Path-following barrier and penalty methods for linearly constrained problems

D. Al-Mutairi\textsuperscript{a)}, C.Grossmann\textsuperscript{b,\ast)}, and Vu Kim Tuan\textsuperscript{b)}

Abstract

In the present paper some barrier and penalty methods (e.g. logarithmic barriers, SUMT, exponential penalties), which define a continuously differentiable primal and dual path, applied to linearly constrained convex problems are studied. In particular, the radius of convergence of Newton’s method depending on the barrier and penalty parameter is estimated. Unlike using self-concordance properties the convergence bounds are derived by direct estimations of the solutions of the Newton equations. The obtained results establish parameter selection rules which guarantee the overall convergence of the considered barrier and penalty techniques with only a finite number of Newton steps at each parameter level. Moreover, the obtained estimates support scaling method which uses approximate dual multipliers as available in barrier and penalty methods.

AMS Subject Classifications: 49M15, 65K10, 65H10, 90C25

Key words: barrier methods, penalty methods, SUMT, Newton’s method, path-following, convex programming, interior-point methods, complexity

1 Introduction

Barrier and penalty methods form ones of the basic techniques for treating constraint optimization problems. The basic idea of these methods consists in embedding the original problem into a family of specifically constructed unconstrained auxiliary problems. Unlike the classic approach, e.g. as in the monographs Fiacco/McCormick \cite{8}, Grossmann/Kaplan \cite{12}, Lootsma \cite{17}, which concentrates on properties of the auxiliary problems and their solutions, more recently the main focus in research on barrier and penalty methods has shifted to path-following concepts as done in various interior point investigations (cf. \cite{9}, \cite{13}, \cite{14}, \cite{15}, \cite{16}, \cite{19}, \cite{18}). The concept of self-concordance introduced by Nesterov/Nemirovskii \cite{19} (see also \cite{18}) provides a quite general tool for studying the convergence behavior of Newton’s method in the case of asymptotically ill-conditioned limit problems which are

\textsuperscript{a)} Kuwait University, Department of Statistics and Operations Research,

\textsuperscript{b)} Kuwait University, Department of Mathematics and Computer Science,

\textsuperscript{\ast)} on leave from TU Dresden, Germany
typical for the ones generated in barrier and penalty methods. However, in this approach
the major task is shifted to verification of the self-concordance property. It may be diffi-
cult, particularly in sequential unconstraint minimization techniques which do not base on
logarithmic barrier functions.

Taking notice of linear invariance properties of Newton’s method, which are not reflected
in the standard Kantorovich analysis of Newton’s method, Deuflhard (cf. [4], [5]) proposed
variant and contravariant transformations to adapt convergence proofs of Newton’s method
to possible ill-posedness of a root finding problem.

The aim of the present paper is to study different types of barrier and penalty methods
applied to linearly constrained convex optimization problems as path-following algorithms
(cf. [1], [3]), where the radius of convergence of Newton’s method (compare [2]) is estimated
in dependence of the barrier-penalty parameter using informations about the continuous path
of the related primal and dual problem (see [8], [17]). Finally, a parameter selection rule,
which guarantees the overall convergence of the considered barrier and penalty technique
with only one Newton step at each parameter level, is proposed. It enables us to estimate
the complexity of the method.

A similar study of the convergence behavior of barrier methods using refined estimates
for the radius of convergence of Newton’s method has been recently published by Wright
[20]. In that paper, on one hand a more general problem with nonlinear constraints has been
investigated, but on the other hand the type of barrier functions is limited to logarithmic
ones.

A quite general approach to parameter selection in singular path-following with trun-
cated Newton techniques was studied in [11], while in the present paper the main focus
is directed to special cases of barrier and penalty techniques for problem (1.1). Unlike in
[10], the investigations here do not restrict to logarithmic barriers, as often in interior-point
algorithms. Moreover, the proposed approach uses neither self-concordance (cf. [18], [19])
nor affine transformations (cf. [4], [5]) directly.

Let $f : \mathbb{R}^n \to \mathbb{R}$ and $g_i : \mathbb{R}^n \to \mathbb{R}$, $i = 1, \ldots, m$, denote a twice continuously
differentiable convex functions, where $f$ is assumed to be strongly convex. In the first part
of our investigation we deal with convex programming problems

$$f(x) \to \min \quad \text{subject to} \quad x \in G := \{ x \in \mathbb{R}^n : g(x) \leq 0 \} \quad (1.1)$$

under the assumption that $G^0 \neq \emptyset$, where

$$G^0 := \{ x \in \mathbb{R}^n : g(x) < 0 \}. $$

Throughout this paper ‘$\leq$’ and ‘$<$’ are understood as the natural componentwise semi-
ordering and $g(x) := (g_1(x), \ldots, g_m(x))^T$. The strong convexity of $f$ guarantees that the
level set $G(\hat{x}) := \{ x \in G : f(x) \leq f(\hat{x}) \}$ is bounded for any $\hat{x} \in G$. Hence, problem (1.1)

2
has a unique solution $x^*$ which can be characterized by the Karush-Kuhn-Tucker (KKT) conditions

$$
\exists y^* \in \mathbb{R}^m : \nabla f(x^*) + \sum_{i=1}^{m} y_i^* \nabla g_i(x^*) = 0,
\quad g(x^*) \leq 0,
\quad y^* \geq 0,
\quad y^*^T g(x^*) = 0.
$$

(1.2)

Under the made assumptions the KKT-conditions are necessary and sufficient for the optimality of $x^*$. In the sequel the active constraints at $x^*$ will be denoted by $I_0$, i.e.

$$
I_0 := \{ i \in \{1, \ldots, m\} : g_i(x^*) = 0 \}.
$$

2 Some basic properties of barrier-penalty methods

Before studying the behavior of a truncated Newton method applied to the barrier-penalty techniques for the numerical solution of problem (1.1) we discuss some important properties of barrier-penalty methods.

Let $\mathbb{R}_{++} := \{ t \in \mathbb{R} : t > 0 \}$ and let $\bar{\mathbb{R}} := \mathbb{R} \cup \{+\infty\}$, i.e. $\bar{\mathbb{R}}$ denotes the set of real numbers completed by the improper number $+\infty$. In the present paper we study the iterative treatment of barrier-penalty methods which incorporate the constraints from problem (1.1) into unconstrained auxiliary problems

$$
f(x) + \sum_{i=1}^{m} \varphi (g_i(x), s) \rightarrow \min \ \
$$

subject to $x \in B_s := \{ x \in \mathbb{R}^n | \ varphi (g_i(x), s) < +\infty, i = 1, \ldots, m \}$, by means of a parametric barrier-penalty function $\varphi (\cdot, \cdot) : \mathbb{R} \times \mathbb{R}_{++} \rightarrow \bar{\mathbb{R}}$. Here $s \in \mathbb{R}_{++}$ denotes a fixed barrier-penalty parameter.

We concentrate our investigation to those functions $\varphi (\cdot, s) : \mathbb{R} \rightarrow \bar{\mathbb{R}}$ which are differentiable at any $t \in \text{dom} \varphi (\cdot, s)$ and satisfy

$$
\frac{\partial}{\partial t} \varphi (t, s) = \psi^t(s), \quad \forall t \in \text{dom} \varphi (\cdot, s), \quad s > 0,
$$

(2.2)

with some $\psi : \mathbb{R} \rightarrow \bar{\mathbb{R}}$, $\psi \neq 0$. For function $\psi$ throughout the paper we make the following assumptions:

1. The domain of $\psi$ is either $\psi = (-\infty, 0)$ or $\psi = (-\infty, +\infty)$;

2. $\psi : \mathbb{R} \rightarrow \bar{\mathbb{R}}$ is continuous in the generalized sense of $\bar{\mathbb{R}}$.

3. $\psi$ is convex and continuously differentiable in $\text{dom} \psi$;
4. \( \psi'(r) \geq 0, \ \forall r \in \text{dom } \psi, \) and \( \lim_{r \to -\infty} \psi(r) = 0. \) (2.3)

5. \( \psi' \) is Lipschitz continuous on \( \text{dom } \psi \) in the sense that \( \psi' \) is locally Lipschitz continuous

\[
|\psi'(\rho_1) - \psi'(\rho_2)| \leq L_1(r) |\rho_1 - \rho_2|, \quad \forall \rho_1, \rho_2 \leq r, \ r \in \text{dom } \psi,
\]

and at \(-\infty\) satisfies

\[
|\psi'(\rho_1) - \psi'(\rho_2)| \leq L_2(r) \left| \frac{1}{\rho_1} - \frac{1}{\rho_2} \right|, \quad \forall \rho_1, \rho_2 \leq r < 0. \quad (2.5)
\]

6. \( \lim_{r \to -\infty} r^2 \psi'(r) \) exists (in the proper sense).

