt{n}}{\mu} \right) u_i^{(0)},$$

with $\sigma_i = r_i(x^*)$.

Proof

According to the update formula for the Lagrange multipliers, we have for the active constraints ($i \in I$):

$$\begin{aligned} u_i^{(1)} &= \frac{\mu u_i^{(0)}}{r_i(x^{(1)}) + \mu} = \frac{\mu u_i^{(0)}}{r_i(x^{(1)}) - r_i(x^*) + \mu} \\ &\ge \frac{\mu u_i^{(0)}}{|r_i(x^{(1)}) - r_i(x^*)| + \mu} \ge \frac{\mu u_i^{(0)}}{\|x^{(1)} - x^*\| + \mu} \\ &\ge \frac{\mu u_i^{(0)}}{\sqrt{n}\,\|x^{(1)} - x^*\|_\infty + \mu} \ge \frac{\mu u_i^{(0)}}{\gamma\sqrt{n}\,\|u^{(0)} - u^*\|_\infty + \mu} \\ &\ge \left(1 + \frac{\gamma\omega\sqrt{n}}{\mu}\right)^{-1} u_i^{(0)}. \end{aligned}$$

Continuing to update in this way, we obtain

$$u_i^{(s)} \geq \left(1+\gamma^s \frac{\omega\sqrt{n}}{\mu}\right)^{-1}\left(1+\gamma^{s-1} \frac{\omega\sqrt{n}}{\mu}\right)^{-1} \cdots \left(1+\gamma \frac{\omega\sqrt{n}}{\mu}\right)^{-1} u_i^{(0)}.$$

Now, since $1+x \leq e^x$ and therefore $/1+x)^{-1} \geq e^{-x}$, we have:

$$u_i^{(s)} \geq \exp\left(-\frac{\gamma\omega\sqrt{n}}{\mu}(1+\gamma+\gamma^2+\cdots+\gamma^{s-1})\right)u_i^{(0)}.$$

Summing the geometric series in γ completes the first part of the proof.

(2) In the same way as before, we can write for the passive constraints ($i \in J$):

$$\begin{aligned} u_i^{(1)} &= \frac{\mu u_i^{(0)}}{r_i(x^{(1)})+\mu} = \frac{\mu u_i^{(0)}}{r_i(x^{(1)})-r_i(x^*)+r_i(x^*)+\mu} \\ &\geq \frac{\mu u_i^{(0)}}{|r_i(x^{(1)})-r_i(x^*)|+r_i(x^*)+\mu} \geq \frac{\mu u_i^{(0)}}{\|x^{(1)}-x^*\|+\sigma_i+\mu} \\ &\geq \frac{\mu u_i^{(0)}}{\gamma\sqrt{n}\,\|u^{(0)}-u^*\|_\infty+\sigma_i+\mu} \geq \left(1+\frac{\sigma_i}{\mu}+\frac{\gamma\omega\sqrt{n}}{\mu}\right)^{-1} u_i^{(0)}. \end{aligned}$$

Continuing to update in this way, we obtain

$$u_i^{(s)} \geq \exp\left(-\frac{s\sigma_i}{\mu}-\frac{\gamma\omega\sqrt{n}}{\mu}(1+\gamma+\gamma^2+\cdots+\gamma^{s-1})\right)u_i^{(0)}.$$

Again, summing the geometric series completes the proof. □

LEMMA A.6 (lemma 5.1)

If $x \in int\,\Omega_\mu$ and $\|d\|_H < \beta_1(x)$, then $x+d \in int\,\Omega_\mu$. Moreover, if $\|d\|_H < \beta_1(x)/2$, then

$$|\phi(x+d,u,\mu)-q_x(d,u,\mu)| < \frac{1}{3\xi_1(x)}\|d\|_H^3,$$

where

$$R_1(x) = \min\left\{\sqrt{\lambda_H}(r_j(x)+\mu) \mid j \in J_\mu\right\}$$

$$\beta_1(x) = \min\left\{\sqrt{\theta}, R_1(x)\right\}$$

$$\xi_1(x) = \frac{1}{2}\left(\frac{R_1(x)\sqrt{\theta}}{R_1(x)+\eta/(R_1^2(x))\sqrt{\theta}}\right).$$

Proof

The proof of the lemma will be along the lines of the proof of lemma 2.1 in [13]. Expanding $\phi(x+d, u, \mu)$ in a Taylor series about x, we can write:

$$\phi(x+d,u,\mu) = \phi(x,u,\mu) + d^T(\nabla_x \phi(x)) + \tfrac{1}{2} d^T(\nabla^2_{xx}\phi(x))d + \sum_{k=3}^{\infty} t_k, \qquad (45)$$

where the first three terms of the RHS constitute the quadratic approximation $q_x(d, u, \mu)$ to $\phi(x+d, u, \mu)$ at x and t_k is the kth order term in the Taylor expansion:

$$t_k = \frac{1}{k!}\sum_{i_1,\ldots,i_k}^{k} \frac{\partial^k \phi(x)}{\partial x_{i_1},\ldots,\partial x_{i_k}} d_{i_1}\ldots d_{i_k}.$$

In this particular case, we find by direct calculation that

$$t_k = \frac{1}{k}\sum_{i=1}^{m}(-1)^k u_i\left(\frac{a_i^T d}{r_i(x)+\mu}\right)^k.$$

