Nonlinear Rescaling as Interior Quadratic Prox Method in Convex Optimization

Roman A. Polyak
Department of SEOR and
Mathematical Sciences Department
George Mason University
Fairfax, VA 22030

This paper is dedicated to Professor Elijah Polak on the occasion of his 75th birthday.

Abstract

A class Ψ of strictly concave and twice continuously differentiable functions $\psi : R \to R$ with particular properties is used for constraint transformation in the framework of a Nonlinear Rescaling (NR) method with "dynamic" scaling parameter updates. We show that the NR method is equivalent to the Interior Quadratic Prox method for the dual problem in a rescaled dual space.

The equivalence is used to prove convergence and to estimate the rate of convergence of the NR method and its dual equivalent under very mild assumptions on the input data for a wide class Ψ of constraint transformations. It is also used to estimate the rate of convergence under strict complementarity and under the standard second order optimality condition.

We proved that for any $\psi \in \Psi$, which corresponds to a well-defined dual kernel $\varphi = -\psi^*$, the NR method applied to LP generates a quadratically convergent dual sequence if the dual LP has a unique solution.

1 Introduction

The intimate relationship between multiplier methods based on Quadratic Augmented Lagrangians [10], [25] and Quadratic Prox methods for the dual problem [14], [16] was established by R.T. Rockafellar in the 70's (see [26]–[28]).

In this paper we show that a similar relationship exists between Nonlinear Rescaling multiplier methods, with "dynamic" scaling parameter updates (see [2], [4], [21], [30]), and Interior Quadratic Prox methods for the dual problem in the rescaled dual space.

We consider a class Ψ of monotone increasing, concave and sufficiently smooth functions $\psi : R \to R$ with particular properties. We use the functions to transform constraints of a given constrained optimization problem into an equivalent set of constraints. The transformation is scaled by a vector of positive scaling parameters, one for each constraint. The unconstrained minimization of the Lagrangian for the equivalent problem in the primal space followed by both the Lagrange multipliers and scaling parameters update forms the general NR multiplier method. We update the scaling parameter vector by the formula suggested by P. Tseng and D. Bertsekas for the exponential multiplier method [30].

It is well known that the NR multiplier method with "dynamic" scaling parameters update leads to the Prox method with second order φ-divergence distance for the dual problem (see [2], [4], [21], [30]). It is also well known that the convergence analysis of the NR method with "dynamic" scaling parameter updates and its dual equivalent turned out to be rather difficult, even for a particular exponential transformation (see [30]). The first convergence result for the NR method with "dynamic" scaling parameter updates was obtained by A. Ben-Tal and M. Zibulevsky [4]. They proved that, for a particular class of constraint transformations, the primal and the dual sequences generated by NR type methods are bounded and any convergent primal-dual subsequence converges to the primal-dual solution. The result in [4] was extended by A. Auslender et al. in [2], establishing that the inexact proximal version of the multiplier method with "dynamic" scaling parameter updates generates a bounded sequence with every limit point being an optimal solution. Moreover, for a particular kernel $\varphi(t) = 0.5\nu(t-1)^2 + \mu(t - \ln t - 1)$, which is a regularized logarithmic MBF kernel (see [23]), the authors in [2] proved the global convergence of the dual sequence and established under very mild assumptions on the input data that the rate of convergence is $\mathcal{O}\left((ks)^{-1}\right)$.

The regularization, which provides the strong convexity of the dual kernel, was an important element in the analysis given in [2]. Unfortunately, such a modification of the dual kernel in some instances leads to substantial difficulties when it comes to finding the primal transformation, which is a Fenchel conjugate for the dual kernel. For example, in case of the exponential transformation, it leads to solving a transcendental equation. Therefore, the results of [2] cannot be applied for the exponential multiplier method [30]. In case of the logarithmic MBF kernel, there is a closed form solution for the corresponding equation, but the primal transformation (see Section 7 in [2]) is substantially different from the original logarithmic MBF transformation [19], which is proven to be very efficient numerically (see [3], [6], [17], [22]). In general, the regularization of the dual kernel changes substantially the properties of the original transformation, which are critical for convergence and rate of convergence of the NR methods.

Therefore, in this paper we consider an alternative approach. We guarantee the strong convexity of the dual kernels on $\mathbb{R}_+$ by a slight modification of the wide class of well known primal transformations $\psi \in \Psi$, using the "gluing" idea (see for example [5]). Such a modification makes the primal transformation well defined on $\mathbb{R}$, provides the original transformations with important properties and allows us to show that eventually only original transformations are responsible for convergence, rate of convergence and numerical efficiency of the NR method.

Our first contribution is the new convergence proof and the rate of convergence estimate of the general NR method for a wide class of transformations $\psi \in \Psi$, under very mild assumptions on the input data. The key component of the convergence proof is the equivalence of the NR method to an Interior Quadratic Prox method for the dual problem.

We prove that under strict complementarity conditions, the NR method converges with rate $o\left((ks)^{-1}\right)$. Such an estimate is typical for the Classical Quadratic Prox method (see [9] and references therein), where k is the penalty parameter and s is the number of steps. This is our second contribution. We show also that under the standard second order optimality condition, the NR method converges with Q-linear rate without unbounded increase of the scaling parameters, which correspond to the active constraints. This means that a Q-linear rate can be achieved without compromising the condition number of the Hessian of the Lagrangian for the equivalent problem. We introduced a stopping criterion that allows us to replace the primal minimizer by an approximation and to retain the Q-linear rate of convergence.

Our third contribution is the quadratic rate of convergence of the NR method with "dynamic" scaling parameter updates for Linear Programming (LP) problems. We proved that for any $\psi \in \Psi$, that corresponds to the well defined dual kernel, the NR method converges with quadratic rate, under the assumption that one of the dual LPs has a unique solution.

We also provide numerical results, which are consistent with the theory.

The paper is organized as follows. In the second section, we state the problem and describe the basic assumptions on the input data. In the third section, we introduce a class Ψ of smooth, strictly concave transformations $\psi : R \to R$ with special properties. We consider their Fenchel conjugate $\varphi^*(s) = \inf_t \{st - \psi(t)\}$ and establish properties of the dual kernels $\varphi = -\psi^*$ that play the key role in our analysis. We also describe the NR method and prove its equivalence to the Interior Quadratic Prox method for the dual problem in the rescaled dual space. In Section 4, we establish convergence and estimate the rate of convergence for the NR method under very mild assumptions on the input data. In Section 5, we establish the rate of convergence of the NR method under the strict complementarity condition and under the standard second order optimality condition. In Section 6, we establish quadratic convergence of the NR method for LP problems for a wide class of transformations $\psi \in \Psi$, which correspond to the well defined kernels $\varphi \in \mathcal{\varphi}$. In Section 7, we provide numerical results, which support the theory for both the NLP and LP calculations. We conclude the paper by discussing issues related to future research.

2 Statement of the Problem and Basic Assumptions

Let $f : \mathbb{R}^n \to \mathbb{R}^1$ be convex and all $c_i : \mathbb{R}^n \to \mathbb{R}^1$, $i = 1, \ldots, q$ be concave and smooth functions. We consider a convex set $\Omega = \{x \in \mathbb{R}^n_+ : c_i(x) \geq 0,\ i = 1, \ldots, q\}$ and the following convex optimization problem

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

We assume that:

A: The optimal set X^* is not empty and bounded.

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

We consider the Lagrangian $L(x, \lambda) = f(x) - \sum_{i=1}^q \lambda_i c_i(x)$, the dual function $d(\lambda) = \inf_{x \in \mathbb{R}^n} L(x, \lambda)$ and the dual problem

$$(\mathcal{D}) \qquad \lambda \in L^* = \text{Argmax}\{d(\lambda) | \lambda \in \mathbb{R}^q_+\}$$

It follows from **B** that the Karush-Kuhn-Tucker's (K-K-T's) conditions hold true and the dual optimal set

$$L^* = \left\{ \lambda \in \mathbb{R}^m_+ : \nabla f(x^*) - \sum_{i=1}^q \lambda_i \nabla c_i(x^*) = 0,\ x^* \in X^* \right\} \tag{2.1}$$

is bounded.

3 Nonlinear Rescaling - Interior Quadratic Prox

We consider a class Ψ of twice continuously differentiable functions $\psi : (-\infty, \infty) \to R$ with the following properties

1^o $\psi(0) = 0$

2^o a) $\psi'(t) > 0$, b) $\psi'(0) = 1$, c) $\psi'(t) \leq at^{-1}$, $\forall\, t \in (0, \infty)$, $a > 0$

3^o $-m^{-1} \leq \psi''(t) < 0$, $\forall\, t \in (-\infty, \infty)$

4^o $\psi''(t) \leq -M^{-1}$, $\forall\, t \in (-\infty, 0]$ and $0 < m < M < \infty$.

5^o $-\psi''(t) \geq 0.5t^{-1}\psi'(t)$, $\forall\, t \in [1, \infty)$.

Several examples of $\psi \in \Psi$ are given at the end of this section.

For any given vector $\boldsymbol{k} = (k_1, \ldots, k_q) \in \mathbb{R}^q_{++}$ due to 1^o and 2^o (a) we have

$$c_i(x) \geq 0 \Leftrightarrow k_i^{-1}\psi(k_i c_i(x)) \geq 0, \quad i \equiv 1 \ldots, q\,. \tag{3.1}$$

Therefore, the problem

$$x^* \in X^* = \text{Argmin}\{f(x)/k_i^{-1}\psi(k_i c_i(x)) \geq 0,\ i = 1, \ldots, q\} \tag{3.2}$$

is equivalent to the primal problem (P). The Lagrangian $\mathcal{L} : \mathbb{R}^n \times \mathbb{R}^q_{++} \times \mathbb{R}^q_{++} \to R$ for the equivalent problem, which is given by formula

$$\mathcal{L}(x, \lambda, \boldsymbol{k}) = f(x) - \sum_{i=1}^{q} \lambda_i k_i^{-1}\psi(k_i c_i(x))$$

is our main tool.