Before continuing, we derive some properties of functions \( \psi \) that will be used later on.

a) Function \( \psi \) maps \( \text{dom } \psi \) onto either \( R_+ \) or \( R_+ \).

In fact, by condition (2.3) function \( \psi \) is monotone increasing on \( \text{dom } \psi \). By property 1, if \( r \in \text{dom } \psi \), then \((-\infty, r] \subset \text{dom } \psi \), and

\[
\psi(r) \geq \lim_{\rho \to -\infty} \psi(\rho) = 0.
\]

(2.6)

Since \( \psi \) is convex, it is continuous in \( \text{dom } \psi \). In the case \( \text{dom } \psi = (-\infty, 0) \) the continuity in the generalized sense of \( IR \) (property 2) yields \( \lim_{r \to 0^-} \psi(r) = +\infty \). Let us consider the case \( \text{dom } \psi = (-\infty, +\infty) \). Since \( \psi \not\equiv 0 \), \( \psi' \geq 0 \) and \( \lim_{r \to -\infty} \psi(r) = 0 \) (property 4) some \( \hat{r} \in \text{dom } \psi \) exists with \( \psi'(\hat{r}) > 0 \). Due to the convexity of \( \psi \) we have

\[
\psi(r) \geq \psi(\hat{r}) + \psi'(\hat{r})(r - \hat{r}), \quad \forall r \in IR,
\]

(2.7)

and therefore, \( \lim_{r \to -\infty} \psi(r) = +\infty \). Thus, in both cases \( IR_+ \) is contained in the range of \( \psi \).

b) \( \psi' \circ \psi^{-1} \) maps \( R_+ \) into \( R_+ \) and is monotone increasing.

Indeed, let \( t > 0 \) be in the range of \( \psi \). Then there exists \( r \in \text{dom } \psi \) such that \( \psi(r) = t \). As before \((-\infty, r] \subset \text{dom } \psi \). If \( \psi'(r) = 0 \) then the convexity of \( \psi \) yields

\[
0 \leq \psi'(\rho) \leq \psi'(r), \quad \forall \rho \in (-\infty, r].
\]

Thus, we have \( \psi'(\rho) = 0, \ \forall \rho \leq r \). Consequently,

\[
t = \psi(r) = \lim_{\rho \to -\infty} \left( \psi(\rho) + \int_{\rho}^{r} \psi'(\tau) \, d\tau \right) = 0,
\]

that is a contradiction. Hence \( \psi'(r) \) is positive, and therefore, \( \psi(r) \) is strictly increasing. So \( \psi^{-1} \) exists and is strictly increasing on \( R_+ \), and \( (\psi' \circ \psi^{-1})(t) \in R_+ \).
Let \( c := \lim_{r \to -\infty} r^2 \psi'(r) \). Thus, we have
\[
\psi'(r) = r^{-2}c + o(r^{-2}), \quad r \to -\infty.
\]
This yields
\[
\psi(r) = \int_{-\infty}^{r} \psi'\left(\frac{\rho}{s}\right) d\rho = -r^{-1}c + o(r^{-1}), \quad r \to -\infty.
\]
Hence, \( \lim_{r \to -\infty} r \psi(r) = -c \).

\[d\] \( B_s \) is nonempty and does not depend on \( s \), and therefore, is denoted throughout as \( B \).

In fact, \( \varphi(g_i(x), s) < +\infty \) if and only if \( \psi\left(\frac{g_i(x)}{s}\right) < +\infty \). This is equivalent to \( \psi(g_i(x)) < +\infty \), since \( \frac{g_i(x)}{s} \in \text{dom} \psi \), \( s \in R_{+} \), if and only if \( g_i(x) \in \text{dom} \psi \). Hence, \( B_s \) does not depend on \( s \).

Let \( x \in G^0 \), then \( g(x) < 0 \). Consequently, \( g_i(x) < 0 \), that means \( g_i(x) \in \text{dom} \psi \) for all \( i = 1, \ldots, m \). Therefore, \( G^0 \subset B \), and since \( G^0 \) is nonempty, so is \( B \).

It should be noticed that function \( \psi \) defines \( \varphi \) up to some constant w.r.t. the variable \( t \). Hence it would be sufficient to consider functions \( \psi \) only. Following the classical barrier-penalty concept (cf. [8], [17], [12]), here first we had described the barrier-penalty methods using functions \( \varphi \). However, in further investigation only exclusively properties of functions \( \psi \) are required. In particular, the barrier-penalty function \( \varphi \) itself could be defined by

\[
\varphi(t, s) := \int_{t}^{r} \psi\left(\frac{\rho}{s}\right) d\rho, \quad \text{with some } r < 0, \text{ fixed.} \quad (2.8)
\]

The assumptions upon \( \psi \) and (2.8) imply the barrier-penalty property

\[
\lim_{s \to 0+} \varphi(t, s) = \begin{cases} 
0, \text{ if } t < 0, \\
+\infty, \text{ if } t > 0,
\end{cases} \quad (2.9)
\]

independently of \( r < 0 \) chosen in (2.8). Indeed, if \( t < 0 \) we have
\[
\lim_{s \to 0+} \varphi(t, s) = \lim_{s \to 0+} \int_{t}^{r} \psi\left(\frac{\rho}{s}\right) d\rho = \int_{t}^{r} \lim_{s \to 0+} \psi\left(\frac{\rho}{s}\right) d\rho = \int_{t}^{r} \lim_{\xi \to -\infty} \psi(\xi) d\rho = 0.
\]

In case \( t > 0 \) conditions \( \psi \geq 0 \) and \( r < 0 \) yield
\[
\lim_{s \to 0+} \varphi(t, s) = \lim_{s \to 0+} \int_{t}^{r} \psi\left(\frac{\rho}{s}\right) d\rho \geq \int_{t}^{r} \lim_{s \to 0+} \psi\left(\frac{\rho}{s}\right) d\rho = \int_{t}^{r} \lim_{\xi \to +\infty} \psi(\xi) d\rho = +\infty.
\]

The strong convexity of the objective function \( f \), the continuity and convexity of the barrier-penalty functions \( \varphi(\cdot, s) : \mathbb{R} \to \mathbb{R} \) for any \( s > 0 \) imply that the auxiliary problem
(2.1) possesses a unique solution $x(s) \in B$ for any $s > 0$. This solution is defined by the necessary first order optimality criterion

$$x(s) \in B, \quad \nabla f(x(s)) + \sum_{i=1}^{m} \psi(g_i(x(s))/s) \nabla g_i(x(s)) = 0. \quad (2.10)$$

Under the made assumptions this condition is also sufficient for the optimality of $x(s) \in B$ of the auxiliary problem (2.1). Let $y(s) \in \mathbb{R}_m^m$ be defined by

$$y_i(s) := \psi(g_i(x(s))/s), \quad \forall s > 0, \quad i = 1, \ldots, m. \quad (2.11)$$

Property (2.9) and $G^0 \neq \emptyset$ imply (cf. [8], [17], [6], [7]) that the solutions $x(s) \in B$ of the auxiliary problems (2.1) approximate the wanted solution $x^* \in G$ of (1.1) in the limit

$$\lim_{s \to 0^+} x(s) = x^*. \quad (2.12)$$

Hence, $x(\cdot) : \mathbb{R}_{++} \to B$ defines a path leading to $x^*$. The properties of this path can be further specified if second order sufficiency and strict complementarity conditions hold at the solution $x^*$ of the original problem (1.1). Before investigating these properties we specify some types of barrier-penalty functions $\varphi$ and related functions $\psi$ which satisfy all the conditions listed above.