Setting

$$\chi_i = \left|\frac{a_i^T d}{r_i(x)+\mu}\right|,$$

we now compute a bound on $|t_k|$ by computing a bound on

$$\frac{1}{k}\sum_{i=1}^{m} u_i\chi_i^k :$$

$$\sum_{i=1}^{m} u_i\chi_i^k \le \theta\sum_{i\in I_\mu}\left(\left(\frac{u_i}{\theta}\right)^{2/k}\chi_i^2\right)^{k/2} + \sum_{i\in J_\mu} u_i\chi_i^k \le \theta\sum_{i\in I_\mu}\left(\left(\frac{u_i}{\theta}\right)\chi_i^2\right)^{k/2} + \sum_{i\in J_\mu} u_i\chi_i^k$$

$$\le \theta\left(\sum_{i\in I_\mu}\left(\frac{u_i}{\theta}\right)\chi_i^2\right)^{k/2} + \eta_1\sum_{i\in J_\mu}\chi_i^k \le \theta\left(\frac{d^TQd}{\mu\theta} + \sum_{i\in I_\mu}\left(\frac{u_i}{\theta}\right)\chi_i^2\right)^{k/2} + \eta_1\sum_{i\in J_\mu}\chi_i^k$$

$$\le \theta\left(\frac{\|d\|_H^2}{\theta}\right)^{k/2} + \eta_1\sum_{i\in J_\mu}\chi_i^k.$$

For the last step we used the fact that

$$\|d\|_H^2 = \frac{d^TQd}{\mu} + \sum_{i=1}^{m} u_i\chi_i^2.$$

Now, if $\|d\|_H < R_1$, then

$$|\chi_i| = \left|\frac{a_i^T d}{r_i(x)+\mu}\right| \le \frac{\|d\|}{r_i(x)+\mu} \le \frac{\|d\|_H}{\sqrt{\lambda_H}(r_i(x)+\mu}$$

and therefore

$$|\chi_i| \le \frac{\|d\|_H}{R_1(x)} < 1.$$

This means that

$$\sum_{i=1}^{m} u_i\chi_i^k \le \theta\left(\frac{\|d\|_H^2}{\theta}\right)^{k/2} + \eta\left(\frac{\|d\|_H^2}{R_1^2(x)}\right)^{k/2}.$$

We are now ready to compute an upper bound on the magnitude of the difference between $\phi(x+d, u, \mu)$ and its quadratic approximation. From (45), this difference is given by $\sum_{k=1}^{\infty} t_k$, so we compute:

$$\left|\sum_{k=3}^{\infty} t_k\right| \le \sum_{k=3}^{\infty} \frac{1}{k}\left\{\sum_{i=1}^{m} u_i\chi_i^k\right\}$$

$$\le \frac{\theta}{3}\left(\frac{\|d\|_H}{\sqrt{\theta}}\right)^3\left(1+\frac{\|d\|_H}{\sqrt{\theta}}+\dots\right) + \frac{\eta}{3}\left(\frac{\|d\|_H}{R_1(x)}\right)^3\left(1+\frac{\|d\|_H}{R_1(x)}+\dots\right)$$

$$\le \left(\frac{\theta^{-1/2}}{3}\right)\frac{\|d\|_H^3}{1-(\|d\|_H/\sqrt{\theta}} + \left(\frac{\eta}{3R_1^3(x)}\right)\frac{\|d\|_H^3}{1-(\|d\|_H/R_1(x))}.$$

Since this upper bound is definitely finite as long as $\|d\|_H < \beta_1(x)$, we conclude that $x+d$ lies in the interior of the extended feasible set Ω_μ.

Moreover, if $\|d\|_H < \sqrt{\theta}/2$ and $\|d\|_H < R_1(x)/2$, then:

$$\left|\sum_{k=3}^{\infty} t_k\right| \le \frac{2}{3}\left(\frac{1}{\sqrt{\theta}} + \frac{\eta}{R_1^3(x)}\right)\|d\|_H^3$$

$$\le \frac{2}{3}\left(\frac{1}{\sqrt{\theta}} + \frac{\eta/R_1^2(x)}{R_1(x)}\right)\|d\|_H^3 \le \frac{1}{3\xi_1(x)}\|d\|_H^3.$$

□

LEMMA A.7 (lemma 5.2)

If

$$\|p\|_H < \beta_2(x) = \min\left\{\frac{R_1(x)}{2}, \frac{\xi_1(x)}{5}\right\},$$

then

$$\|x-\hat{x}(u,\mu)\|_H \le \tfrac{5}{2}\|p\|_H.$$

Proof

Take an arbitrary h such that $\|h\|_H = \frac{3}{2}\|p\|_H$. We then consider the values of ϕ on the ellipsoid: $\{x + p + h| \ \|h\|_H = \frac{3}{2}\|p\|_H\}$. We have

$$\|p + h\|_H \le \tfrac{5}{2}\|p\|_H.$$

Since $\xi_1(x) \le \sqrt{\theta}/2$ and $\|p\|_H \le \xi_1(x)/5$, $\|p\|_H$ satisfies the conditions of lemma 5.1. Using this lemma and the fact that $p = \arg\min_y q_x(y, u, \mu)$, we have

$$\begin{aligned}
\phi(x + p + h, u, \mu) &> q_x(p + h, u, \mu) - \frac{1}{3\xi_1(x)}\|p + h\|_H^3 \\
&> q_x(p, u, \mu) + \tfrac{1}{2}\|h\|_H^2 - \left(\tfrac{5}{2}\right)^3 \frac{1}{3\xi_1(x)}\|p\|_H^3 \\
&> q_x(p, u, \mu) + \tfrac{9}{8}\|p\|_H^2 - \frac{125}{24\xi_1(x)}\|p\|_H^3 \\
&> q_x(p, u, \mu) + \left(\frac{9}{8\|p\|_H} - \frac{125}{24\xi_1(x)}\right)\|p\|_H^3 \\
&> q_x(p, u, \mu) + \left(\frac{45}{8\xi_1(x)} - \frac{125}{24\xi_1(x)}\right)\|p\|_H^3 \\
&> q_x(p, u, \mu) + \frac{5}{12\xi_1(x)}\|p\|_H^3.
\end{aligned}$$

