Algorithmic Operations Research Vol.3 (2008) 12–29

Global convergence of a primal-dual interior-point method for nonlinear programming

Igor Griva

George Mason University, Departments of Math Sciences and CDS, Fairfax, VA 22030

David F. Shanno

Rutgers University, RUTCOR, New Brunswick, NJ 08903

Robert J. Vanderbei

Princeton University, Department of ORFE, Princeton NJ 08544

Hande Y. Benson

Drexel University, Department of Decision Sciences, Philadelphia, PA 19104

Abstract

Many recent convergence results obtained for primal-dual interior-point methods for nonlinear programming, use assumptions of the boundedness of generated iterates. In this paper we replace such assumptions by new assumptions on the NLP problem, develop a modification of a primal-dual interior-point method implemented in software package LOQO *and analyze convergence of the new method from any initial guess.*

Key words: Interior-point method, primal-dual, convergence analysis.

1. Introduction

The primal-dual interior-point algorithm implemented in LOQO proved to be efficient for solving nonlinear optimization problems ([1–3,15,18]). 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 [18] implemented in LOQO at each step minimized a penalty barrier merit function to attempt to ensure that the algorithm converged to a local minimum. 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 convergence to a first-order KKT point from an arbitrary initial guess 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. The latter appear in many convergence analyses (see e.g. [5,6,9,12,16,19]). The full implementation of the studied algorithm in the LOQO framework remains for future work. An implemented preliminary version of the algorithm converges to a minimum of some problems, on which LOQO previously failed.

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

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):

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

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

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

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 $e = (1, \ldots, 1)^T \in \mathbb{R}^m$.

3. Assumptions

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

We invoke the following assumptions throughout the paper.

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

A2. The constraints $h_i(x)$ satisfy the following conditions

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

and

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

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

A3. The minima (local and global) of problem (1) satisfy the standard second order optimality conditions.

A4. For each $\mu > 0$ the minima (local and global) of problem (3), satisfy the standard second order optimality conditions.

A5. 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 (A2) 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. The cases when functions $h_i(x)$ do not satisfy assumption (A2) may involve exponentially growing functions $h_i(x)$. Let us consider the following example. The feasible set $\Omega_1 = [-1, 1] \subset \mathbb{R}^1$ can be defined using two inequalities: $h_1(x) = x + 1 \geq 0$ and $h_2(x) = 1 - x \geq 0$. In this case functions $h_1(x)$ and $h_2(x)$ satisfy assumption (A2). 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 (A2).

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 (A2). We believe 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 (A2) to ensure that the merit function has bounded level sets.

Let us assume that the active constraint set at x^* is $I^* = \{i : h_i(x^*) = 0\} = \{1, \ldots, r\}$. We consider the vectors functions $h_{(r)}^T(x) = (h_1(x), \ldots, h_r(x))$, and its Jacobian $A_{(r)}(x)$. The sufficient regularity conditions

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

together with the sufficient condition for the minimum

x^* to be isolated

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

where $H(x, y)$ is a Hessian of the Lagrangian for problem (1), comprise the standard second order optimality conditions, or Assumption (A3).

Assumption (A4) is equivalent to the following condition to hold $\forall \mu > 0$:

$$s^T (H\left(x_\mu, \mu C^{-1}(x_\mu)e\right) + \mu A^T(x_\mu)C^{-2}(x_\mu)A(x_\mu))s \geq \rho_\mu s^T s,$$
$$\rho_\mu > 0,\ \forall s \neq 0,$$

where (x_μ, w_μ) is the solution of the barrier subproblem (3), $H(x, y)$ is a Hessian of the Lagrangian for problem (1) and $C(x)$ is a diagonal matrix with the elements $c_i(x) = h_i(x)$, $i = 1, \dots, m$.

Remark 1 *It follows from Assumption A3 that the Slater's condition holds: there exists $\bar{x} \in \mathbb{R}^n$ such that $h_i(\bar{x}) > 0$, $i = 1, \dots, m$.*

All the assumptions (A1)-(A5) are imposed on the problem, not on the sequence generated by the algorithm. Our intention is to identify a class of nonconvex problems for which the interior-point algorithm is convergent. The following lemma follows from the assumptions.

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

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

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

where $B(x, \mu) = f(x) - \mu \sum_{i=1}^m \log h_i(x)$. It follows from assumption (A2) that the feasible set Ω is bounded. Let $\bar{x}$ be the point that exists by Remark 1 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.

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

$$\sigma = \nabla f(x) - A(x)^T y,$$

$$\gamma = \mu W^{-1} e - y,$$

$$\rho = w - h(x).$$

$$b(z) = (\sigma^T, (WYe)^T, -\rho^T)^T,$$

$$b_\mu(z) = (\sigma^T, (WYe)^T - \mu e^T, -\rho^T)^T,$$

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). We show later that 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

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

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

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

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 (8) are given in [18] (Theorem 1):

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

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

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

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

where I is the identity matrix in $\mathbb{R}^{n,n}$ to guarantee that the smallest eigenvalue of $\hat{N}$ is greater than some $\lambda_0 > 0$. The parameter λ_p 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)

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

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

$$D_{\lambda_d}(z)\Delta z = -b_\mu(z),$$

where

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

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

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

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

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

where I is the identity matrix in $\mathbb{R}^{n,n}$ to guarantee that the smallest eigenvalue of $\hat{N}_{\lambda_d}$ 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 interior-point method (IPM) algorithm $(x,w,y) \to (\hat{x},\hat{w},\hat{y})$ is as follows

$$\hat{x} = x + \alpha_p \Delta x, \quad (14)$$

$$\hat{w} = w + \alpha_p \Delta w, \quad (15)$$

$$\hat{y} = y + \alpha_d \Delta y, \quad (16)$$

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:

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

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

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

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

where $0 < \delta < 1$. Then the algorithm calculates the merit function $\nu_\mu(z)$ and the dual regularization parameter as follows:

$$\lambda_d = \min\{\lambda_{d\max}, \nu_\mu(z)\}, \quad (20)$$

where $\lambda_{d\max}$ is a fixed largest dual regularization parameter chosen by considerations following from convergence analysis (Lemma 2 and 3). Such a choice of the dual regularization is needed to guarantee that it vanishes in the neighborhood of the solution of problem (3), where due to Assumption A4 the dual regularization is not required. 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 [17]. It is possible that while performing the factorization the algorithm learns that the matrix $N_{\lambda_d}(x, w, y)$ is not positive definite. In this case the algorithm regularizes the matrix $N_{\lambda_d}(x, w, y)$ by formula (13) and begins the factorization again. It keeps increasing the parameter λ_p 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 $\hat{y} := y + \alpha_d \Delta y$.

