Optimization Methods and Software
Vol. 22, No. 2, April 2007, 339–360

Dual convergence of the proximal point method with Bregman distances for linear programming

J. X. CRUZ NETO†, O. P. FERREIRA‡, A. N. IUSEM§ and R. D. C. MONTEIRO*¶

†DM, Universidade Federal do Piauí, Teresina, PI 64049-500, Brazil
‡IME, Universidade Federal de Goiás, Goiânia, GO 74001-970, Brazil
§IMPA, Estrada Dona Castorina 110, Rio de Janeiro, RJ, 22460-320, Brazil
¶School of ISyE, Georgia Tech, Atlanta, Georgia 30332, USA

(Received 19 October 2004; revised 25 October 2005; in final form 6 January 2006)

In this article, 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.

1. Introduction

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

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

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

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

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 non-negative orthant $\mathbb{R}^n_+$ (see (7) for the definition of D_φ). The optimality condition for (2) determines the dual

*Corresponding author. Email: orizon@mat.ufg.br

Optimization Methods and Software
ISSN 1055-6788 print/ISSN 1029-4937 online
http://www.tandf.co.uk/journals
DOI: 10.1080/10556780600565083

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 [1], which has the form of (2) for $\varphi(x) = \|x\|^2$ (note that this φ is not a barrier for the non-negative orthant). Particular cases, corresponding to a special form of φ, were introduced in Eriksson [2], Eggermont [3] and Tseng and Bertsekas [4]. General Bregman distances were studied in several papers, for example, Censor and Zenios [5], Chen and Teboulle [6], Iusem [7] and Kiwiel [8]. Similar methods, using φ-divergences instead of Bregman distances in (2), appear in Iusem and Teboulle [9], Jensen and Polyak [10], Iusem and Teboulle [11], Powell [12] and Polyak and Teboulle [13] (see Iusem, Svaiter and Teboulle [14] 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 article is to prove the convergence of this sequence for linear programming with Bregman distance D_φ, where φ satisfies an appropriate condition, which holds, for example 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 [4], Powell [12], Jensen and Polyak [10] and Polyak and Teboulle [13]. Most of these results are described in a somewhat different framework, for example, with φ-divergences instead of Bregman distances in (2). It is worthwhile to mention that, up to additive and/or multiplicative constants, the only φ-divergence which is also a Bregman distance corresponds to the case of the entropic barrier, given by $\varphi(x) = \sum_{j=1}^n \phi(x_j)$ with $\phi(t) = t\log t - t + 1$. 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 [15]. In this article, 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 [10], where it is proved that some cluster point of $\{\bar{s}^k\}$ is a dual optimal solution. This result was improved upon in Polyak and Teboulle [13], where it is proved that all cluster points of $\{\bar{s}^k\}$ are dual optimal solutions. In this article, it is also proved that dual functional values converge linearly. The convergence of the whole sequence $\{\bar{s}^k\}$ appeared for the first time in Powell [12], but only for linear programming with the shifted logarithmic barrier. None of these articles present results on the convergence of the whole sequence $\{s^k\}$. For linear programming with a certain non-degeneracy assumption, which implies uniqueness of the dual solution, and with φ-divergences instead of Bregman distances, it has been proved in Iusem and Teboulle [11] that the sequence $\{s^k\}$ converges to the dual solution. A new class of barriers, called second order φ-divergences, was introduced in Auslender, Teboulle and Ben-Tiba [16]. They proved quadratic convergence of the dual sequence generated by the proximal point method with these barriers applied to linear programming. Quadratic convergence of the sequence of dual functional values with a related class of barriers was established in Polyak [17].

In this article, 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 $\mu > 0$:

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

We also study the limiting behavior of an associated dual path $s(\mu)$ defined in (9) below. More specifically, our main goal is to 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 [18]. 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.

At this point, it is important to emphasize that, though our convergence analysis of the dual sequence is limited to the case of linear programming, by no means we advocate the use of the proximal point method for solving linear programs, be it with the classical quadratic regularization or with Bregman functions. The method is intended rather for the non-linear problem (1), and if we restrict our analysis to the case of a linear f, it is just because our analytical tools do not allow us to go further. We expect that the results of this article will be the first step towards a convergence analysis of the dual sequence in the non-linear case.

This said, it is worthwhile to make some comments on the proximal method with Bregman barriers viz-a-viz the same method with the classical quadratic regularization, and also interior point methods for convex programming.

The use of barriers (which force the generated sequence to remain in the interior of their domains) in the proximal method, instead of the classical regularization given by $\varphi(x) = \|x\|^2$, has the following purpose. The proximal method, in principle, is just a regularization device, which replaces a possibly ill-conditioned problem by a sequence of better conditioned subproblems of the same nature. Thus, if the original problem has positivity constraints, so do the subproblems. If we replace the quadratic function given earlier by a function defined only in a certain region, and such that its gradient diverges in the boundary of this region, then it will serve not only as a regularization device but also as a penalization one, forcing the generated sequence to remain in the interior of the region. In such a situation, constraints demanding that the solution must belong to this region can be omitted from the subproblems. Thus, a constrained problem is solved through a sequence of unconstrained subproblems, or, as in the case of interest here, a problem with inequality constraints reduces to a sequence of subproblems whose constraints are just linear equalities. This procedure opens the possibility of solving the subproblems with faster methods (e.g. Newton's type of methods). In our case, with linear constraints, the elimination of the positivity constraints from the subproblems eradicates also the combinatorial component of the problem, with all the complications resulting from the need to identify the optimal face of the feasible polyhedral region. Another advantage of the use of barriers appears in a very important application of the proximal method: when it is applied to the dual of a constrained convex optimization problem, the proximal point method gives rise to primal-dual methods, called augmented Lagrangian algorithms, whose subproblems are always unconstrained, but with objective functions, which, with the classical quadratic regularization, are differentiable but never twice differentiable. The subproblems of the augmented Lagrangian methods resulting from proximal method with Bregman barriers have objective functions which are as smooth as the original constraints, and which can be minimized with Newton's method.

If we compare now the proximal method with Bregman barriers for linearly constrained convex optimization with the so called *interior point* algorithms, we observe that both share the same feature discussed earlier: unconstrained subproblems (after dealing in an appropriate way with the linear system $Ax = b$) which can be solved with fast second-order methods. The difference lies in the fact that the specific logarithmic barrier, typical of interior point methods, has a property, namely self-concordance, which allows estimates of the number of iterations needed to achieve a given accuracy, ensuring that the running time of the algorithm is bounded by a polynomial function of an adequate measure of the size of the problem. This feature is not shared by most Bregman barriers considered here, which in general are not self-concordant. Nevertheless, the proximal point method with Bregman functions, and the resulting smooth augmented Lagrangians, have proved to be efficient tools in several specific instances, thus justifying the study of its convergence properties, as has been done in many articles mentioned earlier.

