NUMERICAL ALGEBRA,
CONTROL AND OPTIMIZATION
Volume **1**, Number **2**, June **2011**

doi:10.3934/naco.2011.1.283

pp. **283–299**

PROXIMAL POINT NONLINEAR RESCALING METHOD FOR CONVEX OPTIMIZATION

Igor Griva

Department of Mathematical Sciences
George Mason University
Fairfax, VA 22030, USA

Roman A. Polyak

Departments of Mathematical Sciences and SEOR
George Mason University
Fairfax, VA 22030, USA

(Communicated by Song Wang)

Abstract. Nonlinear rescaling (NR) methods alternate finding an unconstrained minimizer of the Lagrangian for the equivalent problem in the primal space (which is an infinite procedure) with Lagrange multipliers update.

We introduce and study a proximal point nonlinear rescaling (PPNR) method that preserves convergence and retains a linear convergence rate of the original NR method and at the same time does not require an infinite procedure at each step.

The critical component of our analysis is the equivalence of the NR method with dynamic scaling parameter update to the interior quadratic proximal point method for the dual problem in the rescaled from step to step dual space.

By adding the classical quadratic proximal term to the primal objective function the PPNR step can be viewed as a primal-dual proximal point mapping. This allows analyzing a wide variety of non-quadratic augmented Lagrangian methods from unique and general point of view using tools typical for the classical quadratic proximal-point technique.

We proved convergence of the primal-dual PPNR sequence under minimum assumptions on the input data and established a q-linear rate of convergence under the standard second-order optimality conditions.

1. **Introduction.** In this paper we introduce and study a proximal point nonlinear rescaling (PPNR) method, which eliminates the main drawback of the NR methods (see [6]-[9]) the necessity of finding an exact unconstrained minimizer at each step. We show that the PPNR method retains convergence and rate of convergence properties of the original NR method without finding unconstrained minimizer at each step.

The distinct feature of this approach is the equivalence of the NR methods with a "dynamic" scaling parameters update to a interior quadratic proximal point method for the dual problem in the rescaled from step to step dual space. The equivalence allows transforming each step of the PPNR method into a step of the primal-dual

2000 *Mathematics Subject Classification.* 90C25; 90C30.
Key words and phrases. Nonlinear rescaling, primal-dual mapping, proximal point methods.
The research is supported by the NSF grant CCF-0836338.

quadratic proximal point method for finding a saddle point of the classical Lagrangian for the original problem.

The correspondent primal-dual quadratic proximal mapping leads to the rescaled monotone operator. The properties of the rescaled operator were one of our main concerns. The properties allowed extending the results typical for quadratic augmented Lagrangian [11] for a wide class of non-quadratic augmented Lagrangians. Unfortunately such an extension complicates the analysis, but, on the other hand, makes the analysis universal for a wide class of non-quadratic augmented Lagrangians.

The paper is organized as follows. In the next section we state the problem and introduce basic assumptions. In Section 3 we introduce the proximal point nonlinear rescaling algorithm (PPNR). In Section 4 we analyze convergence of the PPNR algorithm under mild assumptions on the input data. In Section 5 we establish an asymptotic q-linear rate of convergence under the standard second-order optimality conditions. We conclude the paper by pointing out some directions of future research.

2. **Statement of the problem and basic assumptions.** Let $f : \mathbb{R}^n \to \mathbb{R}^1$ be convex, all $c_i : \mathbb{R}^n \to \mathbb{R}^1$, $i = 1, \ldots, m$ be concave and twice continuously differentiable functions. We consider a convex set $\Omega = \{x \in \mathbb{R}^n : c_i(x) \geq 0,\ i = 1, \ldots, m\}$ and the following convex optimization problem

$$(\mathcal{P}) \qquad f(x^*) = \min\{f(x) | x \in \Omega\}$$

We assume that:

A. The optimal primal set $X^* = \{x \in \Omega : f(x) = f(x^*)\}$ is not empty and bounded.

B. The Slater's condition holds, i.e. there exists $\hat{x} \in \mathbb{R}^n : c_i(\hat{x}) > 0$, $i = 1, \ldots, m$.

We use the norm $\|x\| = \sqrt{x^T x}$ everywhere in the manuscript unless another norm is explicitly specified.

Under the assumption B, the Karush-Kuhn-Tucker's (K-K-T's) conditions hold true, i.e. there exists vector $y^* = (y_1^*, ..., y_m^*) \in \mathbb{R}^m_+$ such that

$$\nabla_x L(x^*, y^*) = \nabla f(x^*) - \sum_{i=1}^m y_i^* \nabla c_i(x^*) = 0 \tag{1}$$

where $L(x, y) = f(x) - \sum_{i=1}^m y_i c_i(x)$ is the Lagrangian for the primal problem $\mathcal{P}$, and the complementary slackness conditions

$$y_i^* c_i(x^*) = 0,\ i = 1, \ldots, m \tag{2}$$

hold true.

Let us consider also the dual function $d : \mathbb{R}^m_+ \to \mathbb{R}$ which is defined by formula

$$d(y) = \inf_{x \in \mathbb{R}^n} L(x, y)$$

and the dual problem

$$(\mathcal{D}) \qquad d(y^*) = \max\{d(y) | y \in \mathbb{R}^m_+\}.$$

Due to the Slater condition **B** the dual optimal set $Y^* = \{y \in \mathbb{R}^m_+ : d(y) = d(y^*)\}$ is bounded, and due to (2) we have

$$f(x^*) = d(y^*).$$

3. **Proximal point nonlinear rescaling method.** Consider a class Ψ of transformations $\psi : (t_1, t_2) \to (a, b)$, where $-\infty < t_1 < 0 < t_2 < \infty$ and $-\infty < a < 0 < b < \infty$ that satisfy the following properties

1^0 $\psi(0) = 0$.
2^0 a) $\psi'(t) > 0$, b) $\psi'(0) = 1$, c) $\psi'(t) \le a_1(t+1)^{-1}$, $\forall\, t \in (0, \infty)$, $a_1 > 0$.
3^0 $\psi''(t) < 0$, $\forall\, t \in (t_1, t_2)$.
4^0 $-\psi''(t) \ge \alpha e^{-\beta t}$, $\forall\, t \in [1, \infty)$, $\alpha > 0$, $\beta > 0$.

One can verify that well known transformations $\psi_1(t) = 1 - e^{-t}$ [2], $\psi_2(t) = \ln(t+1)$, $\psi_3(t) = t/(t+1)$ [6] belong to Ψ. Other transformations $\psi \in \Psi$ one can find in [8].

For any given vector $\boldsymbol{k} = (k_1, \dots, k_m) \in \mathbb{R}^m_{++}$ due to 1^0 and 2^0 (a) we have

$$c_i(x) \ge 0 \Leftrightarrow k_i^{-1}\psi(k_i c_i(x)) \ge 0, \quad i = 1 \dots, m\,. \tag{3}$$

Therefore, the problem

$$f(x^*) = \min\{f(x) \,|\, k_i^{-1}\psi(k_i c_i(x)) \ge 0,\, i = 1, \dots, m\} \tag{4}$$

is equivalent to the primal problem $\mathcal{P}$.

The Lagrangian $\mathcal{L} : \mathbb{R}^n \times \mathbb{R}^m_{++} \times \mathbb{R}^m_{++} \to \mathbb{R}$ for the equivalent problem, is given by formula

$$\mathcal{L}(x, y, \boldsymbol{k}) = f(x) - \sum_{i=1}^m y_i k_i^{-1}\psi(k_i c_i(x)).$$

We are ready to describe the general PPNR method. Let $x^0 \in \mathbb{R}^n$ and $y^0 \in \mathbb{R}^m_{++}$ are initial primal and dual approximations, $\{k_s > 0\}$ is the nondecreasing sequence.

