Convergence analysis of a primal-dual interior-point method for nonlinear programming *

Igor Griva † David F. Shanno ‡ Robert J. Vanderbei §

July 21, 2004

Abstract

We analyze a primal-dual interior-point method for nonlinear programming. We prove the global convergence for a wide class of problems under the standard assumptions on the problem.

Keywords. Interior-point method, primal-dual, convergence analysis.

1 Introduction

The primal-dual interior-point algorithm implemented in LOQO proved to be very efficient for solving nonlinear optimization problems ([1, 2, 3, 10, 13]). The algorithm applies Newton's method to the perturbed Karush-Kuhn-Tucker system of equations on each step to find the next primal-dual approximation of the solution. The original algorithm [13] implemented in LOQO at each step minimized a penalty barrier merit function to attempt to ensure that the algorithm converged, and to a local minimum rather than

*Research of the first and the third authors supported by NSF grant DMS-9870317 and ONR grant N00014-98-1-0036. Research of the second and author supported by NSF grant DMS-0107450

†George Mason University, Department of Mathematical Sciences, Fairfax, VA 22030 (Corresponding author)

‡Rutgers University, RUTCOR, New Brunswick, NJ 08903

§Princeton University, Department of ORFE, Princeton NJ 08544

any other first order optimality point such as a maximum or a saddle point. A more recent version of LOQO [2] utilizes a memoryless filter to attempt to achieve the same goal. Neither method has been proven convergent under general conditions. In this paper, we analyze the global convergence to a first order optimality point for a general algorithm combining features of the previously mentioned versions of LOQO. This is done under assumptions made only on the problem under consideration, rather than assumptions about the performance of the algorithm. We do not assume that the sequence of primal variables or the Lagrange multipliers remain bounded, two assumptions that appear in many convergence analyses (see e.g. [11] et al.) The algorithm studied here is theoretical. Its implementation in the LOQO framework remains for future work.

2 Problem formulation

The paper considers a method for solving the following optimization problem

$$\begin{aligned} &\min f(x), \\ &\text{s.t. } x \in \Omega, \end{aligned} \tag{1}$$

where the feasible set is defined as $\Omega = \{x \in \mathbb{R}^n : h(x) \geq 0\}$, and $h(x) = (h_1(x), \ldots, h_m(x))$ is a vector function. We assume that $f : \mathbb{R}^n \to \mathbb{R}^1$ and all $h_i : \mathbb{R}^n \to \mathbb{R}^1$, $i = 1, \ldots, m$ are twice continuously differentiable functions. To simplify the presentation we do not consider the equality constraints in this paper. This will be done in the subsequent paper.

After adding nonnegative slack variables $w = (w_1, \ldots, w_m)$, we obtain an equivalent formulation of the problem (1):

$$\begin{aligned} &\min f(x), \\ &\text{s.t. } h(x) - w = 0, \\ &w \geq 0. \end{aligned} \tag{2}$$

The interior-point method places the slacks in a barrier term leading to the following problem

$$\min f(x) - \mu \sum_{i=1}^{m} \log w_i,$$
$$\text{s.t. } h(x) - w = 0, \tag{3}$$

where $\mu > 0$ is a barrier parameter. The solution to this problem satisfies the following primal-dual system

$$\begin{aligned} \nabla f(x) - A(x)^T y &= 0, \\ -\mu e + WYe &= 0, \\ h(x) - w &= 0, \end{aligned} \tag{4}$$

where $y = (y_1, \ldots, y_m)$ is a vector of the Lagrange multipliers or dual variables for problem (3), $A(x)$ is the Jacobian of vector function $h(x)$, Y and W are diagonal matrices with elements y_i and w_i respectively and $e = (1, \ldots, 1) \in \mathbb{R}^m$.

3 Assumptions

We endow $\mathbb{R}^n$ with the l^∞ norm $\|x\| = \max_{1 \le i \le n} |x_i|$, and we endow the space $\mathbb{R}^{m,n}$ with the associated operator norm $\|Q\| = \max_{1 \le i \le m} \left(\sum_{j=1}^{n} |q_{ij}| \right)$.

We invoke the following assumptions throughout the paper.

A1. The objective function is bounded from below: $f(x) \ge \bar{f}$ *for all* $x \in \mathbb{R}^n$.

A2. Slater's condition holds: there exists $\bar{x} \in \mathbb{R}^n$ *such that* $h_i(\bar{x}) > 0$, $i = 1, \ldots, m$.

A3. The constraints $h_i(x)$ *satisfy the following conditions*

$$\lim_{\|x\| \to \infty} \min_{1 \le i \le m} h_i(x) = -\infty. \tag{5}$$

and

$$\sqrt{\log \left(\left| \max_{1 \le i \le m} h_i(x) \right| + 1 \right)} \le - \min_{1 \le i \le m} h_i(x) + C \tag{6}$$

for all $x \in \mathbb{R}^n$, *where* $0 < C < \infty$ *depends only on the problem's data.*

A4. The minima (local and global) of problem (1) satisfy the standard second order optimality conditions.

A5. For each $\mu > 0$ the minima (local and global) of problem (3), satisfy the standard second order optimality conditions.

A6. Hessians $\nabla^2 f(x)$ and $\nabla^2 h_i(x)$, $i = 1, \ldots, m$ satisfy Lipschitz conditions on $\mathbb{R}^n$.

Several comments about the assumptions: assumption (A1) does not restrict the generality. In fact, one can always transform function $f(x)$ using monotone increasing transformation $f(x) := \log(1 + e^{f(x)})$, which is bounded from below.

Assumption (A3) not only implies that the feasible set Ω is bounded, but also implies some growth conditions for the functions $h_i(x)$. In fact, it tells us that there is no function $h_{i_0}(x)$ that grows significantly faster than some other functions $h_i(x)$, $i \neq i_0$, decrease on any unbounded sequence. Most practical problems, including problems with linear and quadratic constraints, convex problems (when functions $h_i(x)$ are concave), nonconvex quadratic and many others satisfy assumption (A3).

The cases when functions $h_i(x)$ do not satisfy assumption (A3) normally involve exponentially fast growing functions $h_i(x)$. Let's consider the following example. The feasible set $\Omega_1 = [-1, 1] \subset \mathbb{R}^1$ can be defined as follows: $h_1(x) = x + 1 \geq 0$ and $h_2(x) = 1 - x \geq 0$. In this case functions $h_1(x)$ and $h_2(x)$ satisfy assumption (A3). However, the same set Ω_1 can be defined differently: $h_1(x) = e^x - e^{-1} \geq 0$ and $h_2(x) = e^{-x} - e^{-1} \geq 0$. In this case, for example, if x increases unboundedly function $h_1(x)$ grows exponentially, but function $h_2(x)$ stays always bounded from below and does not decrease fast enough. Therefore functions $h_1(x)$ and $h_2(x)$ do not satisfy assumption (A3). We believe that failure of a problem to satisfy assumption (A3) is usually a case of bad modeling and thus argue that this assumption does not greatly restrict the generality. The assumption is critical for the convergence analysis because the interior-point algorithm decreases a value of a penalty-barrier merit function and we need assumption (A3) to ensure that the merit function has bounded level sets.

All the assumptions (A1)-(A6) are imposed on the problem, not on the sequence generated by the algorithm. The following lemma follows from the assumptions.

Lemma 1 *Under assumptions (A1)-(A3) a global solution $(x(\mu), w(\mu), y(\mu))$ to the problem (3) exists for any $\mu > 0$.*

Proof. Problem (3) is equivalent to the following problem:

$$\begin{aligned}&\min B(x,\mu)\\&x \in \mathbb{R}^n,\end{aligned}$$

where $B(x,\mu) = f(x) - \mu \sum_{i=1}^m \log h_i(x)$. It follows from assumption (A3) that the feasible set Ω is bounded. Let $\bar{x}$ be the point that exists by assumption (A2) and a constant $M_\mu = 2B(\bar{x},\mu)$. It is easy to show that the set $\Omega_\mu = \{x \in \Omega : B(x,\mu) \leq M_\mu\}$ is a closed bounded set. Therefore due to continuity of $B(x,\mu)$ there exists a global minimizer x_μ such that $B(x,\mu) \geq B(x_\mu,\mu)$ on the set Ω_μ and consequently on the feasible set Ω. Lemma 1 is proven.

4 Interior-point algorithm

In the following we use the following notations.

$$p = (x,w), \quad z = (p,y) = (x,w,y),$$

$$\begin{aligned}\sigma &= \nabla f(x) - A(x)^T y,\\ \gamma &= \mu W^{-1} e - y,\\ \rho &= w - h(x).\\ b(z) &= (\sigma^T, WYe^T, -\rho^T)^T,\\ b_\mu(z) &= (\sigma^T, WYe^T - \mu e^T, -\rho^T)^T,\end{aligned}$$

To control the convergence we need the following merit functions:

$$\nu(z) = \|b(z)\| = \max\{\|\sigma\|, \|\rho\|, \|WYe\|\}$$

$$\nu_\mu(z) = \|b_\mu(z)\| = \max\{\|\sigma\|, \|\rho\|, \|W\gamma\|\},$$

$$\mathcal{L}_{\beta,\mu}(z) = f(x) - \mu \sum_{i=1}^m \log w_i + y^T\rho + \frac{\beta}{2}\rho^T\rho.$$

The function $\nu(z)$ measures the distance between the current approximation and a KKT point of the problem (1). The function $\nu_\mu(z)$ measures the distance between the current approximation and a KKT point of the barrier problem (3). The penalty-barrier function $\mathcal{L}_{\beta,\mu}(z)$ is the augmented Lagrangian for the barrier problem (3). The primal direction decreases the value of $\mathcal{L}_{\beta,\mu}(z)$, which makes the algorithm descend to a minimum rather than another first order optimality point.