- **logarithmic barrier function**

$$\varphi(t, s) := \begin{cases} -s \ln(-t), & \text{if } t < 0, \\ +\infty, & \text{if } t \geq 0, \end{cases} \quad \forall s > 0, \quad \psi(r) = \begin{cases} |r|^{-1}, & \text{if } r < 0, \\ +\infty, & \text{if } r \geq 0. \end{cases} \quad (2.13)$$

- **$p$-th power barrier function** ($p > 0$ is a fixed parameter)

$$\varphi(t, s) := \begin{cases} \frac{1}{p} \frac{s^{p+1}}{|t|^p}, & \text{if } t < 0, \\ +\infty, & \text{if } t \geq 0, \end{cases} \quad \forall s > 0, \quad \psi(r) = \begin{cases} |r|^{-(p+1)}, & \text{if } r < 0, \\ +\infty, & \text{if } r \geq 0. \end{cases} \quad (2.14)$$

- **$p$-th power loss function** ($p \geq 2$ is a fixed parameter)

$$\varphi(t, s) := \frac{1}{p} s^{-(p-1)} \max\{0, t\}, \quad \forall s > 0, \quad \psi(r) = \max^{p-1}\{0, r\}. \quad (2.15)$$

- **exponential penalty function**

$$\varphi(t, s) := s \exp(t/s), \quad \forall s > 0, \quad \psi(r) = \exp(r). \quad (2.16)$$
It can be easily seen that the given $\psi$ are generating functions for the related barrier-penalty functions $\varphi$ according to (2.2) and that $\psi$ are convex and (except for $p = 2$ in the case of the loss function) continuously differentiable in their domain $\text{dom} \psi$ and satisfy the assumed conditions. In the case $p = 1$ the barrier method related to the $p$-th power barrier function is just the well known SUMT method proposed by Fiacco/McCormick (cf. [8]).

Next, we show that the barrier-penalty methods under consideration generate near $s = 0$ differentiable primal and dual trajectories $x(\cdot)$ and $y(\cdot)$, respectively. For the classical barrier and penalty methods, as logarithmic barrier, SUMT or loss functions, these properties have been derived in [8], [17]. In the present paper, however, we base our investigation exclusively on properties of the generating functions $\psi$. While the first part of the proof of Theorem 1 which deals with the convergence of the barrier-penalty method is quite standard, the second part cannot directly follow the approach given in [8], [17]. To make it self-contained the complete proof will be given.

**Theorem 1** If the vectors $\nabla g_i(x^*)$, $i \in I_0$, are linearly independent and the optimal Lagrange multipliers $y^*_i \geq 0$, $i = 1, \ldots, p$, satisfy the strict complementarity condition

$$y^*_i > 0 \iff i \in I_0,$$

then the path $x(s) \in B$ defined by (2.10) for $s > 0$ and $x(0) := x^*$ and the related mapping $y(\cdot) : \mathbb{R}_+ \rightarrow \mathbb{R}^m_+$ defined by (2.11) and $y(0) := y^*$, where $y^* \in \mathbb{R}^m_+$ denotes the optimal Lagrange vector $y^* \in \mathbb{R}^m_+$, are differentiable for sufficiently small $s > 0$ and possess a right side derivative at $s = 0$.

**Proof:** First, we show that for any $s > 0$ the auxiliary problem (2.1) possesses a unique solution. Let $F_s : \mathbb{R}^n \rightarrow \mathbb{R}$ denote the objective function in barrier-penalty problem (2.1) for fixed parameters $s > 0$, i.e.

$$F_s(x) := f(x) + \sum_{i=1}^{m} \varphi(g_i(x), s).$$

Since $\varphi(\cdot, s)$ is monotone increasing and convex and $g_i$, $i = 1, \ldots, m$, are convex, the composite functions $\varphi(g_i(\cdot), s)$, $i = 1, \ldots, m$, are convex for any $s > 0$. Hence, the strong convexity of $f$ implies that some $c > 0$ exists with

$$F_s(x) \geq F_s(z) + \nabla F_s(z)^T(x - z) + c\|x - z\|^2, \quad \forall x, z \in B. \quad (2.18)$$

As shown in property d) we have $G^0 \subset B$. Furthermore, the convexity of $g_i$, $i = 1, \ldots, m$, and $G^0 \neq \emptyset$ yield $G = \overline{G^0}$.

Let $z \in G^0$ be some fixed, but arbitrary element. Then from (2.18) we obtain

$$F_s(x) \leq F_s(z) \quad \Rightarrow \quad \|x - z\| \leq c^{-1} \|
abla F_s(z)\|, \quad (2.19)$$
where $c > 0$ is independent of $s > 0$. Thus, for any fixed $z \in G^0$ the level sets $Q(z) := \{ x \in \mathbb{R}^n : F_s(x) \leq F_s(z) \}$ are uniformly bounded. Keeping in mind the lower semi-continuity and the strong convexity of $F_s(\cdot)$ we see that $Q(z) \subset \mathbb{R}^n$ is closed and bounded, and the auxiliary problem (2.1) possesses a unique solution $x(s)$ for any $s > 0$.

Next, we prove that the limits

$$
\bar{x} = \lim_{s \to 0+0} x(s), \quad \text{and} \quad \bar{y} = \lim_{s \to 0+0} \psi \left( \frac{g_i(x(s))}{s} \right), \quad i = 1, \ldots, m,
$$

exist and satisfy the KKT-conditions. Let $z^0 \in G^0$ be fixed. Since $x(s)$ is the minimizer of $F_s$ we have $x(s) \in Q(z^0)$. Estimate (2.19) yields $\|x(s) - z^0\| \leq c$ with some $c > 0$. Let $\{s_k\}_{k=1}^\infty$ be any sequence of positive numbers converging to 0. Because the related sequence $\{x(s_k)\}$ is bounded, without loss of generality we can assume that it converges to some $\bar{x} \in G$. For any $x \in G^0$ we have

$$
f(x) + \sum_{i=1}^m \varphi (g_i(x), s_k) = F_{s_k}(x) \\
\geq F_{s_k}(x(s_k)) = f(x(s_k)) + \sum_{i=1}^m \varphi (g_i(x(s_k)), s_k), \quad k = 1, 2, \ldots.
$$

(2.20)

Since $x \in G^0$ and $\lim_{k \to \infty} s_k = 0$, property (2.9) yields

$$
\lim_{k \to \infty} \varphi (g_i(x), s_k) = 0, \quad i = 1, \ldots, m.
$$

(2.21)

Furthermore the convergence of $\{x(s_k)\}$, the continuity of the functions $g_i$, $i = 1, \ldots, m$, the monotonicity of $\varphi (\cdot, s)$ and (2.9) result in

$$
\lim_{k \to \infty} \varphi (g_i(x(s_k)), s_k) \geq 0, \quad i = 1, \ldots, m.
$$

(2.22)

Combining (2.20) - (2.22), and recalling the continuity of $f$ and $G^0 = G$ we obtain

$$
f(x) \geq f(\bar{x}), \quad \forall x \in G.
$$

(2.23)

Since the sequence $\{s_k\}$ was an arbitrary one with a related convergent sequence $\{x(s_k)\} \subset Q(z^0)$ of minimizers of $F_{s_k}$ and any minimizer of $F_{s_k}$ belongs to the bounded set $Q(z^0)$, with (2.23) we have

$$
\lim_{s \to 0+0} x(s) = x^*.
$$

(2.24)

Because $B \subset \mathbb{R}^n$ is an open set the solution $x(s)$ is characterized by

$$
\nabla f(x(s)) + \sum_{i=1}^m \psi \left( \frac{g_i(x(s))}{s} \right) \nabla g_i(x(s)) = 0.
$$

(2.25)
Due to property (2.9) and the continuity of the functions \(g_i, i = 1, \ldots, m\), it holds
\[
\lim_{s \to 0^+} \psi\left(\frac{g_i(x(s))}{s}\right) = 0, \quad i \neq I_0.
\]  
(2.26)

With \(\nabla g_i(x^*)^T(z^0 - x^*) \leq g_i(z^0) < 0\), for \(i \in I_0\), formulas (2.24) - (2.26) result in the boundedness of functions \(y_i(s) := \psi\left(\frac{g_i(x(s))}{s}\right), i \in I_0\), as \(s \to 0^+\). Similar to the discussion above for \(x(s)\), for any sequence \(\{s_k\} \subset \mathbb{R}_{++}\) with \(\lim_{k \to \infty} s_k = 0\) and related convergent sequences \(\{y_i(s_k)\}_{k=0}^{\infty}, i \in I_0\), we have
\[
\nabla f(x^*) + \sum_{i \in I_0} \bar{y}_i \nabla g_i(x^*) = 0.
\]  
(2.27)

Since the gradients \(\nabla g_i(x^*), i \in I_0\), are linearly independent, the limits \(\bar{y}_i, i \in I_0\), are unique. Thus, with the boundedness of \(y_i(s)\) for \(s \to 0^+\) there exist the limits
\[
y_i^* = \lim_{s \to 0^+} y_i(s), \quad i \in I_0.
\]

As already mentioned, we have also \(\lim_{s \to 0^+} y_i(s) = 0, i \not\in I_0\). Taking into account \(y_i(s) \geq 0, i = 1, \ldots, m\), and formula (2.27) we see that \((x^*, y^*)\) satisfy the KKT-conditions for the original problem (1.1).