The fundamental difference between this algorithm and basic interior-point algorithms is we require the candidate $\hat{z} = (\hat{x}, \hat{w}, \hat{y})$ to satisfy two criteria. First, $\hat{z}$ must reduce the merit function $\nu(z)$ by a chosen *a priori* desired factor $0 < q < 1$. If it succeeds in obtaining this reduction, after that the factorization of $N_{\lambda_d}(\hat{x}, \hat{w}, \hat{y})$, which will be used to calculate the search direction at $\hat{z}$, is done. If $N_{\lambda_d}(\hat{x}, \hat{w}, \hat{y})$ is positive definite, $\hat{z}$ is accepted and the algorithm continues. If $N_{\lambda_d}(\hat{x}, \hat{w}, \hat{y})$ is not positive definite, $\hat{z}$ fails as it does if a sufficient reduction of the merit function $\nu(z)$ is not obtained. In both of these cases, the Lagrange multipliers y are not changed. The primal direction $\Delta p = (\Delta x, \Delta w)$, which will be shown to be a descent direction for $\mathcal{L}_{\beta,\mu}(p, y)$, is used to update the primal iterates, where the primal steplength α_p is backtracked to satisfy the Armijo rule

$$\mathcal{L}_{\beta,\mu}(p + \alpha_p \Delta p, y) - \mathcal{L}_{\beta,\mu}(p, y) \leq \eta \alpha_p \left\langle \nabla_p \mathcal{L}_{\beta,\mu}(p, y), \Delta p \right\rangle, \tag{22}$$

where $0 < \eta < 1$.

The convergence analysis of the algorithm shows that under the assumptions (A1)-(A5) 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 the 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 in nonconvex regions, poor dual approximations may result from the solution of the primal-dual equation. These approximations can hamper convergence and even prevent it, which happens often in practice when they become unbounded. If the algorithm reaches an approximation $\hat{p} = (\hat{x}, \hat{w})$ of the first order optimality point of the unconstrained minimization of the merit function $\mathcal{L}_{\beta,\mu}(p, y)$, it then changes the dual variables by the formula

$$\hat{y} := y + \beta \rho(\hat{x}, \hat{w}), \tag{23}$$

where $\rho(x, w) = w - h(x)$, to obtain a better dual approximation. If $\hat{z}$ reduces the merit function $\nu_\mu(z)$, the algorithm accepts $\hat{z}$ as new primal-dual approximation. Otherwise, the algorithm keeps the Lagrange multipliers unchanged and increases the penalty parameter β.

It is appropriate to say several words about the choice of the dual regularization parameter λ_d and the penalty parameter β. These parameters are chosen to satisfy two conditions: a) the primal Newton direction $(\Delta x, \Delta w)$ must be a descent direction for the merit function $\mathcal{L}_{\beta,\mu}(z)$ and b) the regularization parameter $\lambda_d > 0$ must become zero when the trajectory of the algorithm approaches the primal-dual solution of the barrier subproblem (3).

To prove global convergence of the algorithm we use the following choice of the parameters at each iteration: $\lambda_d = \min\{\lambda_{d\max}, \nu_\mu(z)\}$, $\beta = 1/\lambda_d$, $\lambda_{d\max} = \frac{1}{\beta_0}$, where β_0 is the smallest value of the penalty parameter estimated in Lemma 2. It will be shown later that such choice of the parameters satisfies the conditions (a) and (b) and allows us to prove global convergence of the algorithm.

The formal description of the algorithm is in Figure 1.

5. Convergence analysis

We need the following auxiliary lemmas for the convergence analysis.

Lemma 2 *For any* $y \in \mathbb{R}^m$, $\beta \geq \beta_0 = 2m\mu$ *and* $\mu > 0$, *there exists a global minimum*

$$S_{\beta,\mu}(y) = \min_{x \in \mathbb{R}^n,\, w \in \mathbb{R}^m_{++}} \mathcal{L}_{\beta,\mu}(x, w, y) > -\infty. \tag{24}$$

Proof. 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 Slater's condition

Initialization:

An initial primal-dual approximation $z^0 = (p^0, y^0) = (x^0, w^0, y^0)$ is given
An accuracy $\varepsilon > 0$ and largest tolerance for the penalty parameter BIG are given
Parameters $0 < \eta < 0.5$, $0 < \delta < q < 1$, $\tau > 0$, $\theta > 0$ are given
Set $z := z^0$, $r := \nu(z^0)$, $\mu := \min\{\delta r, r^2\}$, $r_\mu = \nu_\mu(z^0)$, $\beta := \beta_0 \geq 2m\mu$, $\lambda_d := \min\{\nu(z^0), \frac{1}{\beta_0}\}$,
Iterations counter $s := 0$, Dual iterates update counter $k := 0$, $primal := 0$.

Computations:

While $r > \varepsilon$ and $\beta < BIG$ *do*
BOL (Beginning of the loop)
Factorize the system (11), Start with $\lambda_p = 0$, Increase λ_p until the factorization of (11) is successful
Find directions: $\Delta z := PrimalDualDirection(z, \lambda_d)$
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 $\hat{p} := p + \alpha_p \Delta p$, $\hat{y} := y + \alpha_d \Delta y$
l1: *If* $primal = 0$, $\nu(\hat{z}) \leq qr$ and the Cholesky factorization of (11) is successful with $\lambda_p = 0$ and *then*
Set $z := \hat{z}$, $r := \nu(\hat{z})$, $\mu := \min\{\delta r, r^2\}$, $r_\mu := \nu_\mu(\hat{z})$, $\lambda_d := \min\{\nu_\mu(\hat{z}), \frac{1}{\beta}\}$, $k := k + 1$
Else
l2: Set $primal := 1$, $\beta = 1/\lambda_d$
Backtrack α_p until $\mathcal{L}_{\beta,\mu}(p + \alpha_p \Delta p, y) - \mathcal{L}_{\beta,\mu}(p, y) \leq \eta \alpha_p \langle \nabla_p \mathcal{L}_{\beta,\mu}(p, y), \Delta p \rangle$
Set $\hat{p} := p + \alpha_p \Delta p$
l3: *If* $\|\nabla_p \mathcal{L}_{\beta,\mu}(\hat{p}, y)\| \leq \min\{\tau \|\rho(\hat{p})\|, \beta/k\}$, and $y + \beta\rho(\hat{p}) \geq \delta\mu \hat{W}^{-1} e$, *then*
l4: $\hat{y} := y + \beta\rho(\hat{p})$
l5: *If* $\nu_\mu(\hat{z}) \leq q r_\mu$, *then*
Set $z := \hat{z}$, $r_\mu := \nu_\mu(\hat{z})$, $k := k + 1$, $primal := 0$
If $\nu(\hat{z}) \leq qr$, *then*
Set $r := \nu(\hat{z})$, $\mu := \min\{\delta r, r^2\}$, $\lambda_d := \min\{\nu_\mu(\hat{z}), \frac{1}{\beta}\}$, $r_\mu = \nu_\mu(\hat{z})$
Else
l6: Set $p := \hat{p}$, $\beta := 2\beta$, $\lambda_d := \frac{1}{\beta}$
Else
Set $p := \hat{p}$
EOL (End of the loop)
OUTPUT z

Fig. 1. IPM algorithm.