Newton's method applied to the system (4) leads to the following linear system for the Newton directions

$$\begin{bmatrix} H(x,y) & 0 & -A(x)^T \\ 0 & Y & W \\ A(x) & -I & 0 \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta w \\ \Delta y \end{bmatrix} = \begin{bmatrix} -\nabla f(x) + A(x)^T y \\ \mu e - WYe \\ -h(x) + w \end{bmatrix}, \qquad (7)$$

where $H(x,y)$ is the Hessian of the Lagrangian of problem (1). Using the notations introduced at the beginning of this section, the system (7) can be rewritten as

$$D(z)\Delta z = -b_\mu(z),$$

where

$$D(z) = \begin{bmatrix} H(x,y) & 0 & -A(x)^T \\ 0 & Y & W \\ A(x) & -I & 0 \end{bmatrix}.$$

After eliminating Δw from this system we obtain the following reduced system

$$\begin{bmatrix} -H(x,y) & A(x)^T \\ A(x) & WY^{-1} \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta y \end{bmatrix} = \begin{bmatrix} \sigma \\ \rho + WY^{-1}\gamma \end{bmatrix}. \qquad (8)$$

After finding Δy, we can obtain Δw by the following formula

$$\Delta w = WY^{-1}(\gamma - \Delta y).$$

The explicit formulas for the solution to the primal-dual system are given in [13] (Theorem 1):

$$\begin{aligned} \Delta x &= N^{-1}\left(-\sigma + A^T(W^{-1}Y\rho + \gamma)\right) \\ \Delta w &= -\rho - A\Delta x \\ \Delta y &= \gamma + W^{-1}Y(\rho - A\Delta x) \end{aligned} \qquad (9)$$

where $N(x,y,w) = H(x,y) + A(x)^T W^{-1} Y A(x)$.

If matrix $N(x,w,y)$ is not positive definite the algorithm replaces it with the regularized matrix

$$\hat{N}(x,w,y) = N(x,w,y) + \lambda I, \quad \lambda \geq 0, \qquad (10)$$

where I is the identity matrix in $\mathbb{R}^{n,n}$ to guarantee that mineigenvalue of $\hat{N}$ is greater than some $\lambda_0 > 0$. Parameter λ is chosen big enough to guarantee that $\hat{N}(x, w, y)$ is positive definite.

Together with the primal regularization we consider also the dual regularization of system (7)

$$\begin{bmatrix} H(x,y) & 0 & -A(x)^T \\ 0 & Y & W \\ A(x) & -I & \epsilon I \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta w \\ \Delta y \end{bmatrix} = \begin{bmatrix} -\nabla f(x) + A(x)^T y \\ \mu e - WYe \\ -h(x) + w \end{bmatrix}, \tag{11}$$

where $\epsilon > 0$ is a regularizing parameter. Clearly, for $\epsilon = 0$ the system is the original one. Using the notations introduced at the beginning of this section, we can rewrite (11) as follows

$$D_\epsilon(z)\Delta z = -b_\mu(z),$$

where

$$D_\epsilon(z) = \begin{bmatrix} H(x,y) & 0 & -A(x)^T \\ 0 & Y & W \\ A(x) & -I & \epsilon I \end{bmatrix}.$$

The explicit formulas for finding primal and dual directions are similar to (9)

$$\begin{array}{rcl} \Delta x & = & N_\epsilon^{-1}\left(-\sigma + A^T\left[WY^{-1} + \epsilon I\right]^{-1}(\rho + WY^{-1}\gamma)\right), \\ \Delta y & = & \left[WY^{-1} + \epsilon I\right]^{-1}(\rho + WY^{-1}\gamma - A\Delta x), \\ \Delta w & = & -\rho - A\Delta x - \epsilon\Delta y, \end{array} \tag{12}$$

where $N_\epsilon(x, y, w) = H(x, y) + A(x)^T\left[WY^{-1} + \epsilon I\right]^{-1} A(x)$. Again, if the matrix $N_\epsilon(x, w, y)$ is not positive definite the algorithm replaces it with the regularized matrix

$$\hat{N}_\epsilon(x, w, y) = N_\epsilon(x, w, y) + \lambda I, \quad \lambda \geq 0, \tag{13}$$

where I is the identity matrix in $\mathbb{R}^{n,n}$ to guarantee that mineigenvalue of $\hat{N}_\epsilon$ is greater than some $\lambda_0 > 0$.

As it will be shown later the primal and the dual regularizations ensure that the primal directions is descent for the penalty-barrier merit function.

One pure step of the IPM algorithm $(x, w, y) \to (\hat{x}, \hat{w}, \hat{y})$ is as follows

$$\hat{x} = x + \alpha_p \Delta x, \tag{14}$$

$$\hat{w} = w + \alpha_p \Delta w, \tag{15}$$

$$\hat{y} = y + \alpha_d \Delta y, \tag{16}$$

where α_p and α_d are primal and dual steplengths. The primal and dual steplengths are chosen to keep slack and dual variables strictly positive:

$$\alpha_p = \min\left\{1; -\kappa \frac{w_i}{\Delta w_i} : \Delta w_i < 0\right\}, \tag{17}$$

$$\alpha_d = \min\left\{1; -\kappa \frac{y_i}{\Delta y_i} : \Delta y_i < 0\right\}, \tag{18}$$

where $0 < \kappa < 1$.

As we show later the pure interior point method converges to the primal-dual solution only locally in the neighborhood of the solution. However, far away from the solution the algorithm does not update dual variables at each step and often uses only primal direction $(\Delta x, \Delta w)$ to find the next approximation.

Let us describe the algorithm in more detail. The algorithm starts each iteration by computing the merit function $\nu(z)$, the barrier parameter μ by the following formula

$$\mu := \min\{\delta\nu(z),\, \nu(z)^2\}, \tag{19}$$

the merit function $\nu_\mu(z)$ and the dual regularization parameter

$$\epsilon = \min\{10^{-2}, \nu_\mu(z)\}. \tag{20}$$

Then the algorithm solves the primal-dual system (11) for the primal-dual Newton directions $(\Delta x, \Delta w, \Delta y)$. To solve the system (11) the algorithm uses a sparse Cholesky factorization developed in [12]. It is possible that while performing the factorization the algorithm learns that matrix $N_\epsilon(x, w, y)$ is not positive definite. It this case the algorithm regularizes the matrix $N_\epsilon(x, w, y)$ by formula (13) and begins the factorization again. It keeps increasing the parameter λ in formula (13) until a positive definite factorization is completed.

The algorithm then selects primal and dual steplengths α_p and α_d by formulas (17)-(18) for the parameter κ chosen by formula

$$\kappa = \max\{0.95,\, 1 - \nu(z)\} \tag{21}$$

and finds the next primal-dual candidate $\hat{x} := x + \alpha_p \Delta x$, $\hat{w} := w + \alpha_p \Delta w$ and $\hat{y} := y + \alpha_d \Delta y$. If the candidate $\hat{z} = (\hat{x}, \hat{w}, \hat{y})$ does not reduce the value of merit function $\nu(z)$, by a chosen *a priori* desired factor $0 < q < 1$ then $\hat{z}$ fails the test and is not longer a candidate. Otherwise, the algorithm begins solving the primal-dual system for Newton directions $\Delta\hat{z}$ at new approximation $\hat{z}$. If the factorization detects that the matrix $N_\epsilon(\hat{x}, \hat{w}, \hat{y})$ is positive definite, then the candidate $\hat{z}$ passes the final test and becomes the next primal-dual approximation and the Newton direction $\Delta\hat{z}$ is used for the next step. However if the matrix $N_\epsilon(\hat{x}, \hat{w}, \hat{y})$ is not positive definite then the candidate $\hat{z}$ fails the final test. In any case when the candidate fails either test, the algorithm does not change the dual approximation y and uses the primal direction $\Delta p = (\Delta x, \Delta w)$ to find the next primal approximation. It will be proven later that the primal direction $\Delta p = (\Delta x, \Delta w)$ is a descent direction for the merit function $\mathcal{L}_{\beta,\mu}(z)$. The primal steplength α_p is backtracked to satisfy the Armijo rule

$$\mathcal{L}_{\beta,\mu}(p + \alpha_p \Delta p, y) - \mathcal{L}_{\beta,\mu}(p, y) \leq \eta\alpha_p < \nabla_p \mathcal{L}_{\beta,\mu}(p, y), \Delta p >, \tag{22}$$

where $0 < \eta < 1$.

