Global and Local Convergence of a Reduced Space Quasi–Newton Barrier Algorithm for Large–Scale Nonlinear Programming

Andreas Wächter* and Lorenz T. Biegler†
Department of Chemical Engineering
Carnegie Mellon University
Pittsburgh, PA 15213

CAPD Technical Report B-00-06
(August 14, 2000)

Abstract

A primal–dual quasi-Newton algorithm for the solution of barrier problems, as they arise in the context of interior point methods for nonlinear programming, is presented. After a detailed description of the algorithm, which is given as BFGS and SR1 version, the global and local convergence properties of the BFGS version is examined.

1 Introduction

This paper addresses some convergence properties of an algorithm for solving the barrier problem that arises when we want to solve the following (NLP) with an interior point algorithm:

$$\min \quad f(x) \tag{1.1a}$$
$$\text{s.t.} \quad c(x) = 0 \tag{1.1b}$$
$$x^{(i)} \geq 0 \qquad \text{for } i \in \mathcal{I} \tag{1.1c}$$

where $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ and $c : \mathbb{R}^n \longrightarrow \mathbb{R}^m$ with $m < n$ are sufficiently smooth, and $\mathcal{I} \subseteq \{1, \ldots, n\}$ denotes the set of indices of those components $x^{(i)}$ of x that are bounded.

In order to obtain a local solution of this problem, we solve a sequence of *barrier problems*

$$\min \quad f(x) - \mu \sum_{i \in \mathcal{I}} \ln(x^{(i)}) \tag{1.2a}$$
$$\text{s.t.} \quad c(x) = 0 \tag{1.2b}$$

for decreasing values of the *barrier parameter* $\mu > 0$, so that $\mu \to 0$.

*E-mail: `andreasw@andrew.cmu.edu`
†E-mail: `lb01@andrew.cmu.edu`

The paper is organized as follows: Section 2 motivates the design of the algorithm for solving the barrier problem and presents a detailed description. In Section 3 we prove a global convergence result for the BFGS version of our algorithm for the barrier problem. Section 4 shows convergence to a strict local minimizer of the barrier problem, which is followed by a discussion of R-linear convergence in Section 5 and 2-step superlinear convergence in Section 6.

In the following, $\|\cdot\|$ denotes the ℓ_2-norm for vectors or the corresponding induced matrix norm. For a vector $x \in \mathbb{R}^n$, an inequality is to be understood component-wise, e.g. $x > 0$ means that all components are positive. We further use the notation $\mathbb{R}^n_+ := \{x \in \mathbb{R}^n : x > 0\}$ and $\overline{\mathbb{R}}^n_+ := \{x \in \mathbb{R}^n : x \geq 0\}$.

2 The Algorithm

2.1 The Search Directions

In order to motivate the choice of the search direction, let's look at the first order optimality conditions (KKT-conditions) of the barrier problem (1.2):

$$\begin{aligned} \nabla\varphi_\mu(x) + A(x)\lambda &= 0 \qquad (2.3a)\\ c(x) &= 0, \qquad (2.3b) \end{aligned}$$

where we introduced $A(x) := \nabla c(x)$, and where λ denotes the Lagrangian multipliers for the equality constraints (1.2b). Solving this system of equations directly by a Newton-type method, leads to a *primal method*, which treats only the *primal variables* x as iterates. Instead, we introduce new *dual variables* v defined by

$$v^{(i)} := \begin{cases} \mu/x^{(i)} & \text{for } i \in \mathcal{I} \\ 0 & \text{otherwise.} \end{cases}$$

In addition, we define v^I as the subvector of v consisting only of the elements $v^{(i)}$ with $i \in \mathcal{I}$, and define x^I in a similar way. With this definition, it is easy to verify that (2.3) is equivalent to the *perturbed KKT conditions* for (1.1)

$$\begin{aligned} \nabla f(x) + A(x)\lambda - v &= 0 \qquad (2.4a)\\ c(x) &= 0 \qquad (2.4b)\\ X^I V^I e - \mu e &= 0, \qquad (2.4c) \end{aligned}$$

where e is the vector of all ones of appropriate dimension. Here and later in the paper, the capital letters X, V, X^I, and V^I denote diagonal matrices with the vector elements of x, v, x^I, and v^I, respectively, in the diagonal. Note, that for $\mu = 0$ these conditions together with the inequalities

$$x^I \geq 0 \qquad \text{and} \qquad v^I \geq 0 \qquad (2.5)$$

are the first order optimality conditions for the *original* problem (1.1), and that the dual variables v correspond to the multipliers for the bound constraints (1.1c) in that case. As *Primal-dual method*, our method solves system (2.4) for a fixed $\mu > 0$ by a Newton-type approach, and the iterates are always required to strictly satisfy the inequalities (2.5).

Our algorithm generates search directions for the primal variables x and the dual variables v^I. After the computation of a primal-dual search direction, a line search will be done in order to guarantee global convergence (see below). We will refer to the primal-dual iterates as $z := (x, v^I)$ and the corresponding search directions as $d := (d^x, d^{v^I})$. For our approach it will not be necessary to compute the multipliers λ for the original equality constraints explicitly.

In order to solve for a search direction we reduce the linear system arising from a linearization of (2.4) and obtain

$$\begin{bmatrix} W_k & A_k \\ A_k^T & 0 \end{bmatrix} \begin{pmatrix} d_k^x \\ \lambda_{k+1} \end{pmatrix} = -\begin{pmatrix} \nabla\varphi_\mu(x_k) \\ c_k \end{pmatrix} \tag{2.6}$$

$$d_k^{v^I} = \mu \left(X_k^I\right)^{-1} e - v_k^I - \left(X_k^I\right)^{-1} V_k^I d_k^{x^I}. \tag{2.7}$$

In terms of notation we introduced here $A_k := A(x_k)$, $c_k := c(x_k)$, and the *barrier function* $\varphi_\mu(x) := f(x) - \mu \sum_{i\in\mathcal{I}} \ln(x^{(i)})$. We further define the diagonal matrix $\Sigma_k = \mathrm{diag}(\tilde{\sigma}_k)$ by

$$\tilde{\sigma}_k^{(i)} := \begin{cases} v_k^{(i)}/x_k^{(i)} & \text{if } i \in \mathcal{I} \\ 0 & \text{otherwise,} \end{cases}$$

denote the Lagrangian of the original NLP (1.1) as $\mathcal{L} := f(x) + c(x)^T\lambda - v$, and introduce $W_k := \nabla^2_{xx}\mathcal{L}(x_k, \lambda_k) + \Sigma_k$, and $\lambda_{k+1} := \lambda_k + d_k^\lambda$.

It is well known that obtaining d_k^x from (2.6) is equivalent to solving the quadratic problem

$$\min_{d_x\in\mathbb{R}^n} \quad \nabla\varphi_\mu(x_k)^T d_x + \frac{1}{2} d_x^T W_k d_x \tag{2.8a}$$

$$\text{s.t.} \quad A_k^T d_x + c_k = 0, \tag{2.8b}$$

if the matrix W_k is positive definite in the null space of A_k^T.

Since in our applications we usually have few degrees of freedom compared to the total number of variables, we follow a *reduced space* approach to solve this QP (see e.g. [3]). For this, we choose a null space matrix Z_k of A_k^T, i.e. a $n \times (n-m)$ matrix of full rank with $A_k^T Z_k = 0$, and a range space matrix Y_k, such that the columns of $[Y_k\ Z_k]$ form a basis of $\mathbb{R}^n$. In our implementation we use a *coordinate decomposition* approach, i.e. we choose m columns of $A(x)^T$ that form a square, non-singular submatrix $C(x)$, and denote the remaining columns as $N(x)$. Without loss of generality we assume that $A(x)^T = [N(x)\ C(x)]$. With this we define

$$Z_k = \begin{bmatrix} I \\ -C_k^{-1} N_k \end{bmatrix} \text{ and } Y_k = \begin{bmatrix} 0 \\ I \end{bmatrix}.$$

The following derivations apply also to more general choices of Z_k and Y_k, which satisfy certain regularity assumptions.

Decomposing the primal step as

$$d_k^x = Y_k p_Y + Z_k p_Z \tag{2.9}$$

for vectors p_Y and p_Z, it can easily be seen that the solution of the QP (2.8) can be obtained by computing

$$p_Y = -\left[A_k^T Y_k\right]^{-1} c_k \tag{2.10}$$

$$p_Z = -H_k^{-1}\left(Z_k^T \nabla\varphi_\mu(x_k) + w_k\right). \tag{2.11}$$

Here, H_k stands for the reduced Hessian $Z_k^T W_k Z_k$ or a positive definite approximation to it, and w_k denotes the cross term $Z_k^T W_k Y_k p_Y$ or an approximation to it. Looking closer at (2.10) for the coordinate decomposition, we note that $A_k^T Y_k = C_k$ which shows the advantage of this choice of the range space matrix, since it allows to exploit the structure in A_k.

Since in our applications second order derivative information (in particular $\nabla^2_{xx}\mathcal{L}(x_k, \lambda_k)$) is often unavailable, we want to explore the possibilities of quasi-Newton updates for the reduced Hessian B_k within an interior-point framework (see Section 2.3). For similar reasons, we choose

$$w_k = Z_k \Sigma_k Y_k p_Y \tag{2.12}$$

in the following analysis, since this information is easily available. Neglecting this part of the cross term by choosing $w_k = 0$ led to convergence problems in practice, since the effect of the barrier term is undermined and one could observe that often variables crashed into bounds and get stuck there. On the other hand, we could in addition estimate the second part of the cross term, $Z_k \nabla^2_{xx}\mathcal{L}(x_k, \lambda_k) Y_k p_Y$, by finite differences, as in [3], which would give 1-step superlinear convergence as opposed to 2-step superlinear convergence.

Regarding our notation, we omit the iteration index on the vectors p_Y and p_Z. A solution of the barrier problem (1.2) is denoted by x_*, and we define the distance from the solution (in the primal space) as

$$e_k := \|x_k - x_*\| \qquad \text{and} \qquad \sigma_k := \max\{e_{k+1}, e_k\}.$$

2.2 A primal-dual merit function

Having computed the primal-dual search directions $d_k := (d_k^x, d_k^{v^I})$ from (2.10)-(2.11), (2.9) and (2.7), we need to decide for a step length $\alpha_k \in (0, 1]$ in order to obtain the next iterate

$$z_{k+1} = z_k + \alpha_k d_k,$$

where $z_k = (x_k, v_k^I)$. Among other things we need to ensure that the implicit positivity constraints $x_{k+1}^I > 0$ and $v_{k+1}^I > 0$ are satisfied, since a full step $x_{k+1} = x_k + d_k^x$, $v_{k+1}^I = v_k^I + d_k^{v^I}$ might violate these constraints.