The previous lemma yields:

$$\phi(x + p, u, \mu) \le q_x(p, u, \mu) + \frac{1}{3\xi_1(x)}\|p\|_H^3.$$

For $\phi(x + p + h, u, \mu)$, this means

$$\begin{aligned}
\phi(x + p + h, u, \mu) &> \phi(x + p, u, \mu) + \frac{1}{\xi_1(x)}\left(\frac{5}{12} - \frac{1}{3}\right)\|p\|_H^3 \\
&> \phi(x + p, u, \mu) + \frac{1}{12\xi_1(x)}\|p\|_H^3.
\end{aligned}$$

This means that the value of ϕ is less in the center $x + p$ of the ellipsoid than on the boundary and since ϕ is strictly convex, its minimum has to be in the interior of the ellipsoid. In other words:

$$\|x - \hat{x}(u, \mu)\|_H \le \|h + p\|_H \le \tfrac{5}{2}\|p\|_H. \qquad \square$$

LEMMA A.8 (lemma 5.3)

If

$$\| p \|_H > \beta_2(x) = \min\left\{\frac{R_1(x)}{2}, \frac{\xi_1(x)}{5}\right\},$$

then the reduction $\Delta\phi$ in the MBF after a linesearch along the Newton direction p satisfies

$$\Delta\phi > \tfrac{2}{5}\beta_2^2(x).$$

Proof

Let ℓ be a steplength such that $\|\ell\|_H \le \beta_2(x)$. Then from lemma 5.1, we have

$$\phi(x + \ell p, u, \mu) \le q_x(\ell p, u, \mu) + \frac{1}{3\xi_1(x)}\ell^3 \| p \|_H^3,$$

and from the definition of $q_x(\ell p, u, \mu)$, we obtain

$$\begin{aligned}\phi(x) - \phi(x + \ell p, u, \mu) &\ge -\ell p^T g - \tfrac{1}{2}\ell^2 p^T H p - \frac{1}{3\xi_1(x)}\ell^3 \| p \|_H^3 \\ &\ge \ell \| p \|_H^2 - \tfrac{1}{2}\ell^2 \| p \|_H^2 - \frac{\ell^3}{3\xi_1(x)} \| p \|_H^3.\end{aligned}$$

Taking for ℓ the value $\beta_2(x)/\| p \|_H$ gives

$$\begin{aligned}\phi(x) - \phi(x + \ell p, u, \mu) &\ge \beta_2(x) \| p \|_H - \frac{\beta_2^2(x)}{2} - \frac{\beta_2(x)c}{3\xi_1(x)} \\ &\ge \beta_2^2(x)\left(1 - \frac{1}{2} - \frac{\beta_2(x)}{3\xi_1(x)}\right) \ge \frac{13}{30}\beta_2^2(x).\end{aligned}$$

□

LEMMA A.9 (lemma 5.4)

If

$$\| x - y \| \le \frac{B(x)}{\alpha} \quad (\alpha > 1), \tag{46}$$

then

$$\left(\frac{\alpha}{\alpha + 1}\right)\frac{1}{r_i(x) + \mu} \le \frac{1}{r_i(y) + \mu} \le \left(\frac{\alpha}{\alpha - 1}\right)\frac{1}{r_i(x) + \mu}.$$

Proof

$$\begin{aligned}\frac{1}{r_i(y) + \mu} &\le \frac{1}{r_i(x) - | r_i(x) - r_i(y) | + \mu} \le \frac{1}{r_i(x) - \| x - y \| + \mu} \\ &\le \frac{1}{r_i(x) + \mu - \frac{1}{\alpha}(r_i(x) + \mu)} \le \left(\frac{\alpha}{\alpha - 1}\right)\frac{1}{r_i(x) + \mu}.\end{aligned}$$

On the other hand, we have

$$\frac{1}{r_i(y)+\mu} \geq \frac{1}{r_i(x)+|r_i(y)-r_i(x)|+\mu} \geq \frac{1}{r_i(x)+\|x-y\|+\mu}$$
$$\geq \frac{1}{r_i(x)+\mu+\frac{1}{\alpha}(r_i(x)+\mu)} \geq \left(\frac{\alpha}{\alpha+1}\right)\frac{1}{r_i(x)+\mu}.$$

This completes the proof. □

LEMMA A.10 (lemma 5.5)

If

$$\|x-y\| \leq \frac{B(x)}{\alpha} \quad (\alpha > 1), \tag{47}$$

then

$$\left(\frac{\alpha}{\alpha+1}\right)^2 d^T H(x)d \leq d^T H(y)d \leq \left(\frac{\alpha}{\alpha-1}\right)^2 d^T H(x)d. \tag{48}$$

