Dual Convergence of the Proximal Point Method with Bregman Distances for Linear Programming

Cruz Neto, J. X. * Ferreira, O. P. † Iusem, A. N. ‡ Monteiro, R. D. C. §

March 15, 2003

Abstract

In this paper we consider the proximal point method with Bregman distance applied to linear programming problems, and study the dual sequence obtained from the optimal multipliers of the linear constraints of each subproblem. We establish the convergence of this dual sequence, as well as convergence rate results for the primal sequence, for a suitable family of Bregman distances. These results are obtained by studying first the limiting behavior of a certain perturbed dual path and then the behavior of the dual and primal paths.

Keywords: generalized proximal point methods, barrier function, Bregman distances, dual path, primal path, convergence of dual sequence, dual convergence rate, primal convergence rate.

1 Introduction

The proximal point algorithm with Bregman distances for solving the linearly constrained problem

$$\min\{f(x) : Ax = b,\, x \geq 0\}, \tag{1}$$

where $f : \mathbb{R}^n \to \mathbb{R}$ is a differentiable convex function, A is an $m \times n$ real matrix, b is a real m-vector and the variable x is a real n-vector, generates a sequence $\{x^k\}$ according to the iteration

$$x^{k+1} \equiv \operatorname{argmin}\{f(x) + \lambda_k D_\varphi(x, x^k) : Ax = b\}, \tag{2}$$

*DM, Universidade Federal do Piauí, Teresina, PI 64049-500, BR (Email: `jxavier@ufpi.br`). This author was supported in part by CAPES and PRONEX-Otimização(FAPERJ/CNPq).

†IME, Universidade Federal de Goiás, Goiânia, GO 74001-970, BR (Email: `orizon@mat.ufg.br`). This author was supported in part by CAPES, FUNAPE/UFG and PRONEX-Otimização(FAPERJ/CNPq).

‡Instituto de Matemática Pura e Aplicada, Estrada Dona Castorina 110, Rio de Janeiro, RJ, 22460-320, Brazil (Email: `iusp@impa.br`). This author was supported in part by CNPq grant no. 301280/86 and PRONEX-Otimização(FAPERJ/CNPq).

§School of ISyE, Georgia Tech, Atlanta, Georgia 30332, USA (Email: `monteiro@isye.gatech.edu`). This author was supported in part by NSF Grants CCR-0203113, INT-9910084 and ONR grant N00014-03-1-0401.

where $x^0 > 0$ is arbitrary, $\{\lambda_k\}$ is a sequence of positive scalars satisfying $\sum_{k=0}^{\infty} \lambda_k^{-1} = +\infty$ and D_φ is a Bregman distance determined by a convex barrier φ for the nonnegative orthant $\mathbb{R}^n_+$ (see (8) for the definition of D_φ). The optimality condition for (2) determines the dual sequence $\{s^k\}$ defined as

$$s^k \equiv \lambda_k \left(\nabla\varphi(x^k) - \nabla\varphi(x^{k+1}) \right). \tag{3}$$

This method is a generalization of the classical proximal point method studied in Rockafellar (1976), which has the form of (2) for $\varphi(x) = \|x\|^2$ (note that this φ is not a barrier for the nonnegative orthant). Particular cases, corresponding to a special form of φ, were introduced in Eriksson (1985), Eggermont (1990) and Tseng and Bertsekas (1993). General Bregman distances were studied in several papers, e.g. Censor and Zenios (1992), Chen and Teboulle (1993), Iusem (1995) and Kiwiel (1997). Similar methods, using φ-divergences instead of Bregman distances in (2), appear in Iusem and Teboulle (1993), Jensen and Polyak (1994), Iusem and Teboulle (1995), Powell (1995) and Polyak and Teboulle (1997) (see Iusem, Svaiter and Teboulle (1994) for a definition of φ-divergence). These papers contain a complete study of the primal sequence $\{x^k\}$. However, the convergence of the whole sequence $\{s^k\}$ was lacking, even for linear programming. Our aim in this paper is to prove the convergence of this sequence for linear programming with Bregman distance D_φ, where φ satisfies an appropriate condition, which holds e.g. in the following cases:
1) $\varphi(x) = \sum_{j=1}^n x_j^\alpha - x_j^\beta$, with $\alpha \geq 1$ and $\beta \in (0,1)$,
2) $\varphi(x) = -\sum_{j=1}^n \log x_j$,
3) $\varphi(x) = \sum_{j=1}^n x_j^{-\alpha}$, with $\alpha > 0$.

Some authors, instead of studying the sequence $\{s^k\}$, have considered an averaged dual sequence $\{\bar{s}^k\}$ constructed from $\{s^k\}$. Partial results regarding the behavior of $\{\bar{s}^k\}$ have been obtained in a few papers, see Tseng and Bertsekas (1993), Powell (1995) Jensen and Polyak (1994) and Polyak and Teboulle (1997). Most of these results are described in a somewhat different framework, e.g. with φ-divergences instead of Bregman distances in (2). The convergence of the whole averaged dual sequence $\{\bar{s}^k\}$ for the proximal point method with Bregman distances has been obtained in Iusem and Monteiro (2000). In this paper it was showed that $\{\bar{s}^k\}$, under appropriate conditions including the examples above, converges to the centroid of the dual optimal set of problem (1). The case of the shifted logarithmic barrier was considered in Jensen and Polyak (1994), where is was proved that some cluster point of $\{\bar{s}^k\}$ is a dual optimal solution. This result was improved upon in Polyak and Teboulle (1997), where it is proved that all cluster points of $\{\bar{s}^k\}$ are dual optimal solutions. The convergence of the whole sequence $\{\bar{s}^k\}$ appeared for the first time in Powell (1995), but only for linear programming with the shifted logarithmic barrier. None of these papers present results on the convergence of the whole sequence $\{s^k\}$. For linear programming with a certain nondegeneracy assumption, which implies uniqueness of the dual solution, and with φ-divergences instead of Bregman distances, it has been proved in Iusem and Teboulle (1995) that the sequence $\{s^k\}$ converges to the dual solution.

In this paper we first study the limiting behavior of the path $x(\mu)$ consisting of the optimal solutions of the following family of problems parametrized by a parameter $\mu > 0$:

$$\min \left\{ c^T x + \mu D_\varphi(x, x^1) : Ax = b \right\}.$$

We also study the limiting behavior of an associated dual path $s(\mu)$ defined in (10) below. More specifically, our main goal is obtain a characterization of the limiting behavior of the derivatives of these paths. Our analysis in this part uses several ideas from Adler and Monteiro (1991). Using these results, we then establish convergence of the sequence $\{s^k\}$ as well as convergence rate results for $\{x^k\}$ and $\{\bar{s}^k\}$. We show that both sequences $\{x^k\}$ and $\{\bar{s}^k\}$ have sublinear convergence rates if $0 < \limsup_{k\to+\infty} \lambda_k$ and also give examples of sequences $\{\lambda_k\}$ such that $\lim_{k\to\infty} \lambda_k = 0$ for which the corresponding sequences $\{x^k\}$ and $\{\bar{s}^k\}$ both converge either linearly or superlinearly.

The organization of our paper is as follows. In Subsection 1.1 we list some basic notation and terminology used in our presentation. In Section 2 we review some known concepts, introduce the assumptions that will be used in our presentation and state some basic results. In Section 3, we study the limiting behavior of a perturbed dual path, and use the derived results to analyze the limiting behavior of the derivatives of the paths $x(\mu)$ and $s(\mu)$. The convergence of $\{s^k\}$ is obtained in Section 4 as well as convergence rate results for $\{x^k\}$ and $\{\bar{s}^k\}$. In the Section 5 we make some remarks. We conclude this paper by given in Appendix the proofs of some technical results.

1.1 Notations

We will use the following notation throughout this paper. $\mathbb{R}^n$ denotes the n-dimensional Euclidean space. The Euclidean norm is denoted by $\|\cdot\|$. Define $\mathbb{R}^n_+ \equiv \{x \in \mathbb{R}^n : x_j \geq 0\,,\, j = 1, \dots, n\}$ and $\mathbb{R}^n_{++} \equiv \{x \in \mathbb{R}^n : x_j > 0\,,\, j = 1, \dots, n\}$. The i-th component of a vector $x \in \mathbb{R}^n$ is denoted by x_i for every $i = 1, \dots, n$. Given an index set $J \subseteq \{1, \dots, n\}$, $\mathbb{R}^J$ will denote the set of vectors indexed by J and a vector $x \in \mathbb{R}^J$ is often denoted by x_J. For $J \subseteq \{1, \dots, n\}$ and a vector $x \in \mathbb{R}^n$, we also denote the subvector $[x_i]_{i\in J}$ by x_J and for $x_J > 0$. Given $x, y \in \mathbb{R}^n$, their Hadamard product, denoted by xy, is defined as $xy = (x_1y_1, \dots, x_ny_n) \in \mathbb{R}^n$ and for $\lambda \in \mathbb{R}$ the vector $[{x_i}^\lambda]_{i\in J}$ will denote by $(x_J)^\lambda$. The set of all $m \times n$ matrices with real entries is denoted by $\mathbb{R}^{m\times n}$. If x is a lower case letter that denotes a vector $x \in \mathbb{R}^n$, then the capital letter will denote the diagonal matrix with the components of the vector on the diagonal, i.e., $X = \operatorname{diag}(x_1, \dots, x_n)$. For a matrix A, we let A^T denote its transpose, $\operatorname{Im} A$ denote the subspace generated by the columns of A and $\operatorname{Null} A$ denote the subspace orthogonal to the rows of A. Given $A \in \mathbb{R}^{m\times n}$ and $J \subseteq \{1, \dots, n\}$, we denote by A_J the submatrix of A consisting of all columns of A indexed by indices in J.

2 Preliminaries

In this section we define the notion of the primal and dual central path for an LP problem in standard form wih respect to a given Bregman barrier and recall some results about the limiting behavior of these paths. We also describe the class of Bregman functions considered in this paper and state its basic properties.