In order to determine α_k, we define a primal-dual ℓ_2-penalty function

$$\phi_\nu(x, v) := \varphi_\mu(x) + \mathcal{V}_\mu(x, v) + \nu\|c(x)\|, \tag{2.13}$$

where

$$\mathcal{V}_\mu(x, v) := \sum_{i \in \mathcal{I}} \left(x^{(i)} v^{(i)} - \mu \ln(x^{(i)} v^{(i)}) \right)$$

(see [1]) achieves its global minimum of $|\mathcal{I}| \cdot \mu\,(1 - \ln(\mu))$ for all points satisfying $X^I V^I e = \mu e$; in other words, $\mathcal{V}$ is minimized, if and only if the relaxed complementarity condition (2.4c) is satisfied. It is understood that the value of $\phi_\nu(z)$ is infinity, if at least one of the components of x^I or v^I is non-positive, and as a consequence such a point will never be accepted in an Armijo line search, as it is performed in our algorithm. The value of the penalty parameter $\nu \geq 0$ is adapted during the optimization in order to ensure that the directions generated by the algorithms are descent directions for the merit function, if H_k in (2.11) is positive definite.

It is useful to note, that for d^x and d^v satisfying

$$X^I d^{v^I} + V^I d^{x^I} = \mu e - X^I V^I e,$$

as for example for d_k^v and d_k^x in (2.7), we have [1, p. 8]

$$(\nabla_z \mathcal{V}_\mu(z))^T d = -\left\| \left(X^I\right)^{-1/2} \left(V^I\right)^{-1/2} \left(X^I V^I e - \mu e\right) \right\|_2^2. \tag{2.14}$$

The merit function itself is not differentiable at all points, but the directional derivatives, denoted as $D\phi_\nu(z; d)$, exist.

2.3 Quasi-Newton Approximation of the reduced Hessian

As mentioned before, we want to explore the possibilities to employ quasi-Newton approximations of the reduced Hessian $H_k \approx Z_k^T W_k Z_k$ in (2.11) within this interior-point framework. Despite the fact that we then do not need to compute second order derivatives, this also has the advantage that there is no need to compute the multipliers λ to the original equality constraints explicitly.

A straightforward way to apply a quasi-Newton method would be to use BFGS updates to estimate the matrix $Z_k^T W_k Z_k$, what essentially corresponds to a solution of the barrier problem (1.2) for a fixed barrier parameter μ by a reduced Hessian SQP algorithm such as the one described by Biegler, Nocedal, and Schmid [3]. Despite the fact, that for such an approach good local convergence results (for fixed barrier parameter μ!) would already be guaranteed and that it would allow to bypass the explicit computation of the matrix Z_k, we think that this approach has two major drawbacks: First, after decreasing μ during the overall optimization H_k would be the approximation of the reduced Hessian for a different optimization problem, and some iterations will be needed before the estimate H_k is good enough to provide fast local convergence to the optimal point of the new barrier problem. This contradicts our goal to solve the barrier problems increasingly fast in order to have a fast local convergence rate of the overall algorithm for the original NLP. We could observe this behavior in practical tests. Secondly, during the solution of a barrier problem it can happen that iterates come close to their bounds and need to be "pushed away". In such a case, the term Σ_k, which is part of W_k, becomes very large and a quasi-Newton approximation is not appropriate in this case.

Alternatively, we note that the reduced Hessian consists of the two terms

$$Z_k^T W_k Z_k = Z_k^T \nabla_{xx}^2 \mathcal{L}(x_k, \lambda_k) Z_k + Z_k^T \Sigma_k Z_k$$

and only estimate the first term $B_k \approx Z_k^T \nabla_{xx}^2 \mathcal{L}(x_k, \lambda_k) Z_k$ by means of a quasi-Newton method, whereas the second term can be determined explicitly, since it is easily available, i.e. we choose

$$H_k = B_k + Z_k^T \Sigma_k Z_k. \tag{2.15}$$

The approach described by Armand, Gilbert, and Jan-Jégou [1] has similar motivations.

In our implementation we use BFGS or SR1 quasi-Newton updates for B_k. The advantage of using positive definite BFGS estimates is that then the reduced Hessian H_k of the barrier problem will always be positive definite and no modifications are necessary in order to obtain a descent direction for our merit function. On the other hand, the reduced Hessian of the Lagrangian of the original NLP, which is approximated by B_k, does not have to be positive definite at the local solution of the barrier problem even for small μ, if there are active bounds at the local solution of the original NLP. In this case, BFGS might produce very poor estimates. For this reason, we also tried SR1 to estimate B_k. Since then the overall estimate of the reduced Hessian H_k for the barrier problem can become indefinite (as in the case when exact second order derivatives are used away from a local solution), we perform an eigenvalue decomposition of H_k and correct negative or very small eigenvalues to be sufficiently positive. Alternatives to this approach are subject to further research.

In this paper we only analyze the BFGS option, and under the (strong) assumption that the reduced Hessian of the original NLP at the optimal point of the barrier problem, $Z_*^T \nabla_{xx}^2 \mathcal{L}(x_*, \lambda_*) Z_*$, is positive definite, we can show that our algorithm is globally convergent and is locally 2-step superlinearly convergent. Our analysis is very closely related to [3]. In particular, we use a similar criterion to determine whether an BFGS update should be performed at a certain iteration to the effect that we can guarantee the boundedness of B_k and B_k^{-1}.

We should emphasize here, that our quasi-Newton updates do not require the explicit computation of the multipliers λ for the equality constraints (see also [2]).

2.4 Description of the Algorithm

Algorithm:

Given: Fixed barrier parameter $\mu > 0$, initial point $z_0 = (x_0, v_0)$ with $x_0^I, v_0^I > 0$, initial estimate of the reduced Hessian B_0 (positive definite in the BFGS case), initial penalty parameter $\nu_{-1} > 0$, positive constants $0 < \tau < \bar{\tau} < 1$, $\eta \in (0,1)$, $\rho > 0$. Also, choose a summable sequence of positive number $\{\gamma_k\}$, i.e. $\sum_{k=0}^{\infty} \gamma_k < 0$.

1. Initialize iteration counter $k := 0$ and evaluate $f(x_0)$, $\nabla f(x_0)$, c_0, A_0; compute Y_0 and Z_0.
2. Compute range space p_Y step from (2.10).
3. Compute correction term w_k from (2.12).
4. Compute null space step p_Z from (2.11) (for SR1-version only: if necessary, make H_k sufficiently positive definite).
5. Compute primal search direction d_k^x from (2.9), and dual search direction $d_k^{v^I}$ from (2.7).
6. Do the line search:
   (a) Update penalty parameter using
   $$\nu_k := \begin{cases} \nu_{k-1} & \text{if } \nu_{k-1}\|c_k\| - \nabla\varphi_\mu(x_k)^T Y_k p_Y \\ & \qquad + \min\{0, w_k^T p_Z\} \geq \rho\|c_k\| \\ \frac{\nabla\varphi_\mu(x_k)^T Y_k p_Y - \min\{0, w_k^T p_Z\}}{\|c_k\|} + 2\rho & \text{otherwise} \end{cases} \tag{2.16}$$
   (b) Set $\alpha_k := 1$. (This is what we really do:
   $$\alpha_k := \max\{\alpha \in (0,1] : \alpha d_k^{x^I} \geq -\iota x_k^I \text{ and } \alpha d_k^{v^I} \geq -\iota v_k^I\}$$
   for $\iota \in (0,1)$ close to 1.)
   (c) Evaluate $f(x_k + \alpha_k d_k^x)$ and $c(x_k + \alpha_k d_k^x)$.
   (d) Check if Armijo sufficient decrease condition
   $$\phi_{\nu_k}(z_k + \alpha_k d_k) \leq \phi_{\nu_k}(z_k) + \eta\alpha_k D\phi_{\nu_k}(z_k; d_k) \tag{2.17}$$
   is satisfied (note, that the value of the merit function is assumed to be ∞, if variables are not in the interior). If not, choose a new trial step size $\alpha_k \in [\tau\alpha_k, \bar{\tau}\alpha_k]$ and go back to step 6c.
7. If (2.17) is satisfied, accept the new iterate
   $$z_{k+1} := z_k + \alpha_k d_k. \tag{2.18}$$
   Note, that from the definition of ϕ_ν we have $x_{k+1}^I > 0$ and $v_{k+1}^I > 0$.

8. Evaluate $f(x_{k+1})$, $\nabla f(x_{k+1})$, c_{k+1}, A_{k+1}; compute Y_{k+1} and Z_{k+1}.

9. Perform the quasi-Newton update of B_k:

   (a) Define

$$\begin{aligned} s_k &:= \alpha_k p_Z & (2.19) \\ y_k &:= Z_{k+1}^T \nabla f(x_{k+1}) - Z_k^T \nabla f(x_k). & (2.20) \end{aligned}$$

   (b) BFGS-version:
   If $s_k^T y_k > 0$ (we actually check $s_k^T y_k > 10^{-8} \|s_k\| \|y_k\|$) and the update criterion

$$\|p_Y\| \leq \gamma_k \|p_Z\| \tag{2.21}$$

   is satisfied, perform the BFGS update

$$B_{k+1} := B_k - \frac{B_k s_k s_k^T B_k}{s_k^T B_k s_k} + \frac{y_k y_k^T}{y_k^T s_k}, \tag{2.22}$$

   otherwise skip the update and set $B_{k+1} := B_k$.

   (c) SR1-version:
   If $(y_k - B_k s_k)^T s_k \neq 0$ (we actually check $|(y_k - B_k s_k)^T s_k| > 10^{-8} \|y_k - B_k s_k\| \|s_k\|$), perform SR1 update

$$B_{k+1} := B_k + \frac{(y_k - B_k s_k)(y_k - B_k s_k)^T}{(y_k - B_k s_k)^T s_k}$$

10. Increase iteration counter $k \leftarrow k+1$ and go back to step 2.

For the remainder of this paper we will assume the BFGS case.

3 Global Convergence Analysis

Since we do not want to assume the boundedness of B_k and B_k^{-1} beforehand we make use of the concept of *good iterates*, which was introduced Byrd and Nocedal [4]. In that paper, the following theorem [4, Theorem 2.1] was proven:

Theorem 3.1 *Let $\{B_k\}$ be generated by the BFGS formula (2.22) where, for all $k \geq 1$, $s_k \neq 0$ and*

$$\begin{aligned} \frac{y_k^T s_k}{s_k^T s_k} &\geq m > 0 & (3.1) \\ \frac{\|y_k\|^2}{y_k^T s_k} &\leq M & (3.2) \end{aligned}$$

for constants m, M. Then, there exist constants $\beta_1, \beta_2, \beta_3 > 0$ such that, for any $k \geq 1$, the relations

$$\begin{aligned} \cos\theta_j := \frac{s_j^T B_j s_j}{\|s_j\| \, \|B_j s_j\|} &\geq \beta_1 & (3.3) \\ \beta_2 \leq \frac{\|B_j s_j\|}{\|s_j\|} &\leq \beta_3 & (3.4) \end{aligned}$$

hold for at least $\lceil \frac{1}{2} k \rceil$ values of $j \in [1, k]$. We will call those iterates j the "good iterates" and denote the set of all good iterates as J and define $J_k := J \cap [1, k]$.