Proof

Using the previous lemma, we can write

$$d^T H(y)d = \frac{d^T Qd}{\mu} + \sum_{i=1}^{m} u_i \frac{(a_i^T d)^2}{(r_i(y)+\mu)^2}$$
$$\leq \frac{d^T Qd}{\mu} + \left(\frac{\alpha}{\alpha-1}\right)^2 \sum_{i=1}^{m} u_i \frac{(a_i^T d)^2}{(r_i(x)+\mu)^2} \leq \left(\frac{\alpha}{\alpha-1}\right)^2 d^T H(x)d.$$

On the other hand, we have

$$d^T H(y)d = \frac{d^T Qd}{\mu} + \sum_{i=1}^{m} u_i \frac{(a_i^T d)^2}{(r_i(y)+\mu)^2}$$
$$\geq \frac{d^T Qd}{\mu} + \left(\frac{\alpha}{\alpha+1}\right)^2 \sum_{i=1}^{m} u_i \frac{(a_i^T d)^2}{(r_i(x)+\mu)^2} \geq \left(\frac{\alpha}{\alpha+1}\right)^2 d^T H(x)d.$$

This completes the proof. □

LEMMA A.11 (lemma 5.6)

For $x, x+p \in int\, \Omega_\mu$ with $\|p\| \leq B(x)/\alpha$ $(\alpha > 1)$, the following inequality holds:

$$\|\nabla\phi(x+p)\| \leq \left(\frac{\alpha^3\sqrt{n}}{(\alpha-1)^3 B(x)}\right)\|p\|_H^2.$$

Proof

We have by expanding in a Taylor series that

$$\frac{\partial \phi(x+p)}{\partial x_j} = \frac{\partial \phi(x)}{\partial x_j} + \sum_{i=1}^{n} \frac{\partial^2 \phi(x)}{\partial x_i \partial x_j} p_i + \frac{1}{2} \sum_{s,t=1}^{n} \frac{\partial^3 \phi(\tilde{x})}{\partial x_s \partial x_t \partial x_j} p_s p_t,$$

where $\tilde{x} = x + \zeta p$ $(0 < \zeta < 1)$.

From the definition of p, this means

$$\left| \frac{\partial \phi(x+p)}{\partial x_j} \right| = \frac{1}{2} \left| \sum_{s,t=1}^{n} \frac{\partial^3 \phi(\tilde{x})}{\partial x_s \partial x_t \partial x_j} p_s p_t \right|.$$

We now compute a bound on the expression on the RHS:

$$\frac{1}{2} \left| \sum_{s,t=1}^{n} \left(\frac{\partial^3 \phi(x+\zeta p)}{\partial x_s \partial x_t \partial x_j} p_s p_t \right) \right| = \left| \sum_{i=1}^{m} u_i \frac{(a_i^T p)^2}{(r_i(x+\zeta p)+\mu)^3} (a_i)_j \right|$$

$$\leq \sum_{i=1}^{m} u_i \frac{(a_i^T p)^2}{(r_i(x+\zeta p)+\mu)^2} \left| \frac{(a_i)_j}{r_i(x+\zeta p)+\mu} \right|.$$

Because $|(a_i)_j| \leq \|a_i\| = 1$ and $\|x + \zeta p - x\| \leq \|p\|$ and because of the assumption on $\|p\|$, we can use lemmas 5.4 and 5.5. This gives

$$\sum_{i=1}^{m} u_i \frac{(a_i^T p)^2}{(r_i(x+\zeta p)+\mu)^2} \left| \frac{(a_i)_j}{r_i(x+\zeta p)+\mu} \right|$$

$$\leq \sum_{i=1}^{m} u_i \frac{(a_i^T p)^2}{(r_i(x+\zeta p)+\mu)^2} \left(\frac{\alpha}{(\alpha-1)(r_i(x)+\mu)} \right)$$

$$\leq \frac{\alpha}{(\alpha-1)B(x)} \sum_{i=1}^{m} u_i \frac{(a_i^T p)^2}{(r_i(x+\zeta p)+\mu)^2}$$

$$\leq \frac{\alpha}{(\alpha-1)B(x)} p^T H(x+\zeta p) p \leq \frac{\alpha^3}{(\alpha-1)^3 B(x)} p^T H(x) p.$$

We therefore have

$$\left| \frac{\partial \phi(x+p)}{\partial x_j} \right| \leq \frac{\alpha^3}{(\alpha-1)^3 B(x)} \|p\|_H^2.$$

Squaring the LHS and summing over j gives:

$$\sum_{j=1}^{n} \left| \frac{\partial \phi(x+p)}{\partial x_j} \right|^2 \leq n \left(\frac{\alpha^3}{(\alpha-1)^3 B(x)} \right)^2 \| p \|_H^4.$$

Taking the square root on both sides completes the proof. □

LEMMA A.12 (lemma 5.8)

If

$$\| p \|_H \leq \frac{B(x)\sqrt{\lambda_H}}{\alpha},$$

then

(1) $$\left(\frac{\alpha-1}{\alpha}\right) B(x) \leq B(\bar{x}) \leq \left(\frac{\alpha+1}{\alpha}\right) B(x),$$

(2) $$\left(\frac{\alpha}{\alpha+1}\right)^2 \leq \frac{\lambda_{\bar{H}}}{\lambda_H} \leq \left(\frac{\alpha}{\alpha-1}\right)^2,$$

(3) $$\left(\frac{\alpha-1}{\alpha+1}\right) R_1(x) \leq R_1(\bar{x}) \leq \left(\frac{\alpha+1}{\alpha-1}\right) R_1(x),$$

(4) $$\left(\frac{\alpha-1}{\alpha+1}\right)^3 \xi_1(x) \leq \xi_1(\bar{x}) \leq \left(\frac{\alpha+1}{\alpha-1}\right)^3 \xi_1(x).$$

Proof

We start by noting that

$$\| x - \bar{x} \| = \| p \| \leq \frac{1}{\sqrt{\lambda_H}} \| p \|_H. \tag{49}$$

(1) Because of (49), the first part of the proof follows immediately from lemma 5.4 and the definition of $B(x)$.

