COMPLEXITY ANALYSIS FOR THE NEWTON MODIFIED BARRIER FUNCTION METHOD

Thesis by

Aharon Melman

In Partial Fulfillment of the Requirements

for the Degree of

Doctor of Philosophy

California Institute of Technology

Pasadena, California

1992

(Defended February 18, 1992)

to my parents

Acknowledgements

It is a great pleasure to express my gratitude to Professor Herb Keller and Dr. Roman Polyak for their constant support during my research, with special thanks to Dr. Polyak for accepting to co-supervise my work in addition to his many duties at IBM.

I am also grateful to IBM for letting me work at the T.J. Watson Research Center. I gained valuable experience during my stay there and especially enjoyed working with Dr. David Jensen. I also thank the Polyak family for taking such good care of me during this period.

I am thankful for the support of my parents, my friends here at Caltech and in the outside world and particularly the Remennik-Katz family where I was always welcome whenever I wanted to escape.

Finally, I gratefully acknowledge the financial support from the NSF and DOE.

Abstract

The modified barrier function (MBF) is examined for linear, convex quadratic and other convex nonlinear constrained optimization problems. This new method of transforming a constrained problem into a sequence of unconstrained ones has elements of both Lagrangian function and barrier function methods. At each step, the method updates multipliers, which converge to the optimal Lagrange multipliers. Each such update entails a minimization using Newton's method.

We show that there is a ball around the primal-dual solution of the optimization problem, a so-called "hot start" ball, such that starting from any point in this ball, Newton's method converges quadratically and continues to do so after each subsequent update. We characterize the "hot start" ball in terms of the primal-dual solution of the optimization problem.

This means that from the "hot start" on, only $\mathcal{O}(\ln\ln\epsilon^{-1})$ Newton steps are necessary after each update in order to reach the next update ($\epsilon > 0$ is the desired accuracy for the solution). Taking into account the basic MBF convergence properties, one obtains that the number of Newton steps from a "hot start" to the solution is $\mathcal{O}((\ln\ln\epsilon^{-1})(\ln\epsilon^{-1}))$.

To reach the "hot start" one has to spend $\mathcal{O}(\sqrt{m}\ln\kappa)$, where $\kappa > 0$ is defined by the condition number of the constrained optimization problem, which in turn can be characterized explicitly in terms of quantities defined at the solution.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Many methods for solving constrained optimization problems rely on transforming the problem to an equivalent unconstrained one, or to a sequence of such unconstrained problems, which are usually easier to solve.

Two important classes of such methods are the so called "Lagrangian function" and "Barrier function" methods. We elucidate these ideas at the hand of the following inequality constrained problem :

$$(\mathrm{P}) \qquad \min\{f_0(x) \,|\, f_i(x) \geq 0 \quad i = 1, ..., m\}\,,$$

where $f_0(x)$ and $f_i(x) \in C^2(I\!R^n)$. A local solution to this problem will be denoted by x^*. We also define $\Omega = \{x \,|\, f_i(x) \geq 0\}$, the feasible region, $I = \{i \,|\, f_i(x^*) = 0\}$, the active constraint set and $J = \{i \,|\, f_i(x^*) > 0\}$, the passive constraint set.

Under the assumption that at a local solution x^* the gradients of the active constraint functions $(i \in I)$ are linearly independent , the first order necessary optimality conditions are given by :

$\exists\ u_i^* \geq 0\ \ (i = 1, ..., m)$ such that

$$f_i(x^*) \geq 0 \tag{1.1}$$

$$u_i^* f_i(x^*) = 0 \tag{1.2}$$

$$\nabla f_0(x^*) - \sum_{i=1}^{m} u_i^* \nabla f_i(x^*) = 0\ . \tag{1.3}$$

These conditions are usually referred to as the Karush-Kuhn-Tucker (KKT) conditions.

Note that since the gradients of the constraint functions at x^* were assumed to be linearly independent, the positive constants u_i^* associated with x^* are unique.

Before formulating the second order optimality conditions, we define the following sets :

$$\begin{aligned} T &= \{y \in I\!R^n \mid \nabla f_i(x) \cdot y \leq 0 \ \ \forall i \in I\} \\ T_1 &= \{y \in T \mid \nabla f_i(x) \cdot y = 0 \ \text{ if } \ u_i^* > 0\}\ . \end{aligned}$$

The second order necessary optimality conditions are then given by :

$$\forall v \in T_1 \ \ : \ \ v^T \left(\nabla^2 f_i(x^*) - \sum_{i=1}^{m} u_i^* \nabla^2 f_i(x^*) \right) v \geq 0\ . \tag{1.4}$$

Sufficient conditions for x^* to be an isolated local solution to (P) are given by the first order necessary conditions together with :

$$\forall v \neq 0\ \in T_1 \ \ : \ \ v^T \left(\nabla^2 f_i(x^*) - \sum_{i=1}^{m} u_i^* \nabla^2 f_i(x^*) \right) v > 0\ . \tag{1.5}$$

Lagrangian functions

Defining the "Classical Lagrangian" (CL) as the following real function on $I\!R^n \times I\!R^m$:

$$L(x,u) = \begin{cases} f_0(x) - \sum_{i=1}^m u_i f_i(x) & \text{if } u_i \geq 0 \ \ (i = 1, ..., m) \\ -\infty & \text{otherwise} \ , \end{cases}$$

the KKT first order necessary conditions can be expressed equivalently as :
$\exists \ u^* \in I\!R^m$ such that

$$L(x^*, u^*) = \max_u L(x^*, u) \text{ and } \nabla_x L(x^*, u^*) = 0 \ . \tag{1.6}$$

The positive constants u_i are the dual variables or Lagrange multipliers.

In the case of a convex problem, i.e., one where the objective function is convex and all the constraint functions concave, $L(x,u)$ is also a convex function, and the first order necessary KKT conditions are equivalent to the following saddlepoint condition for the Lagrangian :
$\exists \ u^* \in I\!R^m$ such that

$$L(x^*, u^*) = \max_u L(x^*, u) \text{ and } \min_x L(x, u^*) \ . \tag{1.7}$$

This means that knowledge of the optimal Lagrange multipliers u_i^* enables us in this case to consider the unconstrained problem $\min_x L(x, u^*)$ in stead of (P). Of course, this is only a conceptual idea since u^* is not known a priori. We also note that since this whole approach was based on necessary optimality conditions, the new unconstrained problem is not necessarily equivalent to the original constrained one. This can easily be demonstrated by the following one-dimensional minimization problem :

$$\min\{x \mid x \geq 0\} \ . \tag{1.8}$$

The pair (x^*, u^*) for this problem is given by $(0,1)$. This gives for the CL : $L(x, u^*) \equiv 0$. Clearly, minimizing $L(x, u^*)$ w.r.t. x is not equivalent to the original problem.

Such deficiencies and the limitation of (1.7) to convex problems can basically be overcome by introducing the so-called "Augmented Lagrangian" (AL), a real function on $\mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}_+$, as suggested by Rockafellar [28] :

$$\hat{L}(x,u,c) = \begin{cases} f_0(x) + \sum_{i=1}^m \left(\frac{1}{2}cf_i^2(x) - u_i f_i(x)\right) & \text{if } f_i(x) < \frac{u_i}{c} \\ f_0(x) - \sum_{i=1}^m \frac{u_i^2}{2c} & \text{otherwise .} \end{cases}$$

The constant $c \geq 0$ is chosen such that $\hat{L}(x,u,c)$ is convex in x. For $c = 0$, we obtain the CL.

Even though the AL is an improvement over the CL, it too has its drawbacks. The main motivation behind the AL is to be able to establish a saddlepoint condition like (1.7) which would also be valid for nonconvex problems. The AL almost achieves this. However (1.7), with $L(x,u)$ replaced by $\hat{L}(x,u,c)$, is not exactly equivalent to the first and second order necessary optimality conditions (see [29]). It might therefore still happen in some cases where (x^*, u^*) satisfies the necessary but not sufficient optimality conditions, that (x^*, u^*, c) is not a saddlepoint for $\hat{L}(x,u,c)$ for any c.

Furthermore, we note that the Hessian of $\hat{L}(x,u,c)$ has jump discontinuities at points x for which $f_i(x) = \frac{u_i}{c}$ for some i, which means that the AL does not have the same degree of smoothness as the original objective and constraint functions.

Computing the AL for our example (1.8), we find :

$$\hat{L}(x,u^*,c) = \hat{L}(x,1,c) = \begin{cases} \frac{1}{2}cx^2 & \text{if } x < \frac{1}{c} \\ x - \frac{1}{2c} & \text{otherwise .} \end{cases}$$

We see that in this case $\min_x\{\hat{L}(x,u^*,c)\}$, for any $c > 0$ really is equivalent to the original constrained problem. However we now have to minimize a function with discontinuous second derivatives.

The Lagrangian function concept is important theoretically and also practically, as a basis for numerical procedures. For a good overview of such methods, see for example the book by Bertsekas [2].

Barrier functions

Another important way of transforming the original problem into an unconstrained one can be obtained by using a so called "Barrier function" (BF). One such widely used function is the logarithmic BF introduced by Frisch [9] [10], which for (P) is defined as the following real function on $I\!R^n \times I\!R_+$:

$$B(x,k) = \begin{cases} f_0(x) - \frac{1}{k}\sum_{i=1}^m \ln(f_i(x)) & \text{if } x \in \mathit{int}\,\Omega \\ \infty & \text{otherwise .} \end{cases}$$

We call $x(k)$ a local solution of min $B(x,k)$. Then under suitable conditions it can be proved (see [7]) that, as $k \to \infty$, $x(k) \to x^*$, a local solution of (P).

For convex problems, $B(x,k)$ is also convex and has a global minimum, converging to the global minimum of (P) as $k \to \infty$.

The first order optimality conditions for the problem of minimizing $B(x,k)$ can be writen as follows :

$\exists\ u_i \geq 0 \ \ (i = 1, ..., m)$ such that

$$\nabla f_i(x(k)) - \sum_{i=1}^{m} u_i \nabla f_i(x(k)) = 0$$

$$u_i f_i(x(k)) = \frac{1}{k}$$

$$f_i(x(k)) > 0 \ .$$

This is clearly a perturbation of the same conditions for the original problem (P).

Another well known BF was proposed by Carroll [3] as :

$$C(x,k) = \begin{cases} f_0(x) + \frac{1}{k}\sum_{i=1}^{m}(f_i(x))^{-1} & \text{if } x \in int\,\Omega \\ \infty & \text{otherwise .} \end{cases}$$

This function, together with the logarithmic BF, was extensively studied in Fiacco and McCormick [7].

The way in which one generally proceeds to find x^* using BF's is as follows : starting with a particular $k > 0$ an approximate minimization is performed. This approximate minimum is then used as the starting point for the next minimization of the BF with an increased value for k. This process is repeated for increasing k until a suitable accuracy is reached (see [21]).

A common feature of BF's is that they grow without bound as one approaches the boundary of the feasible region, and methods based on minimizing a BF are therefore called "interior point " methods. They proceed through the interior of the feasible region towards a solution.

Barrier functions have a serious deficiency in that they are not defined at the solution of the original problem if this solution lies on the boundary of the feasible region, as it always does for linear programming problems. This causes numerical difficulties for the minimization as $k \to \infty$ and $x(k) \to x^*$.

Let us now go back to the one-dimensional problem (1.8). The logarithmic BF for this problem is given by : $x - \frac{1}{k}\ln(x)$. Minimizing w.r.t. x, one obtains $x(k) = \frac{1}{k}$. We see that as $k \to \infty$, $x \to 0 = x^*$, as required. However, since this minimization will in general be carried out numerically, the process will become increasingly ill-conditioned as we approach the solution x^*.

Modified barrier functions

The Lagrangian and barrier function concepts lead to a new idea, the theory of which was developed by Polyak in 1981 (see [24] [25] [26]), namely the "Modified barrier function" (MBF). We call $\{x \mid f_i(x) \geq -\frac{1}{k}\}$ the "extended feasible region" and denote it by Ω_k. The logarithmic MBF is then defined as the following real function on $\mathbb{R}^n \times \mathbb{R}_+^m \times \mathbb{R}_+$:

$$F(x,u,k) = \begin{cases} f_0(x) - \frac{1}{k}\sum_{i=1}^m u_i \ln(kf_i(x)+1) & \text{if } x \in \mathit{int}\,\Omega_k \\ \infty & \text{otherwise .} \end{cases}$$

Modified barrier functions can analogously be based on any other classical barrier function. However in what follows we will concentrate on the logarithmic MBF.

Observing that $\{x \mid k^{-1}\ln(kf_i(x)+1) \geq 0\} \equiv \{x \mid f_i(x) \geq 0\}$, one has that (P) is equivalent to : $\min\{f_0(x) \mid k^{-1}\ln(kf_i(x)+1) \geq 0\}$. The MBF $F(x,u,k)$ is therefore a classical Lagrangian for a problem that is equivalent to (P) !

Another way of looking at the MBF is as a barrier function on Ω_k, with different weights for different constraints. However, contrary to the classical barrier functions, the MBF's exist at x^*, the solution to (P), even if it is obtained on the boundary of the feasible set. We have that $F(x^*,u^*,k) = f_0(x^*)$ and

$$\nabla_x F(x^*,u^*,k) = \nabla f_0(x^*) - \sum_{i=1}^m u_i^* \nabla f_i(x^*) = 0 \ . \tag{1.9}$$

This means that for a convex problem, $F(x, u^*, k)$ attains its minimum at x^* for any $k > 0$. Thus, if we know u^*, then (P) can be solved by one smooth unconstrained minimization. As we saw before, this is generally not true for the CL.

For nonconvex problems, it can easily be shown (see Polyak [26]) that there exists a $k_0 > 0$ such that $F(x, u^*, k)$ is strongly convex in a neighborhood of x^* for all $k \geq k_0$. Because of this, $x^* = argmin_x\{F(x, u^*, k)\}$ remains true for the nonconvex case if $k \geq k_0$. The MBF therefore behaves very much like an augmented Lagrangian.

We now apply the MBF idea to problem (1.8). For this problem, the logarithmic MBF $F(x, u^*, k)$ is given by : $x - \frac{1}{k}\ln(kx + 1)$. Minimizing this function yields $x^* = 0$, the correct solution just as in the AL case. However, unlike in the AL case, the function to be minimized here is twice continuously differentiable.

Numerically, the procedure for finding x^* is rather different from the BF case. Here, we have two quantities controlling the process : k and u. There are two basic ways to proceed. First, starting with a particular k and Lagrange multipliers u, an approximate minimization is performed. This approximate minimum is then used both to update the Lagrange multipliers and as a starting point for the next minimization of the MBF with the same k and updated Lagrange multipliers. Under suitable conditions, these Lagrange multipliers will converge to u^*.

The second possibility is not only to update the Lagrange multipliers, but to increase k as well after each minimization. In this work, we will investigate the first method.

Complexity in Convex Optimization

Lately, barrier function techniques in constrained optimization have gained in importance after it was discovered that certain implementations of these methods have polynomial complexity when used for convex optimization and in particular for linear programming (LP).

The long established simplex method for LP, exhibits exponential complexity in the number of iterations, even though in practice it usually requires a number of iterations that is close to linear in the dimension of the problem.

Barrier function methods are not the first nor are they the only algorithms for LP with polynomial complexity. The first such method, the ellipsoid method, was developed by Yudin and Nemirovsky (1976) [32] as an outgrowth of the method of central sections of Levin (1965) [18] and Newman (1965) [20] and the method of generalized gradient descent with space dilatation of Shor (1970) [31]. The contribution of Khachian (1979) [17] was to prove that this method could be used to provide a polynomial algorithm for LP (see [5] for a good overview). However, this method turned out to be extremely slow in practice.

Another method, which was both polynomial and fast in practice was discovered by Karmarkar (1984) [15] (and, for more insight, see also [8] and [5]). This new method seemed to appear out of nowhere until 1986, when Gill et al. [11] showed that this method was in fact equivalent to a BF method. It was this discovery and new numerical linear algebra results which had been developed in the 70's and 80's, which generated renewed interest in BF techniques.

Gonzaga (1987) [12] was the first to directly prove polynomiality for a particular version of the logarithmic BF method for LP. Denoting by L the input of the problem, i.e., the number of bits necessary to specify the problem, then an accuracy of 2^{-L} is necessary to find the exact solution, i.e., to determine which vertex yields the optimal value. For this accuracy, Gonzaga proved a complexity bound of $O(mL)$ in the number of iterations, where m is the number of constraints in the problem.

Since then many authors have proved similar results for different versions of BF methods. The best complexity bound for LP at the moment is $\mathcal{O}(\sqrt{m}L)$ (see [28] [27] [30] to name but a few). For a unifying theory using the concept of "self-concordant functions", see [19].

In all these methods, one tries to stay close to the "central trajectory", the trajectory of $x(k)$ as $k \to \infty$. In order not to stray too far away from this central trajectory each time k is increased, k can only be increased by small increments. This causes those methods to be very slow in practice and k is therefore usually changed more abruptly, destroying the polynomial complexity, but yielding a faster algorithm. This has the effect that a lot of computational effort is spent in reaching a point close to the central trajectory, only to increase k drastically, move away from the central trajectory and start all over again.

Objectives of the thesis

In the MBF method with fixed k, the situation is quite different. We recall that in this method, the Lagrange multipliers are updated after each minimization, rather than k. There comes a point when, after having reached the "Kantorovich Ball" (the region in which Newton's method converges quadratically) the iterates remain in this region after the Lagrange multipliers are updated. This point will be called a 'hot start" for the method. Such a "hot start" was also observed during numerical experiments with linear programing problems at IBM's T.J. Watson Research Center by the author.

This property leads to a significant improvement in the complexity estimate. In this thesis we will determine a ball with its center at the solution of the optimization problem and such that each point in the ball is a "hot start" for the MBF method. The radius of this ball will depend on the condition number of the constrained optimization problem at the solution, as defined in Polyak [26], and estimated in chapter 2 of this thesis. We will investigate the consequences of the existence of the "hot start" on the complexity estimate for nondegenerate linear, quadratic and nonlinear convex optimization problems.

Organisation of the chapters

In the second chapter we state the basic properties of the MBF method together with some new results. In chapter 3, we will determine the "hot start" region for a problem with a convex quadratic objective function and linear constraints. Chapters 4 and 5 do the same for LP and for a general nonlinear convex problem respectively. Chapter 6 is the appendix.

Chapter 2

Basic MBF properties

2.1 Introduction

In this chapter we give an overview of the basic properties of MBF's, obtained by Polyak [26] for optimization problems of the form :

$$\text{(P)} \qquad \min \{ f_0(x) \mid f_i(x) \geq 0 \;\; i = 1, ..., m \} ,$$

where for $i = 1, ..., m$: $f_0(x)$, $-f_i(x)$ are convex and belong to $C^2(I\!R^n)$. We also prove some additional results.

The chapter will have three sections. The first section contains the main convergence theorem for the logarithmic MBF method, the second section deals with an upper bound on the condition number of (P) and in the third section we prove a few inequalities involving the Lagrange multipliers and some parameters of the MBF method, which will be needed later on.

Let us start with some general notation. As usual, a local solution to problem (P) will be denoted by x^*. Also, as before, $L(x, u)$ will be used for the Lagrangian of this problem, Ω for the feasible set and I for the set of active constraint indices. These indices will w.l.o.g. be taken to be $1, ..., r$. The set of the passive constraints is denoted by J. We define $\sigma = \min\{f_i(x^*) \mid i \in J\}$, $\theta^* = \min_{1 \leq i \leq r}\{u_i\}$ and $\rho^* = \max_{1 \leq i \leq r}\{u_i\}$. We will also use $f'_{(r)}(x)$ for the $(r \times n)$- matrix, whose i-th row is given by the gradient of the i-th active constraint and we define $f'_{(m-r)}(x)$ analogously for the passive constraints.

Throughout this thesis we shall use "LHS" for left-hand side and "RHS" for right-hand side.

2.2 The basic theorem

Before formulating the main result from Polyak [26], we will need some preliminary definitions.

Given $\epsilon, \delta, k_0 > 0$, $u^* \in I\!R^m_+$ and $0 < \epsilon < \min\{u_i^* \mid i \in I\}$, we define for $i = 1, ..., r$:

$$D_i(u^*, k_0, \delta, \epsilon) = \{(w, k) \in I\!R^2 \mid w \geq \epsilon \ , \ |w - u_i^*| \leq \delta k \ , \ k \geq k_0\} \ ,$$

and for $i = r + 1, ..., m$:

$$D_i(u^*, k_0, \delta, \epsilon) = \{(w, k) \in I\!R^2 \mid 0 \leq w \leq \delta k \ , \ k \geq k_0\} \ .$$

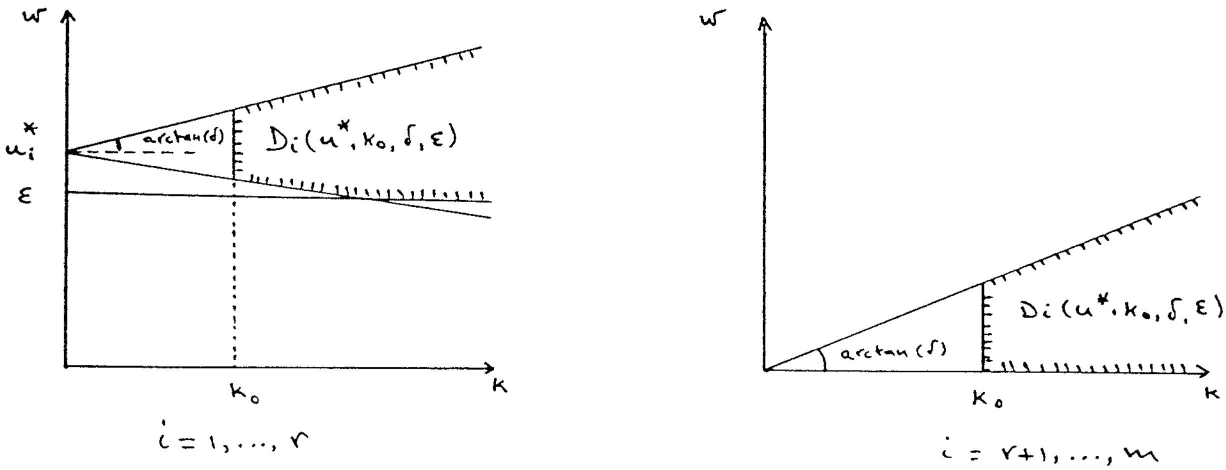
The sets D_i are in fact cut cones and are represented in figure 2.2. Setting $z = (u^*, k_0, \delta, \epsilon)$, we define the direct product of the sets we just defined as :

$$D(z) = D_1(z) \times D_2(z) \times ... D_m(z) \ .$$

We define, whenever they exist :

$$\begin{aligned} \hat{x}(u, k) &= \operatorname{argmin}\{F(x, u, k) \mid x \in I\!R^n\} \\ \hat{u}(u, k) &= \left[\operatorname{diag}(kf_i(\hat{x}) + 1)^{-1}\right]_{i=1}^m u \ , \end{aligned}$$

Figure 2.1: The sets $D_i(u^*, k_0, \delta, \epsilon)$.



where $[diag(a_i)]_{i=1}^{p}$ stands for the $p \times p$ diagonal matrix with the (i, i)-th element equal to a_i.

We note that this last expression is the update formula for the Lagrange multipliers, as will become clear in the following theorem.

We are now ready for the "Basic Theorem", which is a restatement of the part of the basic theorem in Polyak [26], pertaining to convex problems.

For the convex optimization problem (P), the following theorem holds :

Theorem 2.2.1 *(Polyak [26])*

(1) *If* $\Omega^* = \{x \in \Omega \mid f_0(x) = f_0(x^*)\}$ *is compact, then for any* $(u, k) \in I\!R_+^m \times I\!R_+$*, there exists* $\hat{x}(u, k)$ *such that* $\nabla_x F(\hat{x}, u, k) = 0$.

(2) $\hat{x}(u^*, k) = x^*$ *and* $\hat{u}(u^*, k) = u^*$ *for any* $k > 0$.

