

Nonlinear rescaling and proximal-like methods in convex optimization

Roman Polyak^{a,1} Marc Teboulle^{b,*}

^a *Department of Operations Research and Statistics, George Mason University, Fairfax, VA, USA*

^b *School of Mathematical Sciences, Tel-Aviv University, Ramat-Aviv 69978, Israel*

Received 21 March 1994; revised manuscript received 25 October 1995

Abstract

The nonlinear rescaling principle (NRP) consists of transforming the objective function and/or the constraints of a given constrained optimization problem into another problem which is equivalent to the original one in the sense that their optimal set of solutions coincides. A nonlinear transformation parameterized by a positive scalar parameter and based on a smooth scaling function is used to transform the constraints. The methods based on NRP consist of sequential unconstrained minimization of the classical Lagrangian for the equivalent problem, followed by an explicit formula updating the Lagrange multipliers. We first show that the NRP leads naturally to proximal methods with an entropy-like kernel, which is defined by the conjugate of the scaling function, and establish that the two methods are dually equivalent for convex constrained minimization problems. We then study the convergence properties of the nonlinear rescaling algorithm and the corresponding entropy-like proximal methods for convex constrained optimization problems. Special cases of the nonlinear rescaling algorithm are presented. In particular a new class of exponential penalty-modified barrier functions methods is introduced.

Keywords: Convex optimization; Nonlinear rescaling; Modified barrier functions; Augmented Lagrangians; Proximal methods

1. Introduction

One of the most important applications of the proximal algorithm [12,21], is when applied to the dual of a convex programming problem. Rockafellar [20] shows that

* Corresponding author. E-mail: teboulle@math.tau.ac.il. Partially supported by the National Science Foundation, under Grants DMS-9201297, and DMS-9401871. Currently on leave from Department of Mathematics and Statistics University of Maryland, Baltimore County, Baltimore, USA.

¹ E-mail: rpolyak@gmuvax.gmu.edu. Partially supported by NASA Grant NAG3-1397 and NSF Grant DMS-9403218.

applying a proximal algorithm to the dual of a convex program is equivalent to the quadratic augmented Lagrangian method (also called multiplier methods). Augmented Lagrangian methods have many advantages over penalty methods, see for example the monograph of Bertsekas [1].

Recently, *nonquadratic* augmented Lagrangian methods have received much attention, see e.g. [6,9,16,17,22,24]. It has been already recognized in [1, Ch. 5] that the use of penalty terms other than quadratic in an augmented Lagrangian functional, can affect the rate of convergence of the associated iterative methods and consequently their performance. Moreover, while the classical quadratic augmented Lagrangian is differentiable only once (even if the problem's data possess higher differentiability), nonquadratic augmented Lagrangians are often C^2 if the objective /constraints are also twice continuously differentiable. This is an important advantage since Newton type methods can then be applied.

Nonquadratic multiplier methods can be derived from many different approaches. This paper will concentrate on the interplay between two of these approaches: the nonlinear rescaling principle and proximal-like mapping. Given a constrained convex optimization problem, the nonlinear rescaling principle [15] consists of transforming the objective and/or the constraints into another equivalent optimization problem, namely one which has the same set of minimizers. The nonlinear transformation is parameterized by a positive parameter and based on a smooth function, called the scaling function. The basic steps of the nonlinear rescaling (NR) algorithm emerging from this are then based on a primal-dual scheme, namely to minimize the classical Lagrangian associated with the equivalent problem, followed by an explicit update formula for the dual variables. The NR algorithm leads to various classes of nonquadratic multiplier methods which include as special cases several known multiplier methods such as for example: the exponential multiplier method, the class \hat{P}_l of multiplier methods introduced by Bertsekas [1], and the Modified-Barrier Functions (MBF) and Modified Interior Distance Functions introduced by Polyak [15,16].

Recently, Teboulle [22] introduced a new class of proximal-like mappings, where the usual quadratic term in the Moreau [13] proximal approximation of a convex function is replaced by an entropy-like distance, defined in terms of a convex function. It was shown in [22] that applying a proximal-like map to the dual of the given convex optimization problem provides a unified framework for constructing nonquadratic augmented Lagrangians, and allows for recovering the class of multiplier methods proposed by Bertsekas [1], and the modified barrier functions methods of Polyak [16]. The two approaches, namely the nonlinear rescaling principle and the proximal-like framework are seemingly completely different, but in fact they are completely equivalent. To show this will be the first contribution of this paper. More precisely it is shown that the nonlinear rescaling principle leads naturally to one of the class of entropy-like distance functions introduced in [22] with kernel given in terms of the conjugate of the scaling function. This will be developed in Section 3, after introducing the NR method in Section 2.

The other contribution of this paper is to study the convergence properties of the NR algorithm and corresponding entropy-like proximal algorithms. Under very mild assumptions on the given constrained convex optimization problem, we will prove that for any positive barrier parameter, and for a wide class of scaling functions, the dual sequence generated by the NR method globally converges to an optimal dual solution, while the corresponding primal sequence approaches optimality in an ergodic sense. These results extend and complement some of the convergence results recently obtained for various classes of nonquadratic augmented Lagrangian methods discussed below. In Section 5 we give some examples and in Section 6 we introduce a new class of multipliers method based on a scaling function which combines an exponential penalty with modified barrier functions, and for which our convergence results can be applied. Finally, Section 7 draws some brief conclusions and outlines some directions for future research.