We now state the assumptions made in this section:

Assumption 3.1 *The sequence $\{x_k\}$ generated by the algorithm is contained in a* bounded *convex set D with the following properties:*

I. The functions $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ and $c : \mathbb{R}^n \longrightarrow \mathbb{R}^m$ and their first and second derivatives are bounded in norm over D.

II. The matrix $A(x)$ has full column rank for all $x \in D$, and there exist constants γ_0, β_0, and $\tilde{\beta}_0$ such that

$$\|Y(x)[A(x)^T Y(x)]^{-1}\| \leq \gamma_0, \quad \|Z(x)\| \leq \beta_0, \quad \sigma_{min}(Z(x)) \geq \tilde{\beta}_0 > 0 \tag{3.5}$$

for all $x \in D$, where σ_{min} denotes the smallest singular value.

III. For all $k \geq 0$ for which B_k is updated, conditions (3.1) and (3.2) hold.

Note that these assumptions are only made in the "primal space".

Admittedly, Assumption 3.1 III may not be satisfied for general NLPs. But as we will show below in Lemma 4.1, these conditions are valid in a neighborhood of a solution of the barrier problem (satisfying the strong Assumption 4.1 III, see below) whenever BFGS updating takes place. As a consequence of this assumption, we see from Theorem 3.1, that $|J| = \infty$ if BFGS updating takes place an infinite number of times. If BFGS updates are performed only in a limited number of iterations, *all* iterates are good iterates, so that we again obtain $|J| = \infty$.

We start the global analysis by showing that the algorithm generates directions of descent for the merit function as long as the current iterate is no KKT point. This assures that the Armijo line search will be successful in every iteration.

Lemma 3.1 *At an iterate $z_k = (x_k, v_k^I)$ with $(x_k^I, v_k^I) > 0$, which* does not *satisfy the KKT conditions, i.e. it* does not *satisfy*

$$\begin{aligned} Z(x_k)^T \nabla\varphi_\mu(x_k) &= 0 \\ X_k^I V_k^I e - \mu e &= 0 \\ c(x_k) &= 0, \end{aligned}$$

let $d_k = (d_k^x, d_k^{v^I})$ be the search direction computed by the algorithm. Then, it is $D\phi_{\nu_k}(z_k, d_k) < 0$.

Proof: From (2.14), Equation (2.24) in [5], (2.9), and (2.11) we have

$$\begin{aligned} D\phi_{\nu_k}(z_k; d_k) &= \nabla\varphi_\mu(x_k)^T d_k^x + \nabla_z \mathcal{V}(z_k)^T d_k - \nu_k \|c(x_k)\| \\ &= \nabla\varphi_\mu(x_k)^T Z_k p_Z + \nabla\varphi_\mu(x_k)^T Y_k p_Y + \nabla_z \mathcal{V}(z_k)^T d_k - \nu_k \|c(x_k)\| \\ &= -p_Z^T H_k p_Z - w_k^T p_Z + \nabla_z \mathcal{V}(z_k)^T d_k + \nabla\varphi_\mu(x_k)^T Y_k p_Y - \nu_k \|c(x_k)\| \\ &\leq -p_Z^T H_k p_Z - \| \left(X^I\right)^{-\frac{1}{2}} \left(V^I\right)^{-\frac{1}{2}} \left(X^I V^I e - \mu e\right) \|^2 - \rho \|c(x_k)\|, \end{aligned}$$

where the last inequality follows from (2.16). Since H_k is assumed to be positive definite and $(x_k^I, v_k^I) > 0$, the claim follows immediately. $\square$

An important role plays the asymptotic behavior of penalty parameter. The analysis later in this paper will assume, that the sequence $\{\nu_k\}$ is bounded. The following two lemmas provide some insight under what conditions this is the case.

Lemma 3.2 *Assume the algorithm uses the update rule (2.16) for the penalty parameter* ν_k *and Assumptions 3.1 are valid. If* $\{x_k\}$ *has a limit point* x_* *at the boundary of the feasible region, i.e.* $x_*^I \not> 0$*, then* $\nu_k \to \infty$*.*

Proof: We prove the lemma by contradiction: Assume, that $\{\nu_k\}$ is bounded. Then, by the update rule, we have $\nu_k \equiv \bar{\nu}$ for sufficiently large k. Let $\{x_{k_l}\}$ be a subsequence of $\{x_k\}$ with $\lim_{l\to\infty} x_{k_l} = x_*$. If L is such that for $l \geq L$ we have $\nu_{k_l} \equiv \bar{\nu}$, then for $l \geq L$

$$\begin{aligned} \phi_{\bar{\nu}}(z_{k_L}) &\geq \phi_{\bar{\nu}}(z_{k_l}) \\ &= f(x_{k_l}) - \mu \sum_{i\in\mathcal{I}} \ln(x_{k_l}^{(i)}) + \mathcal{V}(z_{k_l}) + \bar{\nu}\|c_{k_l}\| \\ &\geq f_L + |\mathcal{I}| \cdot \mu\,(1 - \ln(\mu)) - \mu \sum_{i\in\mathcal{I}} \ln(x_{k_l}^{(i)}), \end{aligned}$$

where f_L is a lower bound for $f(x)$ on D, and it is easy to verify, that $|\mathcal{I}| \cdot \mu\,(1 - \ln(\mu))$ is the global minimum of $\mathcal{V}(z)$. Thus, the last term has to be bounded, which implies that all components of $\{x_{k_l}\}$ are bounded away from zero. This contradicts that x_*^I lies at the boundary of $\overline{\mathbb{R}}_+^{|\mathcal{I}|}$. □

Lemma 3.3 *Assume the algorithm uses the update rule (2.16) for the penalty parameter* ν_k *and Assumptions 3.1 are valid. Suppose further, that there exist positive constants* c_x *and* c_H*, so that*

$$\left\| \left(X_k^I\right)^{-1} \right\| \leq c_x \qquad \text{and} \qquad \left\| H_k^{-1} \right\| \leq c_H.$$

Then, the sequence of penalty parameters $\{\nu_k\}$ *is bounded.*

Proof: We first show, that there exists $c_\Sigma > 0$ such that $\|\Sigma_k\| \leq c_\Sigma$ for all k. For this, we need to show that $x_k^{(i)} v_k^{(i)}$ is bounded for all $i \in \mathcal{I}$. Note, that from Assumption 3.1 I it is $x_k \leq m_x$ for some $m_x > 0$ and all k.

Let $i \in \mathcal{I}$ and $k \in \{0, 1, 2, \ldots\}$. For simplicity, we will write x, v, and α for $x_k^{(i)}$, $v_k^{(i)}$, and α_k, x_+ and v_+ for $x_{k+1}^{(i)}$ and $v_{k+1}^{(i)}$, and d^x and d^v for $(d_k^x)^{(i)}$ and $(d_k^v)^{(i)}$. We distinguish two cases:

Case a) : $xv \geq \mu$. Define $g(t,s) := ts$. It then follows from the mean value theorem that

$$\begin{aligned} x_+v_+ - xv &= g(x + \alpha d^x, v + \alpha d^v) - g(x, v) \\ &= \nabla g(x + \xi d^x, v + \xi d^v)^T \begin{pmatrix} d^x \\ d^v \end{pmatrix}^T \qquad \text{for some } \xi \in (0, \alpha) \\ &= \underbrace{(v + \xi d^v)}_{>0}{}^T d^x + \underbrace{(x + \xi d^x)}_{>0}{}^T d^v \end{aligned} \tag{3.6}$$

This is non-positive if both $d^x < 0$ and $d^v < 0$. For the other case we note from

$$xd^v + vd^x = \mu - xv \leq 0, \tag{3.7}$$

(see (2.7)) that d^x and d^v have opposite signs, so that with (3.6) and (3.7)

$$\begin{aligned} x_+v_+ - xv &= (v + \xi d^v)^T d^x + (x + \xi d^x)^T d^v \\ &= vd^x + xd^v + 2\xi d^x d^v \\ &\overset{(3.7)}{\leq} 2\xi d^x d^v \leq 0. \end{aligned}$$

In summary, for this case we obtain

$$x_k^{(i)} v_k^{(i)} \geq \mu \qquad \Longrightarrow \qquad x_{k+1}^{(i)} v_{k+1}^{(i)} \leq x_k^{(i)} v_k^{(i)} \tag{3.8}$$

Case b) : $xv < \mu$. It then follows that

$$\begin{aligned} x_+ v_+ &= (x + \alpha d^x)(v + \alpha d^v) \\ &= xv + \alpha(xd^v + vd^x) + \alpha^2 d^x d^v \\ &\overset{(2.7)}{=} xv + \alpha(\mu - xv) + \alpha^2 d^x d^v \\ &= (1-\alpha) \underbrace{xv}_{<\mu} + \alpha\mu + \alpha^2 d^x d^v \\ &\leq \mu + |\alpha d^x| \; |\alpha d^v| \,. \end{aligned}$$

Since $x \leq m_x$, it also follows $|\alpha d^x| \leq m_x$. Furthermore, if $d^v \leq 0$, it follows from $v_+ = v + \alpha d^v > 0$

$$|\alpha d^v| = -\alpha d^v < v < \frac{\mu}{x} \leq \mu c_x,$$

whereas if $d^v > 0$

$$|\alpha d^v| = \alpha d^v \overset{(2.7)}{=} \alpha \frac{\mu}{x} - \alpha v - \frac{v}{x} \alpha d^x \leq \mu c_x + 0 + \mu c_x^2 \, |\alpha d^x| \leq \mu(c_x + c_x^2 m_x),$$

so that in summary

$$x_k^{(i)} v_k^{(i)} < \mu \qquad \Longrightarrow \qquad x_{k+1}^{(i)} v_{k+1}^{(i)} \leq \mu(1 + m_x(c_x + c_x^2 m_x)). \tag{3.9}$$

Combining cases a) and b) it finally follows from induction that

$$x_k^{(i)} v_k^{(i)} \leq \max_{j \in \mathcal{I}} \{ x_0^{(j)} v_0^{(j)}, \mu(1 + m_x(c_x + c_x^2 m_x)) \} =: c_\Sigma,$$

and thus $\|\Sigma_k\| \leq c_\Sigma$ for all k.

In order to prove the boundedness of the penalty parameter we show that the two terms in the update rule (2.16) are bounded above. We first note, that no update takes place, if $c_k = 0$, since this implies $p_Y = 0$ from (2.10) and $w_k = 0$ from (2.12). Thus we can assume, that $c_k \neq 0$.