The PPNR method generates three sequences $\{\boldsymbol{k}^s\} \subset \mathbb{R}^m_{++}$, $\{x^s\} \subset \mathbb{R}^n$, $\{y^s\} \subset \mathbb{R}^m_{++}$, by formulas

$$\begin{aligned} \boldsymbol{k}^s &= (k_i^s = k_s(y_i^s)^{-1}, \quad i = 1, \dots, m), && (5)\\ x^{s+1} &= \arg\min\{\mathcal{L}(x, y^s, \boldsymbol{k}^s) + \frac{1}{2k_s}\|x - x^s\|^2 \,|\, x \in \mathbb{R}^n\} && (6)\\ y^{s+1} &= (y_i^{s+1} = y_i^s\psi'(k_i^s c_i(x^{s+1})), \quad i = 1, \dots, m.) && (7) \end{aligned}$$

The numerical realization of the PPNR method (5)-(7) requires replacing the primal minimizer with its approximation.

To simplify the presentation we use the same notation for the PPNR method (5)-(7) and its modification (8)-(10). To avoid confusion we specify for each statement the method that generates the sequences $\{\boldsymbol{k}^s\}$, $\{x^s\}$, $\{y^s\}$.

Let $\{\varepsilon_s > 0\}$ be a sequence such that $\sum_{s=0}^{\infty}\varepsilon_s < \infty$. Then the modified PPNR method generates the following three sequences $\{\boldsymbol{k}^s\} \subset \mathbb{R}^m_{++}$, $\{x^s\} \subset \mathbb{R}^n$, $\{y^s\} \subset \mathbb{R}^m_{++}$,

$$\boldsymbol{k}^s = (k_i^s = k_s(y_i^s)^{-1}, \quad i = 1, \dots, m), \tag{8}$$

$$x^{s+1} \,:\, \|\nabla_x\mathcal{L}(x^{s+1}, y^s, \boldsymbol{k}^s) + \frac{1}{k_s}(x^{s+1} - x^s)\| \le \frac{\varepsilon_s}{k_s} \tag{9}$$

$$y^{s+1} = (y_i^{s+1} = y_i^s\psi'(k_i^s c_i(x^{s+1})), \quad i = 1, \dots, m.) \tag{10}$$

The minimizer x^{s+1} given by (6) or its approximation (9) exists and it is uniquely defined for any given $y^s \in \mathbb{R}^m_{++}$ and $\boldsymbol{k}^s \in \mathbb{R}^m_{++}$ due to convexity f, concavity c_i, properties $1^0 - 4^0$ of $\psi \in \Psi$ and the proximal term $\frac{1}{2k_s}\|x - x^s\|^2$ in (6).

Due to 2^0a) we have $y^s \in \mathbb{R}^m_{++} \Rightarrow y^{s+1} \in \mathbb{R}^m_{++}$, therefore the modified PPNR method (8)-(10) is well defined.

The formula (5) has been introduced in [14] for the exponential multipliers method and used on several occasions (see [1, 7, 8]) in the framework of NR methods. In particular, the equivalence of the general NR multipliers method with a "dynamic" scaling parameter update to the interior quadratic proximal point method in the rescaled dual space plays a critical role in the convergence analysis of the NR method (see [8]).

Our analysis extends the convergence scheme, developed in [11, 12] for the quadratic augmented Lagrangians, to a wide class of nonquadratic augmented Lagrangians. The key ingredient of our analysis is the equivalence of the modified PPNR method (8)-(10) to a quadratic proximal point method for finding a saddle point of the classical Lagrangian for the original problem. The primal proximal term $\frac{1}{2k_s}\|x - x^s\|^2$ is standard while the dual proximal term is the quadratic term in the rescaled dual space.

In the following section we prove convergence of the modified PPNR method (8)-(10) under mild assumptions on the input data.

4. **Convergence of the modified PPNR Method.**

In this section we establish convergence properties of the modified PPNR method (8) –(10). The convergence of the PPNR method (5)–(7) follows from the convergence the PPNR method (8)–(10) if $\varepsilon_s = 0$ $s = 0, 1, \ldots$.

The following several lemmas outline convergence properties of the method (8)–(10).

Lemma 4.1. *One step of the PPNR method (5)–(7) is equivalent to finding a saddle point (x^{s+1}, y^{s+1}) of the following function*

$$M(x, y, x^s, y^s) = L(x, y) + \frac{1}{2k_s}\|x - x^s\|^2 - \frac{1}{2k_s}\|y - y^s\|^2_{R^{-1}_{s+1}}$$

where $\|y\|^2_{R^{-1}_{s+1}} = y^T R^{-1}_{s+1} y$, and R^{-1}_{s+1} is a diagonal matrix with positive entries, which we will specify later; i.e.

$$(x^{s+1}, y^{s+1}) : \max_{y \in \mathbb{R}^m_+} \min_{x \in \mathbb{R}^n} M(x, y, x^s, y^s) = \min_{x \in \mathbb{R}^n} \max_{y \in \mathbb{R}^m_+} M(x, y, x^s, y^s). \tag{11}$$

Proof. The saddle point (x^{s+1}, y^{s+1}) exists and unique because $M(x, y, x^s, y^s)$ is strongly convex in x and strongly concave in y. Therefore the approximation (x^{s+1}, y^{s+1}) can be found by solving for (x, y) the following system of equations

$$\nabla_x M(x, y, x^s, y^s) = \nabla_x L(x, y) + \frac{1}{k_s}(x - x^s) = 0, \tag{12}$$

$$\nabla_y M(x, y, x^s, y^s) = -c(x) - \frac{1}{k_s} R^{-1}_{s+1}(y - y^s) = 0. \tag{13}$$

We emphasize that we do not suggest finding (x^{s+1}, y^{s+1}) from (12) – (13) because such a step requires knowledge of R_{s+1}. Our purpose is to show the equivalence of (12) – (13) to (5) – (7), which is critical for further analysis.

Let us show that the solution of (12) – (13) produce the same pair of vectors (x^{s+1}, y^{s+1}) as (5)–(7). Since $\mathcal{L}(x, y^s, \boldsymbol{k}^s) + \frac{1}{2k_s}\|x - x^s\|^2$ is strongly convex in x, finding x^{s+1} from (6) is equivalent to solving the following system for x

$$\nabla_x \mathcal{L}(x, y^s, \boldsymbol{k}^s) + \frac{1}{k_s}(x - x^s) = 0.$$

Keeping in mind (5)–(7) we have

$$\begin{aligned} & \nabla_x \mathcal{L}(x^{s+1}, y^s, \boldsymbol{k}^s) + \frac{1}{k_s}(x^{s+1} - x^s) \\ = \; & \nabla f(x^{s+1}) - \sum_{i=1}^{m} y_i^s \psi'(k_i^s c_i(x^{s+1}) \nabla c_i(x^{s+1}) + \frac{1}{k_s}(x^{s+1} - x^s) \\ = \; & \nabla f(x^{s+1}) - \sum_{i=1}^{m} y_i^{s+1} \nabla c_i(x^{s+1}) + \frac{1}{k_s}(x^{s+1} - x^s) \\ = \; & \nabla_x L(x^{s+1}, y^{s+1}) + \frac{1}{k_s}(x^{s+1} - x^s) = 0. \end{aligned} \tag{14}$$

Also, using (5), (7), property $2^0 b)$ of the transformation $\psi(t)$ and the mean value theorem we have

$$y_i^{s+1} - y_i^s = y_i^s(\psi'(k_i^s c_i(x^{s+1}) - \psi'(0)) = y_i^s \psi''(\theta_i^s k_i^s c_i(x^{s+1})) k_i^s c_i(x^{s+1}) =$$

$$k_s \psi''(\theta_i^s k_i^s c_i(x^{s+1})) c_i(x^{s+1}), \quad i = 1, \ldots, m,$$

which is equivalent to

$$-c(x^{s+1}) - \frac{1}{k_s} R_{s+1}^{-1}(y^{s+1} - y^s) = 0, \tag{15}$$

where $R_{s+1} = \operatorname{diag}(r_i^{s+1})_{i=1}^m$, $r_i^{s+1} = -\left[\psi''(\theta_i^s k_i^s c_i(x^{s+1}))\right] > 0$ and $0 < \theta_i^s < 1$.