We are ready to describe the NR method. Let $(\lambda^0, \boldsymbol{k}^0) \in \mathbb{R}^q_{++} \times \mathbb{R}^q_{++}$, for example, we can take $k > 0$, $\lambda^0 \in \mathbb{R}^q_{++}$ and $\boldsymbol{k}^0 = (k_1^0, \ldots, k_q^0) : k_i^0 = k/\lambda_i^0, \quad i = 1, \ldots, q$. The NR method generates three sequences $\{x^s\} \subset \mathbb{R}^n$, $\{\lambda^s\} \subset \mathbb{R}^q_{++}$, $\{\boldsymbol{k}^s\} \subset \mathbb{R}^q_{++}$ by formulas

$$x^{s+1} = \arg\min\{\mathcal{L}(x, \lambda^s, \boldsymbol{k}^s)/x \in \mathbb{R}^n\} \tag{3.3}$$

$$\lambda_i^{s+1} = \lambda_i^s \psi'(k_i^s c_i(x^{s+1})) \tag{3.4}$$

$$k_i^{s+1} = k(\lambda_i^{s+1})^{-1}, \quad i = 1, \ldots, q\,. \tag{3.5}$$

The minimizer x^{s+1} in (3.3) exists for any $\lambda^s \in \mathbb{R}^q_{++}$ and $\boldsymbol{k}^s \in \mathbb{R}^q_{++}$ due to the boundedness of X^*, convexity f, concavity c_i and property 4^0 of $\psi \in \Psi$. It can be proven using consideration similar to those in [1]. So the NR method (3.3)-(3.5) is well defined.

Our main motivation is to understand from a general view point what kind of properties of the original transformation $\psi_i \in \Psi$ are responsible for convergence and the rate of convergence of the NR method (3.3)-(3.5).

First, we show that properties of 3^0 and 4^0 are critical for both theoretical analysis and numerical performance. On the other hand, for all well known transformations including exponential, logarithmic, hyperbolic and parabolic MBF as well as for those related to the smoothing technique [21], at least one of these properties is not satisfied. This is, in our opinion, the main source of difficulties for both theoretical analysis and numerical performance of the NR methods.

Second, we show that a slight modification of the classical transformations guarantees properties 3^0 and 4^0. Moreover, it became clear from the convergence proof that such modification does not affect the original transformations, because practically only the original transformations govern the NR method (3.3)-(3.5).

Third, the critical part of our analysis is based on the properties of the dual kernel $\varphi_i = -\psi_i^*$, induced by 3^0 and 4^0.

Fourth, the main ingredient of the convergence proof is the equivalence of the NR method (3.3)-(3.5) to the Interior Quadratic Prox (IQP) for the dual problem.

Fifth, the equivalence is also important for establishing the rate of convergence of the NR method for both convex and linear optimization under some extra assumptions on the input date.

Along with $\psi \in \Psi$, we consider the Fenchel conjugate $\psi^*(s) = \inf_t \{st - \psi(t)\}$. It follows from 4^o that $\lim_{t \to -\infty} \psi'(t) = \infty$. Therefore, for any $0 < s < \infty$ due to 3^o there exist ψ'^{-1}. Thus, the equation

$$(st - \psi(t))'_t = s - \psi'(t) = 0$$

has a unique solution for t, i.e., $t(s) = \psi'^{-1}(s)$. Using the well known formula $\psi'^{-1} = \psi^{*\prime}$ and $s = \psi'(t)$ we obtain the following identity

$$s \equiv \psi'(\psi^{*\prime}(s))\,. \tag{3.6}$$

With each $\psi \in \Psi$ we associate a smooth, strongly convex and nonnegative function $\varphi(s) = -\psi^*(s)$ which is defined together with its derivatives on $(0,\infty)$. So, with the class Ψ of constraints transformations we associate the class φ of barrier type kernels $\varphi : R_+ \to R_+$.

It is well known (see [2],[4],[21],[30]) that the NR method (3.3)-(3.5) is equivalent to the following prox method for the dual problem

$$\lambda^{s+1} = \text{argmax}\left\{d(\lambda) - k^{-1}D(\lambda,\lambda^s)/\lambda \in \mathbb{R}^q\right\} \tag{3.7}$$

where the second order φ-divergence distance function $D : \mathbb{R}^q_+ \times \mathbb{R}^q_{++} \to R_+$ is given by formula $D(u,v) = \sum v_i^2 \varphi(u_i/v_i)$. The function $\varphi : \mathbb{R}_{++} \to R_+$ is the kernel of the φ-divergence distance function $D(u,v)$. Also, in view of $\lim_{s\to 0}\varphi'(s) \to -\infty$ we have $\lambda^{s+1} \in \mathbb{R}^q_{++}$. So, the Prox method (3.7) is an Interior Prox method with second order φ-divergence distance function $D(u,v)$. Now we will prove that the NR (3.3)-(3.5) method is equivalent to IQP for the dual problem.

Theorem 3.1 *If (P) is a convex programming problem and the assumptions A and B are satisfied, then for any given $k > 0$ and any given pair $(\lambda^0, \boldsymbol{k}^0) \in \mathbb{R}^q_{++} \times \mathbb{R}^q_{++}$ the NR method (3.3)–(3.5) is equivalent to the Interior Quadratic Prox method for the dual problem.*

Proof. From (3.4), 2^ob) and the mean value formula we obtain

$$\begin{aligned}\lambda_i^{s+1} - \lambda_i^s &= \lambda_i^s(\psi'(k_i^s c_i(x^{s+1})) - \psi'(0)) \\ &= \lambda_i^s k_i^s \psi''(\theta_i^s k_i^s c_i(x^{s+1}))c_i(x^{s+1}) \\ &= \lambda_i^s k_i^s \psi''_{[s,i]}(\cdot)c_i(x^{s+1}), \quad i = 1,\ldots,q\end{aligned}$$

where $0 < \theta_i^s < 1$. Using (3.5) we can rewrite the multiplicative formula (3.4) in an additive form

$$\lambda^{s+1} = \lambda^s + k\Psi''_{[s]}(\cdot)c(x^{s+1}) \tag{3.8}$$

where $\Psi''_{[s]}(\cdot) = \text{diag}\,(\psi''_{[s,i]}(\cdot))_{i=1}^q$. The equation (3.8) can be rewritten as follows

$$-c(x^{s+1}) - k^{-1}(-\Psi''_{[s]}(\cdot))^{-1}(\lambda^{s+1} - \lambda^s) = 0\,. \tag{3.9}$$

Keeping in mind $-c(x^{s+1}) \in \partial d(\lambda^{s+1})$ we can view the equation (3.9) as optimality criteria for the vector λ^{s+1} in the following problem:

$$\begin{aligned} \lambda^{s+1} &= \arg\max\left\{d(\lambda) - \frac{1}{2}k^{-1}\sum_{i=1}^q(-\psi''_{[s,i]}(\cdot))^{-1}(\lambda_i - \lambda_i^s)^2/\lambda \in \mathbb{R}^q\right\} \\ &= \arg\max\left\{d(\lambda) - \frac{1}{2}k^{-1}\|\lambda - \lambda^s\|^2_{R_s}/\lambda \in \mathbb{R}^q\right\}\,, \end{aligned} \tag{3.10}$$

where $R_s = (-\Psi''_{[s]}(\cdot))^{-1}$. In other words, the NR method is equivalent to the Quadratic Prox method in the rescaled from step to step dual space. At the same time, the Quadratic Prox (3.10) produces a positive dual sequence $\{\lambda^s\} \subset \mathbb{R}^q_{++}$. Therefore, (3.10) is, in fact, an Interior Quadratic Prox for the dual problem in the rescaled dual space. The properties of the kernel $\varphi \in \mathcal{\varphi}$ are playing the key role in our further analysis. Therefore we start by characterizing the class $\mathcal{\varphi}$.

Theorem 3.2 *The kernels $\varphi \in \mathcal{\varphi}$ are convex, twice continuously differentiable and possess the following properties*

1. $\varphi(s) \geq 0,\ \forall\, s \in (0,\infty)$ *and* $\min_{s\geq 0}\varphi(s) = \varphi(1) = 0$
2. *(a)* $\lim\limits_{s\to 0+}\varphi'(s) = -\infty$, *(b)* $\varphi'(s)$ *is monotone increasing, and (c)* $\varphi'(1) = 0$.
3. *(a)* $\varphi''(s) \geq m > 0,\ \forall\, s \in (0,\infty)$, *(b)* $\varphi''(s) \leq M < \infty,\ \forall\, s \in [1,\infty)$.

Proof.

1) Due to concavity $\psi(t)$, $\psi(0) = 0$ and $\psi'(0) = 1$ for any $0 < s < 1$ there is $t > 0 : \psi(t) > st$. Therefore, $\psi^*(s) = \inf\limits_t\{st - \psi(t)\} < 0,\ \forall\, s \in (0,1)$. For the same reasons, for any given $1 < s < \infty$, there is $t < 0 : \psi(t) > st$. Therefore, $\psi^*(s) < 0,\ \forall\, s \in (1,\infty)$. For $s = 1$ due to 2^o b), we have $\psi^*(1) = 0$. Thus, $\varphi(s) = -\psi^*(s) \geq 0,\ \forall\, s \in (0,\infty)$ and $\varphi(1) = 0$.

2) From the definition of $\varphi(s)$ and 2^o c) follows $\lim\limits_{s\to 0_+} \varphi'(s) = -\infty$. The monotonicity $\varphi'(s)$ follows directly from the strong convexity of $\varphi(s)$, which we will prove later. From 2^o b) we have $\varphi'(1) = 0$. So, $\min\{\varphi(s)/0 < s < \infty\} = \varphi(1) = 0$.

3) By differentiating the identity (3.6) we obtain

$$1 = \psi''(\psi^{*\prime}(s)) \cdot \psi^{*\prime\prime}(s)\,.$$

Using again $t = \psi'^{-1}(s) = \psi^{*\prime}(s)$ we have

$$\psi^{*\prime\prime}(s) = [\psi''(t)]^{-1} \tag{3.11}$$

From 3^o we have $[-\psi''(t)]^{-1} \geq m$. Therefore, using (3.11) we obtain

$$\varphi''(s) = -\psi^{*\prime\prime}(s) = [-\psi''(t)]^{-1} \geq m, \ \forall\, s \in (0, \infty)\,.$$

From 4^o we have $[-\psi''(t)]^{-1} \leq M$. Using again (3.11) we obtain $\varphi''(s) \leq M$, $\forall s \in [1, \infty)$. The proof is complete.

As we pointed out earlier, several well known transformations (see [13], [19]-[21], [23]) do not satisfy 3^o or 4^o.

Let us consider some of them.

Exponential [13]: $\hat{\psi}_1(t) = 1 - e^{-t}$

Logarithmic MBF [19]: $\hat{\psi}_2(t) = \ln(t+1)$

Hyperbolic MBF [19]: $\hat{\psi}_3(t) = t(t+1)^{-1}$

Log-sigmoid [21]: $\hat{\psi}_4(t) = 2(\ln 2 + t - \ln(1+e^t))$

Modified CHKS [21]: $\hat{\psi}_5(t) = t - \sqrt{t^2 + 4\eta} + 2\sqrt{\eta} \quad \eta > 0$.