(2) Again, because of (49), the assumption in the statement of the lemma means that

$$\| x - \bar{x} \| \leq \frac{B(x)}{\alpha}.$$

Lemma 5.5 then gives

$$\left(\frac{\alpha}{\alpha+1}\right)^2 d^T H(x) d \leq d^T H(\bar{x}) d \leq \left(\frac{\alpha}{\alpha-1}\right)^2 d^T H(x) d,$$

and therefore

$$\left(\frac{\alpha}{\alpha+1}\right)^2 \frac{d^T H(x) d}{\| d \|^2} \leq \frac{d^T H(\bar{x}) d}{\| d \|^2} \leq \left(\frac{\alpha}{\alpha-1}\right)^2 \frac{d^T H(x) d}{\| d \|^2}.$$

Recalling that $\lambda_H = \min_d d^T Hd/\|d\|^2$ (and the analog for $\lambda_{\bar{H}}$), the proof of the second part follows.

(3) From lemma 5.4 and from the definition of $R_1(x)$, we have that

$$\left(\frac{\alpha-1}{\alpha}\right)(r_i(x)+\mu) \le r_i(\bar{x})+\mu \le \left(\frac{\alpha+1}{\alpha}\right)(r_i(x)+\mu),$$

$$\left(\frac{\alpha-1}{\alpha}\right)\min_{i\in J_\mu}\{r_i(x)+\mu\} \le \min_{i\in J_\mu}\{r_i(\bar{x})+\mu\} \le \left(\frac{\alpha+1}{\alpha}\right)\min_{i\in J_\mu}\{r_i(x)+\mu\},$$

$$\left(\frac{\alpha-1}{\alpha}\right)\left(\frac{\lambda_{\bar{H}}}{\lambda_H}\right)^{1/2} R_1(x) \le R_1(\bar{x}) \le \left(\frac{\alpha+1}{\alpha}\right)\left(\frac{\lambda_{\bar{H}}}{\lambda_H}\right)^{1/2} R_1(x).$$

With lemma 5.5, the proof of part (3) follows immediately.

(4) For the proof of part (4), we start with part (3):

$$\left(\frac{\alpha-1}{\alpha+1}\right)^3 R_1^3(x) \le R_1^3(\bar{x}) \le \left(\frac{\alpha+1}{\alpha-1}\right)^3 R_1^3(x),$$

$$\left(\frac{\alpha+1}{\alpha-1}\right)^3 \frac{\eta\sqrt{\theta}}{R_1^3(x)} \ge \frac{\eta\sqrt{\theta}}{R_1^3(\bar{x})} \ge \left(\frac{\alpha-1}{\alpha+1}\right)^3 \frac{\eta\sqrt{\theta}}{R_1^3(x)},$$

$$\left(\frac{\alpha+1}{\alpha-1}\right)^3\left(\frac{\eta\sqrt{\theta}}{R_1^3(x)}+1\right) \ge \frac{\eta\sqrt{\theta}}{R_1^3(\bar{x})}+1 \ge \left(\frac{\alpha-1}{\alpha+1}\right)^3\left(\frac{\eta\sqrt{\theta}}{R_1^3(x)}+1\right).$$

Taking the reciprocal of all three expressions, multiplying through by $\sqrt{\theta}$ and recalling the definition of $\xi_1(x)$ completes the proof. □

LEMMA A.13 (lemma 5.10)

Let $(x, u) \in S_1$ and let the rate of convergence of the algorithm be given by $\gamma = 1/(2\sqrt{n})$.

Then

(1) for the active constraints ($i \in I$):

$$\tfrac{1}{2}u_i^* < u_i < \tfrac{3}{2}u_i^*$$

$$\frac{\mu}{2} < r_i(x)+\mu < \frac{3\mu}{2},$$

and

(2) for the passive constraints ($i \in J$):

$$r_i(x) + \mu > \sigma + \frac{\mu}{2}.$$

Proof

(1) From theorem 4.2 with $\gamma = 1/(2\sqrt{n})$, we have that $\mu \le \theta^*$. Therefore, for any u such that

$$\|u - u^*\|_\infty < \frac{\mu}{2\sqrt{n}} \le \frac{\theta^*}{2\sqrt{n}},$$

the following inequalities hold:

$$u_i^* - \frac{\theta^*}{2\sqrt{n}} < u_i < u_i^* + \frac{\theta^*}{2\sqrt{n}}.$$

For the active constraints, this means

$$\tfrac{1}{2} u_i^* < u_i < \tfrac{3}{2} u_i^* \quad (i \in I).$$

We now turn to the second set of inequalities. Since

$$r_i(x) + \mu = r_i(x) - r_i(x^*) + \mu,$$

we can write

$$\mu - |r_i(x) - r_i(x^*)| \le r_i(x) + \mu \le \mu + |r_i(x) - r_i(x^*)|,$$

$$\mu - \|x - x^*\| \le r_i(x) + \mu \le \mu + \|x - x^*\|.$$

The proof then follows because $\|x - x^*\| < \mu/2$.