(3) *If there exists $u^* \in I\!R^m_+$ such that*

$$\nabla_x L(x^*, u^*) = 0 \;\; and \;\; f_i(x^*)u_i^* = 0$$

$$rank f'_{(r)}(x^*) = r \;\; and \;\; u_i^* > 0 \;\; for \;\; i \in I$$

$$y^T \nabla_x^2 L(x^*, u^*) y \geq \tau \|y\|^2 > 0 \;\; \forall y \neq 0 \;\; such \; that \;\; f'_{(r)}(x^*) y = 0 \; ,$$

then there exist $\delta, k_0 > 0$ such that for any $0 < \epsilon < \min_{1 \leq i \leq r} u_i^*$ and any $(u, k) \in D(u^*, k_0, \delta, \epsilon)$:

$$\max\{\|\hat{x} - x^*\|_\infty, \|\hat{u} - u^*\|_\infty\} \leq \frac{C}{k} \|u - u^*\|_\infty \; , \tag{2.1}$$

with the constant C independent of k and $F(x, u, k)$ is strongly convex in a neighborhood of $\hat{x}$.

□

An important result, and one we will use throughout this work, is the expression for the rate of convergence in this theorem. It states that the new minimum as well as the updated Lagrange multipliers are closer to the solution by a factor of $\frac{C}{k}$. A word of explanation is in order regarding the constant C. It is defined as $C = \max\{2\sigma^{-1}, c_0\}$, where $c_0 = 2\|\Phi_\mu^{-1} R_0\|_\infty$ and the matrices Φ_μ and R_0 are given by :

$$\Phi_\mu = \begin{pmatrix} L_{xx} & -f'^T_{(r)} \\ -U^* f'_{(r)} & -\mu I_r \end{pmatrix} , \; R_0 = \begin{pmatrix} 0^{(n,r)} & -f'_{(m-r)} \, [\mathrm{diag}(\sigma_i + \mu)^{-1}]_{r+1}^m \\ -I_r & 0^{(r,m-r)} \end{pmatrix} .$$

The expression for C was derived in the proof of the basic theorem in [26].

We have used L_{xx} for $\nabla_x^2 L(x^*, u^*)$, $f'_{(r)}$ for $f'_{(r)}(x^*)$, $f'_{(m-r)}$ for $f'_{(m-r)}(x^*)$, U^* for $[diag(u_i^*)]_{i=1}^r$ and μ for $\frac{1}{k}$. 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 this thesis.

The constant C is very important since it determines how small μ has to be in order to achieve convergence. In fact this constant can be considered to be the condition number of the constrained optimization problem (P). The larger it is, the smaller μ has to be and the more numerically difficult it will be to minimize the MBF.

We conclude this section with another result from Polyak [26], which we will use in later chapters. In [26] it is shown that it is possible to define an indicator function $\nu(x,u,k)$ as follows :

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

which converges to 0 at a rate of convergence slower than or equal to the rate of convergence at which (x,u) converges to (x^*,u^*).

2.3 Upper bound on C

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

$$\Phi_\mu^{-1} = \begin{pmatrix} F^{-1} & -\frac{1}{\mu} F^{-1} f_{(r)}'^T \\ \frac{1}{\mu} U^* f_{(r)}' F^{-1} & -\frac{1}{\mu}\left(I_r - \frac{1}{\mu} U^* f_{(r)}' F^{-1} f_{(r)}'^T \right) \end{pmatrix} \ ,$$

where $F = L_{xx} + \frac{1}{\mu} f'^T_{(r)} U^* f'_{(r)}$. $\Phi_\mu^{-1} R_0$ is therefore given by :

$$\Phi_\mu^{-1} R_0 = \begin{pmatrix} \frac{1}{\mu} F^{-1} f'^T_{(r)} & F^{-1} f'^T_{(m-r)} \left[\mathrm{diag}(\sigma_i + \mu)^{-1}\right]_{r+1}^{m} \\ \frac{1}{\mu}\left(I_r - \frac{1}{\mu} U^* f'_{(r)} F^{-1} f'^T_{(r)}\right) & \frac{1}{\mu} U^* f'_{(r)} F^{-1} f'^T_{(m-r)} \left[\mathrm{diag}(\sigma_i + \mu)^{-1}\right]_{r+1}^{m} \end{pmatrix} .$$

To estimate $\|\Phi_\mu^{-1} R_0\|_\infty$, we shall use the fact that for a matrix, composed of other matrices A,B,C and D :

$$\left\| \begin{pmatrix} A & B \\ C & D \end{pmatrix} \right\|_\infty \leq \max\left\{ \|A\|_\infty + \|B\|_\infty \ , \ \|C\|_\infty + \|D\|_\infty \right\} \ ,$$

and that for a matrix $M \in I\!R^{(p,s)}$:

$$\|A\|_\infty \leq \sqrt{s}\|A\| \ .$$

In order to compute an upper bound for the 2-norm of each of the four matrices in $\Phi_\mu^{-1} R_0$, we shall use three lemma's from [22], which are restated in the appendix as lemma's 6.2.1, 6.2.2 and 6.2.3. The result will be that for μ "small enough", the upper bound on each matrix will be independent of μ.

We rewrite $f'^T_{(r)} U^* f'_{(r)}$ as $\left(U^{*\frac{1}{2}} f'_{(r)}\right)^T \left(U^{*\frac{1}{2}} f'_{(r)}\right)$. With

$$\|f'^T_{(r)} y\| \geq m_0 \|y\| \quad \text{and} \quad \| \left(f'_{(r)} f'^T_{(r)}\right)^{-1} \| \leq \frac{1}{m_0^2}$$

implying

$$\|U^{*\frac{1}{2}} f'^T_{(r)} y\| \geq \sqrt{\theta^*} m_0 \|y\| \quad \text{and} \quad \| \left(f'_{(r)} U^* f'^T_{(r)}\right)^{-1} \| \leq \frac{1}{\theta^* m_0^2} \ ,$$

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

$$\frac{1}{\mu}F^{-1}f'^{T}_{(r)} = \frac{1}{\mu}\left(L_{xx} + \frac{1}{\mu}\left(U^{*\frac{1}{2}}f'_{(r)}\right)^{T}\left(U^{*\frac{1}{2}}f'_{(r)}\right)\right)^{-1}\left(U^{*\frac{1}{2}}f'_{(r)}\right)^{T}U^{*-\frac{1}{2}} .$$

Taking norms yields :

$$\left\|\frac{1}{\mu}F^{-1}f'^{T}_{(r)}\right\| \le \frac{\|U^{*-\frac{1}{2}}\|}{\mu}\left\|\left(L_{xx} + \frac{1}{\mu}\left(U^{*\frac{1}{2}}f'_{(r)}\right)^{T}\left(U^{*\frac{1}{2}}f'_{(r)}\right)\right)^{-1}\left(U^{*\frac{1}{2}}f'_{(r)}\right)^{T}\right\| .$$

From lemma's A1 and A2 we have that for

$$\frac{1}{\mu} \ge \frac{2\|L_{xx}\|\|U^{*\frac{1}{2}}f'_{(r)}\|}{\theta^{*\frac{3}{2}}m_0^3}\max\left\{\frac{\|U^{*\frac{1}{2}}f'_{(r)}\|}{\sqrt{\theta^*}m_0}\left(1+\frac{2}{l_o}+\frac{l_0}{2\|L_{xx}\|}\right) , \left(1+\frac{\|L_{xx}\|^2}{l_o^2}\right)^{\frac{1}{2}}\right\} \tag{2.2}$$

the following estimate holds :

$$\left\|\left(L_{xx} + \frac{1}{\mu}\left(U^{*\frac{1}{2}}f'_{(r)}\right)^{T}\left(U^{*\frac{1}{2}}f'_{(r)}\right)\right)^{-1}\left(U^{*\frac{1}{2}}f'_{(r)}\right)^{T}\right\| \le \frac{2\mu\|U^{*\frac{1}{2}}f'_{(r)}\|}{\theta^{*\frac{3}{2}}m_0^3}\left(1+\frac{\|L_{xx}\|^2}{l_o^2}\right)^{\frac{1}{2}} .$$

This, together with (2.3) , gives

$$\left\|\frac{1}{\mu}F^{-1}f'^{T}_{(r)}\right\| \le \frac{2\|U^{*\frac{1}{2}}f'_{(r)}\|\|U^{*-\frac{1}{2}}\|}{\theta^{*\frac{3}{2}}m_0^3}\left(1+\frac{\|L_{xx}\|^2}{l_o^2}\right)^{\frac{1}{2}} .$$

Finally,

$$\left\|\frac{1}{\mu}F^{-1}f'^{T}_{(r)}\right\| \le \alpha_1 ,$$

where we have defined

$$\alpha_1 = \frac{2\|f'_{(r)}\|\sqrt{\rho^*}}{\theta^{*2}m_0^3}\left(1+\frac{\|L_{xx}\|^2}{l_o^2}\right)^{\frac{1}{2}} .$$

(2) We now look at the second submatrix :

$$\|F^{-1}f'^{T}_{(m-r)}\left[\mathrm{diag}(\sigma_i+\mu)^{-1}\right]_{r+1}^{m}\| \le \|F^{-1}\|\|f'^{T}_{(m-r)}\left[\mathrm{diag}(\sigma_i+\mu)^{-1}\right]_{r+1}^{m}\| .$$

Lemma A1 yields :

$$\|F^{-1}f_{(m-r)}^{\prime T}\left[\operatorname{diag}(\sigma_i+\mu)^{-1}\right]_{r+1}^{m}\| \leq \frac{2}{l_0}\|F^{-1}\|(\sigma+\mu)^{-1} ,$$

and therefore :

$$\|F^{-1}f_{(m-r)}^{\prime T}\left[\operatorname{diag}(\sigma_i+\mu)^{-1}\right]_{r+1}^{m}\| \leq \alpha_2 ,$$

where

$$\alpha_2 = \frac{2}{\sigma l_0}\|f_{(m-r)}^{\prime T}\| .$$

(3) The third submatrix gives :

$$\begin{aligned}
&\left\|\frac{1}{\mu}\left(I_r - \frac{1}{\mu}U^* f_{(r)}' F^{-1} f_{(r)}^{\prime T}\right)\right\| \\
&\quad = \left\|\frac{U^{*\frac{1}{2}}}{\mu}\left(I_r - \frac{1}{\mu}\left(U^{*\frac{1}{2}}f_{(r)}'\right)^T F^{-1}\left(U^{*\frac{1}{2}}f_{(r)}'\right)^T\right)U^{*-\frac{1}{2}}\right\| \\
&\quad \leq \frac{\|U^{*\frac{1}{2}}\|\|U^{*-\frac{1}{2}}\|}{\mu}\left\|I_r - \frac{1}{\mu}\left(U^{*\frac{1}{2}}f_{(r)}'\right)^T F^{-1}\left(U^{*\frac{1}{2}}f_{(r)}'\right)^T\right\| . \qquad (2.3)
\end{aligned}$$

For μ satisfying (2.2), lemma A3 gives

$$\left\|I_r - \frac{1}{\mu}\left(U^{*\frac{1}{2}}f_{(r)}'\right)^T F^{-1}\left(U^{*\frac{1}{2}}f_{(r)}'\right)^T\right\| \leq \frac{2\mu\|L_{xx}\|\|U^{*\frac{1}{2}}f_{(r)}'\|^2}{(\sqrt{\theta^*}m_0)^5}\left(1+\frac{\|L_{xx}\|^2}{l_o^2}\right)^{\frac{1}{2}} .$$

This, together with (2.3), yields

$$\left\|\frac{1}{\mu}\left(I_r - \frac{1}{\mu}U^* f_{(r)}' F^{-1} f_{(r)}^{\prime T}\right)\right\| \leq \alpha_3 ,$$

with

$$\alpha_3 = \left(\frac{\rho^*}{\theta^*}\right)^{\frac{3}{2}}\frac{\|L_{xx}\|\|f_{(r)}'\|^2}{\theta^{*\frac{3}{2}}m_0^5}\left(1+\frac{\|L_{xx}\|^2}{l_o^2}\right)^{\frac{1}{2}} .$$

(4) Finally, we consider the fourth submatrix :

$$
\begin{aligned}
&\left\| \frac{1}{\mu} U^* f'_{(r)} F^{-1} f'^T_{(m-r)} \left[\operatorname{diag}(\sigma_i + \mu)^{-1} \right]_{r+1}^{m} \right\| \\
&\quad \le \left\| \frac{1}{\mu} U^* f'_{(r)} F^{-1} \right\| \left\| f'^T_{(m-r)} \right\| (\sigma + \mu)^{-1} \\
&\quad \le \left\| \frac{1}{\mu} f'_{(r)} F^{-1} \right\| \left\| f'^T_{(m-r)} \right\| \|U^*\| \, \sigma^{-1} \\
&\quad \le \left\| \frac{1}{\mu} \left(\left(F^{-1}\right)^T f'^T_{(r)} \right)^T \right\| \left\| f'^T_{(m-r)} \right\| \|U^*\| \, \sigma^{-1} \\
&\quad \le \left\| \frac{1}{\mu} F^{-1} f'^T_{(r)} \right\| \left\| f'^T_{(m-r)} \right\| \|U^*\| \, \sigma^{-1} \\
&\quad \le \frac{\rho^*}{\sigma} \left\| f'^T_{(m-r)} \right\| \alpha_1 \\
&\quad \le \alpha_4 \alpha_1 \ ,
\end{aligned}
$$

with

$$\alpha_4 = \frac{\rho^*}{\sigma} \left\| f'^T_{(m-r)} \right\| \ .$$

Since $c_0 \le 2 \left\| \Phi_\mu^{-1} R_0 \right\|_\infty$, we have obtained that

$$c_0 \le 2\sqrt{n} \max\{\alpha_1 + \alpha_2 \ , \ \alpha_3 + \alpha_1 \alpha_4\} \ . \tag{2.4}$$

2.4 Some inequalities for MBF parameters

In this section 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.

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

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

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

Lemma 2.4.1 *Let $\|u^{(0)} - u^*\|_\infty \leq \omega$ for some $\omega > 0$ and let $\gamma < 1$ be the rate of convergence, associated with $\mu > 0$, as in the basic convergence result (2.1).*

Furthermore, let the constraint functions satisfy the following Lipschitz condition with $L > 0$:

$$\forall x, y \in \Omega_k \ , \ \forall i \ : \ |f_i(x) - f_i(y)| \leq L\|x - y\| \ .$$

Then the following inequalities hold :

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

$$u_i^{(s)} \geq exp\left(-\left(\frac{1-\gamma^s}{1-\gamma}\right)\frac{\gamma L \omega \sqrt{n}}{\mu}\right) u_i^{(0)} \ . \tag{2.5}$$

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

$$u_i^{(s)} \geq exp\left(-\frac{s\sigma_i}{\mu}\right) \ exp\left(-\left(\frac{1-\gamma^s}{1-\gamma}\right)\frac{\gamma L \omega \sqrt{n}}{\mu}\right) u_i^{(0)} \ , \tag{2.6}$$

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

Proof : (1) According to the update formula for the Lagrange multipliers, we have for the active constraints ($i \in I$) :

$$u_i^{(1)} = \frac{\mu u_i^{(0)}}{f_i(x^{(1)}) + \mu} \ = \ \frac{\mu u_i^{(0)}}{f_i(x^{(1)}) - f_i(x^*) + \mu}$$

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

Continuing to update in this way, we obtain

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

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

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

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 $(i \in J)$:

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

Continuing to update in this way, we obtain

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

Again, summing the geometric series completes the proof.

□

We now use this lemma to prove the following theorem.

Theorem 2.4.1 *Under the same conditions as in the previous lemma, the following two inequalities hold :*

(1)

$$\mu \leq \left(\frac{\gamma L\sqrt{n}}{1-\gamma}\right)\theta^* \tag{2.7}$$

(2)

$$\mu \leq \frac{\sigma}{|\ln\gamma|} . \tag{2.8}$$

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

$$u_i^{(0)} = u_i^* + \omega \qquad i = 1, ..., m . \tag{2.9}$$

Then, taking the limit for $s \to \infty$ in (2.5) , we obtain :

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

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

$$u_i^* \geq \exp\left(-\frac{\gamma L\omega\sqrt{n}}{\mu(1-\gamma)}\right)u_i^{(0)} . \tag{2.10}$$

We now pick an index $\bar{\imath}$ for which $u_{\bar{\imath}}^{(0)}$ converges to θ^*. For this index, (2.10) becomes :

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

and therefore

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

or :

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

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

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

The first part of the theorem follows immediately.

(2) Turning to the passive constraints, our choice of $u^{(0)}$ in (2.9) means that

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

We now pick an index $\bar{\jmath}$ for which $\sigma_{\bar{\jmath}} = \sigma = \min\{\sigma_i \mid i \in J\}$. The basic convergence result in theorem 2.2.1 then gives

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

With the previous lemma we can therefore write :

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

This gives :

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

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

$\square$

Chapter 3

The quadratic case

3.1 Introduction

(i) In this chapter we shall consider the MBF method for solving the following problem :

$$\text{(QP)} \qquad \begin{array}{ll} \min & \frac{1}{2}x^T Q x - s^T x \\ \text{s.t.} & r_i(x) \geq 0 \qquad i = 1, ..., m \end{array} \quad ,$$

where $Q \in I\!R^{(n,n)}$ is positive semi-definite and :

$$\begin{array}{l} r_i(x) = a_i^T x - b_i \\ x, s, a_i \in I\!R^n \\ b_i \in I\!R \\ \|a_i\| = 1. \end{array}$$

Because of this last assumption, which does not cause any loss of generality, we have the following Lipschitz condition :

$$\forall x, y \in I\!R^n \ : \ |r_i(x) - r_i(y)| \leq \|x - y\|. \tag{3.1}$$

We assume that *int* $\{x \mid r_i(x) \geq 0\}$ is nonempty and bounded, and that all conditions in the basic theorem 2.2.1 are satisfied. We are therefore assuming that the problem is nondegenerate.

(ii) The MBF for this problem is given by :

$$\frac{1}{2}x^TQx - s^Tx - \frac{1}{k}\sum_{i=1}^{m} u_i \ln(kr_i(x) + 1) ,$$

which can be rewritten as follows :

$$\frac{1}{2}x^TQx - s^Tx - \frac{1}{k}\sum_{i=1}^{m} u_i \left(\ln\left(r_i(x) + \frac{1}{k}\right) + \ln k\right) .$$

Setting $\mu = \frac{1}{k}$, this becomes

$$\frac{1}{2}x^TQx - s^Tx - \mu\sum_{i=1}^{m} u_i \ln(r_i(x) + \mu) + \mu \ln \mu \sum_{i=1}^{m} u_i .$$

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

Definition 3.1.1 *The Modified Barrier Function for QP is defined as*

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

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. As a result, 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

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

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

The Hessian $\nabla_x^2\phi(x,u,\mu)$ is strictly positive definite on the extended feasible set (see [14] [13]) and, as a result of (3) in the basic theorem, its smallest eigenvalue is bounded away from zero in a neighborhood of the minimum of $\phi(x,u,\mu)$ for fixed (u,k).

(iii) We now state some assumptions about MBF parameters. Throughout this chapter we shall assume that (u,k) lies in the cut cone as described in the basic theorem and we shall also assume that μ is small enough to obtain a suitable γ, the parameter responsible for the rate of convergence in the basic theorem. As was mentioned in chapter 2, this can be achieved by reducing μ and then checking an indicator function $\nu(x,u,\mu)$.

In the algorithm, we minimize $\phi(x,u,\mu)$ for fixed u and μ up to a certain accuracy ϵ_1, before updating u. When we are far away from this minimum, we perform a linesearch along the Newton direction. The Newton direction $p(x,u,\mu)$ at a given point x for fixed u and μ is given by :

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

Whenever there can be no confusion, we shall write p,g,H in stead of $p(x,u,\mu), \nabla_x\phi(x,u,\mu), \nabla_x^2\phi(x,u,\mu)$ respectively.

We stop doing linesearches at a point where $\|p\|_H \leq \beta$, for a certain β which will be determined later. This β depends on x, but is bounded away from zero, so that eventually this inequality will be satisfied. From this point on, the algorithm takes full Newton steps, without linesearch, until the aforementioned accuracy is reached.

Thus the algorithm contains an inner iteration, where $\phi(x, u, \mu)$ is approximately minimized, and an outer iteration, where the Lagrange multipliers are updated. The inner iteration contains two phases. In the first phase, a linesearch is performed at each iteration whereas in the the second phase full Newton steps are taken without a linesearch.

We are now ready to formally state the algorithm.

Input :

- $(x^{(0)}, u^{(0)}, \mu)$ satisfying the conditions of the basic theorem;
- μ such that the rate of convergence γ is at least $\frac{1}{2\sqrt{n}}$;
- D_0 and $\tilde{\epsilon}$ are accuracy parameters;

begin

$x := x^{(0)}$; $u := u^{(0)}$; $D := 1$

while $D > D_0$ **do**

begin (begin outer iteration)

while $\|p\|_H > \tilde{\epsilon}$ **do**

begin (begin inner iteration)

while $\|p\|_H > \beta$ **do**

begin

$\bar{\ell} := argmin_{\ell>0}\{\phi(x + \ell p, u, \mu) \mid x + \ell p \in int\ \mathcal{F}_1\}$

$x := x + \bar{\ell} p$

end

$x := x + p$

end (end inner iteration)

$\forall i \;:\; u_i := \frac{u_i \mu}{r_i(x)+\mu}$; $D := D\gamma$

end (end outer iteration)

end

The parameter D_0 is chosen in such a way that the algorithm terminates with $\max\{\|x - x^*\|_\infty, \|u - u^*\|_\infty\} \leq \epsilon$, where ϵ is the final accuracy we want to obtain.

(iv) The aim of this chapter is to show that for (QP) , there comes a point, a so-called "hot start", from which all iterates remain well-defined in primal and dual space, i.e., the iterates remain in the Kantorovich ball for the new function after each Lagrange multiplier update and the pairs (u, k) remain in the cut cone as defined in the basic theorem. We call "Kantorovich ball" the region in which $\|p\|_H$ converges to zero quadratically, after Kantorovich's results in this area (see [15]). This point will be characterized in terms of μ and other quantities which depend on the solution.

In order to do this we will first compute $\beta(x)$, which determines the Kantorovich ball. We will then show that $\beta(x)$ is bounded from below by a positive number β^* in a neighborhood of the solution (x^*, u^*). Finally we characterize the region where $\|p\|_H$ falls below this value, and continues to be below this value for all subsequent iterations. Any point in this area will therefore be a "hot start".

3.2 Basic lemmas for the inner iteration

In this section we shall prove some basic lemmas that will be needed to determine β, which was mentioned in the introduction. We start with some notation :

$$
\begin{aligned}
I &= \{i \mid r_i(x)\ \mathit{is\ active}\} \\
J &= \{i \mid r_i(x)\ \mathit{is\ passive}\} \\
I_\mu &= \{i \mid u_i \geq \frac{\mu}{2}\} \\
J_\mu &= \{i \mid u_i < \frac{\mu}{2}\} \\
|J_\mu| &= \mathit{the\ number\ of\ elements\ in}\ J_\mu \\
\theta &= \min\{u_i \mid i \in I_\mu\} \\
\rho &= \max\{u_i\} \\
\eta &= |J_\mu| \max\{u_i \mid i \in J_\mu\} \\
\eta_1 &= \max\{u_i \mid i \in J_\mu\} \\
\lambda_H &= \mathit{the\ smallest\ eigenvalue\ of}\ H \\
\Omega_k &= \{x \mid \forall i\ :\ r_i(x) + \mu \geq 0\}\ .
\end{aligned}
$$

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

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

We will frequently use the H-norm $\|.\|_H$, defined as

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

Because H is positive definite, $\|.\|_H$ defines a norm.

As mentioned in the introduction to this chapter, we want to be able to determine the region in which full Newton steps can be taken. In order to do this, we have to be able to determine under what conditions, for $x \in int\ \Omega_k$ and $d \in I\!R^n$, $x + d$ still lies in $int\ \Omega_k$.

For the inner iteration, we also have to find a way of measuring the error in the approximation to the minimum of the MBF, since this error will have an influence, among other things, on the update of the Lagrange multipliers.