Therefore if (x^{s+1}, y^{s+1}) satisfies (5)–(7) then (x^{s+1}, y^{s+1}) is the solution of the minimax problem (11). Also, the solution of the minimax problem (11) satisfies (5)–(7) because (x^{s+1}, y^{s+1}) is unique. □

Keeping in mind that $\nabla_y L(x, y) = -c(x)$ we obtain from (14) and (15) that finding (x^{s+1}, y^{s+1}) from (5)–(7) is equivalent to solving for (x, y) the following system

$$k \nabla_x L(x, y) + (x - x^s) = 0, \tag{16}$$

$$-kR \nabla_y L(x, y) + (y - y^s) = 0 \tag{17}$$

with $k = k_s$ and $R = R_{s+1}$. Therefore by introducing an operator $T_R : \mathbb{R}^n \times \mathbb{R}^m_{++} \to \mathbb{R}^n \times \mathbb{R}^m_{++}$ defined by the formula

$$T_R(z) = \begin{pmatrix} \nabla_x L(x, y) \\ -R \nabla_y L(x, y) \end{pmatrix}.$$

we can rewrite the system (16) – (17) as follows

$$(I + kT_R)(z) = z^s,$$

$z = (x, y)$, where the nonlinear operator $I + kT_R$ is the sum of the identity operator I and nonlinear operator kT_R.

Thus one can view the solution of the proximal minimax problem (11) as an application of the inverse operator $P_R(z) = (I + kT_R)^{-1}$ to z^s. Therefore the sequence generated by the PPNR method (5)–(7) can be described as

$$z^{s+1} = P_{R_{s+1}}(z^s), \; s = 0, 1, \ldots$$

with a starting point $z^0 = (x^0, y^0)$, $x^0 \in \mathbb{R}^n$, $y^0 \in \mathbb{R}^m_{++}$.

The operators T_R and P_R are defined on $\mathbb{R}^n \times \mathbb{R}^m_{++}$. For convergence analysis, however, we need these operators to be defined on $\mathbb{R}^n \times \mathbb{R}^m_{+}$. We define the generalized operators $T_R(z)$ and consequently P_R in the following way:

$$T_R(z) = \begin{pmatrix} \nabla_x L(x,y) \\ -R\nabla_y L(x,y) \end{pmatrix}, \quad P_R(z) = (I + kT_R)^{-1}(z)$$

where $R = \operatorname{diag}(r_i)_{i=1}^m$, and $y_i = 0 \Rightarrow r_i = 0$.

The modified operators described above satisfy the following properties at any primal-dual solution $z^* = (x^*, y^*) \in X^* \times Y^*$:

$$T_R(z^*) = 0,$$

and

$$P_R(z^*) = z^*. \tag{18}$$

The convergence proof of the modified PPNR method (8)–(10) relies on properties of the family of proximal operators $P_{R_{s+1}}$, $s = 1, 2, \ldots$, where $R_{s+1} = \operatorname{diag}(r_i^{s+1})_{i=1}^m$ and

$$r_i^{s+1} = \begin{cases} -\left[\psi''(\theta_i^s k_i^s c_i(x^{s+1}))\right], & 0 < \theta_i^s < 1, \quad \text{if} \quad y_i > 0 \\ 0, & \qquad\qquad\quad\ \text{if} \quad y_i = 0 \end{cases}$$

The latter, in turn, are closely related to properties of monotone operators $T_{R_{s+1}}$. Our next goal is proving that $T_{R_{s+1}}(z)$ are monotone operators for all $s \geq 0$.

An operator $T : \mathbb{R}^p \to \mathbb{R}^p$, which maps $z \in \mathbb{R}^p$ into a closed bounded convex set $T(z)$ is a *monotone operator* if

$$(z - z')^T(w - w') \geq 0 \quad \text{whenever} \quad w \in T(z),\ w' \in T(z').$$

Let $F : X \times Y \to \mathbb{R}$ be convex in $x \in X$ and concave in $y \in Y$, $X \subset \mathbb{R}^n$, $Y \subset \mathbb{R}^m$. The subdifferential of F at (x, y) is defined as follows

$$\partial F(x,y) = \{(u,v) : F(x',y) - F(x,y) \geq u^T(x'-x),$$
$$F(x,y') - F(x,y) \leq v^T(y'-y),\ \forall x' \in X, y' \in Y\}.$$

It is known (see [13]) that the mapping $T : (x,y) \to \{(u,v) : (u,-v) \in \partial F(x,y)\}$ is a maximal monotone operator.

Let $R = \operatorname{diag}(r_i)_{i=1}^{n+m}$ be a diagonal matrix with $r_i \geq 0$, $i = 1, \ldots, n+m$. We consider a mapping $RT : (x,y) \to \{R(u,v)^T : (u,-v) \in \partial F(x,y)\}$, which one obtains by multiplying an image of T by the diagonal matrix R.

Generally speaking, RT is not necessarily a monotone operator for a convex in x and concave in y function F. However, if F is smooth enough then RT is a monotone operator. We show this in the following lemma.

Lemma 4.2. *Let twice continuously differentiable function $F : \mathbb{R}^n \times \mathbb{R}^m_+ \to \mathbb{R}$ be convex in x and concave in y. Then the rescaled pseudo-gradient*

$$R\nabla F(z) = R\begin{pmatrix} \nabla_x F(x,y) \\ -\nabla_y F(x,y) \end{pmatrix}$$

is a monotone operator.

Proof. For a vector $a \in \mathbb{R}^n \times \mathbb{R}^m_+$ consider a scalar function $\varphi_a : \mathbb{R}^n \times \mathbb{R}^m_+ \to \mathbb{R}$ defined as follows

$$\varphi_a(z) \equiv \varphi_a(x,y) = (R\nabla F(x,y))^T a.$$

Let $z = (x, y) \in \mathbb{R}^n \times \mathbb{R}^m_+$ and $z' = (x', y') \in \mathbb{R}^n \times \mathbb{R}^m_+$ and $a \in \mathbb{R}^n \times \mathbb{R}^m_+$. Then using the mean value formula for the scalar function φ_a one can find $0 \leq \theta \leq 1$ such that

$$\varphi_a(z) - \varphi_a(z') = \nabla_z \varphi_a(z' + \theta(z - z'))^T(z - z'),$$

where $\nabla_z \varphi_a(z) = R\nabla^2 F(z)a$, and

$$\nabla^2 F(z) = \nabla^2 F(x, y) = \begin{bmatrix} \nabla^2_{xx} F(x, y) & \nabla^2_{xy} F(x, y) \\ -\nabla^2_{yx} F(x, y) & -\nabla^2_{yy} F(x, y) \end{bmatrix}.$$

In other words for a given $z = (x, y)$, $z' = (x', y')$ and $a = z - z'$ there exists $0 \leq \theta \leq 1$ such that

$$(R\nabla F(z) - R\nabla F(z'))^T(z - z') = \left\langle R\nabla^2 F(z' + \theta(z - z'))(z - z'), (z - z') \right\rangle.$$

The matrix $\nabla^2 F(z)$ is positive semidefinite for any $z = (x, y)$ because $F(x, y)$ is convex in x and concave in y. Therefore $R\nabla^2 F(z)$ is also positive semidefinite, i.e.

$$(R\nabla F(z) - R\nabla F(z'))^T(z - z') \geq 0$$

for any z and z'. Therefore the rescaled pseudo-gradient $R\nabla F(x, y)$ is a monotone operator. □

The Largangian $L(x, y)$ satisfies the conditions of Lemma 4.2. Therefore it follows from Lemma 4.2 that for a given vector $r \in \mathbb{R}^m_+$ and $R = \text{diag}\,(r_i)_{i=1}^m$ the operator $T_R : \mathbb{R}^n \times \mathbb{R}^m_+$ defined by a formula

$$T_R(z) = \begin{pmatrix} \nabla_x L(x, y) \\ -R\nabla_y L(x, y) \end{pmatrix},$$

is a monotone operator. i.e. the inequality

$$(T_R(z_1) - T_R(z_2))^T(z_1 - z_2) \geq 0 \tag{19}$$

holds for any pair $(z_1, z_2) : z_1 \in \mathbb{R}^n \times \mathbb{R}^m_+, z_2 \in \mathbb{R}^n \times \mathbb{R}^m_+$.

Let us show that the proximal operator $P_{R_{s+1}}$ associated with the monotone operator $T_{R_{s+1}}$ is nonexpansive.