The organization of our article 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 section 5 we make some remarks. We conclude this article with an Appendix which contains the proofs of some technical results.

1.1 *Notation*

We will use the following notation throughout this article. $\mathbb{R}^n$ 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. 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 be denoted 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, that is, $X = \operatorname{diag}(x_1, \dots, x_n)$. For a matrix A, we let A^{T} denote its transpose, $\Im A$ denote the subspace generated by the columns of A and 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 with 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 article and state its basic properties.

We consider the linear programming problem

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

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

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

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

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

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

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 non-negative orthant $\mathbb{R}^n_+$, that is,

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

satisfying certain assumptions described below. The first assumption we make on φ is the following:
(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 $j \in \{1, \ldots, n\}$;
(ii) $\lim_{t\to 0} \varphi_j'(t) = -\infty$ for each $j \in \{1, \ldots, n\}$.

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 $j \in \{1, \ldots, n\}$.

Our second assumption on φ is:
(H2) There exist $\gamma \in (0, 1)$ such that

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

From now on, we consider barrier functions φ satisfying assumptions H1 and H2. We present several examples next, with the corresponding values of γ and r_j. In item (ii) we have used a special definition of φ_j for values of t far from 0, so that $\operatorname{dom}(\varphi_j)$ 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 2.1 For each $j \in \{1, \ldots, n\}$, 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)

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

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 $r_j = [\alpha\,(\alpha+1)^{-(\alpha+1)}]^{1/(\alpha+2)}$.

Finally, we make a last hypothesis on φ. It will be used only in subsection 3.2, whereas hypotheses H1, H2 are used throughout the article.
(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 $j \in \{1, \ldots, n\}$.
We next present some examples of functions φ satisfying hypotheses H1, H2 and H3, with the corresponding values of ν.

Example 2.2 For each $j \in \{1, \ldots, n\}$ we take:

(i) $\varphi_j(t)$ as in Example 2.1 (i). Then, $\nu = 1/(1-\beta)$;
(ii) $\varphi_j(t)$ as in Example 2.1 (iii). Then, $\nu = 1$;
(iii) $\varphi_j(t)$ as in Example 2.1 (v). Then, $\nu = 1/(1+\alpha)$.

We remark that φ as defined in Examples 2.1 (ii) and 2.1 (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

$$D_\varphi(x, y) \equiv \varphi(x) - \varphi(y) - \nabla\varphi(y)^{\mathrm{T}}(x - y), \quad \forall (x, y) \in \mathbb{R}^n \times \mathbb{R}^n_{++}. \tag{7}$$

Observe that $D_\varphi(x, y) = +\infty$ for every $x \notin \mathbb{R}^n_+$ and $y \in \mathbb{R}^n_{++}$. Now, let $x^1 = (x^1_1, \ldots, x^1_n) \in \mathcal{F}^0$ 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

$$x(\mu) \equiv \arg\min\{c^{\mathrm{T}}x + \mu D_\varphi(x, x^1) : Ax = b\}, \quad \forall \mu > 0, \tag{8}$$

and

$$s(\mu) \equiv -\mu(\nabla\varphi(x(\mu)) - \nabla\varphi(x^1)), \quad \forall \mu > 0. \tag{9}$$

It is well-known that, for every $\mu > 0$, problem (8) has a unique optimal solution which is strictly positive, see Proposition 2 of Iusem, Svaiter and Cruz Neto [19]. This clearly implies that both $x(\mu)$ and $s(\mu)$ are well defined for every $\mu > 0$. The curves $x(\mu)$ and $s(\mu)$ are differentiable in $(0, \infty)$ as an elementary consequence of the implicit function theorem. Moreover, the optimality conditions for (8) imply that

$$s(\mu) \in c + \operatorname{Im} A^{\mathrm{T}}, \quad \forall \mu > 0. \tag{10}$$

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

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

The following result characterizes the limiting behavior of the primal and dual central paths (8) and (9).

PROPOSITION 2.3 *The following statements hold*:

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

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

$$\min\{\sigma_N(s_N) : s \in c + \operatorname{Im} A^{\mathrm{T}}, \quad s_B = 0\}, \tag{11}$$

and σ_N is the strictly convex function defined as

$$\sigma_N(s_N) = \begin{cases} \dfrac{\gamma^2}{(1-2\gamma)(1-\gamma)} \displaystyle\sum_{j\in N} r_j^{1/\gamma} s_j^{2-1/\gamma}, & \textit{if } \gamma \in (0,1)\setminus\{1/2\}; \\ -\displaystyle\sum_{j\in N} r_j^2 \log s_j, & \textit{if } \gamma = 1/2. \end{cases} \tag{12}$$

Proof The statement related to the primal central path was proved in Theorem 1 of Iusem, Svaiter and Cruz Neto [19]. The statement related to the dual central path follows from Proposition 7 in Iusem and Monteiro [15] and the fact that items (i) and (ii) of Corollary A3 in the Appendix imply that

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

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

$$g_B(\mu) \equiv \nabla^2\varphi_B(x_B(\mu))e, \qquad h(\mu) \equiv \mu^{-1/\gamma}\left[\nabla^2\varphi(x(\mu))\right]^{-1} e, \tag{13}$$

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

Corollary 2.4 *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* $h_N^* \equiv (r_N(s_N^*)^{-1})^{1/\gamma}$;

(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 2.3 and (9), we have that $s^* = \lim_{\mu\to 0} -\mu\nabla\varphi(x(\mu))$. This, together with (13), the fact that $s_B^* = 0$ and hypothesis H2, implies (i). Statement (ii) follows from the twice continuous differentiability 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 first study the behavior of the derivatives of the dual and primal paths. The motivation is that the primal (respectively, average dual) proximal sequence is contained in the primal (respectively, dual) path corresponding, see Proposition 4.2.