The following two lemmas deal with these issues. In the first one, we determine a condition on d for $x + d$ to lie in $int\ \Omega_k$ 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 3.2.1 *If $x \in int\ \Omega_k$ and $\|d\|_H < \beta_1(x)$, then $x + d \in int\ \Omega_k$. 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 :

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

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

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

Proof : The proof of the lemma will be along the lines of the proof of lemma 2.1 in [14]. Expanding $\phi(x+d,u,\mu)$ in a Taylor series about x, we can write :

$$\phi(x+d,u,\mu)=\phi(x,u,\mu)+d^T(\nabla_x\phi(x))+\frac{1}{2}d^T(\nabla_{xx}\phi(x))d+\sum_{k=3}^{\infty}t_k\ , \quad (3.2)$$

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 k-th order term in the Taylor expansion:

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

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

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

$$\begin{aligned}\sum_{i=1}^{m}u_i\chi_i^k &\le \theta\sum_{i\in I_\mu}\left(\left(\frac{u_i}{\theta}\right)^{\frac{2}{k}}\chi_i^2\right)^{\frac{k}{2}}+\sum_{i\in J_\mu}u_i\chi_i^k\\ &\le \theta\sum_{i\in I_\mu}\left(\left(\frac{u_i}{\theta}\right)\chi_i^2\right)^{\frac{k}{2}}+\sum_{i\in J_\mu}u_i\chi_i^k\\ &\le \theta\left(\sum_{i\in I_\mu}\left(\frac{u_i}{\theta}\right)\chi_i^2\right)^{\frac{k}{2}}+\eta_1\sum_{i\in J_\mu}\chi_i^k\end{aligned}$$

$$
\begin{aligned}
&\le \theta\left(\frac{d^TQd}{\mu\theta}+\sum_{i\in I_\mu}\left(\frac{u_i}{\theta}\right)\chi_i^2\right)^{\frac{k}{2}}+\eta_1\sum_{i\in J_\mu}\chi_i^k \\
&\le \theta\left(\frac{\|d\|_H^2}{\theta}\right)^{\frac{k}{2}}+\eta_1\sum_{i\in J_\mu}\chi_i^k.
\end{aligned}
$$

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

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

and therefore

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

This means that

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

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 (3.2), this difference is given by $\sum_{k=1}^{\infty}t_k$, so we compute :

$$
\begin{aligned}
\left|\sum_{k=3}^{\infty}t_k\right| &\le \sum_{k=3}^{\infty}\frac{1}{k}\left\{\sum_{i=1}^m u_i\chi_i^k\right\} \\
&\le \frac{\theta}{3}\left(\frac{\|d\|_H}{\sqrt{\theta}}\right)^3\left(1+\frac{\|d\|_H}{\sqrt{\theta}}+\dots\right)+\frac{\eta}{3}\left(\frac{\|d\|_H}{R_1(x)}\right)^3\left(1+\frac{\|d\|_H}{R_1(x)}+\dots\right) \\
&\le \left(\frac{\theta^{-\frac{1}{2}}}{3}\right)\frac{\|d\|_H^3}{1-(\|d\|_H/\sqrt{\theta})}+\left(\frac{\eta}{3R_1^3(x)}\right)\frac{\|d\|_H^3}{1-(\|d\|_H/R_1(x))}.
\end{aligned}
$$

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_k.

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

$$\begin{aligned}
\left|\sum_{k=3}^{\infty} t_k\right| &\leq \frac{2}{3}\left(\frac{1}{\sqrt{\theta}} + \frac{\eta}{R_1^3(x)}\right)\|d\|_H^3 \\
&\leq \frac{2}{3}\left(\frac{1}{\sqrt{\theta}} + \frac{\eta/R_1^2(x)}{R_1(x)}\right)\|d\|_H^3 \\
&\leq \frac{1}{3\xi_1(x)}\|d\|_H^3 .
\end{aligned}$$

□

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 [14]. We recall that p denotes the Newton direction.

Lemma 3.2.2 *If*

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

then

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

Proof :

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

$$\|p + h\|_H \leq \frac{5}{2}\|p\|_H.$$

Since $\xi_1(x) \leq \sqrt{\theta}/2$ and $\|p\|_H \leq \xi_1(x)/5$, $\|p\|_H$ satisfies the conditions of lemma 3.2.1 . Using this lemma and the fact that $p = argmin_y q_x(y,u,\mu)$ we have

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

The previous lemma yields :

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

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 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 :

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

□

As a complement to the previous lemma, we also prove the following result, similar to lemma 5 in [4]. 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 3.2.3 *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

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

Proof :

Let ℓ be a steplength such that

$$\|\ell p\|_H \leq \beta_2(x).$$

Then from lemma 3.2.1 we have

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

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

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

Taking for ℓ the value $\frac{\beta_2(x)}{\|p\|_H}$ gives

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

□

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 3.2.7.

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 3.2.1

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

Lemma 3.2.4 *If*

$$\|x-y\| \leq \frac{B(x)}{\alpha} \quad (\alpha > 1)\,, \tag{3.3}$$

then

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

Proof :

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

On the other hand we have

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

This completes the proof.

Lemma 3.2.5 *If*

$$\|x-y\| \le \frac{B(x)}{\alpha} \quad (\alpha > 1)\,, \tag{3.4}$$

then

$$\left(\frac{\alpha}{\alpha+1}\right)^2 d^T H(x) d \le d^T H(y) d \le \left(\frac{\alpha}{\alpha-1}\right)^2 d^T H(x) d\,. \tag{3.5}$$

Proof : Using the previous lemma, we can write

$$\begin{aligned} d^T H(y) d &= \frac{d^T Q d}{\mu} + \sum_{i=1}^{m} u_i \frac{\left(a_i^T d\right)^2}{(r_i(y)+\mu)^2} \\ &\le \frac{d^T Q d}{\mu} + \left(\frac{\alpha}{\alpha-1}\right)^2 \sum_{i=1}^{m} u_i \frac{\left(a_i^T d\right)^2}{(r_i(x)+\mu)^2} \\ &\le \left(\frac{\alpha}{\alpha-1}\right)^2 d^T H(x) d\,. \end{aligned}$$

On the other hand we have

$$\begin{aligned} d^T H(y) d &= \frac{d^T Q d}{\mu} + \sum_{i=1}^{m} u_i \frac{\left(a_i^T d\right)^2}{(r_i(y)+\mu)^2} \\ &\ge \frac{d^T Q d}{\mu} + \left(\frac{\alpha}{\alpha+1}\right)^2 \sum_{i=1}^{m} u_i \frac{\left(a_i^T d\right)^2}{(r_i(x)+\mu)^2} \\ &\ge \left(\frac{\alpha}{\alpha+1}\right)^2 d^T H(x) d\,. \end{aligned}$$

This completes the proof.

□

Lemma 3.2.6 *For $x, x+p \in int\ \Omega_k$ with $\|p\| \le \frac{B(x)}{\alpha}$ $(\alpha > 1)$ the following inequality holds :*

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

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

From the definition of p this means

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

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

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

Because $|(a_i)_j| \le \|a_i\| \le 1$ and $\|x+\zeta p - x\| \le \|p\|$ and because of the assumption on $\|p\|$, we can use lemmas 3.2.4 and 3.2.5. This gives

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

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

$$\begin{aligned}
&\leq \frac{\alpha}{(\alpha-1)\,B(x)}\sum_{i=1}^{m} u_i \frac{\left(a_i^T p\right)^2}{\left(r_i(x+\zeta p)+\mu\right)^2} \\
&\leq \frac{\alpha}{(\alpha-1)\,B(x)} p^T H(x+\zeta p)p \\
&\leq \frac{\alpha^3}{(\alpha-1)^3\,B(x)} p^T H(x)p \,.
\end{aligned}$$

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.

□

The next lemma will determine the rate of convergence for the norms of the Newton directions in the inner iteration.

Lemma 3.2.7 *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 Ω_k and $\|p\| \leq \frac{B(x)}{\alpha}$* $(\alpha > 1)$.

Then :

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

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

$$\begin{aligned} q^T \overline{H}^2 q &= (\overline{H}^{\frac{1}{2}} q)^T \overline{H} (\overline{H}^{\frac{1}{2}} q) \\ &\ge \lambda_{\overline{H}} \|\overline{H}^{\frac{1}{2}} q\|^2 \\ &\ge \lambda_{\overline{H}} \|q\|_{\overline{H}}^2 . \end{aligned}$$

Therefore we can write

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

which gives

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

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

□

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

Definition 3.2.2

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

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

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

$$\|p(x,u,\mu)\|_{H(x,u,\mu)} \leq \beta_3(x,\alpha,\bar{\alpha}) \;, \tag{3.6}$$

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 :

$$\begin{aligned} &(1) \quad y \in int\,\Omega_k \\ &(2) \quad \|p(y,u,\mu)\|_{H(y,u,\mu)} \leq \beta_3(y,\alpha,\bar{\alpha}) \;. \end{aligned}$$

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 $\beta_3(x,\alpha,\bar{\alpha})$. The next lemma will provide this.

We will use the following notation :

$$\begin{aligned} p &= p(x,u,\mu) \\ \bar{x} &= x+p \\ q &= p(\bar{x},u,\mu) \\ \lambda_H &= \textit{the smallest eigenvalue of } H(x,u,\mu) \\ \lambda_{\overline{H}} &= \textit{the smallest eigenvalue of } H(\bar{x},u,\mu) \;. \end{aligned}$$

Lemma 3.2.8 *If*

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

then

$$\begin{aligned}
&(1) \quad \left(\frac{\alpha-1}{\alpha}\right) B(x) \le B(\bar{x}) \le \left(\frac{\alpha+1}{\alpha}\right) B(x)\\
&(2) \quad \left(\frac{\alpha}{\alpha+1}\right)^2 \le \frac{\lambda_{\overline{H}}}{\lambda_H} \le \left(\frac{\alpha}{\alpha-1}\right)^2\\
&(3) \quad \left(\frac{\alpha-1}{\alpha+1}\right) R_1(x) \le R_1(\bar{x}) \le \left(\frac{\alpha+1}{\alpha-1}\right) R_1(x)\\
&(4) \quad \left(\frac{\alpha-1}{\alpha+1}\right)^3 \xi_1(x) \le \xi_1(\bar{x}) \le \left(\frac{\alpha+1}{\alpha-1}\right)^3 \xi_1(x)\,.
\end{aligned}$$

Proof :

We start by noting that

$$\|x-\bar{x}\| = \|p\| \le \frac{1}{\sqrt{\lambda_H}}\|p\|_H\,. \tag{3.7}$$

(1) Because of (3.7), the first part of the proof follows immediately from lemma 3.2.4 and the definition of $B(x)$.

(2) Again, because of (3.7), the assumption in the statement of the lemma means that

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

Lemma 3.2.5 then gives

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

and therefore

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

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

(3) From lemma 3.2.4 and from the definition of $R_1(x)$ we have that

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

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

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

With lemma 3.2.5, the proof of part three 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.

□

We are now ready for the main results of this section. In the following result we will use the previous lemma and lemma 3.2.4 to find a relation between α and $\bar{\alpha}$ so that (3.6) will be satisfied at the point $\bar{x}$ also.

Lemma 3.2.9 *If* $\|p\|_H \leq \beta_3(x, \alpha, \bar{\alpha})$ *with*

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

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

Proof :

From lemma 3.2.7 we have that :

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

Now, we assumed that

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

so from lemmas 3.2.8 and 3.2.5 we have

$$\begin{aligned}
R_1(x) &\leq \left(\frac{\alpha+1}{\alpha-1}\right) R_1(\bar{x}) \\
\xi_1(x) &\leq \left(\frac{\alpha+1}{\alpha-1}\right)^3 \xi_1(\bar{x})
\end{aligned}$$

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

and therefore

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

We have obtained that

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

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

□

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

Definition 3.2.3

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

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

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

Theorem 3.2.1 (1) *If* $\|p\|_H \le \beta(x)$, *then*

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

(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* $\Delta\phi > \frac{2}{5}\beta^2(x)$.

□

3.3 The Lagrange multiplier update

The expression for the rate of convergence of the algorithm in theorem 2.2.1 was derived under the assumption that the minimization w.r.t. x of $\phi(x, u, \mu)$ (fixed u and μ) is carried out exactly. This exact minimum is then used to update the Lagrange multipliers.

However, since in practice this is impossible, we will have to investigate the effect of inexact minimization. There are in fact two problems to consider. First we must decide on the accuracy we want to achieve and secondly we have to be able to measure this accuracy.

The answer to both questions is given by the following lemma. In this lemma we assume that we are close enough to the minimum so that we can apply lemma 3.2.2 to measure the distance to the minimum in terms of $\|p\|_H$.

We also recall that the accuracy required of the solution to (QP) was given by ϵ, i.e., we want to obtain

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

In the lemma we will now prove, we will use the following notation :

$\bar{u}_i$: *old multipliers*

$$
\begin{array}{rcl}
\hat{u}_i & : & \textit{exact new multipliers} \\
\bar{\rho} & : & \max_i \{\bar{u}_i\} \\
u_i & : & \textit{approximate new multipliers} \\
x & : & \textit{approximate minimum of } \phi(x,\bar{u},\mu) \\
\hat{x} & : & \textit{exact minimum of } \phi(x,\bar{u},\mu)\ .
\end{array}
$$

Lemma 3.3.1 *Let $\|\bar{u}-u^*\|_\infty > \epsilon$ and let μ be such that if exact minimization of $\phi(x,\bar{u},\mu)$ were to be performed, the convergence rate would be given by $\frac{1}{4\sqrt{n}}$.*

If

$$\|p\|_H \leq \frac{\epsilon\sqrt{\lambda_H}}{20\sqrt{n}} \min\left\{2, \frac{B^2(x)}{\bar{\rho}\mu}, \beta(x)\right\} ,$$

then

$$\max\{\|x-x^*\|_\infty, \|u-u^*\|_\infty\} \leq \frac{1}{2\sqrt{n}}\|\bar{u}-u^*\|_\infty\ .$$

Proof :

We start by writing :

$$
\begin{aligned}
\|u-u^*\|_\infty &\leq \|u-\hat{u}\|_\infty + \|\hat{u}-u^*\|_\infty \\
&\leq \|u-\hat{u}\|_\infty + \frac{1}{4\sqrt{n}}\|\bar{u}-u^*\|_\infty\ .
\end{aligned}
\tag{3.8}
$$

Now, for all i we have

$$
\begin{aligned}
|u_i-\hat{u}_i| &= \left|\frac{\bar{u}_i\mu}{r_i(x)+\mu} - \frac{\bar{u}_i\mu}{r_i(\hat{x})+\mu}\right| \\
&= \frac{\bar{u}_i\mu|r_i(x)-r_i(\hat{x})|}{(r_i(x)+\mu)(r_i(\hat{x})+\mu)}
\end{aligned}
$$

$$\leq \frac{\bar{u}_i \mu \|x - \hat{x}\|}{(r_i(x) + \mu)(r_i(\hat{x}) + \mu)}, \tag{3.9}$$

and since $\|p\|_H \leq \beta(x)$, lemma 3.2.2 yields

$$\|x - \hat{x}\| \leq \frac{1}{\sqrt{\lambda_H}} \|x - \hat{x}\|_H \leq \frac{5}{2\sqrt{\lambda_H}} \|p\|_H \,. \tag{3.10}$$

Substituting this back into (3.9) gives

$$|u_i - \hat{u}_i| \leq \frac{5\bar{u}_i \mu \|p\|_H}{2\sqrt{\lambda_H}(r_i(x) + \mu)(r_i(\hat{x}) + \mu)} \,. \tag{3.11}$$

With the definition of $\beta(x)$ we also have from (3.10) that

$$\|x - \hat{x}\| \leq \frac{5}{2\sqrt{\lambda_H}} \frac{B(x)\sqrt{\lambda_H}}{6\sqrt{n}} \leq \frac{B(x)}{2} \,. \tag{3.12}$$

From lemma 3.2.4 we therefore have

$$\frac{1}{r_i(\hat{x}) + \mu} \leq \frac{2}{r_i(x) + \mu} \,.$$

Substituting this back into (3.11) yields

$$|u_i - \hat{u}_i| \leq \frac{5\bar{u}_i \mu \|p\|_H}{\sqrt{\lambda_H}(r_i(x) + \mu)^2} \,.$$

The assumption on $\|p\|_H$ in the statement of the lemma then gives

$$\|u - \hat{u}\|_\infty \leq \frac{\epsilon}{4\sqrt{n}} \leq \frac{1}{4\sqrt{n}} \|\bar{u} - u^*\|_\infty \,.$$

Finally, using this last inequality in (3.8), we obtain

$$\|u - u^*\|_\infty \leq \frac{1}{2\sqrt{n}} \|\bar{u} - u^*\|_\infty \,.$$

On the other hand, recalling that

$$\|x - \hat{x}\|_\infty \le \|x - \hat{x}\| \le \frac{1}{\sqrt{\lambda_H}}\|x - \hat{x}\|_H \le \frac{5}{2\sqrt{\lambda_H}}\|p\|_H ,$$

and with the assumption on $\|p\|_H$, we also have

$$\begin{aligned}
\|x - x^*\|_\infty &\le \|x - \hat{x}\|_\infty + \|\hat{x} - x^*\|_\infty \\
&\le \frac{\epsilon}{4\sqrt{n}} + \frac{1}{4\sqrt{n}}\|\bar{u} - u^*\|_\infty \\
&\le \frac{1}{4\sqrt{n}}\|\bar{u} - u^*\|_\infty + \frac{1}{4\sqrt{n}}\|\bar{u} - u^*\|_\infty \\
&\le \frac{1}{2\sqrt{n}}\|\bar{u} - u^*\|_\infty .
\end{aligned}$$

This completes the proof.

□

Definition 3.3.1 *We define the quantity $\epsilon_1(x)$, used to determine the accuracy of the minimization w.r.t. x of $\phi(x, u, \mu)$ (fixed u and μ) :*

$$\epsilon_1(x) \triangleq \frac{\epsilon\sqrt{\lambda_H}}{20\sqrt{n}} \min\left\{2, \frac{B^2(x)}{\bar{\rho}\mu}\right\} . \tag{3.13}$$

Note : In practice, we minimize up to accuracy ϵ_1, with μ small enough to achieve a convergence rate of $\frac{1}{2\sqrt{n}}$. This lemma then shows that in the worst case, μ should be such that for exact minimization of $\phi(x, u, \mu)$, the convergence rate would be $\frac{1}{4\sqrt{n}}$.

3.4 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 3.4.1

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

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 first lemma gives bounds on the Lagrange multipliers and on the quantity $r_i(x) + \mu$ in S_1.

Lemma 3.4.1 *Let $(x, u) \in S_1$ and let the rate of convergence of the algorithm be given by $\gamma = \frac{1}{2\sqrt{n}}$.*

Then

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

$$\frac{1}{2}u_i^* < u_i < \frac{3}{2}u_i^* \tag{3.14}$$

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

and

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

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

Proof :

(1) From theorem 2.4.1 with $\gamma = \frac{1}{2\sqrt{n}}$, and recalling that $L = 1$ (L was the Lipschitz constant for the constraints), we have that $\mu \leq \theta^*$. Therefore, for any u such that

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

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

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

We now turn to the second set of inequalities. Since

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

we can write

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

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

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

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

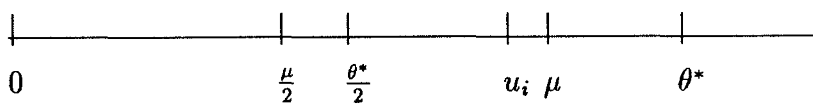
$$\begin{aligned} r_i(x) + \mu &= r_i(x) - r_i(x^*) + r_i(x^*) + \mu \\ &\geq \sigma + \mu - |r_i(x) - r_i(x^*)| \\ &\geq \sigma + \mu - \|x - x^*\| \, . \end{aligned}$$

Again, the proof follows from the asumption that $\|x - x^*\| < \frac{\mu}{2}$.

$\square$

We can use the results from the lemma we just proved to draw the following picture.

Figure 3.1: Location of Lagrange multipliers.



It shows that for $(u, \mu) \in S_1$, 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 $\frac{\theta^*}{2}$.

We will now have a look at the eigenvalues of the Hessian 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 3.4.2 *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 :*

$$\frac{2}{9}\lambda^* \leq \lambda_H \leq 6\lambda^* + \frac{\mu(m - r)}{2\sqrt{n}\left(\sigma + \frac{\mu}{2}\right)^2}$$

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

Proof :

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

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

and therefore

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

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 :

$$\begin{aligned} \min_{\|d\|=1} \left\{ \frac{d^T Q d}{\mu} + \sum_{i=1}^{m} u_i \frac{(a_i^T d)^2}{(r_i(x)+\mu)^2} \right\} &\ge \min_{\|d\|=1} \left\{ \frac{d^T Q d}{\mu} + \sum_{i=1}^{r} u_i \frac{(a_i^T d)^2}{(r_i(x)+\mu)^2} \right\} \\ &\ge \frac{2}{9} \min_{\|d\|=1} \left\{ \frac{d^T Q d}{\mu} + \sum_{i=1}^{r} u_i^* \frac{(a_i^T d)^2}{\mu^2} \right\} . \end{aligned}$$

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

Similarly, we have for the upper bounds :

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

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

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

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

□

We now define the following two quantities :

Definition 3.4.2

$$
\begin{aligned}
\tilde{\lambda} &\triangleq \frac{2}{9} \lambda^* \, . \\
\tilde{\Lambda} &\triangleq 6\Lambda^* + \frac{\mu(m-r)}{2\sqrt{n}\left(\sigma + \frac{\mu}{2}\right)^2} \, .
\end{aligned}
$$

Here, λ^* and Λ^* are the smallest and largest eigenvalues of $\nabla_x^2 \phi(x^*, u^*, \mu)$ respectively.

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

Definition 3.4.3 *Definition of ξ^* and β^*.*

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

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

Lemma 3.4.3 *For* $(x,u) \in S_1$, $\beta(x) \geq \beta^*$.

Proof :

From the definition of $B(x)$ and lemmas 3.4.1 and 3.4, we have immediately that

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

For $R_1(x)$ we have with the previous lemma , since $J_\mu = J$:

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

whereas for η we can write :

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

We now use these bounds and lemma 3.4.1 to compute the lower bound on $\xi_1(x)$ in S_1 :

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

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

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

□

In the final lemma of this section, we compute a lower bound on the accuracy ϵ_1 in S_1.

Lemma 3.4.4 *In S_1, the accuracy ϵ_1 is bounded from below as follows :*

$$\epsilon_1(x) \geq \frac{\epsilon\mu\sqrt{2\lambda^*}}{360\rho^*\sqrt{n}} . \tag{3.15}$$

Proof : With the previous lemmas we have for $(x, u) \in S_1$:

$$\epsilon_1(x) = \frac{\epsilon\sqrt{\lambda_H}}{20\sqrt{n}} \min\left\{2, \frac{B^2(x)}{\bar{\rho}\mu}\right\} \geq \frac{\sqrt{2}\epsilon\sqrt{\lambda^*}}{60\sqrt{n}} \min\left\{2, \frac{\mu}{4\bar{\rho}}\right\} . \tag{3.16}$$

We know from lemma 3.4.1 that $\frac{2}{3} \leq \frac{\rho^*}{\bar{\rho}} \leq 2$, and from lemma 2.4.1 that $\mu \leq \theta^* \leq \rho^*$. Therefore

$$\frac{\mu}{4\bar{\rho}} = \frac{\mu}{4\rho^*}\left(\frac{\rho^*}{\bar{\rho}}\right) \leq \frac{1}{2}\left(\frac{\mu}{\rho^*}\right) \leq \frac{1}{2} .$$

This means that $\min\left\{2, \frac{\mu}{4\bar{\rho}}\right\} = \frac{\mu}{4\bar{\rho}}$.

Substituting all this back into (3.16) gives as a lower bound for ϵ_1 in S_1 :

$$\epsilon_1(x) \geq \frac{\sqrt{2}\epsilon\sqrt{\lambda^*}\mu}{360\sqrt{n}\rho^*} .$$

This completes the proof.

□

We now define this lower bound as ϵ_1^*.

Definition 3.4.4 *The lower bound ϵ_1^* on ϵ in S_1 is defined as*

$$\epsilon_1^* \triangleq \frac{\epsilon\mu\sqrt{2\lambda^*}}{360\rho^*\sqrt{n}} .$$

3.5 Behavior of the algorithm in a neighborhood of the solution and final results

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 3.5.1

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

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 s-th 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.
- $\lambda_s(x)$ is the smallest eigenvalue of H^s at the point x.

Figure 3.5 illustrates the labelling of the iterates.