(Remark 1). The function $\mathcal{L}_{\beta,\mu}(x, w, y)$ is continuous on $(x, w) \in \mathbb{R}^n \times \mathbb{R}^m_{++}$ therefore to prove the lemma it is enough to show the following set

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

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

$$\lim_{l\to\infty} \mathcal{L}_{\beta,\mu}(x^l, w^l, y) = \infty, \tag{25}$$

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 (25) it is enough to show that

$$\lim_{l\to\infty} \sum_{i=1}^m \varphi_i^l = \infty. \tag{26}$$

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 corresponding subsequences). Since for any index $i = 1, \ldots, m$ the sequence $\{h_i(x^l)\}$ is bounded, we have $\lim_{l\to\infty} \rho_i^l = \infty$ 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 (26) 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 $1 \leq i \leq m$. In case

$h_i(x^l) \leq 1$, then

$$\varphi_i^l = \frac{\beta}{2}{\rho_i^l}^2 + y_i\rho_i^l - \mu\log(h_i(x^l) + \rho_i^l)$$
$$\geq \frac{\beta}{2}{\rho_i^l}^2 + y_i\rho_i^l - \mu\log(1 + \rho_i^l) \geq -B_1 \qquad (27)$$

for some $B_1 \geq 0$ large enough.

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

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

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

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

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

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

$$\geq -\mu\log\left(\left|\max_{1\leq i\leq m} h_i(x^l)\right| + 1\right) - B_2$$
$$\geq -\mu(C - \min_{1\leq i\leq m} h_i(x^l))^2 - B_2$$
$$= -\mu(C - h_{i_0}(x^l))^2 - B_2$$
$$\geq \begin{cases} -\mu C^2 - B_2, & \text{if } h_{i_0}(x^l) \geq 0 \\ -\mu(C + \rho_{i_0}^l)^2 - B_2, & \text{if } h_{i_0}(x^l) < 0 \end{cases}$$
$$\geq -\mu\max\left\{C^2, (C + \rho_{i_0}^l)^2\right\} - B_2,$$

where $i_0(x) \in \text{Argmin}_{1\leq i\leq m} h_i(x)$ and $i_0 = i_0(x^l)$.

Keeping in mind that $w_{i_0}^l > 0$, it follows from (5) that

$$\lim_{l\to\infty} \rho_{i_0}^l = +\infty.$$

Hence for all sequence numbers l large enough we have

$$\varphi_i^l \geq -\mu(C + \rho_{i_0}^l)^2 - B_2. \qquad (28)$$

Combining (27) and (28), we obtain for l large enough (that $h_{i_0}(x^l) < 0$)

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

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

The inequality $\beta > 2\mu m$ guarantees that for such β condition (26) holds. Thus, condition (25) 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 2 *Following the proof of Lemma 2 we can show that there exists a global minimum*

$$S_\infty = \min_{x\in\mathbb{R}^n,\, w\in\mathbb{R}^m_+} \|\rho(x, w)\|^2 > -\infty, \qquad (29)$$

and that any set

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

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 $\lambda_d = \frac{1}{\beta}$, is a descent direction for $\mathcal{L}_{\beta,\mu}(p, y)$ and*

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

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

$$\begin{bmatrix} H(x,y) & 0 & -A(x)^T \\ 0 & Y & W \\ A(x) & -I & \frac{1}{\beta}I \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta w \\ \Delta y \end{bmatrix} = \begin{bmatrix} -\nabla f(x) + \nabla h(x)^T y \\ \mu e - WYe \\ -h(x) + w \end{bmatrix} \qquad (30)$$

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

$$\begin{bmatrix} H + \beta A^T A & -\beta A^T \\ -\beta A & W^{-1}Y + \beta I \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta w \end{bmatrix} = \begin{bmatrix} -\sigma + \beta A^T\rho \\ \gamma - \beta\rho \end{bmatrix} \qquad (31)$$

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 increases λ_p such that the smallest eigenvalue of matrix N_β exceeds parameter $\lambda_0 > 0$.), we have by Lemma A1 from the Appendix

$$\begin{aligned} &\begin{bmatrix} \nabla_x \mathcal{L}_{\beta,\mu} \\ \nabla_w \mathcal{L}_{\beta,\mu} \end{bmatrix}^T \begin{bmatrix} \Delta x \\ \Delta w \end{bmatrix} = \\ &- \begin{bmatrix} \Delta x \\ \Delta w \end{bmatrix}^T \begin{bmatrix} H + \beta A^T A & -\beta A^T \\ -\beta A & W^{-1} Y + \beta I \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta w \end{bmatrix} \\ &\le -\alpha \max\{\|\Delta x\|, \|\Delta w\|\}^2, \end{aligned} \tag{32}$$

where α depends only on parameters λ_0, β and $\|A(x)\|$. 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

$$\|D^{-1}(z^*)\| \le M^*. \tag{33}$$

Proof. The proof is straightforward (see e.g. [8]).

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*

$$L_1\|z - z^*\| \le \nu(z) \le L_2\|z - z^*\|. \tag{34}$$

Proof. Keeping in mind that $\nu(z^*) = 0$ the right inequality (34) follows from Lipschitz continuity of $\nu(z)$ on the bounded set Ω_{ε_0}. Therefore there exists $L_2 > 0$ such that

$$\nu(z) \le L_2\|z - z^*\|.$$

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

$$\|\sigma\| \le \nu(z), \tag{35}$$

$$WYe \le \nu(z), \tag{36}$$

$$\|\rho\| \le \nu(z). \tag{37}$$

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

$$\begin{aligned} \sigma(z) =& \sigma(z^*) + H(x^*, y^*)(x - x^*) - A^T(x^*)(y - y^*) \\ &+ \mathcal{O}\|x - x^*\|^2 \\ WYe =& W^*Y^*e + Y^*(w - w*) + W^*(y - y^*) \\ &+ \mathcal{O}\|w - w^*\|\|y - y^* \\ -\rho(z) =& -\rho(z^*) + A^T(x^*)(x - x^*) - (w - w^*) \\ &+ \mathcal{O}\|x - x^*\|^2. \end{aligned}$$

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 $\|D^{-1}(z^*)\| \le M^*$. Therefore we have

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

Choosing $L_1 = 1/(2M^*)$, we obtain the left inequality (34), i.e.

$$L_1\|z - z^*\| \le \nu(z).$$

Lemma 5 is proven.

Also, we need the following Banach Lemma (see e.g. [10] for a proof).