For the first term in the update rule, it is

$$\begin{aligned} \frac{\nabla \varphi_\mu(x_k)^T Y_k p_Y}{\|c_k\|} &= \frac{-\left(\nabla \varphi_\mu(x_k)\right)^T Y_k [A_k^T Y_k]^{-1} c_k}{\|c_k\|} \\ &\leq \frac{\|\nabla \varphi_\mu(x_k)\| \; \|Y_k [A_k^T Y_k]^{-1}\| \; \|c_k\|}{\|c_k\|} \\ &\leq \gamma_0 \left(\|\nabla \varphi_\mu(x_k)\| \right) \\ &\leq \gamma_0 \left(\|\nabla f(x_k)\| + \mu \left\| \left(X_k^I\right)^{-1} e \right\| \right) \leq \gamma_0 \left(\|\nabla f(x_k)\| + \mu c_x \right), \end{aligned}$$

where γ_0 comes from (3.5). This is bounded above since $\nabla f(x)$ is bounded over D by assumption.

Dealing with the second term, we first note that

$$\|w_k\| = \left\| Z_k^T \Sigma_k Y_k p_Y \right\| \leq \|Z_k^T\| \; \|\Sigma_k\| \; \|Y_k [A_k^T Y_k]^{-1}\| \; \|c_k\| \leq \beta_0 c_\Sigma \gamma_0 \|c_k\| =: \tilde{c} \|c_k\|.$$

Since further by (2.11)

$$\begin{aligned}
|w_k^T p_Z| &= \left|w_k^T H_k^{-1}\left(Z_k^T \nabla\varphi_\mu(x_k) + w_k\right)\right| \\
&\le \left|w_k^T H_k^{-1} Z_k^T \nabla\varphi_\mu(x_k)\right| + \left|w_k^T H_k^{-1} w_k\right| \\
&\le \|w_k\|\,\|H_k^{-1}\|\,\|Z_k^T\|\left(\|\nabla f(x_k)\| + \mu\|\left(X_k^I\right)^{-1} e\|\right) + \|H_k^{-1}\|\,\|w_k\|^2 \\
&\le \tilde{c}\, c_H \beta_0 \left(\|\nabla f(x_k)\| + \mu c_x\right)\|c_k\| + c_H\, \tilde{c}^2 \|c_k\|^2,
\end{aligned}$$

it follows, that $|w_k^T p_Z|/\|c_k\|$ is bounded above. This concludes the proof. □

For now, let's explore the case, when the penalty parameter ν_k stays bounded. We first show that the dual variables are bounded and stay away from the boundary of the nonnegative orthant.

Lemma 3.4 *If ν_k is bounded and Assumptions 3.1 hold, then the sequences $\{\left(X_k^I\right)^{-1} e\}$, $\{v_k^I\}$ and $\{\left(V_k^I\right)^{-1} e\}$ are bounded.*

Proof: Since ν_k is bounded, it follows from the update rule that there exist a K such that for all $k \ge K$ we have $\nu_k \equiv \bar{\nu}$. By the line search, $\{\phi_{\bar{\nu}}(z_k)\}_{k \ge K}$ is a decreasing sequence which implies for all $k \ge K$

$$\phi_{\bar{\nu}}(z_K) \ge \phi_{\bar{\nu}}(z_k) \ge \varphi_\mu(x_k) + \mathcal{V}_\mu(z_k).$$

By assumption, f is bounded below and x_k is bounded in norm, for which $\varphi_\mu(x_k)$ is bounded below, and as a consequence, $\mathcal{V}_\mu(z_k)$ has to be bounded above. If we define $g(t) := t - \mu \ln(t)$, it is $\mathcal{V}_\mu(z_k) = \sum_{i \in \mathcal{I}} g(x_k^{(i)} v_k^{(i)})$, and from $\lim_{t\to 0} g(t) = \lim_{t\to\infty} g(t) = \infty$ we see that the boundedness of $\mathcal{V}_\mu$ implies

$$l^{(i)} \le x_k^{(i)} v_k^{(i)} \le u^{(i)} \qquad \text{for } i \in \mathcal{I} \tag{3.10}$$

for positive constants $l^{(i)}$ and $u^{(i)}$ independent of k. Using the boundedness of $\{\nu_k\}$, Lemma 3.2 guarantees the boundedness of $\{(X_k^I)^{-1} e\}$, which together with (3.10) and the boundedness of $\{x_k\}$ proves the claim. □

This boundedness allows to show that for good iterates conditions similar to (3.3) and (3.4) are also valid for the overall reduced Hessian $H_k = B_k + Z_k^T \Sigma_k Z_k$:

Lemma 3.5 *If ν_k is bounded and Assumptions 3.1 hold, then there exist positive constants $\bar{\beta}_1, \bar{\beta}_2, \bar{\beta}_3$ such that for all $j \in J$*

$$\begin{aligned}
\cos\bar{\theta}_j := \frac{s_j^T H_j s_j}{\|s_j\|\,\|H_j s_j\|} &\ge \bar{\beta}_1 \\
\bar{\beta}_2 \le \frac{\|H_j s_j\|}{\|s_j\|} &\le \bar{\beta}_3
\end{aligned}$$

Proof: We first note that for all $j \in J$

$$\|Z_j^T \Sigma_j Z_j s_j\| \le \beta_0^2 c_1 \|s_j\| \le \frac{\beta_0^2 c_1}{\beta_2}\|B_j s_j\| =: c_2 \|B_j s_j\|,$$

where we used (3.5) and (3.4), and defined $c_1 := \max_{j \in J}\{\|\Sigma_j\|\}$ which is finite by Lemma 3.4. Using this, (2.15) and (3.3), we obtain

$$\frac{s_j^T H_j s_j}{\|s_j\|\,\|H_j s_j\|} = \underbrace{\frac{s_j^T Z_j^T \Sigma_j Z_j s_j}{\|s_j\|\,\|H_j s_j\|}}_{\ge 0} + \frac{s_j^T B_j s_j}{\|s_j\|\,\|Z_j^T \Sigma_j Z_j s_j + B_j s_j\|}$$

$$\geq \quad \frac{s_j^T B_j s_j}{\|s_j\| \cdot (1+c_2)\|B_j s_j\|}$$
$$\geq \quad \frac{\beta_1}{1+c_2} =: \bar{\beta}_1.$$

Next, we see from (3.4), (3.5) and Lemma 3.4, that

$$\frac{\|H_j s_j\|}{\|s_j\|} \leq \frac{\|B_j s_j\|}{\|s_j\|} + \|Z_j^T \Sigma_j Z_j\| \leq \beta_3 + \beta_0^2 c_1 =: \bar{\beta}_3.$$

Finally, since

$$\|s_j\| \, \|H_j s_j\| \geq s_j^T H_j s_j = s_j^T B_j s_j + s_j^T Z_j^T \Sigma_j Z_j s_j \geq s_j^T B_j s_j,$$

it follows from (3.3) and (3.4)

$$\frac{\|H_j s_j\|}{\|s_j\|} \geq \frac{s_j^T B_j s_j}{\|s_j\|^2} \geq \frac{s_j^T B_j s_j}{\|s_j\| \, \|B_j s_j\|} \cdot \frac{\|B_j s_j\|}{\|s_j\|} \geq \beta_1 \beta_2 =: \bar{\beta}_2.$$

□

This allows is to show that the algorithm generates directions of sufficient descent for good iterates $j \in J$:

Lemma 3.6 *Let $j \in J$ and assume, that $s_j = \alpha_j p_Z$ for the BFGS update. Then there exists $b_2 > 0$ such that*

$$D\phi_{\nu_j}(z_j; d_j) \leq -b_2 \left(\|Z_j^T \nabla \varphi_\mu(x_j)\|^2 + \|(X_j^I)^{-\frac{1}{2}} (V_j^I)^{-\frac{1}{2}} \left(X_j^I V_j^I e - \mu e \right) \|^2 + \rho \|c_j\| \right). \tag{3.11}$$

Proof: Can be shown similar to pp. 10-12 in [2]. Detailed proof in later version of paper. □

The next Lemma is a generalization of Lemma 4.1 in [3] for the primal-dual merit function and shows that for good iterates the reduction in the merit function is a fraction of the KKT-error.

Lemma 3.7 *If Assumptions 3.1 hold and if $\nu_j \equiv \bar{\nu}$ for all sufficiently large $j \in J$, then there is a positive constant $\gamma_{\bar{\nu}}$ such that for all large $j \in J$,*

$$\phi_{\bar{\nu}}(z_j) - \phi_{\bar{\nu}}(z_{j+1}) \geq \gamma_{\bar{\nu}} \left(\|Z_j^T \nabla \varphi_\mu(x_j)\|^2 + \|X_j^I V_j^I e - \mu e\|^2 + \|c_j\| \right). \tag{3.12}$$

Proof: Since $\{\nu_k\}$ is bounded with $\lim \nu_k = \bar{\nu}$, it follows from Lemma 3.4 that

$$\begin{aligned} c_1 &:= \min_{j \in J} \{ \|(X_j^I)^{-\frac{1}{2}} (V_j^I)^{-\frac{1}{2}}\|^2 \} > 0, \\ c_2 &:= \max_{j \in J} \{ \|(X_j^I)^{-1}\|^2 \} < \infty, \qquad \text{and} \\ c_3 &:= \max_{j \in J} \{ \|(X_j^I)^{-1} V_j^I\|^2 \} < \infty \end{aligned}$$

for some constants c_1, c_2, c_3.