The transformations $\hat{\psi}_1$–$\hat{\psi}_3$ do not satisfy 3^o $(m = 0)$, while the transformations $\hat{\psi}_4$ and $\hat{\psi}_5$ do not satisfy 4^o $(M = \infty)$. However, a slight modification of $\hat{\psi}_i(t)$, $i = 1, \ldots, 5$

leads to transformations which satisfy 1^o–5^o. We consider $-1 < \tau < 0$ and define the modified transformations $\psi_i : R \to R$ as follows

$$\psi_i(t) := \begin{cases} \hat{\psi}_i(t), & t \geq \tau \\ q_i(t), & t \leq \tau \end{cases} \tag{3.12}$$

where $q_i(t) = a_i t^2 + b_i t + c_i$ and $a_i = 0.5\hat{\psi}_i''(\tau)$, $b_i = \hat{\psi}_i'(\tau) - \tau\hat{\psi}_i''(\tau)$, $c_i = \hat{\psi}_i(\tau) - \tau\hat{\psi}_i'(\tau) + 0.5\tau^2\hat{\psi}_i''(\tau)$. The coefficients a_i, b_i, c_i we found by solving the following system for a_i, b_i, c_i

$$\hat{\psi}_i(\tau) = q_i(\tau),\ \hat{\psi}_i'(\tau) = q_i'(\tau),\ \hat{\psi}_i''(\tau) = q_i''(\tau)\,.$$

So, the transformations given by (3.12) are twice continuously differentiable, strictly concave on R and satisfy 1^o–5^o, i.e. $\psi_i \in \Psi$.

The truncated logarithmic MBF $\psi_2(t)$ given by formula (3.12) was successfully used (see [3], [5], [6], [17]) for solving large-scale real world NLP problems, including the COPS set (see [22]).

For transformations $\psi \in \Psi$ given by (3.12), we consider their Fenchel conjugate functions

$$\psi_i^*(s) := \begin{cases} \hat{\psi}_i^*(s), & s \leq \hat{\psi}_i'(\tau) \\ q_i^*(s) = (4a_i)^{-1}(s - b_i)^2 - c, & s \geq \hat{\psi}_i'(\tau)\,. \end{cases} \tag{3.13}$$

The class $\varphi = \{\varphi = -\psi^* : R_+ \to R_+\}$, where ψ^* is defined by (3.13) consists of kernels φ with properties established in Theorem 3.2.

On the other hand, the following kernels

$$\begin{aligned} \hat{\varphi}_1(s) &= s\ln s + 1 - s \\ \hat{\varphi}_2(s) &= -\ln s + s - 1 \\ \hat{\varphi}_3(s) &= -2\sqrt{s} + s + 1 \\ \hat{\varphi}_4(s) &= (2 - s)\ln(2 - s) + s\ln s \\ \hat{\varphi}_5(s) &= -2\sqrt{\eta}\left(\sqrt{(2 - s)s} - 1\right) \end{aligned}$$

that correspond to original transformations $\hat{\psi}_1$–$\hat{\psi}_5$ do not satisfy 3(a) and 3(b) because either $m = 0$ or $M = \infty$. We consider $\varphi_i(s) = -\psi_i^*(s)$, $i = 1, \ldots, 5$. The following statement can be verified directly.

Proposition 3.3 *Transformations* ψ_1–ψ_5 *defined by (3.12) satisfy properties* 1^o–5^o *and correspondent kernels* φ_1–φ_5 *possess all properties established in Theorem 3.2.*

We will call the kernel $\varphi \in \varphi$ *well defined* if $0 < \varphi(0) < \infty$. Assuming $t \ln t = 0$ for $t = 0$, we can see that the kernels φ_1 and φ_3–φ_5 are well defined, while the logarithmic MBF kernel φ_2 is not well defined.

4 Convergence of the NR Method

In this section, we present a new convergence proof and estimate the rate of convergence for a wide class of constraint transformation $\psi \in \Psi$ under very mild assumption on the input data. The important ingredients of the convergence proof are the equivalence of the NR method (3.3)–(3.5) to Interior Quadratic Prox (3.10) and to the Prox method (3.7). The critical factors in the convergence proof are properties 3(a) and 3(b) (see Theorem 3.2) of the dual kernel. The proof extends one given in [21] for the special case of Log-Sigmoid transformation.

Let $d = d(\lambda^*) - d(\lambda^0)$, the dual level set $\Lambda_0 = \{\lambda \in \mathbb{R}^q_+ : d(\lambda) \geq d(\lambda^0)\}$ is bounded due to concavity of $d(\lambda)$ and boundedness of L^* (see Corollary 20 in [8]), $L_0 = \max\{\max_{1\leq i\leq q} \lambda_i : \lambda \in \Lambda_0\}$, $I_l^- = \{i : c_i(x^l) < 0\}$, $I_l^+ = \{i : c_i(x^l) \geq 0\}$. We consider the maximum constraints violation

$$\mathrm{v}_l = \max\{-c_i(x^l)/i \in I_l^-\}$$

and the upper bound for the duality gap

$$d_l = \sum_{i=1}^{q} \lambda_i^l |c_i(x^l)| \geq \sum_{i=1}^{q} \lambda_i^l c_i(x^l)$$

at the step l. Let $\bar{\mathrm{v}}_s = \min_{1\leq l\leq s} \mathrm{v}_l$, $\bar{d}_s = \min_{1\leq l\leq s} d_l$. For a bounded closed set $Y \in \mathbb{R}^n$ and $y_0 \bar{\in} Y$ we consider distance $\rho(y_0, Y) = \min\{||y_0 - y||/y \in Y\}$.

Remark 4.1. By adding one extra constraint $c_0(x) = N - f(x) \geq 0$ from Assumption **A** and Corollary 20 in [8] follows the boundedness of Ω. For $N > 0$ large enough the extra constraint does not affect the solution.

Theorem 4.1 *If the standard assumption A and B are satisfied, then*

1) *the dual sequence $\{\lambda^s\}$ is monotone increasing in value, bounded and the following estimation*
$$d(\lambda^{s+1}) - d(\lambda^s) \geq mk^{-1}\|\lambda^{s+1} - \lambda^s\|^2$$
holds true.

2) *the primal sequence $\{x^s\}$ is bounded, $\lim_{l\to\infty} \mathrm{v}_l = 0$, and the following estimation holds*
$$d(\lambda^{s+1}) - d(\lambda^s) \geq kmM^{-2} \sum_{i\in I^-_{s+1}} c_i^2(x^{s+1})$$

3) *for the constraints violation and the duality gap the following estimations*
$$\bar{\mathrm{v}}_s \leq \mathcal{O}\left((sk)^{-0.5}\right), \quad \bar{d}_s \leq \mathcal{O}\left((sk)^{-0.5}\right)$$
hold

4) *the primal-dual sequence $\{x^s, \lambda^s\}$ converges to the primal-dual solution in value, i.e.*
$$f(x^*) = \lim_{s\to\infty} f(x^s) = \lim_{s\to\infty} d(\lambda^s) = d(\lambda^*)\,.$$
and
$$\lim_{s\to\infty} \rho(x^s, X^*) = 0, \quad \lim_{s\to\infty} \rho(\lambda^s, L^*) = 0.$$

Proof. 1) The dual monotonicity follows immediately from (3.10), i.e.,
$$d(\lambda^{s+1}) - \frac{1}{2}k^{-1}||\lambda^{s+1} - \lambda^s||^2_{R_s} \geq d(\lambda^s) - \frac{1}{2}k^{-1}||\lambda^s - \lambda^s||^2_{R_s} = d(\lambda^s)\,. \tag{4.1}$$
Taking into account $||\lambda^{s+1} - \lambda^s||^2_{R_s} > 0$ for $\lambda^{s+1} \neq \lambda^s$, we obtain
$$d(\lambda^{s+1}) \geq d(\lambda^s) + \frac{1}{2}k^{-1}||\lambda^{s+1} - \lambda^s||^2_{R_s} > d(\lambda^s)\,. \tag{4.2}$$
If $d(\lambda^{s+1}) = d(\lambda^s)$, then $||\lambda^{s+1} - \lambda^s||^2_{R_s} = 0 \Rightarrow \lambda^{s+1} = \lambda^s$. Due to the formula (3.4) for the Lagrange multipliers update, $\lambda_i^{s+1} = \lambda_i^s$ leads to $c_i(x^{s+1}) = 0$, $i = 1, \ldots, m$. Hence for the pair (x^{s+1}, λ^{s+1}), the K-K-T's conditions hold. Therefore, $x^{s+1} = x^* \in X^*$,

$\lambda^{s+1} = \lambda^* \in L^*$. In other words, we can either have $d(\lambda^{s+1}) > d(\lambda^s)$ or $\lambda^{s+1} = \lambda^s = \lambda^*$ and $x^{s+1} = x^*$. Recall that due to the boundedness of L^* and concavity of the dual function $d(\lambda)$ the set $\Lambda_0 = \{\lambda : d(\lambda) \geq d(\lambda^0)\}$ is bounded and so is the dual sequence $\{\lambda^s\} \subset \Lambda_0$. Let us find the lower bound for $d(\lambda^{s+1}) - d(\lambda^s)$. From the concavity of $d(\lambda)$ and $-c(x^{s+1}) \in \partial d(\lambda^{s+1})$ we obtain $d(\lambda) - d(\lambda^{s+1}) \leq (-c(x^{s+1}), \lambda - \lambda^{s+1})$ or

$$d(\lambda^{s+1}) - d(\lambda) \geq (c(x^{s+1}), \lambda - \lambda^{s+1})\,. \tag{4.3}$$

Keeping in mind 3^o, we conclude that ψ'^{-1} exists. Using the formula (3.4), we have

$$c_i(x^{s+1}) = (k_i^s)^{-1}\psi'^{-1}(\lambda_i^{s+1}/\lambda_i^s)\,,$$

then using $\psi'^{-1} = \psi^{*\prime}$, we obtain

$$c_i(x^{s+1}) = (k_i^s)^{-1}\psi^{*\prime}(\lambda_i^{s+1}/\lambda_i^s), \quad i = 1, \ldots, m. \tag{4.4}$$

Using $\psi^{*\prime}(1) = \psi^{*\prime}(\lambda_i^s/\lambda_i^s) = 0$ and (4.3) for $\lambda = \lambda^s$ we obtain

$$d(\lambda^{s+1}) - d(\lambda^s) \geq \sum_{i=1}^{q}(k_i^s)^{-1}\left(\psi^{*\prime}\left(\frac{\lambda_i^{s+1}}{\lambda_i^s}\right) - \psi^{*\prime}\left(\frac{\lambda_i^s}{\lambda_i^s}\right)\right)\,(\lambda_i^s - \lambda_i^{s+1})\,. \tag{4.5}$$

Using the mean value formula we have

$$\psi^{*\prime}\left(\frac{\lambda_i^{s+1}}{\lambda_i^s}\right) - \psi^{*\prime}\left(\frac{\lambda_i^s}{\lambda_i^s}\right) = -\psi^{*\prime\prime}(\cdot)(\lambda_i^s)^{-1}(\lambda_i^s - \lambda_i^{s+1}) = \varphi''(\cdot)(\lambda_i^s)^{-1}(\lambda_i^s - \lambda_i^{s+1})\,.$$

Therefore we can rewrite (4.5) as follws

$$d(\lambda^{s+1}) - d(\lambda^s) \geq \sum_{i=1}^{q}(k_i^s\lambda_i^s)^{-1}\varphi''(\cdot)(\lambda_i^s - \lambda_i^{s+1})^2.$$

Keeping in mind the update formula (3.5) and 3(a) from Theorem 3.2, we obtain the following inequality

$$d(\lambda^{s+1}) - d(\lambda^s) \geq mk^{-1}\|\lambda^s - \lambda^{s+1}\|^2, \tag{4.6}$$

which is typical for the Quadratic Prox method (see [9]).