To put the results obtained in this paper in the perspective of recent achievements, we conclude this section by briefly surveying the convergence results currently available for special realizations of the NR algorithm. In a recent paper [24], Tseng and Bertsekas proved the convergence of the dual sequence to an optimal dual solution and showed that the primal sequence approaches optimality in an ergodic sense for the exponential multiplier method (see Section 5.1). Convergence results for the class of logarithmic/hyperbolic modified barrier function methods (see Section 5.2) for linear and nonlinear programming including nonconvex constrained optimization, have been developed by Polyak in [15,16], under nondegeneracy assumptions. It was proven that if the primal and dual problems have unique solutions, then the MBF method converges with linear rate and the ratio can be made as small as desired by choosing a fixed, but large enough barrier parameter. In a more recent work, Jensen and Polyak [11] proved that the primal and dual sequences generated by the MBF method converge in values for both linear and convex programming problems for any positive barrier parameter, when the primal and dual feasible sets are bounded. It was also shown in [11] that the primal MBF sequence converges to an optimal solution in an ergodic sense. The recent work of Iusem, Svaiter and Teboulle [9] gives convergence results for a general class of multiplier methods generated by a kernel which is required to satisfy a certain inequality on the derivative of the kernel (see Section 4). Further convergence and rate of convergence results for the case of linear programming have been obtained by Tseng-Bertsekas [24] for the exponential multiplier method and by Iusem-Teboulle [10] for a more general class of multiplier methods which includes the examples of Section 5. In all of these works, the convergence of the whole primal sequence to an optimal solution remains an open question. Only very recently, Powell [17] has proved that for linear programs, the primal sequence produced by the logarithmic MBF method converges under the assumption that the primal feasible set is bounded. In particular, he also proved that the primal logarithmic MBF sequence converges to the Chebyshev center on the optimal face with linear R-rate. Convergence results of the primal sequence for a large family of methods (which includes the logarithmic MBF) in the linear programming case has been proved in [10], under different assumptions to those used in [17].

the NR algorithm generates a sequence of feasible points for the dual problem (D). On the other hand, the primal sequence $\{x^s\}$ produced by (2.2) need not be a feasible one. Therefore, one of our main tasks in the convergence analysis of the NR method is to make sure that either the sequence $\{x^s\}$ or a byproduct of it will converge toward a primal feasible solution. One important step in that direction is developed in the following section where we show that in fact, the NR algorithm leads to a special type of proximal method applied to the dual problem of (P).

3. Nonlinear rescaling and proximal-like maps

Recently, it has been shown by Teboulle [22] that when applying an entropy-like proximal regularization to the dual functional, one obtains the class of NR algorithms. In this section we show that the nonlinear rescaling principle gives rise naturally to this class of entropy-like proximal approximation introduced in [22], and thus demonstrate that while the two approaches appear to be seemingly completely different, they can be seen essentially as equivalent.

Starting with the NR method, by writing the optimality conditions for (2.2) we obtain

$$\nabla f(x^{s+1}) - \sum_{i=1}^m u_i^s \psi'(\mu g_i(x^{s+1})) \nabla g_i(x^{s+1}) = 0.$$

Substituting in the above the dual update given in (2.3) we then have

$$\nabla f(x^{s+1}) - \sum_{i=1}^m u_i^{s+1} \nabla g_i(x^{s+1}) = 0,$$

showing that x^{s+1} is the minimizer of $L(x, u^{s+1})$. On the other hand, from the definition of the dual functional, we know that

$$(-g_1(x^{s+1}), \dots, -g_m(x^{s+1}))^T \in \partial h(u^{s+1}), \tag{3.1}$$

where $\partial h(u^{s+1})$ denotes the subdifferential of h at u^{s+1} . By (A3), the inverse function $(\psi')^{-1}$ exists. Once again, using (2.3) we obtain

$$g_i(x^{s+1}) = \mu^{-1} (\psi')^{-1}(u_i^{s+1}/u_i^s) = \mu^{-1} (\psi^*)'(u_i^{s+1}/u_i^s) \quad i = 1, \dots, m,$$

where the second equality follows from the well known relation $(\psi')^{-1} = (\psi^*)'$ and where $\psi^*(s) = \inf_t \{st - \psi(t)\}$ is the concave conjugate of ψ , see e.g., [19]. Therefore, the above inclusion can be rewritten as

$$0 \in \partial h(u^{s+1}) + \mu^{-1} \left((\psi^*)' \left(\frac{u_1^{s+1}}{u_1^s} \right), \dots, (\psi^*)' \left(\frac{u_m^{s+1}}{u_m^s} \right) \right)^T.$$

Given $\psi \in \Psi$, we now introduce the function φ as

$$\varphi(t) = -\psi^*(t). \tag{3.2}$$

Then the last inclusion can be written as

$$0 \in \partial h(u^{s+1}) - \mu^{-1} \left(\varphi' \left(\frac{u_1^{s+1}}{u_1^s} \right), \dots, \varphi' \left(\frac{u_m^{s+1}}{u_m^s} \right) \right)^T. \tag{3.3}$$

The next result gives the properties of the function φ inherited from ψ .

Proposition 3.1. *Let $\psi \in \Psi$. Then the function φ is a strictly convex differentiable function on \mathbb{R}_{++} which satisfies*

$$\varphi(1) = \varphi'(1) = 0, \tag{3.4}$$

$$\lim_{t \rightarrow 0^+} \varphi'(t) = -\infty. \tag{3.5}$$

Proof. The function ψ^* is concave and hence $\varphi = -\psi^*$ is convex. Using [19, Section 24] we have

$$\text{ridom } \psi^* \subset \text{range } \psi' \subset \text{dom } \psi^*.$$

By (A2), the function ψ is monotone increasing, i.e., $\text{range } \psi' \subset \mathbb{R}_{++}$. Therefore we have $\text{ridom } \psi^* \subset \mathbb{R}_{++}$ and hence $\text{dom } \varphi = \text{dom } \psi^* \subset \mathbb{R}_{++}$. Since ψ is strictly concave on \mathbb{R} , it is essentially strictly concave; hence by [19, Theorem 26.3] its conjugate ψ^* is essentially smooth and hence by [19, Theorem 26.1] differentiable. Thus, $\varphi(t)$ is differentiable for $t > 0$. Now, to prove (3.4), we compute

$$\varphi(1) = -\psi^*(1) = -\inf(t - \psi(t)) = -\psi(0) = 0 \quad (\text{by (A1)}).$$

Also, using again the well known relation for conjugate functions:

$$(\psi^*)' = (\psi')^{-1} \tag{3.6}$$

we have from (3.2) $\varphi'(1) = -(\psi^*)'(1) = -(\psi')^{-1}(1) = 0$, by (A1). Finally, to prove (3.5) using (3.6) and (A4) we obtain

$$\lim_{t \rightarrow 0} \varphi'(t) = -\lim_{t \rightarrow 0} (\psi^*)'(t) = -\lim_{t \rightarrow 0} (\psi')^{-1}(t) = -\infty. \quad \square$$

The class of functions satisfying Proposition 3.1 will be denoted by Φ . Given $\varphi \in \Phi$, we define $d_\varphi : \mathbb{R}_+^n \times \mathbb{R}_+^n \rightarrow \mathbb{R}$ as

$$d_\varphi(x, y) = \sum_{j=1}^n y_j \varphi \left(\frac{x_j}{y_j} \right). \tag{3.7}$$

The function d_φ is called the φ -divergence and is used as a kind of “distance” between two vectors in \mathbb{R}_+^n (see [22] for further details). This is justified in the following result.