Next, we show that some \(s_0 > 0\) exists such that \(x(\cdot), y(\cdot)\) are continuously differentiable in \((0, s_0)\) and differentiable from the right at \(s = 0\). Let us consider a perturbed KKT-system
\[
\nabla f(x) + \sum_{i \in I_0} y_i \nabla g_i(x) = r,
\]
with \(s \psi^{-1}(y_i) = g_i(x), i \in I_0\),
\[
(2.28)
\]
and
\[
\nabla f(x^*) + \sum_{i \in I_0} y_i^* \nabla g_i(x^*) = 0,
\]
\[
0 \cdot \psi^{-1}(y_i^*) = g_i(x^*), \quad i \in I_0,
\]
(2.29)
i.e. the optimal solution \(x^*\) and its related dual multipliers \(y_i^*, i \in I_0\), satisfy (2.28) for \(s = 0, r = 0\). Without loss of generality, the constraints can be numbered such that \(I_0 = \{1, \ldots, m_0\}\) with some \(m_0 \leq m\). Since the Hessian \(\nabla^2 f(x^*)\) is positive definite and the gradients \(\nabla g_i(x^*), i \in I_0\), are linearly independent, the \((n + m_0, n + m_0)\)-matrix
\[
\begin{pmatrix}
\nabla^2 f(x^*) + \sum_{i \in I_0} y_i^* \nabla^2 g_i(x^*) & \nabla g_1(x^*) & \nabla g_2(x^*) & \cdots & \nabla g_{m_0}(x^*) \\
\n\nabla g_1(x^*)^T & 0 & 0 & \cdots & 0 \\
0 & 0 & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots & \vdots \\
\n\nabla g_{m_0}(x^*)^T & 0 & 0 & \cdots & 0 \\
\end{pmatrix}
\]

9
is regular (cf. [8], [17]). Now, the implicit function theorem defines differentiable functions $\hat{x}(s, r)$ and $\hat{y}_i(s, r)$, $i \in I_0$, locally near $s = 0$, $r = 0$ by

$$\nabla f(\hat{x}(s, r)) + \sum_{i \in I_0} \hat{y}_i(s, r) \nabla g_i(\hat{x}(s, r)) = r,$$

$$s \psi^{-1}(\hat{y}_i(s, r)) = g_i(\hat{x}(s, r)), \quad i \in I_0. \tag{2.30}$$

For $s > 0$ we define $y_i(s) := \psi\left(\frac{g_i(x(s))}{s}\right)$, $i \in \{1, \ldots, m\}$ where $x(s)$ denotes the solution of the barrier-penalty problem (2.1). The necessary and due to the convexity also sufficient optimality conditions are

$$\nabla f(x(s)) + \sum_{i \in I_0} \psi\left(\frac{g_i(x(s))}{s}\right) \nabla g_i(x(s)) = -\sum_{i \notin I_0} y_i(s) \nabla g_i(x(s)), \quad s > 0. \tag{2.31}$$

Let $r := r(s) = -\sum_{i \notin I_0} y_i(s) \nabla g_i(x(s))$. From (2.30) and (2.31) it follows that $x(s) = \hat{x}(s, r(s))$, $y_i(s) = \hat{y}_i(s, r(s))$. In fact,

$$\lim_{s \to 0^+} \frac{r(s) - r(0)}{s - 0} = \lim_{s \to 0^+} \sum_{i \notin I_0} \psi\left(\frac{g_i(x(s))}{s}\right) \nabla g_i(x(s))$$

$$= \sum_{i \notin I_0} \nabla g_i(x^*) \lim_{s \to 0^+} \frac{g_i(x(s))}{s} \psi\left(\frac{g_i(x(s))}{s}\right) \frac{1}{g_i(x(s))}$$

$$= \sum_{i \notin I_0} \nabla g_i(x^*) \lim_{\rho \to -\infty} \rho \psi(\rho) = -c \sum_{i \notin I_0} \frac{\nabla g_i(x^*)}{g_i(x^*)}.$$

Thus, $r$ is differentiable from the right at $s = 0$, and the implicit function theorem guarantees the differentiability of functions $\hat{x}(s, r(s)) = x(s)$, $\hat{y}_i(s, r(s)) = y_i(s)$, $i \in I_0$, at $s = 0$ from the right. As already shown, also $y_i(\cdot)$, $i \notin I_0$, are differentiable from the right at $s = 0$, namely $D_+y_i(0) = 0$, $i \notin I_0$.

Now, we consider the case $s > 0$. Then $x(s)$, $y(s)$ satisfy the system

$$\nabla f(x(s)) + \sum_{i=1}^m y_i(s) \nabla g_i(x(s)) = 0,$$

$$\psi\left(\frac{g_i(x(s))}{s}\right) - y_i(s) = 0, \quad i = 1, \ldots, m. \tag{2.32}$$

Differentiation of system (2.32) w.r.t. the parameter $s$ yields

$$H(s) \begin{pmatrix} \dot{x}(s) \\ \dot{y}(s) \end{pmatrix} = \begin{pmatrix} 0 \\ q(s) \end{pmatrix}, \tag{2.33}$$
where $H(s)$ denotes the $(n + m, n + m)$-matrix

$$
H(s) := \begin{pmatrix}
\nabla^2 f(x(s)) + \sum_{i \in I_0} y_i(s) \nabla^2 g_i(x(s)) & \nabla g_1(x(s)) & \nabla g_2(x(s)) & \cdots & \nabla g_m(x(s)) \\
\psi' \left( \frac{g_1(x(s))}{s} \right) \nabla g_1(x(s))^T & -s & 0 & \cdots & 0 \\
\psi' \left( \frac{g_2(x(s))}{s} \right) \nabla g_2(x(s))^T & 0 & -s & \cdots & 0 \\
\ddots & \ddots & \ddots & \ddots & \ddots \\
\psi' \left( \frac{g_m(x(s))}{s} \right) \nabla g_m(x(s))^T & 0 & 0 & \cdots & -s
\end{pmatrix},
$$

$q(s) := (q_1(s), \ldots, q_m(s))^T$, with $q_i(s) := s^{-1} \psi' \left( \frac{g_i(x(s))}{s} \right) g_i(x(s))$, and $\dot{x}(s), \dot{y}(s)$ denote the derivatives of $x(s)$ and $y(s)$, respectively. These derivatives exist due to the implicit function theorem, since the matrix $H(s)$ is regular. Next, we estimate the derivatives $\dot{x}(s), \dot{y}(s)$ by using the structure and properties of $H(s)$. With the abbreviations

- $R := \nabla^2 f(x(s)) + \sum_{i=1}^m y_i(s) \nabla^2 g_i(x(s))$,
- $T_1 := \{\nabla g_1(x(s), \ldots, \nabla g_{m_0}(x(s))\}$, $T_2 := \{\nabla g_{m_0+1}(x(s), \ldots, \nabla g_m(x(s))\}$,
- $D_1 := \text{diag} \left\{ \psi' \left( \frac{g_1(x(s))}{s} \right), \ldots, \psi' \left( g_{m_0}(x(s)) \right) \right\}$,
- $D_2 := \text{diag} \left\{ \psi' \left( \frac{g_{m_0+1}(x(s))}{s} \right), \ldots, \psi' \left( \frac{g_m(x(s))}{s} \right) \right\}$,

system (2.33) has the form

$$
R \dot{x} + T_1 \dot{y}^1 + T_2 \dot{y}^2 = 0, \\
D_1 T_1^T \dot{x} - s \dot{y}^1 = q^1, \\
D_2 T_2^T \dot{x} - s \dot{y}^2 = q^2. 
$$

(2.34)