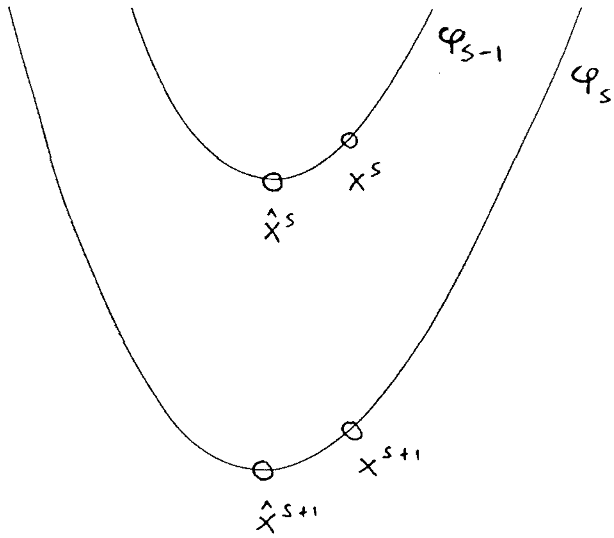
All the iterates will lie in S because of the basic convergence theorem. They satisfy :

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

where $\gamma = \frac{1}{2\sqrt{n}}$ and $\omega = \frac{\mu}{2\sqrt{n}}$.

We now recall that, starting with $x^{(0)}$, the algorithm first checks whether this point is in the Kantorovich ball for ϕ_0. If not, we perform a linesearch

Figure 3.2: Labelling of the iterates.



and continue to do so until the Kantorovich ball 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 Kantorovich ball 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 Kantorovich ball for $\phi_{\bar{s}}$ and whether this will remain so for subsequently generated iterates.

Lemma 3.5.1 *Assuming that* $\epsilon \leq 4n\gamma^{s+1}\omega$, *we have*

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

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

Proof :

(1) We begin with the first part.

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

where we have used the following :

$$\begin{aligned}
\|x^{(s+1)} - \hat{x}^{(s+1)}\| &\le \frac{1}{\sqrt{\lambda_s(x^{(s+1)})}} \|x^{(s+1)} - \hat{x}^{(s+1)}\|_{H^s(x^{(s+1)})} \\
&\le \frac{5}{2\sqrt{\lambda_s(x^{(s+1)})}} \|p_s(x^{(s+1)})\|_{H^s(x^{(s+1)})} \\
&\le \frac{5}{2\sqrt{\lambda_s(x^{(s+1)})}} \frac{\epsilon\sqrt{\lambda_s(x^{(s+1)})}}{10\sqrt{n}} \qquad (\textit{ accuracy of the minimization }) \\
&\le \frac{\epsilon}{4\sqrt{n}}\,.
\end{aligned}$$

(2) The second part follows almost immediately from part (1) and from a standard theorem about convex functions which is stated in the appendix as lemma 6.5.1 :

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

In the first inequality we used the fact that $\phi(x, u, \mu)$ is strongly convex.

This completes the proof.

□

We are now ready to state the main results of this chapter.

Theorem 3.5.1 *When the algorithm reaches a point $(x^{(\bar{s})}, u^{(\bar{s})})$ satisfying*

$$\max\{\|x^{(\bar{s})} - x^*\|_\infty, \|u^{(\bar{s})} - u^*\|_\infty\} < \frac{1}{2\sqrt{n}} \min\left\{\mu, \frac{\sqrt{\tilde{\lambda}}}{\tilde{\Lambda}}\beta^*\right\}, \qquad (3.17)$$

with $\epsilon < \left(\frac{1}{2\sqrt{n}}\right)^{\bar{s}} \frac{\mu}{2\sqrt{n}}$ (otherwise we have reached the desired accuracy and there is no point in continuing the algorithm) , then $x^{(\bar{s})}$ will lie in the "Kantorovich Ball" for $\phi_{\bar{s}}$ and the same will be true for each subsequently generated pair of primal and dual iterates.

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 this means :

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

We therefore have :

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

This completes the proof.

□

Following this theorem we define the following set :

Definition 3.5.2

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

Theorem 3.5.2 *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 (fixed u and μ) will be quadratic with the rate of convergence given by :*

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

The same notation was used as in lemma 3.2.7.

Proof :

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

$$\|\bar{y} - \hat{y}\| \leq \frac{\epsilon}{4\sqrt{n}} ,$$

as is shown in the course of the proof of lemma 3.5.1.

We therefore also have

$$\begin{aligned}
\|\hat{y} - x^*\| &\leq \|\bar{y} - \hat{y}\| + \|\bar{y} - x^*\| \\
&\leq \|\bar{y} - \hat{y}\| + \sqrt{n}\|\bar{y} - x^*\|_\infty \\
&\leq \frac{\epsilon}{4\sqrt{n}} + \frac{1}{2}\|v - u^*\|_\infty \\
&\leq \frac{1}{4\sqrt{n}}(\mu + \epsilon) .
\end{aligned}$$

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

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

$$\begin{aligned}
\|z - \hat{y}\| &\leq \frac{5}{2\sqrt{\lambda_v(z)}}\|p_v(z)\|_{H_v(z)} \\
&\leq \frac{5}{2\sqrt{\lambda_v(z)}}\frac{2}{5}\|p_v(y)\|_{H_v(y)} \quad (\textit{ from theorem } 3.2.1\ ) \\
&\leq \frac{\sqrt{\lambda_v(y)}}{\sqrt{\lambda_v(z)}}\frac{1}{\sqrt{\lambda_v(y)}}\|p_v(y)\|_{H_v(y)} \\
&\leq \frac{7}{6}\frac{1}{\sqrt{\lambda_v(y)}}\beta^* \quad (\textit{ from lemma } 3.2.8\ ) \\
&\leq \frac{7}{6}\frac{1}{\sqrt{\lambda_v(y)}}\frac{\mu\sqrt{\tilde{\lambda}}}{12\sqrt{n}} \\
&\leq \frac{7}{72\sqrt{n}}\mu\,.
\end{aligned}$$

We can therefore write

$$\begin{aligned}
\|z - x^*\|_\infty \leq \|z - x^*\| &\leq \|z - \hat{y}\| + \|\hat{y} - x^*\| \\
&\leq \frac{7}{72\sqrt{n}}\mu + \frac{1}{4\sqrt{n}}(\mu + \epsilon) \\
&\leq \left(25 + 18\frac{\epsilon}{\mu}\right)\frac{\mu}{72\sqrt{n}} \\
&< \frac{\mu}{2\sqrt{n}}\,.
\end{aligned}$$

This means that $(z, v) \in S$. The exact same procedure can now be carried out with z instead of y 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 3.2.7. 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

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

This completes the proof.

□

3.6 Conclusions

With theorem 3.5.1, we have found the region in which the "hot start" occurs. This region depends on several parameters and in order to have a clearer picture, we will make a few simplifying assumptions. First, from the form of the Hessian of $\phi(x, u, \mu)$, it is reasonable to assume that $\tilde{\lambda} \sim \mathcal{O}(\mu^{-1})$ and $\tilde{\Lambda} \sim \mathcal{O}(\mu^{-2})$. We also know from theorem 2.4.1 that for $\gamma = \frac{1}{2\sqrt{n}}$ (as we assumed), $\theta^* = \mu$ and $\sigma = \mu \ln(2\sqrt{n})$ in the worst case. We will assume this worst case. We note that we wrote "worst case", because those values will make β^* smaller.

For ξ^* this gives

$$\xi^* \geq \frac{\mu^{\frac{1}{2}}}{10}\left(\sqrt{2} + \frac{27(m-r)}{4\sqrt{2n}(\ln 2\sqrt{n} + \frac{1}{2})^3(\mu\lambda^*)^{\frac{3}{2}}}\right)^{-1} . \tag{3.18}$$

This means that $\xi^* \sim \mathcal{O}\left(\mu^{\frac{1}{2}}\right)$ or $\xi^* \sim \mathcal{O}\left(\frac{\sqrt{n}(\ln n)^3}{m-r}\mu^{\frac{1}{2}}\right)$, depending on whether the first or the second term in the sum in the RHS of (3.18) is larger.

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

If we put all of this together, we obtain for the RHS in (3.17) :

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

and we define the following :

Definition 3.6.1

$$\begin{aligned} \kappa &\triangleq \frac{1}{2\sqrt{n}}\min\left\{\mu, \frac{\sqrt{\tilde{\lambda}}}{\tilde{\Lambda}}\beta^*\right\} \\ \zeta &\triangleq \min\left\{\frac{1}{n}, \frac{(\ln n)^3}{m-r}\right\} . \end{aligned}$$

From any point in the set T, each outer iteration will need $\mathcal{O}(\ln\ln\epsilon^{-1})$ (since $\mathcal{O}(\ln\ln(\epsilon_1^*)^{-1}) \sim \mathcal{O}(\ln\ln\epsilon^{-1})$) inner iterations, whereas $\mathcal{O}\left(\frac{\ln\left(\frac{\epsilon}{\kappa}\right)^{-1}}{\ln n}\right)$ outer iterations are needed to obtain an accuracy of ϵ.

Since it takes $\mathcal{O}(\sqrt{m}\ln\kappa)$ iterations to reach a point in T using the logarithmic barrier function (see [4]), the overall complexity to reach a point with accuracy ϵ, using the logarithmic BF until T is reached and then continuing with the MBF method, is given by

$$\mathcal{O}\left(\sqrt{m}\ln(\mu\zeta)^{-1}\right) + \mathcal{O}\left(\left(\frac{\ln\epsilon^{-1} - \ln(\mu\zeta)^{-1}}{\ln n}\right)\ln\ln\epsilon^{-1}\right) .$$

In view of the definition of C in chapter 2, this can also be written as

$$\mathcal{O}\left(\sqrt{m}\ln C + \sqrt{m}\ln\zeta^{-1} + \left(\frac{\ln\epsilon^{-1} - \ln C - \ln\zeta^{-1}}{\ln n}\right)\ln\ln\epsilon^{-1}\right) . \qquad (3.19)$$

Under the assumption that $\epsilon \ll \mu$, this represents a significant improvement over the classical barrier method where the complexity is given by $\mathcal{O}(\sqrt{m}\ln\epsilon^{-1})$.

The implications for linear programming problems will be dealt with in the next chapter.

Chapter 4

The linear case

In this chapter we consider the following problem :

$$\text{(LP)} \qquad \begin{array}{ll} \min & b^T x \\ \text{s.t.} & r_i(x) \geq 0 \quad i = 1, ..., m \end{array} \quad ,$$

where, as before :

$$\begin{array}{l} r_i(x) = a_i^T x - b_i \\ a_i \in I\!R^n \ \ and \ \ \|a_i\| = 1 \\ b_i \in I\!R \ . \end{array}$$

We again assume that $int\ \{x \mid r_i(x) \geq 0\}$ is nonempty and bounded, and that all conditions in the basic theorem 2.2.1 are satisfied. We are therefore considering a nondegenerate linear programming problem.

For Karmarkar's algorithm (with $n \geq m$), the number of iterations needed to reach an accuracy of 2^{-L}, where L is the number of bits necessary to specify (LP), is given by $\mathcal{O}\,(nL)$. The best bound on the number of iterations for the Classical Logarithmic Barrier Function at the moment is $\mathcal{O}\,(\sqrt{m}L)$. We now compare these bounds with the bound we will obtain for the MBF method.

We proceed exactly as in the quadratic case and define :

Definition 4.0.2 *The Modified Barrier Function for LP is defined as*

$$\phi(x,u,\mu) \triangleq \frac{b^T x}{\mu} - \sum_{i=1}^{m} u_i \ln(r_i(x)+\mu).$$

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

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

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

The Hessian $\nabla_x^2 \phi(x,u,\mu)$ is strictly positive definite on the extended feasible set as in the quadratic case and, again, because of (3) in the basic theorem, its smallest eigenvalue is bounded away from zero in a neigborhood of the minimum of $\phi(x,u,\mu)$ for fixed (u,k).

All lemmas and theorems for the quadratic case remain valid for LP. However, the conclusions about the order of magnitude of the set T are now slightly different, since in this case it is reasonable to assume from the form of the Hessian of $\phi(x,u,\mu)$, that $\tilde{\lambda} \sim \mathcal{O}(\mu^{-2})$ in stead of $\mathcal{O}(\mu^{-1})$.

In the worst case, $\beta^* \sim \mathcal{O}\left(\min\left\{\frac{1}{\sqrt{n}}, \sqrt{\mu}\right\}\right)$, and therefore

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

Now suppose that we want to achieve an accuaracy of 2^{-L} in order to find an exact solution to the LP problem. In that case, (3.19) will give us an overall complexity of

$$\mathcal{O}\left(\sqrt{m}\ln C + \sqrt{m}\ln n + \left(\frac{\ln L - \ln C - \ln n}{\ln n}\right) L\right) .$$

Name	Size	No. of iterations till "Kantorovich Ball"
SUHLPIPU	48,15	9 , 6 , 2 , 1
LSEUNEW	134,28	8 , 6 , 3 , 1
GM291U	543,250	8 , 10 , 3 ,1
IBM118	657,359	9 , 8 , 2 , 1
GMLP	3801,1065	16 , 23 , 8 , 2 , 1

Table 4.1: The number of iterations until the "Kantorovich Ball" is reached for five IBM problems.

Since, typically, $L \sim \mathcal{O}(mn)$, this gives a significantly better complexity bound than what so far has been obtained, as long as $\ln C \ll L$. If this were not the case, then we would obtain the same complexity bound as for the classical logarithmic barrier function.

We conclude this chapter with some numerical results obtained at IBM by the author. In table 4.1, the first column gives the name of the problem as it appears in the IBM problem library, the second column contains the number of variables of the problem (first number) and the number of constraints (second number), whereas the third column gives the number of iterations necessary to reach the "Kantorovich Ball" after each Lagrange multiplier update (the updates are separated by commas).

Over the past year, more than 60 more LP problems have been solved at IBM with the MBF method (not by the author), all yielding similar results.

Chapter 5

The nonlinear case

5.1 Introduction

(i) In this chapter we shall consider the MBF method for solving the following problem :

$$\begin{array}{ll} \min & f_0(x) \\ \text{s.t.} & f_i(x) \geq 0 \quad i = 1, \ldots, m \quad , \end{array}$$

where $x \in I\!R^n$, the objective function is convex and all constraint functions are concave. We will assume w.l.o.g. that $f_0(x)$ is linear. In case it is not, we can add an additional variable x_{n+1} and an additional constraint

$$f_0(x) - x_{n+1} \leq 0 \ ,$$

and take x_{n+1} as the objective function to be minimized. We will therefore consider the following problem :

$$\text{(NLP)} \quad \begin{array}{ll} \min & -b^T x \\ \text{s.t.} & f_i(x) \geq 0 \quad i = 1, \ldots, m \quad , \end{array}$$

We define the extended feasible set Ω_k completely analogously to Ω_k in the quadratic case.

We shall assume that all constraint functions are twice continuously differentiable on Ω_k and that they all satisfy the following Lipschitz conditions on Ω_k :

$$\exists L > 0 \; : \; \forall x, y \in \Omega_k \; : \; |f_i(x) - f_i(y)| \leq L\|x - y\| \tag{5.1}$$

$\exists M > 0 \; : \; \forall x, y \in \Omega_k, \forall d \in I\!R^n$:

$$|d^T \nabla^2 f_i(x) d - d^T \nabla^2 f_i(y) d| \leq M\|x - y\|(-d^T \nabla^2 f_i(x) d) \, . \tag{5.2}$$

This condition is very similar to the one used by Jarre [14].

As in the quadratic case, it is assumed that *int* $\{x \mid f_i(x) \geq 0\}$ is nonempty and bounded and that all conditions in the basic theorem from chapter 2 are satisfied.

(ii) We now define the MBF for (NLP) :

Definition 5.1.1 *The Modified Barrier Function for (NLP) is defined as*

$$\phi(x, u, \mu) \triangleq \frac{-b^T x}{\mu} - \sum_{i=1}^{m} u_i \ln(f_i(x) + \mu).$$

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

$$\nabla_x \phi(x, u, \mu) = g(x, u, \mu) = \frac{-b}{\mu} - \sum_{i=1}^{m} u_i \frac{\nabla f_i(x)}{f_i(x) + \mu}$$

$$\nabla_x^2 \phi(x, u, \mu) = H(x, u, \mu) = \sum_{i=1}^{m} u_i \left(\frac{-\nabla^2 f_i(x)}{f_i(x) + \mu} + \frac{\nabla f_i(x) \nabla f_i(x)^T}{(f_i(x) + \mu)^2} \right) \, .$$

The Hessian $\nabla_x^2 \phi(x, u, \mu)$ is strictly positive definite as in the quadratic case, and because of condition (3) in the basic theorem, its smallest eigenvalue will be bounded away from zero in a neighborhood of the minimum of $\phi(x, u, \mu)$ for fixed (u, k).

(iii) The algorithm is exactly the same as in the quadratic case, with $f_i(x)$ replacing $r_i(x)$, except that μ should now be such that γ is at least $\frac{1}{2(L+1)\sqrt{n}}$.

(iv) In this chapter we will show that, as in the quadratic case, there exists a "hot start" for (NLP), which we will again characterize in terms of μ and other quantities which depend on the solution.

This chapter will resemble chapter 3 very much and the steps leading to the final result are exactly the same as in that chapter. The proofs are a little bit more involved, but the motivation behind the lemmas hasn't changed. We will therefore not repeat all the explanations from chapter 3 in the same detail.

5.2 Basic lemmas for the inner iteration

In this section we shall prove some basic lemmas, analogous to the ones proved in the quadratic case, that will be needed to determine β.

In the following two lemmas we compute a bound on the error in the quadratic approximation to the MBF. This is done in two stages, first for the case where all functions are linear or quadratic, and then this result is extended to nonlinear functions, using the relative Lipschitz condition (5.2).

We will follow the same notation as in the previous chapter.

Lemma 5.2.1 *If all functions $f_i(x)$ are linear or quadratic with positive definite Hessian matrix,and if $x \in int\ \Omega_k$ and $\|d\|_H \ < \ \beta_1(x)$,* *then $x + d \in int\ \Omega_k$. 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 :

$$R_1(x) = min \left\{ \frac{\sqrt{\lambda_H}(f_j(x)+\mu)}{\|\nabla f_j(x)\| + \|\nabla^2 f_j(x)\|} \mid j \in J_\mu \right\}$$

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

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

Proof : Expanding $\phi(x+d,u,\mu)$ in a Taylor series about x , we can write :

$$\phi(x+d,u,\mu) = \phi(x,u,\mu) + d^T(\nabla_x\phi(x)) + \frac{1}{2}d^T(\nabla_{xx}\phi(x))d + \sum_{k=3}^{\infty} t_k .$$

The first three terms of the RHS constitute the quadratic approximation $q(d,u,\mu)$ to $\phi(x+d,u,\mu)$ and t_k is the k-th order term in the Taylor expansion:

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

We now use the following results which follow immediately from lemma 2 in [4] :

Setting $g_i(x) = f_i(x) + \mu$, and

$$\psi_j = \frac{d^T\nabla^2 g_j(x)d}{-g_j(x)} \quad ; \quad \chi_j = \frac{d^T\nabla g_j(x)}{-g_j(x)} ,$$

we have that

$$t_k = \frac{1}{k!}\sum_{j=1}^{m} u_j \sum_{i=0}^{\lfloor\frac{k}{2}\rfloor} a_{k,i}\chi_j^{k-2i}\psi_j^i .$$

The $a_{k,i}$'s are constants that are nonzero only for $0 \le i \le \lfloor \frac{k}{2} \rfloor$.

We are therefore interested in finding a bound on the value of the objective function at the solution in the following constrained maximization problem :

$$\begin{aligned}
max \quad & \frac{1}{k!}\sum_{j=1}^{m} u_j \sum_{i=0}^{\lfloor \frac{k}{2} \rfloor} a_{k,i}\chi_j^{k-2i}\psi_j^i \\
s.t. \quad & \sum_{j=1}^{m} u_j(\chi_j^2 + \psi_j) = \|d\|_H^2 \\
& \chi_j \geq 0 \\
& \psi_j \geq 0 \, .
\end{aligned}$$

This problem was examined in [4] for the case where $\forall i \; : \; u_i = 1$ and it was shown that the solution to this problem satisfies : $\chi_j > 0 \; and \; \psi_j = 0$. This gives as the maximum value :

$$\frac{1}{k}\sum_{j=1}^{m} u_j \chi_j^k \, ,$$

and therefore : $t_k \leq \frac{1}{k}\sum_{j=1}^{m} u_j \chi_i^k$. It is not difficult to see that this result is equally true in our case. We shall now compute a bound on t_k exactly as in chapter 3, the only difference being the definition of $R_1(x)$. We note here that the definition of $R_1(x)$ implies that

$$R_1(x) \leq min \left\{ \frac{\sqrt{\lambda_H}(f_j(x)+\mu)}{\|\nabla f_j(x)\|} \mid j \in J_\mu \right\} .$$

The rest of the proof is exactly as the proof of lemma 3.2.1.

Lemma 5.2.2 *If the functions f_i satisfy the relative Lipschitz condition (5.2), with constant M, if x is feasible and if*

$$\|d\|_H \leq \beta_2(x) \, ,$$

then $x+d$ is feasible and

$$|\phi(x+d,u,\mu)-q_x(d,u,\mu)| < \frac{1}{3\xi_2(x)}\|d\|_H^3 ,$$

where $\xi_1(x)$ is as before and

$$\xi_2(x) = \left(\frac{1}{\xi_1}+\frac{24M}{7\sqrt{\lambda_H}}\right)^{-1}$$

$$R_2(x) = min \left\{\frac{\lambda_H^{\frac{3}{2}}(f_j(x)+\mu)}{\|\nabla^2 f_j(x)\|} \mid j \in J_\mu\right\}$$

$$\beta_2(x) = min \left\{\frac{\sqrt{\theta}}{2}, \frac{R_1(x)}{2}, \frac{1}{2}\left(\frac{R_2(x)}{M}\right)^{\frac{1}{3}}, \frac{1}{2}\left(\frac{\theta\sqrt{\lambda_H}}{M}\right)^{\frac{1}{3}}\right\} .$$

Proof :

Defining

$$\begin{aligned} q_i(d) &\triangleq g_i(x) + d^T(\nabla g_i(x)) + \frac{1}{2}d^T(\nabla^2 g_i(x))d \\ r_i &\triangleq \frac{g_i(x+d)-q_i(d)}{q_i(d)} , \end{aligned}$$

we have $g_i(x+d) = (1+r_i)q_i(d)$, and

$$\begin{aligned} |g_i(x+d)-q_i(d)| &= \frac{1}{2}\left|d^T\left(\nabla^2 g_i(x+\zeta d) - \nabla^2 g_i(x)\right)d\right| \\ &\le \frac{1}{2}M\zeta\|d\|_H|d^T\nabla^2 g_i d| \qquad (\zeta \in (0,1)) . \end{aligned}$$

Now

$$q_i(d) \ge \left(1-\frac{\|d\|_H}{\beta_1(x)}\right)q_i(0) + \frac{\|d\|_H}{\beta_1(x)}q_i\left(\beta_1(x)\frac{d}{\|d\|_H}\right) ,$$

and, since $q_i\left(\beta_1(x)\frac{d}{\|d\|_H}\right) \geq 0$:

$$q_i(d) \geq \left(1 - \frac{\|d\|_H}{\beta_1(x)}\right) q_i(0) .$$

We now estimate $|r_i|$:

$$\begin{aligned} |r_i| &\leq \frac{1}{2}M\|d\|\frac{|d^T\nabla^2 g_i(x)d|}{q_i(x+d)} \\ &\leq \frac{1}{2}M\|d\|\frac{|d^T\nabla^2 g_i(x)d|}{\left(1-\frac{\|d\|_H}{\beta_1(x)}\right)g_i(x)} . \qquad (since\ g_i(x) = q_i(0)) \end{aligned}$$

This means that for $\|d\|_H < \beta_1(x)/2$:

$$|r_i| \leq M\|d\|\frac{-d^T\nabla^2 g_i(x)d}{g_i(x)} .$$

For $i \in I_\mu$ this means:

$$|r_i| \leq \frac{M}{\theta\sqrt{\lambda_H}}\|d\|_H^3 ,$$

whereas for $i \in J_\mu$ we have (with $R_2(x)$ as in the statement of the lemma) :

$$|r_i| \leq \frac{M}{R_2}\|d\|_H^3 .$$

Therefore, if we choose

$$\|d\|_H \leq min \left\{\frac{\sqrt{\theta}}{2}, \frac{R_1(x)}{2}, \frac{1}{2}\left(\frac{R_2(x)}{M}\right)^{\frac{1}{3}}, \frac{1}{2}\left(\frac{\theta\sqrt{\lambda_H}}{M}\right)^{\frac{1}{3}}\right\} ,$$

then $\forall i\ :\ |r_i| < \frac{1}{8}$.