Lemma 4.3. *For any vectors a, $b \in \mathbb{R}^n \times \mathbb{R}^m_+$ and the operator $P_{R_{s+1}}$ the following estimation holds*

$$\|P_{R_{s+1}}(a) - P_{R_{s+1}}(b)\| \leq \|a - b\|, \quad \forall s \geq 0.$$

Proof. Let $\hat{a} = P_{R_{s+1}}(a)$ and $\hat{b} = P_{R_{s+1}}(b)$. We use the fact that $\hat{a}$ and $\hat{b}$ are uniquely defined and can be found from the system

$$k_s T_{R_{s+1}}(\hat{a}) + \hat{a} = a,$$

$$k_s T_{R_{s+1}}(\hat{b}) + \hat{b} = b.$$

Therefore

$$k_s \left(T_{R_{s+1}}(\hat{a}) - T_{R_{s+1}}(\hat{b}) \right) + (\hat{a} - \hat{b}) = a - b$$

and hence

$$(k_s)^2 \|T_{R_{s+1}}(\hat{a}) - T_{R_{s+1}}(\hat{b})\|^2 + 2k_s \left(T_{R_{s+1}}(\hat{a}) - T_{R_{s+1}}(\hat{b}) \right)^T (\hat{a} - \hat{b}) + \|\hat{a} - \hat{b}\|^2 = \|a - b\|^2$$

Keeping in mind (19) we have

$$(k_s)^2 \|T_{R_{s+1}}(\hat{a}) - T_{R_{s+1}}(\hat{b})\|^2 + 2k_s \left(T_{R_{s+1}}(\hat{a}) - T_{R_{s+1}}(\hat{b}) \right)^T (\hat{a} - \hat{b}) \geq 0,$$

therefore

$$\|\hat{a} - \hat{b}\| \leq \|a - b\|.$$

□

Lemma 4.4. *For the modified PPNR method (8)–(10) the following estimation holds*

$$\|z^{s+1} - P_{R_{s+1}}(z^s)\| \leq \varepsilon_s.$$

Proof. Consider the vector

$$N_{s+1}(z) = T_{R_{s+1}}(z) + (k_s)^{-1}(z - z^s).$$

Note that

$$z^s + k_s N_{s+1}(z^{s+1}) = z^{s+1} + k_s T_{R_{s+1}}(z^{s+1}) = (I + k_s T_{R_{s+1}})(z^{s+1}).$$

Therefore

$$z^{s+1} = (I + k_s T_{R_{s+1}})^{-1}(z^s + k_s N_{s+1}(z^{s+1})) = P_{R_{s+1}}(z^s + k_s N_{s+1}(z^{s+1})).$$

Due to (9) and (15) we have $\|N_s(z^{s+1})\| \leq \frac{\varepsilon_s}{k_s}$. Therefore since the operator $P_{R_{s+1}}$ is nonexpansive (Lemma 4.3), we have

$$\begin{aligned} \|z^{s+1} - P_{R_{s+1}}(z^s)\| =& \|P_{R_{s+1}}(z^s + k_s N_{s+1}(z^{s+1})) - P_{R_{s+1}}(z^s)\| \\ \leq& \|k_s N_{s+1}(z^{s+1})\| \leq \varepsilon_s. \end{aligned} \tag{20}$$

□

The next lemma establishes the boundedness of the sequence generated by the modified PPNR method (8)–(10).

Lemma 4.5. *The sequence $\{z^s = (x^s, y^s)\}$ generated by the modified PPNR method (8)–(10) is bounded, i.e. there exists $M > 0$ that*

$$\|z^s\| \leq M < \infty. \tag{21}$$

Proof. Due to Assumptions **A** and **B** there exists $z^* = (x^*, y^*) \in X^* \times Y^*$ that KKT's conditions (1)–(2) hold. Therefore for all $s \geq 0$ we have $P_{R_{s+1}}(z^*) = z^*$. By Lemmas 4.3, 4.4, triangle inequality and (18), we have

$$\|z^{s+1} - z^*\| - \varepsilon_s \leq \|P_{R_{s+1}}(z^s) - z^*\| = \|P_{R_{s+1}}(z^s) - P_{R_{s+1}}(z^*)\| \leq \|z^s - z^*\|.$$

Keeping in mind that $\sum_{s=0}^{\infty} \varepsilon_s < +\infty$ we can find such $\alpha > 0$ that

$$\|z^s - z^*\| \leq \|z^0 - z^*\| + \sum_{l=0}^{s-1} \varepsilon_k \leq \|z^0 - z^*\| + \alpha.$$

Therefore $\{z^s\}_{s=0}^{\infty}$ is bounded. □

Let us consider the mapping $Q_{R_{s+1}} = I - P_{R_{s+1}}$, then for any $s \geq 0$ we have

$$Q_{R_{s+1}}(z^*) = 0,$$

where z^* is the solution to the problem $\mathcal{P}$.

We need the following auxiliary statements, which will be used later.

Lemma 4.6. *For the operators $P_{R_{s+1}}$ and $Q_{R_{s+1}}$, $s = 0, 1, 2 \dots$ the following is true*

a) $z = P_{R_{s+1}}(z) + Q_{R_{s+1}}(z)$ *for all* z.

b) $(k_s)^{-1} Q_{R_{s+1}}(z) = (k_s)^{-1}(z - P_{R_{s+1}}(z)) = T_{R_{s+1}}(P_{R_{s+1}}(z))$ *for all* z.

c) $(P_{R_{s+1}}(z) - P_{R_{s+1}}(z'))^T (Q_{R_{s+1}}(z) - Q_{R_{s+1}}(z')) \geq 0$ *for all* z, z'.

d) $\|P_{R_{s+1}}(z) - P_{R_{s+1}}(z')\|^2 + \|Q_{R_{s+1}}(z) - Q_{R_{s+1}}(z')\|^2 \leq \|z - z'\|^2$ *for all* z, z'.

The proof of Lemma 4.6 is similar to the proof Proposition 1 in [11].

Lemma 4.7. *The sequence $\{z^s = (x^s, y^s)\}$ generated by the modified PPNR method (8)–(10) satisfies the following condition*

$$\lim_{s \to \infty} T_{R_{s+1}}(z^s) = 0. \tag{22}$$

Proof. The proof follows closely the proof of Theorem 1 in [11]. Without restricting generality we assume that $\lim_{s \to \infty} z^s = \bar{z}$, otherwise we can consider a convergent subsequence.

For any solution z^* we have $P_{R_{s+1}}(z^*) = z^*$, therefore $Q_{R_{s+1}}(z^*) = 0$ for all $s = 0, 1, 2, \dots$. For $z = z^s$ and $z' = z^*$ by Lemma 4.6 d) we have

$$\|P_{R_{s+1}}(z^s) - z^*\|^2 + \|Q_{R_{s+1}}(z^s)\|^2 \leq \|z^s - z^*\|^2 \quad \forall s \geq 0. \tag{23}$$

Therefore keeping in mind Lemma 4.3 we have

$$\begin{aligned}
& \|Q_{R_{s+1}}(z^s)\|^2 - \|z^s - z^*\|^2 + \|z^{s+1} - z^*\|^2 \\
\leq\ & \|z^{s+1} - z^*\|^2 - \|P_{R_{s+1}}(z^s) - z^*\|^2 \\
=\ & \left(z^{s+1} - P_{R_{s+1}}(z^s)\right)^T \left(z^{s+1} - z^* + P_{R_{s+1}}(z^s) - z^*\right) \\
\leq\ & \left\|z^{s+1} - P_{R_{s+1}}(z^s)\right\| \left(\|z^{s+1} - z^*\| + \|P_{R_{s+1}}(z^s) - z^*\|\right) \\
\leq\ & \|z^{s+1} - P_{R_{s+1}}(z^s)\| (\|z^{s+1} - z^*\| + \|z^s - z^*\|),
\end{aligned}$$

and hence using (20) and (21) we obtain

$$\|Q_{R_{s+1}}(z^s)\|^2 \leq \|z^s - z^*\|^2 - \|z^{s+1} - z^*\|^2 + 2\varepsilon_s(M + \|z^*\|)$$

