SIAM J. OPTIM.
Vol. 12, No. 1, pp. 238–261

RESCALING AND STEPSIZE SELECTION IN PROXIMAL METHODS USING SEPARABLE GENERALIZED DISTANCES*

PAULO JOSÉ DA SILVA E SILVA†, JONATHAN ECKSTEIN‡, AND CARLOS HUMES, JR.†

Abstract. This paper presents a convergence proof technique for a broad class of proximal algorithms in which the perturbation term is separable and may contain barriers enforcing interval constraints. There are two key ingredients in the analysis: a mild regularity condition on the differential behavior of the barrier as one approaches an interval boundary and a lower stepsize limit that takes into account the curvature of the proximal term. We give two applications of our approach. First, we prove subsequential convergence of a very broad class of proximal minimization algorithms for convex optimization, where different stepsizes can be used for each coordinate. Applying these methods to the dual of a convex program, we obtain a wide class of multiplier methods with subsequential convergence of both primal and dual iterates and independent adjustment of the penalty parameter for each constraint. The adjustment rules for the penalty parameters generalize a well-established scheme for the exponential method of multipliers. The results may also be viewed as a generalization of recent work by Ben-Tal and Zibulevsky [*SIAM J. Optim*, 7 (1997), pp. 347–366] and Auslender, Teboulle, and Ben-Tiba [*Comput. Optim. Appl.*, 12 (1999), pp. 31–40; *Math. Oper. Res.*, 24 (1999), pp. 645–668] on methods derived from φ-divergences. The second application established full convergence, under a novel stepsize condition, of Bregman-function-based proximal methods for general monotone operator problems over a box. Prior results in this area required strong restrictive assumptions on the monotone operator.

Key words. proximal algorithms, Bregman distances, φ-divergence, convex programming, variational inequalities

AMS subject classifications. 90C25, 90C33

PII. S1052623499365784

1. Introduction.

Let $B \subseteq \mathbb{R}^n$ denote the possibly unbounded n-dimensional "box" $([a_1, b_1] \times \cdots \times [a_n, b_n]) \cap \mathbb{R}^n$, where $-\infty \le a_i < b_i \le +\infty$ for all $i = 1, \ldots, n$. This paper considers two closely related problems: the minimization problem

$$\min_{x \in B} f(x), \tag{1.1}$$

where $f : \mathbb{R}^n \to (-\infty, +\infty]$ is a closed proper convex function, and the variational inequality

$$0 \in T(x) + N_B(x), \tag{1.2}$$

where T is a (possibly set-valued) maximal monotone operator and $N_B(x)$ denotes the cone of vectors normal to the set B at x. It is well known that, under mild regularity conditions, (1.1) is the special case of (1.2) for which $T = \partial f$, the subgradient mapping of f.

*Received by the editors December 7, 1999; accepted for publication (in revised form) February 23, 2001; published electronically October 18, 2001. This paper was partially supported by FAPESP, processo 96/09939-0, the Rutgers University Board of Trustees Research Fellowship for Scholarly Excellence, and PRONEX, convênio 76.97.1008.00.

http://www.siam.org/journals/siopt/12-1/36578.html

†Department of Computer Science, Instituto de Matemática e Estatística—University of São Paulo, São Paulo, Brazil (rsilva@ime.usp.br, humes@ime.usp.br). The first author was a visiting student at RUTCOR from July 1, 1999 through December 31, 1999.

‡Faculty of Management and RUTCOR, 640 Bartholomew Road, Busch Campus, Rutgers University, Piscataway, NJ 08854 (jeckstei@rutcor.rutgers.edu).

The last decade has seen considerable progress in the theory of proximal point methods based on generalized distances [11, 13, 19, 5, 21, 31, 14, 2, 3, 17]. Such methods use a scalar-valued regularization function to derive better-behaved versions of problems (1.1) and (1.2). In this article, we consider separable regularization terms of the form

$$D(x, y) = \sum_{i=1}^{n} d_i(x_i, y_i),$$

where $d_1, \dots, d_n$ are scalar functions conforming to very general assumptions (see Assumption 2.1 below). In particular, we assume that as $x \in \operatorname{int} B$ approaches the boundary of B, $\|\nabla_1 D(x, y)\| \to \infty$, where ∇_1 denotes the gradient with respect to the first vector argument. The distance-like measure D can be, for example, the squared Euclidean distance, a Bregman distance [8], or a φ-divergence [19] (see section 2.2 below).