The convergence analysis of the algorithm shows that under the assumptions (A1)-(A6) in the neighborhood of the solution the candidate $\hat{z}$ never fails the tests (Lemma 8) and the algorithm always uses the primal-dual direction Δz to find the next approximation. On the other hand to ensure convergence, the algorithm changes dual variables y only when the next dual approximation $\hat{y}$ is closer to the dual solution either to the original problem (1) or to the barrier problem (3). The motivation for such careful treatment of the dual variables lies in the fact that in nonlinear programming the computational work needed to obtain a better dual approximation y generally speaking requires solving a minimization problem unless the primal-dual approximation is close to the primal-dual solution. Therefore changing the dual variables on each step can result in divergence of the algorithm. If the algorithm reaches the unconstrained minimum $\hat{p}$ of the merit function $\mathcal{L}_{\beta,\mu}(p, z)$ it then changes the dual variables by the formula $\hat{y} := y + \beta\rho(\hat{x}, \hat{w})$ to obtain a better dual approximation.

It is appropriate to say several words about the choice of the dual regularization parameter ϵ and the penalty parameter β. These parameters are chosen to satisfy two conditions: a) the primal Newton direction $(\Delta x, \Delta w)$ must be descent for the merit function $\mathcal{L}_{\beta,\mu}(z)$ and b) the regularization parameter $\epsilon > 0$ must become zero when the trajectory of the algorithm approaches the primal-dual solution.

To prove the convergence of the algorithm we use the following choice of the parameters at each iteration: $\epsilon = \nu_\mu(z)$, $\beta = 1/\epsilon$. It will be shown later that such choice of the parameters satisfies the conditions (a) and (b) and allows us to prove convergence of the algorithm.

The formal description of the algorithm is in Figure 1.

Step 1: Initialization:
An initial primal-dual approximation $z^0 = (p^0, y^0) = (x^0, w^0, y^0)$ is given.
An accuracy $\varepsilon > 0$, initial penalty parameter $\beta_0 \geq 2m\mu$ are given.
Parameters $0 < \eta < 0.5$, $0 < \delta < q < 1$, $\tau > 0$, $\theta > 0$ are given.
Set $z := z^0$, $r := \nu(z^0)$, $\mu := \min\{\delta r, r^2\}$, $r_\mu = \nu_\mu(z^0)$, $\beta := \beta_0$, $\epsilon := \min\{\nu_\mu(z^0), \frac{1}{\beta}\}$, $s := 0$.
Step 2: If $r \leq \varepsilon$, Stop, **Output**: z.
Step 3: Factorize the system, Increase λ until success.
Find direction: $\Delta z := PrimalDualDirection(z, \epsilon)$.
Set $s := s + 1$.
Set $\kappa := \max\{0.95,\ 1 - r\}$.
Choose primal and dual steplengths: α_p and α_d by the formulas (17)-(18).
Set $\hat{p} := p + \alpha_p \Delta p$, $\hat{y} := y + \alpha_d \Delta y$.
Step 4: If $\nu(\hat{z}) \leq qr$, Goto Step 10.
Step 5: Set $\beta = \max\{\beta, 1/\epsilon\}$.
Backtrack α_p until $\mathcal{L}_{\beta,\mu}(p + \alpha_p \Delta p, y) - \mathcal{L}_{\beta,\mu}(p, y) \leq \eta \alpha_p < \nabla_p \mathcal{L}_{\beta,\mu}(p, y), \Delta p >$
Set $\hat{p} := p + \alpha_p \Delta p$.
Step 6: If $\|\nabla_p \mathcal{L}_{\beta,\mu}(\hat{p}, y)\| > \min\left\{\tau \|\rho(\hat{p})\|,\ 1/(2\beta s^3),\right\}$, or $y + \beta\rho(\hat{p}) \not\geq 0$, Goto Step 3.
Step 7: $\hat{y} := y + \beta\rho(\hat{p})$.
If $\nu_\mu(\hat{z}) > qr_\mu$, Set $p := \hat{p}$, $\beta := 2\beta$, Goto Step 3.
Step 8: If $\nu(\hat{z}) \leq qr$, Set $r := \nu(\hat{z})$, $\mu := \min\{\delta r, r^2\}$.
Set $z := \hat{z}$. $r_\mu := \nu_\mu(\hat{z})$, $\epsilon := \min\{\nu_\mu(z), \frac{1}{\beta}\}$.
Step 9: Goto Step 3.
Step 10: Factorize the system in $\hat{z}$ with $\lambda = 0$, If $N(\hat{z})$ is not p.d., Goto step 5.
Step 11: Set $z := \hat{z}$, $r := \nu(\hat{z})$, $\mu := \min\{\delta r, r^2\}$, $\epsilon := \min\{\nu_\mu(z), \frac{1}{\beta}\}$, $r_\mu := \nu_\mu(\hat{z})$.
If $r \leq \varepsilon$, Stop, **Output**: z.
Step 12: Find direction: $\Delta z := PrimalDualDirection(\hat{z}, \epsilon)$.
Set $s := s + 1$.
Choose primal and dual steplengths: α_p and α_d by the formulas (17)-(18).
Set $\hat{p} := p + \alpha_p \Delta p$, $\hat{y} := y + \alpha_d \Delta y$.
Goto Step 4.

Figure 1: IPM algorithm.

5 Convergence analysis

We need the following auxiliary lemmas for the convergence analysis.

Lemma 2 *1) For any $y \in \mathbb{R}^m$, $\beta \geq 2m\mu$ and $\mu > 0$, there exists a global minimum*

$$S_{\beta,\mu}(y) = \min_{x \in \mathbb{R}^n,\, w \in \mathbb{R}^m_{++}} \mathcal{L}_{\beta,\mu}(x, w, y) > -\infty. \qquad (23)$$

Proof. 1) Let us fix any $\bar{w} \in \mathbb{R}^m_{++}$ and set $M = 2\mathcal{L}_{\beta,\mu}(\bar{x}, \bar{w}, y)$, where $\bar{x}$ exists by assumption (A2). The function $\mathcal{L}_{\beta,\mu}(x, w, y)$ is continuous on $\mathbb{R}^n \times \mathbb{R}^m_{++}$ therefore to prove the lemma it is enough to show the following set

$$\mathcal{R}_\beta = \Big\{(x, w) \in \mathbb{R}^n \times \mathbb{R}^n_{++} : \mathcal{L}_{\beta,\mu}(x, w, y) \leq M\Big\}$$

is a bounded and closed set.

First we show that the set $\mathcal{R}_\beta$ is bounded. Let us assume that $\mathcal{R}_\beta$ is unbounded. Then there exists an unbounded sequence $\{p^l\} = \{(x^l, w^l)\}$ defined on $\mathbb{R}^n \times \mathbb{R}^m_{++}$ such that

(a) $x^0 = \bar{x}$, $w^0 = \bar{w}$,

(b) $\lim_{l\to\infty} \|p^l - p^0\| = \infty$,

(c) $\lim_{l\to\infty} \mathcal{L}_{\beta,\mu}(x^l, w^l, y) \leq M$.

We are going to show that for any sequence satisfying (a) and (b) we have

$$\lim_{l\to\infty} \mathcal{L}_{\beta,\mu}(x^l, w^l, y) = \infty, \tag{24}$$

which contradicts (c).

Let $P = \{p^l\} = \{(x^l, w^l)\}$ be a sequence satisfying conditions (a) and (b). Let us introduce sequences $\{\rho_i^l\} = \{w_i^l - h_i(x^l)\}$ and $\{\varphi_i^l\} = \{\frac{\beta}{2}{\rho_i^l}^2 + y_i\rho_i^l - \mu\log(h_i(x^l) + \rho_i^l)\}$, $i = 1, \ldots, m$. Since $f(x)$ is bounded from below, to prove (24) it is enough to show that

$$\lim_{l\to\infty} \sum_{i=1}^m \varphi_i^l = \infty. \tag{25}$$

Let us first consider the simpler case when the sequence $\{x^l\}$ corresponding to the sequence P is bounded. In this case, the corresponding sequence $\{w^l\}$ is unbounded. We can assume that there exists a nonempty index set of constraints I_+ such that for any index $i \in I_+$ we have $\lim_{l\to\infty} w_i^l = \infty$ (otherwise we consider the correspondent subsequences). Since for any index $i = 1, \ldots, m$ a sequence $\{h_i(x^l)\}$ is bounded, we have $\lim_{l\to\infty} \rho_i(t) = \infty$ for $i \in I_+$, and hence

$$\lim_{l\to\infty} \varphi_i^l = \lim_{l\to\infty} \frac{\beta}{2}{\rho_i^l}^2 + y_i\rho_i^l - \mu\log(h_i(x^l) + \rho_i^l)) = \infty, \quad i \in I_+,$$

and (25) holds true.

Now we study the case when the sequence $S = \{x^l\}$ corresponding to the sequence P is unbounded. Let us first estimate separately φ_i^l for any $1 \le i \le m$. In case $h_i(x^l) \le 1$, then

$$\varphi_i^l = \frac{\beta}{2}{\rho_i^l}^2 + y_i\rho_i^l - \mu\log(h_i(x^l) + \rho_i^l) \ge \frac{\beta}{2}{\rho_i^l}^2 + y_i\rho_i^l - \mu\log(1 + \rho_i^l) \ge -B_1 \quad (26)$$

for some $B_1 \ge 0$ large enough.