Lemma 6 *Let matrix $A \in \mathbb{R}^{n,n}$ be nonsingular and $\|A^{-1}\| \le M$. Then there exists $\varepsilon > 0$ small enough such that any matrix $B \in \mathbb{R}^{n,n}$ satisfying $\|A - B\| \le \varepsilon$ is nonsingular and the following bound holds*

$$\|B^{-1}\| \le 2M.$$

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 $\lambda_d \le \varepsilon_0$ the matrix*

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

has an inverse and its norm satisfies

$$\|D_{\lambda_d}^{-1}(z)\| \le M_2. \tag{38}$$

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

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

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

$$\|D_{\lambda_d}(z)^{-1}\| \le M_2.$$

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

The following assertion is a slight modification of the Debreu theorem [7].

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 $\theta > 0$ such that $\xi^T H\xi \ge \theta\xi^T\xi$, $\forall\xi : A\xi = 0$. Then there exists $k_0 > 0$ large enough that for any $0 < \theta_1 < \theta$ the inequality*

$$\xi^T\left(H + kA^T\Lambda A\right)\xi \ge \theta_1\xi^T\xi, \quad \forall\xi \in \mathbb{R}^n \tag{39}$$

holds for any $k \ge k_0$.

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 $z = (x, w, y) \in \Omega_{\varepsilon_0}(z^*)$, $z \ne z^*$, $\lambda_d = \nu_\mu(z)$ and $\mu = \min\{\delta\nu(z),\, \nu(z)^2\}$, the matrix $N_{\lambda_d}(x, y, w)$ is positive definite.*

Proof. Let us assume that the active constraint set at x^* is $I^* = \{i : h_i(x^*) = 0\} = \{1, \ldots, r\}$. Also, we consider the vector function $h_{(r)}^T(x) = (h_1(x), \ldots, h_r(x))$ and 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

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

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 $k \ge k_0$ and therefore the matrix $M(x, y)$ remains positive definite in some ε_0 neighborhood of the solution (x^*, y^*).

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

$$\begin{aligned} N_{\lambda_d}(x, y, w) = H(x, y) +& \\ A_{(r)}(x)^T\left[W_{(r)}Y_{(r)}^{-1} + \lambda_d I\right]^{-1}A_{(r)}(x)& \\ +A_{(m-r)}(x)^T\left[W_{(m-r)}Y_{(m-r)}^{-1} + \lambda_d I\right]^{-1}A_{(m-r)}(x),& \end{aligned} \tag{40}$$

where the second and the third terms correspond to active and inactive constraints. Keeping in mind (34), we have

$$\lambda_d = \nu_\mu(z) \le (1 + \delta)\nu(z) \le L_2(1 + \delta)\varepsilon_0.$$

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

$$\left[W_{(r)}Y_{(r)}^{-1} + \lambda_d I\right]^{-1} \ge \frac{\tau_a}{1 + 2\tau_a(1 + \delta)L_2}\varepsilon_0^{-1}I_{(r)}, \tag{41}$$

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

The third term of (40) corresponding to the inactive constraints is positive semi-definite. Therefore, by choosing $\varepsilon_0 > 0$ small enough we can make the elements of the diagonal $\left[W_{(r)}Y_{(r)}^{-1} + \lambda_d I\right]^{-1}$ as large as necessary. Therefore the positive definiteness of the matrix $N_{\lambda_d}(x, y, w)$ follows from the result that $M(x, y)$ and $N_{\lambda_d}(x, y, w) - M(x, y)$ are positive definite provided that ε_0 is sufficiently small.

Remark 3 *It follows from Lemma 8 that in the neighborhood of the solution the interior-point algorithm does not perform the primal regularization of the Hessian $H(x, y)$ when solve the system (11) for finding the primal-dual directions.*

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 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*

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

where $\hat{z}$ is the next primal-dual approximation obtained by formulas (14)-(16) and $c > 0$.

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 $\|z - z^*\| = \varepsilon \le \varepsilon_0$. For ε_0 small enough and using (34), we have

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

It follows from formulas (34), (38) 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 $\Delta z = -D_{\lambda_d}(z)^{-1}b_\mu(z)$, then keeping in mind (38), 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

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

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

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

or

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

By Assumption (A3) 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 $i : \Delta w_i < 0$ 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 (34) 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 $\alpha_i = \alpha_p$, $i = 1, \ldots, n+m$ and $\alpha_i = \alpha_d$, $i = n+m+1, \ldots, n+2m$. 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

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

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

$$\begin{aligned}
&\hat{z} - z^* = z - \mathcal{A}D^{-1}_{\lambda_d}(z)b_\mu(z) - z^* \\
&= \mathcal{A}(z - z^*) - \mathcal{A}D^{-1}_{\lambda_d}(z)b_\mu(z) + (I - \mathcal{A})(z - z^*) \\
&= \mathcal{A}D^{-1}_{\lambda_d}(z)(D_{\lambda_d}(z)(z - z^*) - b_\mu(z)) + (I - \mathcal{A})(z - z^*) \\
&= \mathcal{A}D^{-1}_{\lambda_d}(z)[D(z)(z - z^*) - b(z) + \\
&\quad (D_{\lambda_d}(z) - D(z))(z - z^*) + b(z) - b_\mu(z)] \\
&\quad + (I - \mathcal{A})(z - z^*).
\end{aligned}$$

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), (34), (42) and (48), we have

$$\begin{aligned}
\|\hat{z} - z^*\| \le\; & M_2[\|D(z)(z - z^*) - b(z)\| + \|D_{\lambda_d}(z) - D(z)\| \\
& \|z - z^*\| + \|b_\mu(z) - b(z)\|] + c_5\|z - z^*\|^2 \\
& = M_2[c_6\varepsilon^2 + \nu_\mu(z)\varepsilon + \mu] + c_5\varepsilon^2 \\
& \le M_2[c_6\varepsilon^2 + L_2\varepsilon^2 + L_2^2\varepsilon^3 + L_2^2\varepsilon^2] + c_5\varepsilon^2 \\
& \le c\varepsilon^2,
\end{aligned}$$

where $c = M_2(c_6 + 3L_2) + c_5$. 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)-(A5), 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. We considere several possible cases.

Case 1. The approximation z^s is such that the conditions in line l1 of the algorithm (see Figure 1) hold for all $k \ge s$. Such possibility exists due to Lemmas 5, 9 and Remark 3 in some neighborhood of a local or global minimizer. In this case

$$\lim_{k \to \infty} \nu(z^k) = 0.$$

The trajectory of the algorithm corresponds to the pure interior-point method.

Case 2. Either of the last two conditions in line l1 of the algorithm does not hold for some approximation z^s. In this case, the algorithm switches to the unconstrained minimization mode ($primal = 1$.) The penalty parameter $\beta > 0$ is chosen by the rule given in line l2 of the algorithm description (see Figure 1) so that the primal direction Δp is a descent direction for the augmented Lagrangian by Lemma 3, therefore the algorithm descends to an approximation of the first order optimality point of the minimization of the augmented Lagrangian $\mathcal{L}_{\beta,\mu}(p, y^s)$ in p.