Using these regularization terms, proximal methods for (1.1) take the form

$$x^{k+1} = \arg\min_{x} \left\{ f(x) + \sum_{i=1}^{n} \frac{1}{\alpha_i^k} d_i(x_i, x_i^k) \right\}, \tag{1.3}$$

where α^k is a positive n-dimensional vector whose elements are called *stepsizes*. Note that we allow different stepsizes for each coordinate, as suggested by a variety of computational and theoretical studies [32, 5, 2, 3]. Moreover, since $\|\nabla_1 D(x, x^k)\| \to \infty$ as x approaches the boundary of B, the regularization acts not only as a stabilizing proximal term but also as a kind of barrier function keeping the iterates within $\operatorname{int} B$.

In the case of the variational inequality (1.2), (1.3) generalizes to finding x^{k+1} satisfying the recursion

$$0 \in T(x^{k+1}) + \operatorname{diag}(\alpha^k)^{-1} \nabla_1 D(x^{k+1}, x^k). \tag{1.4}$$

We derive some general results for these types of algorithms in section 2, assuming that the stepsizes conform to a special rule that takes into account the curvature of the proximal term. This rule, although restrictive, appears to cover cases of the greatest practical interest; as we shall see, it covers the stepsize/penalty selection rules proposed in [32, 5, 2, 3].

Section 3 uses the results of section 2 to obtain subsequential convergence results for the generalized proximal minimization algorithm (1.3).

A critical application of (1.3), considered in section 3.2, is when f is minus the dual function of a convex program such as

$$\begin{array}{lll} \min & g_0(y) & \\ \text{such that (s.t.)} & g_i(y) \le 0, & i = 1, \dots, n, \end{array} \tag{1.5}$$

where $g_0, \dots, g_n : \mathbb{R}^m \to \mathbb{R}$ are differentiable convex functions.[1] We also assume that this problem is feasible; i.e., there is a $\bar{y} \in \mathbb{R}^m$ such that $g_i(\bar{y}) \le 0, i = 1, \dots, n$. Choosing B to be any box containing the nonnegative orthant and f to be the negative of the dual function of (1.5), we may implement (1.3) via a multiplier method in

[1] Actually, the results of section 3.2 continue to hold [28] if one supposes only that $g_0, \dots, g_n : \mathbb{R}^m \to (-\infty, \infty]$ are closed proper convex and assumes appropriate conditions on the effective domains of the objective and constraints, as in [24, Chapter 28]. However, this further generality makes the proofs more convoluted and is dropped for the sake of simplicity in the exposition.

which a sequence of unconstrained penalized versions of (1.5) must be solved. This construction leads to a class of multiplier methods that is extremely broad, subsuming both the classical quadratic augmented Lagrangian and the exponential method of multipliers [32, 6].

For these multiplier methods, our stepsize choice ensures that for indices i with $x_i^k \to 0$ the corresponding penalty term is augmented so that it does not become so "flat" as to permit infeasibility of primal limit points. Empirically, the technique speeds convergence, and it also appears in a convergence *rate* analysis in [32] for the exponential method of multipliers case. Ben-Tal and Zibulevsky [5] have proved the optimality of the accumulation points of the exponential method, together with a class of proximal terms closely related to φ-divergences, and their results are extended in [3]. Section 3 places such results in a broader context that includes Bregman distances.