Using the assumption that $\lim_{s \to \infty} = \bar{z}$ and passing s to the limit we have

$$\lim_{s \to \infty} \|Q_{R_{s+1}}(z^s)\| = 0,$$

which implies

$$\lim_{s \to \infty} (k_s)^{-1} Q_{R_{s+1}}(z^s) = 0 \tag{24}$$

and from Lemma 4.6b) we have

$$\lim_{s \to \infty} \|z^{s+1} - P_{R_{s+1}}(z^s)\| = 0.$$

The latter limit, in turn, implies that

$$\lim_{s \to \infty} P_{R_{s+1}}(z^s) = P_{\bar{R}}(\bar{z}) = \bar{z}, \tag{25}$$

where $\bar{R} = \operatorname{diag}(\bar{r}_i)_{i=1}^m$, and

$$\bar{r}_i = \begin{cases} -\left[\psi''(\theta_i^s k_i^s c_i(\bar{x}))\right], & 0 < \theta_i^s < 1, \quad \text{if} \quad \bar{y}_i > 0 \\ 0, & \qquad\qquad\quad\ \text{if} \quad \bar{y}_i = 0 \end{cases}$$

Since $\bar{z}$ is a fixed point of $P_{\bar{R}}$ we have $(I + T_{\bar{R}})^{-1}(\bar{z}) = \bar{z}$, or $(I + T_{\bar{R}})(\bar{z}) = \bar{z}$, or $\bar{z} + T_{\bar{R}}(\bar{z}) = \bar{z}$, or $T_{\bar{R}}(\bar{z}) = 0$. ∎

Lemma 4.8. *The sequence $\{z^s = (x^s, y^s)\}$ generated by the modified PPNR method (8)–(10) satisfies the following conditions*

$$\lim_{s\to\infty} \|z^{s+1} - z^s\| = 0. \tag{26}$$

Proof. The result follows from Lemma 4.7 and formulas (14) and (15). □

Lemma 4.9. *Let $\{k_s\}$ be a nondecreasing bounded sequence of positive parameters. Then the sequence $\{z^s = (x^s, y^s)\}$ generated by the modified PPNR method (8)–(10) satisfies the following conditions*

a) $\lim_{s\to\infty} \nabla_x L(x^s, y^s) = 0,$

b) $\lim_{s\to\infty} y_i^s \geq 0,\ i = 1, \ldots, m.$

c) $\lim_{s\to\infty} c_i(x^s) \geq 0,\ i = 1, \ldots, m.$

d) $\lim_{s\to\infty} c_i(x^s) y_i^s = 0,\ i = 1, \ldots, m,$

Proof. Let $\bar{z} = (\bar{x}, \bar{y})$ is any limit point of the sequence $\{z^s\}$. Without restricting generality we assume that $\lim_{s\to\infty} z^s = \bar{z}$, otherwise we can consider a converging subsequence of $\{z^s\}$.

a) Focusing on the first n components of $T_{R_{s+1}}(z^s)$ from (22) we have immediately

$$\lim_{s\to\infty} \nabla_x L(x^{s+1}, y^{s+1}) = \nabla_x L(\bar{x}, \bar{y}) = 0.$$

b) Since $y_i^0 > 0$ and $\psi'(t) > 0$ for any t, we have by (10) $y_i^s > 0 \Rightarrow y_i^{s+1} > 0$, therefore

$$\lim_{s\to\infty} y_i^s = \bar{y}_i \geq 0 \quad i = 1, \ldots, m.$$

c) and d) Let us consider last m components of $T_{R_{s+1}}(z^s)$. From (22) we have

$$\lim_{s\to\infty} -\psi''(\theta_i^s k_i^s c_i(x^{s+1})) c_i(x^{s+1}) = 0, \quad 0 < \theta_i^s < 1, \quad i = 1, \ldots, m.$$

Therefore for each $1 \leq i \leq m$ we have either

$$\lim_{s\to\infty} c_i(x^s) = c_i(\bar{x}) = 0,$$

or

$$\lim_{s\to\infty} -\psi''(\theta_i^s k_i^s c_i(x^{s+1})) = 0, \quad 0 < \theta_i < 1.$$

Let us show that the latter implies

$$\lim_{s\to\infty} c_i(x^s) = c_i(\bar{x}) \geq 0.$$

and

$$\lim_{s\to\infty} y_i^s = \bar{y}_i = 0.$$

From Property 4^0 of transformation $\psi(t)$ follows that

$$\lim_{s\to\infty} -\psi''(\theta_i^s k_i^s c_i(x^{s+1})) = 0$$

implies

$$\lim_{s\to\infty} \theta_i^s k_i^s c_i(x^{s+1})) = \infty \tag{27}$$

Since $k_i^s > 0$ and $0 < \theta_i^s < 1$ we have

$$\lim_{s\to\infty} c_i(x^s) = c_i(\bar{x}) \geq 0.$$

Also from $0 < \theta_i^s < 1$ follows $\lim_{s\to\infty} k_i^s = \infty$. Boundedness of $\{k_s\}$ implies

$$\lim_{s\to\infty} y_i^{s+1} = 0.$$

Therefore for any $i = 1, \ldots, m$ we have either

$$\lim_{s\to\infty} c_i(x^s) = 0,$$

and

$$\lim_{s\to\infty} y_i^s \geq 0 \quad \text{(based on already proven b))}$$

or

$$\lim_{s\to\infty} c_i(x^s) \geq 0.$$

and

$$\lim_{s\to\infty} y_i^s = 0$$

Thus c) and d) are proven. □

The following theorem takes place.

Theorem 4.10. *If assumptions* **A** *and* **B** *are satisfied and* $\{k_s > 0\}$ *is a nondecreasing bounded sequence, then any limit point* $\bar{z} = (\bar{x}, \bar{y})$ *of the sequence* $\{z^s = (x^s, y^s)\}$ *generated by the modified PPNR method (8)–(10) is the primal-dual solution, i.e.* $(\bar{x}, \bar{y}) \in X^* \times Y^*$.

Proof. By Lemma 4.9 $\bar{z} = (\bar{x}, \bar{y})$ satisfies the KKT's optimality conditions (1)-(2). Therefore for the convex optimization problem $\mathcal{P}$, we have $\bar{x} \in X^*$ and $\bar{y} \in Y^*$, i.e. $\bar{z} = (\bar{x}, \bar{y})$ is the primal-dual solution. □

Remark 1. The convergence of the method (5)–(7) follows from Theorem 4.10.

5. **Rate of convergence of the modified PPNR Method.** In this section we prove q-linear rate of convergence of the modified PPNR method (8)–(10) under standard second-order optimality conditions.

Let us assume that the active constraint set is $I^* = \{i : c_i(x^*) = 0\} = \{1, \ldots, r\}$. We consider the vectors functions $c^T(x) = (c_1(x), \ldots, c_m(x))$, $c_{(r)}^T(x) = (c_1(x), \ldots, c_r(x))$, and their Jacobians $\nabla c(x) = J(c(x))$ and $\nabla c_{(r)}(x) = J(c_{(r)}(x))$. The sufficient regularity conditions

$$rank \nabla c_{(r)}(x^*) = r, y_i^* > 0,\, i \in I^* \tag{28}$$

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

$$(\nabla_{xx}^2 L(x^*, y^*)y, y) \geq \rho(y, y),\, \rho > 0, \forall y \neq 0 : \nabla c_{(r)}(x^*)y = 0 \tag{29}$$

comprise the standard second-order optimality conditions.

From the standard second-order optimality conditions follows the existence of a constant $\sigma > 0$:

$$\min\{\min\{y_i^* \,|\, i = 1, \ldots, r\};\, \min\{c_i(x^*) \,|\, i = r+1, \ldots, m\}\} = \sigma > 0. \tag{30}$$

From this point on we assume:

C. The standard second-order optimality conditions (28)–(29) are satisfied.

To establish the rate of convergence we modify the stopping criteria for finding an inexact unconstrained minimizer. Let $\{\delta_s\}$ be a positive sequence that $\sum_{s=0}^{\infty} \delta_s < \infty$. We consider the modified PPNR method that generates the following sequences

$$\boldsymbol{k}^s = (k_i^s = k_s(y_i^s)^{-1}, \quad i = 1, \ldots, m), \tag{31}$$

$$\begin{aligned} x^{s+1} &: \|\nabla_x \mathcal{L}(x^{s+1}, y^s, \boldsymbol{k}^s) + \tfrac{1}{k_s}(x - x^s)\| \le \min\left\{\delta_s,\ \tfrac{\delta_s}{k_s}\|z^{s+1} - z^s\|_\infty\right\} \\ &= \min\left\{\delta_s,\ \tfrac{\delta_s}{k_s}\max\left\{\|x^{s+1} - x^s\|_\infty, \max_{1\le i\le m}|y_i^s\psi'(k_i^s c_i(x^{s+1})) - y_i^s|\right\}\right\} \end{aligned} \tag{32}$$

$$y_i^{s+1} = y_i^s \psi'(k_i^s c_i(x^{s+1})), \quad i = 1, \dots, m, \tag{33}$$

where for a given vector $a \in \mathbb{R}^p$ the norm $\|a\|_\infty = \max_{1\le i\le p}|a_i|$ and for $A : \mathbb{R}^q \to \mathbb{R}^p$ the associated matrix norm $\|A\|_\infty = \max\limits_{1\le i\le p}\left(\sum\limits_{j=1}^q |a_{ij}|\right)$.