Here $y^1, y^2$ and $q^1, q^2$ denote the splittings of $y(s)$ and $q(s)$ respectively, adjusted to the submatrices of $H(s)$. Further, the parameter $s$ has been omitted to simplify the notations. Eliminating $\dot{y}^2$ from the last equation in (2.34)

$$
\dot{y}^2 = s^{-1} (D_2 T_2^T \dot{x} - q^2)
$$

(2.35)

leads system (2.34) to the reduced form

$$
(R + s^{-1} T_2 D_2 T_2^T) \dot{x} + T_1 \dot{y}^1 = p, \\
D_1 T_1^T \dot{x} - s \dot{y}^1 = q^1.
$$

(2.36)
where
\[ p = p(s) := s^{-1}T_2q^2 = \sum_{i \not\in I_0} s^{-2} \psi' \left( \frac{g_i(x(s))}{s} \right) g_i(x(s)) \nabla g_i(x(s)). \] (2.37)

Now, the condition \( \psi'(\rho) = O(|\rho|^{-2}) \) for \( \rho \to -\infty \), the convergence \( x(s) \to x^* \) and the continuous differentiability of the constraint functions \( g_i \) guarantee that a constant \( c_0 > 0 \) and some \( s_0 > 0 \) exist with
\[
\|s^{-1}q^2\| \leq c_0, \quad \|p\| \leq c_0, \quad \forall s \in (0, s_0]. \] (2.38)

Since \( R \) is uniformly positive definite, from (2.36) we obtain
\[
\dot{x} = (R + s^{-1}T_2D_2T_2^T)^{-1}(p - T_1\dot{y}^1). \] (2.39)

Because \( s^{-1}T_2D_2T_2^T \) is semi-positive definite, some \( c_1 > 0 \) exists with \( \| (R + s^{-1}T_2D_2T_2^T)^{-1} \| \leq c_1 \). Thus,
\[
\|\dot{x}\| \leq \|(R + s^{-1}T_2D_2T_2^T)^{-1}\| (\|p\| + \|T_1\| \|\dot{y}^1\|) \leq c_1 (\|p\| + \|T_1\| \|\dot{y}^1\|). \] (2.40)

Now, (2.36), (2.40) yield
\[
(T_1^T (R + s^{-1}T_2D_2T_2^T)^{-1}T_1 + s D_1^{-1}) \dot{y}^1 = T_1^T (R + s^{-1}T_2D_2T_2^T)^{-1}p - D_1^{-1}q^1. \] (2.41)

The linear independence of the vectors \( \nabla g_i(x^*), i \in I_0 \), the continuous differentiability of the functions \( g_i \), the uniformly positive definiteness of the matrices \( R + s^{-1}T_2D_2T_2^T \), and \( x(s) \to x^* \) for \( s \to 0 + 0 \) provide some \( c_2 > 0 \) and \( s_1 \in (0, s_0] \) such that
\[
\|(T_1^T (R + s^{-1}T_2D_2T_2^T)^{-1}T_1 + s D_1^{-1})^{-1}\| \leq c_3, \quad \forall s \in (0, s_1]. \] (2.42)

With \( y_i = \psi \left( \frac{g_i(x(s))}{s} \right), i = 1, \ldots, m \), and \( y(s) \to y^* \) for \( s \to 0 + 0 \), and with the strict complementarity \( y_i^* > 0, i \in I_0 \) we have \( D_1 = \text{diag} \{ \psi'(\psi^{-1}(y_1(s))), \ldots, \psi'(\psi^{-1}(y_m(s))) \} \).

Hence, a constant \( c_4 > 0 \) and some \( s_2 \in (0, s_1] \) exist so that
\[
\|D_1^{-1}\| \leq c_3, \quad \forall s \in (0, s_2]. \]

Using (2.41) and the above estimates, we obtain
\[
\|\dot{y}^1\| \leq c_2 (c_1 T_1^T \|p\| + c_3 \|q^1\|). \]

The boundedness of \( p \) and \( q^1 \) implies that \( \dot{y}^1 \) is bounded. Finally the boundedness of \( \dot{x} \) that follows from (2.40) and (2.38) yields the boundedness of \( \dot{y}^2 \).

**Corollary 1** If the vectors \( a_i, i \in I_0 \), are linearly independent and the optimal Lagrange multipliers \( y_i^* \geq 0, i = 1, \ldots, m \), satisfy the strict complementarity condition (2.17) then there exist some constants \( s_0 > 0 \) and \( c_L > 0 \) such that
\[
\begin{align*}
\|x(s) - x(t)\| &\leq c_L |s - t|, \\
\|y(s) - y(t)\| &\leq c_L |s - t|,
\end{align*}
\] \forall s, t \in [0, s_0]. \] (2.43)
3 Newton’s method applied to barrier problems

Here and in the sequel we restrict our investigations to linearly constrained problems, i.e. we consider optimization problems

\[ f(x) \rightarrow \min \quad \text{subject to} \quad x \in G := \{ x \in \mathbb{R}^n : a_i^T x \leq b_i, \ i = 1, \ldots, m \} \]

with vectors \( a_i \in \mathbb{R}^n \) and constants \( b_i \in \mathbb{R}, i = 1, \ldots, m \). Hence, further the constraint functions have the specific form \( g_i(x) = a_i^T x - b_i, \ i = 1, \ldots, m \). For any fixed parameter \( s > 0 \) the related barrier-penalty problem (2.1) for (3.1) is (locally) equivalent to the system of nonlinear equations

\[ x(s) \in B, \quad \nabla f(x(s)) + \sum_{i=1}^{m} \psi((a_i^T x(s) - b_i)/s) a_i = 0, \quad (3.2) \]

which has a unique solution \( x(s) \in B \). Next, we investigate the convergence behavior of Newton’s method applied to (3.2). Let \( x \in B \) denote some approximation of the solution \( x(s) \) of (2.1). Related to \( x \in B \) we define vectors \( u(x, s), v(x, s) \in \mathbb{R}_+^m \) by

\[ \begin{aligned}
  u_i(x, s) & := \psi((a_i^T x - b_i)/s), \\
  v_i(x, s) & := s^{-1} \psi'((a_i^T x - b_i)/s), \\
\end{aligned} \quad \forall \ s > 0, \ i = 1, \ldots, m, \quad (3.3) \]

respectively. In particular, (2.11), (3.3) yield \( y(s) = u(x(s), s), \forall s > 0 \).

Starting from \( x \in B \), one step of Newton’s method defines a new approximate \( \tilde{x} \) of \( x(s) \) as solution of the linear system

\[ \nabla f(x) + \sum_{i=1}^{m} u_i(x, s) a_i + \left( \nabla^2 f(x) + \sum_{i=1}^{m} v_i(x, s) a_i a_i^T \right) (\tilde{x} - x) = 0. \quad (3.4) \]

The strong convexity of the function \( f \) guarantees that for any \( x \in B \) the system matrix of (3.4) is positive definite. Hence, the linear system (3.4) has a unique solution \( \tilde{x} \in \mathbb{R}^n \). It remains to provide bounds for \( \|x - x(s)\| \) which guarantee that \( \tilde{x} \in B \) and to establish an estimate of \( \|\tilde{x} - x(s)\| \) for the new iterate \( \tilde{x} \). For fixed barrier-penalty parameter Newton’s method is locally quadratically convergent. Hence, some functions \( \delta(\cdot), \rho(\cdot) : \mathbb{R}_+ \rightarrow \mathbb{R}_+ \) exist such that

\[ \|\tilde{x} - x(s)\| \leq \delta(s) \|x - x(s)\|^2, \quad \forall x \in B, \quad \|x - x(s)\| \leq \rho(s). \quad (3.5) \]

Before we determine these functions, we provide some estimate for the solutions of a certain family of linear systems.
Lemma 1 Let \( \{A_i\}_{i=1}^q \) be given fixed, symmetric and positive semi-definite \((n,n)\)-matrices. Then for any symmetric, positive definite \((n,n)\)-matrix \(C\), for any \(u \in \mathbb{R}^n\), for any \(\alpha_i > 0, i = 1, \ldots, q\), and for any \(\beta_i \in \mathbb{R}, i = 1, \ldots, q\), the linear system
\[
(C + \sum_{i=1}^q \alpha_i A_i) x = \left(\sum_{i=1}^q \beta_i A_i\right) u
\]
possesses a unique solution \(x \in \mathbb{R}^n\). Moreover, \(x\) can be estimated by
\[
\|x\|_C \leq \left(1 + 2\sigma \frac{\sum_{i=1}^q |\beta_i - \alpha_i|}{\alpha_i} \right) \|u\|_C.
\]
Here \(\|\cdot\|_C\) denotes the energy norm induced by \(C\) and \(\sigma > 0, \gamma, \bar{\gamma} > 0\) are defined by
\[
\sigma := \max_{1 \leq i \leq q} \alpha_i / \min_{1 \leq i \leq q} \alpha_i, \quad \gamma := \min_{A^1 x \neq 0} \frac{x^T A^1 x}{x^T C x}, \quad \bar{\gamma} := \max_{1 \leq i \leq q} \|A_i\|_C,
\]
with \(A^1 := \sum_{1 \leq i \leq q} A_i\).