(2) For the proof of the second part, we write

$$\begin{aligned} r_i(x) + \mu &= r_i(x) - r_i(x^*) + r_i(x^*) + \mu \\ &\ge \sigma + \mu - |r_i(x) - r_i(x^*)| \\ &\ge \sigma + \mu - \|x - x^*\|. \end{aligned}$$

Again, the proof follows from the asumption that $\|x - x^*\| < \mu/2$. □

LEMMA A.14 (lemma 5.11)

For $(x, u) \in S_1$, the smallest and largest eigenvalues of the Hessian of $\phi(x, u, \mu)$ at x for fixed u and μ are bounded as follows:

$$\tfrac{2}{9}\lambda^* \le \lambda_H \le 6\lambda^* + \frac{\mu(m-r)}{2\sqrt{n}(\sigma + \frac{\mu}{2})^2},$$

$$\tfrac{2}{9}\Lambda^* \le \Lambda_H \le 6\Lambda^* + \frac{\mu(m-r)}{2\sqrt{n}(\sigma + \frac{\mu}{2})^2}.$$

Proof

We use the previous lemma to obtain the following inequalities, valid for $i \in I$ and for any d:

$$\tfrac{1}{2} u_i^* \frac{(a_i^T d)^2}{\left(\tfrac{3}{2}\mu\right)^2} \le u_i \frac{(a_i^T d)^2}{(r_i(x)+\mu)^2} \le \tfrac{3}{2} u_i^* \frac{(a_i^T d)^2}{\left(\tfrac{1}{2}\mu\right)^2},$$

and therefore

$$\tfrac{2}{9} u_i^* \frac{(a_i^T d)^2}{\mu^2} \le u_i \frac{(a_i^T d)^2}{(r_i(x)+\mu)^2} \le 6 u_i^* \frac{(a_i^T d)^2}{\mu^2}.$$

With the help of these inequalities, and recalling that

$$\lambda_H = \min_{\|d\|=1} d^T H d \quad \text{and} \quad \Lambda_H = \max_{\|d\|=1} d^T H d,$$

we have for the lower bounds:

$$\min_{\|d\|=1}\left\{\frac{d^T Q d}{\mu} + \sum_{i=1}^{m} u_i \frac{(a_i^T d)^2}{(r_i(x)+\mu)^2}\right\} \ge \min_{\|d\|=1}\left\{\frac{d^T Q d}{\mu} + \sum_{i=1}^{r} u_i \frac{(a_i^T d)^2}{(r_i(x)+\mu)^2}\right\}$$

$$\ge \tfrac{2}{9} \min_{\|d\|=1}\left\{\frac{d^T Q d}{\mu} + \sum_{i=1}^{r} u_i^* \frac{(a_i^T d)^2}{\mu^2}\right\}.$$

The exact same procedure goes through for the largest eigenvalue, if we take max instead of min in the above expressions.

Similarly, we have for the upper bounds:

$$\min_{\|d\|=1}\left\{\frac{d^T Q d}{\mu} + \sum_{i=1}^{m} u_i \frac{(a_i^T d)^2}{(r_i(x)+\mu)^2}\right\}$$

$$\le \min_{\|d\|=1}\left\{\frac{d^T Q d}{\mu} + 6\sum_{i=1}^{r} u_i^* \frac{(a_i^T d)^2}{\mu^2} + \sum_{i=r+1}^{m} u_i \frac{(a_i^T d)^2}{(\sigma+\mu/2)^2}\right\}$$

$$\le 6 \min_{\|d\|=1}\left\{\frac{d^T Q d}{\mu} + \sum_{i=1}^{r} u_i^* \frac{(a_i^T d)^2}{\mu^2}\right\} + \frac{\mu(m-r)}{2\sqrt{n}(\sigma+\mu/2)^2}.$$

In the last step, we have used the facts that $|a_i^T d| \le \|a_i\| \|d\| \le 1$ and that, since $(x, u) \in S_1$, $u_i < \mu/(2\sqrt{n})$.

Again the same can be done for the largest eigenvalue, with max replacing min and this completes the proof. □

Appendix B: Notation

$f_0(x) = \{\frac{1}{2} x^T Q x - q^T x \mid x \in \Omega\}$

$r_i(x) = a_i^T x - b_i,\ a_i \in \mathbb{R}^n,\ b_i \in \mathbb{R}$

$\Omega = \{x \mid r_i(x) \geq 0,\ i = 1,\dots,m\}$

$\Omega_k = \{x \mid r_i(x) \geq -k^{-1},\ i = 1,\dots,m\}$

$\Omega_\mu = \{x \mid r_i(x) \geq -\mu,\ i = 1,\dots,m\}$

$I = \{1,\dots,r\} = \{i : r_i(x^*) = 0\},\ r < n < m$

$J = \{r+1,\dots,m\} = \{i : r_i(x^*) > 0\}$

$A_r \in \mathbb{R}^{n,r}$ – matrix with columns $a_1,\dots,a_r$

$A_{m-r} \in \mathbb{R}^{n,m-r}$ – matrix with columns $a_{r+1},\dots,a_n$

$u^T = (u_1,\dots,u_m) \in \mathbb{R}^m_+,\ U = [\mathrm{diag}(u_i)]_{i=1}^m$

$u_{(r)}^T = (u_1,\dots,u_r),\ u_{(m-r)}^T = (u_{r+1},\dots,u_m)$

$\theta^* = \min\{u_i^* \mid i \in I\}$

$\sigma = \min\{r_i(x^*) \mid i \in J\}$

$\rho^* = \max\{u_i^* \mid i \in I\}$

$D(\cdot) = D_1(\cdot) \otimes D_2(\cdot) \otimes \dots \otimes D_m(\cdot)$

$D_i(\cdot) = D_i(u^*, k_0, \delta, \varepsilon) = \{(u_i, k) \in \mathbb{R}^2 \mid u_i \geq \varepsilon,\ |u_i - u_i^*| \leq \delta k,\ k \geq k_0\},\ i \in I$