Remark 2. The second-order optimality conditions implies the uniqueness of the primal-dual solution $z^* = (x^*, y^*)$. The method (31)-(33) satisfies the conditions of Theorem 4.10 with $\varepsilon_s = \delta_s$. Therefore the whole primal-dual sequence $\{z^s\}$ generated by (31)-(33)converges to z^*. Also, there exists s' such that

$$\frac{\delta_s}{k_s}\|z^{s+1} - z^s\|_\infty \le \delta_s$$

for $s \ge s'$ and (32) can be replaced with

$$\begin{aligned} x^{s+1} &: \|\nabla_x \mathcal{L}(x^{s+1}, y^s, \boldsymbol{k}^s) + \tfrac{1}{k_s}(x - x^s)\| \le \tfrac{\delta_s}{k_s}\|z^{s+1} - z^s\|_\infty \\ &= \tfrac{\delta_s}{k_s}\max\left\{\|x^{s+1} - x^s\|_\infty, \max_{1\le i\le m}|y_i^s\psi'(k_i^s c_i(x^{s+1})) - y_i^s|\right\} \end{aligned}$$

In the rest of the section we will show that the second-order optimality conditions (28)–(29) imply an asymptotic q-linear rate of convergence of $\{z^s\}$, i.e. there exists $s_0 > 0$ and $0 < q < 1$ that

$$\|z^{s+1} - z^*\| \le q\|z^s - z^*\|, \forall s \ge s_0.$$

We start with the following lemmas.

Lemma 5.1. *The sequence $\{y^s_{(m-r)}\}$, $y^s_{(m-r)} = (y_i^s,\ i = r+1, \dots, m)$ of the Lagrange multipliers that correspond to the passive constraints generated by the proximal nonlinear rescaling method (31)-(33) converges to $y^*_{(m-r)} = 0$ with a quadratic rate, i.e there exist $C_0 > 0$ independent of $\{z^s\}$ and s_0 that for $s \ge s_0$ we have*

$$\|y^{s+1}_{(m-r)} - y^*_{(m-r)}\|_\infty \le C_0\|y^s_{(m-r)} - y^*_{(m-r)}\|^2_\infty \tag{34}$$

Proof. Let $\{z^s\}$ be the sequence generated by the proximal nonlinear rescaling method (31)-(33). By Remark 2 we have $\lim_{s\to\infty} z^s = z^*$. From $z \to z^*$ and the second-order optimality conditions follows the existence of $s_0 > 0$ such that $c_i(x^s) \ge 0.5\sigma > 0$, $i = r+1, \dots, m$ for $s \ge s_0$, where σ is defined in (30). Therefore keeping in mind property $2^0c)$ of transformation $\psi(t)$ for $i = r+1, \dots, m$ we have

$$y_i^{s+1} = y_i^s\psi'(k_i^s c_i(x^{s+1})) \le y_i^s \frac{a_1}{k_i^s c_i(x^{s+1}) + 1} < y_i^s \frac{a_1}{k_i^s c_i(x^{s+1})}$$

$$= \frac{a_1}{k_s c_i(x^{s+1})}(y_i^s)^2 \le \frac{a_1}{k_0 c_i(x^{s+1})}(y_i^s)^2 \le \frac{2a_1}{k_0\sigma}(y_i^s)^2 = C_0(y_i^s)^2,$$

where $C_0 = 2a_1/k_0\sigma$. Keeping in mind that $y^*_{(m-r)} = 0$ and the definition of l_∞-norm we have (34). ∎

Let us consider operators $T^{n+r}_{R_{s+1}}$ and $P^{n+r}_{R_{s+1}} : \mathbb{R}^n \times \mathbb{R}^m_+ \to \mathbb{R}^n \times \mathbb{R}^r_+$, operators $T^{m-r}_{R_{s+1}}$ and $P^{m-r}_{R_{s+1}} : \mathbb{R}^n \times \mathbb{R}^m_+ \to \mathbb{R}^{m-r}_+$, operator $P^n_{R_{s+1}} : \mathbb{R}^n \times \mathbb{R}^m_+ \to \mathbb{R}^n$ and operator $P^m_{R_{s+1}} : \mathbb{R}^n \times \mathbb{R}^m_+ \to \mathbb{R}^m_+$, obtained by truncating the operators $T_{R_{s+1}}$ and $P_{R_{s+1}}$:

$$T_{R_{s+1}} = \begin{bmatrix} T^{n+r}_{R_{s+1}} \\ T^{m-r}_{R_{s+1}} \end{bmatrix}$$

and

$$P_{R_{s+1}} = \begin{bmatrix} P^{n+r}_{R_{s+1}} \\ P^{m-r}_{R_{s+1}} \end{bmatrix} = \begin{bmatrix} P^{n}_{R_{s+1}} \\ P^{m}_{R_{s+1}} \end{bmatrix}.$$

For example, the operator $T^{n+r}_{R_{s+1}}$ is obtained from $T_{R_{s+1}}$ by considering only first $n+r$ components of the latter, while the operator $P^{m-r}_{R_{s+1}}$ is obtained from $P_{R_{s+1}}$ by considering only last $m-r$ components of the latter.

Remark 3. For a particular realization of the method (31)-(33) with $(x^{s+1}, y^{s+1}) = z^{s+1} = P_{R_{s+1}}(z^s)$ from Lemma 5.1 we have

$$\|P^{m-r}_{R_{s+1}}(z^s) - y^*_{(m-r)}\|_\infty \le C_0 \|y^s_{(m-r)} - y^*_{(m-r)}\|^2_\infty. \tag{35}$$

Lemma 5.2. *If the standard second-order optimality conditions are satisfied, $\{k_s\}$ is a nondecreasing bounded sequence of positive parameters, the sequence $\{z^s\} = \{(x^s, y^s)\}$ is generated by (31)-(33) then there exists an index s_1 and $a > 0$ independent of $\{z_s\}$ such that*

$$\|P_{R_{s+1}}(z^s) - z^*\|_\infty \le \frac{a}{k_s} \|P_{R_{s+1}}(z^s) - z^s\|_\infty. \tag{36}$$

holds for all $s \ge s_1$.