The proof is to a large part identical with the proof in [3]: For sufficiently large good iterates $j \in J$ we have from (3.11)

$$D\phi_{\bar{\nu}}(z_j; d_j) \leq -b_2 \left(\|Z_j^T \nabla \varphi_\mu(x_j)\|^2 + \|X_j^I V_j^I e - \mu e\|^2 + \|c_j\| \right), \tag{3.13}$$

where $b_2 := \min\{\beta_1/\beta_3, \rho, c_1\}$. The Armijo condition (2.17) states

$$\phi_{\bar{\nu}}(z_j) - \phi_{\bar{\nu}}(z_{j+1}) \geq -\eta \alpha_j D\phi_{\bar{\nu}}(z_j; d_j).$$

From the last two inequalities it is clear, that (3.12) holds, if α_j, $j \in J$, can be uniformly bounded away from zero. Suppose that $\alpha_j < 1$, which means that for a $\tilde{\alpha} \le \alpha_j/\tau$:

$$\phi_{\bar{\nu}}(z_j + \tilde{\alpha} d_j) - \phi_{\bar{\nu}}(z_j) > \eta \tilde{\alpha} D\phi_{\bar{\nu}}(z_j; d_j).$$

On the other hand, expanding to second order, we have

$$\phi_{\bar{\nu}}(z_j + \tilde{\alpha} d_j) - \phi_{\bar{\nu}}(z_j) \le \tilde{\alpha} D\phi_{\bar{\nu}}(z_j; d_j) + \tilde{\alpha}^2 b_1 \|d_j\|^2,$$

where b_1 depends on $\bar{\nu}$. Combining the last two inequalities gives

$$(\eta - 1) D\phi_{\bar{\nu}}(z_j; d_j) < \tilde{\alpha} b_1 \|d_j\|^2. \tag{3.14}$$

In [3] it was shows (see Equation (4.11) there), that

$$\|d_j^x\|^2 \le b_3 \left(\|Z_j^T \nabla \varphi_\mu(x_j)\|^2 + \|c_j\| \right)$$

for some constant b_3. Further examining

$$\begin{aligned} \|d_j^{v^I}\|^2 &= \|\mu (X_j^I)^{-1} e - v_j^I - (X_j^I)^{-1} (V_j^I) d_j^{x^I}\|^2 \\ &= \|(X_j^I)^{-1} \left(\mu e - X_j^I V_j^I e \right) + (X_j^I)^{-1} V_j^I d_j^{x^I}\|^2 \\ &\le 3 c_2 \|X_j^I V_j^I e - \mu e\|^2 + 3 c_3 \|d_j^x\|^2 \end{aligned}$$

gives for the overall step

$$\begin{aligned} \|d_j\|^2 &= \|d_j^x\|^2 + \|d_j^{v^I}\|^2 \\ &\le (1 + 3c_3) \|d_j^x\|^2 + 3c_2 \|X_j^I V_j^I e - \mu e\|^2 \\ &\le (1 + 3c_3) b_3 \left(\|Z_j^T \nabla \varphi_\mu(x_j)\|^2 + \|c_j\| \right) + 3c_2 \|X_j^I V_j^I e - \mu e\|^2 \\ &\le b_4 \left(\|Z_j^T \nabla \varphi_\mu(x_j)\|^2 + \|X_j^I V_j^I e - \mu e\|^2 + \|c_j\| \right) \end{aligned} \tag{3.15}$$

with $b_4 := \max\{(1 + 3c_3) b_3, 3c_2\}$. Combining (3.14), (3.13), and (3.15), and recalling that $\eta < 1$, we obtain

$$\tilde{\alpha} > \frac{(1 - \eta) b_2}{b_1 b_4}.$$

Thus, $\{\alpha_j\}$ is uniformly bounded away from zero. The claim follows with

$$\gamma_{\bar{\nu}} := \eta b_2 \min\{1, (1 - \eta) \tau b_2 / (b_1 b_4)\}.$$

□

This result allows us to proof the following global convergence result (analogue to Theorem 4.2 in [3]).

Theorem 3.2 *If Assumptions 3.1 hold and $\{\nu_k\}$ is bounded, then*

$$\liminf_{k \to \infty} \left(\|Z_k^T \nabla \varphi_\mu(x_k)\| + \|X_k^I V_k^I e - \mu e\| + \|c_k\| \right) = 0.$$

Proof: Since $\{\nu_k\}$ is bounded, there is a K such that for all $k \ge K$ we have $\nu_k \equiv \bar{\nu}$. Then, by Lemma 3.7, for all $k > K$

$$\begin{aligned} \phi_{\bar{\nu}}(z_K) - \phi_{\bar{\nu}}(z_{k+1}) &= \sum_{j=K}^{k} \left(\phi_{\bar{\nu}}(z_j) - \phi_{\bar{\nu}}(z_{j+1}) \right) \\ &\ge \sum_{j \in J \cup [K,k]} \left(\phi_{\bar{\nu}}(z_j) - \phi_{\bar{\nu}}(z_{j+1}) \right) \\ &\ge \gamma_{\bar{\nu}} \sum_{j \in J \cup [K,k]} \left[\|Z_j^T \nabla \phi_\mu(x_j)\|^2 + \|\mu e - X_j^I V_j^I e\|^2 + \|c_j\| \right]. \end{aligned}$$

Since, by Assumption 3.1, $\phi_{\bar{\nu}}(z)$ is bounded below, the last sum is finite, so that the term inside the square brackets converges to zero. □

Note, that this result can also be shown for *any* approximation H_k under the strong assumption, that the sequences $\{H_k\}$ and $\{H_l^{-1}\}$ are bounded.

4 Local Convergence

In this section we will refer to multipliers λ for the equality constraints (1.2b). We will assume that they are computed as first order multipliers by

$$\lambda(x) = -\left[Y(x)^T A(x)\right]^{-1} Y(x)^T \nabla\varphi_\mu(x), \tag{4.1}$$

although, due to our multiplier-free update of the penalty parameter ν_k and multiplier-free quasi-Newton updates for B_k, an explicit computation of λ_k is not necessary.

First, we state the assumptions made for local convergence.

Assumption 4.1 *The point $x_* \in \mathbb{R}^n$ with $x_*^I > 0$ is a local minimizer for the barrier problem, at which the following conditions hold.*

I. The functions $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ and $c : \mathbb{R}^n \longrightarrow \mathbb{R}^m$ are twice continuously differentiable in a neighborhood of x_*, and their Hessians are Lipschitz continuous in a neighborhood of x_*.

II. The matrix $A(x_*)$ has full column rank. This implies that there exists a unique vector $\lambda_* \in \mathbb{R}^m$ such that the KKT-system

$$\begin{aligned} \nabla_x \mathcal{L}(x_*, \lambda_*) - v_* = \nabla_x f(x_*) + A(x_*)\lambda_* - v_* &= 0 \\ -\mu e + X_*^I V_*^I &= 0 \\ c(x_*) &= 0 \end{aligned}$$

is satisfied with v_* defined by

$$v_*^{(i)} := \begin{cases} \mu / x_*^{(i)} & \text{for } i \in \mathcal{I} \\ 0 & \text{otherwise.} \end{cases}$$

III. The reduced Hessian of the original NLP *(1.1)*

$$G_* := Z_*^T \nabla^2_{xx} \mathcal{L}(x_*, \lambda_*) Z_*$$

is positive definite.

IV. There exist constants γ_0, β_0 and γ_c such that, for all x in a neighborhood of x_*,

$$\|Y(x)[A(x)^T Y(x)]^{-1}\| \le \gamma_0, \quad \|Y(x)\ Z(x)\| \le \beta_0, \tag{4.2}$$

and

$$\|[Y(x)\ Z(x)]^{-1}\| \le \gamma_c. \tag{4.3}$$

V. $Z(x)$ and $\lambda(x)$ are Lipschitz continuous in a neighborhood of x_*, i.e. there are constants γ_Z and γ_λ such that

$$\|\lambda(x) - \lambda(y)\| \leq \gamma_\lambda \|x - y\| \tag{4.4}$$

$$\|Z(x) - Z(y)\| \leq \gamma_Z \|x - y\|, \tag{4.5}$$

for all x, y close to x_*.

Again, all assumptions are made in the "primal space".

The next lemma shows that Assumption 3.1 III is indeed reasonable in a neighborhood of a local solution.

Lemma 4.1 *Let x_* be a solution of the barrier problem satisfying Assumptions 4.1. Then, for large enough k and x_k close enough to x_*, there exist $m, M > 0$, so that conditions (3.1) and (3.2) are valid whenever BFGS updating takes place.*

Proof: (cf. [2]). Defining

$$\tilde{W}_k := \int_0^1 \nabla^2_{xx} \mathcal{L}(x_k + t\alpha_k d_k, \lambda_*) dt, \tag{4.6}$$

we have from (2.20), the mean value theorem, (2.18), (2.9), (2.19), (4.4), and (4.3)

$$\begin{aligned} y_k &= Z_{k+1}^T \nabla f(x_{k+1}) - Z_k^T \nabla f(x_k) \\ &= Z_{k+1}^T \nabla_x \mathcal{L}(x_{k+1}, \lambda_*) - Z_k^T \nabla_x \mathcal{L}(x_k, \lambda_*) \\ &= Z_k^T \tilde{W}_k \alpha_k d_k + (Z_{k+1} - Z_k)^T \nabla_x \mathcal{L}(x_{k+1}, \lambda_*) \\ &= Z_k^T \tilde{W}_k Z_k s_k + \alpha_k Z_k^T \tilde{W}_k Y_k p_Y + O(\sigma_k) \left(\|s_k\| + \alpha_k \|p_Y\|\right) \\ &= Z_k^T \tilde{W}_k Z_k s_k + O(\sigma_k)\|s_k\| + O(\alpha_k \|p_Y\|). \end{aligned}$$

Now, for iterates k at which the BFGS update is performed, we have from the update criterion (2.21)

$$y_k = Z_k^T \tilde{W}_k Z_k s_k + O(\sigma_k)\|s_k\| + O(\gamma_k)\|s_k\|$$

and thus for x_k close enough to x_* and k large enough

$$\begin{aligned} s_k^T y_k &\geq \frac{1}{2} \lambda_{min}(G_*) \|s_k\|^2 - O(\sigma_k + \gamma_k)\|s_k\|^2 \\ &\geq m\|s_k\|^2 \end{aligned}$$

for some positive constant m, which proves (3.1). Since further $y_k = O(\|s_k\|)$, we obtain

$$\frac{\|y_k\|^2}{y_k^T s_k} \leq M$$

for some $M > 0$. □

4.1 Local properties of the primal-dual merit function

The following generalization of Lemma 4.1 in [5] shows that locally the KKT-error at an iterate z_k is of the same order as the distance of z_k to the solution z_*.