Lemma 3.2. Let $\varphi \in \Phi$. Then,

- (a) $d_\varphi(x, y) \geq 0$ and equality holds iff $x = y$.
 (b) $d_\varphi(x, y)$ is jointly convex in (x, y) .

Proof. Property (a) follows easily from the gradient inequality for φ . To prove (b) it is enough to show that $\zeta(s, t) := t\varphi(s/t)$ is jointly convex in (s, t) . By the convexity of φ , $\forall (s_1, t_1), (s_2, t_2) \in \mathbb{R}_{++}^2$, and $\lambda \in (0, 1)$,

$$\begin{aligned} \varphi\left(\frac{\lambda s_1 + (1-\lambda)s_2}{\lambda t_1 + (1-\lambda)t_2}\right) &= \varphi\left(\frac{\lambda t_1}{\lambda t_1 + (1-\lambda)t_2} \frac{s_1}{t_1} + \frac{(1-\lambda)t_2}{\lambda t_1 + (1-\lambda)t_2} \frac{s_2}{t_2}\right) \\ &\leq \frac{\lambda}{\lambda t_1 + (1-\lambda)t_2} t_1 \varphi(s_1/t_1) \\ &\quad + \frac{(1-\lambda)}{\lambda t_1 + (1-\lambda)t_2} t_2 \varphi(s_2/t_2), \end{aligned}$$

which can be rewritten as

$$\zeta(\lambda s_1 + (1-\lambda)s_2, \lambda t_1 + (1-\lambda)t_2) \leq \lambda \zeta(s_1, t_1) + (1-\lambda) \zeta(s_2, t_2),$$

proving the joint convexity of ζ . \square

Going back to the inclusion (3.3), using (3.7), it follows that (3.3) is nothing else but the optimality condition to solve:

$$u^{s+1} = \operatorname{argmax}_{u \geq 0} \{h(u) - \mu^{-1} d_\varphi(u, u^s)\}. \quad (3.8)$$

The latter is precisely the entropy-like proximal algorithm recently proposed in [9] when applied to the dual problem (D) . We have thus shown that the nonlinear rescaling algorithm leads naturally to the class of proximal-like regularization, with distance kernel involving the negative of the conjugate of the nonlinear scaling function ψ . This in turns gives rise to the entropy-like algorithm applied to the dual of (P) , and therefore the two methods are dually equivalent. Using the equivalence between the NR algorithm and the entropy-like proximal method, we will use some convergence results from the latter to prove the convergence of the dual sequence to an optimal dual solution u^* , and an ergodic convergence result for the corresponding primal sequence.

4. Convergence analysis

In this section, we develop convergence results for the NR algorithm with a general nonlinear transformation $\psi \in \Psi$ which will allow us to obtain convergence results for a wide class of multiplier methods. The key steps in the convergence proof of the primal-dual sequence $\{x^s, u^s\}$ generated by the NR algorithm consist of establishing the following properties for the sequences $\{x^s, u^s\}$:

- (i) Boundedness of the dual sequence $\{u^s\}$.
- (ii) Asymptotic complementary slackness.
- (iii) Convergence of the sequence $\{u^s\}$ to an optimal solution.
- (iv) Asymptotic feasibility of the primal sequence $\{x^s\}$.

As we shall see later, the most delicate parts are to establish properties (iii) and (iv). Unless stated otherwise, in this section we assume that ψ satisfies (A1)–(A5) and that assumptions (P1) and (P2) hold.

Proposition 4.1. *The sequence of dual objective function values $\{h(u^s)\}$ is monotone nondecreasing and convergent.*

Proof. From (3.8), we have shown that the sequence u^s is equivalently given by

$$u^{s+1} = \operatorname{argmax}_{u \geq 0} \{h(u) - \mu^{-1}d_\varphi(u, u^s)\}.$$

Therefore,

$$h(u^{s+1}) \geq h(u) - \mu^{-1}d_\varphi(u, u^s), \quad \forall u \geq 0.$$

In particular, with $u = u^s$ in the above inequality, and using the fact that $d_\varphi(u^s, u^s) = 0$ (see Lemma 3.2(a)), we obtain that $h(u^{s+1}) \geq h(u^s)$, proving that $\{h(u^s)\}$ is monotone nondecreasing. Moreover, by the weak duality theorem, the sequence $\{h(u^s)\}$ is bounded above by the optimal primal value f^* , and hence $\{h(u^s)\}$ is convergent. \square

Proposition 4.2. *The dual sequence $\{u^s\}$ generated by the NR algorithm is bounded.*

Proof. By assumption (P2) the set of optimal Lagrange multipliers is nonempty and compact, i.e., one level set of h is compact. Since h is a closed proper concave function, all its level sets are compact, and in particular $L := \{u : h(u) \geq h(u^1)\}$. Since by Proposition 4.1 $u^s \in L, \forall s$, it follows that u^s is bounded. \square

Proposition 4.3. *Let $\{x^s, u^s\}$ be the sequences generated by the NR algorithm. Then,*

$$\lim_{s \rightarrow \infty} u_i^s g_i(x^s) = 0, \quad i = 1, \dots, m. \tag{4.1}$$

Proof. Using the concavity of h and $-g(x^{s+1}) \in \partial h(u^{s+1})$ (cf. (3.1)), we obtain

$$\sum_{i=1}^m (u_i^s - u_i^{s+1}) g_i(x^{s+1}) \leq h(u^{s+1}) - h(u^s). \tag{4.2}$$