Proof. Consider the operator $T^{n+r} : \mathbb{R}^n \times \mathbb{R}^r_+ \to \mathbb{R}^n \times \mathbb{R}^r$ defined by

$$T^{n+r}(z_{(n+r)}) = \begin{pmatrix} \nabla_x L(x, y_{(r)}) \\ -\nabla_{y_{(r)}} L(x, y_{(r)}) \end{pmatrix} = \begin{pmatrix} \nabla f(x) - \sum_{i=1}^r y_i \nabla c_i(x) \\ c_{(r)}(x) \end{pmatrix}.$$

First we prove that there exists small enough $\varepsilon > 0$ and $C_1 > 0$ such that for all $z_{(n+r)} = (x, y_{(r)}) : \|z_{(n+r)} - z^*_{(n+r)}\|_\infty \le \varepsilon$ the following inequality holds

$$\|z_{(n+r)} - z^*_{(n+r)}\|_\infty \le C_1 \|T^{n+r}(z_{(n+r)})\|_\infty. \tag{37}$$

Let $(x^*, y^*) = (x^*, y^*_{(r)}, y^*_{(m-r)})$ is the primal-dual solution. After linearizing $\nabla_x L(x, y_{(r)})$ and $c_{(r)}(x)$ at $(x^*, y^*_{(r)})$, we obtain

$$\begin{aligned} \nabla_x L(x, y_{(r)}) =& \nabla_x L(x^*, y^*_{(r)}) + \nabla^2_{xx} L(x^*, y^*_{(r)})(x - x^*) \\ & - \nabla c^T_{(r)}(x^*)(y_{(r)} - y^*_{(r)}) + \mathcal{O}_{(n)} \|x - x^*\|^2_\infty, \end{aligned} \tag{38}$$

$$c_{(r)}(x) = c_{(r)}(x^*) + \nabla c_{(r)}(x^*)(x - x^*) + \mathcal{O}_{(r)} \|x - x^*\|^2_\infty, \tag{39}$$

Keeping in mind K-K-T conditions we can rewrite (38)–(39) in a matrix form

$$\begin{aligned} & \begin{bmatrix} \nabla^2_{xx} L(x^*, y^*_{(r)}) & -\nabla c^T_{(r)}(x^*) \\ \nabla c_{(r)}(x^*) & 0 \end{bmatrix} \begin{bmatrix} x - x^* \\ y_{(r)} - y^*_{(r)} \end{bmatrix} \\ = & \begin{bmatrix} \nabla_x L(x, y_{(r)}) & +\mathcal{O}_{(n)} \|x - x^*\|^2_\infty \\ c_{(r)}(x) & +\mathcal{O}_{(r)} \|x - x^*\|^2_\infty \end{bmatrix} \\ = & T^{n+r}(z_{(n+r)}) + \mathcal{O}_{(n+r)} \|z_{(n+r)} - z^*_{(n+r)}\|^2_\infty \end{aligned} \tag{40}$$

Due to the standard second order optimality conditions the matrix

$$A(x^*, y^*_{(r)}) = \left[\begin{array}{cc} \nabla^2_{xx} L(x^*, y^*_{(r)}) & -\nabla c^T_{(r)}(x^*) \\ \nabla c_{(r)}(x^*) & 0 \end{array} \right]$$

is nonsingular (see [5], p. 231) and there exists $N > 0$ such that

$$\|A^{-1}(x^*, y^*_{(r)})\|_\infty \leq N. \tag{41}$$

Hence from (41) for small enough $\varepsilon > 0$ and any $z_{(n+r)} = (x, y_r) : \|z_{(n+r)} - z^*_{(n+r)}\| \leq \varepsilon$ we have

$$\begin{aligned} \left\| \begin{array}{c} x - x^* \\ y_{(r)} - y^*_{(r)} \end{array} \right\|_\infty &\leq M \left(\|T^{n+r}(z_{(n+r)})\|_\infty + \mathcal{O}_{(n+r)} \|z - z^*\|^2 \right) \\ &\leq C_1 \|T^{n+r}(z_{(n+r)})\|_\infty, \end{aligned}$$

where $C_1 = 2N$.

Therefore for s_1 large enough and any $s \geq s_1$ we have

$$\|z^{s+1}_{(n+r)} - z^*_{(n+r)}\|_\infty \leq C_1 \|T^{n+r}(z^{s+1}_{(n+r)})\|_\infty. \tag{42}$$

Due to Theorem 4.10, the second-order optimality conditions and Remark 2, there exists $C_2 > 0$ that for s_1 large enough and for all $s \geq s_1$ we have $|c_i(x^{s+1})| \leq C_2$, $i = 1, \ldots, r$ and $y^{s+1}_i \geq \sigma > 0$, $i = 1, \ldots, r$ (σ is defined in (30)). Keeping in mind Property 4^0 of transformation $\psi(t)$, and boundedness of $\{k_s\}$ we have for all $s \geq s_1$

$$0 < b_1 \leq r^{s+1}_i = -\left[\psi''(\theta^s_i k^s_i c_i(x^{s+1}))\right] \leq b_2, \quad i = 1, \ldots, r,$$

$0 < \theta^s_i < 1$.

We can assume that for s_1 large enough and all $s \geq s_1$ we have

$$y^{s+1}_i \leq \|T^{n+r}(z^{s+1}_{(n+r)})\|^{1.5}_\infty, \quad i = r+1, \ldots, m.$$

otherwise it is easy to show using Lemma 5.1 that $\{z^s\}$ converges to z^* with a superlinear rate. We can also assume that for s_1 large enough there exists $C_3 > 0$ that for all $s \geq s_1$ we have $\|\nabla c_i(x^{s+1})\|_\infty \leq C_3$, $i = r+1, \ldots, m$. Therefore we have

$$\begin{aligned} \|T^{n+r}(z^{s+1}_{(n+r)})\|_\infty &= \left\| \begin{array}{c} \nabla_x L(x^{s+1}, y^{s+1}_{(r)}) \\ c_{(r)}(x^{s+1}) \end{array} \right\|_\infty \\ = & \left\| \begin{array}{c} \nabla_x L(x^{s+1}, y^{s+1}) + \sum_{i=r+1}^m y^{s+1}_i \nabla c_i(x^{s+1}) \\ c_{(r)}(x^{s+1}) \end{array} \right\|_\infty \\ \leq & \left\| \begin{array}{c} \nabla_x L(x^{s+1}, y^{s+1}) \\ (R^{(n+r)}_{s+1})^{-1} \left(R^{(n+r)}_{s+1} c_{(r)}(x^{s+1}) \right) \end{array} \right\|_\infty + \| \sum_{i=r+1}^m y^{s+1}_i \nabla c_i(x^{s+1}) \|_\infty \\ \leq & \; b_1^{-1} \|T^{n+r}_{R_{s+1}}(z^{s+1})\|_\infty + C_3 \|T^{n+r}(z^{s+1}_{(n+r)})\|^{1.5}_\infty \\ \leq & \; C_4 \|T^{n+r}_{R_{s+1}}(z^{s+1})\|_\infty \end{aligned}$$

for all $s \geq s_1$, where $C_4 = 2b_1^{-1}$.

Therefore keeping in mind (42) we have

$$\|z^{s+1}_{(n+r)} - z^*_{(n+r)}\|_\infty \leq C_5 \|T^{n+r}_{R_{s+1}}(z^{s+1})\|_\infty, \tag{43}$$

where $C_5 = C_1 C_4$.

By replacing $z^{s+1}_{(n+r)}$ with $P^{n+r}_{R_{s+1}}(z^s)$ and z^{s+1} with $P_{R_{s+1}}(z^s)$ from (43) we obtain

$$\|P^{n+r}_{R_{s+1}}(z^s) - z^*_{(n+r)}\|_\infty \leq C_5 \|T^{n+r}_{R_{s+1}}(P_{R_{s+1}}(z^s))\|_\infty. \tag{44}$$

From the definition of l_∞-norm we have

$$\|T^{n+r}_{R_{s+1}}(P_{R_{s+1}}(z^s))\|_\infty \le \|T_{R_{s+1}}(P_{R_{s+1}}(z^s))\|_\infty \tag{45}$$

From Lemma 4.6 b) we have

$$\|T_{R_{s+1}}(P_{R_{s+1}}(z^s))\|_\infty = \frac{1}{k_s}\|P_{R_{s+1}}(z^s)) - z^s\|_\infty. \tag{46}$$

Therefore combining (44)–(46) we obtain

$$\|P^{n+r}_{R_{s+1}}(z^s) - z^*_{(n+r)}\|_\infty \le \frac{C_5}{k_s}\|P_{R_{s+1}}(z^s)) - z^s\|_\infty. \tag{47}$$

Let $x(z) = P^n_{R_{s+1}}(z)$. then $\lim_{s\to\infty}\psi'(k^s_i c_i(x(z^s))) = 0$, $i = r+1,\dots,m$. Therefore, keeping in mind the boundedness of $\{k_s\}$, there exists C_6 that for s_1 large enough and for all $s \ge s_1$ we have

$$\psi'(k^s_i c_i(x(z^s))) \le \frac{C_6}{k_s}|\psi'(k^s_i c_i(x(z^s))) - 1|$$

or

$$y^s_i\psi'(k^s_i c_i(x(z^s))) \le \frac{C_6}{k_s}|y^s_i\psi'(k^s_i c_i(x(z^s))) - y^s_i|$$

for $i = r+1,\dots,m$. Therefore keeping in mind that $y^*_{(m-r)} = 0$ and the definition of l_∞-norm we have

$$\|P^{m-r}_{R_{s+1}}(z^s) - y^*_{(m-r)}\|_\infty \le \frac{C_6}{k_s}\|P_{R_{s+1}}(z^s)) - z^s\|_\infty. \tag{48}$$