2) We start with the set $I_{l+1}^- = \{i : c_i(x^{l+1}) < 0\}$ of constraints that are violated at the point x^{l+1}. Let's consider $i \in I_{l+1}^-$. Using $\psi^{*\prime}(1) = \psi^{*\prime}(\lambda_i^l/\lambda_i^l) = 0$, the mean value formula, the equation (4.4), the dual kernel property 3(b) from Theorem 3.2 and the update formula (3.5), we obtain

$$-c_i(x^{l+1}) = (k_i^l)^{-1}\left[\psi^{*\prime}\left(\frac{\lambda_i^l}{\lambda_i^l}\right) - \psi^{*\prime}\left(\frac{\lambda_i^{l+1}}{\lambda_i^l}\right)\right] =$$

$$\left(\lambda_i^l k_i^l\right)^{-1} (-\psi^{*\prime\prime}(\cdot)) (\lambda_i^{l+1} - \lambda_i^l) \le k^{-1}\varphi''(\cdot)|\lambda_i^{l+1} - \lambda_i^l| \le k^{-1}M|\lambda_i^{l+1} - \lambda_i^l|$$

or

$$|\lambda_i^{l+1} - \lambda_i^l| \ge kM^{-1}(-c_i(x^{l+1}))\,, \quad i \in I_{l+1}^-.$$

Combining the last bound with (4.6) we obtain

$$d(\lambda^{l+1}) - d(\lambda^l) \ge kmM^{-2} \sum_{i \in I_{l+1}^-} c_i^2(x^{l+1})\,. \tag{4.7}$$

3) Let us consider $\mathrm{v}_{l+1} = \max_{i \in I_{l+1}^-}(-c_i(x^{l+1}))$—the maximum constraints violation at the step $l+1$, then from (4.7) we have

$$d(\lambda^{l+1}) - d(\lambda^l) \ge kmM^{-2}\mathrm{v}_{l+1}^2. \tag{4.8}$$

Summing up (4.8) from $l = 1$ to $l = s$ we obtain

$$d = d(\lambda^*) - d(\lambda^0) \ge d(\lambda^{s+1}) - d(\lambda^0) \ge kmM^{-2}\sum_{l=0}^{s} \mathrm{v}_{l+1}^2\,.$$

Therefore, $\mathrm{v}_l \to 0$. Remembering that $\bar{\mathrm{v}}_s = \min\{\mathrm{v}_l | 1 \le l \le s\}$, we obtain

$$\bar{\mathrm{v}}_s \le M\sqrt{dm^{-1}}(ks)^{-0.5} = \mathcal{O}\left((ks)^{-0.5}\right)\,. \tag{4.9}$$

The primal asymptotic feasibility follows from $\mathrm{v}_l \to 0$.

Using arguments similar to those we used to prove Theorem 2 in [21] we can show the estimate (4.9) for the duality gap $\bar{d}_s$.

4) It follows from (4.6) that the sequence $\{d(\lambda^s)\}$ is monotone icreasing and $d(\lambda^s) \le f(x^*)$, $s \ge 1$. Therefore there is $\lim_{s\to\infty} d(\lambda^s) = d(\lambda^\infty) \le f(x^*)$ and again from (4.6) we have

$$\lim_{s\to\infty} \|\lambda^{s+1} - \lambda^s\| = 0. \tag{4.10}$$

Due to concavity of $d(\lambda)$ and boundedness of L^*, the sequence $\{\lambda^s\}$ is bounded (see Corollary 20 in [8]). Therefore there is a converging subsequence $\{\lambda^{s_l}\}$ that

$$\lim_{s_l\to\infty} \lambda^{s_l} = \lim_{s_l\to\infty} \lambda^{s_l+1} = \bar{\lambda}. \tag{4.11}$$

From assumption **A**, Remark 4.1, Corollary 20 in [8] and $v_l \to 0$ follows the boundedness of the primal sequence $\{x^s\}$. Without loosing the generality we can assume that $\lim_{s_l\to\infty} x^{s_l+1} = \bar{x}$. We consider two sets of indices $I_+ = \{i : \bar{\lambda}_i > 0\}$ and $I_0 = \{i : \bar{\lambda}_i = 0\}$. From (4.11) and (3.5) we have $\bar{k}_i = \lim_{s_l\to\infty} k_i^{s_l} = k \lim_{s_l\to\infty} (\lambda_i^{s_l})^{-1} = k(\bar{\lambda}_i)^{-1}, \quad i \in I_+$. From (3.4) we obtain

$$c_i(x^{s_l+1}) = k^{-1}\lambda_i^{s+l}\psi'^{-1}(\lambda_i^{s_l+1}/\lambda_i^{s_l}) = -k^{-1}\lambda_i^{s+l}\varphi'^{-1}(\lambda_i^{s_l+1}/\lambda_i^{s_l}). \tag{4.12}$$

Passing (4.12) to the limit, we obtain

$$c_i(\bar{x}) = -k^{-1}\bar{\lambda}_i\varphi'(1) = 0, \quad i \in I_+ \tag{4.13}$$

From $v_s \to 0$ we obtain $\lim_{s_l\to\infty} c_i(x^{s_l+1}) = c_i(\bar{x}) \geq 0, \quad i \in I_0$. In view of (3.3)-(3.4) we have $\nabla_x \mathcal{L}(x^{s_l+1}, \lambda^{s_l}, \boldsymbol{k}^{s_l}) = \nabla_x L(x^{s_l+1}, \lambda^{s_l+1}) = 0$. Therefore for any limit point $(\bar{x}, \bar{\lambda})$ the KKT's conditions are satisfied. Hence, $\bar{x} = x^*$, $\bar{\lambda} = \lambda^*$. From the dual monotonicity, we obtain that the entire dual sequence $\{\lambda^s\}$ converges to the dual solution in value, i.e. $\lim_{s\to\infty} d(\lambda^s) = d(\lambda^*)$. Using considerations similar to those we used to prove (4.13) we obtain the asymptotic complementarity conditions for the entire primal-dual sequence, i.e.

$$\lim_{s_l\to\infty} \lambda_i^s c_i(x^s) = 0, \quad i = 1, \ldots, q \tag{4.14}$$

From (4.14) we obtain

$$d(\lambda^*) = \lim_{s\to\infty} d(\lambda^s) = \lim_{s\to\infty} L(x^s, \lambda^s) = \lim_{s\to\infty} f(x^s) = f(x^*)\,. \tag{4.15}$$

So we proved the first part of statement 4). The second part of the statement 4) follows directly from (4.15) boundedness of $X^* = \{x \in \Omega : f(x) \leq f(x^*)\}$ and $L^* = \{\lambda \in \mathbb{R}^r_{++} : d(\lambda) \geq d(\lambda^*)\}$, and Lemma 11 (see [18], Chapter 9, §1).

We would like to mention again that the results in part 4) of Theorem 4.1 have been proven by other means in [4] (see Theorem 1) and [2] (see Theorem 5.1).

Remark 4.2. The estimation (4.7) is the critical part of the convergence proof. It indicates how the NR method (3.3)-(3.5) transforms the primal constraints violation into the increase of the dual function value. Keeping in mind the dual monotonicity and the boundedness of the dual function it became evident from (4.7) that the constraints violation has to vanish. So the estimation (4.7) simplify substantially the proof of the primal asymptotic feasibility, which always has been the most difficult part of the convergence proof. On the other hand, it allows to understand better the pricing mechanism of NR method (3.3)-(3.5). We would like to emphasize that for transformations $\hat{\psi}_1 - \hat{\psi}_5$ either $m = 0$ or $M = \infty$, which makes the estimation (4.7) trivial and useless. The modification (3.12) provides properties 3^o and 4^o for the primal transformation and properties 3a) and 3b) (see Theorem 3.2) for the dual kernels, which are critical for the convergence proof. Also, the strong convexity of the dual kernel (property 3a)) allows to prove (4.6), which is another important element of the convergence proof.

Remark 4.3. It follows from (4.7) that for any given $\tau < 0$ and any $i = 1, \dots, q$ the inequality $c_i(x^{s+1}) \leq \tau$ is possible only for a finite number of steps. So the quadratic branch in the modification (3.12) can be used only a finite number of times. In fact, for $k > 0$ large enough, just once. Therefore, from some point on only original transformations $\hat{\psi}_1 - \hat{\psi}_5$ are used in the NR method (3.3)-(3.5) and only kernels $\hat{\varphi}_1 - \hat{\varphi}_5$ that correspond to the original transformations $\hat{\psi}_1$–$\hat{\psi}_5$ are used in the Interior Prox method with φ-divergence distance (3.7). Transformations $\hat{\psi}_1$–$\hat{\psi}_5$ for $t \geq \tau$ are infinitely differentiable and so are the Lagrangians $\mathcal{L}(x, \lambda, \boldsymbol{k})$ if the input data possesses the correspondent property. This is an important advantage, because it allows to use the Newton method for primal minimization or for solving the primal-dual system [20].

To the best of our knowledge, the strongest result so far under the assumptions A and B for the Interior Prox method (3.7) was obtained in [2]. It was proven that for the regularized MBF kernel $\varphi(t) = 0.5\nu(t-1)^2 + \mu(t - \ln t - 1)$, $\mu > 0$, $\nu > 0$ the method (3.7) produces a convergent sequence $\{\lambda^s\}$ and the rate of convergence in value is $\mathcal{O}(ks)^{-1}$. In the next section, we strengthen the estimation under some extra assumptions on the input data.

5 Rate of Convergence

.