Lemma 4.2 *If Assumptions 4.1 hold, then there exist positive constants* γ_1, $\tilde{\gamma}_1$, γ_2, *and* $\tilde{\gamma}_2$, *such that for all* $z = (x, v^I)$ *sufficiently close to* $z_* = (x_*, v_*^I)$

$$\begin{aligned} \tilde{\gamma}_1 \|x - x_*\| &\leq \|Z(x)^T \nabla \varphi_\mu(x)\| + \|c(x)\| \\ &\leq \tilde{\gamma}_2 \|x - x_*\| \end{aligned} \tag{4.7}$$

$$\begin{aligned} \gamma_1 \left(\|x - x_*\| + \|v^I - v_*^I\|\right) &\leq \|Z(x)^T \nabla \varphi_\mu(x)\| + \|c(x)\| + \|X^I V^I e - \mu e\| \\ &\leq \gamma_2 \left(\|x - x_*\| + \|v^I - v_*^I\|\right) \end{aligned} \tag{4.8}$$

Proof: (4.7) is Lemma 4.1 in [5]. The proof of (4.8) is also along the lines of [5]. Define $H : \mathbb{R}^{n+m+n} \longrightarrow \mathbb{R}^{n+m+n}$ by

$$H(x, \lambda, v) := \begin{pmatrix} \nabla_x \varphi_\mu(x) + A(x)\lambda \\ c(x) \\ X^I V^I e - \mu e \end{pmatrix}.$$

Then $H(x_*, \lambda_*, v_*) = 0$ and $H'(x_*, \lambda_*, v_*)$ is nonsingular. Following the outline in [5] shows the existence of positive constants $\tilde{\gamma}_1$ and $\tilde{\gamma}_2$ such that

$$\begin{aligned} &\tilde{\gamma}_1 \left(\|x - x_*\| + \|\lambda(x) - \lambda_*\| + \|v - v_*\|\right) \\ \leq\ &\|\nabla_x \varphi_\mu(x) + A(x)\lambda(x)\| + \|c(x)\| + \|X^I V^I e - \mu e\| \\ \leq\ &\tilde{\gamma}_2 \left(\|x - x_*\| + \|\lambda(x) - \lambda_*\| + \|v - v_*\|\right) \end{aligned}$$

With the assumptions on the null and range space matrices, the matrix $T(x) := [Z(x)\ Y(x)]$ is nonsingular, and it is

$$T(x)^T \left(\nabla_x \varphi_\mu(x) + A(x)\lambda(x)\right) = \begin{pmatrix} Z(x)^T \nabla \varphi_\mu(x) \\ 0 \end{pmatrix},$$

where we used the special choice of $\lambda(x)$ in (4.1). Thus, using (4.2), it is

$$\|Z(x)^T \nabla \varphi_\mu(x)\| \leq \beta_0 \|\nabla_x \varphi_\mu(x) + A(x)\lambda(x)\|.$$

On the other hand, from (4.3),

$$\begin{aligned} \|Z(x)^T \nabla \varphi_\mu(x)\| &\geq \frac{1}{\|T(x)^{-T}\|} \|\nabla_x \varphi_\mu(x) + A(x)\lambda(x)\| \\ &\geq \frac{1}{\gamma_c} \|\nabla_x \varphi_\mu(x) + A(x)\lambda(x)\|. \end{aligned}$$

Finally recalling (4.4), we see that the claim is valid with $\gamma_1 := \tilde{\gamma}_1 / (\max\{1, \gamma_c\})$ and $\gamma_2 := \left(\tilde{\gamma}_2(1 + \gamma_\lambda)\right) / (\min\{1, 1/\beta_0\})$. □

The following generalization of Lemma 4.2 in [5] shows that the merit function has a strict local minimum at z_*, if the penalty parameter is chosen large enough.

Lemma 4.3 *Suppose, that Assumptions 4.1 hold at* $z_* = (x_*, v_*^I)$, *and that for sufficiently large* k *the penalty parameter* $\nu_k = \bar{\nu}$ *stays constant and is chosen, so that for sufficiently large* k

$$\bar{\nu} \|c(x_k)\| \geq \nabla \varphi_\mu(x_k)^T Y_k p_Y + \rho \|c(x_k)\|. \tag{4.9}$$

Then there exist a neighborhood U of (x_*, v_*^I) and positive constants γ_3 and γ_4, so that

$$\begin{aligned}\gamma_3\left(\|x_k - x_*\|^2 + \|X_k^I V_k^I e - \mu e\|^2\right) &\leq \phi_{\bar\nu}(z_k) - \phi_{\bar\nu}(z_*) \\ &\leq \gamma_4\left(\|Z_k^T \nabla\varphi_\mu(x_k)\|^2 + \|c(x_k)\| + \|X_k^I V_k^I e - \mu e\|^2\right)\end{aligned} \tag{4.10}$$

for (x_k, v_k^I) in U.

Proof: This proof is a generalization of Lemma 3.4 in [2] for the primal-dual ℓ_2-merit function and a tighter update for the penalty parameter (no absolute values in update).

We first show, that for the primal ℓ_2-merit function $\tilde\phi_\nu(x) := \varphi_\mu(x) + \nu\|c(x)\|$ under the above assumptions there exist positive constants $\tilde\gamma_3$ and $\tilde\gamma_4$ such that

$$\tilde\gamma_3\|x_k - x_*\|^2 \leq \tilde\phi_{\bar\nu}(x_k) - \tilde\phi_{\bar\nu}(x_*) \leq \tilde\gamma_4\left(\|Z(x_k)^T\nabla\varphi_\mu(x_k)\|^2 + \|c(x_k)\|\right). \tag{4.11}$$

To show the left inequality in (4.11), we obtain from (4.1) and (2.10)

$$\lambda(x_k)^T c(x_k) = -\nabla\varphi_\mu(x_k)^T Y_k \left[Y_k^T A_k\right]^{-T} c(x_k) = \nabla\varphi_\mu(x_k)^T Y_k p_Y$$

Denoting $\tilde{\mathcal{L}}(x,\lambda) := \varphi_\mu(x) + \lambda^T c(x)$ as the Lagrangian of the *barrier* problem, we can write

$$\begin{aligned}\tilde\phi_{\bar\nu}(x_k) - \tilde\phi_{\bar\nu}(x_*) &= \varphi_\mu(x_k) + \bar\nu\|c(x_k)\| - \varphi_\mu(x_*) \\ &= \varphi_\mu(x_k) - \varphi_\mu(x_*) + \lambda(x_k)^T c(x_k) + \left[\bar\nu\|c(x_k)\| - \nabla\varphi_\mu(x_k)^T Y_k p_Y\right] \\ &\geq \tilde{\mathcal{L}}(x_k,\lambda_*) - \varphi_\mu(x_*) + (\lambda(x_k) - \lambda(x_*))^T c(x_k) + \rho\|c(x_k)\| \\ &\geq \tilde{\mathcal{L}}(x_k,\lambda_*) - \varphi_\mu(x_*) + \frac{\rho}{2}\|c(x_k)\|\end{aligned}$$

where the last inequality is valid by (4.4) for all x_k sufficiently close to x_*. Noting $\varphi_\mu(x_*) = \tilde{\mathcal{L}}(x_*,\lambda_*)$ and $\nabla_x\tilde{\mathcal{L}}(x_*,\lambda_*) = 0$, expanding the last inequality in a Taylor series leads to

$$\begin{aligned}\tilde\phi_{\bar\nu}(x_k) - \tilde\phi_{\bar\nu}(x_*) \geq\ & \frac{1}{2}(x_k - x_*)^T\nabla_{xx}^2\tilde{\mathcal{L}}(x_*,\lambda_*)(x_k - x_*) + \\ & \frac{\rho}{2}\|c(x_k)\| + \frac{\eta}{2}\|c(x_k)\|^2 - \frac{\eta}{2}\|c(x_k)\|^2 + O(\|x_k - x_*\|^3).\end{aligned} \tag{4.12}$$

Now we note that a Taylor expansion of $c(x_k)$ around x_* gives

$$\frac{\eta}{2}\|c(x_k)\|^2 = \frac{\eta}{2}(x_k - x_*)^T A_* A_*^T (x_k - x_*) + O(\|x_k - x_*\|^3). \tag{4.13}$$

In Lemma 4.2 of [5] it is shown that if Assumptions 4.1 are satisfied there exists a sufficiently large value of η such that

$$\frac{1}{2}(x - x_*)^T(\nabla_{xx}^2\tilde{\mathcal{L}}(x_*,\lambda_*) + \eta A_* A_*^T)(x - x_*) \geq 2\tilde\gamma_3\|x - x_*\|^2 \tag{4.14}$$

for all x. Thus, together with (4.12)-(4.14)

$$\begin{aligned}\tilde\phi_{\bar\nu}(x_k) - \tilde\phi_{\bar\nu}(x_*) &\geq 2\tilde\gamma_3\|x_k - x_*\|^2 + \frac{\rho}{2}\|c(x_k)\| - \frac{\eta}{2}\|c(x_k)\|^2 + O(\|x_k - x_*\|^3) \\ &\geq 2\tilde\gamma_3\|x_k - x_*\|^2 + \left(\frac{\rho}{2} - \frac{\eta}{2}\|c(x_k)\|\right)\|c(x_k)\| + O(\|x_k - x_*\|^3) \\ &\geq 2\tilde\gamma_3\|x_k - x_*\|^2 + \frac{\rho}{4}\|c(x_k)\| + O(\|x_k - x_*\|^3) \\ &\geq \tilde\gamma_3\|x_k - x_*\|^2\end{aligned}$$

where the last two inequalities are valid, if x_k is sufficiently close to x_*.

The right inequality in (4.11) follows directly from

$$\begin{aligned}
\tilde\phi_{\bar\nu}(x_k) - \tilde\phi_{\bar\nu}(x_*) &= \varphi_\mu(x_k) + \bar\nu\|c(x_k)\| - \varphi_\mu(x_*) \\
&= \tilde{\mathcal{L}}(x_k,\lambda_k) - \tilde{\mathcal{L}}(x_*,\lambda_*) + [\bar\nu\|c(x_k)\| - \lambda_k^T c(x_k)] \\
&\le O(\|x_k - x_*\|^2) + \rho\|c(x_k)\| \\
&\le \gamma_4[\|Z(x_k)^T\nabla\varphi_\mu(x_k)\|^2 + \|c(x_k)\|]
\end{aligned}$$

for a constant $\gamma_4 > 0$, where we used (4.7), that fact, that $(a+b)^2 \le 3(a^2+b^2)$ for $a, b \ge 0$, and assumed $\|c(x_k)\| \le 1$.

Since $\phi_{\bar\nu}(z) = \tilde\phi_{\bar\nu}(x) + \mathcal{V}_\mu(x,v)$, the proof is done if we show that there are positive constants $\bar\gamma_3$ and $\bar\gamma_4$ such that

$$\bar\gamma_3\|X^I V^I e - \mu e\|^2 \le \mathcal{V}(x,v) - \mathcal{V}(x_*,v_*) \le \bar\gamma_4\|X^I V^I e - \mu e\|^2,$$

for all (x, v^I) sufficiently close to (x_*, v_*^I).