We consider the linear programming problem

$$\min\left\{c^T x : Ax = b,\, x \geq 0\right\}, \tag{4}$$

where $c \in \mathbb{R}^n$, $A \in \mathbb{R}^{m\times n}$ has full row rank and $b \in \mathbb{R}^m$. We make two assumptions on problem (4), whose solution set will be denoted as X^*:

A1) $X^* \neq \emptyset$.

A2) $\mathcal{F}^0 \equiv \left\{x \in \mathbb{R}^n_{++} : Ax = b\right\} \neq \emptyset$.

Associated with problem (4), we have the dual problem

$$\min\left\{\tilde{x}^T s : s \in c + \operatorname{Im} A^T,\, s \geq 0\right\}, \tag{5}$$

where $\tilde{x} \in \mathbb{R}^n$ is any point such that $A\tilde{x} = b$. Under condition A1, the optimal set to the dual problem (5), which we denote by S^*, is a nonempty polyhedral set, namely

$$S^* = \left\{s \in \mathbb{R}^n_+ : s \in c + \operatorname{Im} A^T,\, \bar{x}^T s = 0\right\},$$

where $\bar{x}$ is an arbitrary element of X^*. Moreover, it is known that S^* is bounded when, in addition, A2 holds.

We consider separable barrier functions φ for the nonnegative orthant $\mathbb{R}^n_+$, i.e.,

$$\varphi(x) \equiv \sum_{j=1}^{n} \varphi_j(x_j) \tag{6}$$

satisfying certain assumptions described below. The first assumption we make on φ is as follows.

H1) The function $\varphi : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ is closed, strictly convex, twice continuously differentiable in $\mathbb{R}^n_{++}$, and such that

i) $\lim_{t\to 0} \varphi_j(t) = +\infty$ or $\lim_{t\to 0} \varphi_j(t) = 0$ for each $j \in \{1, \dots, n\}$;

ii) $\lim_{t\to 0} \varphi_j'(t) = -\infty$ for each $j \in \{1, \dots, n\}$.

We mention that assumption **H1**(i) is not really restrictive, because the algorithms we are interested in are invariant through addition of a constant to φ, and therefore, without loss of generality, we can add an appropriate constant to φ so that $\lim_{t\to 0} \varphi_j(t) = 0$ whenever $\lim_{t\to 0} \varphi_j(t) < +\infty$, for each $j \in \{1, \dots, n\}$.

Our second assumption on φ is:

H2) Assume that there exist $\gamma \in (0, 1)$ such that

$$r_j \equiv \lim_{t\to 0} -\frac{\varphi_j'(t)}{\varphi_j''(t)^\gamma} \in (0, \infty), \quad \forall j \in \{1, \dots, n\}. \tag{7}$$

From now on, we consider barrier functions φ satisfying assumptions **H1** and **H2**. We present next several examples, with the corresponding values of γ and r_j. In one case, we have used special definitions for values of t far from 0, so that $\mathrm{dom}(\varphi_j)$ contains the whole half line $(0, +\infty)$, and therefore φ can be used to generate a D_φ whose zone is $\mathbb{R}_{++}$, but we emphasize that both γ and r_j depend only on the behavior of φ near 0.

Example 1. *For each $j \in \{1, \dots, n\}$, let:*

i) $\varphi_j(t) = t^\alpha - t^\beta$, *with* $\alpha \geq 1$ *and* $\beta \in (0,1)$. *Then* $\gamma = (1-\beta)/(2-\beta) \in (0, 1/2)$ *and*

$$r_j = \left(\frac{\beta}{(1-\beta)^{1-\beta}} \right)^{1/(2-\beta)};$$

ii)

$$\varphi_j(t) = \begin{cases} -\left(1-(1-t)^\alpha\right)^{1/\alpha}, & \text{if } t \in (0,1); \\ (t-1)^\alpha/2, & \text{if } t \geq 1, \end{cases}$$

with $\alpha \geq 2$. *Then* $\gamma = (\alpha-1)/(2\alpha-1) \in (0,1/2)$ *and* $r_j = (\alpha-1)^{(1-\alpha)/(2\alpha-1)}$*;*

iii) $\varphi_j(t) = -\log t$. *Then* $\gamma = 1/2$ *and* $r_j = 1$*;*

iv) $\varphi_j(t) = (t-1)\log t$. *Then* $\gamma = 1/2$ *and* $r_j = 1$*;*

v) $\varphi_j(t) = t^{-\alpha}$ *with* $\alpha > 0$. *Then,* $\gamma = (\alpha+1)/(\alpha+2) \in (1/2, 1)$ *and* $r_j = \left[\alpha(\alpha+1)^{-(\alpha+1)}\right]^{1/(\alpha+2)}$.

Finally, we make a last hypothesis on φ, which will be used only in Subsection 3.2 opposite of the hypothesis **H1**, **H2** used in whole paper.

H3) There exists $\nu \neq 0$ such that $\lim_{t \to 0} \varphi_j'(t) + \nu\, t\, \varphi_j''(t) \in \mathbb{R}$ for all $j \in \{1, \dots, n\}$.

We present next some examples of functions φ satisfying hypotheses **H1**, **H2** and **H3**, with the corresponding values of ν.

Example 2. *For each $j \in \{1, \dots, n\}$ we take:*

i) $\varphi_j(t)$ *as in Example 1.i. Then,* $\nu = 1/(1-\beta)$*;*

ii) $\varphi_j(t)$ *as in Example 1.iii. Then,* $\nu = 1$*;*

iii) $\varphi_j(t)$ *as in Example 1.v. Then,* $\nu = 1/(1+\alpha)$.

We remark that φ as defined in Examples 1(ii) and 1(iv) does not satisfy **H3**.

The *Bregman distance* associated with φ is the function $D_\varphi \colon \mathbb{R}^n \times \mathbb{R}^n_{++} \to \mathbb{R} \cup \{+\infty\}$ defined as

$$D_\varphi(x,y) \equiv \varphi(x) - \varphi(y) - \nabla\varphi(y)^T(x-y), \quad \forall (x,y) \in \mathbb{R}^n \times \mathbb{R}^n_{++}. \tag{8}$$

Observe that $D_\varphi(x,y) = +\infty$ for every $x \notin \mathbb{R}^n_+$ and $y \in \mathbb{R}^n_{++}$. Now, let $x^1 = (x^1_1, ..., x^1_n) \in \mathcal{F}^0$ be given and consider the "barrier" function $D_\varphi(\cdot, x^1)$. The *primal central path* $\{x(\mu) : \mu > 0\}$ and the *dual central path* $\{s(\mu) : \mu > 0\}$ for problem (4) with respect to the barrier $D_\varphi(\cdot, x^1)$ are defined as

$$x(\mu) \equiv \operatorname{argmin}\left\{c^T x + \mu D_\varphi(x, x^1) : Ax = b\right\}, \quad \forall \mu > 0, \tag{9}$$

and

$$s(\mu) \equiv -\mu\left(\nabla\varphi(x(\mu)) - \nabla\varphi(x^1)\right), \quad \forall \mu > 0. \tag{10}$$

It is well-known that, for every $\mu > 0$, problem (9) has a unique optimal solution which is strictly positive (see Proposition 2 of Iusem et al. [10]). This clearly implies that both $x(\mu)$ and $s(\mu)$ are well-defined for every $\mu > 0$. Moreover, the optimality conditions for (9) imply that

$$s(\mu) \in c + \operatorname{Im} A^T, \quad \forall \mu > 0. \tag{11}$$

The optimal partition (B, N) for (4) is defined as

$$B \equiv \{j : x^*_j > 0 \text{ for some } x^* \in X^*\} \text{ and } N \equiv \{1, \ldots, n\} \setminus B.$$

The following result characterizes the limiting behavior of the primal and dual central paths (9) and (10).

Proposition 1. *The following statements hold:*

i) $\lim_{\mu\to 0} x(\mu) = x^*$, *where* $x^* = \operatorname{argmin}_{x\in X^*} \sum_{j\in B} D_{\varphi_j}(x_j, x^1_j)$*;*

ii) $\lim_{\mu\to 0} s(\mu) = s^*$, *where* s^* *is the unique optimal solution of the problem*

$$\min\left\{\sigma_N(s_N) : s \in c + \operatorname{Im} A^T,\ s_B = 0\right\}, \tag{12}$$

and σ_N *is the strictly convex function defined as*

$$\sigma_N(s_N) = \begin{cases} \frac{\gamma^2}{(1-2\gamma)(1-\gamma)} \sum_{j\in N} r_j^{1/\gamma} s_j^{2-1/\gamma}, & \text{if } \gamma \in (0,1) \setminus \{1/2\}; \\ -\sum_{j\in N} r_j^2 \log s_j, & \text{if } \gamma = 1/2. \end{cases} \tag{13}$$

Proof. The statement related to the primal central path was proved in Theorem 1 of Iusem, Svaiter and Cruz Neto (1999). The statement related to the dual central path follows from Proposition 7 in Iusem and Monteiro (2000) and the fact that items (i) and (ii) of Corollary 4 in the Appendix imply that

$$\sigma_N(s_N) = \begin{cases} \lim_{\mu\to 0} \frac{1}{\mu^{1/\gamma-2}} \sum_{j\in N} \varphi^*_j\left(\delta_j - \frac{s_j}{\mu}\right), & \text{if } \gamma \in (0,1) \setminus \{1/2\}; \\ \lim_{\mu\to 0} \sum_{j\in N} \left(\varphi^*_j\left(\delta_j - \frac{s_j}{\mu}\right) - \varphi^*_j\left(\delta_j - \frac{1}{\mu}\right)\right), & \text{if } \gamma = 1/2, \end{cases}$$

for any $\delta \in \operatorname{dom}\varphi^*$, where φ^*_j denotes the conjugate function of φ_j. □

We conclude this subsection by giving a result about the limiting behavior of the second derivative of φ along the primal central path. For every $\mu > 0$, define

$$g_B(\mu) \equiv \nabla^2 \varphi_B(x_B(\mu))e, \qquad h(\mu) \equiv \mu^{-1/\gamma} \left[\nabla^2 \varphi(x(\mu))\right]^{-1} e, \tag{14}$$

where e denotes the vector of all ones of appropriate dimension.