For $|\phi(x+d,u,\mu) - q_x(d,u,\mu)|$ this gives :

$$\begin{aligned} &\left|-\frac{b^T(x+d)}{\mu} - \sum_{i=1}^m u_i \ln((1+r_i)q_i(d)) - q_x(d,u,\mu)\right| \\ = &\left|-\frac{b^T(x+d)}{\mu} - \sum_{i=1}^m u_i \ln(q_i(d)) - q_x(d,u,\mu) - \sum_{i=1}^m u_i \ln(1+r_i)\right| \\ \leq &\left|-\frac{b^T(x+d)}{\mu} - \sum_{i=1}^m u_i \ln(q_i(d)) - q_x(d,u,\mu)\right| + \sum_{i=1}^m |u_i \ln(1+r_i)| . \end{aligned}$$

Because q_i is quadratic, the first part can be estimated using Lemma 2.1 and since $|r_i| < \frac{1}{8}$, which implies that $|\ln(1+r_i)| < \frac{8}{7}|r_i|$ (see appendix), the second part can be estimated as follows

$$\begin{aligned}\sum_{i=1}^{m} |u_i \ln(1+r_i)| &\leq \sum_{i=1}^{m} \frac{8}{7} u_i |r_i| \\ &\leq \frac{8}{7} M \|d\| \sum_{i=1}^{m} u_i \frac{d^T \nabla^2 g_i(x) d}{-g_i(x)} .\end{aligned}$$

So finally

$$\begin{aligned}|\phi(x+d,u,\mu) - q(x+d,u,\mu)| &\leq \left(\frac{1}{3\xi_1} + \frac{8M}{7\sqrt{\lambda_H}}\right) \|d\|_H^3 \\ &\leq \frac{1}{3\xi_2(x)} \|d\|_H^3 .\end{aligned}$$

□

The following two lemmas are stated without proof since the proof is exactly the same as in the quadratic case.

Lemma 5.2.3 *If*

$$\|p\|_H < \beta_3(x) = \min\left\{\beta_2(x), \frac{\xi_2(x)}{5}\right\},$$

then

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

Lemma 5.2.4 *If*

$$\|p\|_H > \beta_3(x) = \min\left\{\beta_2(x), \frac{\xi_2(x)}{5}\right\},$$

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

$$\Delta\phi > \frac{2}{5}\beta_3^2(x) .$$

We now define the following quantities :

Definition 5.2.1

$$\begin{aligned} B(x) &\triangleq \min_i\{f_i(x)+\mu\} \\ C(x) &\triangleq \max_i\{\|\nabla f_i(x)\|\} \\ D(x) &\triangleq \max_i\{\|\nabla^2 f_i(x)\|\} \,. \end{aligned}$$

As in the quadratic case, the next two lemmas will determine the rate of convergence of the iterates in the inner iteration.

Lemma 5.2.5 *For $x, x+p \in$ int Ω_k with*

$$\|p\| \leq \min\left\{\frac{B(x)}{\alpha L}, \frac{B(x)}{\alpha}, \frac{1}{\alpha M}, \frac{1}{\alpha}, \frac{1}{2\alpha D(x)}\right\} \qquad (\alpha > 1)\,,$$

the following inequality holds :

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

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

From the definition of p this means

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

We now compute a bound on the expression in the RHS. From lemma 6.4.3 in the appendix we have :

$$\left|\sum_{s,t=1}^{m} \frac{\partial^3\phi(x+\zeta p)}{\partial x_s \partial x_t \partial x_i} p_s p_t\right| \\ \leq \left(\frac{(M+1)B(x+\zeta p)+2C(x+\zeta p)+D(x+\zeta p)}{B(x+\zeta p)}\right) p^T H(x+\zeta p)p \, .$$

Because $\|x+\zeta p - x\| \leq \|p\|$ and because of the assumption on $\|p\|$, we can use lemma 6.4.4 from the appendix. This gives

$$\begin{aligned}
&(M+1)B(x+\zeta p)+2C(x+\zeta p)+D(x+\zeta p) \\
&\quad \leq (M+1)\left(\frac{\alpha+1}{\alpha}\right)B(x)+2C(x)+2\left(1+M\|p\|\right)\|p\|D(x)+\left(\frac{\alpha+1}{\alpha}\right)D(x) \\
&\quad \leq (M+1)\left(\frac{\alpha+1}{\alpha}\right)B(x)+2C(x)+\frac{(\alpha+1)(\alpha+2)}{\alpha^2}D(x) \\
&\quad \leq \frac{(\alpha+1)(\alpha+2)}{\alpha^2}\left((M+1)B(x)+2C(x)+D(x)\right) \, .
\end{aligned}$$

We now use this bound, together with lemma 6.4.6 from the appendix to obtain

$$\left|\sum_{s,t=1}^{m} \frac{\partial^3\phi(x+\zeta p)}{\partial x_s \partial x_t \partial x_i} p_s p_t\right| \\ \leq \frac{2\alpha\left(\alpha+1\right)\left(\alpha+2\right)}{\left(\alpha-1\right)^3}\left(\frac{(M+1)B(x)+2C(x)+D(x)}{B(x)}\right) p^T H(x)p \, .$$

We therefore have

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

Squaring the LHS and summing over j gives :

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

Taking the square root on both sides completes the proof.

□

Lemma 5.2.6 *Let p,q and $H,\overline{H}$ be the Newton directions and Hessians at x and $x+p$ respectively, with $x, x+p \in int\ \Omega_k$ and*

$$\|p\| \leq \min\left\{\frac{1}{2},\frac{B(x)}{\alpha L},\frac{1}{\alpha M},\frac{1}{2D(x)}\right\} \qquad (\alpha > 1)\,,$$

then :

$$\frac{\|q\|_{\overline{H}}}{\|p\|_H^2} \leq \frac{\alpha\,(\alpha+1)\,(\alpha+2)\sqrt{n}}{(\alpha-1)^3}\left(\frac{(M+1)B(x)+2C(x)+D(x)}{B(x)\sqrt{\lambda_{\overline{H}}}}\right) \,.$$

Proof : We have

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

Therefore :

$$q^T\overline{H}^2q = \|\overline{H}q\|^2 = \|\nabla\phi(x+p)\|^2\,.$$

On the other hand one has

$$\begin{aligned} q^T\overline{H}^2q &= (\overline{H}^{\frac{1}{2}}q)^T\overline{H}(\overline{H}^{\frac{1}{2}}q) \\ &\geq \lambda_{\overline{H}}\|\overline{H}^{\frac{1}{2}}q\|^2 \\ &\geq \lambda_{\overline{H}}\|q\|_{\overline{H}}^2\,. \end{aligned}$$

Therefore we can write

$$\lambda_{\overline{H}} \|q\|_{\overline{H}}^2 \leq \|\nabla\phi(x+p)\|^2 \,.$$

With the previous lemma, this gives

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

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

□

We now define a quantity $\beta_4(x,\alpha,\bar{\alpha})$, depending on x and two parameters α and $\bar{\alpha}$:

Definition 5.2.2

$$\beta_4(x,\alpha,\bar{\alpha}) \triangleq \min \left\{ \frac{1}{2}, \frac{R_1(x)}{2}, \frac{\xi_2(x)}{5}, \frac{R_2^{\frac{1}{3}}(x)}{2M^{\frac{1}{3}}}, \frac{\theta^{\frac{1}{3}}\lambda_H^{\frac{1}{6}}}{2M^{\frac{1}{3}}}, \frac{\sqrt{\lambda_H}}{\alpha M}, \frac{\sqrt{\lambda_H}}{\alpha}, \frac{\sqrt{\lambda_H}}{2\alpha D(x)}, \frac{B(x)\sqrt{\lambda_H}}{\alpha L}, \right.$$
$$\left. \frac{B(x)\sqrt{\lambda_H}}{\alpha}, \frac{1}{\bar{\alpha}} \left(\frac{(\alpha-1)^3}{\alpha\,(\alpha+1)\,(\alpha+2)\,\sqrt{n}} \right) \frac{B(x)\sqrt{\lambda_H}}{(M+1)\,B(x)+2C(x)+D(x)} \right\} .$$

As in the quadratic case, we will show that for certain values of α and $\bar{\alpha}$ the following property holds :

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

$$\|p(x,u,\mu)\|_{H(x,u,\mu)} \leq \beta_4(x,\alpha,\bar{\alpha}) \,,$$

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) $\quad y \in int\ \Omega_k$

(2) $\quad \|p(y,u,\mu)\|_{H(y,u,\mu)} \leq \beta_4(y,\alpha,\bar{\alpha})$.

The next lemma will be needed to show this . We will use the same notation as in the quadratic case, which we repeat here for convenience :

$$\begin{aligned} p &= p(x,u,\mu) \\ \bar{x} &= x+p \\ q &= p(\bar{x},u,\mu) \\ \lambda_H &= \textit{the smallest eigenvalue of } H(x,u,\mu) \\ \lambda_{\overline{H}} &= \textit{the smallest eigenvalue of } H(\bar{x},u,\mu)\ . \end{aligned}$$

Lemma 5.2.7 *If* $\|p\|_H \le \beta_4(x,\alpha,\bar{\alpha})$ *with*

$$\begin{aligned} \alpha &\ge 3 \\ \bar{\alpha} &> 4\left(\frac{\alpha+1}{\alpha}\right)\left(\frac{(\alpha+1)^9}{\alpha^6\,(\alpha-1)^3}\right) , \end{aligned}$$

then $\|q\|_{\overline{H}} \le \beta_4(\bar{x},\alpha,\bar{\alpha})$.

Proof :

From lemma 5.2.6 we have that :

$$\|q\|_{\overline{H}} \le \frac{\alpha\,(\alpha+1)\,(\alpha+2)\sqrt{n}}{(\alpha-1)^3}\left(\frac{(M+1)B(x)+2C(x)+D(x)}{B(x)\sqrt{\lambda_{\overline{H}}}}\right)\|p\|_H^2\ . \quad (5.3)$$

Now, we assumed that

$$\begin{aligned} \|p\|_H &\le \beta_4(x,\alpha,\bar{\alpha}) \\ &\le \frac{1}{\bar{\alpha}}\left(\frac{(\alpha-1)^3}{\alpha\,(\alpha+1)\,(\alpha+2)\,\sqrt{n}}\right)\frac{B(x)\sqrt{\lambda_H}}{(M+1)\,B(x)+2C(x)+D(x)}\ . \end{aligned}$$

Substituting this value for one of the $\|p\|_H$'s in the RHS of (5.3) gives

$$\begin{aligned}\|q\|_{\overline{H}} &\leq \frac{1}{\bar{\alpha}}\left(\frac{\lambda_H}{\lambda_{\overline{H}}}\right)^{\frac{1}{2}}\|p\|_H \\ &\leq \frac{\sqrt{2}}{\bar{\alpha}}\left(\frac{\alpha+1}{\alpha}\right)\|p\|_H \, . \end{aligned} \tag{5.4}$$

Since

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

and because of the condition on $\|p\|_H$, we can use lemma 6.4.7 with $y=\bar{x}$. From the statement of this lemma and the definition of $\beta_4(x,\alpha,\bar{\alpha})$, it is clear that $\beta_4(\bar{x},\alpha,\bar{\alpha})$ will in the worst case be smaller than $\beta_4(x,\alpha,\bar{\alpha})$ by a factor of

$$\frac{\alpha^6\,(\alpha-1)^3}{2^{\frac{3}{2}}\,(\alpha+1)^9} \, ,$$

and therefore

$$\|p\|_H \leq \left(\frac{2^{\frac{3}{2}}\,(\alpha+1)^9}{\alpha^6\,(\alpha-1)^3}\right)\beta_4(\bar{x},\alpha,\bar{\alpha}) \, .$$

Together with (5.4), this gives

$$\|q\|_{\overline{H}} \leq \frac{\sqrt{2}}{\bar{\alpha}}\left(\frac{\alpha+1}{\alpha}\right)\left(\frac{2^{\frac{3}{2}}\,(\alpha+1)^9}{\alpha^6\,(\alpha-1)^3}\right)\beta_4(\bar{x},\alpha,\bar{\alpha}) \, .$$

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

□

We now have to choose α and $\bar{\alpha}$ and a convenient choice is 15 and $\frac{343}{34}$ respectively. It can easily be verified that for these values α and $\bar{\alpha}$ satisfy the conditions imposed on them earlier. This leads to the following definition :

Definition 5.2.3

$$\beta(x) \triangleq \beta_4(x, 15, \frac{343}{34}) = \min\left\{\frac{1}{2}, \frac{R_1(x)}{2}, \frac{\xi_2(x)}{5}, \frac{R_2^{\frac{1}{3}}(x)}{2M^{\frac{1}{3}}}, \frac{\theta^{\frac{1}{3}}\lambda_H^{\frac{1}{6}}}{2M^{\frac{1}{3}}}, \frac{\sqrt{\lambda_H}}{15M}, \frac{\sqrt{\lambda_H}}{15}, \frac{\sqrt{\lambda_H}}{30D(x)}, \right.$$
$$\left. \frac{B(x)\sqrt{\lambda_H}}{15L}, \frac{B(x)\sqrt{\lambda_H}}{15}, \frac{B(x)\sqrt{\lambda_H}}{15\sqrt{n}\left((M+1)\,B(x) + 2C(x) + D(x)\right)}\right\}.$$

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

Theorem 5.2.1 **(1)** *If* $\|p\|_H \leq \beta(x)$ *, then*

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

(ii) $\frac{\|q\|_{\overline{H}}}{\|p\|_H^2} \leq \frac{3\sqrt{n}\left((M+1)\,B(x) + 2C(x) + D(x)\right)}{2B(x)\sqrt{\lambda_{\overline{H}}}}$

(iii) $\|q\|_{\overline{H}} \leq \frac{1}{6}\|p\|_H$.

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

□

5.3 The Lagrange multiplier update

In this section we determine the accuracy to which the minimization w.r.t. x of $\phi(x,u,\mu)$ (fixed u and μ) should be computed. It is very similar to the section bearing the same title in chapter 3 and we will state and prove the following lemma without further ado. In it we use the same notation as in the quadratic case, which we repeat here for convenience :

$$
\begin{aligned}
\bar{u}_i &: \textit{old multipliers} \\
\hat{u}_i &: \textit{exact new multipliers} \\
\bar{\rho} &: \max_i \{\bar{u}_i\} \\
u_i &: \textit{approximate new multipliers} \\
x &: \textit{approximate minimum } \phi(x, \bar{u}, \mu) \\
\hat{x} &: \textit{exact minimum of } \phi(x, \bar{u}, \mu)\ .
\end{aligned}
$$

We also recall that the accuracy required of the solution to (NLP) was denoted by ϵ.

Lemma 5.3.1 *Let $\|u - u^*\|_\infty > \epsilon$ and let μ be such that if exact minimization of $\phi(x, u, \mu)$ were to be performed, the convergence rate would be given by $\frac{1}{4(L+1)\sqrt{n}}$.*

If

$$\|p\|_H \leq \frac{\epsilon\sqrt{\lambda_H}}{20(L+1)\sqrt{n}} \min\left\{2, \frac{B^2(x)}{\bar{\rho}\mu L}, \beta(x)\right\} ,$$

then

$$\max\{\|x - x^*\|_\infty, \|u - u^*\|_\infty\} \leq \frac{1}{2(L+1)\sqrt{n}} \|\bar{u} - u^*\|_\infty \ .$$

Proof :

We start by writing :

$$\begin{aligned}\|u-u^*\|_\infty &\le \|u-\hat{u}\|_\infty + \|\hat{u}-u^*\|_\infty \\ &\le \|u-\hat{u}\|_\infty + \frac{1}{4\,(L+1)\,\sqrt{n}}\|\bar{u}-u^*\|_\infty \,.\end{aligned}$$

Now, for all i we have

$$\begin{aligned}|u_i-\hat{u}_i| &= \left|\frac{\bar{u}_i\mu}{f_i(x)+\mu}-\frac{\bar{u}_i\mu}{f_i(\hat{x})+\mu}\right| \\ &= \frac{\bar{u}_i\mu|f_i(x)-f_i(\hat{x})|}{(f_i(x)+\mu)(f_i(\hat{x})+\mu)} \\ &\le \frac{\bar{u}_i\mu L\|x-\hat{x}\|}{(f_i(x)+\mu)(f_i(\hat{x})+\mu)}\,, \end{aligned} \tag{5.5}$$

and since $\|p\|_H \le \beta(x)$, lemma 5.2.3 yields

$$\|x-\hat{x}\| \le \frac{1}{\sqrt{\lambda_H}}\|x-\hat{x}\|_H \le \frac{5}{2\sqrt{\lambda_H}}\|p\|_H \,.$$

Substituting this back into (5.5) gives

$$|u_i-\hat{u}_i| \le \frac{5\bar{u}_i\mu L\|p\|_H}{2\sqrt{\lambda_H}(f_i(x)+\mu)(f_i(\hat{x})+\mu)}\,. \tag{5.6}$$

With the definition of $\beta(x)$ we also have from (5.3) that

$$\|x-\hat{x}\| \le \frac{5}{2\sqrt{\lambda_H}}\frac{B(x)\sqrt{\lambda_H}}{15L} \le \frac{B(x)}{2L}\,. \tag{5.7}$$

From lemma 6.4.1 we therefore have

$$\frac{1}{f_i(\hat{x})+\mu} \le \frac{2}{f_i(x)+\mu}\,.$$

Substituting this back into (5.6) yields

$$|u_i-\hat{u}_i| \le \frac{5\bar{u}_i\mu L\|p\|_H}{\sqrt{\lambda_H}(f_i(x)+\mu)^2}\,.$$

The assumption on $\|p\|_H$ in the statement of the lemma then gives

$$\|u - \hat{u}\|_\infty \leq \frac{\epsilon}{4\,(L+1)\,\sqrt{n}} \leq \frac{1}{4\,(L+1)\,\sqrt{n}}\|\bar{u} - u^*\|_\infty \,.$$

Finally, using inequality (5.5), we obtain

$$\|u - u^*\|_\infty \leq \frac{1}{2(L+1)\sqrt{n}}\|\bar{u} - u^*\|_\infty \,.$$

On the other hand, recalling that

$$\|x - \hat{x}\|_\infty \leq \|x - \hat{x}\| \leq \frac{1}{\sqrt{\lambda_H}}\|x - \hat{x}\|_H \leq \frac{5}{2\sqrt{\lambda_H}}\|p\|_H \,,$$

and with the assumption on $\|p\|_H$, we also have

$$\begin{aligned}
\|x - x^*\|_\infty &\leq \|x - \hat{x}\|_\infty + \|\hat{x} - x^*\|_\infty \\
&\leq \frac{\epsilon}{4\,(L+1)\,\sqrt{n}} + \frac{1}{4\,(L+1)\,\sqrt{n}}\|\bar{u} - u^*\|_\infty \\
&\leq \frac{1}{4\,(L+1)\,\sqrt{n}}\|\bar{u} - u^*\|_\infty + \frac{1}{4\,(L+1)\,\sqrt{n}}\|\bar{u} - u^*\|_\infty \\
&\leq \frac{1}{2(L+1)\sqrt{n}}\|\bar{u} - u^*\|_\infty \,.
\end{aligned}$$

This completes the proof.

□

Definition 5.3.1 *We define the quantity $\epsilon_1(x)$, used to determine the accuracy of the minimization w.r.t. x of $\phi(x,u,\mu)$ (fixed u and μ) :*

$$\epsilon_1(x) \triangleq \frac{\epsilon\sqrt{\lambda_H}}{20(L+1)\sqrt{n}} \min\left\{2, \frac{B^2(x)}{\bar{\rho}\mu L}\right\} .$$

Note : In practice, we minimize up to accuracy ϵ_1, with μ small enough to achieve a convergence rate of $\frac{1}{2(L+1)\sqrt{n}}$. This lemma then shows that in the worst case, μ should be such that for exact minimization of $\phi(x,u,\mu)$, the convergence rate would be $\frac{1}{4(L+1)\sqrt{n}}$.

5.4 Bounds on algorithm parameters in a neighborhood of the solution

As in chapter 3, we now consider the algorithm for (x, u) lying in the set S_1 , which is defined as follows :

Definition 5.4.1

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

In this definition,

$$\begin{aligned}\tilde{L} &= \max\left\{1, L, \frac{3}{2}M\mu, \frac{3}{2}D^*\mu, \frac{3}{2}\mu\right\}\\ D^* &= D(x^*)\,.\end{aligned}$$

In this section we will compute bounds in S_1 for previously defined quantities. We proceed almost exactly as in the quadratic case.

We start with the following lemma :

Lemma 5.4.1 *Let $(x,u) \in S_1$ and let the rate of convergence of the algorithm be given by $\gamma = \frac{1}{2(\tilde{L}+1)\sqrt{n}}$.*

Then

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

$$\frac{1}{2}u_i^* < u_i < \frac{3}{2}u_i^* \tag{5.8}$$

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

and

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

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

Proof :

(1) From theorem 2.4.1 with $\gamma = \frac{1}{2(L+1)\sqrt{n}}$, we have that $\mu \leq \theta^*$. Therefore, for any u such that

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

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

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

We now turn to the second set of inequalities. Since

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

we can write

$$\mu - |f_i(x) - f_i(x^*)| \leq f_i(x) + \mu \leq \mu + |f_i(x) - f_i(x^*)|$$

$$\mu - L\|x - x^*\| \leq f_i(x) + \mu \leq \mu + L\|x - x^*\| .$$

The proof then follows because $\|x - x^*\| < \frac{\mu}{2L}$.

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

$$\begin{aligned} f_i(x) + \mu &= f_i(x) - f_i(x^*) + f_i(x^*) + \mu \\ &\geq \mu - |f_i(x) - f_i(x^*)| + \sigma \\ &\geq \sigma + \mu - L\|x - x^*\| \, . \end{aligned}$$

Again, the proof follows from the asumption that $\|x - x^*\| < \frac{\mu}{2L}$.

$\square$

Exactly as in the quadratic case, the previous lemma means that for $(u, \mu) \in S_1$, the sets I_μ and J_μ are identical to I and J respectively.