If $h_i(x^l) \ge 1$ then, keeping in mind that $h_i(x^l) + \rho_i^l > 0$, we have

$$\varphi_i^l = \frac{\beta}{2}{\rho_i^l}^2 + y_i\rho_i^l - \mu\log(h_i(x^l) + \rho_i^l)$$

$$= \frac{\beta}{2}{\rho_i^l}^2 + y_i\rho_i^l - \mu\log h_i(x^l) - \mu\log\left(1 + \frac{\rho_i^l}{h_i(x^l)}\right)$$

$$\ge \frac{\beta}{2}{\rho_i^l}^2 + y_i\rho_i^l - \mu\log h_i(x^l) - \mu - \mu\frac{\rho_i^l}{h_i(x^l)}$$

$$\ge \frac{\beta}{2}|\rho_i^l|^2 - |y_i||\rho_i^l| - \mu\log h_i(x^l) - \mu - \mu|\rho_i^l| \ge -\mu\log h_i(x^l) - B_2,$$

where B_2 is large enough. Invoking inequality (6) we obtain

$$-\mu\log h_i(x^l) - B_2 \ge -\mu\log\left(\max_{1\le i\le m} h_i(x^l)\right) - B_2$$

$$\ge -\mu\log\left(\left|\max_{1\le i\le m} h_i(x^l)\right| + 1\right) - B_2 \ge -\mu(C - \min_{1\le i\le m} h_i(x^l))^2 - B_2$$

$$= -\mu(C - h_{i_0}(x^l))^2 - B_2 \ge \begin{cases} -\mu C^2 - B_2, & \text{if} \quad h_{i_0}(x^l) \ge 0 \\ -\mu(C + \rho_{i_0}^l)^2 - B_2, & \text{if} \quad h_{i_0}(x^l) < 0 \end{cases} \ge$$

$$-\mu\max\left\{C^2, (C + \rho_{i_0}^l)^2\right\} - B_2,$$

where $i_0 = i_0(x) \in \text{Argmin}_{1\le i\le m} h_i(x^l)$. It follows from (5) and unboundedness of the sequence $\{x^l\}$ that

$$\lim_{\|x\|\to\infty} \rho_{i_0(x)}(x) = +\infty.$$

Hence for all sequence numbers l large enough (that $h_{i_0}(x^l) < 0$) we have

$$\varphi_i^l \ge -\mu(C + \rho_{i_0}^l)^2 - B_2 \quad (27)$$

Combining (26) and (27), we obtain for l large enough (that $h_{i_0}(x^l) < 0$)

$$\sum_{i=1}^{m} \varphi_i^l = \varphi_{i_0}^l + \sum_{i \neq i_0 : h_i(x^l) < 1} \varphi_i^l + \sum_{i : h_i(x^l) \geq 1} \varphi_i^l$$

$$\geq \frac{\beta}{2} {\rho_{i_0}^l}^2 + y_i \rho_{i_0}^l - \mu \log \rho_{i_0}^l - mB_1 - \Big(\mu(C + \rho_{i_0}^l)^2 + B_2\Big)\, m$$

The inequality $\beta > 2\mu m$ guarantees that for such β condition (25) holds. Thus, condition (24) also holds, and we have the contradiction. Therefore the set $\mathcal{R}_\beta$ is bounded.

It is easy to see that the set $\mathcal{R}_\beta$ is closed. Therefore $\mathcal{L}_{\beta,\mu}(x^l, w^l, y)$ reaches its global minimum on $\mathbb{R}^n \times \mathbb{R}^m_{++}$.

Lemma 2 is proven.

Remark 1 *Following the proof of Lemma 2 we can show that there exists a global minimum*

$$S_\infty = \min_{x \in \mathbb{R}^n,\, w \in \mathbb{R}^m_+} \|\rho(x, w)\|^2 > -\infty. \tag{28}$$

and that any set

$$\mathcal{R}_\infty = \Big\{(x, w) \in \mathbb{R}^n \times \mathbb{R}^n_+ : \|\rho(x, w)\|^2 \leq M\Big\}$$

is bounded.

Lemma 3 *For any $\beta > 0$, there exists $\alpha > 0$ such that for any primal-dual approximation (x, w, y) such that $w \in \mathbb{R}^m_{++}$, $y \in \mathbb{R}^m_{++}$, the primal direction $\Delta p = (\Delta x, \Delta w)$, obtained as the solution of the system (11) with the primal regularization rule (13) and the dual regularization parameter $\epsilon = \frac{1}{\beta}$, is a descent direction for $\mathcal{L}_{\beta,\mu}(p, y)$ and*

$$(\nabla_p \mathcal{L}_{\beta,\mu}(p, y), \Delta p) \leq -\alpha \|\Delta p\|^2$$

Proof. For the regularization parameter $\epsilon = 1/\beta$, the primal-dual system (11) is as follows

$$\begin{bmatrix} H(x,y) & 0 & -A(x)^T \\ 0 & Y & W \\ A(x) & -I & \frac{1}{\beta} I \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta w \\ \Delta y \end{bmatrix} = \begin{bmatrix} -\nabla f(x) + \nabla h(x)^T y \\ \mu e - WYe \\ -h(x) + w \end{bmatrix}. \tag{29}$$

After solving the third equation for Δy and eliminating Δy from the first two equations we obtain the following reduced system for the primal directions

$$\begin{bmatrix} H+\beta A^T A & -\beta A^T \\ -\beta A & W^{-1}Y+\beta I \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta w \end{bmatrix} = \begin{bmatrix} -\sigma+\beta A^T\rho \\ \gamma-\beta\rho \end{bmatrix}. \tag{30}$$

On the other hand the gradient of $\mathcal{L}_{\beta,\mu}(x,w,y)$ with respect to x and w is as follows

$$\nabla_x \mathcal{L}_{\beta,\mu}(x,w,y) = \sigma - \beta A^T\rho$$
$$\nabla_w \mathcal{L}_{\beta,\mu}(x,w,y) = -\gamma + \beta\rho$$

Therefore, assuming that matrix $N_\beta = H + A^T\left[\beta^{-1}I + Y^{-1}W\right]^{-1}A$ is positive definite (otherwise the algorithm always regularizes it by adding λI such that the smallest eigenvalue of matrix N_β exceeds parameter $\lambda_0 > 0$.), we have by Lemma A1 from the Appendix

$$\begin{aligned} \begin{bmatrix} \nabla_x \mathcal{L}_{\beta,\mu} \\ \nabla_w \mathcal{L}_{\beta,\mu} \end{bmatrix}^T \begin{bmatrix} \Delta x \\ \Delta w \end{bmatrix} &= -\begin{bmatrix} \Delta x \\ \Delta w \end{bmatrix}^T \begin{bmatrix} H+\beta A^T A & -\beta A^T \\ -\beta A & W^{-1}Y+\beta I \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta w \end{bmatrix} \\ &\le -\alpha \max\{\|\Delta x\|, \|\Delta w\|\}^2, \end{aligned} \tag{31}$$

where α depends on parameters λ_0 and β. Lemma 3 is proven.

We will need also several lemmas about local convergence properties of the algorithm.

Lemma 4 *If $z^* = (x^*, w^*, y^*)$ is a solution to the problem (2) then the matrix*

$$D(z^*) = \begin{bmatrix} H(x^*,y^*) & 0 & -A(x^*)^T \\ 0 & Y^* & W^* \\ A(x^*) & -I & 0 \end{bmatrix}$$

is nonsingular and hence there exists $M^* > 0$ such that

$$\|D^{-1}(z^*)\| \le M^*. \tag{32}$$

Proof. The proof is straightforward (see e.g. [5]).

Let $\Omega_\varepsilon(z^*) = \{z : \|z - z^*\| \le \varepsilon\}$ be the ε-neighborhood of the solution to the problem (2).

Lemma 5 *There exists $\varepsilon_0 > 0$ and $0 < L_1 < L_2$ such that for any primal-dual pair $z \in \Omega_{\varepsilon_0}(z^*)$the merit function $\nu(z)$ satisfies*

$$L_1\|z - z^*\| \leq \nu(z) \leq L_2\|z - z^*\|. \tag{33}$$

Proof. Keeping in mind that $\nu(z^*) = 0$ the right inequality (33) follows from continuity of $\nu(z)$ and the boundedness of Ω_{ε_0}. Therefore there exists $L_2 > 0$ such that

$$\nu(z) \leq L_2\|z - z^*\|$$

Let us prove the left inequality. From a definition of a merit function $\nu(z)$ we obtain

$$\|\sigma\| \leq \nu(z), \tag{34}$$

$$WYe \leq \nu(z), \tag{35}$$

$$\|\rho\| \leq \nu(z), \tag{36}$$

Let us linearize σ, WYe and ρ at the solution $z^* = (x^*, w^*, y^*)$.

$$\sigma(z) = \sigma(z^*) + H(x^*, y^*)(x - x^*) - A^T(x^*)(y - y^*) + \mathcal{O}\|x - x^*\|^2$$

$$WYe = W^*Y^*e + Y^*(w - w*) + W^*(y - y^*) + \mathcal{O}\|w - w^*\|\|y - y^*\|$$

$$-\rho(z) = -\rho(z^*) + A^T(x^*)(x - x^*) - (w - w^*) + \mathcal{O}\|x - x^*\|^2$$

By Lemma 4 the matrix

$$D^* = D(z^*) = \begin{bmatrix} H(x^*, y^*) & 0 & -A(x^*)^T \\ 0 & Y^* & W^* \\ A(x^*) & -I & 0 \end{bmatrix}$$

is nonsingular and there is a constant M^* such that $\|D^{-1}(z^*)\| \leq M^*$. Therefore we have

$$\|z - z^*\| \leq M^*\nu(z) + \mathcal{O}\|z - z^*\|^2$$

Choosing $L_1 = 1/(2M^*)$, we obtain the left inequality (33), i.e.

$$L_1\|z - z^*\| \leq \nu(z)$$

Lemma 5 is proven.