Corollary 1. *The following statements hold:*

i) $\lim_{\mu\to 0} h_j(\mu) = +\infty$ *for all* $j \in B$ *and* $\lim_{\mu\to 0} h_N(\mu) = h_N^* > 0$*, where* $h_N^* \equiv \left(r_N(s_N^*)^{-1}\right)^{1/\gamma}$*;*

ii) $\lim_{\mu\to 0} g_B(\mu) = g_B^* > 0$*, where* $g_B^* \equiv \nabla^2 \varphi_B(x_B^*)e$*.*

Proof. From Proposition 1 and equation (10) we have that $s^* = \lim_{\mu\to 0} -\mu\nabla\varphi(x(\mu))$. This together with (14), the fact that $s_B^* = 0$ and hypothesis **H2** then imply (i). Statement (ii) follows from the twice continuous differentiability of of φ in $\mathbb{R}^n_{++}$ and the fact that $x_B^* > 0$.

□

3 Limiting behavior of the derivatives of the paths

Our aim is to prove the convergence of the dual proximal sequence (3). We are also interested in obtaining convergence rate results for the primal proximal sequence (2). Instead of starting by analyzing the behavior of these sequences, we study first the behavior of the derivatives of the dual and primal paths. The motivation is that the primal (resp. average dual) proximal sequence is contained in the primal (resp. dual) path corresponding, see Proposition 3.

3.1 Limiting behavior of the primal central path

In this section, we study the limiting behavior of the primal central path $x(\mu)$ as μ goes to 0. As an intermediate step, we also study the limiting behavior of a certain perturbed dual path.

The perturbed dual path $s^E(\mu)$ is defined as

$$s^E(\mu) \equiv s(\mu) - \mu\dot{s}(\mu), \quad \forall \mu > 0. \tag{15}$$

Derivating (10) and using (15), we easily see that

$$s^E(\mu) = \mu^2 \nabla^2 \varphi(x(\mu))\dot{x}(\mu), \quad \forall \mu > 0. \tag{16}$$

In view of (14), this is equivalent to

$$h(\mu)s^E(\mu) = \mu^{2-1/\gamma}\,\dot{x}(\mu), \quad \forall \mu > 0. \tag{17}$$

Lemma 1. $\lim_{\mu\to 0} s^E(\mu) = \bar{s}^E$, *where* $\bar{s}^E$ *is the unique optimal solution of the problem*

$$\min\left\{\frac{1}{2}\left\|(h_N^*)^{1/2}s_N\right\|^2 : s \in c + \operatorname{Im} A^T,\ s_B = 0\right\}. \tag{18}$$

Proof. We will first show that $\{s^E(\mu) : \mu \in (0,1]\}$ is bounded and that $\lim_{\mu\to 0} s_B^E(\mu) = 0$. In view of (11), we have that $\dot{s}(\mu) \in \operatorname{Im} A^T$ for all $\mu > 0$. This fact together with (11), (15) then imply that $s^E(\mu) \in c + \operatorname{Im} A^T$ for all $\mu > 0$. Fix some $\bar{s} \in S^*$ and note that $\bar{s} \in c + \operatorname{Im} A^T$ and $\bar{s}_B = 0$. It then follows that

$$s^E(\mu) - \bar{s} \in \operatorname{Im} A^T, \quad \forall \mu > 0. \tag{19}$$

On the other hand, by (17) and the fact that $\dot{x}(\mu) \in \operatorname{Null}(A)$, we conclude that $h(\mu)s^E(\mu) \in \operatorname{Null}(A)$ for all $\mu > 0$. This together with (19) then imply that $(s^E(\mu) - \bar{s})^T h(\mu) s^E(\mu) = 0$, and hence

$$\|h(\mu)^{1/2}s^E(\mu)\|^2 = s^E(\mu)^T h(\mu) s^E(\mu) = \bar{s}^T h(\mu) s^E(\mu) \leqslant \|h(\mu)^{1/2}\bar{s}\|\,\|h(\mu)^{1/2}s^E(\mu)\|,$$

which in turn yields

$$\|h(\mu)^{1/2}s^E(\mu)\| \leqslant \|h(\mu)^{1/2}\bar{s}\| = \|h_N(\mu)^{1/2}\bar{s}_N\|,$$

where the last equality is due to the fact that $\bar{s}_B = 0$. By Corollary 1, we know that $h_B(\mu)$ and $h_N(\mu)$ converges to $+\infty$ and some strictly positive vector, respectively, as μ tends to 0. This observation together with the previous inequality then imply that $\{s^E(\mu) : \mu \in (0,1]\}$ is bounded and $\lim_{\mu\to 0} s_B^E(\mu) = 0$.

We will now show that any accumulation point $\bar{s}$ of $\{s^E(\mu) : \mu \in (0,1]\}$ satisfies the optimality conditions for (18), from which the result follows. Clearly, $\bar{s}$ is feasible for (18). Moreover, by (17) and the fact that $A\dot{x}(\mu) = 0$, we conclude that $A_N\left(h_N(\mu)s_N^E(\mu)\right) \in \operatorname{Im} A_B$ for all $\mu > 0$. This equation together with Corollary 1 then imply that $A_N\left(h_N^*\bar{s}_N\right) \in \operatorname{Im} A_B$. We have thus proved that $\bar{s}$ satisfies the optimality condition for (18). □

Theorem 1. *The following statements hold:*

i) $\lim_{\mu\to 0} s^E(\mu) = s^*$;

ii) $\lim_{\mu\to 0} \mu\dot{s}(\mu) = 0$.

Proof. To prove i), it suffices to show that s^* satisfies the optimality conditions for (18). Since s^* is the optimal solution of (12), it is feasible for (18) and satisfies $A_N\left(\nabla\sigma_N(s_N^*)\right) \in \operatorname{Im} A_B$. Using (13), we easily see that for any $\gamma \in (0,1)$:

$$\nabla\sigma_N(s_N^*) = \frac{-\gamma}{1-\gamma}\left(r_N(s_N^*)^{-1}\right)^{1/\gamma} s_N^* = \frac{-\gamma}{1-\gamma} h_N^* s_N^*.$$

These two observations then imply that $A_N (h_N^* s_N^*) \in \operatorname{Im} A_B$, and hence that s^* satisfies the optimality condition for (18).

Statement ii) follows by noting that statement i), relation (15) and Proposition 1(ii) imply that

$$\lim_{\mu \to 0} \mu \dot{s}(\mu) = \lim_{\mu \to 0} \left[s(\mu) - s^E(\mu) \right] = s^* - s^* = 0.$$

□

We now turn our attention to the analysis of the limiting behavior of the derivative of the primal central path. We start by stating the following technical result, which is essentially one of the many ways of stating Hoffman's lemma for system of linear equations (see Hoffman (1952)).

Lemma 2. *Let a subspace $E \subseteq \mathbb{R}^n$ and an index set $J \subset \{1, \ldots, n\}$ be given and define $\bar{J} \equiv \{1, \ldots, n\} \setminus J$. Then, there exists a constant $M = M(E, J) \geq 0$ with the following property: for each $u = (u_J, u_{\bar{J}}) \in E$, there exists $\tilde{u}_J$ such that $(\tilde{u}_J, u_{\bar{J}})^T \in E$ and $\|\tilde{u}_J\| \leqslant M \|u_{\bar{J}}\|$*

Theorem 2. $\lim_{\mu \to 0} \mu^{2-1/\gamma} \dot{x}(\mu) = d^\infty$, *where d^∞ is the unique optimal solution of the problem*

$$\min \left\{ \frac{1}{2} \left\| (g_B^*)^{1/2} d \right\|^2 : d \in \operatorname{Null} A,\ d_N = h_N^* s_N^* \right\}. \tag{20}$$

Proof. To simplify notation, let $d(\mu) \equiv \mu^{2-1/\gamma} \dot{x}(\mu)$ for all $\mu > 0$. We will first show that the set $\{d(\mu) : \mu \in (0, 1]\}$ is bounded. By (17), Corollary 1 and Theorem 1, we have

$$\lim_{\mu \to 0} d_N(\mu) = \lim_{\mu \to 0} \mu^{2-1/\gamma} \dot{x}_N(\mu) = \lim_{\mu \to 0} h_N(\mu) s_N^E(\mu) = h_N^* s_N^* > 0. \tag{21}$$

Clearly, $Ad(\mu) = A_B d_B(\mu) + A_N d_N(\mu) = 0$ for all $\mu > 0$. Applying Lemma 2 with $E = \operatorname{Null}(A)$ and $J = B$, we conclude that there exists $M \geq 0$ and a function $p_B : \mathbb{R}_{++} \to \mathbb{R}^B$ such that for every $\mu > 0$:

$$A_B p_B(\mu) + A_N d_N(\mu) = 0, \quad \|p_B(\mu)\| \leqslant M \|d_N(\mu)\|.$$

This together with (21) imply that the set $\{p_B(\mu) : \mu \in (0, 1]\}$ is bounded and that $p_B(\mu) - d_B(\mu) \in \operatorname{Null}(A_B)$ for all $\mu > 0$. On the other hand, by (16), (19) and the fact that $\bar{s}_B = 0$, we conclude that $g_B(\mu) d_B(\mu) \in \operatorname{Im}(A_B^T)$ for all $\mu > 0$. The two last conclusions then imply that $(p_B(\mu) - d_B(\mu))^T g_B(\mu) d_B(\mu) = 0$ for all $\mu > 0$. An argument similar to the one used in the proof of Lemma 1 then shows that $\|g_B(\mu)^{1/2} d_B(\mu)\| \leqslant \|g_B(\mu)^{1/2} p_B(\mu)\|$ for all $\mu > 0$. This inequality together with Corollary 1(ii) and the fact that $\{p_B(\mu) : \mu \in (0, 1]\}$ is bounded then imply that $\{d_B(\mu) : \mu \in (0, 1]\}$ is also bounded. We have thus shown that $\{d(\mu) : \mu \in (0, 1]\}$ is bounded.