Using the dual update formula given in (2.3), we have for each $i = 1, \dots, m$:

$$\begin{aligned} u_i^s - u_i^{s+1} &= u_i^s - \psi'(\mu g_i(x^{s+1})) u_i^s = u_i^s (1 - \psi'(\mu g_i(x^{s+1}))) \\ &= u_i^{s+1} r(\mu g_i(x^{s+1})), \end{aligned}$$

where in the last equality we use once more the dual update formula (2.3), and we have defined $r(t) := 1/\psi'(t) - 1$. Substituting this in (4.2) we then obtain:

$$\sum_{i=1}^m u_i^{s+1} g_i(x^{s+1}) r(\mu g_i(x^{s+1})) \leq h(u^{s+1}) - h(u^s),$$

which gives by summation,

$$\sum_{s=1}^{\infty} \sum_{i=1}^m u_i^{s+1} g_i(x^{s+1}) r(\mu g_i(x^{s+1})) \leq \lim_{s \rightarrow \infty} h(u^s) - h(u^1) \leq f^* - h(u^1) < \infty,$$

where the second inequality follows from Proposition 4.1. Therefore,

$$\sum_{i=1}^m u_i^{s+1} g_i(x^{s+1}) r(\mu g_i(x^{s+1})) \rightarrow 0. \tag{4.3}$$

Now, since ψ is concave with $\psi(0) = 0, \psi'(0) = 1$ and $\psi'(t) > 0$ we have $tr(t) \geq 0, \forall t$. Hence, since $u^s > 0, \forall s$, and $\mu > 0$, each term in the summation of the sequence (4.3) is nonnegative and we obtain that

$$\lim_{s \rightarrow \infty} u_i^{s+1} g_i(x^{s+1}) r(\mu g_i(x^{s+1})) = 0, \quad \forall i = 1, \dots, m. \tag{4.4}$$

To complete the proof, we argue by contradiction. Suppose that for any fixed i , there exists a subsequence $S \subset \{1, 2, \dots\}$ and an $\epsilon > 0$ such that

$$|u_i^{s+1} g_i(x^{s+1})| \geq \epsilon > 0, \quad \forall s \in S. \tag{4.5}$$

Then from (4.4) $\{r(\mu g_i(x^{s+1}))\}_S \rightarrow 0$. Since $r(t) = 1/\psi'(t) - 1, \psi'(0) = 1$ and $\mu > 0$, it follows that $\{g_i(x^{s+1})\}_S \rightarrow 0$, and hence from (4.5) that $\{u_i^{s+1}\}_S \rightarrow \infty$, which contradicts the boundedness of $\{u^s\}$ proved in Proposition 4.2. Therefore, $\lim_{s \rightarrow \infty} u_i^{s+1} g_i(x^{s+1}) = 0, i = 1, \dots, m. \quad \square$

We will now prove that the dual sequence converges to an optimal dual solution. To establish this result, it will be convenient to recall the recent result proven in [9] for the entropy-like proximal minimization algorithm. Consider the convex minimization problem

$$(C) \quad \min\{c(u) : u \in \mathbb{R}_+^n\},$$

where $c : \mathbb{R}^n \rightarrow (-\infty, +\infty]$ is a proper closed convex function. Let U_* denote the set of optimal solutions of that problem, and assume that U_* is nonempty and that $\text{dom}c \cap \mathbb{R}_{++}^n \neq \emptyset$.

Given $\varphi \in \Phi$ and $u^0 > 0$, the entropy-like proximal method for solving the above problem generates a sequence u^s via the iteration:

$$u^{s+1} = \underset{u \geq 0}{\operatorname{argmin}}\{c(u) + \mu^{-1} d_\varphi(u, u^s)\}. \tag{4.6}$$

The next result was proved in [9, Theorem 4.2].

Theorem 4.4 ([9]). *Let $\varphi \in \Phi$. Assume that the set of optimal solution for problem (C) is bounded. Then, the sequence $\{u^s\}$ generated by (4.6) converges to a limit. Moreover, if φ satisfies $\varphi'(t) \leq \varphi''(1) \log t, \forall t > 0$, then u^s converges to an optimal solution of problem (C).*

We shall prove the convergence of the dual sequence $\{u^s\}$ to an optimal dual solution u^* under the following additional assumption on the kernel ψ :

(A6) The function ψ' is logarithmic convex.

Recall that a function $k : \mathbb{R} \rightarrow \mathbb{R}_{++}$ is logarithmic convex if and only if $\log k(t)$ is convex.

Proposition 4.5. *Let $\psi \in \Psi$ and assume that (P1) and (P2) hold. Then,*

- (i) *the dual sequence $\{u^s\}$ converges to a limit,*
- (ii) *if (A6) holds, then $\{u^s\}$ converges to an optimal dual solution u^* .*

Proof. From (3.8) we have

$$u^{s+1} = \operatorname{argmax}_{u \geq 0} \{h(u) - \mu^{-1} d_\varphi(u, u^s)\}.$$

Therefore, since by assumption (P2) the set of optimal Lagrange multipliers is nonempty and compact, applying the first part of Theorem 4.4 with $h = -c$, we obtain the convergence of the sequence $\{u^s\}$, proving (i). Now, since by (A6) the function $\log \psi'$ is convex, this implies that

$$\log \psi'(s) - \log \psi'(0) \geq s \frac{\psi''(0)}{\psi'(0)},$$

and since by (A1) $\psi'(0) = 1$, the above can be written as

$$\log \psi'(s) \geq s \psi''(0). \tag{4.7}$$

From (3.2) we have $\psi^*(s) = -\varphi(s)$. Therefore, $-\varphi' = (\psi^*)' = (\psi')^{-1}$, from which we obtain

$$\psi'(-\varphi'(s)) = s, \quad \forall s.$$

Differentiating the above identity we get $-\psi''(-\varphi'(s))\varphi''(s) = 1$. Substituting $s = 1$ and using $\varphi'(1) = 0$ in the latter, we obtain that

$$-\varphi''(1) = 1/\psi''(0). \tag{4.8}$$

Let $\psi'(s) = t$. Then, using (4.8) in (4.7) we obtain $\varphi'(t) \leq \varphi''(1) \log t$. Therefore, we can apply the second part of Theorem 4.4 to conclude that $\{u^s\}$ converges to an optimal dual solution u^* . \square

We will now establish that the primal sequence $\{x^s\}$ converges to the feasible set in an ergodic sense. Ergodic convergence is frequently used in the analysis of fixed points of nonexpansive mappings. The idea is to generate a sequence of (weighted) averages from the original sequence which in general possesses nicer properties than the original one; see e.g., [4] for results and references on ergodic convergence type results.