Lemma 6 *Let the matrix* $A \in \mathbb{R}^{n,n}$ *be nonsingular such that* $\|A^{-1}\| \leq M$ *and the matrix* $B \in \mathbb{R}^{n,n}$ *is such that* $\|A - B\| \leq \varepsilon$ *for some* $\varepsilon > 0$. *Therefore there exists* $\varepsilon > 0$ *such that matrix* B *is nonsingular and we have*

$$\|B^{-1}\| \leq 2M$$

Proof. Since the matrix A is nonsingular, we have

$$B = A - (A - B) = A(I - A^{-1}(A - B)).$$

Let us denote matrix $C = A^{-1}(A - B)$. Since $\|A^{-1}\| \leq M$, we can choose such $\varepsilon > 0$ small enough that

$$\|C\|_2 \leq \frac{1}{2\sqrt{n}}.$$

Therefore there exists matrix $(I - C)^{-1}$ and we have

$$\|(I - C)^{-1}\| \leq \|I\| + \|C\| + \|C\|^2 + \cdots \leq 1 + \left(\frac{1}{2}\right) + \left(\frac{1}{2}\right)^2 + \cdots \leq 2.$$

Thus we have the following estimate

$$\|B^{-1}\| = \|(I - C)^{-1}A^{-1}\| \leq \|(I - C)^{-1}\|\|A^{-1}\| \leq 2M.$$

Lemma 6 is proven.

Lemma 7 *There exists* $\varepsilon_0 > 0$ *and* $M_2 > 0$ *such that for any primal-dual pair* $z = (x, w, y) \in \Omega_{\varepsilon_0}(z^*)$ *and* $\epsilon \leq \varepsilon_0$ *the matrix*

$$D_\epsilon(z) = \begin{bmatrix} H(x,y) & 0 & -A(x)^T \\ 0 & Y & W \\ A(x) & -I & \epsilon I \end{bmatrix}$$

has an inverse and its norm satisfies

$$\|D_\epsilon^{-1}(z)\| \leq M_2. \tag{37}$$

Proof. It follows from the Lipschitz conditions and boundedness of $\Omega_{\varepsilon_0}(z^*)$ that we have

$$\|D_\epsilon(z) - \mathcal{D}(z^*)\| \leq C_1\varepsilon_0,$$

for some $C_1 > 0$. Therefore, by Lemmas 4 and 6 there exists $M_2 > 0$ such that

$$\|D_\epsilon(z)^{-1}\| \leq M_2.$$

for $\varepsilon_0 > 0$ small enough. Lemma 7 is proven.

The following assertion is a slight modification of the Debreu theorem [4].

Assertion 1. *Let H be a symmetric matrix, $A \in \mathbb{R}^{r\times n}$, $\Lambda = diag(\lambda_i)_{i=1}^r$ with $\lambda_i > 0$ and there is $\theta > 0$ that $\xi^T H\xi \geq \theta\xi^T\xi$, $\forall\xi : A\xi = 0$. Then there exists $k_0 > 0$ large enough that for any $0 < \theta_1 < \theta$ the inequality*

$$\xi^T \left((H + kA^T\Lambda A\right)\xi \geq \theta_1\xi^T\xi, \quad \forall\xi \in \mathbb{R}^n \tag{38}$$

holds for any $k \geq k_0$.

The next lemma follows from Assertion 1.

Lemma 8 *There exists $\varepsilon_0 > 0$ small enough that for any approximation of the primal-dual solution $z = (x, w, y) \in \Omega_{\varepsilon_0}(z^*)$, $\epsilon = \nu_\mu(z)$ and $\mu = \min\{\delta\nu(z), \nu(z)^2\}$, the matrix $N_\epsilon(x, y, w)$ is positive definite.*

Proof. Let's assume that the active constraint set at x^* is $I^* = \{i : h_i(x^*) = 0\} = \{1, \ldots, r\}$. We consider the vectors function $h_{(r)}^T(x) = (h_1(x), \ldots, h_r(x))$, its Jacobian $A_{(r)}(x)$. The sufficient regularity conditions

$$rank\, A_{(r)}(x^*) = r, y_i^* > 0,\, i \in I^*$$

together with the sufficient conditions for the minimum x^* to be isolated

$$\xi^T H(x^*, y^*)\xi \geq \theta\xi^T\xi,\, \theta > 0, \forall\xi \neq 0 : A_{(r)}(x^*)\xi = 0$$

comprise the standard second order optimality conditions.

It follows from Assertion 1 and the second order optimality conditions that the matrix $M(x^*, y^*) = H(x^*, y^*) + kA_{(r)}(x^*)^T A_{(r)}(x^*)$ is positive definite for some $k \geq k_0$ and therefore the matrix $M(x, y)$ remains positive definite in some ε_0 neighborhood of the solution (x^*, y^*).

The matrix $N_\epsilon(x, y, w)$ can be written as follows

$$\begin{aligned} N_\epsilon(x, y, w) &= H(x, y) + A_{(r)}(x)^T \left[W_{(r)}Y_{(r)}^{-1} + \epsilon I\right]^{-1} A_{(r)}(x) \\ &+ A_{(m-r)}(x)^T \left[W_{(m-r)}Y_{(m-r)}^{-1} + \epsilon I\right]^{-1} A_{(m-r)}(x), \end{aligned} \tag{39}$$

where the second and the third terms correspond to active and inactive constraints. Keeping in mind (33), we have

$$\epsilon = \nu_\mu(z) \leq (1+\delta)\nu(z) \leq L_2(1+\delta)\varepsilon_0.$$

Also, due to the standard second order optimality conditions for the active constraints, we have $|w_i| \leq \varepsilon_0$ and $\tau_a \leq y_i \leq 2\tau_a$, $i = 1, \ldots, m$ for some $\tau_a > 0$. Therefore, we obtain

$$\left[W_{(r)}Y_{(r)}^{-1} + \epsilon I\right]^{-1} \geq \frac{\tau_a}{1 + 2\tau_a(1+\delta)L_2}\varepsilon_0^{-1} I_{(r)}, \tag{40}$$

where $I_{(r)}$ is the identity matrix.

For the inactive constraints we have $|y_i| \leq \varepsilon_0$, and $w_i \geq \tau_{in}$ for some $\tau_{in} > 0$. Therefore, we have

$$\left[W_{(m-r)}Y_{(m-r)}^{-1} + \epsilon I\right]^{-1} \leq \frac{1}{\tau_{in}}\varepsilon_0 I_{(m-r)}, \tag{41}$$

where $I_{(m-r)}$ is also the identity matrix.

Therefore, by choosing $\varepsilon_0 > 0$ small enough we can make the third term of (39) very small and the elements of the diagonal $\left[W_{(r)}Y_{(r)}^{-1} + \epsilon I\right]^{-1}$ as large as necessary. Therefore the positive definiteness of the matrix $N_\epsilon(x, y, w)$ follows from Assertion 1.

Remark 2 *It follows from Lemma 8 that in the neighborhood of the solution there is no need for the primal regularization of the matrix* $N_\epsilon(x, y, w)$.

Lemma 9 *There exists* $\varepsilon_0 > 0$ *such that if any approximation of the primal-dual solution* $z = (x, w, y) \in \Omega_{\varepsilon_0}(z^*)$, *with the barrier, dual regularization and steplength parameters obtained by the formulas (17)-(21) and the primal-dual direction* $\Delta z = (\Delta x, \Delta w, \Delta z)$ *obtained from the system (11) then*

$$\|\hat{z} - z^*\| \leq c\|z - z^*\|^2,$$

where $\hat{z}$ *is the next primal-dual approximation obtained by formulas (14)-(16) and the constant* $0 < c < \infty$ *depends only on the problem's data.*

Proof. Let $\varepsilon_0 > 0$ be small enough that the conditions of Lemmas 5-8 hold true. Let $z = (x, w, y) \in \Omega_{\varepsilon_0}(z^*)$. Let us denote $\|z - z^*\| = \varepsilon \leq \varepsilon_0$. For ε_0 small enough and using (33), we have

$$\mu = \nu(z)^2 \leq L_2^2\varepsilon^2. \tag{42}$$

It follows from formulas (33), (37) and (42) that

$$\|b_\mu(z)\| = \nu_\mu(z) \le \nu(z) + \mu \le c_1\varepsilon,$$

for some $c_1 > 0$. Since the algorithm computes the primal-dual direction by the formula $\Delta z = -D_\epsilon(z)^{-1} b_\mu(z)$, then keeping in mind (37), we have

$$\|\Delta z\| \le M_2 c_1 \varepsilon. \tag{43}$$

First we prove an estimation for the primal and dual steplengths obtained by formulas (17), (18) and (21). The second equation of the system (11) can be rewritten as follows

$$y_i \Delta w_i + w_i \Delta y_i = \mu - w_i y_i, \quad i = 1, \dots, m.$$

Therefore, keeping in mind that $\mu > 0$ and $w_i y_i > 0$, we have

$$y_i \Delta w_i + w_i \Delta y_i \ge -w_i y_i, \quad i = 1, \dots, m.$$

or

$$-\frac{\Delta w_i}{w_i} \le 1 + \frac{\Delta y_i}{y_i}, \quad i = 1, \dots, m.$$