Now, let $\bar{d}$ be an accumulation point of $\{d(\mu) : \mu \in (0, 1]\}$. Clearly, (21) and the fact that $Ad(\mu) = 0$ for all $\mu > 0$ imply that $\bar{d}$ is feasible for (20). Moreover, the fact that $g_B(\mu) d_B(\mu) \in \operatorname{Im}(A_B^T)$ imply that $g_B^* \bar{d}_B \in \operatorname{Im}(A_B^T)$. We have thus shown that $\bar{d}$ satisfy the optimality condition for (20), and hence that $\bar{d} = d^\infty$. Since this holds for any accumulation point $\bar{d}$ of $\{d(\mu) : \mu \in (0, 1]\}$, the result follows. □

Corollary 2. *The following limits hold:*

$$\lim_{\mu\to 0}\frac{x(\mu)-x^*}{\mu^{1/\gamma-1}} = \frac{d^\infty}{1/\gamma-1} \neq 0, \quad \lim_{\mu\to 0}\frac{c^T(x(\mu)-x^*)}{\mu^{1/\gamma-1}} = \frac{\|(h_N^*)^{1/2}s_N^*\|^2}{1/\gamma-1} > 0. \tag{22}$$

Proof. We will only prove that the second limit holds since the first one one can proved in a similar way. Using the fact that $c - s^* \in \operatorname{Im} A^T$, $x(\mu) - x^* \in \operatorname{Null}(A)$, $x_N^* = 0$ and $s_B^* = 0$, we easily see that $c^T(x(\mu) - x^*) = (s_N^*)^T x_N(\mu)$ for all $\mu > 0$. Using this identity together with L'Hospital's rule of calculus and relation (21), we obtain

$$\lim_{\mu\to 0}\frac{c^T(x(\mu)-x^*)}{\mu^{1/\gamma-1}} = \lim_{\mu\to 0}\frac{(s_N^*)^T x_N(\mu)}{\mu^{1/\gamma-1}} = \lim_{\mu\to 0}\frac{(s_N^*)^T \dot{x}_N(\mu)}{(1/\gamma-1)\mu^{1/\gamma-2}} = \frac{\|(h_N^*)^{1/2}s_N^*\|^2}{1/\gamma-1}.$$

□

3.2 Limiting behavior of the derivative of the dual path

In this section we are interested in the behavior of $\dot{s}(\mu)$ as μ goes to 0. We mention that in this section we do need **H3**.

Letting ν be as in Assumption **H3**, define

$$v(\mu) \equiv \big(\nabla\varphi(x(\mu)) - \nabla\varphi(x^1)\big) + \nu\nabla^2\varphi(x(\mu))x(\mu), \quad \forall \mu > 0. \tag{23}$$

Lemma 3. *The following statements hold:*

i) if Assumption **H3** *holds then $v^* \equiv \lim_{\mu\to 0} v(\mu)$ exists and is finite;*

ii) $h(\mu)(\dot{s}(\mu) + v(\mu)) = \mu^{-1/\gamma}[\nu x(\mu) - \mu\dot{x}(\mu)]$ for all $\mu > 0$;

iii) $A_N[h_N(\mu)(\dot{s}_N(\mu) + v_N(\mu))] \in \operatorname{Im} A_B$ for all $\mu > 0$.

Proof. The fact that $\lim_{\mu\to 0} v_N(\mu)$ exists and is finite is an immediate consequence of Assumption **H3**, while the existence and finiteness of $\lim_{\mu\to 0} v_B(\mu)$ is obvious. Now, derivating (10) and using (23), we obtain for all $\mu > 0$ that

$$\begin{aligned} \dot{s}(\mu) &= -\big(\nabla\varphi(x(\mu)) - \nabla\varphi(x^1)\big) - \mu\nabla^2\varphi(x(\mu))\dot{x}(\mu) \\ &= -v(\mu) + \nabla^2\varphi(x(\mu))[\nu x(\mu) - \mu\dot{x}(\mu)]. \end{aligned} \tag{24}$$

Statement (ii) now follows by rearranging this expression and using (14). Using the fact that $Ax(\mu) = b$, $A\dot{x}(\mu) = 0$ and $b \in \operatorname{Im} A_B$, we easily see that $A_N(\nu x_N(\mu) - \mu\dot{x}_N(\mu)) \in \operatorname{Im} A_B$ for all $\mu > 0$. Statement (iii) now follows from this conclusion in view of statement (ii). □

Theorem 3. *Suppose that Assumption* **H3** *holds. Then,* $\dot{s}^\infty \equiv \lim_{\mu\to 0} \dot{s}(\mu)$ *exists and* $\dot{s}^\infty$ *is characterized as follows:* $\dot{s}^\infty_B = -(\nabla\varphi_B(x^*_B) - \nabla\varphi_B(x^1_B))$ *and* $\dot{s}^\infty_N$ *is the unique optimal solution of the problem*

$$\min\left\{\frac{1}{2}\left\|(h^*_N)^{1/2}\,(p_N + v^*_N)\right\|^2 : \begin{pmatrix}\dot{s}^\infty_B \\ p_N\end{pmatrix} \in \operatorname{Im} A^T\right\}. \tag{25}$$

Proof. By Theorem 2 and the fact that $\gamma \in (0,1)$, we have that $\lim_{\mu\to 0} \mu\dot{x}(\mu) = 0$. Hence, by (24) and the fact that $\lim_{\mu\to 0} x_B(\mu) = x^*_B > 0$, we conclude that $\lim_{\mu\to 0} \dot{s}_B(\mu) = -(\nabla\varphi_B(x^*_B) - \nabla\varphi_B(x^1_B))$. We will now show that the set $\{\dot{s}_N(\mu) : \mu \in (0,1]\}$ is bounded. Indeed, it follows from (11) that $\dot{s}(\mu) = (\dot{s}_B(\mu), \dot{s}_N(\mu))^T \in \operatorname{Im} A^T$ for all $\mu > 0$. Applying Lemma 2 with $E = \operatorname{Im} A^T$ and $J = N$, we conclude that there exist a constant $M_1 \geq 0$ and a function $p_N : \mathbb{R}_{++} \to \mathbb{R}^N$ such that for every $\mu > 0$:

$$(\dot{s}_B(\mu), p_N(\mu))^T \in \operatorname{Im} A^T, \quad \|p_N(\mu)\| \leqslant M_1 \|\dot{s}_B(\mu)\|.$$

This implies that $(0, p_N(\mu) - \dot{s}_N(\mu))^T \in \operatorname{Im} A^T$ for all $\mu > 0$ and that the set $\{p_N(\mu) : \mu \in (0,1]\}$ is bounded in view of the fact that $\lim_{\mu\to 0} \dot{s}_B(\mu)$ exists and is finite. The first conclusion together with Lemma 3(iii) then imply that $(p_N(\mu) - \dot{s}_N(\mu))^T h_N(\mu)(\dot{s}_N(\mu) + v_N(\mu)) = 0$ for all $\mu > 0$. Hence, we have

$$\begin{aligned}\|h_N(\mu)^{1/2}(\dot{s}_N(\mu) + v_N(\mu))\|^2 &= (\dot{s}_N(\mu) + v_N(\mu))^T h_N(\mu)(\dot{s}_N(\mu) + v_N(\mu)) \\ &= (p_N(\mu) + v_N(\mu))^T h_N(\mu)(\dot{s}_N(\mu) + v_N(\mu)) \\ &\leqslant \|h_N(\mu)^{1/2}(p_N(\mu) + v_N(\mu))\|\, \|h_N(\mu)^{1/2}(\dot{s}_N(\mu) + v_N(\mu))\|,\end{aligned}$$

which in turn implies

$$\|h_N(\mu)^{1/2}(\dot{s}_N(\mu) + v_N(\mu))\| \leqslant \|h_N(\mu)^{1/2}(p_N(\mu) + v_N(\mu))\|, \quad \forall\mu > 0.$$

This inequality together with Lemma 3(i), Corollary 1(i) and the fact that $\{p_N(\mu) : \mu \in (0,1]\}$ is bounded immediately implies that $\{\dot{s}_N(\mu) : \mu \in (0,1]\}$ is bounded. Now, with the aid of (i) and (iii) of Lemma 3, Corollary 1(i) and the fact that $\dot{s}(\mu) \in \operatorname{Im} A^T$ and $\dot{s}^\infty_B = \lim_{\mu\to 0} \dot{s}(\mu)$, we easily see that any accumulation point of $\dot{s}_N(\mu)$ is feasible for (25) and satisfies its corresponding optimality condition. Hence, it follows that $\lim_{\mu\to 0} \dot{s}_N(\mu)$ exists and is characterized as in the statement of the theorem. □

Corollary 3. *Under Assumption* **H3**, *we have* $\lim_{\mu\to 0} (s(\mu) - s^*)/\mu = \dot{s}^\infty$.

Proof. The corollary follows immediately from Theorem 3 and the L'Hospital's rule from calculus. □

4 The proximal sequence

In this section we define the primal and dual proximal sequences generated by the proximal point method with Bregman distances, and the associated averaged sequence, in order to prove our main result.

The proximal point method with the Bregman distance D_φ for solving problem (4) generates a sequence $\{x^k\} \subset \mathcal{F}^0$ defined as $x^0 \in \mathcal{F}^0$ and

$$x^{k+1} = \operatorname{argmin}\left\{c^T x + \lambda_k D_\varphi(x, x^k) : Ax = b\right\}, \tag{26}$$

where the sequence $\{\lambda_k\} \subseteq \mathbb{R}^n_{++}$ satisfies

$$\sum_{k=0}^{\infty} \lambda_k^{-1} = +\infty. \tag{27}$$

The following result on the convergence of $\{x^k\}$, as defined in (26), is known.

Proposition 2. *The sequence $\{x^k\}$ generated by* (26) *converges to a solution of problem* (4)*.*

Proof. See, e.g. Theorem 3 of Iusem, Svaiter and Cruz Neto (1999). □

The optimality condition for x^{k+1} to be an optimal solution of (26) is that $s^k \in c + \operatorname{Im} A^T$, where

$$s^k \equiv \lambda_k \left(\nabla\varphi(x^k) - \nabla\varphi(x^{k+1})\right). \tag{28}$$

Note that in principle s^k may fail to be nonnegative, and hence dual feasible.