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

$$s^E(\mu) \equiv s(\mu) - \mu\dot{s}(\mu), \quad \forall \mu > 0. \tag{14}$$

Differentiating (9) and using (14), we easily see that

$$s^E(\mu) = \mu^2 \nabla^2 \varphi(x(\mu))\dot{x}(\mu), \quad \forall \mu > 0. \tag{15}$$

In view of (13), observing that $\nabla^2\varphi(x(\mu))$ is a diagonal matrix, (15) is equivalent to

$$h(\mu)s^E(\mu) = \mu^{2-1/\gamma}\,\dot{x}(\mu), \quad \forall \mu > 0. \tag{16}$$

LEMMA 3.1 $\lim_{\mu\to 0} s^E(\mu) = \bar{s}^E$, *where $\bar{s}^E$ is the unique optimal solution of the problem*

$$\min\left\{\frac{1}{2}\left\|(h_N^*)^{1/2}s_N\right\|^2 : s \in c + \operatorname{Im} A^{\mathrm{T}}, \quad s_B = 0\right\}. \tag{17}$$

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 (10), we have that $\dot{s}(\mu) \in \operatorname{Im} A^{\mathrm{T}}$ for all $\mu > 0$. This fact together, with (10) and (14), implies that $s^E(\mu) \in c + \operatorname{Im} A^{\mathrm{T}}$ for all $\mu > 0$. Fix some $\tilde{s} \in S^*$ and note that $\tilde{s} \in c + \operatorname{Im} A^{\mathrm{T}}$ and $\bar{s}_B = 0$. It then follows that

$$s^E(\mu) - \tilde{s} \in \operatorname{Im} A^{\mathrm{T}}, \quad \forall \mu > 0. \tag{18}$$

On the other hand, by (16) 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 fact, together with (18), implies that $(s^E(\mu) - \tilde{s})^T h(\mu)s^E(\mu) = 0$, and hence

$$\begin{aligned}\|h(\mu)^{1/2}s^E(\mu)\|^2 &= s^E(\mu)^{\mathrm{T}}h(\mu)s^E(\mu) = \tilde{s}^{\mathrm{T}}h(\mu)s^E(\mu)\\ &\leqslant \|h(\mu)^{1/2}\tilde{s}\|\,\|h(\mu)^{1/2}s^E(\mu)\|,\end{aligned}$$

which in turn yields

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

where the last equality is due to the fact that $\tilde{s}_B = 0$. By Corollary 2.4, 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, implies 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 (17), from which the result follows. Clearly, $\bar{s}$ is feasible for (17). Moreover, by (16) and the fact that $A\dot{x}(\mu) = 0$, we conclude that $A_N(h_N(\mu)s_N^E(\mu)) \in \operatorname{Im} A_B$ for all $\mu > 0$. This equation, together with Corollary 2.4, implies that $A_N(h_N^*\bar{s}_N) \in \operatorname{Im} A_B$. We have thus proved that $\bar{s}$ satisfies the optimality condition for (17). ■

THEOREM 3.2 *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 (17). Since s^* is the optimal solution of (11), it is feasible for (17) and satisfies $A_N(\nabla\sigma_N(s_N^*)) \in \operatorname{Im} A_B$. Using (12), we easily see that for any $\gamma \in (0, 1)$:

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

These two observations imply that $A_N(h_N^* s_N^*) \in \operatorname{Im} A_B$, and hence that s^* satisfies the optimality condition for (17).

Statement (ii) follows by noting that statement (i), relation (14) and Proposition 2.3(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 ref. 20].

LEMMA 3.3 *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 $(\tilde{u}_J, u_{\bar{J}})^{\mathrm{T}} \in E$ and $\|\tilde{u}_J\| \leqslant M\|u_{\bar{J}}\|$.*

THEOREM 3.4 $\lim_{\mu\to 0} \mu^{2-1/\gamma}\dot{x}(\mu) = d^\infty$, *where d^∞ is the unique optimal solution of the problem*

$$\min\left\{\frac{1}{2}\left\|(g_B^*)^{1/2} d\right\|^2 : d \in \operatorname{Null} A, \quad d_N = h_N^* s_N^*\right\}. \tag{19}$$

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 (16), Corollary 2.4 and Theorem 3.2, we have

$$\lim_{\mu\to 0} d_N(\mu) = \lim_{\mu\to 0} \mu^{2-1/\gamma}\dot{x}_N(\mu) = \lim_{\mu\to 0} h_N(\mu) s_N^E(\mu) = h_N^* s_N^* > 0. \tag{20}$$

Clearly, $Ad(\mu) = A_B d_B(\mu) + A_N d_N(\mu) = 0$ for all $\mu > 0$. Applying Lemma 3.3 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 fact, together with (20), implies 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 (15), (18) and the fact that $\bar{s}_B = 0$, we conclude that $g_B(\mu)d_B(\mu) \in \operatorname{Im}(A_B^{\mathrm{T}})$ for all $\mu > 0$. The two last conclusions imply that $(p_B(\mu) - d_B(\mu))^{\mathrm{T}} g_B(\mu) d_B(\mu) = 0$ for all $\mu > 0$. An argument similar to the one used in the proof of Lemma 3.1 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 2.4(ii) and the fact that $\{p_B(\mu) : \mu \in (0, 1]\}$ is bounded, implies 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, (20) and the fact that $Ad(\mu) = 0$ for all $\mu > 0$ imply that $\bar{d}$ is feasible for (19). Moreover, the fact that $g_B(\mu)d_B(\mu) \in \mathrm{Im}(A_B^{\mathrm{T}})$ implies that $g_B^* \bar{d}_B \in \mathrm{Im}(A_B^{\mathrm{T}})$. We have thus shown that $\bar{d}$ satisfy the optimality condition for (19), and hence that $\bar{d} = d^\infty$. As this holds for any accumulation point $\bar{d}$ of $\{d(\mu) : \mu \in (0, 1]\}$, the result follows. ■

COROLLARY 3.5 *The following limit holds*:

$$\lim_{\mu \to 0} \frac{x(\mu) - x^*}{\mu^{1/\gamma - 1}} = \frac{d^\infty}{1/\gamma - 1} \neq 0, \quad \lim_{\mu \to 0} \frac{c^{\mathrm{T}}(x(\mu) - x^*)}{\mu^{1/\gamma - 1}} = \frac{\|(h_N^*)^{1/2} s_N^*\|^2}{1/\gamma - 1} > 0. \tag{21}$$

Proof We will only prove that the second limit holds, as the first one one can be proved in a similar way. Using the fact that $c - s^* \in \mathrm{Im}\, A^{\mathrm{T}}$, $x(\mu) - x^* \in \mathrm{Null}(A)$, $x_N^* = 0$ and $s_B^* = 0$, we easily see that $c^{\mathrm{T}}(x(\mu) - x^*) = (s_N^*)^{\mathrm{T}} x_N(\mu)$ for all $\mu > 0$. Using this identity together with L'Hospital's rule of calculus and relation (20), we obtain

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

■

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

$$v(\mu) \equiv \big(\nabla\varphi(x(\mu)) - \nabla\varphi(x^1)\big) + \nu \nabla^2 \varphi(x(\mu)) x(\mu), \quad \forall \mu > 0. \tag{22}$$