By Assumption (A4) for the set of active constraints we have $|w_i| \le \varepsilon$ and $y_i \ge \tau_a > 0$. Therefore keeping in mind (43) for the indices $i : \Delta w < 0$ we have

$$-\frac{w_i}{\Delta w_i} \ge \frac{1}{1 + \frac{\Delta y_i}{y_i}} \ge \frac{1}{1 + \frac{M_2 c_1 \varepsilon}{\tau_a}} \ge 1 - c_2 \varepsilon, \tag{44}$$

where $c_2 = \frac{M_2 c_1}{\tau_a}$. By formulas (21) and (33) we have

$$\kappa \ge 1 - \nu(z) \ge 1 - L_2 \varepsilon. \tag{45}$$

Therefore combining formulas (17), (44) and (45) we obtain

$$1 - c_3 \varepsilon \le \alpha_p \le 1. \tag{46}$$

Following the same scheme we establish a similar estimate for the dual steplength

$$1 - c_4 \varepsilon \le \alpha_d \le 1. \tag{47}$$

Let us denote $\mathcal{A} \in \mathbb{R}^{n+2m}$ the diagonal matrix with the elements $\alpha_i = \alpha_p$, $i = 1, \dots, n+m$ and $\alpha_i = \alpha_d$, $i = n+m+1, \dots, n+2m$. Using $\mathcal{A}$, the next primal-dual approximation $\hat{z}$ is computed by the formula

$$\hat{z} = z + \mathcal{A}\Delta z.$$

Combining formulas (46) and (47) we obtain

$$\|I - \mathcal{A}\| \leq c_5 \varepsilon, \tag{48}$$

where $c_5 = \max\{c_3, c_4\}$. Now we estimate the distance between the next primal-dual approximation $\hat{z}$ and the solution. We have

$$\hat{z} - z^* = z - \mathcal{A}D_\epsilon^{-1}(z)b_\mu(z) - z^* = \mathcal{A}(z - z^*) - \mathcal{A}D_\epsilon^{-1}(z)b_\mu(z) + (I - \mathcal{A})(z - z^*)$$

$$= \mathcal{A}D_\epsilon^{-1}(z)\left(D_\epsilon(z)(z - z^*) - b_\mu(z)\right) + (I - \mathcal{A})(z - z^*)$$

$$= \mathcal{A}D_\epsilon^{-1}(z)\left[D(z)(z - z^*) - b(z) + (D_\epsilon(z) - D(z))(z - z^*) + (b_\mu(z) - b(z)\right]$$

$$+ (I - \mathcal{A})(z - z^*)$$

Using the Taylor expansion of $b(z^*)$ around z we obtain

$$0 = b(z^*) = b(z) + D(z)(z^* - z) + \mathcal{O}\|z - z^*\|^2,$$

or

$$D(z)(z - z^*) - b(z) = \mathcal{O}\|z - z^*\|^2.$$

Therefore, using formulas (19), (20), (33), (42) and (48), we have

$$\|\hat{z} - z^*\| \leq M_2[\|D(z)(z - z^*) - b(z)\| + \|D_\epsilon(z) - D(z)\|\|z - z^*\|$$

$$+ \|b_\mu(z) - b(z)\|] + c_5\|z - z^*\|^2 = M_2[c_6\varepsilon^2 + \nu_\mu(z)\varepsilon + \mu] + c_5\varepsilon^2$$

$$\leq M_2[c_6\varepsilon^2 + L_2\varepsilon^2 + L_2^2\varepsilon^3 + L_2^2\varepsilon^2] + c_5\varepsilon^2 \leq c\varepsilon^2,$$

where $c_2 = M_2(c_6 + 3L_2) + c_5$. Lemma 9 is proven.

Now we are ready to prove the main theorem about convergence properties of the IPM algorithm.

Theorem 1 *Under assumptions (A1)-(A6), the IPM algorithm generates a primal-dual sequence* $\{z^s = (x^s, w^s, y^s)\}$ *such that any limit point* $\bar{x}$ *of the primal sequence* $\{x^s\}$ *is a first-order optimality point for the minimization of* *the l_2 norm of the vector of the constraint violation $v(x) = (v_1(x), \ldots, v_m(x))$, where $v_i(x) = \min\{h_i(x), 0\}$:*

$$V(x) = \|v(x)\|_2.$$

If, in particular, $V(\bar{x}) = 0$ then $\bar{x} = x^*$ is a a first order optimality point of problem (1).

Proof. Let $z^* = (x^*, w^*, y^*)$ be a local or global solution to problem (2) and let the sequence $\{z^s\}$, where $z^s = (x^s, w^s, y^s) = (p^s, y^s)$, be generated by the algorithm with a given an initial approximation z^0. Let $\varepsilon_0 > 0$ be such that conditions of Lemma 9 hold.

We consider three possible scenarios. First, if the initial approximation $z^0 \in \Omega_{\varepsilon_0}(z^*)$ then the algorithm converges to z^* with a quadratic rate by Lemma 9.

Now we consider the case when the trajectory of the algorithm is outside the ε_0-neighborhood of the solution. In this case the algorithm minimizes the merit function $\mathcal{L}_{\beta,\mu}(p, y^s)$ in p. Indeed, it follows from Lemmas 2, 3 and the Armijo rule (22) that for the fixed Lagrange multipliers y and the chosen penalty parameter β the primal direction Δp satisfies the following condition

$$(\nabla_p \mathcal{L}_{\beta,\mu}(p, y), \Delta p) \leq -\bar{\alpha}\|\Delta p\|\|\nabla_p \mathcal{L}_{\beta,\mu}(p, y)\|,$$

for some $\bar{\alpha} > 0$, and also that the gradient of the merit function $\nabla_x \mathcal{L}_{\beta,\mu}(p, y^s)$ satisfies the Lipschitz condition on some open set containing the level set $\{p : \mathcal{L}_{\beta,\mu}(p, y^s) \leq \mathcal{L}_{\beta,\mu}(p^{s_0}, y^s)\}$, where p^{s_0} is a starting point of an unconstrained minimization of $\mathcal{L}_{\beta,\mu}(p, y^s)$ in p. Therefore the algorithm eventually descends to the approximation $\hat{p}$ of the first order optimality point of this unconstrained minimization (see e.g. [7]). After finding such an approximation, the algorithm changes the Lagrange multipliers by formula

$$y^{s+1} := y^s + \beta\rho(\hat{x}, \hat{w}), \tag{49}$$

if this update reduces the value of the merit function $\nu_\mu(z)$ by a chosen factor $0 < q < 1$. Otherwise the algorithm increases the value of penalty parameter β and continues the minimization of $\mathcal{L}_{\beta,\mu}(p, y^s)$ in p.

Here, there are two further possible scenarios. They both occur when the algorithm increases the value of the penalty parameter β after minimization of the merit function $\mathcal{L}_{\beta,\mu}(p, y^s)$ in p.

By the second scenario the minimization of the merit function $\mathcal{L}_{\beta,\mu}(p, y^s)$ in p for a larger β followed by the Lagrange multipliers update brings the trajectory close to the solution to the barrier problem (3) due to the global convergence properties of the Augmented Lagrangian algorithm [8, 9]. In this case the algorithm reduces the value of the merit function $\nu_\mu(z)$. Therefore for the value of the merit function $\nu(z)$ the following estimate takes place

$$\nu(\hat{z}) \le \nu_\mu(\hat{z}) + \mu = \nu_\mu(\hat{z}) + \min\{\delta r,\, r^2\},$$

where r is the previous record value of the merit function $\nu(z)$ and $0 < \delta < 1$. The value of the barrier parameter μ is smaller than the previous record value of the merit function $\nu(z)$ before parameter μ was decreased. Therefore the reduction of the merit function $\nu_\mu(z)$ will guarantee the reduction of the merit function $\nu(z)$. Thus the algorithm reduces the merit function $\nu(z)$ followed by the further reduction of the barrier parameter μ eventually brings its trajectory to the neighborhood of some first order optimality point $\Omega_{\varepsilon_0}(z^*)$. If z^* is a local or global solution of the problem (2) then then the algorithm converges to the solution by the first scenario. In general case, however, we can guarantee that any limit point of the sequence generated by the algorithm is the first order optimality point of the problem (2):

$$\lim_{s\to\infty} \nu(z^s) = 0,$$
$$\lim_{s\to\infty} w^s \ge 0,$$
$$\lim_{s\to\infty} y^s \ge 0.$$