We will now have a look at the eigenvalues of the Hessian 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 5.4.2 *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 :*

$$\frac{1}{9}\lambda^* \leq \lambda_H \leq 12\lambda^* + \frac{\mu(m-r)}{2\tilde{L}\sqrt{n}} \left(\frac{D^*}{\sigma + \frac{\mu}{2}} + \frac{C^*}{\left(\sigma + \frac{\mu}{2}\right)^2} \right)$$

$$\frac{1}{9}\Lambda^* \leq \Lambda_H \leq 12\Lambda^* + \frac{\mu(m-r)}{2\tilde{L}\sqrt{n}} \left(\frac{D^*}{\sigma + \frac{\mu}{2}} + \frac{C^*}{\left(\sigma + \frac{\mu}{2}\right)^2} \right) .$$

Proof :

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

$$\frac{1}{2}u_i^* \frac{(\nabla f_i(x)^T d)^2}{(\frac{3}{2}\mu)^2} \leq u_i \frac{(\nabla f_i(x)^T d)^2}{(f_i(x)+\mu)^2} \leq \frac{3}{2}u_i^* \frac{(\nabla f_i(x)^T d)^2}{(\frac{1}{2}\mu)^2} ,$$

$$\frac{1}{2}u_i^* \frac{(-d^T \nabla^2 f_i(x) d)}{(\frac{3}{2}\mu)} \leq u_i \frac{(-d^T \nabla^2 f_i(x) d)}{f_i(x)+\mu} \leq \frac{3}{2}u_i^* \frac{(-d^T \nabla^2 f_i(x) d)}{(\frac{1}{2}\mu)} .$$

Therefore :

$$\frac{2}{9}u_i^* \frac{(\nabla f_i(x)^T d)^2}{\mu^2} \leq u_i \frac{(\nabla f_i(x)^T d)^2}{(f_i(x)+\mu)^2} \leq 6u_i^* \frac{(\nabla f_i(x)^T d)^2}{\mu^2} ,$$

and

$$\frac{1}{3}u_i^* \frac{(-d^T \nabla^2 f_i(x) d)}{\mu} \leq u_i \frac{(-d^T \nabla^2 f_i(x) d)}{f_i(x)+\mu} \leq 3u_i^* \frac{(-d^T \nabla^2 f_i(x) d)}{\mu} .$$

From lemma 6.4.2 and 6.4.7 with $x = x^*$ and $y = x$ and with the condition we assumed on $\|x - x^*\|$, we have

$$\begin{aligned}
\frac{\left(\nabla f_i(x)^T d\right)^2}{\mu^2} &\geq \frac{1}{2}\frac{\left(\nabla f_i(x^*)^T d\right)^2}{\mu^2} - \frac{D^* \left(1 + M\|x - x^*\|\right)^2 \|x - x^*\|^2 \left(-d^T \nabla^2 f_i(x^*) d\right)}{\mu^2} \\
&\geq \frac{1}{2}\frac{\left(\nabla f_i(x^*)^T d\right)^2}{\mu^2} - \frac{8}{27}\frac{\left(-d^T \nabla^2 f_i(x^*) d\right)}{\mu} ,
\end{aligned}$$

and also

$$\begin{aligned}
\frac{\left(-d^T \nabla^2 f_i(x) d\right)}{\mu} &\geq \frac{\left(1 - M\|x - x^*\|\right)\left(-d^T \nabla^2 f_i(x^*) d\right)}{\mu} \\
&\geq \frac{2}{3}\frac{\left(-d^T \nabla^2 f_i(x^*) d\right)}{\mu} .
\end{aligned}$$

Similarly :

$$\begin{aligned}
\frac{\left(\nabla f_i(x)^T d\right)^2}{\mu^2} &\le 2\frac{\left(\nabla f_i(x^*)^T d\right)^2}{\mu^2} + \frac{2D^*\left(1+M\|x-x^*\|\right)^2\|x-x^*\|^2\left(-d^T\nabla^2 f_i(x^*)d\right)}{\mu^2} \\
&\le 2\frac{\left(\nabla f_i(x^*)^T d\right)^2}{\mu^2} + \frac{16}{27}\frac{\left(-d^T\nabla^2 f_i(x^*)d\right)}{\mu} ,
\end{aligned}$$

and

$$\begin{aligned}
\frac{\left(-d^T\nabla^2 f_i(x)d\right)}{\mu} &\le \frac{\left(1+M\|x-x^*\|\right)\left(-d^T\nabla^2 f_i(x^*)d\right)}{\mu} \\
&\le \frac{4}{3}\frac{\left(-d^T\nabla^2 f_i(x^*)d\right)}{\mu} .
\end{aligned}$$

Combining all of the above inequalities, we obtain

$$\begin{aligned}
u_i\left[\frac{(-d^T\nabla^2 f_i(x)d)}{f_i(x)+\mu} + \frac{(\nabla f_i(x)^T d)^2}{(f_i(x)+\mu)^2}\right] &\ge \frac{1}{3}u_i^*\frac{\left(-d^T\nabla^2 f_i(x)d\right)}{\mu} + \frac{2}{9}u_i^*\frac{\left(\nabla f_i(x)^T d\right)^2}{\mu^2} \\
&\ge \frac{1}{9}u_i^*\left[\frac{\left(-d^T\nabla^2 f_i(x^*)d\right)}{\mu} + \frac{\left(\nabla f_i(x^*)^T d\right)^2}{\mu^2}\right] ,
\end{aligned}$$

and

$$\begin{aligned}
u_i\left[\frac{(-d^T\nabla^2 f_i(x)d)}{f_i(x)+\mu} + \frac{(\nabla f_i(x)^T d)^2}{(f_i(x)+\mu)^2}\right] &\le 3u_i^*\frac{\left(-d^T\nabla^2 f_i(x)d\right)}{\mu} + 6u_i^*\frac{\left(\nabla f_i(x)^T d\right)^2}{\mu^2} \\
&\le 12u_i^*\left[\frac{\left(-d^T\nabla^2 f_i(x^*)d\right)}{\mu} + \frac{\left(\nabla f_i(x^*)^T d\right)^2}{\mu^2}\right] ,
\end{aligned}$$

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 :

$$\begin{aligned}
&\min_{\|d\|=1} \left\{ \sum_{i=1}^{m} u_i \left[\frac{\left(-d^T \nabla^2 f_i(x) d\right)}{f_i(x)+\mu} + \frac{(\nabla f_i(x)^T d)^2}{(f_i(x)+\mu)^2} \right] \right\} \\
&\quad \le \min_{\|d\|=1} \left\{ \sum_{i=1}^{r} u_i \left[\frac{\left(-d^T \nabla^2 f_i(x) d\right)}{f_i(x)+\mu} + \frac{(\nabla f_i(x)^T d)^2}{(f_i(x)+\mu)^2} \right] \right\} \\
&\quad \le \frac{1}{9} \min_{\|d\|=1} \left\{ \sum_{i=1}^{r} u_i^* \left[\frac{\left(-d^T \nabla^2 f_i(x^*) d\right)}{\mu} + \frac{\left(\nabla f_i(x^*)^T d\right)^2}{\mu^2} \right] \right\} .
\end{aligned}$$

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

Similarly, we have for the upper bounds :

$$\begin{aligned}
&\min_{\|d\|=1} \left\{ \sum_{i=1}^{m} u_i \left[\frac{\left(-d^T \nabla^2 f_i(x) d\right)}{f_i(x)+\mu} + \frac{(\nabla f_i(x)^T d)^2}{(f_i(x)+\mu)^2} \right] \right\} \\
&\quad \le \min_{\|d\|=1} \left\{ \sum_{i=1}^{r} u_i \left[\frac{\left(-d^T \nabla^2 f_i(x) d\right)}{f_i(x)+\mu} + \frac{(\nabla f_i(x)^T d)^2}{(f_i(x)+\mu)^2} \right] \right. \\
&\qquad \left. + \sum_{i=r+1}^{m} u_i \left[\frac{\left(-d^T \nabla^2 f_i(x) d\right)}{f_i(x)+\mu} + \frac{(\nabla f_i(x)^T d)^2}{(f_i(x)+\mu)^2} \right] \right\} \\
&\quad \le 12 \min_{\|d\|=1} \left\{ \sum_{i=1}^{r} u_i^* \left[\frac{\left(-d^T \nabla^2 f_i(x^*) d\right)}{\mu} + \frac{\left(\nabla f_i(x^*)^T d\right)^2}{\mu^2} \right] \right\} \\
&\qquad + \frac{\mu(m-r)}{2\tilde{L}\sqrt{n}} \left(\frac{D^*}{\sigma + \frac{\mu}{2}} + \frac{C^*}{\left(\sigma + \frac{\mu}{2}\right)^2} \right) .
\end{aligned}$$

In the last step we have used the following notation :

$$C^* = C(x^*)\,.$$

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

□

We now define the following two quantities :

Definition 5.4.2

$$\begin{aligned}
\tilde{\lambda} &\triangleq \frac{1}{9}\lambda^* \\
\tilde{\Lambda} &\triangleq 12\Lambda^* + \frac{\mu(m-r)}{2\tilde{L}\sqrt{n}}\left(\frac{D^*}{\sigma+\frac{\mu}{2}} + \frac{C^*}{\left(\sigma+\frac{\mu}{2}\right)^2}\right).
\end{aligned}$$

Here, λ^* and Λ^* are the smallest and largest eigenvalues of $\nabla^2\phi(x^*, u^*, \mu)$ respectively.

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

Definition 5.4.3 *Definition of R^*, ξ^* and β^*.*

$$\begin{aligned}
R^* &\triangleq \min\left\{\frac{3\sqrt{\lambda^*}(\sigma+\frac{\mu}{2})}{16(C^*+D^*)}, \frac{\sqrt{\lambda^*}(\sigma+\frac{\mu}{2})^{\frac{1}{3}}}{8M^{\frac{1}{3}}D^*}\right\} \\
\xi^* &\triangleq \frac{2}{5}\left(\frac{\sqrt{2}}{\sqrt{\theta^*}} + \frac{8^3\mu(m-r)(C^*+D^*)^3}{54\tilde{L}\sqrt{n}(\lambda^*)^{\frac{3}{2}}(\sigma+\frac{\mu}{2})^3} + \frac{36}{7\sqrt{\lambda^*}}\right)^{-1} \\
\beta^* &\triangleq \min\left\{\frac{1}{2}, R^*, \xi^*, \frac{(\theta^*)^{\frac{1}{3}}(\lambda^*)^{\frac{1}{6}}}{6^{\frac{1}{3}}}, \frac{\mu\sqrt{\lambda^*}}{90\tilde{L}}, \frac{\mu\sqrt{\lambda^*}}{180\sqrt{n}\left((M+1)\mu+2C^*+D^*\right)}\right\}.
\end{aligned}$$

Lemma 5.4.3 *For* $(x,u) \in S_1$, $\beta(x) \geq \beta^*$.

Proof :

We start by observing that since $(x,u) \in S_1$,

$$\|x - x^*\| \leq \frac{\mu}{2\tilde{L}} \leq \min\left\{\frac{\mu}{2}, \frac{\mu}{2L}, \frac{1}{3D^*}, \frac{1}{3M}, \frac{1}{3}\right\} ,$$

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

and we also recall that $J_\mu = J$.

From the definition of $B(x)$ and lemma 5.4.1, we have immediately that

$$\frac{\mu}{2} \leq B(x) \leq \frac{3\mu}{2} \quad \text{and} \quad B(x)\sqrt{\lambda_H} \geq \frac{\mu\sqrt{\lambda^*}}{6} .$$

We also have from lemma 6.4.4 with $\alpha = 3$ that

$$\begin{aligned} D(x) &\leq \frac{4}{3}D^* \\ C(x) + D(x) &\leq \frac{16}{9}(C^* + D^*) \\ 2C(x) + D(x) &\leq \frac{17}{9}(2C^* + D^*) . \end{aligned}$$

The following inequalities are now straightforward :

$$\begin{aligned} \frac{\sqrt{\lambda_H}}{D(x)} &\geq \frac{3\sqrt{\tilde{\lambda}}}{4D^*} \geq \frac{\sqrt{\lambda^*}}{4D^*} , \\ \frac{B(x)\sqrt{\lambda_H}}{(M+1)B(x) + 2C(x) + D(x)} &\geq \frac{\mu\sqrt{\lambda^*}}{6\left((M+1)\left(\frac{3\mu}{2}\right) + \frac{17}{9}(2C^* + D^*)\right)} \\ &\geq \frac{\mu\sqrt{\lambda^*}}{12\left((M+1)\mu + 2C^* + D^*\right)} , \end{aligned}$$

$$
\begin{aligned}
\theta^{\frac{1}{3}}\lambda_H^{\frac{1}{6}} &\geq \frac{(\theta^*)^{\frac{1}{3}}(\lambda^*)^{\frac{1}{6}}}{2^{\frac{1}{3}}3^{\frac{1}{3}}} \geq \frac{(\theta^*)^{\frac{1}{3}}(\lambda^*)^{\frac{1}{6}}}{6^{\frac{1}{3}}}\,, \\
R_1(x) &\geq \frac{9\sqrt{\tilde{\lambda}}(\sigma+\frac{\mu}{2})}{16(C^*+D^*)} \geq \frac{3\sqrt{\lambda^*}(\sigma+\frac{\mu}{2})}{8(C^*+D^*)}\,, \\
R_2^{\frac{1}{3}}(x) &\geq \frac{3\sqrt{\tilde{\lambda}}(\sigma+\frac{\mu}{2})^{\frac{1}{3}}}{4D^*} \geq \frac{\sqrt{\lambda^*}(\sigma+\frac{\mu}{2})^{\frac{1}{3}}}{4D^*}\,, \\
\xi_1(x) &= \frac{1}{2}\left(\frac{1}{\sqrt{\theta}}+\frac{\eta}{R_1^3(x)}\right)^{-1} \\
&\geq \frac{1}{2}\left(\frac{\sqrt{2}}{\sqrt{\theta^*}}+\frac{8^3\mu(m-r)(C^*+D^*)^3}{54\tilde{L}\sqrt{n}(\lambda^*)^{\frac{3}{2}}(\sigma+\frac{\mu}{2})^3}\right)^{-1}\,, \\
\xi_2(x) &= \left(\frac{1}{\xi_1(x)}+\frac{24}{7\sqrt{\lambda_H}}\right)^{-1} \\
&\geq \left(2\left(\frac{\sqrt{2}}{\sqrt{\theta^*}}+\frac{8^3\mu(m-r)(C^*+D^*)^3}{54\tilde{L}\sqrt{n}(\lambda^*)^{\frac{3}{2}}(\sigma+\frac{\mu}{2})^3}\right)+\frac{72}{7\sqrt{\lambda^*}}\right)^{-1}\,.
\end{aligned}
$$

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

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

□

We conclude this section with a lower bound on the accuracy of the inner iteration ϵ_1. This lower bound is given by :

$$\epsilon_1^* \triangleq \frac{\epsilon\mu\sqrt{\lambda^*}}{360\tilde{L}\,(L+1)\,\rho^*\sqrt{n}}\,.$$

The proof is exactly as in the quadratic case and will be omitted here.

5.5 Behavior of the algorithm in a neighborhood of the solution and final results

In this last section we consider the behavior of the algorithm in a subset of S_1 and present the final results.

The subset $S \subset S_1$ which we will look at is defined as follows :

Definition 5.5.1

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

Starting from a point $(x^{(0)}, u^{(0)})$ in S, let us now examine the iterates $(x^{(s)}, u^{(s)})$. Again we will follow chapter 3 and for convenience restate the notation used there :

- $\phi_{s-1} = \phi(x, u^{(s-1)}, \mu)$.
- $x^{(s)}$ is the s-th 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.
- $\lambda_s(x)$ is the smallest eigenvalue of H^s at the point x.

All the iterates will lie in S because of the basic convergence theorem. They satisfy :

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

where $\gamma = \frac{1}{2(L+1)\sqrt{n}}$ and $\omega = \frac{\mu}{2\tilde{L}\sqrt{n}}$.

We now recall that, starting with $x^{(0)}$, the algorithm first checks whether this point is in the Kantorovich ball for ϕ_0. If not, we perform a linesearch and continue to do so until the Kantorovich ball 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 Kantorovich ball for ϕ_1 and so on.

The next lemma is the last one we need to prove the 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 Kantorovich ball for $\phi_{\bar{s}}$ and whether this will remain so for subsequently generated iterates.

Lemma 5.5.1 *Assuming that $\epsilon \leq 4(L+1)n\gamma^{s+1}\omega$, we have*

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

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

Proof :

(1) We begin with the first part.

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

where we have used the following :

$$\|x^{(s+1)} - \hat{x}^{(s+1)}\| \;\leq\; \frac{1}{\sqrt{\lambda_s(x^{(s+1)})}}\|x^{(s+1)} - \hat{x}^{(s+1)}\|_{H^s(x^{(s+1)})}$$

$$\begin{aligned}
&\le \frac{5}{2\sqrt{\lambda_s(x^{(s+1)})}}\|p_s(x^{(s+1)})\|_{H^s(x^{(s+1)})} \\
&\le \frac{5}{2\sqrt{\lambda_s(x^{(s+1)})}}\frac{\epsilon\sqrt{\lambda_s(x^{(s+1)})}}{(10(L+1)\sqrt{n})} \qquad (\text{ accuracy of the minimization }) \\
&\le \frac{\epsilon}{4\,(L+1)\,\sqrt{n}}\,.
\end{aligned}$$

(2) The second part follows almost immediately from part (1) exactly as in the quadratic case.

This completes the proof.

□

We are now ready to state the main results of this chapter.

Theorem 5.5.1 *When the algorithm reaches a point* $(x^{(\bar{s})}, u^{(\bar{s})})$ *satisfying*

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

with $\epsilon < \left(\frac{1}{2(L+1)\sqrt{n}}\right)^{\bar{s}}\frac{\mu}{2\tilde{L}\sqrt{n}}$ *(otherwise we have reached the desired accuracy and there is no point in continuing the algorithm) , then* $x^{(\bar{s})}$ *will lie in the "Kantorovich Ball" for* $\phi_{\bar{s}}$ *and the same will be true for each subsequently generated pair of primal and dual iterates.*

Proof : The proof is exactly as in the quadratic case.

□

Following this theorem we define the following set :

Definition 5.5.2

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

Theorem 5.5.2 *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 (fixed u and μ) will be quadratic with the rate of convergence given by*

$$\frac{\|q\|_{\overline{H}}}{\|p\|_H^2} \leq \frac{1}{10\beta^*} .$$

The same notation was used as in lemma 3.2.7.

Proof :

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

$$\|\bar{y}-\hat{y}\| \leq \frac{\epsilon}{4\,(L+1)\sqrt{n}} .$$

We therefore also have

$$\begin{aligned}
\|\hat{y}-x^*\| &\leq \|\bar{y}-\hat{y}\| + \|\bar{y}-x^*\| \\
&\leq \|\bar{y}-\hat{y}\| + \sqrt{n}\|\bar{y}-x^*\|_\infty \\
&\leq \frac{\epsilon}{4\,(L+1)\sqrt{n}} + \frac{\sqrt{n}}{2(L+1)\sqrt{n}}\|v-u^*\|_\infty \\
&\leq \frac{\epsilon}{4\,(L+1)\sqrt{n}} + \frac{\mu}{4\tilde{L}(L+1)\sqrt{n}} \\
&\leq \frac{1}{4\,(L+1)\sqrt{n}}\left(\frac{\mu}{\tilde{L}} + \epsilon\right) .
\end{aligned}$$

We now show that any iterate w, obtained during the minimization of $\phi(x, v, \mu)$ starting from $(y, v) \in T$, is such that (w, v) lies in S_1.

We start by looking at what happens to the first iterate z, obtained by taking a full Newton step from y. We show that (z, v) lies inside S_1.

$$
\begin{aligned}
\|z - \hat{y}\| &\leq \frac{5}{2\sqrt{\lambda_v(z)}}\|p_v(z)\|_{H_v(z)} \\
&\leq \frac{5}{2\sqrt{\lambda_v(z)}}\frac{1}{6}\|p_v(y)\|_{H_v(y)} \quad (\; \textit{from theorem } 5.2.1 \;) \\
&\leq \left(\frac{5}{12}\right)\frac{\sqrt{\lambda_v(y)}}{\sqrt{\lambda_v(z)}}\frac{1}{\sqrt{\lambda_v(y)}}\|p_v(y)\|_{H_v(y)} \\
&\leq \left(\frac{16\sqrt{2}}{15}\right)\left(\frac{5}{12}\right)\frac{1}{\sqrt{\lambda_v(y)}}\beta^* \quad (\; \textit{from lemma } 6.4.7 \;) \\
&\leq \left(\frac{16\sqrt{2}}{15}\right)\left(\frac{5}{12}\right)\frac{1}{\sqrt{\lambda_v(y)}}\left(\frac{\mu\sqrt{\tilde{\lambda}}}{30\tilde{L}}\right) \quad (\; \tilde{\lambda} = \frac{1}{9}\lambda^* \;) \\
&\leq \frac{2}{135}\frac{\mu}{\tilde{L}}
\end{aligned}
$$

We can therefore write

$$
\begin{aligned}
\|z - x^*\|_\infty \leq \|z - x^*\| &\leq \|z - \hat{y}\| + \|\hat{y} - x^*\| \\
&\leq \frac{2}{135}\left(\frac{\mu}{\tilde{L}}\right) + \frac{1}{4(L+1)\sqrt{n}}\left(\frac{\mu}{\tilde{L}} + \epsilon\right) \\
&\leq \left(2 + \frac{135}{4} + \frac{135\epsilon\tilde{L}}{4\mu}\right)\frac{\mu}{135\tilde{L}} \\
&\leq \left(2 + \frac{135}{4} + \frac{135}{8}\right)\frac{\mu}{135\tilde{L}}
\end{aligned}
$$

$$< \frac{\mu}{2\tilde{L}} .$$

This means that $(z, v) \in S_1$. The exact same procedure can now be carried out with z instead of y and we obtain in this way that all iterates lie in S_1. This can now be used together with the rate of convergence result in lemma 5.2.6. We can apply the bounds on various quantities which were obtained for S_1.

This gives

$$\begin{aligned}
\frac{\|q\|_{\overline{H}}}{\|p\|_H^2} &\leq \frac{3\sqrt{n}\left((M+1)\, B(x) + 2C(x) + D(x)\right)}{2B(x)\sqrt{\lambda_{\overline{H}}}} \\
&= \left(\frac{2B(x)\sqrt{\lambda_{\overline{H}}}}{3\sqrt{n}\left((M+1)\, B(x) + 2C(x) + D(x)\right)}\right)^{-1} \\
&\leq \left(\frac{\mu\sqrt{\lambda^*}}{18\sqrt{n}\left((M+1)\,\mu + 2C^* + D^*\right)}\right)^{-1} \\
&\leq \frac{1}{10\beta^*} .
\end{aligned}$$

This completes the proof.

□

5.6 Conclusions

Under the same assumptions as in chapter 3 and with the additional assumption that $(M\mu) \sim \mathcal{O}(1)$, we can see from the form of β^* that the orders of magnitude will not change much compared to the quadratic case. The only difference is caused by the factor $\tilde{L}$.

Defining :

$$\tilde{\zeta} \triangleq \frac{1}{\tilde{L}} \left\{ \frac{1}{n}, \frac{(\ln n)^3}{m - r} \right\} ,$$

we obtain, exactly as in the quadratic case, that the overall complexity to reach a point with accuracy ϵ is given by :

$$\mathcal{O}\left(\sqrt{m} \ln \left(\mu\tilde{\zeta}\right)^{-1}\right) + \mathcal{O}\left(\left(\frac{\ln \epsilon^{-1} - \ln \left(\mu\tilde{\zeta}\right)^{-1}}{\ln n}\right) \ln \ln \epsilon^{-1}\right) .$$

As in chapter 3, this can also be written as

$$\mathcal{O}\left(\sqrt{m} \ln C + \sqrt{m} \ln \tilde{\zeta}^{-1} + \left(\frac{\ln \epsilon^{-1} - \ln C - \ln \tilde{\zeta}^{-1}}{\ln n}\right) \ln \ln \epsilon^{-1}\right) .$$

Chapter 6

Appendix

6.1 Frobenius' formula

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

$$\begin{pmatrix} A & B \\ C & D \end{pmatrix}^{-1} = \begin{pmatrix} (A-BD^{-1}C)^{-1} & -(A-BD^{-1}C)^{-1}BD^{-1} \\ -D^{-1}C(A-BD^{-1}C)^{-1} & D^{-1}+D^{-1}C(A-BD^{-1}C)^{-1}BD^{-1} \end{pmatrix}$$

6.2 Some lemmas on the norms of special matrices

For $\quad A \in I\!R^{(n,n)} \ , \ C \in I\!R^{(r,n)} \ , \ x \in I\!R^n \ , \ y \in I\!R^r \quad$ with :

(1) $\left\|C^T y\right\| \geq m_0 \|y\| \quad (m_0 > 0) \quad \forall y \in I\!R^r$, which implies

$$\left\|\left(CC^T\right)^{-1}\right\| \leq \frac{1}{m_0^2}$$

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

the following three lemma's hold (see [23]).