LEMMA 3.6 *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) $A_N[h_N(\mu)(\dot{s}_N(\mu) + v_N(\mu))] \in \mathrm{Im}\, A_B$ *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, differentiating (9) and using (22), we obtain for all $\mu > 0$ that

$$\begin{aligned} \dot{s}(\mu) &= -(\nabla\varphi(x(\mu)) - \nabla\varphi(x^1)) - \mu \nabla^2 \varphi(x(\mu)) \dot{x}(\mu) \\ &= -v(\mu) + \nabla^2 \varphi(x(\mu))[\nu x(\mu) - \mu \dot{x}(\mu)]. \end{aligned} \tag{23}$$

Statement (ii) now follows by rearranging this expression and using (13). Using the fact that $Ax(\mu) = b$, $A\dot{x}(\mu) = 0$ and $b \in \mathrm{Im}\, A_B$, we easily see that $A_N(\nu x_N(\mu) - \mu \dot{x}_N(\mu)) \in \mathrm{Im}\, A_B$ for all $\mu > 0$. Statement (iii) now follows from this conclusion, in view of statement (ii). ■

THEOREM 3.7 *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*: $\dot{s}_B^{\infty} = -(\nabla\varphi_B(x_B^*) - \nabla\varphi_B(x_B^1))$ *and* $\dot{s}_N^{\infty}$ *is the unique optimal solution of the problem*

$$\min\left\{\frac{1}{2}\left\|(h_N^*)^{1/2}(p_N + v_N^*)\right\|^2 : (\dot{s}_B{}^{\infty} p_N) \in \operatorname{Im} A^{\mathrm{T}}\right\}. \tag{24}$$

Proof By Theorem 3.4 and the fact that $\gamma \in (0, 1)$, we have that $\lim_{\mu\to 0} \mu\dot{x}(\mu) = 0$. Hence, by (23) and the fact that $\lim_{\mu\to 0} x_B(\mu) = x_B^* > 0$, we conclude that $\lim_{\mu\to 0} \dot{s}_B(\mu) = -(\nabla\varphi_B(x_B^*) - \nabla\varphi_B(x_B^1))$. We will now show that the set $\{\dot{s}_N(\mu) : \mu \in (0, 1]\}$ is bounded. Indeed, it follows from (10) that $\dot{s}(\mu) = (\dot{s}_B(\mu), \dot{s}_N(\mu))^{\mathrm{T}} \in \operatorname{Im} A^{\mathrm{T}}$ for all $\mu > 0$. Applying Lemma 3.3 with $E = \operatorname{Im} A^{\mathrm{T}}$ 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 $\mu > 0$:

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

This implies that $(0, p_N(\mu) - \dot{s}_N(\mu))^{\mathrm{T}} \in \operatorname{Im} A^{\mathrm{T}}$ 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 3.6(iii), implies that $(p_N(\mu) - \dot{s}_N(\mu))^{\mathrm{T}} h_N(\mu)(\dot{s}_N(\mu) + v_N(\mu)) = 0$ for all $\mu > 0$. Hence, we have

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

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 3.6(i), Corollary 2.4(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 3.6, Corollary 2.4(i) and the fact that $\dot{s}(\mu) \in \operatorname{Im} A^{\mathrm{T}}$ and $\dot{s}_B^{\infty} = \lim_{\mu\to 0} \dot{s}(\mu)$, we easily see that any accumulation point of $\dot{s}_N(\mu)$ is feasible for (24) 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 3.8 *Under Assumption* H3, *we have* $\lim_{\mu\to 0} (s(\mu) - s^*)/\mu = \dot{s}^{\infty}$.

Proof The corollary follows immediately from Theorem 3.7 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

$$x^{k+1} = \arg\min\{c^{\mathrm{T}}x + \lambda_k D_\varphi(x, x^k) : Ax = b\}, \tag{25}$$

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

$$\sum_{k=0}^{\infty} \lambda_k^{-1} = +\infty. \tag{26}$$

The following result on the convergence of $\{x^k\}$, as defined in (25), is known.

Proposition 4.1 *The sequence* $\{x^k\}$ *generated by* (25)*converges to a solution of problem* (4).

Proof See, for example Theorem 3 of Iusem, Svaiter and Cruz Neto [19]. ■

The optimality condition for x^{k+1} to be an optimal solution of (25) is that $s^k \in c + \operatorname{Im} A^{\mathrm{T}}$, where

$$s^k \equiv \lambda_k \left(\nabla\varphi(x^k) - \nabla\varphi(x^{k+1})\right). \tag{27}$$

Note that in principle s^k may fail to be non-negative, 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

$$\bar{s}^k = \mu_k \sum_{i=1}^{k} \lambda_i{}^{-1} s^i, \quad \text{where } \mu_k = \left(\sum_{i=1}^{k} \lambda_i^{-1}\right)^{-1}. \tag{28}$$

Observe that $\{\mu_k\}$ converges to 0 in view of (26). 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 4.2 *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 statement related to the primal sequence was proved in Theorem 3 of Iusem, Svaiter and Cruz Neto [19] and then one related to the dual sequence follows from the final remarks in Iusem and Monteiro [15]. The last conclusion follows immediately from Proposition 2.3 (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 article, shows that this is indeed the case.

Theorem 4.3 $\lim_{k\to+\infty} s^k = s^*$.

Proof Using (28) and Proposition 4.2, it is easy to verify that

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

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

$$\left|s_i^k - \bar{s}_i^k\right| = \left|\frac{\lambda_k}{\mu_{k-1}}(\mu_k - \mu_{k-1})\, \dot{s}_i(\xi_i^{k-1})\right| = \left|\mu_k\, \dot{s}_i(\xi_i^k)\right| \leqslant \left|\xi_i^k \dot{s}_i(\xi_i^k)\right|, \tag{29}$$

where the second equality follows from the definition of μ_k in (28). As $\lim_{k\to+\infty} \mu_k = 0$ and $0 < \xi_i^k \leqslant \mu_{k-1}$ for all $i = 1, \ldots, n$, it follows from Theorem 3.2(ii) that $\lim_{k\to+\infty} \xi_i^k \dot{s}_i(\xi_i^k) = 0$. Using this fact in (29), we conclude that $\lim_{k\to\infty} s^k - \bar{s}^k = 0$, which in turn implies the theorem in view of Proposition 4.2. ■

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 4.4 *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* (25):

$$\limsup_{k\to\infty} \frac{\|x^{k+1} - x^*\|}{\|x^k - x^*\|} = \tau^{1/\gamma-1}, \qquad \limsup_{k\to\infty} \frac{c^{\mathrm{T}}(x^{k+1} - x^*)}{c^{\mathrm{T}}(x^k - x^*)} = \tau^{1/\gamma-1}. \tag{30}$$