Indeed, for any primal-dual point z^{s_k} and fixed $\beta > 0$ from Lemma 2 follows the boundedness of the set $P_{s_k} = \{p : \mathcal{L}_{\beta,\mu}(p, y^{s_k}) \leq \mathcal{L}_{\beta,\mu}(p^{s_k}, y^{s_k})\}$. Let p^{s_k} be a starting point of an unconstrained minimization of $\mathcal{L}_{\beta,\mu}(p, y^{s_k})$ in p. By Lemma 3 the condition number of the matrix in equation (31) is uniformly bounded with respect to $p \in P_{s_k}$ for any fixed Lagrange multipliers y^{s_k} and the penalty parameter β. This boundedness of the condition number and bactracking with the Armijo rule (22) guarantee that the algorithm eventually descends to the approximation $\hat{p}$ of the first order optimality point of the unconstrained minimization $\mathcal{L}_{\beta,\mu}(p, y^s)$ in p (see e.g. [4], Proposition 1.2.2. or [13], Sections 3.1, 3.2). Therefore there exists the iteration number s_{k+1} such that

$$\|\nabla_p \mathcal{L}_{\beta,\mu}(p_{s_{k+1}}, y^{s_k})\| \leq \min\left\{\tau\|\rho(p_{s_{k+1}})\|,\ \beta/k\right\}, \tag{49}$$

The gradient of the augmented Lagrangian then becomes small enough so the conditions in line l3 of the algorithm hold true.

In the following discussion, we assume that this first order optimality point is a minimum (local or global). The case if it is not a minimum is left for the later discussion *(Case 2b)*. After finding an approximation of an unconstrained minimizer of $\mathcal{L}_{\beta,\mu}(p, y^{s_k})$ in p, the algorithm changes the Lagrange multipliers (line l4 of the algorithm) by the formula (23). Let z^{s_k} and $z^{s_{k+1}}$ be two subsequent iterates of the augmented Lagrangian method with the updated Lagrange multipliers: $y^{s_{k+1}} := y^{s_k} + \beta\rho(x^{s_{k+1}}, w^{s_{k+1}})$, $y^s = y^{s_k}$, $s = s_k, \ldots, s_{k+1} - 1$.

Theorem 5 from [14] implies that under Assumptions A4 and A5 there exists a neighborhood of the minimum of the barrier subproblem $\Omega_{\varepsilon_\mu}(z_\mu)$ and the number $\beta_\mu > 0$ such that if $y^{s_k} \in \Omega_{\varepsilon_\mu}(z_\mu)$ and $\beta \geq \beta_\mu$, for the new primal-dual approximation $z^{s_{k+1}}$ the following estimation hold

$$\|z^{s_{k+1}} - z_\mu\| \leq \frac{c_\mu(1+\tau)}{\beta}\|z^{s_k} - z_\mu\|, \tag{50}$$

where $\tau > 0$ is used in condition (49) (line l3) of the algorithm, and $c_\mu > 0$ is a constant depending only on the characteristics of the barrier subproblem (3) at the solution z_μ. The inequality (50) and Lipschitz continuity of $\nu_\mu(z)$ on the bounded set $\Omega_{\varepsilon_\mu}(z_\mu)$ implies that

$$\nu_\mu(z^{s_{k+1}}) \leq q\nu_\mu(z^{s_k}) \tag{51}$$

for β large enough. This can be shown using the considerations similar to those in Lemma 5. Therefore violation of the inequality (51) (line l5) can be due to any of the following reasons: $z^{s_k} \in \Omega_{\varepsilon_\mu}(z_\mu)$, $p^{s_{k+1}}$ is an approximation a minimizer, but β is not large enough (*Case 2a*), $z^{s_k} \notin \Omega_{\varepsilon_\mu}(z_\mu)$ (*Case 2b*). The latter case includes also the situation if p^{s_k} is an approximation of some other than a minimum first order optimality point of the unconstrained minimization of the augmented Lagrangian. According to the algorithm, the case (*2b*) leads to an unbounded increase of the penalty parameter β.

Case 2a. Let $z^{s_k} \in \Omega_{\varepsilon_\mu}(z_\mu)$. Therefore by the Theorem 5 of [14] for the condition (50) to hold, the penalty parameter $\beta > 0$ must be large enough. The algorithm increases β, adjusts the dual regularization parameter (line l6 of the algorithm) and continues the minimization of $\mathcal{L}_{\beta,\mu}(p, y^{s_k})$ in p. Eventually when β becomes large enough and (50) implies (51). Therefore the update of the Lagrange multipliers reduces the value of the merit function $\nu_\mu(z)$ by a chosen factor $0 < q < 1$.

In this case, the minimization of the merit function $\mathcal{L}_{\beta,\mu}(p, y^{s_k})$ in p for a larger β followed by the Lagrange multipliers update attracts the trajectory to the solution to the barrier problem (3). For the value of the merit function $\nu(z)$ the following estimation holds

$$\nu(z^{s_{k+1}}) \leq \nu_\mu(z^{s_{k+1}}) + \mu = \nu_\mu(z^{s_{k+1}}) + \min\{\delta r,\ r^2\}, \tag{52}$$

where r is the previous best value of the merit function $\nu(z)$ and $0 < \delta < 1$. The value of the barrier parameter μ is smaller than the previous best value of the merit function $\nu(z)$ before the 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 reduction of the merit function $\nu(z)$, finding the approximation of a minimizer z_μ followed by the further reduction of the barrier parameter μ eventually brings its trajectory to the neighborhood of some minimizer

$\Omega_{\varepsilon_0}(z^*)$. Then the algorithm converges to the solution by *Case 1* with an asymptotic quadratic rate.

Case 2b. In this case, there is no guarantee that (50) holds and the new approximation $z^{s_{k+1}}$ is close to the minimizer of the barrier subproblem z_μ, i.e the inequality (51) (line l5) may not hold an infinite number of times while the penalty parameter β increases unboundedly. In this case, the algorithm does not change the Lagrange multipliers y by formula (23) since this update does not reduce the value of the merit function $\nu_\mu(z)$. Therefore, the algorithm doubles penalty parameter (line l6 of the algorithm) and eventually 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 penalty parameter β. The vector of the Lagrange multipliers y does not change according to the algorithm. In this case, we show that any limit point of the primal sequence $\{x^s\}$ is actually 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 $v_i(x) = \min\{h_i(x), 0\}$:

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

First we establish that the primal sequence $\{p^s\}$ is bounded. Consider the monotone increasing sequence $2m\mu \le \beta^{s_0} \le \beta^{s_1} \le \ldots \le \beta^{s_k} \le \ldots$. 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 $\mathcal{L}_{\beta^{s_k},\mu}(p,y)$ in p for the monotone nondecreasing sequence $\beta^0 \le \beta^{s_1} \le \ldots \le \beta^{s_k} \le \ldots$ is equivalent to the sequence of unconstrained minimizations of function $\theta_\xi(p,y)$ in p for the monotone nonincreasing sequence $1 = \xi^0 \ge \xi^{s_1} \ge \ldots \ge \xi^{s_k} \ldots > 0$.