We will say that for the primal-dual sequence $\{x^s, \lambda^s\}_{s=0}^{\infty}$, which converges to the primal-dual solution (x^*, λ^*), the complementarity condition is satisfied in the strict form if

$$\max\{\lambda_i^*, c_i(x^*)\} > 0, \quad i = 1, \dots, q. \tag{5.1}$$

Theorem 5.1 *If for the primal-dual sequence generated by (3.3)–(3.5) the complementarity condition is satisfied in strict form (5.1), then for any fixed $k > 0$ the following estimation holds true*

$$d(\lambda^*) - d(\lambda^s) = o(ks)^{-1}.$$

Proof. Recall that $I^* = \{i : c_i(x^*) = 0\} = \{1, \dots, r\}$ is the active constraint set, then $\min\{c_i(x^*) \mid i \bar{\in} I^*\} = \sigma > 0$. Therefore, there is such a number s_0 that $c_i(x^s) \ge \frac{\sigma}{2}$, $s \ge s_0$, $i \bar{\in} I^*$. From 2^oc) and (3.5) we have

$$\begin{aligned} &\text{a)} \qquad \lambda_i^{s+1} = \lambda_i^s \psi'(k_i^s c_i(x^{s+1})) \le a\lambda_i^s (k_i^s c_i(x^{s+1}))^{-1} = 2a(\sigma k)^{-1} (\lambda_i^s)^2 \text{ and} \\ &\text{b)} \qquad k_i^{s+1} = k(\lambda_i^{s+1})^{-1} \ge \frac{\sigma}{2a}(k_i^s)^2 \text{ and } k_i^s \to \infty, \quad i \bar{\in} I^* = \{1, \dots, r\}. \end{aligned} \tag{5.2}$$

Then,

$$\mathcal{L}(x, \lambda^s, \boldsymbol{k}^s) = f(x) - k^{-1} \sum_{i=1}^{r} (\lambda_i^s)^2 \psi(k_i^s c_i(x)) - k^{-1} \sum_{i=r+1}^{m} (\lambda_i^s)^2 \psi(k_i^s c_i(x))\,.$$

Let us estimate the last term. We have

$$\begin{aligned} k^{-1} \sum_{i=r+1}^{q} (\lambda_i^s)^2 \psi(k_i^s c_i(x)) &= k^{-1} \sum_{i=r+1}^{q} (\lambda_i^s)^2 (\psi(k_i^s c_i(x)) - \psi(0)) \\ &= k^{-1} \sum_{i=r+1}^{q} (\lambda_i^s)^2 k_i^s c_i(x) \psi'(\theta_i^s k_i^s c_i(x)) \end{aligned}$$

and $0 < \theta_i^s < 1$. For $k_i^s \to \infty$ and $c_i(x) \ge 0.5\sigma$, we obtain $\theta_i^s \to 1$. Therefore, for s_0 large enough and any $s \ge s_0$, we have $\theta_i^s \ge 0.5$. Thus, using 2^oc) and (3.5) we obtain

$$\begin{aligned} k^{-1} \sum_{i=r+1}^{q} (\lambda_i^s)^2 \psi(k_i^s c_i(x)) &= \sum_{i=r+1}^{q} \lambda_i^s c(x) \psi'(\theta_i^s k_i^s c(x)) \\ &\le a \sum_{i=r+1}^{q} \lambda_i^s c(x) (\theta_i^s k_i^s c_i(x))^{-1} \\ &\le 2a \sum_{i=r+1}^{q} \lambda_i^s (k_i^s)^{-1} = 2ak^{-1} \sum_{i=r+1}^{q} (\lambda_i^s)^2. \end{aligned}$$

From (5.2) we have

$$\lambda_i^{s+1} = \mathcal{O}(\lambda_i^s)^2, \quad i = 1, \dots, r\,.$$

So, for s_0 large enough and any $s \geq s_0$, the last term is negligibly small, and instead of $\boldsymbol{\mathcal{L}}(x, \lambda, \boldsymbol{k})$ we can consider the truncated Lagrangian for the equivalent problem $\boldsymbol{\mathcal{L}}(x, \lambda, \boldsymbol{k}) := f(x) - \sum_{i=1}^r (k_i^s)^{-1}(\lambda_i^s)\psi(k_i^s c_i(x))$ and the correspondent truncated Lagrangian $L(x, \lambda) := f(x) - \sum_{i=1}^r \lambda_i c_i(x)$ for the original problem (P). Accordingly, instead of the original dual function and the second order φ-divergence distance, we consider the dual function $d(\lambda) := \inf_{x \in \mathbb{R}^n} L(x, \lambda)$ and the second order φ-divergence distance $D_2(u, v) := \sum_{i=1}^r v_i^2 \varphi(u_i/v_i)$ in the truncated dual space $\mathbb{R}^r$.

For simplicity, we retain the previous notations for the truncated Lagrangians, for both the original and the equivalent problems, correspondent dual function and the second order φ-divergence distance. Below we will assume that $\{\lambda^s\}_{s=1}^\infty$, is the truncated dual sequence, i.e., $\lambda^s = (\lambda_1^s, \dots, \lambda_r^s)$. Let us consider the optimality criteria for the truncated Interior Prox method

$$\lambda^{s+1} = \arg\max\left\{ d(\lambda) - k^{-1} \sum_{i=1}^r (\lambda_i^s)^2 \varphi(\lambda_i/\lambda_i^s) \mid \lambda \in \mathbb{R}^r \right\}.$$

We have

$$c(x^{s+1}) + k^{-1} \sum_{i=1}^r \lambda_i^s \varphi'(\lambda_i^{s+1}/\lambda_i^s) e_i = 0\,, \tag{5.3}$$

where $e_i = (0, \dots, 1, \dots, 0) \in \mathbb{R}^r$. Using $\psi^{*\prime}(1) = \psi^{*\prime}(\lambda_i^s/\lambda_i^s) = 0$ we can rewrite (5.3) as follows

$$c(x^{s+1}) + k^{-1} \sum_{i=1}^r \lambda_i^s (\varphi'(\lambda_i^{s+1}/\lambda_i^s) - \varphi'(\lambda_i^s/\lambda_i^s)) e_i = 0\,.$$

Using the mean value formula, we obtain

$$c(x^{s+1}) + k^{-1} \sum_{i=1}^r \varphi'' \left(1 + \left(\frac{\lambda_i^{s+1}}{\lambda_i^s} - 1 \right) \theta_i^s \right) (\lambda_i^{s+1} - \lambda_i^s) e_i = 0\,, \tag{5.4}$$

where $0 < \theta_i^s < 1$. We recall that $-c(x^{s+1}) \in \partial d(\lambda^{s+1})$, so (5.4) is the optimality criteria for the following problem in the truncated dual space

$$\lambda^{s+1} = \arg\max\{ d(\lambda) - 0.5k^{-1} \|\lambda - \lambda^s\|_{R_s}^2 \mid \lambda \in \mathbb{R}^r \}\,, \tag{5.5}$$

where $\|x\|_R = x^T R x$, $R_s = \mathrm{diag}(r_i^s)_{i=1}^r$ and $r_i^s = \varphi'' \left(1 + \left(\frac{\lambda_i^{s+1}}{\lambda_i^s} - 1 \right) \theta_i^s \right)$, $i = 1, \dots, r$. Due to $\lambda_i^s \to \lambda_i^* > 0$, we have $\lim_{s\to\infty} \lambda_i^{s+1} (\lambda_i^s)^{-1} = 1$, $i = 1, \dots, r$. Keeping in mind

Theorem 3.2 we have

$$\lim_{s\to\infty} r_i^s = \lim_{s\to\infty} \varphi''\left(1+\left(\frac{\lambda_i^{s+1}}{\lambda_i^s}-1\right)\theta_i^s\right) = \varphi''(1) \geq m > 0\,.$$

Therefore, due to the continuity of $\varphi''(\cdot)$ for $s_0 > 0$ large enough and any $s > s_0$ we have $m \leq r_i^s < 2\varphi''(1) = m_1$ and $||R_s|| \leq m_1$.

We will show now that the convergence analysis, which is typical for the Quadratic Prox method (see [9]), can be extended for the Interior Quadratic Prox method (3.10) in the truncated dual space. From (4.6) we have

$$d(\lambda^*) - d(\lambda^s) - (d(\lambda^*) - d(\lambda^{s+1})) \geq mk^{-1}\|\lambda^s - \lambda^{s+1}\|^2$$

or

$$v_s - v_{s+1} \geq mk^{-1}\|\lambda^s - \lambda^{s+1}\|^2, \tag{5.6}$$

where $v_s = d(\lambda^*) - d(\lambda^s) > 0$. Using the concavity of $d(\lambda)$ we obtain

$$d(\lambda) - d(\lambda^{s+1}) \leq (-c(x^{s+1}), \lambda - \lambda^{s+1})$$

or $d(\lambda^{s+1}) - d(\lambda) \geq (c(x^{s+1}), \lambda - \lambda^{s+1})$. Using (5.4), we obtain

$$d(\lambda^{s+1}) - d(\lambda) \geq -k^{-1}(R_s(\lambda^{s+1} - \lambda^s), \lambda - \lambda^{s+1})\,.$$

So, for $\lambda = \lambda^*$ we have

$$k^{-1}(R_s(\lambda^{s+1} - \lambda^s), \lambda^* - \lambda^{s+1}) \geq d(\lambda^*) - d(\lambda^{s+1}) = v_{s+1}$$

or

$$k^{-1}(R_s(\lambda^{s+1} - \lambda^s), \lambda^* - \lambda^s) - k^{-1}\|\lambda^{s+1} - \lambda^s\|^2_{R_s} \geq v_{s+1}.$$

Hence,

$$\|R_s\| \cdot \|\lambda^{s+1} - \lambda^s\| \cdot \|\lambda^* - \lambda^s\| \geq kv_{s+1}$$

or

$$\|\lambda^{s+1} - \lambda^s\| \geq \frac{1}{m_1}kv_{s+1}\|\lambda^s - \lambda^*\|^{-1}. \tag{5.7}$$

From (5.6) and (5.7), it follows that

$$v_s - v_{s+1} \geq \frac{m}{m_1^2}kv_{s+1}^2\|\lambda^s - \lambda^*\|^{-2}$$

or

$$v_s \geq v_{s+1}\left(1+\frac{m}{m_1^2}kv_{s+1}\|\lambda^s-\lambda^*\|^{-2}\right).$$

By inverting the last inequality we obtain

$$v_s^{-1} \leq v_{s+1}^{-1}\left(1+\frac{m}{m_1^2}kv_{s+1}\|\lambda^s-\lambda^*\|^{-2}\right)^{-1}. \tag{5.8}$$

Furthermore, from (5.5) we obtain

$$d(\lambda^{s+1})-0.5k^{-1}\|\lambda^{s+1}-\lambda^s\|^2_{R_s} \geq d(\lambda^*)-0.5k^{-1}\|\lambda^*-\lambda^s\|^2_{R_s}$$

or

$$v_{s+1} \leq 0.5k^{-1}\|\lambda^s-\lambda^*\|^2_{R_s} \leq 0.5k^{-1}m_1\|\lambda^s-\lambda^*\|^2.$$

Therefore, $m_1^{-1}kv_{s+1}\|\lambda^*-\lambda^s\|^{-2} \leq 1$ or $mm_1^{-2}kv_{s+1}\|\lambda^*-\lambda^s\|^{-2} \leq mm_1^{-1} < 1$.