Given the sequence $\{x^s\}$ generated by the NR algorithm one defines the sequence of averages

$$\bar{x}^s = s^{-1} \sum_{k=1}^s x^k. \quad (4.9)$$

Before establishing the ergodic feasibility of the sequence $\{\bar{x}^s\}$, we will first prove a technical result.

Lemma 4.6. *Let $w(t) : \mathbb{R} \rightarrow \mathbb{R}_{++}$, be a monotone decreasing continuously differentiable logarithmic convex function, with $w(0) = 1$. Then, for any $t_i \in \mathbb{R}$, $i = 1, \dots, s$ and any $\gamma \in [0, 1)$, we have*

$$\prod_{i=1}^s w(t_i) \leq (1 - \gamma) \Rightarrow \sum_{i=1}^s t_i \geq -\gamma(w'(0))^{-1}.$$

Proof. The left inequality in the lemma is equivalent to

$$-\sum_{i=1}^m \log w(t_i) \geq -\log(1 - \gamma), \quad \forall t_i \in \mathbb{R}, \forall \gamma \in [0, 1).$$

Using $\log z \leq z - 1$, $z > 0$ in the above inequality implies that

$$-\sum_{i=1}^m \log w(t_i) \geq \gamma, \quad \forall \gamma \in [0, 1). \quad (4.10)$$

Now, since $w(t)$ is logarithmic convex, then $\log w(t)$ is convex. Applying the gradient inequality to the convex function $\log w(t)$ and using $w(0) = 1$ we obtain:

$$\log w(t) \geq tw'(0), \quad \forall t \in \mathbb{R}. \quad (4.11)$$

Combining (4.11) with (4.10) and recalling that $-w'(0) > 0$, the result is proved. \square

We are now ready to establish the ergodic feasibility of the sequence $\{\bar{x}^s\}$.

Proposition 4.7. *Let ψ satisfy (A1)–(A6). For the sequence \bar{x}^s defined by (4.9), we have*

$$\liminf_{s \rightarrow \infty} g_i(\bar{x}^s) \geq 0, \quad i = 1, \dots, m. \quad (4.12)$$

Proof. By Proposition 4.5(i), $\{u^s\}$ is a convergent sequence, say to the point \bar{u} . Define $I^0 = \{i : \bar{u}_i = 0\}$ and $I^+ = \{i : \bar{u}_i > 0\}$, so that $I^0 \cup I^+ = \{1, \dots, m\}$. We will first consider the set I^+ . Using the concavity of g_i we have

$$g_i(\bar{x}^s) = g_i\left(s^{-1} \sum_{k=1}^s x^k\right) \geq s^{-1} \sum_{k=1}^s g_i(x^k), \quad i = 1, \dots, m. \tag{4.13}$$

From Proposition 4.3 we obtain

$$\liminf_{s \rightarrow \infty} g_i(x^s) = 0, \quad \forall i \in I^+. \tag{4.14}$$

and therefore $\liminf_{s \rightarrow \infty} s^{-1} \sum_{k=1}^s g_i(x^k) \geq 0, i \in I^+$. Hence from (4.13) we obtain

$$\liminf_{s \rightarrow \infty} g_i(\bar{x}^s) \geq 0, \quad i \in I^+.$$

Next, we consider the set I^0 . Using the dual update formula (2.3) and the fact that $u_i^s \rightarrow \bar{u}_i = 0$, for $i \in I^0$, we obtain

$$u_i^s = u_i^0 \prod_{k=1}^s \psi'(\mu g_i(x^k)) \rightarrow 0, \quad i \in I^0.$$

Therefore, there exists $\{\gamma_i^s\}$ with $\lim \gamma_i^s = 1$ such that

$$\prod_{k=1}^s \psi'(\mu g_i(x^k)) \leq 1 - \gamma_i^s, \quad i \in I^0.$$

It is now easy to verify that all the assumptions of Lemma 4.6 are satisfied with the choice $w := \psi'$, and hence the latter inequality implies that

$$\sum_{k=1}^s g_i(x^k) \geq -\gamma_i^s (\mu \psi''(0))^{-1}, \quad i \in I^0$$

and using once again (4.13) we thus obtain

$$g_i(\bar{x}^s) \geq -\gamma_i^s (\mu s \psi''(0))^{-1} (> 0), \quad i \in I^0,$$

from which it follows that $\liminf_{s \rightarrow \infty} g_i(\bar{x}^s) \geq 0, i \in I^0$. \square

Remark 4.8. From the proof above, we see that the asymptotic ergodic feasibility is obtained without assumption (A6) for $i \in I^+$. However, in the boundary case (i.e., $u_i^s \rightarrow 0, i \in I^0$), assumption (A6) is essential to complete the proof. We have not been able to prove the result without it or without making additional assumptions on the problem’s data (f, g) . Fortunately, assumption (A6) is satisfied for many interesting choices of the kernel ψ , see the examples given in Sections 5 and 6.

We are now in position to establish the main result of this section.

Theorem 4.9. Let ψ satisfy (A1)–(A6) and assume that (P1) and (P2) hold. Let $\{x^s\}, \{u^s\}$ be the sequences generated by the NR algorithm. Then,

- (i) $\{u^s\}$ converges to an optimal dual solution u^* ,
- (ii) the average sequence $\{\bar{x}^s\}$ defined by (4.9) is bounded,
- (iii) every limit point of the sequence $\{\bar{x}^s\}$ converges to an optimal primal solution x^* ,
- (iv) $\lim_{s \rightarrow \infty} f(x^s) = \lim_{s \rightarrow \infty} h(u^s) = f(x^*)$.