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

$$\lim_{l\to\infty} \sup_{0\le s\le l} g_2^s = +\infty. \tag{53}$$

We will show that (53) implies that

$$\lim_{l\to\infty} \inf_{0\le s\le l} g_1^s = -\infty \tag{54}$$

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

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

$$p^0 = p^{s_0}, p^{s_0+1}, \ldots, p^{s_0+d_0} = p^{s_1}, p^{s_1+1}, \ldots, p^{s_1+d_1}$$

$$= \cdots = p^{s_k}, p^{s_k+1}, \ldots, p^{s_k+d_k}, \ldots$$

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

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

or, equivalently

$$g_1^s - g_1^{s+1} \ge \frac{1-\xi^{s_k}}{\xi^{s_k}}(g_2^{s+1} - g_2^s). \tag{55}$$

After the summation of the inequality (55) over all $s = s_k, \ldots, s_k+d_k-1$, we obtain

$$g_1^{s_k} - g_1^{s_k+d_k} \ge \frac{1-\xi^{s_k}}{\xi^{s_k}}(g_2^{s_k+d_k} - g_2^{s_k}). \tag{56}$$

After the summation of the inequality (56) for all $k = 0, 1, \ldots j$ and keeping in mind that $g_1^{s_k+d_k} = g_1^{s_{k+1}}$ and $g_2^{s_k+d_k} = g_2^{s_{k+1}}$ for $k = 0, 1, \ldots, j-1$, we obtain

$$g_1^0 - g_1^{s_j+d_j} \ge \sum_{i=1}^{j} \frac{1-\xi^{s_i}}{\xi^{s_i}}(g_2^{s_i+d_i} - g_2^{s_i}). \tag{57}$$

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

$$\lim_{l\to\infty} \sup_{0\le s\le l} g_2^s = \lim_{l\to\infty} \sup_{0\le s\le l} \sum_{i=1}^{j}(g_2^{s_i+d_i} - g_2^{s_i}) = +\infty.$$

Since the sequence $\{\xi^{s_k}\}$ is monotonically decreasing to zero, the sequence $\left\{\frac{1-\xi^{s_k}}{\xi^k}\right\}$ is monotone, increasing and unbounded and greater than or equal to one starting with $k = k_0$. Without restricting the generality, we can assume that $k_0 = s_1$. Therefore by Lemma (A3) from the Appendix we have

$$\lim_{l\to\infty} \sup_{0\le s\le l} \sum_{i=1}^{j} \frac{1-\xi^{s_i}}{\xi^{s_i}} (g_2^{s_i+d_i} - g_2^{s_i}) = +\infty.$$

Therefore using (57) we obtain

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

or equivalently

$$\lim_{l\to\infty} \inf_{0\le s\le l} g_1^s = -\infty,$$

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

$$\begin{aligned} \nabla f(\hat{x}) - A(\hat{x})^T(y + \beta\rho(\hat{p})) &= 0, \\ -\mu\hat{W}^{-1}e + y + \beta\rho(\hat{p}) &= 0. \end{aligned} \qquad (58)$$

Therefore the only reason that the merit function $\nu_\mu(\hat{z})$ is not zero for the triple $\hat{z} = (\hat{x}, \hat{w}, \hat{y})$, where $\hat{y} = y + \beta\rho$, 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 $s \ge s_0$. 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

$$\begin{aligned} \nabla f(x^{s_k}) - A(x^{s_k})^T(y + \beta^{s_k}\rho(x^{s_k}, w^{s_k})) &= \beta^{s_k}\Upsilon_n^{s_k}, \\ -\mu W_k^{-1}e + y + \beta^{s_k}\rho(x^{s_k}, w^{s_k}) &= \beta^{s_k}\Upsilon_m^{s_k}, \end{aligned} \qquad (59)$$

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

If the primal sequence (x^{s_k}, w^{s_k}) satisfy the system (59), then it satisfies the following system

$$\begin{aligned} &\nabla f(x^{s_k})/\beta^{s_k} - A(x^{s_k})^T y/\beta^{s_k} + A(x^{s_k})\rho(x^{s_k}, w^{s_k})) \\ &= \Upsilon_n^{s_k}, \\ &- \mu/\beta^{s_k} + W^{s_k} y/\beta^{s_k} + W^{s_k}\rho(x^{s_k}, w^{s_k}) \\ &= W^{s_k}\Upsilon_m^{s_k}, \end{aligned} \qquad (60)$$

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

$$\lim_{k\to\infty} A(x^{s_k})\rho(x^{s_k}, w^{s_k}) = 0, \qquad (61)$$

$$\lim_{k\to\infty} (w_i^{s_k} - h_i(x^{s_k}))w_i^{s_k} = 0, \quad i = 1, \ldots, m. \qquad (62)$$

and

$$\lim_{k\to\infty} w_i^{s_k} \ge 0, \quad i = 1, \ldots, m. \qquad (63)$$

It is easy to verify that conditions (61)-(63) are also the first-order optimality conditions for the problem

$$\begin{aligned} &\min \|w - h(x)\|_2^2, \\ &\text{s.t. } w \ge 0. \end{aligned} \qquad (64)$$

and, in turn, for the problem

$$\min\, [V(x)]^2, \quad x \in \mathbb{R}^n.$$

The theorem is proven.

6. Numerical testing

As it follows from Theorem 1, in the worst case, the algorithm minimizes the constrain violation of the nonlinear problem. However, we believe that in most cases the algorithm finds the first order optimality point. To demonstrate this, we implemented the algorithm within LOQO software package and tested the code using the Hock and Schittkowski [11] problems.

We consider differentiable problems only with inequality constraints and bounds. The results are shown in the Tables 1 and 2. In the tables for each problem we show the name of the problem, the number of variables, the number of constraints, the number of iterations, running time on IBM Laptop with Red Hat Linux Fedora Core 2.0, 1GB of main memory and 1.3GHz clock speed, optimal objective value, and if the problem convex or not. The iteration limit was set to 1,000,000 iterations. All the problems were formulated in AMPL.

Out of 65 differentiable problems with inequality constraints the algorithm solved 61 including all 20 convex problems. It follows from the proof of Theorem 1 that in case of solving convex problems with bounded feasible sets the algorithm is guaranteed to find a global minimum of the problem because *Case 2b* in the proof of Theorem 1 is not possible under the imposed assumptions. Further, the algorithm solved 41 out of 45 nonconvex optimization problems.

We would like to mention that among the unsolved problems hs013 violates the assumption A3. For problems hs097, hs098 and hs116 the iterates of the algorithm are attracted to the areas where infeasibility is around 10^{-4}. Such behavior of the algorithms is in agreement with Theorem 1.