In order to show this, let's examine the strictly convex function $g : \mathbb{R}_+^{|\mathcal{I}|} \longrightarrow \mathbb{R}$, defined as $g(w) := \sum_{i=1}^{|\mathcal{I}|}\left(w^{(i)} - \mu\ln(w^{(i)})\right)$. It has its global minimum at $w_* = \mu e$. Since the Hessian $\nabla^2 g(w_*) = \frac{1}{\mu}I$ is positive definite, a Taylor expansion around w_* shows, that there are positive constants $\bar\gamma_3$ and $\bar\gamma_4$, so that

$$\bar\gamma_3\|w - w_*\|^2 \le g(w) - g(w_*) \le \bar\gamma_4\|w - w_*\|^2$$

for all w close to w_*. Since $\mathcal{V}(x,z) = g(X^I V^I e)$ and $X_*^I V_*^I e = \mu e$, we obtain the desired result. (Note, that "(x, v^I) close to (x_*, v_*^I)" implies "$X^I V^I e$ close to μe".) □

Finally, we can prove convergence to a solution z_*, if the iterates are close enough to this solution. As in [3], it is necessary to make an assumption for the line search:

Assumption 4.2 *The line search has the property that, for all large k, $\phi_{\bar\nu}\left((1-\theta)z_k + \theta z_{k+1}\right) \le \phi_{\bar\nu}(z_k)$, i.e. for a sufficiently large k all succeeding iterates are contained in the connected component of the level set of $\phi_{\bar\nu}(z_k)$.*

Theorem 4.1 *Suppose, that the assumptions of Lemma 4.3 and Assumption 4.2 are valid. Then, if an iterate z_{k_0} is sufficiently close to z_*, then the sequence z_k converges to z_*.*

Proof: We first note, that for $z = (x, v^I)$ with $\|(X^I)^{-1}\| \le C_1$

$$\begin{aligned}
\|v^I - v_*^I\| &\le \|(X^I)^{-1}\|\left(\|X^I V^I e - X^I V_*^I e\|\right) \\
&\le C_1\left(\|X^I V^I e - X_*^I V_*^I e\| + \|(X^I - X_*^I)v_*^I\|\right) \\
&\le C_1\left(\|X^I V^I e - \mu e\| + \|v_*^I\|\|x - x_*\|\right),
\end{aligned}$$

from which (using equivalence of norms)

$$\begin{aligned}
\|z - z_*\| &\le C_2\left(\|v^I - v_*^I\| + \|x - x_*\|\right) \\
&\le C_2 C_1\|X^I V^I e - \mu e\| + C_2\left(C_1\|v_*^I\| + 1\right)\|x - x_*\| \\
&\le C_3\left(\|X^I V^I e - \mu e\| + \|x - x_*\|\right) \\
&\le C_4\left(\|X^I V^I e - \mu e\|^2 + \|x - x_*\|^2\right)^{1/2} \qquad (4.15)
\end{aligned}$$

Now, let U be the neighborhood from Lemma 4.3, which we can without loss of generality assume to be bounded and bounded away from the boundary of the positive orthant, and in addition small enough, so that Lemma 4.2 is valid in U. We then have from (4.10)

$$\|z - z_*\| \leq C_4 \left(\frac{1}{\gamma_3} \left(\phi_\nu(z) - \phi_\nu(z_*) \right) \right)^{1/2}$$

for all $z \in U$. This shows that z_* is a unique minimizer of the merit function in U. Thus, since $\phi_\nu(z)$ is a continuous function of z and since $z_* \in U$, we can find a $\bar{z} \in U$, such that U contains the connected component L of the level set of $\phi_\nu(\bar{z})$ that contains z_*. If now z_{k_0} lies within L, it follows from the line search assumption, that all succeeding iterates are contained in $L \subset U$. It then follows from Theorem 3.2 and (4.8), that z_k converges to z_*. □

5 R-Linear Convergence

For the rest of the paper we assume that the line search satisfies Assumption 4.2. We also assume that the iterates generated by our algorithm converge to a local solution x_* at which Assumptions 4.1 hold. This implies, that the penalty parameter $\nu_k \equiv \bar{\nu}$ for large k.

We now define U to be the set of iterates at which BFGS updating takes place, i.e.

$$U := \{k : B_{k+1} = BFGS(B_k, s_k, y_k)\},$$

and let

$$U_k := U \cap \{1, 2, \ldots, k\}.$$

The number of elements in U_k will be denoted by $|U_k|$.

As in [3] we can show

Theorem 5.1 *Suppose that the iterates $\{x_k\}$ generated by our algorithm converge to a point x_* that satisfies Assumptions 4.1. Then for any $k \in U$ and any $j \geq k$*

$$\|z_j - z_*\| \leq C r^{|U_k|}, \tag{5.1}$$

for some constants $C > 0$ and $0 \leq r < 1$.

Proof: The proof follows the one of Theorem 5.4 in [3] using (3.12) and (4.10) and considering (4.15). □

As in [3], we make use of the matrix function ψ defined by

$$\psi(B) = \mathrm{tr}(B) - \ln(\det(B)), \tag{5.2}$$

where tr denotes the trace, and det the determinant. It can be shown that

$$\ln \mathrm{cond}(B) < \psi(B),$$

for any positive definite matrix B (see [4]). We also make use of the weighted quantities

$$\tilde{y}_k = G_*^{-1/2} y_k, \qquad \tilde{s}_k = G_*^{1/2} s_k, \tag{5.3}$$

$$\tilde{B}_k = G_*^{-1/2} B_k G_*^{-1/2}, \tag{5.4}$$

$$\cos\tilde{\theta}_k = \frac{\tilde{s}_k^T \tilde{B}_k \tilde{s}_k}{\|\tilde{B}_k \tilde{s}_k\| \|\tilde{s}_k\|}, \tag{5.5}$$

and

$$\tilde{q}_k = \frac{\tilde{s}_k^T \tilde{B}_k \tilde{s}_k}{\tilde{s}_k^T \tilde{s}_k}. \tag{5.6}$$

One can show (see Equation (3.22) of [4]) that if B_k is updated by the BFGS formula then

$$\begin{aligned} \psi(\tilde{B}_{k+1}) &= \psi(\tilde{B}_k) + \frac{\|\tilde{y}_k\|^2}{\tilde{y}_k^T \tilde{s}_k} - 1 - \ln\frac{\tilde{y}_k^T \tilde{s}_k}{\tilde{s}_k^T \tilde{s}_k} + \ln\cos^2\tilde{\theta}_k \\ &\quad + \left[1 - \frac{\tilde{q}_k}{\cos^2\tilde{\theta}_k} + \ln\frac{\tilde{q}_k}{\cos^2\tilde{\theta}_k}\right]. \end{aligned} \tag{5.7}$$

This expression characterizes the behavior of the BFGS matrices B_k, and will be crucial to the proof of the following theorem, whose proof is similar to the one in [2].

Theorem 5.2 *Suppose that the iterates $\{z_k\}$ generated by our algorithm converge to a solution point z_* that satisfies Assumptions 4.1. Then $\{\|B_k\|\}$ and $\{\|B_k^{-1}\|\}$ are bounded, and for all $k \in U$*

$$\|(B_k - G_*)p_Z^{(k)}\| = o(\|d_k^x\|). \tag{5.8}$$

Proof: Consider only iterates k for which BFGS updating of B_k takes place. We have from (2.20), (2.9), (4.6), (2.19), and (4.5)

$$\begin{aligned} y_k &= Z_{k+1}^T \nabla f(x_{k+1}) - Z_k^T \nabla f(x_k) \\ &= Z_{k+1}^T \nabla \mathcal{L}(x_{k+1}, \lambda_*) - Z_k^T \nabla \mathcal{L}(x_k, \lambda_*) \\ &= (Z_{k+1} - Z_k)^T \nabla \mathcal{L}(x_{k+1}, \lambda_*) + Z_k^T (\nabla \mathcal{L}(x_{k+1}, \lambda_*) - \nabla \mathcal{L}(x_k, \lambda_*)) \\ &= (Z_{k+1} - Z_k)^T \nabla \mathcal{L}(x_{k+1}, \lambda_*) + \left[Z_k^T \int_0^1 \nabla_{xx}^2 \mathcal{L}(x_k + t\alpha_k d_k, \lambda_*) dt\right] \alpha_k d_k \\ &= O(\sigma_k)(\|s_k\| + \alpha_k \|p_Y\|) + \alpha_k Z_k^T \tilde{W}_k (Z_k p_Z + Y_k p_Y) \\ &= Z_k^T \tilde{W}_k Z_k s_k + \alpha_k Z_k^T \tilde{W}_k Y_k p_Y + O(\sigma_k)(\|s_k\| + \alpha_k \|p_Y\|). \end{aligned} \tag{5.9}$$

Since the at iterate k BFGS updating is performed, the update criterion (2.21) has to be satisfied, giving $\|\alpha_k p_Y\| = O(\gamma_k)\|s_k\|$. Thus, we obtain

$$\begin{aligned} y_k &= \left[Z_k^T \tilde{W}_k Z_k - G_*\right] s_k + G_* s_k + O(\sigma_k + \gamma_k)\|s_k\| \\ &= G_* s_k + O(\sigma_k + \gamma_k)\|s_k\|, \end{aligned}$$

since the term in square brackets is $O(\sigma_k)$. Recalling (5.3) and noting that $\tilde{y}_k^T \tilde{s}_k = y_k^T s_k$ we have

$$\tilde{y}_k^T \tilde{s}_k = \|\tilde{s}_k\|^2 + O(\sigma_k + \gamma_k)\|\tilde{s}_k\|^2,$$

since $\|\tilde{s_k}\|$ and $\|s_k\|$ are of the same order. Therefore

$$\frac{\tilde{y}_k^T \tilde{s}_k}{\|\tilde{s}_k\|^2} = 1 + O(\sigma_k + \gamma_k). \tag{5.10}$$

Similarly,

$$\tilde{y}_k^T \tilde{y}_k = \|\tilde{s}_k\|^2 + O(\sigma_k + \gamma_k)\|\tilde{s}_k\|^2$$

and thus

$$\frac{\|\tilde{y}_k\|^2}{\|\tilde{s}_k\|^2} = 1 + O(\sigma_k + \gamma_k).$$

Therefore

$$\frac{\|\tilde{y}_k\|^2}{\tilde{y}_k^T \tilde{s}_k} = \frac{\|\tilde{y}_k\|^2}{\|\tilde{s}_k\|^2} \frac{\|\tilde{s}_k\|^2}{\tilde{y}_k^T \tilde{s}_k} = 1 + O(\sigma_k + \gamma_k). \tag{5.11}$$

We now consider $\psi(\tilde{B}_{k+1})$ given by (5.7). A simple expansion shows that for large k, $\ln(1 + O(\sigma_k + \gamma_k)) = O(\sigma_k + \gamma_k)$. Using this, (5.10) and (5.11) we have

$$\psi(\tilde{B}_{k+1}) = \psi(\tilde{B}_k) + O(\sigma_k + \gamma_k) + \ln\cos^2\tilde{\theta}_k + \left[1 - \frac{\tilde{q}_k}{\cos^2\tilde{\theta}_k} + \ln\frac{\tilde{q}_k}{\cos^2\tilde{\theta}_k}\right]. \tag{5.12}$$

Note that for $x \geq 0$ the function $1 - x + \ln x$ is non-positive, implying that the term in square brackets is non-positive, and that $\ln\cos^2\tilde{\theta}_k$ is also non-positive. We can therefore delete these terms to obtain

$$\psi(\tilde{B}_{k+1}) \leq \psi(\tilde{B}_k) + O(\sigma_k + \gamma_k). \tag{5.13}$$

Summing over the set of iterates in U_k, using (5.13) , and noting that $B_{j+1} = B_j$ for $j \notin U_k$, we have

$$\psi(\tilde{B}_{k+1}) \leq \psi(\tilde{B}_1) + C_1 \sum_{j \in U_k} \sigma_j + C_2 \sum_{j \in U_k} \gamma_j, \tag{5.14}$$

for some constants C_1, C_2. By (5.1),

$$\begin{aligned} \sum_{j \in U} \sigma_j &\leq C \sum_{j \in U} r^{|U_j|} \\ &= C \sum_{i=1}^{|U|} r^i \\ &< \infty, \end{aligned}$$

and since $\{\gamma_k\}$ is summable we conclude from (5.14) that $\{\psi(\tilde{B}_k)\}$ is bounded above. By (5.2) $\psi(\tilde{B}_k) = \sum_{i=1}^n (l_i - \ln l_i)$, where l_i are the eigenvalues of $\tilde{B}_k$, and it is easy to see that this implies that both $\|B_k\|$ and $\|B_k^{-1}\|$ are bounded.