Proof: Let denote \(A^\alpha := \sum_{i=1}^q \alpha_i A_i\) and \(A^\beta\) similarly. Since \(C\) is positive definite and \(A^\alpha\) is positive semi-definite, for any \(u \in \mathbb{R}^n\) system (3.6) possesses a unique solution \(x \in \mathbb{R}^n\).

To obtain an estimate for \(x\), we split the right hand side into two parts according to
\[
(C + A^\alpha) x = A^\alpha u + (A^\beta - A^\alpha) u. \tag{3.8}
\]
To treat the first part, a generalized eigenfunction expansion w.r.t. \(C\) is used. Since \(A^\alpha\) is symmetric, positive semi-definite and \(C\) is symmetric, positive definite, eigenvalues \(\lambda_j \geq 0\) and related eigenvectors \(z_j \in \mathbb{R}^n, j = 1, \ldots, n\), exist with
\[
A^\alpha z_j = \lambda_j C z_j, \quad j = 1, \ldots, n, \quad \text{and} \quad (z_i, z_j)_C = \delta_{ij}, \quad i, j = 1, \ldots, n, \tag{3.9}
\]
where \((x, y)_C := x^T C y, \forall x, y \in \mathbb{R}^n\). Let us expand \(u\) as well as the solution \(x^1 \in \mathbb{R}^n\) of
\[
(C + A^\alpha) x^1 = A^\alpha u \tag{3.10}
\]
over the basis \(\{z_j\}_{j=1}^n\), i.e. \(u = \sum_{j=1}^n \mu_j z_j\) and \(x^1 = \sum_{j=1}^n \xi_j z_j\). Now, (3.9), (3.10) yield
\[
\sum_{j=1}^n (1 + \lambda_j) \xi_j C z_j = \sum_{j=1}^n \lambda_j \mu_j C z_j.
\]
Using the \(C\)-orthogonality of the eigenvectors \(z_j, j = 1, \ldots, n\), we obtain
\[
(1 + \lambda_j) \xi_j = \lambda_j \mu_j, \quad j = 1, \ldots, n.
\]
Due to $\lambda_j \geq 0$, $j = 1, \ldots, n$ this results in $|\xi_j| \leq |\mu_j|$, $j = 1, \ldots, n$. Thus, we have
\[
\|x^1\|^2_C = (x^1, x^1)_C = \sum_{j=1}^n \xi_j^2 \leq \sum_{j=1}^n \mu_j^2 = (u, u)_C = \|u\|^2_C. \tag{3.11}
\]
Next, we estimate the solution $x^2 \in \mathbb{R}^n$ of the linear system
\[
(C + A^\alpha)x^2 = (A^\beta - A^\alpha)u. \tag{3.12}
\]
We notice that the nullspaces $\mathcal{N}(A^\alpha)$ and $\mathcal{N}(A^\beta)$ can be described by
\[
\mathcal{N}(A^\alpha) = \bigcap_{j=1}^q \mathcal{N}(A_j) \subset \mathcal{N}(A^\beta).
\]
Let us represent $x^2$ in the form
\[
x^2 = x_N + x_R \quad \text{with} \quad x_N \in \mathcal{N}(A^\alpha), \quad x_R \in \mathcal{N}(A^\alpha)^\perp,
\]
where $\mathcal{N}(A^\alpha)^\perp := \{x \in \mathbb{R}^n : (x, v)_C = 0, \forall v \in \mathcal{N}(A^\alpha)\}$ denotes the $C$-orthogonal complement of $\mathcal{N}(A^\alpha)$. Multiplying (3.6) with $x_N^T$ results in
\[
x_N^T C x_N + x_N^T C x_R = 0.
\]
Now, Cauchy’s inequality for the scalar product $(\cdot, \cdot)_C$ yields the estimate
\[
\|x_N\|_C \leq \|x_R\|_C. \tag{3.13}
\]
It remains to find an upper bound for $\|x_R\|_C$. With the positive definiteness of $C$ the equation (3.12) results in
\[
x_R^T A^\alpha x_R = (x^2)^T A^\alpha x^2 \leq (x^2)^T (C + A^\alpha) x^2 = (x^2)^T (A^\beta - A^\alpha) u = x_R^T (A^\beta - A^\alpha) u.
\]
Hence, we have
\[
\min_{1 \leq i \leq q} \alpha_i x_R^T \left( \sum_{i=1}^q A_i \right) x_R \leq x_R^T A^\alpha x_R \leq \max_{1 \leq i \leq q} \alpha_i \max_{1 \leq i \leq q} \left| x_R^T A_i u \right| \sum_{i=1}^q \frac{\beta_i - \alpha_i}{\alpha_i},
\]
and consequently
\[
x_R^T \left( \sum_{i=1}^q A_i \right) x_R \leq \sigma \max_{1 \leq i \leq q} \left| x_R^T A_i u \right| \sum_{i=1}^q \frac{\beta_i - \alpha_i}{\alpha_i}. \tag{3.14}
\]
With $\gamma \|x_R\|^2_C \leq x_R^T \left( \sum_{i=1}^q A_i \right) x_R$ and $\max_{1 \leq i \leq q} \left| x_R^T A_i u \right| \leq \tau \|x_R\|_C \|u\|_C$ this results in
\[
\|x_R\|_C \leq \sigma \frac{\tau}{\gamma} \sum_{i=1}^q \frac{\beta_i - \alpha_i}{\alpha_i} \|u\|_C. \tag{3.15}
\]
Using (3.11), (3.13), now, we obtain
\[
\|x\|_C \leq \|x^1\|_C + \|x^2\|_C \leq \left( 1 + 2 \sigma \frac{\tau}{\gamma} \sum_{i=1}^q \frac{\beta_i - \alpha_i}{\alpha_i} \right) \|u\|_C.
\]
Theorem 2  For any \( s > 0 \) and \( x \in B \) the linear system (3.4) possesses a unique solution \( \hat{x} \in \mathbb{R}^n \) and some constants \( s_0, \delta, c_\rho > 0 \) exist such that
\[
\begin{align*}
    x \in \mathbb{R}^n, \quad s \in (0, s_0], \\
    \|x - x(s)\|_C \leq c_\rho s, \\
\end{align*}
\]
\[
\begin{align*}
    \|x - x(s)\|_C \leq c_\rho s, \quad \implies \quad x, \hat{x} \in B, \\
    \|\hat{x} - x(s)\|_C \leq c_\delta s^{-1} \|x - x(s)\|_C^2, \tag{3.16}
\end{align*}
\]
with
\[
C := C(x, s) := \nabla^2 f(x) + \sum_{i \in I_0} v_i(x, s) a_i a_i^T. \tag{3.17}
\]

Proof: Because of the strong convexity of function \( f \) the Hessian \( \nabla^2 f(x) \) is uniformly positive definite. The definition \( v_i(x, s) = s^{-1}(a_i^T x - b_i) \), and the monotonicity of \( \psi \) yield \( v_i(x, s) \geq 0 \), \( i = 1, \ldots, m \). With the positive semi-definiteness of the dyadic matrices \( A_i := a_i a_i^T \) this guarantees the system matrix in (3.4) to be positive definite. Hence \( \hat{x} \in \mathbb{R}^n \) is uniquely defined by (3.4) for any \( s > 0 \) and any \( x \in B \).

Next, we derive some preliminary results which are later used to estimate \( \|\hat{x} - x(s)\|_C \).