7. Concluding remarks

In this paper we analyzed convergence of the primal-dual interior-point algorithm for nonlinear optimization problems with inequality constraints. The important features of the algorithm are the primal and dual regularizations, which guarantee that the algorithm decreases the merit function $\mathcal{L}_{\beta,\mu}(x,w,y)$ in (x,w) in order to drive the trajectory of the algorithm down to the neighborhood of a first order optimality point.

Another important feature of the algorithm is that it stabilizes a sequence of primal iterates in the sense that at the worst case the algorithm finds a first order optimality point of the l_2-norm of the constraint violation without any assumptions on the sequence of primal and dual iterates. Such assumptions have been common in recent convergence proofs.

In the worst case the algorithm can be “trapped” in areas of where the constraints of the problem are inconsistent. For example, the constraints $x^2 - x \geq 0$, $x \geq 10^{-5}$ are inconsistent around $x = 0$. Therefore if an initial guess $x^0 = 0$ the algorithm stays in the neighborhood of 0 and never converges to a feasible point.

We believe that in similar situation, most of interior-point algorithms will generate a sequence of unbounded Lagrange multipliers. Having an assumption of boundedness of iterates simply eliminates such cases. By dropping the assumption of the boundedness of the iterates in this paper we bring these cases into consideration and guarantee that the sequence of primal iterates does not diverge.

The next important step is to generalize the theory for equality constraints and to work on numerical performance of the algorithm. Currently LOQO implements only a primal regularization. Therefore in the future we will 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.

8. Appendix

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

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

is also positive definite with the lower bound for the smallest eigenvalue $\lambda_M > 0$ depending on λ_N, λ_C and $\|B\|$.

Proof. Let the size of A and N be $n \times n$, the sizes of B and C be $m \times n$ and $m \times m$ respectively. 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

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

$$\geq \lambda_N x^T x + x^T B^T C^{-1} B x + y^T C y + 2 y^T B x$$
$$= \lambda_N x^T x + (C^{-1}Bx + y)^T C (C^{-1}Bx + y)$$
$$\geq \lambda_N x^T x + \lambda_C (C^{-1}Bx + y)^T (C^{-1}Bx + y)$$
$$\geq \lambda_{\min} \left(x^T x + (C^{-1}Bx + y)^T (C^{-1}Bx + y) \right)$$
$$= \lambda_{\min} \left(x^T \left((I + (C^{-1}B)^T (C^{-1}B) \right) x + y^T y \right.$$
$$\left. + 2y^T (C^{-1}B) x \right)$$
$$= \lambda_{\min} [x^T y^T] \begin{bmatrix} I + (C^{-1}B)^T (C^{-1}B) & (C^{-1}B)^T \\ C^{-1}B & I \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}$$
$$= \lambda_{\min} [x^T y^T] Q \begin{bmatrix} x \\ y \end{bmatrix} \geq \lambda_{\min} \lambda_Q [x^T y^T] \begin{bmatrix} x \\ y \end{bmatrix}.$$

where $\lambda_{\min} = \min\{\lambda_N, \lambda_C\}$ and λ_Q is the smallest eigenvalue of Q. Keeping in mind that $\lambda_Q^{-1} = \|Q^{-1}\|_2 \leq \sqrt{n+m}\|Q^{-1}\|$, we estimate $\|Q^{-1}\|$:

$$\|Q^{-1}\| = \left\| \begin{bmatrix} I + (C^{-1}B)^T (C^{-1}B) & (C^{-1}B)^T \\ C^{-1}B & I \end{bmatrix}^{-1} \right\|$$

Table 1

Hock and Schittkowski nonlinear problems with inequality constraints

Name	Variables	Constraints	Iterations	Runtime	f(x*)	Type
hs001	2	0	34	0.003	1.03E-20	convex
hs002	2	0	21	0.002	4.941229318	convex
hs003	2	0	6	0.001	4.46E-09	convex
hs004	2	0	7	0	2.666666668	nonconvex
hs005	2	0	8	0	-1.913222955	nonconvex
hs010	2	1	9	0.001	-1	convex
hs011	2	1	8	0.001	-8.498464222	convex
hs012	2	1	11	0.001	-30	convex
hs013	2	1	(IL)			nonconvex
hs015	2	2	38	0.003	306.5000015	nonconvex
hs016	2	2	17	0.001	0.25	nonconvex
hs017	2	2	28	0.002	1.00000001	nonconvex
hs018	2	2	19	0.002	5	nonconvex
hs019	2	2	19	0.002	-6961.814063	nonconvex
hs020	2	3	16	0.001	40.19872981	nonconvex
hs021	2	1	11	0.001	-99.96	convex
hs022	2	2	7	0.001	1.000000035	convex
hs023	2	5	17	0.002	2	nonconvex
hs024	2	2	11	0.001	-1	nonconvex
hs025	3	0	34	0.015	1.15E-18	nonconvex
hs029	3	1	10	0.001	-22.62741655	nonconvex
hs030	3	1	8	0.001	1	convex
hs031	3	1	10	0.001	5.999999375	nonconvex
hs033	3	2	21	0.002	2	nonconvex
hs034	3	2	13	0.001	-0.834032445	convex
hs035	3	1	8	0.001	0.111111117	convex
hs036	3	1	134	0.027	-3300.000002	nonconvex
hs037	3	1	17	0.002	-3456	nonconvex
hs038	4	0	40	0.004	7.26E-24	nonconvex
hs043	4	3	9	0.001	-44	convex
hs044	4	6	13	0.002	-13	nonconvex
hs045	5	0	28	0.003	1	nonconvex
hs057	2	1	2365	1.335	0.030647619	nonconvex

$$= \left\| \begin{bmatrix} I & -(C^{-1}B)^T \\ -C^{-1}B & I + (C^{-1}B)(C^{-1}B)^T \end{bmatrix} \right\|$$

$$\leq 1 + \|(C^{-1}B)^T\| + \|(C^{-1}B)\| + \|(C^{-1}B)(C^{-1}B)^T\|$$

$$= L < +\infty, \text{ if } \|B\| < +\infty.$$

Therefore, we have

$$[x^T y^T] M \begin{bmatrix} x \\ y \end{bmatrix} \geq \lambda_M [x^T y^T] \begin{bmatrix} x \\ y \end{bmatrix},$$

where $\lambda_M \geq \min\{\lambda_N, \lambda_C\}(L\sqrt{n+m})^{-1}$. The lemma is proven.