In section 4, we restrict our attention to Bregman distances. It has been known for the better part of a decade that, when $D(\cdot,\cdot)$ is any Bregman distance and the stepsizes do not vary by coordinate, the recursion (1.4) converges to a solution of the variational inequality (1.2) in various special cases: when $T = \partial f$, the subdifferential of a closed proper convex function f, or when $\operatorname{dom} T \subseteq \operatorname{int} B$, meaning that all constraints must already be embedded in the operator T. In [9], these results were extended to "paramonotone" operators T, a category which includes $T = \partial f$ as a special case. Unfortunately, many interesting practical cases, such as the subdifferential maps of saddle functions, are not paramonotone. More recently, Auslender, Teboulle, and Ben-Tiba [2] have obtained strong results for general maximal monotone T, but only for a specific φ-divergence choice of $D(\cdot,\cdot)$. As noted in [4], these results can be extended to the (generally non-Bregman) case in which $D(\cdot,\cdot)$ is obtained by adding a quadratic to any member of the class Φ_2 of [3].

Section 4 shows convergence, for general maximal monotone T, of the proximal method (1.4), where $D(\cdot,\cdot)$ is a Bregman distance, to a solution of (1.2). We do impose some additional assumptions, derived from those of section 2. First, we assume that the Bregman function used to construct the distance is twice-differentiable, which is not part of the standard Bregman function setup. Second, in addition to our general stepsize rule, we also require that the stepsizes do not vary by coordinate, that is, $\alpha_1^k = \cdots = \alpha_n^k$ for all k. The resulting condition is stronger than the usual requirement that the stepsize is simply bounded away from zero, but is crucial to the analysis, which blends the techniques of section 2 with traditional Fejér monotonicity arguments. Still, we have managed to substitute conditions on $D(\cdot,\cdot)$ and α^k, which are parts of the algorithm, for conditions on T, which is part of the problem to be solved.

Finally, we allow the calculations required for the recursions (1.3) and (1.4) to be performed approximately, as is likely to be necessary in practice. For the rescaling minimization case of section 3, we adopt a constructive approximation criterion inspired by [17] and [29]. However, our criterion, which is tailored to the proximal minimization case, appears to be new. In the variational inequality analysis of section 4, we use the simple, verifiable criterion of [14], although extension to the more sophisticated criterion of [29] may well be possible.

In summary, the primary contributions of this paper are

- a novel convergence proof framework for a broad class of proximal algorithms;
- using this framework to establish subsequential convergence of a wide range of proximal minimization algorithms (1.3) with differing stepsize parameters for each coordinate—this result in turn leads to subsequential convergence

of a broad class of multiplier methods with differing penalty parameters for each constraint;

- using the framework to show convergence of "interior" Bregman proximal point algorithms for maximal monotone operators, with a novel stepsize condition, but without the usual restrictive assumptions on the operator T.

The new proximal minimization approximation criterion of section 3 constitutes an additional contribution.

2. Fundamental analysis. This section develops the fundamental analysis necessary for our results. We concentrate our attention on the variational problem (1.2), since it subsumes the minimization problem (1.1) under mild assumptions.

In order to simplify the notation, we denote, for $i = 1, \ldots, n$,

$$d_i'(x_i, y_i) \overset{\text{def}}{=} \frac{\partial d_i}{\partial x_i}(x_i, y_i),$$

$$d_i''(x_i, y_i) \overset{\text{def}}{=} \frac{\partial^2 d_i}{\partial x_i^2}(x_i, y_i).$$

We are now able to present the necessary assumptions on the functions d_i.

ASSUMPTION 2.1. *For $i = 1, \ldots, n$, the function $d_i : \mathbb{R} \times (a_i, b_i) \to (-\infty, \infty]$ has the following properties:*

2.1.1. *For all $y_i \in (a_i, b_i)$, $d_i(\cdot, y_i)$ is closed and strictly convex, with its minimum at y_i. Moreover,* $\operatorname{int} \operatorname{dom} d_i(\cdot, y_i) = (a_i, b_i)$.

2.1.2. *d_i is continuously differentiable over $(a_i, b_i) \times (a_i, b_i)$, and, for all $y_i \in (a_i, b_i)$, $d_i''(y_i, y_i)$ exists and is strictly positive.*

2.1.3. *For all $y_i \in (a_i, b_i)$, $d_i(\cdot, y_i)$ is essentially smooth* [24, Chapter 26].