In this section, we are interested in describing the convergence properties of the dual sequence $\{s^k\}$. Instead of dealing directly with the sequence $\{s^k\}$, we first study the behavior of the averaged sequence $\{\bar{s}^k\}$ defined as

$$\bar{s}^k = \mu_k \sum_{i=1}^{k} {\lambda_i}^{-1} s^i, \quad \text{where } \mu_k = \left(\sum_{i=1}^{k} \lambda_i^{-1}\right)^{-1}. \tag{29}$$

Observe that $\{\mu_k\}$ converges to 0 in view of (27). The following result describes how the sequences $\{x^k\}$ and $\{\bar{s}^k\}$ relate to the primal and dual central paths, respectively, for problem (4) with respect to the barrier $D_\varphi(\cdot, x^1)$.

Proposition 3. *Let $x(\mu)$ and $s(\mu)$ denote the primal and dual central paths, respectively, for problem* (4) *with respect to the barrier $D_\varphi(\cdot, x^1)$. Then, for every $k \geq 1$, $x^{k+1} = x(\mu_k)$ and $\bar{s}^k = s(\mu_k)$. As a consequence,* $\lim_{k\to\infty} \bar{s}^k = s^*$.

Proof. The part related to primal sequence was proved in Theorem 3 of Iusem, Svaiter and Cruz Neto (1999) and dual sequence is proved in the Final Remarks section of Iusem and Monteiro (2000). The last conclusion of the proposition follows immediately from Proposition 1(ii). ☐

In view of the fact that $\{\bar{s}^k\}$ converges to s^*, it is natural to conjecture that $\{s^k\}$ does too. The next result, which is one of the main results of this paper, shows that this is indeed the case.

Theorem 4. $\lim_{k\to+\infty} s^k = s^*$.

Proof. Using (29) and Proposition 3, it is easy to verify that

$$s^k - \bar{s}^k = s^k - s(\mu_k) = \frac{\lambda_k}{\mu_{k-1}} \left(s(\mu_k) - s(\mu_{k-1}) \right), \quad \forall k \geq 2.$$

Hence, by the mean value theorem, for each $k \geq 2$ and $i = 1, \ldots, n$, there exists $\xi_i^k \in (\mu_k, \mu_{k-1})$ such that

$$\left| s_i^k - \bar{s}_i^k \right| = \left| \frac{\lambda_k}{\mu_{k-1}} (\mu_k - \mu_{k-1})\, \dot{s}_i(\xi_i^{k-1}) \right| = \left| \mu_k\, \dot{s}_i(\xi_i^k) \right| \leqslant \left| \xi_i^k \dot{s}_i(\xi_i^k) \right|, \tag{30}$$

where the second equality follows from the definition of μ_k in (29). Since $\lim_{k\to+\infty} \mu_k = 0$ and $0 < \xi_i^k \leqslant \mu_{k-1}$ for all $i = 1, \ldots n$, it follows from Theorem 1(ii) that $\lim_{k\to+\infty} \xi_i^k \dot{s}_i(\xi_i^k) = 0$. Using this fact in (30), we conclude that $\lim_{k\to\infty} s^k - \bar{s}^k = 0$, which in turn implies the theorem in view of Proposition 3. ☐

We will now see how the results of Section 3 can be used to obtain convergence rate results with respect to the primal and dual (averaged) sequences.

Theorem 5. *Define* $\tau \equiv \limsup_{k\to\infty} \mu_{k+1}/\mu_k \in [0,1]$. *Then, the following holds for the primal proximal sequence* $\{x^k\}$ *given by* (26)*:*

$$\limsup_{k\to\infty} \frac{\left\| x^{k+1} - x^* \right\|}{\left\| x^k - x^* \right\|} = \tau^{1/\gamma - 1}, \quad \limsup_{k\to\infty} \frac{c^T (x^{k+1} - x^*)}{c^T (x^k - x^*)} = \tau^{1/\gamma - 1}. \tag{31}$$

If, in addition, Assumption **H3** *holds and the limit* $\dot{s}^\infty$ *of Theorem 3 is nonzero, then the following holds for the proximal dual average sequence* $\{\bar{s}^k\}$ *given by* (29)*:*

$$\limsup_{k\to\infty} \frac{\left\| \bar{s}^{k+1} - s^* \right\|}{\left\| \bar{s}^k - s^* \right\|} = \tau. \tag{32}$$

Proof. The three limits in (31) and (32) can be easily derived using Corollaries 2 and 3 together with the fact that $x^{k+1} = x(\mu_k)$ and $\bar{s}^k = s(\mu_k)$ for all $k \geq 1$ in view of Proposition 3. ☐

Using the definition of μ_k in (29), we easily see that $\mu_{k+1}/\mu_k = (1 + \mu_k/\lambda_{k+1})^{-1}$. Thus, if the condition $\limsup_{k\to\infty} \lambda_k > 0$ holds, then we have $\tau \equiv \limsup_{k\to\infty} \mu_{k+1}/\mu_k = 1$, and by Theorem 5, we conclude that the two sequences $\{x^k\}$ and $\{\bar{s}^k\}$ both converge Q-sublinearly. Faster convergence can only be achieved if the consition $\lim_{k\to\infty} \lambda_k = 0$ is imposed. For example, if for some $\beta \in (0, 1)$, we have $\lambda_k = \beta^k$ for all k, then $\tau = \beta$, implying that $\{x^k\}$ and $\{\bar{s}^k\}$ both converge Q-linearly. On the other hand, if $\lambda_k = 1/k!$ for all k, then $\tau = 0$, implying that $\{x^k\}$ and $\{\bar{s}^k\}$ both converge Q-superlinearly.

5 Final Remarks

We finish this paper with some remarks about the case where $\gamma = 0$. This case is indeed relevant because it includes the entropic Bregman distance (known to statisticians as the Kullback-Leibler information divergence), corresponding to the barrier in Example 3(i) below. This is the prototypical example of a Bregman distance, and the only one considered, either explicitly or implicitly, in early references, like Eriksson (1985), Eggermont (1990) and Tseng and Bertsekas (1993). Precisely in this reference a linear convergence was established for the primal sequence generated by the proximal method with this Bregman distance applied to linear programming, assuming $\bar{\lambda} \equiv \limsup_{k\to\infty} \lambda_k > 0$. This result was extended in Iusem and Teboulle (1995) to a large class of φ-divergences, including the Kullback-Leibler one, which incidentally, is the only barrier which gives rise both to a Bregman distance and to a φ-divergence, up to additive and/or multiplicative constants.

In these references the linear convergence rate is established not for the sequence $\{\|x^k - x^*\|\}$, but for the sequence of functional values $\{c^T(x^k - x^*)\}$ and for the distance from x^k to the primal solution set, but the result can be extended without trouble to the primal sequence itself. On the other hand, we have seen in the previous section that with $\bar{\lambda} > 0$ the convergence rate of $\{\|x^k - x^*\|\}$ is sublinear for all separable Bregman distances with $\gamma > 0$. It follows that **H1**, i.e. demanding $\gamma > 0$, makes it possible a convergence analysis of the dual sequence, but it also has the negative side effect of worsening the convergence rate, from linear to sublinear. We conjecture indeed that with $\bar{\lambda} > 0$ the convergence rate is linear for all separable Bregman distance with $\gamma_1 = \cdots = \gamma_n = 0$.

We also mention that the case of $\bar{\lambda} > 0$ is the most important one. In view of the definition of the method (equation (2)), it is natural that by taking sequences $\{\lambda_k\}$ which go to zero fast enough one can get convergence rates as high as desired (e.g. the examples at the end of the previous section), but such rates are somewhat deceiving, because when λ_k goes to zero, for large k the regularization term $\lambda_k D_\varphi(x, x^k)$ in (2) becomes numerically negligible, and the kth subproblem becomes equivalent to solving the original problem, which, if it can be solved in a straightforward way, makes the whole proximal method somewhat superfluous. One could assume that $\bar{\lambda}$ is always strictly positive in actual implementations of the method.

We establish now a result for the case of $\gamma = 0$. In this case, we do not assume **H2** and **H3**, only **H1**, and we can obtain the following result.

Proposition 4. *Assume that* $\varphi_1 = \cdots = \varphi_n \equiv \bar{\varphi}$, $\lim_{t\to 0} \bar{\varphi}(t) = 0$ *and that* $\xi \equiv \lim_{t\to 0} t\bar{\varphi}''(t)$ *exists and belongs to* $(0, +\infty)$. *Then, for every* $\delta \in \operatorname{dom}(\varphi^*) \subset \mathbb{R}^n$, $s \in \mathbb{R}^n_{++}$ *and* $J \subset \{1, \dots, n\}$, *we have:*

$$\lim_{\mu\to 0} \left(\sum_{j\in J} \bar{\varphi}^* \left(\delta_j - \frac{s_j}{\mu} \right) \right)^{\mu} = \max \left\{ e^{-s_j/\xi} : j \in J \right\}.$$

Proof. See the proof in the Appendix. □

We present next two examples of functions φ satisfying the hypotheses of Proposition 4, with the corresponding values of γ, r and ξ .

Example 3. *If we take:*

i) $\varphi_1(t) = t \log t$. *Then, in this case,* $\gamma = 0$, $\theta_1 = +\infty$, $\xi = 1$;

ii) $\varphi_1(t) = (t^2 + t) \log t$. *Then, in this case,* $\gamma = 0$, $\theta_1 = +\infty$, $\xi = 1$.

The next result stated below gives characterization of the limit point of the dual central path, namely an specific solution of the problem

$$\min \left\{ \sigma_N(s_N) : s \in c + \operatorname{Im} A^T,\ s_B = 0 \right\}, \tag{33}$$

called the centroid s^c of the dual solution set S^* (see e.g. Cominetti and San Martín (1994) and Iusem and Monteiro (2000).)

Proposition 5. *Suppose that the assumptions of Proposition 4 hold. Then* $\lim_{\mu\to 0} s(\mu) = s^c$, *where* s^c *is the centroid of* S^*.