Lemma A2. Let the numbers $a_1, \dots, a_n$, $n \geq 2$ be such that

$$\sum_{i=1}^{l} a_i \leq 0 \quad \text{for} \quad l = 1, \dots, n-1 \tag{65}$$

and

$$\sum_{i=1}^{n} a_i > 0, \tag{66}$$

the numbers $b_1, \dots, b_n$, $n \geq 2$ such that $1 \leq b_1 \leq b_2 \leq \dots \leq b_n$ then the following estimation holds

$$\sum_{i=1}^{n} a_i b_i \geq \left(\sum_{i=1}^{n} a_i \right) b_1 > 0. \tag{67}$$

Table 2

Hock and Schittkowski nonlinear problems with inequality constraints

Name	Variables	Constraints	Iterations	Runtime	f(x*)	Type
hs059	2	3	53	0.012	-6.749505274	nonconvex
hs064	3	1	19	0.001	6299.84205	convex
hs065	3	1	13	0.001	0.953528857	convex
hs066	3	2	12	0.001	0.518163274	convex
hs070	4	1	22	0.011	0.17517448	nonconvex
hs072	4	2	267	0.028	727.6793469	convex
hs076	4	3	9	0.001	-4.681818182	convex
hs083	5	3	22	0.002	-30665.53897	nonconvex
hs084	5	3	29332	40.357	-5280335.069	nonconvex
hs085	5	36	16748	24.066	-1.905155258	nonconvex
hs086	5	6	13	0.001	-32.3486783	nonconvex
hs088	2	1	810	0.598	1.362656814	nonconvex
hs089	3	1	393	0.633	1.362656814	nonconvex
hs090	4	1	1109	2.257	1.362656814	nonconvex
hs091	5	1	1075	4.408	1.362656777	nonconvex
hs092	6	1	1106	4.846	1.362656767	nonconvex
hs093	6	2	20	0.002	135.0759628	nonconvex
hs095	6	4	3128	3.908	0.01561953	nonconvex
hs096	6	4	2809	1.087	0.01561953	nonconvex
hs097	6	4	(IL)			nonconvex
hs098	6	4	(IL)			nonconvex
hs100	7	4	10	0.001	680.6300599	convex
hs101	7	6	239	0.223	1809.764762	nonconvex
hs102	7	6	222	0.215	911.8805717	nonconvex
hs103	7	6	309	0.225	543.6679738	nonconvex
hs105	8	0	23	0.216	1136.360984	nonconvex
hs106	8	6	44	0.009	7049.248019	nonconvex
hs108	9	13	38	0.007	-0.866025404	nonconvex
hs110	10	0	7	0.001	-45.77846971	convex
hs113	10	8	12	0.001	24.30620911	convex
hs116	13	15	(IL)			nonconvex
hs117	15	5	16	0.003	32.34867896	nonconvex

Proof. First, we notice that for the given numbers $a_1, \ldots, a_n$, we have

$$\sum_{i=l}^{n} a_i > 0 \quad \text{for} \quad l = 1, \ldots, n, \tag{68}$$

otherwise we come to contradiction to (65) and (66).

Also, we notice the specifics of the trivial case: if the numbers $\bar{a}_1$ and $\bar{a}_2$ are such that $\bar{a}_1 + \bar{a}_2 > 0$ and $\bar{a}_2 > 0$, ($\bar{a}_1$ can be either negative or nonnegative) then

$$\bar{a}_1 b_1 + \bar{a}_2 b_2 \geq (\bar{a}_1 + \bar{a}_2) b_1 > 0, \tag{69}$$

if $1 \leq b_1 \leq b_2$.

Using inequality (69) recursively and keeping in mind (68), we have

$$\sum_{i=1}^{k} a_i b_i = \sum_{i=1}^{k-2} a_i b_i + (a_{k-1} b_{k-1} + a_k b_k)$$

$$\geq \sum_{i=1}^{k-2} a_i b_i + (a_{k-1} + a_k) b_{k-1}$$

$$= \sum_{i=1}^{k-2} a_i b_i + \bar{a}_{k-1} b_{k-1}$$

$$= \sum_{i=1}^{k-3} a_i b_i + a_{k-2} b_{k-2} + \bar{a}_{k-1} b_{k-1}$$

$$\geq \sum_{i=1}^{k-3} a_i b_i + (a_{k-2} + \bar{a}_{k-1}) b_{k-2}$$
$$= \sum_{i=1}^{k-3} a_i b_i + \bar{a}_{k-2} b_{k-2} \geq \cdots$$
$$\geq a_1 b_1 + \bar{a}_2 b_2 \geq (a_1 + \bar{a}_2) b_1 = \left(\sum_{i=1}^{k} a_i \right) b_1 > 0,$$

where $\bar{a}_l = a_l + \cdots + a_k > 0$, $l = 1, \ldots, k$. The lemma is proven.

Lemma A3. 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.

$$\lim_{k \to \infty} s_k = +\infty. \quad (70)$$

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 (70). 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 segmented into an infinite number of groups where all the elements of each individual groups are equal.

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 $a_{q_i + 1} \leq 0$, if $a_{q_i + 1} \neq a_{q_{i+1}}$, and $a_{q_i} > 0$, $i = 0, 1, 2, \ldots$. Moreover, we have

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

and

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

Therefore using Lemma A2, 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,* Math. Programming, 99(1), 35-48, 2004.

[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] D.P. Bertsekas, *Nonlinear Programming,* Athema Scientific, Belmont, Massachusetts, 1995.

[5] R. Byrd, J. Gilbert and J. Nocedal, *A trust region method based on interior point techniques for nonlinear programming,* Math. Programming, 89, 149-185, 2000.

[6] L. Chen, D. Goldfarb, *Interior-point l(2)-penalty methods for nonlinear programming with strong global convergence properties*, Math. Programming, 108(1), 1-36, 2006.

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

[8] 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, 507-541, 1996.

[9] A. Forsgren and P. Gill, *Primal-dual interior methods for nonconvex nonlinear programming*, SIAM Journal on Optimization, 8(4), 1132-1152, 1998.

[10] I. Griva, R. Polyak, *Primal-dual nonlinear rescaling method with dynamic scaling parameter update*, Math. Programming, 106(2), 237-259, 2006.

[11] W. Hock and K. Schittkowski, *Test examples for nonlinear programming codes*, volume 187 of Lecture Notes in Economics and Mathematical Systems, Springer Verlag: Heidelberg, 1981.

[12] X. Liu and J. Sun, *A robust primal-dual interior-point algorithm for nonlinear programs*, SIAM Journal on Optimization, 14(4), 1163-1186, 2004.

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

[14] B.T. Polyak, N.V. Tretjakov, *A penalty method for constrained optimization problems*, Journal of Computational Mathematics and Mathematical Physics, 13(1), 34-46.

[15] 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.

[16] M. Ulbrich, S. Ulbrich, and L. N. Vicente, *A globally convergent primal-dual interior-point filter method for nonlinear programming,* Math. Programming, 100(2), 379-410, 2004.

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

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

[19] A. Wächter and L. Biegler, *Line search filter methods for nonlinear programming: motivation and global convergence*, SIAM Jounal on Computing, 16(1), 1-31.

Received 11 May, 2007; accepted 14 February 2008