Finally combining (47) and (48) we obtain for $s \ge s_1$

$$\|P_{R_{s+1}}(z^s) - z^*\|_\infty \le \frac{a}{k_s}\|P_{R_{s+1}}(z^s)) - z^s\|_\infty,$$

where $a = \max\{C_5, C_6\}$. □

Remark 4. Keeping in mind the relations between l_∞ and l_2 norm we have

$$\begin{aligned}\|P_{R_{s+1}}(z^s) - z^*\| &\le \sqrt{n+m}\|P_{R_{s+1}}(z^s) - z^*\|_\infty \\ &\le \frac{a\sqrt{n+m}}{k_s}\|P_{R_{s+1}}(z^s)) - z^s\|_\infty \\ &\le \frac{\bar a}{k_s}\|P_{R_{s+1}}(z^s)) - z^s\|,\end{aligned} \tag{49}$$

where $\bar a = a\sqrt{n+m}$.

Now we are ready to establish the q-linear rate of convergence.

Theorem 5.3. *If the problem ($\mathcal{P}$) satisfy the standard second order optimality conditions* **C***, the nondecreasing sequence $\{k_s\}$ is bounded and k_0 is large enough then for the primal-dual sequence $\{z^s\}$ is generated by the PPNR method (31)-(33), there exist numbers $0 < q < 1$ and $\bar s$ such that for $s \ge \bar s$ the following bound holds*

$$\|z^{s+1} - z^*\| \le q\|z^s - z^*\|. \tag{50}$$

Proof. The sequence $\{z^s\}$ generated by PPNR method (31)-(33) satisfies the conditions of Lemma 4.4 with $\varepsilon_s = \delta_s\|z^{s+1} - z^s\|_\infty$. We remind that $\lim_{s\to\infty}\varepsilon_s = 0$, $\sum\varepsilon_s < \infty$. Therefore

$$\|P_{R_{s+1}}(z^s) - z^{s+1}\| \le \delta_s\|z^{s+1} - z^s\|_\infty \le \delta_s\|z^{s+1} - z^s\| \le \delta_s\|z^{s+1} - z^*\| + \delta_s\|z^s - z^*\| \tag{51}$$

Invoking Lemma 4.6 c) with $z = z^*$, $z' = z^s$ we obtain

$$\|P_{R_{s+1}}(z^s) - z^*\|^2 + \|Q_{R_{s+1}}(z^s)\|^2 \leq \|z^s - z^*\|^2.$$

It follows from Lemma 4.6 a) that $\|Q_{R_{s+1}}(z^s)\|^2 = \|P_{R_{s+1}}(z^s) - z^s\|^2$, therefore

$$\|P_{R_{s+1}}(z^s) - z^*\|^2 + \|P_{R_{s+1}}(z^s) - z^s\|^2 \leq \|z^s - z^*\|^2. \tag{52}$$

Combining (49) and (52) yields

$$\|P_{R_{s+1}}(z^s) - z^*\|^2 \leq (\bar{a}/k_s)^2/(1 + (\bar{a}/k_s)^2)\|z^s - z^*\|^2,$$

or

$$\|P_{R_{s+1}}(z^s) - z^*\| \leq \gamma_1 \|z^s - z^*\|, \tag{53}$$

where $\gamma_1 = \sqrt{(\bar{a}/k_s)^2/(1 + (\bar{a}/k_s)^2)}$. Note that taking k_s large enough makes γ_1 small enough.

Therefore combining (51) and (53) and keeping in mind (32) and Lemma 4.4 we have

$$\begin{aligned}\|z^{s+1} - z^*\| &\leq \|z^{s+1} - P_{R_{s+1}}(z^s)\| + \|P_{R_{s+1}}(z^s) - z^*\| \\ &\leq \delta_s\|z^{s+1} - z^*\| + \delta_s\|z^s - z^*\| + \gamma_1\|z^s - z^*\|,\end{aligned}$$

or

$$\|z^{s+1} - z^*\| \leq \frac{\delta_s + \gamma_1}{1 - \delta_s}\|z^s - z^*\|.$$

By taking $k_0 > 0$ large enough for any $k_s \geq k_0$ we obtain $0 < \gamma_1 < 1$ small enough. Keeping in mind that $\delta_s \to 0$ from the last formula follows the existence of $0 < q < 1$ such that $(\delta_s + \gamma_1)/(1 - \delta_s) \leq q$ and formula (50) holds. Moreover, for any given small enough $0 < q < 1$ one can find such k_q and s_q that (50) holds true for any $k_s \geq k_q$ and $s \geq s_q$. □

6. **Concluding remarks.** The PPNR method (8)-(10) generates a bounded primal-dual sequence $\{x^s, y^s\}$ that any limit point $(\bar{x}, \bar{y}) \in X^* \times Y^*$. The results are true under mild assumptions on the input data. The PPNR method (8)-(10) does not require solving an unconstrained minimization problems at each step.

However, it may require several steps of an unconstrained minimization method for finding the primal-dual approximation. Reducing numerical effort for finding the approximation in (9) is critical for the numerical efficiency of PPNR method (8)-(10).

One step of the PPNR method (5)-(7) is equivalent to solving for (x, y) the primal-dual nonlinear system of equations (12)-(13). Application of Newton's method for solving the system (12)-(13) leads to the primal-dual PPNR method. Such approach has proven to be efficient for the NR methods (see [3, 10]). We expect that the primal-dual PPNR method will reduce the computational effort per step and at the same time improve the asymptotic convergence rate.

REFERENCES

[1] A. Auslender, M. Teboulle and S. Ben-Tiba, *Interior proximal and multipliers methods based on second-order homegeneous kernels*, Mathematics of Operations Research, **24 (3)** (1999), 645–668.

[2] D. Bertsekas, "Constrained Optimization and Lagrange Multipliers Methods," Academic Press, New York, 1982.

[3] I. Griva and R. Polyak, *Primal-Dual Nonlinear Rescaling Method with Dynamic Scaling Parameter Update*, Mathematical Programming, **106 (2)** (2006), 237–259.

[4] G. J. Minty, *Monotone (nonlinear) Operators in Hilbert Space*, Duke Math Journal, **29** (1962), 341–346.
[5] B. T. Polyak, "Introduction to Optimization," Software Inc., New York, 1987.
[6] R. Polyak, *Modified Barrier Functions (theory and methods)*, Mathematical Programming, **54** (1992), 177–222.
[7] R. Polyak, *Nonlinear Rescaling vs. Smoothing Technique in convex optimization*, Mathematical Programming, **92**, (2002), 197–235.
[8] R. Polyak, *Nonlinear rescaling as Interior Quadratic Prox method in convex optimization*, Computational Optimization and Applications, **35** (2006), 347–373.
[9] R. Polyak and M. Teboulle, *Nonlinear Rescaling and Proximal-like Methods in convex optimization*, Mathematical Programming, **76** (1997), 265–284.
[10] R. Polyak and I. Griva, *Primal-Dual Nonlinear Rescaling method for convex optimization*, Journal of Optimization Theory and Applications, **122** (2004), 111–156.
[11] R. T. Rockafellar, *Monotone Operators and The Proximal Point Algorithm*, SIAM Journal of Control and Optimization, **14** (1976), 887–898.
[12] R. T. Rockafellar, *Augmented Lagrangians and Applications of the Proximal Point algorithm in convex programming*, Mathematics of Operations Research, **1** (1976), 97–116.
[13] R. T. Rockafellar, "Convex analysis," Princeton University Press, Princeton, 1996.
[14] P. Tseng and D. Bertsekas *On the convergence of exponential multipliers method for convex programming*, Mathematical Programming, **76** (1993), 1–19.

Received November 2010; revised April 2011.

E-mail address: igriva@gmu.edu
E-mail address: rpolyak@gmu.edu