Proof. The first statement of the theorem was proved in Proposition 4.5(ii). From Proposition 4.3 we have $u_i^s g_i(x^s) \rightarrow 0$, $i = 1, \dots, m$, and from the definition of the dual objective h together with the fact that x^s minimizes the Lagrangian $L(x, u^s)$, we have $h(u^s) = f(x^s) - \sum u_i^s g_i(x^s)$. Hence it follows that

$$\lim_{s \rightarrow \infty} f(x^s) = \lim_{s \rightarrow \infty} h(u^s). \quad (4.15)$$

Using (4.9) and the convexity of f , we obtain

$$f(\bar{x}^s) = f\left(s^{-1} \sum_{k=1}^s x^k\right) \leq s^{-1} \sum_{k=1}^s f(x^k). \quad (4.16)$$

By the weak duality theorem we have

$$h(u^s) \leq f(x^*) \quad \forall s. \quad (4.17)$$

Therefore using (4.17), (4.15), it follows that

$$\limsup_{s \rightarrow \infty} f(\bar{x}^s) \leq \lim_{s \rightarrow \infty} h(u^s) \leq f(x^*). \quad (4.18)$$

By Proposition 4.7 we also have

$$\liminf_{s \rightarrow \infty} g_i(\bar{x}^s) \geq 0, \quad i = 1, \dots, m. \quad (4.19)$$

Since by assumption (P1), the optimal set of (P) is nonempty and bounded, invoking [7, Corollary 20, p. 94] it follows from (4.18)–(4.19) that $\{\bar{x}^s\}$ lies in a compact set, which proves (ii). Let \bar{x} be any limit point of $\{\bar{x}^s\}$. By continuity of f and g_i (4.18), (4.19) thus implies that

$$f(\bar{x}) \leq \lim_{l \rightarrow \infty} h(u^s) \leq f(x^*), \quad g_i(\bar{x}) \geq 0, \quad i = 1, \dots, m. \quad (4.20)$$

Since \bar{x} is feasible, we also have $f(\bar{x}) \geq f(x^*)$, therefore it follows from (4.17)–(4.20) that $f(\bar{x}) = f(x^*)$, i.e., \bar{x} is primal optimal and $\lim_{s \rightarrow \infty} f(x^s) = \lim_{s \rightarrow \infty} h(u^s) = f(x^*)$, proving (iii) and (iv). \square

This result extends the recent convergence results of [24] proved for the exponential multiplier method, to a general class of multiplier methods. It also complements the recent convergence results derived in [9]. We note that the convergence of the whole sequence x^s itself to an optimal primal solution remains an open question, even for special realizations of the NR algorithm discussed in the next section.

5. Examples

The convergence result proved in the previous section can be applied to various realizations of the NR algorithm. This section gives some examples for the choice of the scaling function ψ satisfying (A1)–(A6), and the corresponding kernel $\varphi = -\psi^*$ which generates the corresponding entropy-like distance d_φ , and for which our convergence result can be applied. The specific realizations of the NR algorithm for each of these examples, giving rise to particular methods are also given.

5.1. The exponential multiplier method

Let $\psi(t) = 1 - e^{-t}$, $t \in \mathbb{R}$. Then, $\psi^*(s) = s - 1 - s \log s$, $s \geq 0$. The corresponding NR algorithm (2.2)–(2.3) is then

$$x^{s+1} \in \operatorname{argmin} \left\{ f(x) + \mu^{-1} \sum_{i=1}^m u_i^s e^{-\mu g_i(x)} - \sum_{i=1}^m u_i^s : x \in \mathbb{R}^n \right\},$$

$$u_i^{s+1} = u_i^s e^{-\mu g_i(x^{s+1})}, \quad i = 1, \dots, m,$$

which is the exponential multiplier method [1,24]. This method is then equivalent to the following entropy-like proximal method:

$$u^{s+1} = \operatorname{argmax}_{u>0} \left\{ h(u) - \mu^{-1} \sum_{i=1}^m u_i \log \frac{u_i}{u_i^s} - u_i + u_i^s \right\}.$$

5.2. Logarithmic modified barrier functions

Let us consider the shifted logarithmic barrier transformation $\psi(t) = \log(1 + t)$, $t > -1$. Then, $\psi^*(s) = \log s - s + 1$, $s > 0$. The NR algorithm in this case takes the form

$$x^{s+1} \in \operatorname{argmin} \left\{ f(x) - \mu^{-1} \sum_{i=1}^m u_i^s \log(1 + \mu g_i(x)) : x \in \mathbb{R}^n \right\},$$

$$u_i^{s+1} = \frac{u_i^s}{1 + \mu g_i(x^{s+1})}, \quad i = 1, \dots, m,$$

which is exactly the modified barrier function (Frish-type) method proposed by Polyak [16]. This method is in turn equivalent (cf. (3.7)) to the following proximal method:

$$u^{s+1} = \operatorname{argmax}_{u>0} \left\{ h(u) - \mu^{-1} \sum_{i=1}^m (u_i - u_i^s) + \mu^{-1} \sum_{i=1}^m u_i^s \log u_i \right\}.$$

From the above, it follows that the MBF in primal space is equivalent to the weighted classical barrier method in dual space with a *shift* in the dual objective function. In the case of linear programming a similar observation has been made in [11,23]. Note that due to the weighted classical barrier term $\mu^{-1} \sum_{i=1}^m u_i^s \log u_i$ (and since $\log t = -\infty$

for $t \leq 0$), one can see that the proximal method requires solving only an *unconstrained* optimization problem at every step. Therefore in contrast with the classical quadratic proximal method (and associated classical augmented Lagrangian methods), the nonnegativity of the Lagrange multipliers is taken care of automatically.