$D_i(\cdot) = D_i(u^*, k_0, \delta, \varepsilon) = \{(u_i, k) \in \mathbb{R}^2 \mid 0 \leq u_i \leq \delta k,\ k \geq k_0\},\ i \in U$

$U_k^i = \{u_i : u_i^* - \delta k \leq u_i \leq u_i^* + \delta k\},\ i \in I$

$U_k^i = \{u_i : 0 \leq u_i \leq \delta k\},\ i \in J$

$U_k = U_k^1 \otimes U_k^2 \otimes \dots \otimes U_k^m$

$\alpha_k = \max\{(r_i(x^*) + k^{-1})^{-1} \mid i = r+1,\dots,m\}$

$\beta_k = \|\Phi_k^{-1} R_k\|_\infty,\ c_0 = \|\Phi_\infty^{-1} R_\infty\|_\infty$

$C = \max\{\sigma^{-1}, c_0\},\ \gamma = C k^{-1} = C\mu,\ \mu = k^{-1}$

$p(x, u, \mu) = -(\nabla_{xx}^2 \phi(x, u, \mu))^{-1} \nabla_x \phi(x, u, \mu)$

$\|x\|_H = \sqrt{x^T H x}$

$I_\mu = \{i \mid u_i \geq \mu/2\}$

$J_\mu = \{i \mid u_i < \mu/2\}$

$|J_\mu|$ = the number of elements in J_μ

$\theta = \min\{u_i \mid i \in I_\mu\}$

$\rho = \max\{u_i\}$

$\eta = |J_\mu| \max\{u_i | \, i \in J_\mu\}$

$\eta_1 = \max\{u_i | \, i \in J_\mu\}$

$\lambda_H = \text{mineigval}\, H(x, u, \mu)$, $\Lambda_H = \text{maxeigval}\, H(x, u, \mu)$

$q_x(d, u, \mu) = \phi(x, u, \mu) + g^T d + \frac{1}{2} d^T H d$

$R_1(x) = \min\left\{\sqrt{\lambda_H}\,(r_j(x) + \mu) \,|\, j \in J_\mu\right\}$

$\beta_1(x) = \min\left\{\sqrt{\theta}, R_1(x)\right\}$

$\beta_2(x) = \min\left\{\frac{R_1(x)}{2}, \frac{\xi_1(x)}{5}\right\}$

$\beta_3(x, \alpha, \bar{\alpha}) = \min\left\{\frac{R_1(x)}{2}, \frac{\xi_1(x)}{5}, \frac{1}{2}, \frac{B(x)\sqrt{\lambda_H}}{\alpha}, \left(\frac{\alpha-1}{\alpha}\right)^3 \frac{B(x)\sqrt{\lambda_H}}{\bar{\alpha}\sqrt{n}}\right\}$

$\beta(x) = \beta_3(x, 6, \frac{125}{36}) = \min\left\{\frac{\xi_1(x)}{5}, \frac{1}{2}, \frac{B(x)\sqrt{\lambda_H}}{6\sqrt{n}}\right\}$

$\beta^* = \min\left\{\frac{1}{2}, \xi^*, \frac{\mu\sqrt{\lambda^*}}{18\sqrt{2}\sqrt{n}}\right\}$

$\xi_1(x) = \frac{1}{2}\left(\frac{R_1(x)\sqrt{\theta}}{R_1(x) + \frac{\eta}{R_1^2(x)}\sqrt{\theta}}\right)$

$B(x) = \min\{r_i(x) + \mu \,|\, i = 1, \ldots, m\}$

$p = p(x, u, \mu)$

$\bar{x} = x + p$

$q = p(\bar{x}, u, \mu)$

$S_1 = \{(x, u) | \max\{\| x - x^* \|, \sqrt{n}\, \| u - u^* \|_\infty\} < \frac{\mu}{2}\}$

$\lambda^* = \text{mineigval}\, H(x^*, u^*, \mu)$, $\Lambda^* = \text{maxeigval}\, H(x^*, u^*, \mu)$

$\xi^* = \frac{1}{10}\left(\frac{\sqrt{2}}{\sqrt{\theta^*}} + \frac{27(m-r)\mu}{4\sqrt{2n}\lambda^{*3/2}(\sigma + \frac{\mu}{2})^3}\right)^{-1}$

$S = \left\{(x, u) : \max\{\| x - x^* \|_\infty, \| u - u^* \|_\infty\} < \frac{\mu}{2\sqrt{n}}\right\}$

$T = \left\{(x, u) : \max\{\| x - x^* \|_\infty, \| u - u^* \|_\infty\} < \frac{1}{2\sqrt{n}} \min\left\{\mu, \frac{\sqrt{\bar{\lambda}}}{\bar{\Lambda}} \beta^*\right\}\right\}$

References

[1] M. Abramowitz and I.A. Stegun, *Handbook of Mathematical Functions* (Dover, New York, 1972.

[2] A. Ben-Tal, I. Yuzefovich and M. Zibulevsky, Penalty/barrier multiplier methods for minimax and constrained smooth convex programs, Research Report 9/92, Optimization Laboratory, Technion, Haifa, Israel (1992).