Lemma 6.2.1

$$\begin{array}{ll} \textit{If} & \dfrac{1}{\mu} \geq \dfrac{2\|A\|\|C\|^2}{m_0^4}\left(1+\dfrac{2}{l_0}+\dfrac{l_0}{2\|A\|}\right) , \\ \textit{then} & \left\|\left(A+\dfrac{1}{\mu}C^TC\right)^{-1}\right\| \leq \dfrac{2}{l_0} . \end{array}$$

Lemma 6.2.2

$$\begin{array}{ll} \textit{If} & \dfrac{1}{\mu} \geq \dfrac{2\|A\|\|C\|}{m_0^3}\left(1+\dfrac{\|A\|^2}{l_0^2}\right)^{\frac{1}{2}} , \\ \textit{then} & \left\|\left(A+\dfrac{1}{\mu}C^TC\right)^{-1} C^T\right\| \leq \dfrac{2\mu\|C\|}{m_0^3}\left(1+\dfrac{\|A\|^2}{l_0^2}\right)^{\frac{1}{2}} . \end{array}$$

Lemma 6.2.3

$$\begin{array}{ll} \textit{If} & \dfrac{1}{\mu} \geq \dfrac{2\|A\|\|C\|}{m_0^3}\left(1+\dfrac{\|A\|^2}{l_0^2}\right)^{\frac{1}{2}} , \\ \textit{then} & \left\|I_r - \dfrac{1}{\mu}C\left(A+\dfrac{1}{\mu}C^TC\right)^{-1} C^T\right\| \leq \dfrac{2\mu\|A\|\|C\|^2}{m_0^3}\left(1+\dfrac{\|A\|^2}{l_0^2}\right)^{\frac{1}{2}} . \end{array}$$

6.3 An inequality for $\ln(x+1)$

We want to estimate $|\ln(1+x)|$ for $|x| \leq \frac{1}{8}$. From inequalities 4.1.33 in Abramowitz & Stegun [1] we have

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

Therefore, since $|x| \leq \frac{1}{8}$:

$$\begin{aligned} |\ln(1+x)| &\leq \max\left\{1, \frac{1}{1-|x|}\right\}|x| \\ &\leq \max\left\{1, \frac{8}{7}\right\}|x| \,. \end{aligned}$$

This means that

$$|\ln(1+x)| \leq \frac{8}{7}|x| \,. \tag{6.1}$$

6.4 Some inequalities based on the relative Lipschitz condition

Lemma 6.4.1 *If the functions f_i satisfy the relative Lipschitz condition with constant M, then the following inequalities hold :*

(1) $(1 - M\|x-y\|)\|\nabla^2 f_i(x)\| \leq \|\nabla^2 f_i(y)\| \leq (1 + M\|x-y\|)\|\nabla^2 f_i(x)\|$

(2) $\|\nabla f_i(y)\| \leq \|\nabla f_i(x)\| + (1 + M\|x-y\|)\,\|x-y\|\|\nabla^2 f_i(x)\,.$

or :

$$\|\nabla f_i(y)\| \leq \|\nabla f_i(x)\| + (1 + M\|x-y\|)\,\|x-y\|\|\nabla^2 f_i(y)\,.$$

Proof :

We start by observing that for any positive definite symmetric matrix A, the largest eigenvalue of which we denote by Λ_A, the following is valid :

$$\|A\| = \max_d \frac{\|Ad\|}{\|d\|} = \Lambda_A = \max_d \frac{d^T A d}{\|d\|^2}\,.$$

We also note that for the same matrix A :

$$d^T A^2 d = \left(A^{\frac{1}{2}} d\right)^T A \left(A^{\frac{1}{2}} d\right) \leq \Lambda_A \|A^{\frac{1}{2}} d\|^2 = \|A\| \left(d^T A d\right) \; .$$

Using the relative Lipschitz condition on $\nabla^2 f_i(x)$ and noting that $-\nabla^2 f_i(x)$ is positive definite, this means that

$$\begin{aligned}
\|\nabla^2 f_i(y)\| &= \max_d \frac{-d^T \nabla^2 f_i(y) d}{\|d\|^2} \\
&= \max_d \frac{\left(-d^T \nabla^2 f_i(y) d + d^T \nabla^2 f_i(x) d\right) - d^T \nabla^2 f_i(x) d}{\|d\|^2} \\
&\leq \max_d \frac{\left|-d^T \nabla^2 f_i(y) d + d^T \nabla^2 f_i(x) d\right| - d^T \nabla^2 f_i(x) d}{\|d\|^2} \\
&\leq \max_d \frac{(1 + M\|x-y\|)\left(-d^T \nabla^2 f_i(x) d\right)}{\|d\|^2} \\
&\leq (1 + M\|x-y\|)\|\nabla^2 f_i(x)\| \; .
\end{aligned}$$

On the other hand we also have

$$\begin{aligned}
\|\nabla^2 f_i(y)\| &= \max_d \frac{-d^T \nabla^2 f_i(y) d}{\|d\|^2} \\
&= \max_d \frac{\left(-d^T \nabla^2 f_i(y) d + d^T \nabla^2 f_i(x) d\right) - d^T \nabla^2 f_i(x) d}{\|d\|^2} \\
&\geq \max_d \frac{-\left|-d^T \nabla^2 f_i(y) d + d^T \nabla^2 f_i(x) d\right| - d^T \nabla^2 f_i(x) d}{\|d\|^2} \\
&\geq \max_d \frac{(1 - M\|x-y\|)\left(-d^T \nabla^2 f_i(x) d\right)}{\|d\|^2} \\
&\geq (1 - M\|x-y\|)\|\nabla^2 f_i(x)\| \; .
\end{aligned}$$

This proves the first part of the lemma.

For the second part, we develop $\nabla f_i(y)$ in a Taylor series about x, which gives :

$$\|\nabla f_i(y) - \nabla f_i(x)\| \leq \|\nabla^2 f_i(\tilde{x})\| \|y - x\| ,$$

where $\tilde{x}$ is a point between x and y. Using the first part of the lemma yields

$$\begin{aligned}
\|\nabla f_i(y)\| &\leq \|\nabla f_i(x)\| + \|\nabla^2 f_i(\tilde{x})\| \|x - y\| \\
&\leq \|\nabla f_i(x)\| + \|\nabla^2 f_i(\tilde{x})\| \|x - y\| \\
&\leq \|\nabla f_i(x)\| + (1 + M\|x - y\|) \|x - y\| \|\nabla^2 f_i(x)\| .
\end{aligned}$$

or :

$$\leq \|\nabla f_i(x)\| + (1 + M\|x - y\|) \|x - y\| \|\nabla^2 f_i(y)\| .$$

This completes the proof of the lemma.

□

Lemma 6.4.2 *If the functions f_i satisfy the relative Lipschitz condition with constant M, then the following inequalities hold :*

$$\begin{aligned}
&\left(d^T \nabla f_i(y)\right)^2 \\
&\quad \leq 2\left(d^T \nabla f_i(x)\right)^2 + 2\|\nabla^2 f_i(x)\| (1 + M\|x - y\|)^2 \|x - y\|^2 \left(-d^T \nabla^2 f_i(x) d\right) .
\end{aligned}$$

$$\begin{aligned}
&\left(d^T \nabla f_i(y)\right)^2 \\
&\quad \geq \frac{1}{2}\left(d^T \nabla f_i(x)\right)^2 - (1 + M\|x - y\|)^2 \|x - y\|^2 \|\nabla^2 f_i(x)\| \left(-d^T \nabla^2 f_i(x) d\right) .
\end{aligned}$$

Proof :

Developing $d^T\nabla f_i(y)$ in a Taylor series about x gives :

$$d^T\nabla f_i(y) = d^T\nabla f_i(x) + d^T\nabla^2 f_i(\tilde{x})(y-x)\ ,$$

where $\tilde{x}$ is a point between x and y. Using the previous lemma and bearing in mind that $(a+b)^2 \leq 2(a^2+b^2)$, we can write

$$\begin{aligned}
&\left(d^T\nabla f_i(y)\right)^2 \\
&\quad\leq 2\left[\left(d^T\nabla f_i(x)\right)^2 + \left(d^T\nabla^2 f_i(\tilde{x})(y-x)\right)^2\right] \\
&\quad\leq 2\left[\left(d^T\nabla f_i(x)\right)^2 + \|x-y\|^2\|\nabla^2 f_i(\tilde{x})d\|^2\right] \\
&\quad\leq 2\left[\left(d^T\nabla f_i(x)\right)^2 + \|x-y\|^2 d^T\left(\nabla^2 f_i(\tilde{x})\right)^2 d\right] \\
&\quad\leq 2\left[\left(d^T\nabla f_i(x)\right)^2 + \|x-y\|^2\|\nabla^2 f_i(\tilde{x})\|\left(-d^T\nabla^2 f_i(\tilde{x})d\right)\right] \\
&\quad\leq 2\left[\left(d^T\nabla f_i(x)\right)^2 + \left(1+M\|x-y\|\right)^2\|x-y\|^2\|\nabla^2 f_i(x)\|\left(-d^T\nabla^2 f_i(x)d\right)\right]\ .
\end{aligned}$$

The last inequality comes from

$$\|\tilde{x}-x\| \leq \|x-y\|\ .$$

The first part of the lemma is therefore proven.

For the second inequality, we develop $d^T\nabla f_i(x)$ in a Taylor series around y. Analogously to the first part, we obtain

$$\begin{aligned}
&\left(d^T\nabla f_i(x)\right)^2 \\
&\quad\leq 2\left[\left(d^T\nabla f_i(y)\right)^2 + \|x-y\|^2\|\nabla^2 f_i(\tilde{x})\|\left(-d^T\nabla^2 f_i(\tilde{x})d\right)\right]
\end{aligned}$$

$$\leq 2\left[\left(d^T\nabla f_i(y)\right)^2 + (1+M\|x-y\|)^2\|x-y\|^2\|\nabla^2 f_i(x)\|\left(-d^T\nabla^2 f_i(x)d\right)\right] .$$

This last inequality is based on

$$\|x-\tilde{x}\| \leq \|x-y\| .$$

We therefore have

$$\begin{aligned}&\left(d^T\nabla f_i(y)\right)^2\\ &\quad\geq \frac{1}{2}\left(d^T\nabla f_i(x)\right)^2 - (1+M\|x-y\|)^2\|x-y\|^2\|\nabla^2 f_i(x)\|\left(-d^T\nabla^2 f_i(x)d\right) .\end{aligned}$$

This completes the proof.

□

Lemma 6.4.3 *If the functions f_i satisfy the relative Lipschitz condition with constant M, then the following inequality holds :*

$$\left|\sum_{s,t=1}^{m}\frac{\partial^3\phi(x)}{\partial x_s\partial x_t\partial x_i}d_s d_t\right| \leq \left(\frac{(M+1)B(x)+2C(x)+D(x)}{B(x)}\right)\|d\|_H^2 .$$

Proof :

We start by computing an expression for the third derivative mentioned in the statement of the lemma.

$$\sum_{s,t=1}^{m} \frac{\partial^3 \phi(x)}{\partial x_s \partial x_t \partial x_i} d_s d_t =$$
$$- \sum_{s,t=1}^{m} \sum_{j=1}^{m} u_j \left(\frac{\frac{\partial^3 f_j(x)}{\partial x_s \partial x_t \partial x_i}}{f_j(x)+\mu} + \frac{\left(-\frac{\partial^2 f_j(x)}{\partial x_s \partial x_t}\right)\left(\frac{\partial f_j(x)}{\partial x_i}\right)}{(f_j(x)+\mu)^2} + \frac{\left(-\frac{\partial^2 f_j(x)}{\partial x_s \partial x_i}\right)\left(\frac{\partial f_j(x)}{\partial x_t}\right)}{(f_j(x)+\mu)^2} \right.$$
$$\left. + \frac{\left(-\frac{\partial^2 f_j(x)}{\partial x_t \partial x_i}\right)\left(\frac{\partial f_j(x)}{\partial x_s}\right)}{(f_j(x)+\mu)^2} + 2\frac{\left(\frac{\partial f_j(x)}{\partial x_s}\right)\left(\frac{\partial f_j(x)}{\partial x_t}\right)\left(\frac{\partial f_j(x)}{\partial x_i}\right)}{(f_j(x)+\mu)^3} \right) d_s d_t \qquad (6.2)$$

We now compute bounds on the different terms making up the RHS of the above equation. We start with the following expression :

$$\begin{aligned}
\left| \sum_{s,t=1}^{m} \frac{\partial^3 f_j(x)}{\partial x_s \partial x_t \partial x_i} d_s d_t \right| &= \left| \sum_{s,t=1}^{m} \left(\lim_{\eta \to 0} \frac{1}{\eta} \left(\frac{\partial^2 f_j(x+\eta e_{(i)})}{\partial x_s \partial x_t} - \frac{\partial^2 f_j(x)}{\partial x_s \partial x_t} \right) \right) d_s d_t \right| \\
&\le \left| \lim_{\eta \to 0} \frac{1}{\eta} \left(d^T \nabla^2 f_j(x+\eta e_{(i)}) d - d^T \nabla^2 f_j(x) d \right) \right| \\
&\le \lim_{\eta \to 0} \frac{1}{|\eta|} \left| \left(d^T \nabla^2 f_j(x+\eta e_{(i)}) d - d^T \nabla^2 f_j(x) d \right) \right| \\
&\le \lim_{\eta \to 0} \frac{M}{|\eta|} \|\eta e_{(i)}\| \left(-d^T \nabla^2 f_j(x) d \right) \\
&\le M \left(-d^T \nabla^2 f_j(x) d \right) .
\end{aligned}$$

This gives the following bound on the first term in the RHS of (6.2).

$$\begin{aligned}
\left| \sum_{s,t=1}^{m} \sum_{j=1}^{m} u_j \frac{\frac{\partial^3 f_j(x)}{\partial x_s \partial x_t \partial x_i}}{f_j(x)+\mu} d_s d_t \right| &\le \left| \sum_{j=1}^{m} \frac{u_j}{f_j(x)+\mu} \sum_{s,t=1}^{m} \frac{\partial^3 f_j(x)}{\partial x_s \partial x_t \partial x_i} d_s d_t \right| \\
&\le \sum_{j=1}^{m} \frac{u_j}{f_j(x)+\mu} \left| \sum_{s,t=1}^{m} \frac{\partial^3 f_j(x)}{\partial x_s \partial x_t \partial x_i} d_s d_t \right|
\end{aligned}$$

$$
\begin{aligned}
&\leq M \sum_{j=1}^{m} u_j \frac{-d^T \nabla^2 f_j(x) d}{f_j(x)+\mu} \,. \\
&\leq M \|d\|_H^2 \,.
\end{aligned}
$$

For the following two terms in (6.2) we obtain :

$$
\begin{aligned}
&\left| \sum_{s,t=1}^{m} \sum_{j=1}^{m} u_j \left(\frac{\left(-\frac{\partial^2 f_j(x)}{\partial x_s \partial x_i}\right)\left(\frac{\partial f_j(x)}{\partial x_t}\right)}{(f_j(x)+\mu)^2} + \frac{\left(-\frac{\partial^2 f_j(x)}{\partial x_t \partial x_i}\right)\left(\frac{\partial f_j(x)}{\partial x_s}\right)}{(f_j(x)+\mu)^2} \right) d_s d_t \right| \\
&\leq \sum_{j=1}^{m} \frac{u_j}{(f_j(x)+\mu)^2} \left| \sum_{s,t=1}^{m} \left(\left(-\frac{\partial^2 f_j(x)}{\partial x_s \partial x_i}\right)\left(\frac{\partial f_j(x)}{\partial x_t}\right) + \left(-\frac{\partial^2 f_j(x)}{\partial x_t \partial x_i}\right)\left(\frac{\partial f_j(x)}{\partial x_s}\right) \right) d_s d_t \right| \\
&\leq \sum_{j=1}^{m} \frac{u_j}{(f_j(x)+\mu)^2} \left| 2\left(d^T \nabla f_j(x)\right)\left(-\nabla^2 f_j(x) d\right)_i \right| \\
&\leq \sum_{j=1}^{m} u_j \frac{\left(d^T \nabla f_j(x)\right)^2}{(f_j(x)+\mu)^2} + \sum_{j=1}^{m} u_j \frac{(\nabla^2 f_j(x) d)_i^2}{(f_j(x)+\mu)^2} \\
&\leq \sum_{j=1}^{m} u_j \frac{\left(d^T \nabla f_j(x)\right)^2}{(f_j(x)+\mu)^2} + \frac{1}{B(x)} \sum_{j=1}^{m} u_j \frac{\|\nabla^2 f_j(x) d\|^2}{f_j(x)+\mu} \\
&\leq \sum_{j=1}^{m} u_j \frac{\left(d^T \nabla f_j(x)\right)^2}{(f_j(x)+\mu)^2} + \frac{D(x)}{B(x)} \sum_{j=1}^{m} u_j \frac{-d^T \nabla^2 f_j(x) d}{f_j(x)+\mu} \\
&\leq \left(1 + \frac{D(x)}{B(x)}\right) \|d\|_H^2 \,.
\end{aligned}
$$

For the last two terms in (6.2) we can compute the following bound :

$$
\left| \sum_{s,t=1}^{m} \sum_{j=1}^{m} u_j \left(\frac{\left(-\frac{\partial^2 f_j(x)}{\partial x_s \partial x_t}\right)\left(\frac{\partial f_j(x)}{\partial x_i}\right)}{(f_j(x)+\mu)^2} + 2 \frac{\left(\frac{\partial f_j(x)}{\partial x_s}\right)\left(\frac{\partial f_j(x)}{\partial x_t}\right)\left(\frac{\partial f_j(x)}{\partial x_i}\right)}{(f_j(x)+\mu)^3} \right) d_s d_t \right|
$$

$$
\begin{aligned}
&\le \left|\sum_{j=1}^{m} u_j \left(\sum_{s,t=1}^{m}\left(\frac{\left(-\frac{\partial^2 f_j(x)}{\partial x_s \partial x_t}\right)\left(\frac{\partial f_j(x)}{\partial x_i}\right)}{(f_j(x)+\mu)^2} + 2\frac{\left(\frac{\partial f_j(x)}{\partial x_s}\right)\left(\frac{\partial f_j(x)}{\partial x_t}\right)\left(\frac{\partial f_j(x)}{\partial x_i}\right)}{(f_j(x)+\mu)^3}\right) d_s d_t\right)\right| \\
&\le \left|\sum_{j=1}^{m} u_j \left(\frac{\left(-d^T\nabla^2 f_j(x) d\right)(\nabla f_j(x))_i}{(f_j(x)+\mu)^2} + 2\frac{\left(d^T\nabla f_j(x)\right)^2(\nabla f_j(x))_i}{(f_j(x)+\mu)^3}\right)\right| \\
&\le \sum_{j=1}^{m} u_j \left(\frac{\left(-d^T\nabla^2 f_j(x) d\right)\|\nabla f_j(x)\|}{(f_j(x)+\mu)^2} + 2\frac{\left(d^T\nabla f_j(x)\right)^2\|\nabla f_j(x)\|}{(f_j(x)+\mu)^3}\right) \\
&\le 2\frac{C(x)}{B(x)}\sum_{j=1}^{m} u_j \left(\frac{\left(-d^T\nabla^2 f_j(x) d\right)}{f_j(x)+\mu} + \frac{\left(d^T\nabla f_j(x)\right)^2}{(f_j(x)+\mu)^2}\right) \\
&\le 2\frac{C(x)}{B(x)}\|d\|_H^2 .
\end{aligned}
$$

Taking all of these bounds together, we obtain :

$$
\left|\sum_{s,t=1}^{m} \frac{\partial^3\phi(x)}{\partial x_s \partial x_t \partial x_i} d_s d_t\right| \le \left(\frac{(M+1)B(x)+2C(x)+D(x)}{B(x)}\right)\|d\|_H^2
$$

This completes the proof.

□

Lemma 6.4.4 *If*

$$
\|x-y\| \le \min\left\{\frac{1}{\alpha M}, \frac{1}{\alpha}\right\} \qquad (\alpha > 3)\ , \tag{6.3}
$$

then

$$
(1) \quad \|\nabla f_j(y)\| + \|\nabla^2 f_j(y)\| \ge \left(\frac{\alpha^2-2\alpha-1}{\alpha^2}\right)\left(\|\nabla f_j(x)\| + \|\nabla^2 f_j(x)\|\right)
$$

$$
(2) \quad \|\nabla f_j(y)\| + \|\nabla^2 f_j(y)\| \le \left(\frac{\alpha+1}{\alpha}\right)^2\left(\|\nabla f_j(x)\| + \|\nabla^2 f_j(x)\|\right)
$$

$$\text{(3)} \quad C(y) + D(y) \geq \left(\frac{\alpha^2 - 2\alpha - 1}{\alpha^2} \right) (C(x) + D(x))$$

$$\text{(4)} \quad C(y) + D(y) \leq \left(\frac{\alpha + 1}{\alpha} \right)^2 (C(x) + D(x))$$

$$\text{(5)} \quad 2C(y) + D(y) \geq \left(\frac{\alpha^2 - 5\alpha - 2}{\alpha^2} \right) (2C(x) + D(x))$$

$$\text{(6)} \quad 2C(y) + D(y) \leq \left(\frac{\alpha^2 + 3\alpha + 2}{\alpha^2} \right) (2C(x) + D(x)) \ .$$

Proof :

From lemma 6.4.1, we have

$$\|\nabla f_j(x)\| \leq \|\nabla f_j(y)\| + (1 + M\|x - y\|) \|x - y\| \|\nabla^2 f_j(x)\| \ .$$

Therefore

$$\begin{aligned}
&\|\nabla f_j(y)\| + \|\nabla^2 f_j(y)\| \\
&\quad \geq \|\nabla f_j(x)\| - (1 + M\|x - y\|) \|x - y\| \|\nabla^2 f_j(x)\| + \|\nabla^2 f_j(y)\| \\
&\quad \geq \|\nabla f_j(x)\| + [1 - M\|x - y\| - (1 + M\|x - y\|) \|x - y\|] \|\nabla^2 f_j(x)\| \\
&\quad \geq \|\nabla f_j(x)\| + \left(\frac{\alpha - 1}{\alpha} - \frac{1}{\alpha} \left(\frac{\alpha + 1}{\alpha} \right) \right) \|\nabla^2 f_j(x)\| \\
&\quad \geq \left(\frac{\alpha^2 - 2\alpha - 1}{\alpha^2} \right) \left(\|\nabla f_j(x)\| + \|\nabla^2 f_j(x)\| \right) \ .
\end{aligned}$$

For the second inequality, we again use lemma 6.4.1, which yields :

$$\|\nabla f_j(y)\| + \|\nabla^2 f_j(y)\|$$

$$
\begin{aligned}
&\leq \|\nabla f_j(x)\| + (1 + M\|x-y\|)\,\|x-y\|\|\nabla^2 f_j(x)\| + \|\nabla^2 f_j(y)\| \\
&\leq \|\nabla f_j(x)\| + (1 + M\|x-y\|)\,(1 + \|x-y\|\|)\,\|\nabla^2 f_j(x)\| + \|\nabla^2 f_j(y)\| \\
&\leq \|\nabla f_j(x)\| + \left(\frac{\alpha+1}{\alpha}\right)^2 \|\nabla^2 f_j(x)\| \\
&\leq \left(\frac{\alpha+1}{\alpha}\right)^2 \left(\|\nabla f_j(x)\| + \|\nabla^2 f_j(x)\|\right) .
\end{aligned}
$$

The third and fourth inequalities follow from observing that if we take the maximum over all j's in the inequalities in the statement of lemma 6.4.1, then :

$$C(x) \leq C(y) + (1 + M\|x-y\|)\,\|x-y\|D(x)$$

$$(1 - M\|x-y\|)\,D(x) \leq D(y) \leq (1 + M\|x-y\|)\,D(x).$$

The rest of the proof is exactly as the proof for the first two inequalities.

For the fifth and sixth inequalities we rewrite $2C(x) + D(x)$ as :

$$2\,(C(x) + D(x)) - D(x) .$$

We then have

$$
\begin{aligned}
&2C(y) + D(y) \\
&\quad\geq 2\left(\frac{\alpha^2 - 2\alpha - 1}{\alpha^2}\right)(C(x) + D(x)) - D(y) \\
&\quad\geq 2\left(\frac{\alpha^2 - 2\alpha - 1}{\alpha^2}\right)C(x) + 2\left(\frac{\alpha^2 - 2\alpha - 1}{\alpha^2}\right)D(x) - (1 + M\|x-y\|)\,D(x) \\
&\quad\geq 2\left(\frac{\alpha^2 - 2\alpha - 1}{\alpha^2}\right)C(x) + \left(2\left(\frac{\alpha^2 - 2\alpha - 1}{\alpha^2}\right) - \frac{\alpha+1}{\alpha}\right)D(x)
\end{aligned}
$$

$$\geq \left(\frac{\alpha^2 - 5\alpha - 2}{\alpha^2}\right)(2C(x) + D(x)) \ .$$

On the other hand we have

$$\begin{aligned}
&2C(y) + D(y) \\
&\quad \leq 2\left(\frac{\alpha+1}{\alpha}\right)^2 (C(x) + D(x)) - D(y) \\
&\quad \leq 2\left(\frac{\alpha+1}{\alpha}\right)^2 C(x) + 2\left(\frac{\alpha+1}{\alpha}\right)^2 D(x) - (1 - M\|x-y\|)\, D(x) \\
&\quad \leq 2\left(\frac{\alpha+1}{\alpha}\right)^2 C(x) + \left(2\left(\frac{\alpha+1}{\alpha}\right)^2 - \frac{\alpha-1}{\alpha}\right) D(x) \\
&\quad \leq \left(\frac{\alpha^2 + 3\alpha + 2}{\alpha^2}\right)(2C(x) + D(x)) \ .
\end{aligned}$$

This completes the proof.