Let us consider the function $(1+t)^{-1}$. It is easy to see that $(1+t)^{-1} \leq -0.5t+1$, $0 \leq t \leq 1$. Using the last inequality for $t = mm_1^{-2}kv_{s+1}\|\lambda^{s+1}-\lambda^*\|^{-2}$ from (5.8) we obtain

$$v_s^{-1} \leq v_{s+1}^{-1}\left(1-0.5mm_1^{-2}kv_{s+1}\|\lambda^{s+1}-\lambda^*\|^{-2}\right)$$

or

$$v_i^{-1} \leq v_{i+1}^{-1}-0.5mm_1^{-2}k\|\lambda^{i+1}-\lambda^*\|^{-2}, \quad i=0,\dots,s-1. \tag{5.9}$$

Summing up (5.9) for $i=1,\dots,s-1$ we obtain

$$v_s^{-1} \geq v_s^{-1}-v_0^{-1} \geq 0.5mm_1^{-2}k\sum_{i=1}^{s-1}\|\lambda^i-\lambda^*\|^{-2}.$$

By inverting the last inequality we obtain

$$v_s = d(\lambda^*)-d(\lambda^s) \leq \frac{2m^{-1}m_1^2}{k\sum_{i=0}^{s-1}\|\lambda^i-\lambda^*\|^{-2}}$$

or

$$ksv_s = \frac{2m^{-1}m_1^2}{s^{-1}\sum_{i=0}^{s-1}\|\lambda^i-\lambda^*\|^{-2}}.$$

From $\|\lambda^s-\lambda^*\| \to 0$ it follows $\|\lambda^s-\lambda^*\|^{-2} \to \infty$. Using the Silverman-Toeplitz theorem [12], we have $\lim_{s\to\infty} s^{-1}\sum_{i=1}^{s}\|\lambda^i-\lambda^*\|^{-2} = \infty$. Therefore, there exists $\alpha_s \to 0$ such that

$$v_s = m^{-1}m_1^2\alpha_s\frac{1}{ks} = o\left((ks)^{-1}\right). \tag{5.10}$$

We completed the proof of Theorem 5.1.

The estimation (5.10) can be strengthened. Under the standard second order optimality conditions, the method NR (3.3)–(3.5) converges with Q-linear rate if $k > 0$ is fixed but large enough.

First of all, due to the standard second order optimality conditions, the primal-dual solution is unique. Therefore, the primal-dual sequence $\{x^s, \lambda^s\}\}$ converges to the primal-dual solution (x^*, λ^*), for which the complementarity conditions are satisfied in a strict form (5.1). Therefore, the Lagrange multipliers for the passive constraints converge to zero quadratically. From (3.5) we have $\lim_{s\to\infty} k_i^s = k(\lambda_i^*)^{-1}$, $i = 1, \ldots, r$, i.e., the scaling parameters, which correspond to the active constraints, grow linearly with $k > 0$. Therefore, the technique used in [19], [20] can be applied for the asymptotic analysis of the method (3.3)–(3.5).

For a given small enough $\delta > 0$, we define the extended neighborhood of λ^* as follows

$$D(\lambda^*, \boldsymbol{k}, \delta) = \{(\lambda, \boldsymbol{k}) \in \mathbb{R}^m_+ \times \mathbb{R}^m_+ : \lambda_i \geq \delta,\ |\lambda_i - \lambda_i^*| \leq \delta k_i,\ i = 1, \ldots, r;\ k \geq k_0$$
$$0 \leq \lambda_i \leq k_i \delta,\ i = r+1, \ldots, m\}\,.$$

Proposition 5.2 *If f and all $c_i \in C^2$ and the standard second order optimality conditions hold, then there exists such small $\delta > 0$ and large enough $k_0 > 0$ that for any $(\lambda, \boldsymbol{k}) \in D(\cdot)$:*

1. *There exists $\hat{x} = \hat{x}(\lambda, \boldsymbol{k}) = \arg\min\{\mathcal{L}(x, \lambda, \boldsymbol{k}) \mid x \in \mathbb{R}^n\}$ such that*

$$\nabla_x \mathcal{L}(\hat{x}, \lambda, \boldsymbol{k}) = 0$$

and

$$\hat{\lambda}_i = \lambda_i \psi'(k_i c_i(\hat{x})),\ \hat{k}_i = k\hat{\lambda}_i^{-1},\ i = 1, \ldots, m.$$

2. *For the pair $(\hat{x}, \hat{\lambda})$ the estimate*

$$\max\{\|\hat{x} - x^*\|,\ \|\hat{\lambda} - \lambda^*\|\} \leq ck^{-1}\|\lambda - \lambda^*\| \tag{5.11}$$

holds and $c > 0$ is independent on $k \geq k_0$.

3. The Lagrangian $\mathcal{L}(x, \lambda, \boldsymbol{k})$ is strongly convex in the neighborhood of $\hat{x}$.

Theorem 5.2 can be proven by a slight modification of the correspondent proof in [19] (see also [20]).

Finding x^{s+1} requires solving an unconstrained minimization problem (3.3), which is generally speaking an infinite procedure. To make the multipliers method (3.3)-(3.5) practical one has to replace x^{s+1} by and approximation $\bar{x}^{s+1}$, which can be found by a finite number of Newton steps. It turns out that if $\bar{x}^{s+1}$ is used in the formula (3.4) for the Lagrange multipliers update instead of x^{s+1} then the bound similar to (5.11) remtains true.

For a given $\sigma > 0$ let us consider the sequence $\{\bar{x}^s, \bar{\lambda}^s, \boldsymbol{k}^s\}$ generated by the following formulas.

$$\bar{x} \, : \, \|\nabla_x \mathcal{L}(\bar{x}^{s+1}, \bar{\lambda}^s, \boldsymbol{k}^s)\| \leq \sigma k^{-1} \|\Psi' \left(k(\bar{\lambda}^s)^{-1} c(\bar{x}^{s+1}) \right) \bar{\lambda}^s - \bar{\lambda}^s\| \tag{5.12}$$

$$\bar{\lambda}^{s+1} = \Psi' \left(k(\bar{\lambda}^s)^{-1} c(\bar{x}^{s+1}) \right) \bar{\lambda}^s, \tag{5.13}$$

where

$$\Psi' \left(k(\bar{\lambda}^s)^{-1} c(\bar{x}^{s+1}) \right) = \text{diag} \left(\psi(k(\bar{\lambda}_i^s)^{-1} c_i(\bar{x}^{s+1})) \right)_{i=1}^q$$

and

$$\bar{\boldsymbol{k}}^{s+1} = (\bar{k}_i^{s+1} = k(\lambda_i^{s+1})^{-1}, \, i = 1, \ldots, q).$$

By using arguments similar similar to those in [20], we can prove the following proposition.

Proposition 5.3 *If the standard second order optimality conditions hold and the Hessians $\nabla^2 f$ and $\nabla^2 c_i$, $i = 1, \ldots, m$ satisfy Lipschitz conditions*

$$\|\nabla^2 f(x) - \nabla^2 f(y)\| \leq L_0 \|x - y\|, \; \|\nabla^2 c_i(x) - \nabla^2 c_i(y)\| \leq L_i \|x - y\|, \; i = 1, \ldots, m \tag{5.14}$$

then there is $k_0 > 0$ large enough, that for the primal-dual sequence $\{\bar{x}^s, \bar{\lambda}^s\}$ generated by the formulas (5.12)-(5.13) the following estimations hold true and $c > 0$ is independent on $k \geq k_0$ for $s \geq 0$.

$$\|\bar{x} - x^*\| \leq c(1 + \sigma) k^{-1} \|\lambda - \lambda^*\|, \quad \|\bar{\lambda} - \lambda^*\| \leq c(1 + \sigma) k^{-1} \|\lambda - \lambda^*\|. \tag{5.15}$$

We used the stopping criteria (5.12)for the inner iteration and formula (5.13) for the Lagrange multipliers update. To measure the distance between the current approximation $y = (x, \lambda)$ and the primal-dual solution $y^* = (x^*, \lambda^*)$ we use the following merit function

$$\nu(x, \lambda) = \max \{\|\nabla_x L(x, \lambda)\|, -c_i(x), \lambda_i |c_i(x)|, i = 1, \dots, q\}.$$

It is easy to see that $\nu(x, \lambda) = 0 \equiv x = x^*, \lambda = \lambda^*$. We used the following final stopping criteria

$$\nu(x, \lambda) \le \varepsilon, \tag{5.16}$$

where $\varepsilon > 0$ the desired accuracy.

In the next section, we apply the NR method (5.1)-(5.3) for Linear Programming. The convergence under very mild assumption follows from Theorem 4.1. Under the dual uniqueness, we prove the global quadratic convergence. The key ingredients of the proof are the A. Hoffman-type lemma ([11], see also [18], Ch. 10, §1) and the properties of the well defined kernels $\varphi \in \mathcal{\varphi}$.

6 Nonlinear Rescaling Method for Linear Programming

Let $A : \mathbb{R}^n \to \mathbb{R}^m$, $a \in \mathbb{R}^n$, $b \in \mathbb{R}^m$. We assume that

$$X^* = Argmin\{(a, x) \mid c_i(x) = (Ax - b)_i = (a_i, x) - b_i \ge 0, \ i = 1, \dots, q\} \tag{6.1}$$

and

$$L^* = Argmax\{(b, \lambda) \mid A^T \lambda - a = 0, \ \lambda_i \ge 0, \ i = 1, \dots, q\} \tag{6.2}$$

are nonempty and bounded.

We consider $\psi \in \Psi$, which corresponds to the well-defined kernel $\varphi \in \mathcal{\varphi}$, i.e., $0 < \varphi(0) < \infty$.