Theorem 1 guarantees the existence of constants \( s_0, \mu, \overline{\mu}, \delta > 0 \) such that
\[
\begin{align*}
    0 < \mu &\leq u_i(x(s), s) \leq \overline{\mu}, \quad i \in I_0, \\
    a_i^T x(s) - b_i &\leq -2\delta < 0, \quad i \not\in I_0, \forall s \in (0, s_0]. \tag{3.18}
\end{align*}
\]
Hence, some \( \rho_1 > 0 \) exists with
\[
a_i^T x - b_i \leq -\delta < 0, \quad i \not\in I_0, \quad \forall x \in \mathbb{R}^n \text{ with } \|x - x(s)\| \leq \rho_1, \quad \forall s \in (0, s_0]. \tag{3.19}
\]
Since \( \psi \) is monotone, \( \mathcal{R}_+ \) is contained in the range of \( \psi \) and \( \lim_{r \to -\infty} \psi(r) = 0 \) some constants \( \underline{r}, \overline{r} \in \text{dom } \psi, \underline{r} < \overline{r} \) exist with
\[
\underline{\mu} \leq \psi(r) \leq \overline{\mu} \quad \iff \quad \underline{r} \leq r \leq \overline{r}. \tag{3.20}
\]
Now, the continuity of \( \psi \) at \( r = \underline{r} \) and at \( r = \overline{r} \) guarantees that
\[
\frac{1}{2} \underline{\mu} \leq \psi(r) \leq 2 \overline{\mu} \quad \iff \quad \underline{r} - \theta \leq r \leq \overline{r} + \theta, \tag{3.21}
\]
with some \( \theta > 0 \). By construction we have \( u_i(x(s), s) = \psi(\frac{a_i^T x(s) - b_i}{s}), \quad i = 1, \ldots, m \).
Hence, due to (3.20) the first part of (3.18) is equivalent to
\[
\underline{r} \leq \frac{a_i^T x(s) - b_i}{s} \leq \overline{r}, \quad s \in (0, s_0], \quad i \in I_0.
\]

With \( \omega := \theta / \max_{i \in I_0} \|a_i\| \) from (3.21) we obtain
\[
\frac{1}{2} \underline{\mu} \leq \psi(\frac{a_i^T x - b_i}{s}) \leq 2 \overline{\mu}, \quad \forall x \in \mathbb{R}^n, \text{ with } \|x - x(s)\| \leq \omega s, \quad s \in (0, s_0], \quad i \in I_0. \tag{3.22}
\]
With (3.19) this yields $x \in \text{dom } \psi$, that means $x \in B$, if
\[
\|x - x(s)\| \leq \min\{\rho_1, \omega s\}, \ s \in (0, s_0].
\] (3.23)
Throughout the remaining part of the proof we assume that
\[
x \in \mathbb{R}^n, \ \|x - x(s)\| \leq \min\{\rho_1, \omega s\}, \ s \in (0, s_0].
\] (3.24)
By definition holds $v_i(x, s) = s^{-1} \psi' \left( \frac{a_i^T x - b_i}{s} \right)$, $i = 1, \ldots, m$. With property b) and (3.22) we obtain
\[
0 < s^{-1} \psi' \left( \psi^{-1}(\frac{\mu}{2}) \right) \leq v_i(x, s) \leq s^{-1} \psi' \left( \psi^{-1}(2\mu) \right), \ i \in I_0.
\] (3.25)
Hence, for $\alpha_i := v_i(x, s)$, $i = 1, \ldots, m$ this leads to the estimate
\[
\sigma := \frac{\max_{i \in I_0} \alpha_i}{\min_{i \in I_0} \alpha_i} \leq \frac{\psi' \left( \psi^{-1}(2\mu) \right)}{\psi' \left( \psi^{-1}(\frac{\mu}{2}) \right)},
\] (3.26)
which will be used later in the application of Lemma 1. Taking into account the monotonicity of $\psi'$ and (3.19) we have
\[
0 \leq s^{-1} \psi' \left( \frac{a_i^T x - b_i}{s} \right) \leq s^{-1} \psi' \left( \frac{-\delta}{s} \right), \ i \notin I_0.
\]
Using the Lipschitz continuity at $-\infty$ of $\psi'$ we have
\[
\psi' \left( \frac{-\delta}{s} \right) - \psi'(r) \leq L_2 \left( \frac{-\delta}{s_0} \right) \left| \frac{s}{\delta} + \frac{1}{r} \right|, \ r < -\frac{\delta}{s} \leq -\frac{\delta}{s_0}.
\]
Taking $r \to -\infty$ and using the fact that $\lim_{r \to -\infty} \psi'(r) = 0$, we arrive at
\[
s^{-1} \psi' \left( \frac{-\delta}{s} \right) \leq L_2 \left( \frac{-\delta}{s_0} \right) \delta^{-1}.
\]
Thus, constants $\gamma^* \geq \gamma_* > 0$ exist with
\[
\gamma_* \|z\|^2 \leq z^T C(x, s) z \leq \gamma^* \|z\|^2, \ \forall z \in \mathbb{R}^n.
\] (3.27)
Now, we study the linear system (3.4) arising in Newton’s method, i.e. the system
\[
\nabla f(x) + \sum_{i=1}^m u_i(x, s) a_i + \left( \nabla^2 f(x) + \sum_{i=1}^m v_i(x, s) a_i a_i^T \right) (\tilde{x} - x) = 0.
\] (3.28)
The necessary and sufficient first order optimality conditions (3.2), which define the solution $x(s) \in B$, can be written equivalently in the form
\[
\nabla f(x(s)) + \sum_{i=1}^m u_i(x(s), s) a_i + \left( \nabla^2 f(x) + \sum_{i=1}^m v_i(x(s), s) a_i a_i^T \right) (x(s) - x(s)) = 0.
\] (3.29)
Subtracting this system from (3.28) yields

\[
\begin{align*}
\left( \nabla^2 f(x) + \sum_{i=1}^m v_i(x, s) a_i a_i^T \right) (\dot{x} - x(s)) &= \nabla f(x(s)) - \nabla f(x) - \nabla^2 f(x) (x(s) - x) \\
&+ \sum_{i=1}^m \left( (u_i(x(s), s) - u_i(x, s)) a_i - v_i(x, s) a_i a_i^T (x(s) - x) \right).
\end{align*}
\] (3.30)

By Taylor’s formula we have

\[
\nabla f(x(s)) - \nabla f(x) - \nabla^2 f(x) (x(s) - x)
= \int_0^1 \left( \nabla^2 f(x + \tau(x(s) - x)) - \nabla^2 f(x) \right) d\tau \ (x(s) - x).
\] (3.31)

Taking into account \( \nabla_x u_i(x, s) = v_i(x, s) a_i \) similarly holds

\[
(u_i(x(s), s) - u_i(x, s)) a_i - v_i(x, s) a_i a_i^T (x(s) - x)
= \int_0^1 (v_i(x + \tau(x(s) - x), s) - v_i(x, s)) d\tau \ a_i a_i^T (x(s) - x), \quad i = 1, \ldots, m.
\] (3.32)

Now, we split the right hand side of (3.30) into three components \( r_j \in \mathbb{R}^n, \ j = 1, 2, 3, \) according to

\[
\begin{align*}
\text{r}_1 &:= \int_0^1 \frac{1}{(\nabla^2 f(x + \tau(x(s) - x)) - \nabla^2 f(x))} d\tau \ (x(s) - x), \\
\text{r}_2 &:= \sum_{\substack{i \neq \{I_0\}}} \left( \int_0^1 [v_i(x, s) - v_i(x + \tau(x(s) - x), s)] d\tau \right) a_i a_i^T (x - x(s)), \\
\text{r}_3 &:= \sum_{i \in I_0} \left( \int_0^1 [v_i(x, s) - v_i(x + \tau(x(s) - x), s)] d\tau \right) a_i a_i^T (x - x(s)),
\end{align*}
\] (3.33)

and we estimate the solutions \( d_j \) of the linear systems

\[
\left( \nabla^2 f(x) + \sum_{i=1}^m v_i(x, s) a_i a_i^T \right) d_j = r_j, \quad j = 1, 2, 3,
\] (3.34)

separately.