5.3. Hyperbolic modified barrier functions

Let $\psi(t) = t/(1+t)$, $t > -1$. Then, $\psi^*(s) = 2\sqrt{s} - s - 1$, $s \geq 0$. The NR algorithm in this case reduces to

$$x^{s+1} \in \operatorname{argmin} \left\{ f(x) - \sum_{i=1}^m u_i^s \frac{g_i(x)}{(1 + \mu g_i(x))} : x \in \mathbb{R}^n \right\},$$

$$u_i^{s+1} = \frac{u_i^s}{(1 + \mu g_i(x^{s+1}))^2}, \quad i = 1, \dots, m,$$

which is the other modified barrier (Carroll-type) function method given in [16]. The equivalent proximal method (3.7) in this case takes the form

$$u^{s+1} = \operatorname{argmax}_{u \geq 0} \left\{ h(u) - \mu^{-1} \sum_{i=1}^m (u_i - u_i^s) + 2\mu^{-1} \sum_{i=1}^m \sqrt{u_i^s u_i} \right\}.$$

6. An exponential-penalty modified barrier function

The original idea behind the nonlinear rescaling principle and the corresponding modified barrier function, was to allow for extending the feasible set of the constraints. While the MBF theoretically provides such an extension, from the computational point of view this might not be enough, because for large enough $\mu > 0$, the shifted logarithmic barrier for example, has almost the same behavior as the classical barrier. Another difficulty is the fact that since the minimization step in $F(x, u^s, \mu)$ is performed approximately, it is possible for x^s to leave the extended feasible set. One approach to overcome this difficulty is to switch to an *exterior* penalty formulation when infeasible iterates occur. The idea of combining barrier (logarithmic or hyperbolic) functions with exterior penalty functions was proposed in [5], where a linear function was used to handle infeasible points and in [8], where a quadratic function was suggested to penalize the infeasible iterates. This approach was used to develop a class of pure barrier-penalty algorithms (i.e., with no multiplier updates). We refer the reader to the references just cited and to [18] and references therein, for other possible choices of the penalty functions.

More recently in [2], this idea has been used in the context of multiplier methods, where a shifted logarithmic barrier function is combined with an exterior quadratic penalty to penalize infeasible points. This corresponds for example in the NR algorithm to the choice:

$$\psi(t) = \begin{cases} \log(t + 1) & \text{if } t \geq -\frac{1}{2}, \\ -2t^2 + \frac{1}{2} - \log 2 & \text{if } t \leq -\frac{1}{2}. \end{cases} \tag{6.1}$$

This approach appears to lead to computationally more efficient algorithms than the ones based only on modified barrier functions. See for example the recent numerical experiments reported by Ben-Tal, Yuzefovich and Zibulevsky [2], Breitfeld and Shanno [3], and Nash, Polyak and Sofer [14]. The conclusions drawn in these papers from extensive computational testing, is that the Penalty Modified Barrier Function (PMBF) exhibits better performance than the modified barrier function, and appears quite promising for solving large scale nonlinear constrained optimization problems. However, thus far there is no theoretical proof that such methods generate primal-dual sequences of points which are globally convergent to an optimal primal-dual solution, without further assumptions on the problem’s data.

Note that Propositions 4.1, 4.2, 4.3 and 4.5 hold for the NR algorithm with the kernel (6.1). However, the quadratic part of the penalty term in the example (6.1) violates assumption (A6), namely, the quadratic function is not logarithmic convex, and therefore Theorem 4.4 is not applicable. However, since as explained previously, the basic idea of constructing a PMBF is to penalize the infeasibility of the current iterate, one could think of choosing any other appropriate reasonable function which will penalize the constraints and behave almost like a quadratic near the cut-off parameter η where the branching between the penalty and the barrier occurs. This motivates us to introduce a new class of PMBF, where the penalty branch is taken as an *exponential function*, and which accordingly will be called EPMBF. More precisely we thus propose exponential penalty-modified barrier kernels of the form:

$$\psi(t) = \begin{cases} b(t) & \text{if } t \geq -\eta, \\ p(t) & \text{if } t \leq -\eta, \end{cases} \tag{6.2}$$

where $b(t)$ above is a modified barrier function in the class Ψ , while $p(t)$ is an “exponential” penalty type function defined by

$$p(t) = ae^{bt} + c, \tag{6.3}$$

which extends into the infeasible region, and $0 < \eta < 1$, is the matching parameter. To preserve the smoothness of the scaling function ψ we require that the following conditions be satisfied:

$$b(-\eta) = p(-\eta); \quad b'(-\eta) = p'(-\eta); \quad b''(-\eta) = p''(-\eta). \tag{6.4}$$

Two examples of this class can be obtained by using the shifted-log barrier and shifted hyperbolic barrier for $b(t)$. To illustrate this consider the following specific examples with $\eta = \frac{1}{2}$. Using (6.2), (6.3) and (6.4) we obtain:

$$(a) \quad \psi(t) = \begin{cases} \log(1 + t) & \text{if } t \geq -\frac{1}{2}, \\ -e^{-2t-1} + 1 - \log 2 & \text{if } t \leq -\frac{1}{2}, \end{cases}$$

with corresponding conjugate

$$\psi^*(s) = \begin{cases} \log s - s + 1 & \text{if } 0 < s \leq 2, \\ -\frac{1}{2}s \log s + (\frac{1}{2}s + 1) \log 2 - 1 & \text{if } s \geq 2. \end{cases}$$

Note that if we take the quadratic approximation of $e^{-2t-1} + 1 - \log 2$ at $t = -\frac{1}{2}$ we obtain

$$\hat{p}(t) = -2t^2 + \frac{1}{2} - \log 2,$$

namely exactly the quadratic penalty branch used in (6.1).

$$(b) \quad \psi(t) = \begin{cases} \frac{t}{(1+t)} & \text{if } t \geq -\frac{1}{2}, \\ -e^{-4t-2} & \text{if } t \leq -\frac{1}{2}, \end{cases}$$

with corresponding conjugate

$$\psi^*(s) = \begin{cases} 2\sqrt{s} - s - 1 & \text{if } 0 < s \leq 4, \\ -\frac{1}{4}s - \frac{1}{4}s \log(\frac{1}{4}s) & \text{if } s \geq 4. \end{cases}$$

The corresponding NR algorithm with example (a) leads to an exponential penalty-modified logarithmic barrier multiplier method with Lagrangian function (cf. Eq. (2.1)) obtained by choosing the function ψ given above in (a). The corresponding equivalent proximal method then takes the form:

$$u^{s+1} = \operatorname{argmax}_{u>0} \begin{cases} h(u) - \mu^{-1} \sum_{i=1}^m (u_i - u_i^s) + \mu^{-1} \sum_{i=1}^m u_i^s \log u_i & \text{if } 0 < u_i \leq 2u_i^s, \\ h(u) - (2\mu)^{-1} \sum_{i=1}^m \left\{ u_i \log \frac{u_i}{u_i^s} - (u_i \log 2 - (1 - \log 2)u_i^s) \right\} & \text{if } u_i \geq 2u_i^s. \end{cases}$$

Similarly, one can generate another exponential penalty modified barrier function method and its equivalent proximal method, by using the example (b).