[3] M.G. Breitfeld and D.F. Shanno, Preliminary computational experience with modified log-barrier functions for large-scale nonlinear programming, in: *Large Scale Optimization, State of the Art*, ed. W. Hager, D. Hearn and P. Pardalos (Kluwer Academic, 1994).

[4] D.P. Bertsekas, *Constrained Optimization and Lagrange Multiplier Methods* (Academic Press, New York, 1982).

[5] D. den Hertog, C. Roos and T. Terlaky, On the classical logarithmic barrier function method for a class of smooth convex programming problems, Technical Report 90-28, Delft University of Technology, Delft, The Netherlands (1990).

[6] V.N. Fadeeva, *Computational Methods of Linear Algebra* (Dover, New York, 1959).

[7] A.V. Fiacco and G.P. McCormick, *Nonlinear Programming: Sequential Unconstrained Minimization Techniques* (Wiley, New York, 1968).

[8] R.M. Freund, Theoretical efficiency of a shifted barrier function algorithm for linear programming, Linear Algebra and its Applications 152(1991)19–41.

[9] P.E. Gill, W. Murray, M.A. Saunders, J.A. Tomlin and M.H. Wright, On projected Newton barrier methods for linear programming and an equivalence to Karmarkar's projective method, Mathematical Programming 36(1986)183–209.

[10] P.E. Gill, W. Murray, M.A. Saunders, J.A. Tomlin and M.H. Wright, Shifted barrier methods for linear programming, Technical Report SOL 88-9, Systems Optimization Laboratory, Dept. of Operations Research, Stanford University, Stanford, CA 94305 (August 1988).

[11] C.C. Gonzaga, Path-following methods for linear programming, SIAM Review 34(1992) 167–224.

[12] D. Goldfarb and S. Liu, An $O(n^3L)$ primal interior point algorithm for convex quadratic programming, Mathematical Programming 49(1990/1991)325–340.

[13] F. Jarre, The method of analytic centers for smooth convex programs, Ph.D. Thesis, Institut für Angewandte Mathematik und Statistik, Universität Würzburg, Würzburg, Germany (1989).

[14] D. Jensen, R. Polyak and R. Schneur, Numerical experience with MBF methods for LP, Research Report, IBM, T.J. Watson Research Center, Yorktown Heights, NY 10598 (1992) pp. 1–31.

[15] D. Jensen and R. Polyak, Modified Barrier Method convergence in quadratic programming, *16th IFIP Conference on System Modelling and Optimization*, vol. 2 (1993).

[16] D. Jensen and R. Polyak, The convergence of a Modified Barrier Method for convex programming, IBM Journal of Research and Development 38(1994)307–321.

[17] L. Kantorovich and G. Akilov, *Functional Analysis in Normed Spaces* (McMillan, New York, 1964).

[18] S. Nash, R. Polyak and A. Sofer, A numerical comparison of Barrier and Modified Barrier Methods for large-scale bound-constrained optimization, in: *Large Scale Optimization, State of the Art*, ed. W. Hager, D. Hearn and P. Pardalos (Kluwer Academic, 1994).

[19] Yu.E. Nesterov and A.S. Nemirovsky, *Interior Point Polynomial Methods in Convex Programming: Theory and Algorithms* (SIAM, Philadelphia, 1994).

[20] B.T. Polyak, *Introduction to Optimization* (Nauka, Moscow, 1983) (in Russian).

[21] B.T. Polyak and N.V. Tret'yakov, The method of penalty estimates for conditional extremum problems, Zhurnal Vycheslitel'noi Matematiki i Matematicheskoi Fiziki 13(1973)34–46 (in Russian).

[22] R. Polyak and I. Polyak, *Constrained Optimization Problems and Multiplier Methods* (Znanie, Kiev, 1985) (in Russian).

[23] R. Polyak, Controlled processes in extremal and equilibrium problems, VINITI, Deposited manuscript, Moscow (1986) 420 p. (in Russian).

[24] R. Polyak, Modified barrier functions (theory and methods), Mathematical Programming 54 (1992)177–222.

[25] R. Polyak, Modified barrier functions in LP, IBM Research Report RC 17790 (#78331), Yorktown Heights, NY (1992).

[26] R. Polyak and M. Teboulle, Nonlinear rescaling principle and proximal-like methods in convex optimization, Research Report, Dept. Operations Research, George Mason University, Fairfax, VA (1995), to appear in Mathematical Programming.

[27] M. Powell, Some convergence properties of a shifted log barrier method for linear programming, Technical Report NA7, DAMTP, Cambridge, UK (1992).

[28] J. Renegar and M. Shub, Unified complexity analysis for Newton LP methods, Mathematical Programming 53(1992)1–16.

[29] R.T. Rockafellar, Augmented Lagrange multiplier functions and duality in nonconvex programming, SIAM Journal on Control 12(1974)268–285.

[30] C. Roos and J.-Ph. Vial, Long steps with the logarithmic penalty barrier function in linear programming, Technical Report No. 89-44, Faculty of Mathematics and Informatics/Computer Science, Delft University of Technology, Delft, The Netherlands (1989).

[31] S. Smale, Newton method estimates from data at one point, The merging of disciplines in pure, applied and computational mathematics, ed. R.E. Ewing et al. (Springer, New York/Berlin, 1986).

[32] S. Smale and M. Shub, private communication.

[33] P. Wolfe, A duality theorem for nonlinear programming, Quarterly of Applied Mathematics 19(1961)239–244.