2.1.4. *There exist $\rho, \epsilon > 0$ such that if either $-\infty < a_i < y_i \le x_i < a_i + \epsilon$ or $b_i - \epsilon < x_i \le y_i < b_i < +\infty$, then $\rho\,|d_i'(x_i, y_i)| \le d_i''(y_i, y_i)\,|x_i - y_i|$.*

The assumption of strict convexity is standard in generalized proximal methods. The assumption of twice-differentiability is also quite common, although many existing results require only a once-differentiable d_i. The essential smoothness assumption makes the distance D act like a barrier function, forcing the iterates defined by the recursion (1.4), and hence its approximate version (2.1) below, to remain in the interior of the box B. In section 2.2, we specialize these assumptions to the case of Bregman distances and φ-divergences, where similar comments can be made.

Finally, the fourth part of the assumption is new to the theory of generalized proximal methods, but is not very restrictive in practice. In particular, we show in section 2.2 that, for Bregman distances and φ-divergences, this condition can be written in terms of the kernels used to obtain the regularizations, and that it holds for most of the examples of which we are aware.

In addition, we make the following standard regularity assumption which, in view of the barrier function properties of d_i, is required for any sensible application of (1.4).

ASSUMPTION 2.2. $\operatorname{dom} T \cap \operatorname{int} B \ne \emptyset$.

We are now able to present the proximal minimization algorithm.

RESCALING PROXIMAL METHOD FOR VARIATIONAL INEQUALITY (RPMVI).

1. **Initialization:** *Let $k = 0$. Choose a scalar $c > 0$ and an initial iterate $x^0 \in \operatorname{int} B$.*
2. **Iteration:**
   (a) *Choose $\alpha^k \in \mathbb{R}^n_{++}$ such that $\alpha_i^k \ge c \max\left\{1, d_i''(x_i^k, x_i^k)\right\}$ for $i = 1, \ldots, n$.*
   (b) *Find x^{k+1} and e^{k+1} such that*

$$e^{k+1} \in T(x^{k+1}) + \operatorname{diag}(\alpha^k)^{-1} \nabla_1 D(x^{k+1}, x^k). \tag{2.1}$$

(c) *Let $k = k+1$, and repeat the iteration.* □

To guarantee the convergence of the RPMVI, we need additional assumptions on the stepsizes $\{\alpha_i^k\}$ and the error sequence $\{e^k\}$; see Assumption 2.3 below.

We define

$$\gamma^k \stackrel{\text{def}}{=} e^k - \operatorname{diag}(\alpha^{k-1})^{-1} \nabla_1 D(x^k, x^{k-1}), \tag{2.2}$$

whence it is clear from (2.1) that $\gamma^k \in T(x^k)$ for all $k \geq 1$.

ASSUMPTION 2.3. *Let $\{\beta_k\}$ be a real sequence converging to zero. The error sequence $\{e^k\}$, the regularization functions $d_1, \ldots, d_n$, and the stepsizes $\{\alpha_i^k\}$, $i = 1, \ldots, n$, must be chosen in order to guarantee the following:*

2.3.1. $\left|e_i^k\right| \leq \frac{1}{\alpha_i^{k-1}} \left|d_i'(x_i^k, x_i^{k-1})\right| + \beta_k$.

2.3.2. *If $\bar{x}$ is an accumulation point of $\{x^k\}$, i.e., there is an infinite set $\mathcal{K} \subseteq \mathbb{N}$ such that $x^k \to_{\mathcal{K}} \bar{x}$, then, for each $i = 1, \ldots, n$, either $\gamma_i^k \to_{\mathcal{K}} 0$ or there is an infinite set $\mathcal{K}' \subseteq \mathcal{K}$ such that $x_i^{k-1} \to_{\mathcal{K}'} \bar{x}_i$.*

Assumption 2.3 may seem artificial at this point, but sections 3 and 4 will describe settings in which it is easily verifiable.