If, in addition, Assumption H3 *holds and the limit* $\dot{s}^\infty$ *of Theorem* 3.7 *is non-zero, then the following holds for the proximal dual average sequence* $\{\bar{s}^k\}$ *given by* (28):

$$\limsup_{k\to\infty} \frac{\|\bar{s}^{k+1} - s^*\|}{\|\bar{s}^k - s^*\|} = \tau. \tag{31}$$

Proof The three limits in (30) and (31) can be easily derived using Corollaries 3.5 and 3.8, 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 4.2. ■

Using the definition of μ_k in (28), 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 4.4, we conclude that the two sequences $\{x^k\}$ and $\{\bar{s}^k\}$ both converge Q-sublinearly. Faster convergence can only be achieved if the condition $\lim_{k\to\infty} \lambda_k = 0$ 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 the article by giving some results for the case where the barrier function (6) satisfies, instead of H2, the following assumption:
(H2′) For every $\alpha \in (0, 1)$ and $j = 1, \ldots, n$, we have:

$$\lim_{t\to 0} -\frac{\varphi_j'(t)}{\varphi_j''(t)^\alpha} = 0.$$

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 5.2(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 [2], Eggermont [3] and Tseng and Bertsekas [4]. 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 [11] 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 $\{c^{\mathrm{T}}(x^k - x^*)\}$ 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. In contrast, 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 satisfying conditions H1 and H2. Thus, these conditions allow the convergence analysis of the dual average 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 determined by a barrier satisfying H1 and H2$'$.

We also mention that the case of $\bar{\lambda} > 0$ is the most important one. In view of the definition of the method (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 k-th 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 where the barrier φ satisfies conditions H1 and H2$'$. In fact, using only H1, we can obtain the following result.

PROPOSITION 5.1 *Assume that* H1 *holds*, $\varphi_1 = \cdots = \varphi_n \equiv \bar{\varphi}$ *and that* $\xi \equiv \lim_{t\to 0} t\bar{\varphi}''(t)$ *exists and belongs to* $(0, +\infty)$. *Then,* φ *satisfies* H2$'$. *If, in addition,* $\lim_{t\to 0}\bar{\varphi}(t) = 0$, *then for every* $\delta_j \in \operatorname{dom}(\bar{\varphi}^*) \subset \mathbb{R}^n$, *where* $j \in \{1, \ldots, n\}$, $s = (s_1, \ldots, s_n) \in \mathbb{R}^n_{++}$ *and* $J \subset \{1, \ldots, n\}$, *we have*:

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

where $\bar{\varphi}^*$ *is the adjoint of* $\bar{\varphi}$.

Proof See the proof in the Appendix. ■

We next present two examples of functions φ satisfying the hypotheses of Proposition 5.1, with the corresponding values of ξ.

Example 5.2 If we take:

(i) $\varphi_1(t) = t\log t$ then $\xi = 1$;
(ii) $\varphi_1(t) = (t^2 + t)\log t$ then $\xi = 1$.

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

$$\min\left\{\sigma_N(s_N) : s \in c + \operatorname{Im} A^{\mathrm{T}}, \quad s_B = 0\right\}, \tag{32}$$

called the centroid s^c of the dual solution set S^*, see for example, Cominetti and San Martín [21] and Iusem and Monteiro [15].

PROPOSITION 5.3 *Suppose that the assumptions of Proposition* 5.1 *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 [15] and Proposition 5.1. ■

Note that, as σ_N is not strictly convex in S^*, problem (32) 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 (32). 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 (17), because in this case $r_1 = +\infty$.

Acknowledgements

Cruz Neto was supported in part by CAPES and PRONEX-Otimização(FAPERJ-CNPq). Ferreira was supported in part by FUNAPE-UFG, CAPES, PADCT-CNPq, and PRONEX-Otimização (FAPERJ-CNPq). Iusem was supported in part by CNPq Grant No 301280/86 and PRONEX-Otimização (FAPERJ-CNPq). Monteiro was supported in part by NSF Grants INT-9910084, CCR-0203113 and CCF-0430644 and ONR Grants N00014-03-1-0401 and N00014-05-1-0183.

References

[1] Rockafellar, R.T., 1976, Monotone operators and the proximal point algorithm. *SIAM Journal on Control and Optimization*, **14**, 877–898.
[2] Eriksson, J., 1985, An iterative primal-dual algorithm for linear programming. Technical Report 85-10, Department of Mathematics, Linkoping University, Linkoping, Sweden.
[3] Eggermont, P.P.B., 1990, Multiplicative iterative algorithms for convex programming. *Linear Algebra and Its Applications*, **130**, 25–42.
[4] Tseng, P. and Bertsekas, D., 1993, On the convergence of the exponential multiplier method for convex programming. *Mathematical Programming*, **60**, 1–19.
[5] Censor, Y. and Zenios, S.A., 1992, Proximal minimization algorithm with D-functions. *Journal of Optimization Theory and Applications*, **73**, 451–464.
[6] Chen, G. and Teboulle, M., 1993, Convergence analysis of a proximal-like optimization algorithm using Bregman functions. *SIAM Journal on Optimization*, **3**, 538–543.
[7] 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.
[8] Kiwiel, K.C., 1997, Proximal minimization methods with generalized Bregman functions. *SIAM Journal on Control and Optimization*, **35**, 1142–1168.
[9] Iusem, A.N. and Teboulle, M., 1993, On the convergence rate of entropic proximal optimization algorithms. *Computational and Applied Mathematics*, **12**, 153–168.
[10] Jensen, D.L. and Polyak, R.A., 1994, The convergence of a modified barrier method for convex programming. *IBM Journal of Research and Development*, **38**, 307–321.
[11] Iusem, A.N. and Teboulle, M., 1995, Convergence rate analysis of nonquadratic proximal methods for convex and linear programming. *Mathematics of Operations Research*, **20**, 657–677.
[12] Powell, M.J.D., 1995, Some convergence properties of the modified log barrier method for linear programming. *SIAM Journal on Optimization*, **5**, 695–739.
[13] Polyak, R. and Teboulle, M., 1997, Nonlinear rescaling and proximal-like methods in convex optimization. *Mathematical Programming*, **76**, 265–284.