By the third scenario the algorithm does not change the Lagrange multipliers y by formula (49) since this update does not reduce the value of the merit function $\nu_\mu(z)$. Therefore, the algorithm turns into the sequence of unconstrained minimizations of the merit function $\mathcal{L}_{\beta,\mu}(p, y)$ in p followed by an increase of the barrier parameter β. The vector of the Lagrange multipliers y does not change according to the algorithm. Let us show that any limit point of the primal sequence $\{x^s\}$ is actually the first order optimality point for the minimization of the l_2 norm of the vector of the constraint violation $v(x) = (v_1(x), \ldots, v_m(x)$, where $v_i(x) = \min\{h_i(x), 0\}$:

$$V(x) = \|v(x)\|_2.$$

First we will show that the primal sequence $\{p^s\}$ is bounded. Consider the monotone increasing sequence $2m\mu \le \beta^0 \le \beta^1 \le \ldots \le \beta^k \le \ldots$. We can rewrite a merit function $\mathcal{L}_{\beta,\mu}(p, y)$ as follows

$$\mathcal{L}_{\beta,\mu}(p, y) = L_\mu(p, y) + \frac{\beta}{2}\rho^T\rho$$

$$= (1+\beta-\beta^0)\left[\frac{1}{1+\beta-\beta^0}\left(L_\mu(p,y)+\frac{\beta^0}{2}\rho^T\rho\right)+\frac{\beta-\beta^0}{2(1+\beta-\beta^0)}\rho^T\rho\right]$$
$$= \frac{1}{\xi}\left[\xi g_1(p,y)+(1-\xi)g_2(p,y)\right] = \frac{1}{\xi}\theta_\xi(p,y),$$

where $L_\mu(p,y) = f(x) - \mu\sum_{i=1}^m \log w_i + y^T\rho$, $\xi = 1/(1+\beta-\beta_0)$, $g_1(p,y) = L_\mu(p,y)+0.5\beta_0\rho^T\rho$, $g_2(p,y) = 0.5\rho^T\rho$ and $\theta_\xi(p,y) = \xi g_1(p,y)+(1-\xi)g_2(p,y)$. Therefore the sequence of unconstrained minimizations of the merit function $\mathcal{L}_{\beta^s,\mu}(p,y)$ in p for the monotone nondecreasing sequence $\beta^0 \le \beta^1 \le \ldots \le \beta^k \le \ldots$ is equivalent to the sequence of unconstrained minimizations of function $\theta_\xi(p,y)$ in p for the monotone nonincreasing sequence $1 = \xi_0 \ge \xi_1 \ge \ldots \ge \xi^k > 0$.

Suppose that the primal sequence $\{p^s\}$ is unbounded. Since $p^s = (x^s, w^s) \in \mathbb{R}^n \times \mathbb{R}^m_{++}$, by Remark 1 following Lemma 2, the sequence $\{g_2^s\}$, where $g_2^s = g_2(p^s, y)$ is unbounded and

$$\lim_{k\to\infty}\sup_{0\le s\le k} g_2^s = +\infty. \tag{50}$$

We will show that (50) implies that

$$\lim_{k\to\infty}\inf_{0\le s\le k} g_1^s = -\infty \tag{51}$$

with $g_1^s = g_1(p^s, y)$, which contradicts again Lemma 2.

First, we renumber the sequence $\{p^s\}$ as follows

$$p^0 = p^{q_0}, p^{q_0+1}, \ldots, p^{q_0+d_0} = p^{q_1}, p^{q_1+1}, \ldots, p^{q_1+d_1} = \cdots = p^{q_k}, p^{q_k+1}, \ldots, p^{q_k+d_k}, \ldots$$

so all p^s, $s = q_k, \ldots, q_k+d_k$ correspond to the same value of ξ^k. For any k, for all $s = q_k, \ldots, q_k+d_k-1$ we have

$$\xi^k g_1^{s+1} + (1-\xi^k)g_2^{s+1} \le \xi^k g_1^s + (1-\xi^k)g_2^s,$$

or, equivalently

$$g_1^s - g_1^{s+1} \ge \frac{1-\xi^k}{\xi^k}(g_2^{s+1} - g_2^s) \tag{52}$$

After the summation of the inequality (52) over all $s = q_k, \ldots, q_k+d_k-1$, we obtain

$$g_1^{q_k} - g_1^{q_k+d_k} \ge \frac{1-\xi^k}{\xi^k}(g_2^{q_k+d_k} - g_2^{q_k}). \tag{53}$$

After the summation of the inequality (53) for all $k = 0, 1, \ldots j$ and keeping in mind that $g_1^{q_k+d_k} = g_1^{q_{k+1}}$ and $g_2^{q_k+d_k} = g_2^{q_{k+1}}$ for $k = 0, 1, \ldots, j-1$, we obtain

$$g_1^0 - g_1^{q_j+d_j} \geq \sum_{i=1}^{j} \frac{1-\xi^i}{\xi^i}(g_2^{q_i+d_i} - g_2^{q_i}). \tag{54}$$

Assuming that $s = q_j + d_j$ we recall that

$$\lim_{k\to\infty} \sup_{0\leq s\leq k} g_2^s = \lim_{k\to\infty} \sup_{0\leq s\leq k} \sum_{i=1}^{j}(g_2^{q_i+d_i} - g_2^{q_i}) = +\infty.$$

Since the sequence $\{\xi^k\}$ is monotonically decreasing to zero, the sequence $\left\{\frac{1-\xi^k}{\xi^k}\right\}$ is monotone, increasing and unbounded and greater than or equal to one starting with $k = 1$. Therefore by Lemma (A2) from the Appendix we have

$$\lim_{k\to\infty} \sup_{0\leq s\leq k} \sum_{i=1}^{j} \frac{1-\xi^i}{\xi^i}(g_2^{q_i+d_i} - g_2^{q_i}) = +\infty$$

Therefore using (54) we obtain

$$\lim_{k\to\infty} \sup_{0\leq s\leq k} (g_1^0 - g_1^s) = +\infty,$$

or equivalently

$$\lim_{k\to\infty} \inf_{0\leq s\leq k} g_1^s = -\infty,$$

which contradicts Lemma 2. Therefore our assumption of unboundedness of the sequence $\{p^s\}$ was not correct and we conclude that the primal sequence $\{p^s\}$ generated by the algorithm is bounded.

Now we show that any limit point of the primal sequence $\{x^s\}$ generated by the algorithm is actually the first order optimality point for minimization of the l_2 norm of the vector of the constraint violation $v(x) = (v_1(x), \ldots, v_m(x))$, where $v_i(x) = \min\{h_i(x), 0\}$:

$$V(x) = \|v(x)\|_2$$

The necessary conditions for the primal pair $\hat{p} = (\hat{x}, \hat{w})$ to be a minimizer of merit function $\mathcal{L}_{\beta,\mu}(p, y)$ in p is the following system

$$\begin{aligned} \nabla f(x) - A(x)^T(y + \beta\rho) &= 0, \\ -\mu W^{-1}e + y + \beta\rho &= 0. \end{aligned} \tag{55}$$

If the triple $\hat{z} = (\hat{x}, \hat{w}, \hat{y})$, where $\hat{y} = y + \beta\rho$ satisfies system (55) then the only reason that merit function $\nu_\mu(\hat{z})$ is not zero is infeasibility: $\rho(\hat{x}, \hat{w}) \neq 0$.

Let us consider the sequence $\{z^s\}$, $z^s = (x^s, w^s, y^s)$ generated by the algorithm. The dual sequence $\{y^s\}$ does not change from some point on. We assume that $y^s = y$ for $s \geq s_0$. Also, the asymptotic infeasibility takes place: $\lim_{s\to\infty} \rho_i(x^s, w^s) \neq 0$ for some index i. We denote I_- the index set of all the indices such that $\lim_{s\to\infty} \rho_i(x^s, w^s) \neq 0$ for $i \in I_-$.

According to the algorithm for the sequence of the primal approximations of exact minimizers, we have

$$\begin{array}{rcl} \nabla f(x^k) - A(x^k)^T(y + \beta^k \rho(x^k, w^k)) & = & \Upsilon_n^k, \\ -\mu W_k^{-1} e + y + \beta^k \rho(x^k, w^k) & = & \Upsilon_m^k \end{array} \tag{56}$$

where $\lim_{k\to\infty} \Upsilon_n^k = 0$ and $\lim_{k\to\infty} \Upsilon_m^k = 0$.

If the primal sequence (x^k, w^k) satisfy the system (56), then it will satisfy the following system

$$\begin{array}{rcl} \nabla f(x^k)/\beta^k - A(x^k)^T y/\beta^k + A(x^k)\rho(x^k, w^k)) & = & \Upsilon_n^k/\beta^k, \\ -\mu/\beta^k + W^k y/\beta^k + W^k \rho(x^k, w^k) & = & W^k \Upsilon_m^k/\beta^k \end{array} \tag{57}$$

Therefore keeping in mind the boundedness of the sequence $\{(x^k, w^k)\}$, we have

$$\lim_{k\to\infty} A(x^k)\rho(x^k, w^k) = 0, \tag{58}$$

$$\lim_{k\to\infty} (w_i^k - h_i(x^k))w_i^k = 0, \quad i = 1, \dots, m. \tag{59}$$

and

$$\lim_{k\to\infty} w_i^k \geq 0, \quad i = 1, \dots, m. \tag{60}$$

It is easy to verify that conditions (58)-(60) are also the first-order optimality conditions for the problem

$$\begin{array}{l} \min \|w - h(x)\|_2^2, \\ \\ \text{s.t. } w \geq 0. \end{array} \tag{61}$$

In turn, the solution to the problem (61) (x^*, w^*) minimizes $V(x)$ (otherwise it would contradict the optimality of the problem (61)).