The supposed twice Lipschitz continuous differentiability of \( f \) and (3.27) yield

\[
\|d_i\|_C \leq c_1 \|x - x(s)\|^2_C,
\] (3.35)

with some constant \( c_1 > 0. \)

Next, we estimate \( \|d_2\|_C, \) i.e. we consider the case \( i \notin I_0. \) With (2.5) and (3.19) we have

\[
|v_i(x, s) - v_i(x + \tau(x(s) - x), s)| \leq L_2(-\frac{\delta}{s_0}) \left\| \frac{\tau a_i^T (x - x(s))}{(a_i^T x - b_i)(a_i^T (x + \tau(x(s) - x)) - b_i)} \right\|
\leq \delta^2 L_2(-\frac{\delta}{s_0}) |a_i^T (x - x(s))|, \quad i \notin I_0.
\] (3.36)

18
Now, (3.19) and (3.27) guarantee the existence of some $c_2 > 0$ such that

$$
\|d_2\|_C \leq c_2 \|x - x(s)\|^2_C. \tag{3.37}
$$

The remaining estimation of $\|d_3\|_C$ cannot follow the above scheme since $\|r_3\|_C$ is not of order $O(\|x - x(s)\|^2_C)$. The Lipschitz continuity (2.4) for $\psi'$ results in

$$
|v_i(x, s) - v_i(x + \tau(x(s) - x), s)| = s^{-1} \left| \psi'(\frac{a_i^T x - b_i}{s}) - \psi'(\frac{a_i^T (x + \tau(x(s) - x)) - b_i}{s}) \right| \leq s^{-2} L_1(\tau + \theta) |a_i^T (x(s) - x)|, \quad i \in I_0. \tag{3.38}
$$

Hence, we have

$$
r_3 = \sum_{i \in I_0} v_i(s) a_i a_i^T (x - x(s)), \tag{3.39}
$$

where $|v_i(s)| \leq c s^{-2} \|x - x(s)\|$ with some $c > 0$. Now, for the linear system (3.34) in case of the right hand side $r_3 \in \mathbb{R}^n$ all the assumptions of Lemma 1 are satisfied for

$$
\alpha_i(s) := \alpha_i = v_i(x, s) = s^{-1} \psi'(\psi^{-1}(u_i(x, s))), \quad i \in I_0.
$$

From (3.22) follows that some $c_\ast > 0$ exists such that

$$
\alpha_i(s) \geq c_\ast s^{-1}, \quad i \in I_0.
$$

Furthermore the quotient $\sigma$ is bounded due to (3.26). Thus, we can estimate

$$
\|d_3\|_C \leq c_3 s^{-2} \|x - x(s)\|^2_C, \tag{3.40}
$$

with some constant $c_3 > 0$. Superposition yields

$$
\tilde{x} - x(s) = \sum_{j=1}^3 d_j,
$$

and with (3.35), (3.37), (3.40) we have

$$
\|\tilde{x} - x(s)\|_C \leq (c_1 + c_2 + c_3 s^{-1}) \|x - x(s)\|^2_C, \quad \forall x \in B, \|x - x(s)\| \leq \omega s, \quad s \in (0, s_0].
$$

Taking again into account (3.27), we can conclude that some $c_\delta, c_\rho > 0$ exist with

$$
\|\tilde{x} - x(s)\|_C \leq c_\delta s^{-2} \|x - x(s)\|^2_C, \quad \forall x \in B, \|x - x(s)\|_C \leq c_\rho s, \quad s \in (0, s_0]. \tag{3.41}
$$
4 Path-following with truncated Newton steps

The convergence estimates given in Theorem 2 lay the foundation for a path-following algorithm with truncated Newton iterations at each level $s_k$ of the barrier parameter $s > 0$ in the auxiliary problems (2.1) or equivalently (3.2). In this paper, in particular, we propose a parameter strategy which guarantees the convergence of the overall iteration to the wanted solution $x^*$ of the original problem (3.1) in the case of only one Newton step at each parameter level.

Truncated Newton path-following algorithm

Step 1: Select parameters $\varepsilon, \rho > 0$, $\nu \in (0,1)$ and $s_0 > 0$.

Select $x_0 \in B$ such that
\[ \|x_0 - x(s_0)\| \leq \rho s_0. \]  
(4.1)

Set $k := 0$.

Step 2: Determine $x_{k+1} \in \mathbb{R}^n$ as solution of the linear system
\[ \nabla f(x_k) + \sum_{i=1}^{m} u_i(x_k, s_k) a_i + \left( \nabla^2 f(x_k) + \sum_{i=1}^{m} v_i(x_k, s_k) a_i a_i^T \right) (x_{k+1} - x_k) = 0. \]  
(4.2)

Step 3: If $s_k \leq \varepsilon$ then stop. Otherwise set $s_{k+1} := \nu s_k$ and go to step 2 with $k := k + 1$.

We notice that due to the convergence $\lim_{s \to 0+0} x(s) = x^*$ some $c > 0$, $s_0 > 0$ exist with
\[ v_i(x, s) \leq c, \quad \forall i \notin I_0, \quad \|x - x(s)\| \leq \rho s, \quad s \in (0, s_0]. \]  
(4.3)

This implies that the energy norms $\| \cdot \|_{C(x,s)}$ considered in Theorem 2 are uniformly equivalent to the euclidean norm $\| \cdot \|$, i.e. with some $\delta, \delta > 0$ holds
\[ \delta \|z\| \leq \|z\|_{C(x,s)} \leq \delta \|z\|, \quad \forall x, z \in \mathbb{R}^n, \quad \|x - x(s)\| \leq \rho s, \quad s \in (0, s_0]. \]  
(4.4)

Theorem 3 If $s_0 > 0$, $\rho > 0$ are sufficiently small and if the parameter $\nu \in (0,1)$ is selected such that
\[ \nu^{-1} \left( c_{\nu} \delta^{-1} \delta^2 \rho^2 + c_L (1 - \nu) \right) \leq \rho, \]  
(4.5)

where $c_L$ is the constant from Corollary 1, then the truncated Newton path-following algorithm given above generates iterates $x_k \in B$, $k = 1, 2, \ldots$, which satisfy
\[ \|x_k - x(s_k)\| \leq \rho s_k, \quad k = 0, 1, \ldots. \]  
(4.6)
Furthermore, the algorithm terminates after at most $k^* := \lceil \ln(\varepsilon/s_0)/\ln(\nu) \rceil$ steps, where $\lceil t \rceil$ denotes the smallest natural number greater or equal $t$, and the following estimates

$$0 \leq f(x_{k^*}) - f(x^*) \leq (p + c) \varepsilon \quad \text{and} \quad \|x_{k^*} - x^*\| \leq (c_L + \rho) s_{k^*} \quad (4.7)$$

hold with some $c > 0$.

**Proof:** We show (4.6) by induction.

For $k = 0$ this inequality holds according to (4.1). Let it be true for some index $k \geq 0$. Since $0 < \rho \leq c\rho$ and $s_k \leq s_0$, Theorem 2 yields $x_{k+1} \in B$ and

$$\|x_{k+1} - x(s_k)\|_{C_{C(x_k, s_k)}} \leq c_\delta s_k^{-1} \|x_k - x(s_k)\|^2_{C_{C(x_k, s_k)}} \leq c_\delta \delta^2 \rho^2 s_k.$$

The triangle inequality and Corollary 1 imply

$$\|x_{k+1} - x(s_{k+1})\| \leq \|x_{k+1} - x(s_k)\| + \|x(s_k) - x(s_{k+1})\|$$

$$\leq \delta^{-1} \|x_{k+1} - x(s_k)\|_{C_{C(x_k, s_k)}} + \|x(s_k) - x(s_{k+1})\|$$

$$\leq c_\delta \delta^{-1} \rho^2 s_k + c_L(s_k - s_{k+1})$$

$$= \nu^{-1}(\delta^{-1} \rho^2 c_\delta \rho^2 + c_L(1 - \nu)) s_{k+1} \leq \rho s_{k+1}.$$

This completes the induction.

Since $s_k = \nu^k s_0$, we have $s_k \leq \varepsilon$ for any $k \geq \ln(\varepsilon/s_0)/\ln(\nu)$. Finally, the well known (cf. [8], [17], [12]) estimate

$$0 \leq f(x(s)) - f(x^*) \leq p s$$

of the logarithmic barrier method and the local Lipschitz continuity of $f$ together with (4.6) prove the first part of (4.7). The second part of (4.7) follows directly from (4.6) and from Corollary 1. ■

**Remark 1** Since

$$\lim_{\nu \to 1^{-1}} \nu^{-1}(c_\delta \delta^{-1} \rho^2 + c_L(1 - \nu)) = c_\delta \delta^{-1} \rho^2 < \rho$$

for any $\rho \in (0, \min\{c_\rho, c_\delta \delta^{-2}\})$ one can find parameters $\rho$ and $\nu \in (0, 1)$ that satisfy the conditions given in Theorem 3. □

**Acknowledgement** Sponsoring by Kuwait University grant SM-153 is gratefully acknowledged. We thank F. Jarre (Würzburg) for bringing the paper [20] to our attention.

**References**