The rest of the proof can be found in [3]. □

This result immediately implies that the iterates are R-linearly convergent, regardless of how often updating takes place.

Theorem 5.3 *Suppose that the iterates $\{z_k\}$ generated by our algorithm converge to a solution point z_* that satisfies Assumptions 4.1, and that $|U_k| \to \infty$. Then the rate of convergence is at least R-linear.*

Proof: See [3]. □

6 2-Step Superlinear Convergence

In order to show the 2-step superlinear convergence of our primal-dual method we will use the analysis by Nocedal and Overton [6] with some minor modifications which are necessary to deal with non-orthogonal basis matrices as well as a cross term in the computation of the null space component; compare (2.11) with p. 833, l. 36 in [6]. In this section we assume that the line search always accepts full steps $\alpha_k = 1$ for k large enough, that Assumptions 4.1 are valid, and that $z_k \to z_*$. The first part of this discussion will only deal with primal variables, and we define $e_k^x := \|x_k - x_*\|$.

In order to generalize the result in [6] for non-orthogonal matrices we define the change of coordinates matrix

$$T_k := [Y_k \ Z_k] .$$

From (4.3) we have, that $\{T_k^{-1}\}$ is bounded, and we define matrices $\bar{Y}_k \in \mathbb{R}^{n\times m}$ and $\bar{Z}_k \in \mathbb{R}^{n\times(n-m)}$, so that

$$T_k^{-T} = \left[\bar{Y}_k \ \bar{Z}_k\right],$$

which implies the relationships

$$\begin{aligned} \bar{Y}_k^T Y_k &= I \\ \bar{Z}_k^T Z_k &= I \\ \bar{Z}_k^T Y_k &= 0 \\ \bar{Y}_k^T Z_k &= 0 \\ Y_k \bar{Y}_k^T + Z_k \bar{Z}_k^T &= I. \end{aligned}$$

Having this in mind, it is easy to generalize Lemma 4.1 in [6]:

Lemma 6.1 *Assume that for given k, $\|e_k^x\|$ and $\|e_{k-1}^x\|$ are sufficiently small. Then there exists a constant C_0 independent of the value of k, such that*

$$\|\bar{Y}_k^T e_{k+1}^x\| \leq C_0 \|e_k^x\|^2 \tag{6.1}$$

$$\|\bar{Y}_k^T e_k^x\| \leq C_0 \left(\|e_k^x\|^2 + \|e_{k-1}^x\|^2\right) \tag{6.2}$$

$$\|p_Y^{(k)}\| = \|\bar{Y}_k^T (x_{k+1} - x_k)\| \leq C_0 \left(\|e_k^x\|^2 + \|e_{k-1}^x\|^2\right). \tag{6.3}$$

Since Σ_k is bounded from Lemma 3.4, as well as B_k and B_k^{-1} by Theorem 5.2, we know that $H_k = B_k + Z_k^T \Sigma_k Z_k$ and its inverse are bounded. As a consequence, if follows in analogy to Theorem 4.1 in [6]:

Lemma 6.2 *There exists constant C_1, so that for k large enough*

$$\|e_k^x\| \leq C_1 \|e_{k-1}^x\| \tag{6.4}$$

$$\|e_{k+1}^x\| \leq C_1 \left(\|e_{k-1}^x\|^2 + \|(H_k - \tilde{G}_*) p_Z^{(k)}\|\right), \tag{6.5}$$

where we defined $\tilde{G}_* := G_* + \frac{1}{\mu} V_*^2$.

Proof: The proof is very similar to the proof in [6] with the difference that we have to include the correction term into the analysis:

It is

$$\|e_k^x\| \le \|x_k - x_{k-1}\| + \|e_{k-1}^x\| \le \|Y_{k-1}p_Y^{(k-1)} + Z_{k-1}H_{k-1}^{-1}\left(Z_{k-1}^T\nabla\varphi_\mu(x_{k-1}) + w_{k-1}\right)\| + \|e_{k-1}^x\|.$$

Looking at the definition of the cross term and considering the boundedness of Σ_k we see that $w_{k-1} = O(\|p_Y^{(k-1)}\|)$. Since further $p_Y^{(k-1)} = O(\|c_{k-1}\|)$, and $c_* = 0$ and $Z_*^T\nabla\varphi_\mu(x_*) = 0$, we obtain

$$\|e_k^x\| = O(\|c_{k-1}\|) + O(\|Z_{k-1}^T\nabla\varphi_\mu(x_{k-1})\|) = O(\|e_{k-1}^x\|).$$

In order to show (6.5), we follow the proof of Theorem 4.1 (iii) in [6] and note, that for the cross term appearing in the computation of the null space component

$$w_k = Z_k\Sigma_k Y_k p_Y = O(\|p_Y\|) = O(\|e_k^x\|^2 + \|e_{k-1}^x\|^2) = O(\|e_{k-1}^x\|^2),$$

where we used (6.3) and (6.4). □

As a result we obtain the following theorem (see Corollary 4.1 in [6]).

Theorem 6.1 *If*

$$\lim_{k\to\infty}\frac{\|(H_k - \tilde{G}_*)p_Z^{(k)}\|}{\|d_k^x\|} = 0, \tag{6.6}$$

then $x_k \to x_*$ at a two-step Q-superlinear rate.

We can apply this result to finally obtain 2-step Q-superlinear convergence for our algorithm:

Corollary 6.1 *If the iterates z_k generated by our algorithm converge to a point z_* satisfying Assumption 4.1 and $\alpha_k = 1$ for all large k, then $x_k \to x_*$ 2-step Q-superlinearly.*

Proof: It is

$$\begin{aligned}\frac{\|(H_k - \tilde{G}_*)p_Z^{(k)}\|}{\|d_k^x\|} &\le \frac{\|(B_k - G_*)p_Z^{(k)}\|}{\|d_k^x\|} + \frac{\|Z_k^T(\Sigma_k - \frac{1}{\mu}V_*^2)Z_k\bar{Z}_k^T d_k^x\|}{\|d_k^x\|}\\ &\le \frac{\|(B_k - G_*)p_Z^{(k)}\|}{\|d_k^x\|} + O(\Sigma_k - \frac{1}{\mu}V_*^2)\end{aligned}$$

which implies (6.6) since $\lim_{k\to\infty}\Sigma_k = \frac{1}{\mu}V_*^2$, and because the first term converges to zero (Theorem 5.2). □

We can also make a statement about the primal-dual iterates:

Corollary 6.2 *Under the same assumptions as Corollary 6.1 also the sequence $\{z_k\}$ converges to z_* 2-step Q-superlinearly.*

Proof: Since $\{(X_k^I)^{-1}e\}$ is bounded, a Taylor expansion around x_k gives

$$\mu(X_*^I)^{-1}e = \mu(X_k^I)^{-1}e - \mu(X_k^I)^{-2}(x_*^I - x_k^I) + O(\|x_k^I - x_*^I\|^2).$$

It then follows, using (2.7),

$$\begin{aligned}\|v_{k+1}^I - v_*^I\| &= \|\mu(X_k^I)^{-1}e - (X_k^I)^{-1}V_k^I d_k^{x^I} - \mu(X_*^I)^{-1}\|\\ &\le \|-\mu(X_k^I)^{-2}(x_k^I - x_*^I) + (X_k^I)^{-1}V_k^I(x_k^I - x_*^I) - (X_k^I)^{-1}V_k^I(x_k^I + d_k^{x^I} - x_*^I)\| +\\ &\qquad O(\|x_k - x_*\|^2)\\ &= \|(X_k^I)^{-1}(V_k^I - \mu(X_k^I)^{-1})(x_k^I - x_*^I) - (X_k^I)^{-1}V_k^I(x_{k+1}^I - x_*^I)\| + O(\|x_k - x_*\|^2)\\ &\le O(\|z_k - z_*\|)\|x_k - x_*\| + O(\|x_{k+1} - x_*\|) + O(\|x_k - x_*\|^2)\end{aligned}$$

From (6.4) we have $\|x_k - x_*\| = O(\|x_{k-1} - x_*\|)$ and thus, also using Corollary 6.1,

$$\|v_{k+1}^I - v_*^I\| = o(\|x_{k-1} - x_*\|). \tag{6.7}$$

The equivalence of norms gives

$$\|z - z_*\| \leq C \left(\|x - x_*\| + \|v^I - v_*^I\|\right)$$

for a positive constant C, so that

$$\begin{aligned} \frac{\|z_{k+1} - z_*\|}{\|z_{k-1} - z_*\|} &\leq \frac{\|z_{k+1} - z_*\|}{\|x_{k-1} - x_*\|} \\ &\leq C \left(\frac{\|v_{k+1}^I - v_*^I\|}{\|x_{k-1} - x_*\|} + \frac{\|x_{k+1} - x_*\|}{\|x_{k-1} - x_*\|} \right), \end{aligned}$$

which together with (6.7) and Corollary 6.1 proves the claim. □

Remark:

1-step superlinear convergence can be shown, if finite differences (as described in [3]) are used to approximate the cross terms w_k.

References

[1] P. Armand, J.-Ch. Gilbert, and S. Jan-Jégou. A feasible BFGS interior point algorithm for solving strongly convex minimization problems. Technical report, December 1999.

[2] L. T. Biegler. Convergence analysis for a multiplier-free reduced Hessian method. Technical Report 06–203–95, EDRC, Carnegie Mellon University, Pittsburgh, PA, USA, 1995. Revised 1999.

[3] L. T. Biegler, J. Nocedal, and C. Schmid. A reduced Hessian method for large-scale constrained optimization. *SIAM Journal on Optimization*, 5(2):314–347, 1995.

[4] R. H. Byrd and J. Nocedal. A tool for the analysis of quasi-Newton methods with application to unconstrainted minimization. *SIAM Journal on Numerical Analysis*, 26(3):727–739, 1989.

[5] R. H. Byrd and J. Nocedal. An analysis of reduced Hessian methods for constrained optimization. *Mathematical Programming*, 49:285–323, 1991.

[6] J. Nocedal and M. Overton. Projected Hessian updating algorithms for nonlinearly constrained optimization. *SIAM Journal on Numerical Analysis*, 22:821–850, 1985.