Proof. The statement follows from Proposition 8 of Iusem and Monteiro (2000) and Proposition 4. □

Note that, since σ_N is not strictly convex in S^*, problem (33) may have multiple solutions. Therefore our technique does not work in this case, because we cannot characterize the limit point of the path $s(\mu)$ as a solution of problem (33). Even if this were possible, we cannot characterize the limit point of the perturbed dual path $s^E(\mu)$ as solution of the related problem (18), because in this case $r_1 = +\infty$.

6 Appendix

In this appendix we give a proof of some technical results, namely, the Corollary 4 used in Proposition 1 and Proposition 4.

We consider functions $\varphi : \mathbb{R} \to \mathbb{R} \cup \{+\infty\}$ satisfying the following assumptions

h1) The function φ is closed, strictly convex, twice continuously differentiable in $\mathbb{R}_{++}$, and such that

i) $\lim_{t\to 0} \varphi(t) = 0$ or $\lim_{t\to 0} \varphi(t) = +\infty$;

ii) $\lim_{t\to 0} \varphi'(t) = -\infty$.

Lemma 4. *The following statement holds:* $\lim_{u\to-\infty} \varphi^*(u) = -\lim_{t\to 0} \varphi(t)$, *where* φ^* *denotes the conjugate function of* φ.

Proof. Since $\lim_{t\to 0} \varphi'(t) = -\infty$, for all $u \in \mathbb{R}$ there exists $\bar{s} > 0$ such that $\varphi'(\bar{s}) < u$, so that

$$-\bar{s}u \leqslant -\varphi'(\bar{s})\bar{s} = \varphi'(\bar{s})(0 - \bar{s}) \leqslant \varphi(0) - \varphi(\bar{s}) = \lim_{t\to 0} \varphi(t) - \varphi(\bar{s}), \tag{34}$$

using the fact that φ is closed and convex. By (34),

$$-\lim_{t\to 0} \varphi(t) \leqslant \bar{s}u - \varphi(\bar{s}) \leqslant \sup_{s\in\mathbb{R}}\{su - \varphi(s)\} = \varphi^*(u). \tag{35}$$

It follows from (35) that

$$-\lim_{t\to 0} \varphi(t) \leqslant \lim_{u\to-\infty} \varphi^*(u). \tag{36}$$

Let $\bar{t} = \lim_{t\to+\infty} \varphi'(t)$, and take any $u < \min\{0, \bar{t}\}$. Since φ' is continuous and increasing, there exists a unique $s_u > 0$ such that $\varphi'(s_u) = u$, in which case

$$\varphi^*(u) = s_u u - \varphi(s_u) \leqslant -\varphi(s_u). \tag{37}$$

Note that $\lim_{u\to-\infty} s_u = 0$, because $\lim_{t\to 0} \varphi'(t) = -\infty$. Taking limits in (37) as $u \to -\infty$, we obtain

$$\lim_{u\to-\infty} \varphi^*(u) \leqslant -\lim_{u\to-\infty} \varphi(s_u) = -\lim_{t\to 0} \varphi(t). \tag{38}$$

The result follows from (36) and (38). ☐

We remark that it is possible to prove that for all $\gamma \geq 1$, $\liminf_{t\to 0}(-\varphi'(t)/\varphi''(t)^\gamma) = 0$, but we not need of this result here.

Our second assumption on φ is:

h2) Assume that there exist $\gamma \in (0, 1)$ such that

$$r \equiv \lim_{t\to 0} -\frac{\varphi'(t)}{\varphi''(t)^\gamma} \in (0, +\infty). \tag{39}$$

Lemma 5. *The following statements hold:*

i) $\lim_{t\to 0} \varphi(t) = 0$ *when* $\gamma \in (0, 1/2)$ *and* $\lim_{t\to 0} \varphi(t) = +\infty$ *when* $\gamma \in [1/2, 1)$*;*

ii) If $\gamma \in (0, 1) \setminus \{1/2\}$ *then, for all* $\delta \in \operatorname{dom}(\varphi^*)$ *and all* $s > 0$,

$$\lim_{\mu\to 0} \frac{\varphi^*(\delta - s/\mu)}{\mu^{1/\gamma-2}} = \frac{\gamma^2 r^{1/\gamma}}{(1-2\gamma)(1-\gamma)} s^{2-1/\gamma}; \tag{40}$$

iii) If $\gamma = 1/2$ *then, for all* $\delta \in \operatorname{dom}(\varphi^*)$ *and all* $s > 0$,

$$\lim_{\mu\to 0} [\varphi^*(\delta - s/\mu) - \varphi^*(\delta - 1/\mu)] = -r^2 \log s. \tag{41}$$

Proof. We start by proving i). Let $\lambda \equiv \gamma/(1-\gamma) > 0$, $\eta \equiv (1-2\gamma)/(1-\gamma)$. We claim first that

$$\lim_{t\to 0} \frac{-\varphi'(t)}{t^{-\lambda}} = \left(\lambda r^{1/\gamma}\right)^\lambda. \tag{42}$$

We proceed to prove the claim. We compute first $\lim_{t\to 0}(-\varphi'(t))^{-1/\lambda}/t$. Since both the numerator and the denominator converge to 0 as $t \to 0$, using **h1** *ii*), **h2**) and that $\lambda > 0$, we may apply L'Hospital's rule:

$$\begin{aligned}
\lim_{t\to 0+} \frac{(-\varphi'(t))^{-1/\lambda}}{t} = \lim_{t\to 0} \frac{\lambda^{-1}(-\varphi'(t))^{-1/\lambda-1}\varphi''(t)}{1} &= \lambda^{-1} \lim_{t\to 0} \left(-\varphi'(t)\right)^{-1/\gamma} \varphi''(t) \\
&= \lambda^{-1} \lim_{t\to 0} \left(\frac{-\varphi'(t)}{\varphi''(t)^\gamma}\right)^{-1/\gamma} = \lambda^{-1} r^{-1/\gamma}.
\end{aligned} \tag{43}$$

By (43), $\lim_{t\to 0}(-\varphi'(t)/t^{-\lambda}) = \left(\lambda r^{1/\gamma}\right)^\lambda$, establishing the claim.

Let $\nu = (\lambda r^{1/\gamma})^\lambda$. By (42), there exists $\bar{t}$ such that, for all $t \in (0, \bar{t})$, $\nu/2 \leqslant -\varphi'(t)/t^{-\lambda} \leqslant 2\nu$, i.e.,

$$\frac{\nu}{2} t^{-\lambda} \leqslant -\varphi'(t) \leqslant 2\nu t^{-\lambda}. \tag{44}$$

Take t, $u \in (0, \bar{t})$. We consider first the case $\lambda \neq 1$, i.e., $\gamma \neq 1/2$. Integrating (44) between t and u and observing that $1-\lambda = \eta$, we obtain

$$\frac{\nu}{2\eta}\left(u^\eta - t^\eta\right) \leqslant \varphi(t) - \varphi(u) \leqslant \frac{2\nu}{\eta}\left(u^\eta - t^\eta\right). \tag{45}$$

Now we consider separately the cases $\gamma \in (0,1/2)$ and $\gamma \in (1/2,1)$, i.e. $\eta > 0$ and $\eta < 0$ respectively. For $\gamma \in (0,1/2)$, we obtain from (45), for all $t \in (0,\bar{t})$, $\varphi(t) \leqslant \varphi(u) + (2\nu/\eta)\left(u^\eta - t^\eta\right)$, which implies, for any $u \in (0,\bar{t})$, $\lim_{t\to 0}\varphi(t) \leqslant \varphi(u) + (2\nu/\eta)u^\eta < +\infty$, because $\eta > 0$. Hence, by **h1** *i*), $\lim_{t\to 0}\varphi(t) = 0$. For $\gamma \in (1/2,1)$, we obain from (45) $\varphi(t) \geq \varphi(u) + \nu/(2|\eta|)\left(t^\eta - u^\eta\right)$, which implies $\lim_{t\to 0}\varphi(t) = +\infty$, because $\lim_{t\to 0} t^\eta = +\infty$, since $\eta < 0$. Finally, we consider the case of $\lambda = 1$, i.e. $\gamma = 1/2$. In this case, integrating (44), we obtain $\varphi(t) \geq \varphi(u) + (\nu/2)(\log u - \log t)$, and thus $\lim_{t\to 0}\varphi(t) = +\infty$.

ii) Note that for $s > 0$, we have

$$\lim_{\mu\to 0}\varphi^*(\delta - s/\mu) = \lim_{u\to -\infty}\varphi^*(u) = -\lim_{t\to 0}\varphi(t), \tag{46}$$

using Lemma 4. Now, for $\gamma \in (0,1/2)$, it holds that $\lim_{t\to 0}\varphi(t) = 0$, by (i). Thus, in view of (46), the numerator of the left hand side of (40) converges to 0 as $t \to 0$. Also the denominator converges to 0 as $t \to 0$, because $\gamma \in (0,1/2)$. On the other hand, for $\gamma \in (1/2,1)$, we have, by (ii) and (46), that the numerator in the left hand side of (40) converges to $-\infty$ as $t \to 0$, while the denominator converges to $+\infty$. In both cases we can apply L'Hospital's rule for the computation of the limit in the left hand side of (40), obtaining

$$\lim_{\mu\to 0}\frac{\varphi^*(\delta - s/\mu)}{\mu^{1/\gamma-2}} = \lim_{\mu\to 0}\frac{s\mu^{-2}(\varphi^*)'(\delta - s/\mu)}{(1/\gamma-2)\mu^{1/\gamma-3}} = \frac{-\gamma s}{2\gamma-1}\lim_{\mu\to 0}\frac{(\varphi^*)'(\delta - s/\mu)}{\mu^{1/\gamma-1}}. \tag{47}$$