The NR method (3.3)–(3.5) being applied to (6.1) produces three sequences $\{x^s\} \subset \mathbb{R}^n$, $\{\lambda^s\} \subset \mathbb{R}^q_{++}$ and $\{\boldsymbol{k}^s\} \in \mathbb{R}^q_{++}$:

$$x^{s+1} : \nabla_x \mathcal{L}(x^{s+1}, \lambda^s, \boldsymbol{k}^s) \;=\; a - \sum_{i=1}^{m} \lambda_i^s \psi' \left(k_i^s c_i(x^{s+1}) \right) a_i = 0\,, \tag{6.3}$$

$$\lambda^{s+1} : \lambda_i^{s+1} \;=\; \lambda_i^s \psi' \left(k_i^s c_i(x^{s+1}) \right) a_i = 0\,,\; i = 1, \ldots, q \tag{6.4}$$

$$\boldsymbol{k}^{s+1} : k_i^{s+1} \;=\; k(\lambda_i^{s+1})^{-1},\; i = 1, \ldots, q\,. \tag{6.5}$$

If X^* and L^* are bounded, then all statements of Theorem 4.1 are taking place for the primal-dual sequence $\{x^s, \lambda^s\}$ generated by (6.3)–(6.5). In particular,

$$\lim_{s\to\infty} (a, x^s) = (a, x^*) = \lim_{s\to\infty} (b, \lambda^s) = (b, \lambda^*)\,.$$

Using Lemma 5 (see [18], Ch. 10, §1), we can find such $\alpha > 0$ that

$$(b, \lambda^*) - (b, \lambda^s) \geq \alpha \rho(\lambda^s, L^*)\,. \tag{6.6}$$

Therefore, $\lim_{s\to\infty} \rho(\lambda^s, L^*) = 0$.

If λ^* is a unique dual solution, then the same Lemma 5 guarantees the existence of $\alpha > 0$ that the following inequality

$$(b, \lambda^*) - (b, \lambda) = \alpha \|\lambda - \lambda^*\| \tag{6.7}$$

holds true for $\forall \lambda \in L = \{\lambda : A^T\lambda = a,\, \lambda \in \mathbb{R}^m_+\}$.

Theorem 6.1 *If the dual problem (6.2) has a unique solution, then for any well defined $\varphi \in \varphi$ the dual sequence $\{\lambda^s\}$ converges in value quadratically, i.e., there is $c > 0$ independent on $k > 0$ that the following estimation*

$$(b, \lambda^*) - (b, \lambda^{s+1}) \leq ck^{-1}[(b, \lambda^*) - (b, \lambda^s)]^2 \tag{6.8}$$

holds true.

Proof. It follows from (6.3)–(6.5) that

$$\nabla_x \mathcal{L}(x^{s+1}, \lambda^s, \boldsymbol{k}^s) = a - \sum_{i=1}^{q} \lambda_i^{s+1} a_i = a - A^T\lambda^{s+1} = 0 \tag{6.9}$$

and $\lambda^{s+1} \in \mathbb{R}^m_{++}$. In other words, the NR method generates a dual interior point sequence $\{\lambda^s\}_{s=0}^{\infty}$.

From (6.9) we obtain

$$0 = (a - A^T\lambda^{s+1}, x^{s+1}) = (a, x^{s+1}) - \sum_{i=1}^{q} \lambda_i^{s+1} c_i(x^{s+1}) - (b, \lambda^{s+1})$$

or

$$(b, \lambda^{s+1}) = L(x^{s+1}, \lambda^{s+1})\,.$$

The multipliers method (6.3)–(6.5) is equivalent to the following Interior Prox for the dual problem

$$\lambda^{s+1} = \arg\max \left\{ (b, \lambda) - k^{-1} \sum_{i=1}^{q} (\lambda_i^s)^2 \varphi\left(\frac{\lambda_i}{\lambda_i^s}\right) \mid A^T\lambda - a = 0 \right\}. \qquad (6.10)$$

Keeping in mind Remark 4.3, we can assume without restricting the generality that only kernels φ_i, $i = 1, \ldots, 5$. which corresponds to the original transformations $\hat{\psi}_1$–$\hat{\psi}_5$ are used in the method (6.10). Moreover, we consider only $\varphi_i \in \mathcal{\varphi}$, which are well defined, i.e. φ_1, φ_3–φ_5.

From (6.10), taking into account $\lambda^* \in \mathbb{R}^m_+$ and $A^T\lambda^* = a$, we obtain

$$(b, \lambda^{s+1}) - k^{-1} \sum_{i=1}^{q} (\lambda_i^s)^2 \varphi\left(\frac{\lambda_i^{s+1}}{\lambda_i^s}\right) \geq (b, \lambda^*) - k^{-1} \sum_{i=1}^{q} (\lambda_i^s)^2 \varphi\left(\frac{\lambda_i^*}{\lambda_i^s}\right).$$

Keeping in mind $k^{-1} \sum_{i=1}^{q} (\lambda_i^s)^2 \varphi\left(\frac{\lambda_i^{s+1}}{\lambda_i^s}\right) \geq 0$, we have

$$k^{-1} \sum_{i=1}^{q} (\lambda_i^s)^2 \varphi\left(\frac{\lambda_i^*}{\lambda_i^s}\right) \geq (b, \lambda^*) - (b, \lambda^{s+1})\,. \qquad (6.11)$$

Let us assume that $\lambda_i^* > 0$, $i = 1, \ldots, r$; $\lambda_i^* = 0$, $i = r+1, \ldots, q$. Then $\varphi\left(\frac{\lambda_i^*}{\lambda_i^s}\right) = \varphi(0) < \infty$, $i = r+1, \ldots, q$. Keeping in mind $\varphi(\lambda_i^s/\lambda_i^s) = \varphi'(\lambda_i^s/\lambda_i^s) = 0$, $i = 1, \ldots, r$ and using the mean value formula twice, we obtain

$$k^{-1} \sum_{i=1}^{q} (\lambda_i^s)^2 \varphi\left(\frac{\lambda_i^*}{\lambda_i^s}\right) = k^{-1} \left[\sum_{i=1}^{r} (\lambda_i^s)^2 \left(\varphi\left(\frac{\lambda_i^*}{\lambda_i^s}\right) - \varphi\left(\frac{\lambda_i^s}{\lambda_i^s}\right) \right) + \varphi(0) \sum_{i=r+1}^{q} (\lambda_i^* - \lambda_i^s)^2 \right] \leq$$

$$k^{-1} \left[\sum_{I=1}^{r} \varphi''\left(1 + \bar{\bar{\theta}}_i^s \bar{\theta}_i^s \frac{\lambda_i^* - \lambda_i^s}{\lambda_i^s}\right) (\lambda_i^* - \lambda_i^s)^2 + \varphi(0) \sum_{i=r+1}^{q} (\lambda_i^* - \lambda_i^s)^2 \right],$$

where $0 < \bar{\theta}_i^s < 1$, $0 < \bar{\bar{\theta}}_i^s < 1$.

Taking into account the dual uniqueness from (6.6), we obtain $\lim_{s\to\infty} \lambda_i^s = \lambda_i^* > 0$, $i = 1,\ldots,r$. Therefore, $\lim_{s\to\infty} \varphi''\left(1 + \bar{\bar{\theta}}_i^s \bar{\theta}_i^s \frac{\lambda_i^* - \lambda_i^s}{\lambda_i^s}\right) = \varphi''(1)$, $i = 1,\ldots,r$. Hence, there is $s_0 > 0$ such that for any $s \geq s_0$ we have $m \leq \varphi''(\cdot) \leq m_1$. Then, for $\varphi_0 = \max\{m_1, \varphi(0)\}$, we obtain

$$k^{-1}\sum_{i=1}^{q}(\lambda_i^s)^2\varphi\left(\frac{\lambda_i^*}{\lambda_i^s}\right) \leq \varphi_0 k^{-1}\|\lambda^* - \lambda^s\|^2. \tag{6.12}$$

Combining (6.11) and (6.12) we have

$$\varphi_0 k^{-1}\|\lambda^* - \lambda^s\|^2 \geq (b, \lambda^*) - (b, \lambda^{s+1})\,. \tag{6.13}$$

From (6.7) with $\lambda = \lambda^s$ we obtain

$$\|\lambda^s - \lambda^*\| = \alpha^{-1}[(b, \lambda^*) - (b, \lambda^s)]\,.$$

Therefore, the following estimation

$$(b, \lambda^*) - (b, \lambda^{s+1}) \leq ck^{-1}[(b, \lambda^*) - (b, \lambda^s)]^2 \tag{6.14}$$

holds true with $c = \varphi_0\alpha^{-2}$ for any $s \geq s_0$.

It follows from Theorem 4.1 that by taking $k > 0$ large enough, we can make $s_0 = 1$. Also one can make the rate of convergence of the NR method for LP superquadratic by increasing the penalty parameter $k > 0$ from step to step.

Remark 6.1. Theorem 6.1 is valid for the NR method (6.3)–(6.5) with exponential, LS, CHKS and hyperbolic MBF transformations because correspondent kernels φ_1, φ_3–φ_5 are well defined.

7 Numerical Results

The numerical realization of the NR method (3.3)–(3.5) requires the replacement of the unconstrained minimizer x^{s+1} by an approximation $\bar{x}^{s+1}$. We used the overall stopping criteria (5.16) with $\varepsilon = 10^{-10}$. To find an approximation $\bar{x}^{s+1}$ we applied damped Newton's method with Armijo steplength rule for minimizing $\mathcal{L}(x, \lambda^s, \boldsymbol{k}^s)$ in x, using the stopping criteria (5.12).

For the NLP calculations we used the truncated MBF transformation $\psi_2(t)$ given by (3.12) and $\lambda^0 = e = (1, \ldots, 1) \in \mathbb{R}^q_{++}$, $k = 10^2$, $\boldsymbol{k}^0 = 10^2 e$ as starting Lagrange multipliers and scaling parameters vectors. For LP calculations we used truncated log-sigmoid transformation $\psi_4(t)$ given by (3.12). Also for LP calculations we used LIPSOL solver (see [32]) to find the first primal-dual approximation with one digit of accuracy for the duality gap, i.e. we find the interior primal-dual approximation, for which $\sum_{i=1}^q \lambda_i c_i(x) \leq 0.1$. We show in the tables below the duality gap and the infeasibility after each Lagrange multipliers and scaling parameters update. We also show the number of Newton steps that is required for each update as well as the total number of Newton steps required to obtain the duality gap and infeasibility with at least ten digits of accuracy.

The numerical results obtained using the Newton NR method allowed us to observe systematically the "hot start" phenomenon (see [15], [19], [22]). Practically speaking, the "hot start" means that from some point on the primal approximation will remain in the Newton area after each Lagrange multiplier and scaling parameters update. Therefore, from this point on only few (often one) Newton steps require to find $\bar{x}^{s+1}$ and to reduce the duality gap and primal infeasibility by a factor $0 < \gamma = ck^{-1} < 1$ (see Proposition 5.2). In our numerical experiments we took $\gamma = 0.5$.