The theorem is proven.

6 Concluding remarks

In this paper we analyzed the convergence of the primal-dual interior-point algorithm for nonlinear optimization problems with inequality constraints. The important feature of the algorithm is the primal and dual regularization, which guarantees that the algorithm minimizes the merit function $\mathcal{L}_{\beta,\mu}(x, w, y)$ in (x, w) in order to drive the trajectory of the algorithm to the neighborhood of a local minimum rather than any other first order optimality point such as a maximum or a saddle point.

Another important feature of the algorithm is that it stabilizes a sequence of primal approximations in the sense that the algorithm minimizes the l_2-norm of the constraint violation without any assumptions on the sequence of primal and dual estimates to the optimal points. Such assumptions have been common in recent convergence proofs.

The next step is to generalize the theory for equality constraints and to work on numerical performance of the algorithm. Currently LOQO implements only primal regularization. Therefore the next important step in the future research would be to modify LOQO to include new features of the algorithm studied in this paper such as the dual regularization and more careful updating of the dual variables. We believe that such modifications can potentially improve the robustness of the solver.

Acknowledgement. The authors would like to thank Roman Polyak for his valuable comments and suggestions, which helped to improve the manuscript.

7 Appendix

Lemma A1. Let matrices $N = A - B^T C^{-1} B$ and C be symmetric positive definite with the smallest eigenvalue $\lambda_N > 0$ and $\lambda_C > 0$ respectively. Then the matrix

$$M = \begin{bmatrix} A & B^T \\ B & C \end{bmatrix}$$

is also positive definite with the smallest eigenvalue $\lambda_M > 0$ depending on λ_N and λ_C.

Proof. Let us show that for any $z = (x, y) \neq 0$ quadratic form $z^T M z$ is positive. Since matrix N is positive definite, we have

$$x^T (A - B^T C^{-1} B) x \geq \lambda_N x^T x$$

Therefore

$$[x^T y^T] \begin{bmatrix} A & B^T \\ B & C \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = x^T A x + y^T C y + 2y^T B x$$

$$\geq \lambda_N x^T x + x^T B^T C^{-1} B x + y^T C y + 2y^T B x$$
$$= \lambda_N x^T x + C^{-1}(Bx + Cy)^2 \geq \lambda_M z^T z$$

where $\lambda_M \geq \alpha \min\{\lambda_N, \lambda_C\}$, $\alpha > 0$. Lemma is proven.

Lemma A2. Let series $\sum_{i=0}^{\infty} a_i$ be such that the sequence of the largest partial sums $\{s_k\}$, where

$$s_k = \sup_{0 \leq l \leq k} \sum_{i=1}^{l} a_i$$

is unbounded monotone and increasing, i.e.

$$\lim_{k\to\infty} s_k = +\infty. \tag{62}$$

Also let a sequence $\{b_k\}$ with $b_k \geq 1$ be monotone increasing and such that $\lim_{k\to\infty} b_k = +\infty$. Then for the series $\sum_{i=0}^{\infty} a_i b_i$ the sequence of the largest partial sums $\{p_k\}$, where

$$p_k = \sup_{0 \leq l \leq k} \sum_{i=1}^{l} a_i b_i$$

is also unbounded monotone increasing, i.e.

$$\lim_{k\to\infty} p_k = +\infty.$$

Proof. To prove the lemma we are going to show that $p_k \geq s_k$ for $k = 0, 1, 2, \ldots$. Without loss of generality we assume that $s_0 = a_0$ are positive, otherwise we can add any positive number in the series $\sum_{i=0}^{\infty} a_i$ as the first term without changing the property (62). Thus the sequence $\{s_k\}$ has the following property

$$0 < s_0 = s_{q_0} \cdots = s_{q_1-1} < s_{q_1} = s_{q_1+1} = \cdots = s_{q_2-1} < \cdots.$$

In other words, the sequence $\{s_k\}$ is shattered on infinite number of groups with equal elements.

Since there is one to one correspondence between the sequences $\{s_k\}$ and $\{a_k\}$, where a_k is the k-th term of the series $\sum_{i=0}^{\infty} a_i$, we can use the same enumeration for $\{a_k\}$ described above and based on the sequence $\{s_k\}$. Consequently, we will the same introduced enumeration of all the rest sequences $\{b_k\}$, $\{a_k b_k\}$ and $\{p_k\}$.

Such enumeration helps us to understand some useful properties of the elements of considered sequences. First of all it is easy to see that $a_{q_i+1} < 0$, if $a_{q_i+1} \neq a_{q_{i+1}}$, and $a_{q_i} > 0$, $i = 0, 1, 2, \ldots$. Moreover, we have

$$\sum_{j=q_i+1}^{q_{i+1}} a_j > 0, \quad i = 1, 2, \ldots.$$

Thus, for any $i = 0, 1, 2, \ldots$ all the negative terms of the sum $\sum_{j=q_i+1}^{q_{i+1}} a_j$ are neutralized by following them positive terms and there is some reserve in the last positive term $a_{q_{i+1}}$ that makes the whole sum positive.

Therefore, due to the monotonicity of the increasing positive sequence $\{b_k\}$ for any $i = 1, 2, \ldots$ all the negative terms of the sum $\sum_{j=q_i+1}^{q_{i+1}} a_j b_j$ are also neutralized by following them positive terms. Moreover, keeping in mind $b_i \geq 1$ for all $i = 1, 2, \ldots,$ we have

$$\sum_{j=q_i+1}^{q_{i+1}} a_j b_j \geq \sum_{j=q_i+1}^{q_{i+1}} a_j b_{q_i+1} = b_{q_i+1} \sum_{j=q_i+1}^{q_{i+1}} a_j \geq \sum_{j=q_i+1}^{q_{i+1}} a_j$$

Since $s_0 = s_{q_0}$ is positive then we have $p_{q_0} \geq s_{q_0}$. Assuming that $p_{q_i} \geq s_{q_i}$, we obtain

$$p_{q_{i+1}} \geq p_{q_i} + \sum_{j=q_i+1}^{q_{i+1}} a_j b_j \geq p_{q_i} + \sum_{j=q_i+1}^{q_{i+1}} a_j \geq s_{q_i} + \sum_{j=q_i+1}^{q_{i+1}} a_j = s_{q_{i+1}}.$$

Therefore by induction we have $p_k \geq s_k$ for $k = 0, 1, 2, \ldots$ and

$$\lim_{k \to \infty} p_k = +\infty.$$

The lemma is proven.

References

[1] H. Benson, D. Shanno and R. Vanderbei, *Interior point methods for nonconvex nonlinear programming: jamming and comparative numerical testing,* to appear in Math programming.

[2] H. Benson, D. Shanno and R. Vanderbei, *Interior point methods for nonconvex nonlinear programming: filter methods and merit functions,* Computational Optimization and Applications, 23:257-272, 2002.

[3] H. Benson, D. Shanno and R. Vanderbei, *A comparative study of large scale nonlinear optimization algorithms,* Technical Report, ORFE-01-04, Department of Operations Research and Financial Engineering, 2001.

[4] G. Debreu, *Definite and semidefinite quadratic forms,* Econometrica, 20 (1952), 295-300.

[5] A.S. El-Bakry, R.A. Tapia, T. Tsuchia, and Y. Zhang, *On the formulation and Theory of the Newton Interior-Point Method for Nonlinear Programming,* Journal of Optimization theory and Applications, Vol. 89, No. 3, pp. 507-541, 1996.

[6] L.V. Kantorovich, *Functional analysis and applied mathematics,* Uspekhi Matematicheskikh Nauk, 3 (1948), pp. 89 - 185.

[7] J. Nocedal, S. Wright (1999): Numerical Optimization, Springer, New York.

[8] B.T. Polyak, N.V. Tretjakov, *A penalty method for constrained optimization problems*, Journal of computational mathematics and mathematical physics, Vol. 13, No. 1, pp. 34-46.

[9] R.T. Rockafellar, *Augmented Lagrangians and Applications of the Proximal Point algorithm in convex programming,* Mathematics of Operations Research, Vol. 1, No. 2, pp.97-116, 1976.

[10] D.F. Shanno and R. J. Vanderbei, *Interior-point methods for nonconvex nonlinear programming: ordering and higher-order methods,* Math programming, 87(2), 303-316, 2000.

[11] M. Ulbrich, S. Ulbrich, and L. N. Vicente, *A globally convergent primal-dual interior-point filter method for nonlinear programming,* to appear in Mathematical Programming.

[12] R. Vanderbei, "Symmetric Quasidefinite Matrices", SIAM Journal on Optimization, vol. 5(1), pp. 100-113, 1995.

[13] R. J. Vanderbei, D.F. Shanno, *An interior-point algorithm for nonconvex nonlinear programming,* COAP vol. 13, pp. 231-252, 1999.