Let now $t = (\varphi^*)'(\delta - s/\mu) = (\varphi')^{-1}(\delta - s/\mu)$, so that $\varphi'(t) = \delta - s/\mu$, i.e. $\mu = s/(\delta - \varphi'(t))$. When $\mu \to 0$, $\delta - \varphi'(t) \to +\infty$, and therefore $\varphi'(t) \to -\infty$, which implies that $t \to 0$. Replacing the variable μ by t, we obtain from (47)

$$\lim_{\mu\to 0}\frac{\varphi^*(\delta - s/\mu)}{\mu^{1/\gamma-2}} = \frac{-\gamma s}{2\gamma-1}\lim_{t\to 0}\frac{t}{\left(s(\delta - \varphi'(t))^{-1}\right)^{1/\gamma-1}} = \frac{-\gamma s^{2-1/\gamma}}{2\gamma-1}\lim_{t\to 0}\frac{t}{\left(\delta - \varphi'(t)\right)^{1-1/\gamma}}. \tag{48}$$

Note that, since $\gamma < 1$, both the numerator and the denominator inside the limit in the rightmost expression of (48) converge to 0 as $t \to 0$, so that we can apply again L'Hospital's rule, obtaining

$$\begin{aligned}\lim_{\mu\to 0}\frac{\varphi^*(\delta - s/\mu)}{\mu^{1/\gamma-2}} &= \frac{\gamma s^{2-1/\gamma}}{2\gamma-1}\lim_{t\to 0}\frac{1}{(1-1/\gamma)\big(\delta - \varphi'(t)\big)^{-1/\gamma}\varphi''(t)} \\ &= \frac{\gamma^2 s^{2-1/\gamma}}{(2\gamma-1)(\gamma-1)}\lim_{t\to 0}\frac{\big(\delta - \varphi'(t)\big)^{1/\gamma}}{\varphi''(t)} = \frac{\gamma^2 s^{2-1/\gamma}}{(2\gamma-1)(\gamma-1)}\lim_{t\to 0}\left(\frac{\delta - \varphi'(t)}{\varphi''(t)^\gamma}\right)^{1/\gamma} \\ &= \frac{\gamma^2 s^{2-1/\gamma}}{(2\gamma-1)(\gamma-1)}\lim_{t\to 0}\left(\frac{-\varphi'(t)}{\varphi''(t)^\gamma}\right)^{1/\gamma} = \frac{\gamma^2 r^{1/\gamma}}{(1-2\gamma)(1-\gamma)}s^{2-1/\gamma},\end{aligned} \tag{49}$$

by using **h2**. Hence (49) establishes (40).

iii) Note that $\gamma = 1/2$ implies $\lambda = 1$, with the notation of (i). Fix $s > 0$ and $\varepsilon \in (0,1)$. By (42), there exists $\bar{t} > 0$ such that, for $t \in (0, \bar{t})$,

$$(1-\varepsilon)r^2 \leqslant -t\varphi'(t) \leqslant (1+\varepsilon)r^2.$$

Thus,

$$\frac{-(1+\varepsilon)r^2}{t} \leqslant \varphi'(t) \leqslant \frac{-(1-\varepsilon)r^2}{t}. \tag{50}$$

Since φ' is increasing and $(\varphi^*)' = (\varphi')^{-1}$, we obtain from (50)

$$(\varphi^*)'\left(\frac{-(1+\varepsilon)r^2}{t}\right) \leqslant t \leqslant (\varphi^*)'\left(\frac{-(1-\varepsilon)r^2}{t}\right). \tag{51}$$

Taking first $u = -(1+\varepsilon)r^2/t$ and then $u = -(1-\varepsilon)r^2/t$, it follows from (51) that

$$\frac{-(1-\varepsilon)r^2}{u} \leqslant (\varphi^*)'(u) \leqslant \frac{-(1+\varepsilon)r^2}{u}, \tag{52}$$

for all $u \leqslant -2r^2\bar{t}^{-1}$, in which case $t \in (0, \bar{t})$ for both choices of u. Take now $v \leqslant w \leqslant -2r^2\bar{t}^{-1}$. Integrating (52) between v and w,

$$-(1-\varepsilon)r^2(\log(-w) - \log(-v)) \leqslant \varphi^*(w) - \varphi^*(v) \leqslant -(1+\varepsilon)r^2(\log(-w) - \log(-v)). \tag{53}$$

Let now $\rho(s,\mu) \equiv \varphi^*(\delta - s/\mu) - \varphi^*(\delta - 1/\mu)$, $\sigma(s) \equiv -r^2 \log s$. Note that for $s = 1$ we have $\rho(1,\mu) = 0$ for all $\mu > 0$, so that

$$\lim_{\mu\to 0} \rho(1,\mu) = 0 = \sigma(1). \tag{54}$$

For $s \in (0,1)$, take $v = \delta - 1/\mu$, $w = \delta - s/\mu$, $\bar{\mu} = s/(2r^2\bar{t}^{-1} + |\delta|)$. Then for all $\mu \in (0, \bar{\mu})$ it holds that $v \leqslant w \leqslant -2r^2\bar{t}^{-1}$. With these values of v and w, (53) becomes

$$-(1-\varepsilon)r^2 \log\left(\frac{s-\delta\mu}{1-\delta\mu}\right) \leqslant \rho(s,\mu) \leqslant -(1+\varepsilon)r^2 \log\left(\frac{s-\delta\mu}{1-\delta\mu}\right), \tag{55}$$

for all $\mu \in (0, \bar{\mu})$. Taking limits in (55) as $\mu \to 0$,

$$(1-\varepsilon)\sigma(s) \leqslant \text{liminf}_{\mu\to 0}\rho(s,\mu) \leqslant \text{limsup}_{\mu\to 0}\rho(s,\mu) \leqslant (1+\varepsilon)\sigma(s). \tag{56}$$

Since (56) holds for all $\varepsilon \in (0,1)$, we take limits as $\varepsilon \to 0$ in (56) and obtain

$$\lim_{\mu\to 0} \rho(s,\mu) = \sigma(s) \quad \forall s \in (0,1). \tag{57}$$

For $s > 1$, we take $v = \delta - s/\mu$, $w = \delta - 1/\mu$, $\hat{\mu} = (2r^2\bar{t}^{-1} + |\delta|)^{-1}$ and we have again $v \leqslant w \leqslant -2r^2\bar{t}^{-1}$ so that (53) holds, but now $\varphi^*(w) - \varphi^*(v) = -\rho(s,\mu)$, in which case the inequalities in (56) reverse, and after taking limits as $\mu \to 0$, we obtain

$$(1-\varepsilon)\sigma(s) \geq \text{limsup}_{\mu\to 0}\rho(s,\mu) \geq \text{liminf}_{\mu\to 0}\rho(s,\mu) \geq (1+\varepsilon)\sigma(s). \tag{58}$$

Then, taking limits in (58) as $\varepsilon \to 0$, we conclude that

$$\lim_{\mu\to 0}\rho(s,\mu) = \sigma(s) \quad \forall s > 1. \tag{59}$$

By (54), (57) and (59), $\lim_{\mu\to 0}\rho(s,\mu) = \sigma(s)$ for all $s > 0$, which is precisely (41). □

Corollary 4. *Take $\varphi : \mathbb{R}^n \to \mathbb{R}\cup\{+\infty\}$ of the form $\varphi(x) = \sum_{j=1}^n \varphi_j(x_j)$, with $\varphi_j : \mathbb{R} \to \mathbb{R}\cup\{+\infty\}$. Assume that the φ_j's satisfy the general hypotheses* **h1** *and* **h2** *with the same constant γ for the correpondent r_j's . Then:*

i) If $\gamma \in (0,1)\setminus\{1/2\}$ then, for all $\delta \in \text{dom}(\varphi^*) \subset \mathbb{R}^n$ and all $s \in \mathbb{R}^n_{++}$,

$$\lim_{\mu\to 0}\frac{\varphi^*(\delta - s/\mu)}{\mu^{1/\gamma-2}} = \frac{\gamma^2}{(1-2\gamma)(1-\gamma)}\sum_{j=1}^n r_j^{1/\gamma}s_j^{2-1/\gamma}. \tag{60}$$

ii) If $\gamma = 1/2$ then, for all $\delta \in \text{dom}(\varphi^*) \subset \mathbb{R}^n$ and all $s \in \mathbb{R}^n_{++}$,

$$\lim_{\mu\to 0}\left(\varphi^*(\delta - s/\mu) - \varphi^*(\delta - 1/\mu)\right) = -\sum_{j=1}^n r_j^2 \log s_j. \tag{61}$$

Proof. By separability of φ, we obtain that $\varphi^* = \sum_{j=1}^n \varphi_j^*$, and thus (i) and (ii) follow immediately from items (ii) and (iii) of Lemma 5. □

The result of Lemma 5 (ii) also hold when $r = 0$ or $r = +\infty$, but they become rather irrelevant, because only when $0 < r < +\infty$ the right hand side of (40) is strictly convex functions of s (otherwise they are identically 0 or $+\infty$). We observe that the condition $0 < r < +\infty$ implies $\gamma \in (0,1)$, because for $\gamma = 1$, either r vanishes or it does not exist, while for $\gamma = 0$ it holds that $r = \lim_{t\to 0}(-\varphi'(t)) = +\infty$. We remark, however, that it may happen that $r = 0$ or $r = +\infty$ even when $\gamma \in (0,1)$.