In the tables below we show numerical results for the some NLP problems from R. Vanderbei webpage (http://www.sor.princeton.edu/ rvdb/ampl/nlmodels/index.html) and LP problems from Netlib library. The number of variables is n and q is the number of constraints.

Name: $esfl_{socp}$; Objective: linear; Constraints: convex quadratic.

$n = 1002$, $q = 1000$

it	gap	inf	# of steps
1	2.20e-01	8.76e-04	18
2	2.41e-03	5.17e-06	2
3	2.78e-04	2.50e-06	2
4	1.51e-05	6.24e-07	1
5	2.05e-08	3.90e-08	1
6	2.22e-11	1.52e-10	1
Total number of Newton steps			25

Name: moonshot; Objective: linear; Constraints: nonconvex quadratic.

$n = 786$, $q = 592$

it	gap	inf	# of steps
1	3.21e+00	4.98e-05	10
2	8.50e-06	5.44e-10	1
3	1.19e-08	1.24e-11	1
4	3.96e-09	4.16e-12	1
Total number of Newton steps			13

Name: markowitz2; Objective: convex quadratic; Constraints: linear .

$n = 1200$, $q = 1201$,

it	gap	inf	# of steps
0	7.032438e+01	1.495739e+00	0
1	9.001130e-02	5.904234e-05	10
2	4.205607e-03	3.767383e-06	12
3	6.292277e-05	2.654451e-05	13
4	1.709659e-06	1.310097e-05	8
5	1.074959e-07	1.381697e-06	5
6	7.174959e-09	3.368086e-07	4
7	4.104959e-10	3.958086e-08	3
8	1.749759e-11	2.868086e-09	2
9	4.493538e-13	1.338086e-10	2
Total number of Newton steps			59

Name: brandy.

$n = 149$, $q = 259$

it	gap	inf	# of steps
0	6.56e+04	3.57e+03	0
1	1.68e-01	2.04e-05	28
2	2.66e-01	3.58e-04	15
3	3.33e-03	2.73e-05	7
4	1.74e-04	6.68e-07	4
5	2.83e-08	3.68e-10	3
6	2.49e-14	1.33e-15	2
Total number of Newton steps			59

Name: Israel.

$n = 174$, $q = 316$

it	gap	primal inf	# of steps
0	1.05e+10	1.21e+06	0
1	6.40e+00	2.77e-09	20
2	7.364041e-02	1.0729e-07	14
3	7.497715e-07	4.2011e-12	6
4	1.628188e-10	1.7764e-15	3
Total number of Newton steps			43

Name: AGG2.

$n = 516,\ q = 758$

it	gap	primal inf	# of steps
0	6.93e+10	7.41e+06	0
1	6.07e+00	4.39e-07	16
2	1.422620e-03	2.2625e-09	3
3	2.630272e-10	7.1054e-15	3
Total number of Newton steps			25

Name: BNL1.

$n = 644,\ p = 1175$

it	gap	primal inf	# of steps
0	3.98e+06	1.83e+04	0
1	9.47e-05	7.13e-09	25
2	2.645905e-07	3.8801e-10	5
3	2.025197e-12	4.5938e-13	4
Total number of Newton steps			34

8 Concluding Remarks

The NR methods provide an exchange of information between the primal and dual variables. However, the calculations are always conducted sequentially: first is the primal minimization, then the Lagrange multiplier and the scaling parameter updates. On the other hand, it has become evident lately that the most efficient methods that are based on the interior point path-following ideas, are the Primal-Dual methods, for which calculations are conducted simultaneously in the primal and dual spaces (see [29], [31]).

For each NR multiplier method, there exists the Primal-Dual equivalent (see [20]). Our experiments with the Primal-Dual NR methods are very encouraging [22].

It seems that NR methods with "dynamic" scaling parameter update are particularly suitable for the Primal-Dual approach. It follows from Theorem 5.1 and Proposition 5.2 that under strict complementarity or standard second-order optimality conditions, the Lagrange multipliers, which correspond to the passive constraints, converge to zero at least quadratically. Therefore, the Primal-Dual NR method asymptotically turns into the Newton method for the Lagrange system of equations, which corresponds to the set of active constraints.

On the other hand, it follows from Theorem 9 (see [18], p. 247) that under the standard second-order optimality condition for sufficiently smooth functions, the Newton method for the Lagrange system of equations converges to the primal-dual solution with quadratic rate.

Therefore, it seems the Primal-Dual NR methods with "dynamic" scaling parameter updates have the potential to be globally convergent with asymptotic quadratic rate. We will cover the corresponding theory and methods in an upcoming paper.

Acknowledgment

This research was conducted during my Fulbright Professorship at the Technion. I am grateful to the Council for International Exchange of Scholars at the US Department of State and to the United States-Israel Educational Foundation for giving me such an opportunity.

I am thankful to Prof. W. Hager and anonymous referees for their valuable suggestions, which helped me in improving the manuscript.

Finally, it gives me great pleasure to thank Ms. Lillian Bluestein for her excellent job in preparing this manuscript for publication.

References

1. A. Auslender, R. Cominetti and M. Haddou. " Asymptotic analysis of penalty and barrier methods in convex and linear programming." Math. Oper. Res., vol. 22, pp. 43–62, 1997.

2. A. Auslender, M. Teboulle and S. Ben-Tiba. "Interior proximal and multipliers methods based on second-order homogeneous kernels." Math. Oper. Res., vol. 24, no. 3, pp. 645–668, 1999.

3. M.P. Bendsøe, A. Ben-Tal and J. Zowe. "Optimization methods for truss geometry and topology design." Structural Optimization, vol. 7, pp. 141–159, 1994.

4. A. Ben-Tal and M. Zibulevsky. "Penalty-barrier methods for convex programming problems." SIAM J. Optim., vol. 7, pp. 347–366, 1997.

5. A. Ben-Tal, B. Yuzefovich and M. Zibulevsky. "Penalty-barrier multipliers methods for minimax and constrained smooth convex optimization." Optimization Laboratory, Technion, Israel. Research Report 9-92. 1992.

6. M. Breitfelt and D. Shanno. "Experience with modified log-barrier method for nonlinear programming." Ann. Oper. Res., vol. 62, pp. 439–464, 1996.

7. D. Bertsekas. Constrained Optimization and Lagrange Multipliers Methods. Academic Press: New York. 1982.

8. A. Fiacco and G. McCormick. Nonlinear Programming: Sequential Unconstrained Minimization Techniques. Classics in Applied Mathematics. SIAM: Philadelphia, PA. 1990.

9. O. Guler. "On the convergence of the proximal point algorithm for convex minimization." SIAM J. Control Optim., vol. 29, pp. 403–419, 1991.

10. M.R. Hestenes. "Multipliers and gradient methods." J. Optim. Theory Appl., vol. 4, pp. 303–320, 1969.

11. A. Hoffman. “On approximate solution of system of linear inequalities.” Journal of Research of the National Bureau of Standards, vol. 49, pp. 263–265, 1952.

12. K. Knopp. Infinite Sequence and Series. Dover Publication Inc.: New York. 1956.

13. B.W. Kort and D.P. Bertsekas. “Multiplier methods for convex programming,” in Proc. IEEE Conf. on Decision and Control. San Diego, CA., 1973, pp. 428–432.

14. B. Martinet. “Regularization d’inequations variationelles par approximations successive.” Revue Francaise d’Automatique et Informatique Rechershe Operationelle, vol. 4, pp. 154–159, 1970.

15. A. Melman and R. Polyak. “The Newton modified barrier method for QP problems.” Ann. Oper. Res., vol. 62, pp. 465–519, 1996.

16. J. Moreau. “Proximité et dualité dans un espace Hilbertien.” Bull. Soc. Math. France, vol. 93, pp. 273–299, 1965.

17. S. Nash, R. Polyak and A. Sofer. “A numerical comparison of barrier and modified barrier method for large-scale bound-constrained optimization,” in Large Scale Optimization, State of the Art. W. Hager, D. Hearn, P. Pardalos (Eds.). Kluwer Academic Publishers, pp. 319–338, 1994.

18. B. Polyak, Introduction to Optimization. Software Inc.: NY. 1987.

19. R. Polyak. “Modified barrier functions (theory and methods).” Math. Programming, vol. 54, pp. 177–222, 1992.

20. R. Polyak. “Log-sigmoid multipliers method in constrained optimization.” Ann. Oper. Res., vol. 101, pp. 427–460, 2001.

21. R. Polyak. “Nonlinear rescaling vs. smoothing technique in convex optimization.” Math. Programming, vol. 92, pp. 197-235, 2002 .

22. R. Polyak and I. Griva. “Primal-dual nonlinear rescaling methods for convex optimization,” J. Optim. Theory Appl., vol 122, pp 111–156, 2004.

23. R. Polyak and M. Teboulle. “Nonlinear rescaling and proximal-like methods in convex optimization.” Math. Programming, vol. 76, pp. 265–284, 1997.

24. R. Polyak, I. Griva and J. Sobieski. The Newton Log-Sigmoid Method in Constrained Optimization. A Collection of Technical Papers. 7th AIAA/USAF/NASA/ISSMO Symposium on Multidisciplinary Analysis and Optimization, vol. 3, pp. 2193–2201, 1998.

25. M.J.D. Powell. “A method for nonlinear constraints in minimization problems,” in Fletcher, ed., Optimization, London Academic Press, pp. 283–298, 1969.

26. R.T. Rockafellar. “A dual approach to solving nonlinear programming problems by unconstrainted minimization.” Math. Programming, vol. 5, pp. 354–373, 1973.

27. R.T. Rockafellar. “Monotone operators and the proximal point algorithm.” SIAM J. Control Optim., vol. 14, pp. 877–898, 1976.

28. R.T. Rockafellar. “Augmented Lagrangians and applications of the proximal points algorithms in convex programming.” Math. Oper. Res., vol. 1, pp. 97–116, 1976.

29. D.F. Shanno and R.J. Vanderbei. “An interior point algorithm for nonconvex nonlinear programming.” COAP, vol. 13, pp. 231–252, 1999.

30. P. Tseng and D. Bertsekas. “On the convergence of the exponential multipliers method for convex programming.” Math. Programming, vol. 60, pp. 1–19, 1993.

31. S. Wright. Primal-Dual Interior-Point Methods, SIAM. 1997.

32. Y. Zhang, “Solving Large-Scale Linear Programs by Interior-Point Methods Under MATLAB Environment.” Dept. of Computational and Applied Mathmatics, Rice University, Houston, TX 77005, 1996.