□

Lemma 6.4.5 *If*

$$\|x - y\| \leq \frac{B(x)}{\alpha L} \quad (\alpha > 3) \,, \tag{6.4}$$

then

$$\left(\frac{\alpha}{\alpha+1}\right)\frac{1}{f_i(x)+\mu} \leq \frac{1}{f_i(y)+\mu} \leq \left(\frac{\alpha}{\alpha-1}\right)\frac{1}{f_i(x)+\mu} \,.$$

Proof : We have

$$\begin{aligned}
f_i(y) + \mu &\leq f_i(x) + |f_i(y) - f_i(x)| + \mu \\
&\leq f_i(x) + \mu + L\|x - y\| \,,
\end{aligned}$$

and also

$$
\begin{aligned}
f_i(y) + \mu &\geq f_i(x) - |f_i(y) - f_i(x)| + \mu \\
&\geq f_i(x) + \mu - L\|x - y\| .
\end{aligned}
$$

The rest of the proof is exactly as in the quadratic case.

□

Lemma 6.4.6 *If*

$$
\|x - y\| \leq \min\left\{\frac{B(x)}{\alpha L}, \frac{B(x)}{\alpha}, \frac{1}{\alpha M}, \frac{1}{2\alpha D(x)}\right\} \quad (\alpha \geq 3) , \tag{6.5}
$$

then

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

Proof : Using the previous lemma, we can write

$$
\begin{aligned}
d^T H(y)d &= \sum_{i=1}^{m} u_i \left(\frac{-d^T\nabla^2 f_i(y)d}{(f_i(y)+\mu)} + \frac{\left(d^T\nabla f_i(y)\right)^2}{(f_i(y)+\mu)^2} \right) \\
&\leq \left(\frac{\alpha}{\alpha-1}\right)^2 \sum_{i=1}^{m} u_i \left(\frac{-d^T\nabla^2 f_i(y)d}{(f_i(x)+\mu)} + \frac{\left(d^T\nabla f_i(y)\right)^2}{(f_i(x)+\mu)^2} \right) .
\end{aligned}
$$

From the assumptions on $\|x - y\|$ and the relative Lipschitz condition we have

$$
\begin{aligned}
-d^T\nabla^2 f_i(y)d &\leq (1 + M\|x - y\|)\left(-d^T\nabla^2 f_i(x)d\right) \\
&\leq \left(\frac{\alpha+1}{\alpha}\right)\left(-d^T\nabla^2 f_i(x)d\right) .
\end{aligned}
$$

Furthermore, with lemma 6.4.2 :

$$
\begin{aligned}
\frac{\left(d^T\nabla f_i(y)\right)^2}{(f_i(x)+\mu)^2} &\le \frac{2\left(d^T\nabla f_i(x)\right)^2 + 2D(x)\left(1+M\|x-y\|\right)^2\|x-y\|^2\left(-d^T\nabla^2 f_i(x)d\right)}{(f_i(x)+\mu)^2} \\
&\le 2\frac{\left(d^T\nabla f_i(x)\right)^2}{(f_i(x)+\mu)^2} + \frac{2D(x)\left(1+M\|x-y\|\right)^2\|x-y\|\left(-d^T\nabla^2 f_i(x)d\right)}{\alpha\left(f_i(x)+\mu\right)} \\
&\le 2\frac{\left(d^T\nabla f_i(x)\right)^2}{(f_i(x)+\mu)^2} + \frac{\left(1+M\|x-y\|\right)^2\left(-d^T\nabla^2 f_i(x)d\right)}{\alpha^2\left(f_i(x)+\mu\right)} \\
&\le 2\frac{\left(d^T\nabla f_i(x)\right)^2}{(f_i(x)+\mu)^2} + \frac{1}{\alpha^2}\left(\frac{\alpha+1}{\alpha}\right)^2\frac{-d^T\nabla^2 f_i(x)d}{(f_i(x)+\mu)}\,.
\end{aligned}
$$

We therefore obtain

$$
d^T H(y)d = \left(\frac{\alpha}{\alpha-1}\right)^2\sum_{i=1}^m u_i\left(\left(\frac{(\alpha+1)\left(\alpha^3+\alpha+1\right)}{\alpha^4}\right)\frac{-d^T\nabla^2 f_i(x)d}{(f_i(x)+\mu)} + 2\frac{\left(d^T\nabla f_i(x)\right)^2}{(f_i(x)+\mu)^2}\right)\,.
$$

Since we assumed that $\alpha \ge 3$, this yields

$$
\begin{aligned}
d^T H(y)d &\le 2\left(\frac{\alpha}{\alpha-1}\right)^2\sum_{i=1}^m u_i\left(\frac{-d^T\nabla^2 f_i(x)d}{(f_i(x)+\mu)} + \frac{\left(d^T\nabla f_i(x)\right)^2}{(f_i(x)+\mu)^2}\right) \\
&\le 2\left(\frac{\alpha}{\alpha-1}\right)^2 d^T H(x)d\,.
\end{aligned}
$$

On the other hand, from the previous lemma we also have

$$
\begin{aligned}
d^T H(y)d &= \sum_{i=1}^m u_i\left(\frac{-d^T\nabla^2 f_i(y)d}{(f_i(y)+\mu)} + \frac{\left(d^T\nabla f_i(y)\right)^2}{(f_i(y)+\mu)^2}\right) \\
&\ge \left(\frac{\alpha}{\alpha+1}\right)^2\sum_{i=1}^m u_i\left(\frac{-d^T\nabla^2 f_i(y)d}{(f_i(x)+\mu)} + \frac{\left(d^T\nabla f_i(y)\right)^2}{(f_i(x)+\mu)^2}\right)\,.
\end{aligned}
$$

As before, with the relative Lipschitz condition and the assumptions on $\|x-y\|$, we can write

$$\begin{aligned}
-d^T\nabla^2 f_i(y)d &\geq (1-M\|x-y\|)\left(-d^T\nabla^2 f_i(x)d\right) \\
&\geq \left(\frac{\alpha-1}{\alpha}\right)\left(-d^T\nabla^2 f_i(x)d\right) .
\end{aligned}$$

Then, with lemma 6.4.2 :

$$\begin{aligned}
\frac{\left(d^T\nabla f_i(y)\right)^2}{\left(f_i(x)+\mu\right)^2} &\geq \frac{\frac{1}{2}\left(d^T\nabla f_i(x)\right)^2 - D(x)\left(1+M\|x-y\|\right)^2\|x-y\|^2\left(-d^T\nabla^2 f_i(x)d\right)}{\left(f_i(x)+\mu\right)^2} \\
&\geq \frac{1}{2}\frac{\left(d^T\nabla f_i(x)\right)^2}{\left(f_i(x)+\mu\right)^2} - \frac{D(x)\left(1+M\|x-y\|\right)^2\|x-y\|\left(-d^T\nabla^2 f_i(x)d\right)}{\alpha\left(f_i(x)+\mu\right)} \\
&\geq \frac{1}{2}\frac{\left(d^T\nabla f_i(x)\right)^2}{\left(f_i(x)+\mu\right)^2} - \frac{\left(1+M\|x-y\|\right)^2\left(-d^T\nabla^2 f_i(x)d\right)}{2\alpha^2\left(f_i(x)+\mu\right)} \\
&\geq \frac{1}{2}\frac{\left(d^T\nabla f_i(x)\right)^2}{\left(f_i(x)+\mu\right)^2} - \frac{1}{2\alpha^2}\left(\frac{\alpha+1}{\alpha}\right)^2\frac{-d^T\nabla^2 f_i(x)d}{\left(f_i(x)+\mu\right)} .
\end{aligned}$$

We therefore have

$$\begin{aligned}
&d^T H(y)d \\
&\geq \left(\frac{\alpha}{\alpha+1}\right)^2\sum_{i=1}^m u_i\left(\left(\frac{2\alpha^4-2\alpha^3-\alpha^2-2\alpha-1}{\alpha^4}\right)\frac{-d^T\nabla^2 f_i(x)d}{\left(f_i(x)+\mu\right)} + \frac{1}{2}\frac{\left(d^T\nabla f_i(x)\right)^2}{\left(f_i(x)+\mu\right)^2}\right) .
\end{aligned}$$

Since we assumed that $\alpha \geq 3$, this yields

$$d^T H(y)d \geq \frac{1}{2}\left(\frac{\alpha}{\alpha+1}\right)^2\sum_{i=1}^m u_i\left(\frac{-d^T\nabla^2 f_i(x)d}{\left(f_i(x)+\mu\right)} + \frac{\left(d^T\nabla f_i(x)\right)^2}{\left(f_i(x)+\mu\right)^2}\right)$$

$$\geq \frac{1}{2}\left(\frac{\alpha}{\alpha+1}\right)^2 d^T H(x) d\,.$$

This complete the proof.

□

In what follows, we will use the following notation :

$$\begin{aligned} H(x) &= H \\ H(y) &= \overline{H}\,. \end{aligned}$$

Lemma 6.4.7 *If*

$$\|x-y\| \leq \min\left\{\frac{B(x)}{\alpha L}, \frac{B(x)}{\alpha}, \frac{1}{\alpha M}, \frac{1}{\alpha}, \frac{1}{2\alpha D(x)}\right\} \qquad (\alpha \geq 3)\,,$$

then :

$$\begin{array}{ll}
(\mathbf{1}) & \left(\frac{\alpha-1}{\alpha}\right) B(x) \leq B(y) \leq \left(\frac{\alpha+1}{\alpha}\right) B(x) \\
(\mathbf{2}) & \left(\frac{\alpha-1}{\alpha}\right) D(x) \leq D(y) \leq \left(\frac{\alpha+1}{\alpha}\right) D(x) \\
(\mathbf{3}) & \frac{1}{\sqrt{2}}\left(\frac{\alpha}{\alpha+1}\right)^2 \frac{\sqrt{\lambda_H}}{D(x)} \leq \frac{\lambda_{\overline{H}}}{D(y)} \leq \sqrt{2}\left(\frac{\alpha}{\alpha-1}\right)^2 \frac{\sqrt{\lambda_H}}{D(x)} \\
(\mathbf{4}) & \frac{1}{2}\left(\frac{\alpha}{\alpha+1}\right)^2 \leq \frac{\lambda_{\overline{H}}}{\lambda_H} \leq 2\left(\frac{\alpha}{\alpha-1}\right)^2 \\
(\mathbf{5}) & \frac{1}{\sqrt{2}}\left(\frac{\alpha-1}{\alpha+1}\right) B(x)\sqrt{\lambda_H} \leq B(y)\sqrt{\lambda_{\overline{H}}} \leq \sqrt{2}\left(\frac{\alpha+1}{\alpha-1}\right) B(x)\sqrt{\lambda_H} \\
(\mathbf{6}) & \left(\frac{\alpha^2(\alpha-1)}{\sqrt{2}(\alpha+1)^3}\right) R_1(x) \leq R_1(y) \leq \left(\frac{\sqrt{2}\alpha^2(\alpha+1)}{(\alpha-1)(\alpha^2-2\alpha-1)}\right) R_1(x) \\
(\mathbf{7}) & \left(\frac{\alpha^{\frac{2}{3}}(\alpha-1)^{\frac{1}{3}}}{2^{\frac{1}{3}}(\alpha+1)}\right) R_2^{\frac{1}{3}}(x) \leq R_2^{\frac{1}{3}}(y) \leq \left(\frac{2^{\frac{1}{3}}\alpha^{\frac{2}{3}}(\alpha+1)^{\frac{1}{3}}}{\alpha-1}\right) R_2^{\frac{1}{3}}(x)
\end{array}$$

$$\text{(8)}\qquad \left(\frac{\alpha^6(\alpha-1)^3}{2^{\frac{3}{2}}(\alpha+1)^9}\right)\xi_1(x)\le\xi_1(y)\le\left(\frac{2^{\frac{3}{2}}\alpha^6(\alpha+1)^3}{(\alpha-1)^3(\alpha^2-2\alpha-1)^3}\right)\xi_1(x)$$

$$\text{(9)}\qquad \left(\frac{\alpha^6(\alpha-1)^3}{2^{\frac{3}{2}}(\alpha+1)^9}\right)\xi_2(x)\le\xi_2(y)\le\left(\frac{2^{\frac{3}{2}}\alpha^6(\alpha+1)^3}{(\alpha-1)^3(\alpha^2-2\alpha-1)^3}\right)\xi_2(x)$$

(10)

$$\frac{B(y)\sqrt{\lambda_{\overline{H}}}}{(M+1)B(y)+2C(y)+D(y)} \le \sqrt{2}\left(\frac{\alpha+1}{\alpha-1}\right)\left(\frac{\alpha^2}{\alpha^2-5\alpha-2}\right)\frac{B(x)\sqrt{\lambda_H}}{(M+1)B(x)+2C(x)+D(x)}$$

$$\frac{B(y)\sqrt{\lambda_{\overline{H}}}}{(M+1)B(y)+2C(y)+D(y)} \ge \frac{1}{\sqrt{2}}\left(\frac{\alpha-1}{\alpha+1}\right)\left(\frac{\alpha^2}{\alpha^2+3\alpha+2}\right)\frac{B(x)\sqrt{\lambda_H}}{(M+1)B(x)+2C(x)+D(x)}\,.$$

Proof : The inequalities (1), (2), and (3) follow immediately from lemmas 6.4.1, 6.4.5 and 6.4.6. Inequalities (4) and (5) follow immediately from the pairs of inequalities (1),(2) and (3).

For (6) we have

$$\frac{\sqrt{\lambda_{\overline{H}}}\,(f_j(y)+\mu)}{\|\nabla f_j(y)\|+\|\nabla^2 f_j(y)\|} \le \left(\frac{\alpha+1}{\alpha}\right)\left(\frac{\alpha^2}{\alpha^2-2\alpha-1}\right)\left(\frac{\sqrt{\lambda_{\overline{H}}}}{\sqrt{\lambda_H}}\right)^{\frac{1}{2}}\frac{\sqrt{\lambda_H}\,(f_j(x)+\mu)}{\|\nabla f_j(x)\|+\|\nabla^2 f_j(x)\|}\,,$$

and therefore, by taking the minimum over $j\in J_\mu$:

$$R_1(y)\le\left(\frac{\alpha+1}{\alpha}\right)\left(\frac{\alpha^2}{\alpha^2-2\alpha-1}\right)\left(\frac{\alpha\sqrt{2}}{\alpha-1}\right)R_1(x)\,.$$

We also have

$$\frac{\sqrt{\lambda_{\overline{H}}}\,(f_j(y)+\mu)}{\|\nabla f_j(y)\| + \|\nabla^2 f_j(y)\|}$$

$$\geq \left(\frac{\alpha-1}{\alpha}\right)\left(\frac{\alpha}{\alpha+1}\right)^2 \left(\frac{\sqrt{\lambda_{\overline{H}}}}{\sqrt{\lambda_H}}\right)^{\frac{1}{2}} \frac{\sqrt{\lambda_H}\,(f_j(x)+\mu)}{\|\nabla f_j(x)\| + \|\nabla^2 f_j(x)\|}\,,$$

which gives

$$R_1(y) \geq \left(\frac{\alpha-1}{\alpha}\right)\left(\frac{\alpha}{\alpha+1}\right)^2 \left(\frac{\alpha}{\sqrt{2}\,(\alpha+1)}\right) R_1(x)\,.$$

For $R_2(x)$ in (7) we obtain

$$R_2^{\frac{1}{3}}(y) = \min_{j\in J_\mu} \left(\frac{\sqrt{\lambda_{\overline{H}}}\,(f_j(y)+\mu)^{\frac{1}{3}}}{\|\nabla^2 f_j(y)\|^{\frac{1}{3}}}\right)$$

$$\leq \left(\frac{\alpha+1}{\alpha}\right)^{\frac{1}{3}} \left(\frac{\alpha}{\alpha-1}\right)^{\frac{1}{3}} \left(\frac{\alpha^{\frac{2}{3}} 2^{\frac{1}{3}}}{(\alpha-1)^{\frac{2}{3}}}\right) R_2^{\frac{1}{3}}(x)\,,$$

and similarly :

$$R_2^{\frac{1}{3}}(y) \geq \left(\frac{\alpha-1}{\alpha}\right)^{\frac{1}{3}} \left(\frac{\alpha}{\alpha+1}\right)^{\frac{1}{3}} \left(\frac{\alpha^{\frac{2}{3}}}{2^{\frac{1}{3}}\,(\alpha+1)^{\frac{2}{3}}}\right) R_2^{\frac{1}{3}}(x)\,,$$

which proves the inequalities (7).

The bounds for $\xi_1(x)$ in (8) follow immediately from those on $R_1(x)$ exactly as in the quadratic case.

We now prove inequalities (9) : From the definition of $\xi_2(x)$, we have that

$$\frac{1}{\xi_2(y)} = \frac{1}{\xi_1(y)} + \frac{24M}{7\sqrt{\lambda_{\overline{H}}}}$$

$$= \frac{1}{\xi_1(y)} + \left(\frac{\lambda_H}{\lambda_{\overline{H}}}\right)^{\frac{1}{2}} \frac{24M}{7\sqrt{\lambda_H}} \, .$$

The inequalities then follow immediately from the previously derived bounds on $\xi_1(x)$ and $\frac{\lambda_H}{\lambda_{\overline{H}}}$.

For the last pair of inequalities, we first look at

$$(M+1)B(y) + 2C(x) + D(x) \, .$$

From the first pair of inequalities in the present lemma and from lemma 6.4.2 we have that

$$\begin{aligned}
&(M+1)B(y) + 2C(y) + D(y) \\
&\quad \leq \left(\frac{\alpha+1}{\alpha}\right)(M+1)B(x) + \left(\frac{\alpha^2+3\alpha+2}{\alpha^2}\right)(2C(x)+D(x)) \\
&\quad \leq \left(\frac{\alpha^2+3\alpha+2}{\alpha^2}\right)((M+1)B(x) + 2C(x) + D(x)) \ ,
\end{aligned}$$

and, analogously :

$$\begin{aligned}
&(M+1)B(y) + 2C(y) + D(y) \\
&\quad \geq \left(\frac{\alpha-1}{\alpha}\right)(M+1)B(x) + \left(\frac{\alpha^2-5\alpha-2}{\alpha^2}\right)(2C(x)+D(x)) \\
&\quad \geq \left(\frac{\alpha^2-5\alpha-2}{\alpha^2}\right)((M+1)B(x) + 2C(x) + D(x)) \ .
\end{aligned}$$

Taking reciprocals and with the help of inequalities (4), this gives :

$$\frac{B(y)\sqrt{\lambda_{\overline{H}}}}{(M+1)B(y)+2C(y)+D(y)}$$
$$\leq \sqrt{2}\left(\frac{\alpha+1}{\alpha}\right)\left(\frac{\alpha}{\alpha-1}\right)\left(\frac{\alpha^2}{\alpha^2-5\alpha-2}\right)\frac{B(x)\sqrt{\lambda_H}}{(M+1)B(x)+2C(x)+D(x)}$$
$$\frac{B(y)\sqrt{\lambda_{\overline{H}}}}{(M+1)B(y)+2C(y)+D(y)}$$
$$\geq \frac{1}{\sqrt{2}}\left(\frac{\alpha-1}{\alpha}\right)\left(\frac{\alpha}{\alpha+1}\right)\left(\frac{\alpha^2}{\alpha^2+3\alpha+2}\right)\frac{B(x)\sqrt{\lambda_H}}{(M+1)B(x)+2C(x)+D(x)} \ .$$

This concludes the proof.

□

6.5 A lemma about convex functions

Lemma 6.5.1 *For a twice continuously differentiable function $\psi(x)$, defined on a convex set $F \subset I\!R^n$, such that*

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

and with

$$\hat{x} = argmin_x \psi(x) \ ,$$

the following inequality holds :

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

Proof :

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

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

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

$$\begin{aligned} \|\nabla\psi(x)\| &\leq \|\nabla^2\psi(\tilde{x})\|\|x - \hat{x}\| \\ &\leq \Lambda\|x - \hat{x}\| . \end{aligned}$$

$\square$

Bibliography

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

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

[3] C.W. Carroll. The created response surface technique for optimizing nonlinear restrained systems. *Operations Research*, 9(2):169–184, 1961.

[4] 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.

[5] D.Goldfarb and M.J. Todd. Linear programming. In G.L. Nemhauser et al., editor, *Handbooks in OR & MS*, volume 1, chapter 2, pages 73–170. Elsevier Science Publishers B.V. (North-Holland), 1989.

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

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

[8] J. Franklin. Convergence in Karmarkar's algorithm for linear programming. *SIAM Journal of Numerical Analysis*, 24(4):928–945, 1987.

[9] R. Frisch. The logarithmic potential method for solving linear programming problems. Memorandum of the University Institute of Economics. Oslo, Norway, 1955.

[10] R. Frisch. The logarithmic potential method of convex programming. Memorandum of the University Institute of Economics. Oslo, Norway, 1955.

[11] 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*, 36:183–209, 1986.

[12] C.C. Gonzaga. An algorithm for solving linear programming problems in $\mathcal{O}(n^3L)$ operations. Memorandum No. UCB/ERL M87/10, Electronics Research Laboratory, College of Engineering, University of California, Berkeley, CA, 1987.

[13] F. Jarre. A new polynomial time algorithm for linear programming. *Combinatorica*, 4:373–395, 1984.

[14] F. Jarre. *The Method of Analytic Centers for Smooth Convex Programs.* PhD thesis, Institut fur Angewandte Mathematik und Statistik, Universitat Wurzburg, Wurzburg, FRG, 1989.

[15] N. Karmarkar. A new polynomial time algorithm for linear programming. *Combinatorica*, 4:373–395, 1984.

[16] H.B. Keller. *Numerical Methods in Bifurcation Problems.* Springer-Verlag, New York, 1987.

[17] L.G. Khachian. A polynomial algorithm in linear programming. *Doklady Akademiia Nauk SSSR*, 244:1093–1096, 1979 (in Russian).

[18] A.Yu. Levin. On an algorithm for the minimization of convex functions. *Doklady Akademiia Nauk SSSR*, 160:1244–1247, 1965 (in Russian).

[19] Yu.E. Nesterov and A.S. Nemirovsky. *Self-Concordant Functions and Polynomial Time Methods in Convex Programming.* CEMI, Moscow, USSR, 1989.

[20] D.J. Newman. Location of the maximum on unimodal surfaces. *Journal of the Association for Computing Machinery*, 12:395–398, 1965.

[21] E. Polak. *Computational Methods in Optimization.* Academic Press, New York, 1971.

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

[23] 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(1):34–46, 1973 (in Russian).

[24] R. Polyak. Smooth optimization methods for solving nonlinear extremal and equilibrium problems with constraints. In *Abstracts of the papers*, Bonn, FRG, 1982. The 11th International Symposium on Mathematical Programming.

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

[26] R. Polyak. Modified barrier functions, theory and methods. Research Report RC15886, IBM T.J.Watson Research Center, Yorktown Heights, NY, 1990. To appear in Mathematical Programming.

[27] J. Renegar and M. Shub. Simplified complexity analysis for Newton LP methods. Technical Report No. 807, School of Operations Research and Industrial Engineering, College of Engineering, Cornell University, Ithaca, NY, 1988.

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

[29] R.T. Rockafellar. Lagrange multipliers in optimization. *SIAM-AMS Proceedings*, 9:145–168, 1976.

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

[31] N.Z. Shor. Utilization of the operation of space dilatation in the minimization of convex functions. *Kibernetika*, 1:6–12, 1970 (in Russian).

[32] D.B. Yudin and A.S. Nemirovsky. Informational complexity and efficient methods for the solution of convex extremal problems. *Ekonomika i Matematicheskie Metody*, 12:357–369, 1976 (in Russian).