Lemma 6. *Let $\xi \equiv \lim_{t\to 0} t\varphi''(t)$. If ξ exists and belongs to $(0,+\infty)$ then $\gamma = 0$. If, additionally, $\lim_{t\to 0}\varphi(t) = 0$, then, for all $\delta \in \text{dom}(\varphi^*)$ and all $s > 0$,*

$$\lim_{\mu\to 0}\left[\varphi^*(\delta - s/\mu)^{\mu}\right] = e^{-s/\xi(\varphi)}. \tag{62}$$

Proof. Since $\lim_{t\to 0} t\varphi''(t) = \xi$, there exists $\bar{t} < 1$, such that, for all $t \in (0, \bar{t})$,

$$\frac{\xi}{2t} \leqslant \varphi''(t) \leqslant \frac{2\xi}{t}. \tag{63}$$

Take $t \in (0, \bar{t})$. Integrating (63) between t and $\bar{t}$,

$$(\xi/2)\log(\bar{t}/t) \leqslant \varphi'(\bar{t}) - \varphi'(t) \leqslant 2\xi \log(\bar{t}/t). \tag{64}$$

From (63) and (64) it follows that, for $t \in (0, \bar{t})$,

$$\begin{aligned}(2\xi)^{-\gamma} t^{\gamma} \left(-\varphi'(\bar{t}) + (\xi/2)\log \bar{t} - (\xi/2)\log t\right) &\leqslant \frac{-\varphi'(t)}{\varphi''(t)^{\gamma}} \\ &\leqslant (\xi/2)^{-\gamma} t^{\gamma} \left(-\varphi'(\bar{t}) + 2\xi \log \bar{t} - 2\xi \log t\right). \end{aligned} \tag{65}$$

Taking limits in (65) as $t \to 0$, and remembering that, for all $\gamma > 0$,

$$\lim_{t\to 0} t^{\gamma} \log t = \lim_{t\to 0} \gamma^{-1} t^{\gamma} \log t^{\gamma} = \gamma^{-1} \lim_{u\to 0} u \log u = 0,$$

we obtain $\lim_{t\to 0}(-\varphi'(t)/\varphi''(t)) = 0$ for all $\gamma > 0$, i.e. $A_{\varphi} = \{0\}$, implying that $\gamma = 0$.

We proceed to prove (62). Let

$$\psi(\mu) = \log\left(\varphi^*(\delta - s/\mu)^{\mu}\right) = \mu \log \varphi^*(\delta - s/\mu). \tag{66}$$

Then

$$\lim_{\mu\to 0} \psi(\mu) = \lim_{\mu\to 0} \frac{\log \varphi^*(\delta - s/\mu)}{\mu^{-1}}. \tag{67}$$

From Lemma 4 and the assumption that $\lim_{t\to 0} \varphi(t) = 0$, it follows that

$$\lim_{\mu\to 0} \varphi^*(\delta - s/\mu) = \lim_{u\to -\infty} \varphi^*(u) = \lim_{t\to 0} \varphi(t) = 0,$$

so that $\lim_{\mu\to 0} \log \varphi^*(\delta - s/\mu) = -\infty$. Since $\lim_{\mu\to 0} \mu^{-1} = +\infty$, we can apply L'Hospital's rule to compute the limit in (67), obtaining

$$\begin{aligned}\lim_{\mu\to 0} \psi(\mu) &= \lim_{\mu\to 0} \frac{(s/\mu^2)\varphi^*(\delta - s/\mu)^{-1}(\varphi^*)'(\delta - s/\mu)}{-\mu^{-2}} \\ &= -s \lim_{\mu\to 0} \left(\frac{(\varphi^*)'(\delta - s/\mu)}{\varphi^*(\delta - s/\mu)}\right) = -s \lim_{t\to 0} \frac{t}{(\varphi^*)(\varphi'(t))}, \end{aligned} \tag{68}$$

with the change of variables $t = (\varphi^*)'(\delta - s/\mu)$, already used in (ii) and (iii). Note that the (sufficient) optimality condition of $\max_{s\in\mathbb{R}}\{s\varphi'(t) - \varphi(s)\}$ is $\varphi'(t) = \varphi'(s)$, satisfied only by $s = t$, because φ' is strictly increasing. Thus, $\varphi^*(\varphi'(t)) = t\varphi'(t) - \varphi(t)$, and we obtain from (68)

$$\lim_{\mu\to 0} \psi(\mu) = -s \lim_{t\to 0} \frac{t}{t\varphi'(t) - \varphi(t)}. \tag{69}$$

Multiplying throughout (64) by t, and taking limits as $t \to 0$, we obtain that $\lim_{t\to 0} t\varphi'(t) = 0$, and thus both the numerator and the denominator in the right hand side of (69) converge to 0 as $t \to 0$, allowing us to apply L'Hospital's rule to (69), which gives

$$\lim_{\mu\to 0} \psi(t) = -s \lim_{t\to 0} \frac{1}{t\varphi''(t)} = \frac{-s}{\xi}. \tag{70}$$

By (70) and (66), we obtain $\lim_{\mu\to 0} \varphi^*(\delta - s/\mu)^\mu = e^{-s/\xi}$, establishing (62). ☐

Proof of the Proposition 4. We cannot apply directly Lemma 6 because powers do not distribute with sums. Let $I = \operatorname{argmin}\{s_j : 1 \leqslant j \leqslant n\}$, $L = \operatorname{argmax}\{\delta_i : i \in I\}$. Take $\ell \in L$. We claim that for μ close enough to 0, $\delta_\ell - s_\ell/\mu \geq \delta_j - s_j/\mu$ for all $j \in \{1, \dots, n\}$. This is certainly true for $j \in I$, by definition of L, because all the s_j's with $j \in I$ have the same value. For $j \notin I$, we have $s_\ell < s_j$, and so the results holds if $\delta_\ell = \delta_j$. Otherwise it suffices to take $\mu \leqslant (s_j - s_\ell)/|\delta_j - \delta_\ell|$. Since $(\varphi^*)' = (\varphi')^{-1}$, we have that $\operatorname{Im}[(\bar\varphi^*)'] = \operatorname{dom}(\bar\varphi') = \mathbb{R}_{++}$. Thus, $(\bar\varphi^*)'(u) > 0$ for all u, i.e. $\bar\varphi^*$ is increasing. It follows that

$$\bar\varphi^*(\delta_\ell - s_\ell/\mu) \geq \bar\varphi^*(\delta_j - s_j/\mu) \tag{71}$$

for all $j \in \{1, \dots, n\}$ and small enough μ. Then

$$\left(\sum_{j=1}^n \bar\varphi^*(\delta_j - s_j/\mu)\right)^\mu = (\bar\varphi^*(\delta_\ell - s_\ell/\mu))^\mu \left(\sum_{j=1}^m \frac{\bar\varphi^*(\delta_j - s_j/\mu)}{\bar\varphi^*(\delta_\ell - s_\ell/\mu)}\right)^\mu. \tag{72}$$

The first factor in the rightmost expression of (72) converges to $\exp(-s_\ell/\xi)$ by Lemma 6. We look now at the summation in the second factor, which we will denote by $S(\mu)$. All terms are positive, the ℓ-th one is 1, and all the others belong to $(0, 1)$ by (71). Thus $1 \leqslant S(\mu) \leqslant n$, and therefore $\lim_{\mu\to 0} S(\mu)^\mu = 1$. It follows that

$$\lim_{\mu\to 0} \left(\sum_{j=1}^n \bar\varphi^*(\delta_j - s_j/\mu)\right)^\mu = e^{-s_\ell/\xi} = \max\left\{e^{-s_j/\xi} : j \in \{1, \dots n\}\right\}$$

because $s_\ell = \min\{s_1, \dots, s_n\}$ by definition of I. ☐

References

[1] Adler, I., Monteiro, R.D.C. (1991). Limiting behavior of the affine scaling continuous trajectories for linear programming problems, *Mathematical Programming* **50** 29–51.

[2] Censor, Y., Zenios, S.A. (1992). Proximal minimization algorithm with D-functions, *Journal of Optimization Theory and Applications* **73** 451-464.

[3] Chen, G., Teboulle, M. (1993). Convergence analysis of a proximal-like optimization algorithm using Bregman functions, *SIAM Journal on Optimization* **3** 538–543.

[4] Comonetti, R., San Martín, J. (1994), Asymptotic Analysis of the exponential penalty trajectory in linear programming, *Mathematical Programming* **67** pp. 169–187.

[5] Eggermont, P.P.B. (1990) Multiplicative iterative algorithms for convex programming, *Linear Algebra and Its Applications* **130** 25-42.

[6] Eriksson, J. An iterative primal-dual algorithm for linear programming. Technical Report 85-10, Department of Mathematics, Linkoping University, Linkoping, Sweden (1985).

[7] Hoffman, A.J. (1952). On approximate solutions of systems of linear inequalities. *Journal of Research of the National Bureau of Standards* **49** 263–265.

[8] Iusem, A.N. (1995). On some properties of generalized proximal point methods for quadratic and linear programming. *Journal of Optimization Theory and Applications* **85** 593-612.

[9] Iusem, A.N., Monteiro, R.D.C. (2000). On dual convergence of the generalized proximal point method with Bregman distance. *Mathematics of Operations Research* **25** 606–624.

[10] Iusem, A.N., Svaiter, B.F., Cruz Neto, J.X. (1999). Generalized proximal point methods and Cauchy trajectories in Riemannian manifolds. *SIAM Journal on Control and Optimization* **37** 566–588.

[11] Iusem, A.N., Svaiter, B.F., Teboulle, M. (1994). Entropy like proximal methods in convex programming. *Mathematics of Operations Research* **19** 790-814.

[12] Iusem, A.N., Teboulle, M. (1993). On the convergence rate of entropic proximal optimization algorithms. *Computational and Applied Mathematics* **12** 153-168.

[13] Iusem, A.N., Teboulle, M. (1995). Convergence rate analysis of nonquadratic proximal methods for convex and linear programming. *Mathematics of Operations Research* **20** 657–677.

[14] Jensen, D.L., Polyak, R.A. (1994). The convergence of a modified barrier method for convex programming. *IBM Journal of Research and Development* **38** 307–321.

[15] Kiwiel, K.C. (1997). Proximal minimization methods with generalized Bregman functions. *SIAM Journal on Control and Optimization* **35** 1142–1168.

[16] Polyak, R., Teboulle, M. (1997). Nonlinear rescaling and proximal-like methods in convex optimization. *Mathematical Programming* **76** 265–284.

[17] Powell, M.J.D. (1995). Some convergence properties of the modified log barrier method for linear programming. *SIAM Journal on Optimization* **5** 695–739.

[18] Rockafellar, R.T. (1976). Monotone operators and the proximal point algorithm. *SIAM Journal on Control and Optimization* **14** 877–898.

[19] Tseng, P., Bertsekas, D. (1993). On the convergence of the exponential multiplier method for convex programming. *Mathematical Programming* **60** 1–19.