Now, it is easy to verify that the EPMBF generated by the kernel (6.2) satisfies the assumptions (A1)–(A6). Therefore, we can apply Theorem 4.4 to conclude

Theorem 6.1. *The dual sequence generated by an EPMBF globally converges to an optimal dual solution, while every limit point of the corresponding primal sequence of averages converges to an optimal primal solution of (P).*

7. Concluding remarks

The equivalence between the NR algorithm and proximal methods with entropy-like kernels makes it possible to obtain an elegant framework and further insights in the development and convergence analysis of nonquadratic augmented Lagrangians.

In the dual space, the NR method is in fact an interior point algorithm where the entropy-like kernel plays the role of a weighted barrier, which automatically takes care of the nonnegativity constraints of the Lagrange multipliers, and thus eliminates the combinatorial nature of the dual problem. Moreover, in both primal and dual spaces we are dealing with smooth unconstrained optimization problems and therefore the whole arsenal of unconstrained smooth optimization techniques, and in particular Newton type methods, can be applied to solve such constrained optimization problems.

Several important aspects of the NR method remain to be studied and are left for future research. This includes both theoretical and computational work. The NR method as stated in this paper is not a practical one, but just a conceptual algorithm. In particular, the method, as stated, requires exact unconstrained minimization at each step of $F(\cdot, u^s, \mu)$. This is obviously numerically impossible, and therefore the effect of inexact unconstrained minimization on the method needs to be analyzed. Another important issue is to study the rate of convergence of the NR method under different assumptions on the problem's data.

References

- [1] D. Bertsekas, *Constrained Optimization and Lagrange Multiplier Methods* (Academic Press, NY, 1982).
- [2] A. Ben-Tal, I. Yuzefovich and M. Zibulevsky, Penalty/barrier multiplier methods for minimax and constrained smooth convex programs, Research Report 9-92, Optimization Laboratory, Technion, Israel.
- [3] M.G. Breitfeld and D.F. Shanno, Computational experience with modified log-barrier methods for nonlinear programming, Rutcor Research Report 17-93, Rutgers University, New Brunswick, NJ.
- [4] R.D. Bruck, Asymptotic behavior of nonexpansive mappings, *Contemporary Mathematics* 18 (1983) 1–47.
- [5] J.H. Cassis, and A. Schmit, On implementation of the extended interior penalty function, *International Journal for Numerical Methods in Engineering* 10 (1976) 3–23.
- [6] J. Eckstein, Nonlinear proximal point algorithms using Bregman functions, with applications to convex programming, *Mathematics of Operations Research* 18 (1993) 202–226.
- [7] A.V. Fiacco, and G.P. McCormick, *Nonlinear Programming. Sequential Unconstrained Minimization Techniques*, SIAM Classics in Applied Mathematics (SIAM, Philadelphia, PA, 1990).
- [8] R.T. Haftka and J.H. Starnes, Applications of a quadratic extended interior penalty function for structural optimization, *AIAA Journal* 14 (1976) 718–724.
- [9] A.N. Iusem, B. Svaiter, M. Teboulle, Entropy-like proximal methods in convex programming, *Mathematics of Operations Research* 19 (1994) 790–814.
- [10] A. Iusem and M. Teboulle, Convergence rate analysis of nonquadratic proximal methods for convex and linear programming, *Mathematics of Operations Research* 20 (1995) 657–677.
- [11] D.L. Jensen and R.A. Polyak, The convergence of a modified barrier method for convex programming, *IBM Journal of Research and Development* 38 (1994) 307–321.
- [12] B. Martinet, Perturbation des méthodes d'optimisation. Application, *R.A.I.R.O. Analyse numérique/Numerical Analysis* (1978).
- [13] J.J. Moreau, Proximité et dualité dans un espace Hilbertien, *Bulletin de la Société Mathématique de France* 93 (1965) 273–299.
- [14] S.G. Nash, R. Polyak and A. Sofer, A numerical comparison of barrier and modified-barrier methods for large scale bound constrained optimization, in: W.W. Hager, D.W. Hearn and P.M. Pardalos, eds., *Large Scale Optimization: State of the Art* (Kluwer, Dordrecht, The Netherlands, 1994).
- [15] R.A. Polyak, Controlled processes in extremal and equilibrium problems, VINITI, deposited manuscript, Moscow (1986) (in Russian)
- [16] R.A. Polyak, Modified barrier functions (theory and methods), *Mathematical Programming* 54 (1992) 177–222.

- [17] M.J.D. Powell, Some convergence properties of the shifted log barrier method for linear programming, Numerical Analysis Report, DAMTP 1992/NA7, University of Cambridge (1992).
- [18] B. Prasad, and R.T. Haftka, A cubic extended interior penalty function for structural optimization, *International Journal for Numerical Methods in Engineering* 14 (1979) 1107–1126.
- [19] R.T. Rockafellar, *Convex Analysis* (Princeton University Press, Princeton, NJ, 1970).
- [20] R.T. Rockafellar, A dual approach to solving nonlinear programming problems by unconstrained optimization, *Mathematical Programming* 5 (1973) 354–373.
- [21] R.T. Rockafellar, Monotone operators and the proximal point algorithm, *SIAM Journal of Control and Optimization* 14 (1976) 877–898.
- [22] M. Teboulle, Entropic proximal mappings with application to nonlinear programming, *Mathematics of Operations Research* 17 (1992) 670–690.
- [23] M.J. Todd, Scaling, shifting and weighting in interior-points methods, Technical Report NO. 1067, School of Operations Research and Industrial Engineering, Cornell University (1993).
- [24] P. Tseng, D. Bertsekas, On the convergence of the exponential multiplier method for convex programming, *Mathematical Programming* 60 (1993) 1–19.