[14] Iusem, A.N., Svaiter, B.F. and Teboulle, M., 1994, Entropy like proximal methods in convex programming. *Mathematics of Operations Research*, **19**, 790–814.
[15] Iusem, A.N. and Monteiro, R.D.C., 2000, On dual convergence of the generalized proximal point method with Bregman distance. *Mathematics of Operations Research*, **25**, 606–624.
[16] Auslender, A., Teboulle, M. and Ben-Tiba, S., 1999, Interior proximal and multiplier methods based on second order homogeneous kernels. *Mathematics of Operations Research*, **24**, 645–668.
[17] Polyak, R., 2002, Nonlinear rescaling vs. smoothing technique in convex optimization. *Mathematical Programming*, **92**, 197–235.
[18] Adler, I. and Monteiro, R.D.C., 1991, Limiting behavior of the affine scaling continuous trajectories for linear programming problems. *Mathematical Programming*, **50**, 29–51.
[19] Iusem, A.N., Svaiter, B.F. and Cruz Neto, J.X., 1999, Generalized proximal point methods and Cauchy trajectories in Riemannian manifolds. *SIAM Journal on Control and Optimization*, **37**, 566–588.
[20] Hoffman, A.J., 1952, On approximate solutions of systems of linear inequalities. *Journal of Research of the National Bureau of Standards*, **49**, 263–265.
[21] Cominetti, R. and San Martín, J., 1994, Asymptotic analysis of the exponential penalty trajectory in linear programming. *Mathematical Programming*, **67**, 169–187.

Appendix A

In this appendix, we give the proofs of some technical results, namely Corollary A3, used in Proposition 2.3 and Proposition 5.1.

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 A1 *If φ satisfies* h1, *then* $\lim_{u\to-\infty} \varphi^*(u) = -\lim_{t\to 0} \varphi(t)$, *where φ^* denotes the conjugate function of φ.*

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

$$-\bar{s}u \leqslant -\varphi'(\bar{s})\bar{s} = \varphi'(\bar{s})(0 - \bar{s}) \leqslant \varphi(0) - \varphi(\bar{s}) = \lim_{t\to 0} \varphi(t) - \varphi(\bar{s}), \tag{A1}$$

using the fact that φ is closed and convex. By (A1),

$$-\lim_{t\to 0} \varphi(t) \leqslant \bar{s}u - \varphi(\bar{s}) \leqslant \sup_{s\in\mathbb{R}}\{su - \varphi(s)\} = \varphi^*(u). \tag{A2}$$

It follows from (A2) that

$$-\lim_{t\to 0} \varphi(t) \leqslant \lim_{u\to-\infty} \varphi^*(u). \tag{A3}$$

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

$$\varphi^*(u) = s_u u - \varphi(s_u) \leqslant -\varphi(s_u). \tag{A4}$$

Note that $\lim_{u\to-\infty} s_u = 0$, because $\lim_{t\to 0} \varphi'(t) = -\infty$. Taking limits in (A4) as $u \to -\infty$, we obtain

$$\lim_{u\to-\infty} \varphi^*(u) \leqslant -\lim_{u\to-\infty} \varphi(s_u) = -\lim_{t\to 0} \varphi(t). \tag{A5}$$

The result follows from (A3) and (A5). ∎

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

Our second assumption on φ is:

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

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

LEMMA A2 *If* φ *satisfies* h1 *and* h2 *then*

(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* $s > 0$,

$$\lim_{\mu\to 0} \frac{\varphi^*(\delta - s/\mu)}{\mu^{1/\gamma - 2}} = \frac{\gamma^2 r^{1/\gamma}}{(1-2\gamma)(1-\gamma)} s^{2-1/\gamma}; \tag{A6}$$

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

$$\lim_{\mu\to 0}[\varphi^*(\delta - s/\mu) - \varphi^*(\delta - 1/\mu)] = -r^2 \log s. \tag{A7}$$

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

$$\lim_{t\to 0} \frac{-\varphi'(t)}{t^{-\lambda}} = \left(\lambda r^{1/\gamma}\right)^{\lambda}. \tag{A8}$$

We proceed to prove the claim. We compute first $\lim_{t\to 0}(-\varphi'(t))^{-1/\lambda}/t$. As 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:

$$\begin{aligned}\lim_{t\to 0+} \frac{(-\varphi'(t))^{-1/\lambda}}{t} &= \lim_{t\to 0} \frac{\lambda^{-1}(-\varphi'(t))^{-1/\lambda - 1}\varphi''(t)}{1} \\ &= \lambda^{-1} \lim_{t\to 0}(-\varphi'(t))^{-1/\gamma}\varphi''(t) \\ &= \lambda^{-1} \lim_{t\to 0}\left(\frac{-\varphi'(t)}{\varphi''(t)^{\gamma}}\right)^{-1/\gamma} = \lambda^{-1} r^{-1/\gamma}.\end{aligned} \tag{A9}$$

By (A9), $\lim_{t\to 0}(-\varphi'(t)/t^{-\lambda}) = (\lambda r^{1/\gamma})^{\lambda}$, establishing the claim.

Let $\nu = (\lambda r^{1/\gamma})^{\lambda}$. By (A8), there exists $\bar{t}$ such that, for all $t \in (0, \bar{t})$, $\nu/2 \leqslant -\varphi'(t)/t^{-\lambda} \leqslant 2\nu$, that is,

$$\frac{\nu}{2} t^{-\lambda} \leqslant -\varphi'(t) \leqslant 2\nu t^{-\lambda}. \tag{A10}$$

Take $t, u \in (0, \bar{t})$. We consider first the case $\lambda \neq 1$, that is, $\gamma \neq 1/2$. Integrating (A10) between t and u and observing that $1 - \lambda = \eta$, we obtain

$$\frac{\nu}{2\eta}(u^{\eta} - t^{\eta}) \leqslant \varphi(t) - \varphi(u) \leqslant \frac{2\nu}{\eta}(u^{\eta} - t^{\eta}). \tag{A11}$$

Now, we consider separately the cases $\gamma \in (0, 1/2)$ and $\gamma \in (1/2, 1)$, that is $\eta > 0$ and $\eta < 0$, respectively. For $\gamma \in (0, 1/2)$, we obtain from (A11), for all $t \in (0, \bar{t})$, $\varphi(t) \leqslant \varphi(u) + (2\nu/\eta)(u^{\eta} - t^{\eta})$, 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 obtain from (A11) $\varphi(t) \geq \varphi(u) + \nu/(2|\eta|)(t^{\eta} - u^{\eta})$, which implies $\lim_{t\to 0}\varphi(t) = +\infty$, because

$\lim_{t\to 0} t^{\eta} = +\infty$, as $\eta < 0$. Finally, we consider the case of $\lambda = 1$, that is, $\gamma = 1/2$. In this case, integrating (A10), 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

$$\lim_{\mu\to 0} \varphi^*(\delta - s/\mu) = \lim_{u\to -\infty} \varphi^*(u) = -\lim_{t\to 0} \varphi(t), \tag{A12}$$

using Lemma A1. Now, for $\gamma \in (0, 1/2)$, it holds that $\lim_{t\to 0} \varphi(t) = 0$, by (i). Thus, in view of (A12), the numerator of the left hand side of (A6) converges to 0 as $t \to 0$. Also the denominator converges to 0 as $t \to 0$, because $\gamma \in (0, 1/2)$. In contrast, for $\gamma \in (1/2, 1)$, we have, by (ii) and (A12), that the numerator in the left hand side of (A6) 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 (A6), obtaining

$$\begin{aligned}\lim_{\mu\to 0} \frac{\varphi^*(\delta - s/\mu)}{\mu^{1/\gamma-2}} &= \lim_{\mu\to 0} \frac{s\mu^{-2}(\varphi^*)'(\delta - s/\mu)}{(1/\gamma - 2)\mu^{1/\gamma-3}} \\ &= \frac{-\gamma s}{2\gamma - 1} \lim_{\mu\to 0} \frac{(\varphi^*)'(\delta - s/\mu)}{\mu^{1/\gamma-1}}.\end{aligned} \tag{A13}$$

Let now $t = (\varphi^*)'(\delta - s/\mu) = (\varphi')^{-1}(\delta - s/\mu)$, so that $\varphi'(t) = \delta - s/\mu$, that is, $\mu = s/(\delta - \varphi'(t))$. 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 (A13)

$$\begin{aligned}\lim_{\mu\to 0} \frac{\varphi^*(\delta - s/\mu)}{\mu^{1/\gamma-2}} &= \frac{-\gamma s}{2\gamma - 1} \lim_{t\to 0} \frac{t}{(s(\delta - \varphi'(t))^{-1})^{1/\gamma-1}} \\ &= \frac{-\gamma s^{2-1/\gamma}}{2\gamma - 1} \lim_{t\to 0} \frac{t}{(\delta - \varphi'(t))^{1-1/\gamma}}.\end{aligned} \tag{A14}$$

Note that, as $\gamma < 1$, both the numerator and the denominator inside the limit in the rightmost expression of (A14) converge to 0 as $t \to 0$, so that we can apply again L'Hospital's rule, obtaining

$$\begin{aligned}\lim_{\mu\to 0} \frac{\varphi^*(\delta - s/\mu)}{\mu^{1/\gamma-2}} &= \frac{\gamma s^{2-1/\gamma}}{2\gamma - 1} \lim_{t\to 0} \frac{1}{(1 - 1/\gamma)\big(\delta - \varphi'(t)\big)^{-1/\gamma} \varphi''(t)} \\ &= \frac{\gamma^2 s^{2-1/\gamma}}{(2\gamma - 1)(\gamma - 1)} \lim_{t\to 0} \frac{\big(\delta - \varphi'(t)\big)^{1/\gamma}}{\varphi''(t)} \\ &= \frac{\gamma^2 s^{2-1/\gamma}}{(2\gamma - 1)(\gamma - 1)} \lim_{t\to 0} \left(\frac{\delta - \varphi'(t)}{\varphi''(t)^{\gamma}}\right)^{1/\gamma} \\ &= \frac{\gamma^2 s^{2-1/\gamma}}{(2\gamma - 1)(\gamma - 1)} \lim_{t\to 0} \left(\frac{-\varphi'(t)}{\varphi''(t)^{\gamma}}\right)^{1/\gamma} \\ &= \frac{\gamma^2 r^{1/\gamma}}{(1 - 2\gamma)(1 - \gamma)} s^{2-1/\gamma},\end{aligned} \tag{A15}$$

using h2. Hence, (A15) establishes (A6).

(iii) Note that $\gamma = 1/2$ implies $\lambda = 1$, with the notation of (i). Fix $s > 0$ and $\varepsilon \in (0, 1)$. By (A8), 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,

$$\frac{-(1+\varepsilon)r^2}{t} \leqslant \varphi'(t) \leqslant \frac{-(1-\varepsilon)r^2}{t}. \tag{A16}$$

As φ' is increasing and $(\varphi^*)' = (\varphi')^{-1}$, we obtain from (A16)

$$(\varphi^*)'\left(\frac{-(1+\varepsilon)r^2}{t}\right) \leqslant t \leqslant (\varphi^*)'\left(\frac{-(1-\varepsilon)r^2}{t}\right). \tag{A17}$$

Taking first $u = -(1+\varepsilon)r^2/t$ and then $u = -(1-\varepsilon)r^2/t$, it follows from (A17) that

$$\frac{-(1-\varepsilon)r^2}{u} \leqslant (\varphi^*)'(u) \leqslant \frac{-(1+\varepsilon)r^2}{u}, \tag{A18}$$

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 (A18) between v and w,

$$-(1-\varepsilon)r^2(\log(-w) - \log(-v)) \leqslant \varphi^*(w) - \varphi^*(v) \leqslant -(1+\varepsilon)r^2(\log(-w) - \log(-v)). \tag{A19}$$

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

$$\lim_{\mu\to 0} \rho(1, \mu) = 0 = \sigma(1). \tag{A20}$$

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, (A19) becomes

$$-(1-\varepsilon)r^2 \log\left(\frac{s-\delta\mu}{1-\delta\mu}\right) \leqslant \rho(s, \mu) \leqslant -(1+\varepsilon)r^2 \log\left(\frac{s-\delta\mu}{1-\delta\mu}\right), \tag{A21}$$

for all $\mu \in (0, \bar{\mu})$. Taking limits in (A21) as $\mu \to 0$,

$$(1-\varepsilon)\sigma(s) \leqslant \liminf_{\mu\to 0}\rho(s, \mu) \leqslant \limsup_{\mu\to 0}\rho(s, \mu) \leqslant (1+\varepsilon)\sigma(s). \tag{A22}$$

As (A22) holds for all $\varepsilon \in (0, 1)$, we take limits as $\varepsilon \to 0$ in (A22) and obtain

$$\lim_{\mu\to 0} \rho(s, \mu) = \sigma(s) \quad \forall s \in (0, 1). \tag{A23}$$

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 (A19) holds, but now $\varphi^*(w) - \varphi^*(v) = -\rho(s, \mu)$, in which case the inequalities in (A22) reverse, and after taking limits as $\mu \to 0$, we obtain

$$(1-\varepsilon)\sigma(s) \geq \limsup_{\mu\to 0}\rho(s, \mu) \geq \liminf_{\mu\to 0}\rho(s, \mu) \geq (1+\varepsilon)\sigma(s). \tag{A24}$$

Then, taking limits in (A24) as $\varepsilon \to 0$, we conclude that

$$\lim_{\mu\to 0} \rho(s, \mu) = \sigma(s) \quad \forall s > 1. \tag{A25}$$

By (A20), (A23) and (A25), $\lim_{\mu\to 0} \rho(s, \mu) = \sigma(s)$ for all $s > 0$, which is precisely (A7). ■

Corollary A3 *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 corespondent* r_j*'s . Then*:

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

$$\lim_{\mu\to 0} \frac{\varphi^*(\delta - s/\mu)}{\mu^{1/\gamma-2}} = \frac{\gamma^2}{(1-2\gamma)(1-\gamma)} \sum_{j=1}^n r_j^{1/\gamma} s_j^{2-1/\gamma}. \tag{A26}$$

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

$$\lim_{\mu\to 0} \big(\varphi^*(\delta - s/\mu) - \varphi^*(\delta - 1/\mu)\big) = -\sum_{j=1}^n r_j^2 \log s_j. \tag{A27}$$

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 A2. ■

The results of Lemma A2(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 (A6) is a strictly convex functions of s (otherwise it is 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 $\gamma \in (0, 1)$.

We now turn our attention towards proving Proposition 5.1. First, we need to establish the following technical result.

Lemma A4 *If* $\xi \equiv \lim_{t\to 0} t\varphi''(t)$ *exists and belongs to* $(0, +\infty)$, *then* φ *satisfies*

$$\lim_{t\to 0} -\frac{\varphi'(t)}{\varphi''(t)^\alpha} = 0, \quad \forall \alpha > 0.$$

If, in addition, $\lim_{t\to 0} \varphi(t) = 0$, *then, for all* $\delta \in \mathrm{dom}(\varphi^*)$ *and all* $s > 0$,

$$\lim_{\mu\to 0} \left[\varphi^*(\delta - s/\mu)^\mu\right] = e^{-s/\xi(\varphi)}. \tag{A28}$$

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

$$\frac{\xi}{2t} \leqslant \varphi''(t) \leqslant \frac{2\xi}{t}. \tag{A29}$$

Take $t \in (0, \bar{t})$. Integrating (A29) between t and $\bar{t}$,

$$(\xi/2)\log(\bar{t}/t) \leqslant \varphi'(\bar{t}) - \varphi'(t) \leqslant 2\xi \log(\bar{t}/t). \tag{A30}$$

From (A29) and (A30) it follows that, for $t \in (0, \bar{t})$ and $\alpha > 0$,

$$\begin{aligned}(2\xi)^{-\alpha} t^\alpha (-\varphi'(\bar{t}) + (\xi/2)\log \bar{t} - (\xi/2)\log t) &\leqslant \frac{-\varphi'(t)}{\varphi''(t)^\alpha} \\ &\leqslant (\xi/2)^{-\alpha} t^\alpha \left(-\varphi'(\bar{t}) + 2\xi \log \bar{t} - 2\xi \log t\right).\end{aligned} \tag{A31}$$

Taking limits in (A31) as $t \to 0$, and remembering that, for all $\alpha > 0$,

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

we obtain $\lim_{t\to 0}(-\varphi'(t)/(\varphi''(t))^\alpha) = 0$ for all $\alpha > 0$.

We proceed to prove (A28). Let

$$\psi(\mu) = \log\left(\varphi^*(\delta - s/\mu)^\mu\right) = \mu \log \varphi^*(\delta - s/\mu). \tag{A32}$$

Then

$$\lim_{\mu\to 0} \psi(\mu) = \lim_{\mu\to 0} \frac{\log \varphi^*(\delta - s/\mu)}{\mu^{-1}}. \tag{A33}$$

From Lemma A1 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 $\lim_{\mu\to 0} \log \varphi^*(\delta - s/\mu) = -\infty$. As $\lim_{\mu\to 0} \mu^{-1} = +\infty$, we can apply L'Hospital's rule to compute the limit in (A33), obtaining

$$\begin{aligned}\lim_{\mu\to 0} \psi(\mu) &= \lim_{\mu\to 0} \frac{(s/\mu^2)\varphi^*(\delta - s/\mu)^{-1}(\varphi^*)'(\delta - s/\mu)}{-\mu^{-2}} \\ &= -s \lim_{\mu\to 0}\left(\frac{(\varphi^*)'(\delta - s/\mu)}{\varphi^*(\delta - s/\mu)}\right) = -s \lim_{t\to 0} \frac{t}{(\varphi^*)(\varphi'(t))},\end{aligned} \tag{A34}$$

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 (A34)

$$\lim_{\mu\to 0} \psi(\mu) = -s \lim_{t\to 0} \frac{t}{t\varphi'(t) - \varphi(t)}. \tag{A35}$$

Multiplying throughout (A30) 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 (A35) converge to 0 as $t \to 0$, allowing us to apply L'Hospital's rule to (A35), which gives

$$\lim_{\mu\to 0} \psi(t) = -s \lim_{t\to 0} \frac{1}{t\varphi''(t)} = \frac{-s}{\xi}. \tag{A36}$$

By (A36) and (A32), we obtain $\lim_{\mu\to 0} \varphi^*(\delta - s/\mu)^\mu = e^{-s/\xi}$, establishing (A28). ∎

We end the appendix with the proof of Proposition 5.1.

Proof of Proposition 5.1 First note that, we cannot apply directly Lemma A4 because powers do not distribute with sums. Let $I = \arg\min\{s_j : 1 \leqslant j \leqslant n\}$ and $L = \arg\max\{\delta_i : i \in I\}$, and fix some $\ell \in L$. We claim that for μ close enough to 0, $\delta_\ell - s_\ell/\mu \geq \delta_j - s_j/\mu$ for all $j \in \{1,\ldots,n\}$. 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 $\mu \leqslant (s_j - s_\ell)/|\delta_j - \delta_\ell|$. As $(\bar{\varphi}^*)' = (\bar{\varphi}')^{-1}$, we have that $\mathrm{Im}[(\bar{\varphi}^*)'] = \mathrm{dom}(\bar{\varphi}') = \mathbb{R}_{++}$. Thus, $(\bar{\varphi}^*)'(u) > 0$ for all u, that is, $\bar{\varphi}^*$ is increasing. It follows that

$$\bar{\varphi}^*(\delta_\ell - s_\ell/\mu) \geq \bar{\varphi}^*(\delta_j - s_j/\mu), \tag{A37}$$

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

$$\left(\sum_{j=1}^n \bar{\varphi}^*(\delta_j - s_j/\mu)\right)^\mu = \left(\bar{\varphi}^*(\delta_\ell - s_\ell/\mu)\right)^\mu \left(\sum_{j=1}^m \frac{\bar{\varphi}^*(\delta_j - s_j/\mu)}{\bar{\varphi}^*(\delta_\ell - s_\ell/\mu)}\right)^\mu. \tag{A38}$$

The first factor in the rightmost expression of (A38) converges to $\exp(-s_\ell/\xi)$ by Lemma A4. 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 (A37). Thus $1 \leqslant S(\mu) \leqslant n$, and therefore $\lim_{\mu \to 0} S(\mu)^{\mu} = 1$. It follows that

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

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