2.1. Convergence analysis. We assume throughout this section that Assumptions 2.1 and 2.2 hold and that sequences $\{\alpha_k\}$, $\{x^k\}$, and $\{e^k\}$ conforming to the recursions of the RPMVI algorithm and Assumption 2.3 exist. In sections 3 and 4 we will present conditions which, in more specific settings, guarantee the existence of such sequences.

LEMMA 2.4. *Let $\bar{x} \in \mathbb{R}^n$ be a limit point of $\{x^k\}$, i.e., $x^k \to_{\mathcal{K}} \bar{x}$ for some infinite set $\mathcal{K} \subseteq \mathbb{N}$. Then for $i = 1, \ldots, n$,*

$$\begin{array}{rcll} \displaystyle\lim_{k \to_{\mathcal{K}} \infty} \gamma_i^k & = & 0 & \text{if } \bar{x}_i \in (a_i, b_i), \\ \displaystyle\liminf_{k \to_{\mathcal{K}} \infty} \gamma_i^k & \geq & 0 & \text{if } \bar{x}_i = a_i, \\ \displaystyle\limsup_{k \to_{\mathcal{K}} \infty} \gamma_i^k & \leq & 0 & \text{if } \bar{x}_i = b_i. \end{array} \tag{2.3}$$

Proof. For each i, we consider the three possible cases.

First, suppose i is such that $\bar{x}_i \in (a_i, b_i)$. For the sake of a contradiction, assume that $\gamma_i^k \not\to_{\mathcal{K}} 0$. Then, using Assumption 2.3.2, there is an infinite set $\mathcal{K}' \subseteq \mathcal{K}$ and a $\zeta > 0$ such that for all $k \in \mathcal{K}'$, $|\gamma_i^k| \geq \zeta$ and $x_i^{k-1} \to_{\mathcal{K}'} \bar{x}_i$. Therefore

$$\begin{array}{rcll} \left|\gamma_i^k\right| & = & \left| e_i^k - \dfrac{1}{\alpha_i^{k-1}} d_i'(x_i^k, x_i^{k-1}) \right| & \\ & \leq & \dfrac{1}{\alpha_i^{k-1}} \left|d_i'(x_i^k, x_i^{k-1})\right| + \left|e_i^k\right| & \\ & \leq & \dfrac{2}{\alpha_i^{k-1}} \left|d_i'(x_i^k, x_i^{k-1})\right| + \beta_k & \text{(Assumption 2.3.1)} \\ & \leq & (2/c) \left|d_i'(x_i^k, x_i^{k-1})\right| + \beta_k & \text{(choice of } \alpha_i^k) \\ & \underset{\mathcal{K}'}{\to} & (2/c) \left|d_i'(\bar{x}_i, \bar{x}_i)\right| + 0 & \\ & = & 0 & \text{(the minimum of } d_i(\cdot, \bar{x}_i) \text{ is } \bar{x}_i). \end{array}$$

This result contradicts $\left|\gamma_i^k\right| > \zeta,\ k \in \mathcal{K}'$.

Next, consider the case $\bar{x}_i = a_i$, and suppose that $\liminf_{k \to_{\mathcal{K}} \infty} \gamma_i^k < 0$. Then, using Assumption 2.3.2, there must be a $\zeta > 0$ and an infinite set $\mathcal{K}' \subseteq \mathcal{K}$ such that for all $k \in \mathcal{K}'$, $\gamma_i^k \leq -\zeta$ and $x_i^{k-1} \to_{\mathcal{K}'} \bar{x}_i$. Then

$$\begin{aligned}
\zeta &\leq \left|\gamma_i^k\right| \\
&= \left| e_i^k - \frac{1}{\alpha_i^{k-1}} d_i'(x_i^k, x_i^{k-1}) \right| \\
&\leq \frac{2}{\alpha_i^{k-1}} \left| d_i'(x_i^k, x_i^{k-1}) \right| + \beta_k \\
&\leq \frac{2}{c d_i''(x_i^{k-1}, x_i^{k-1})} \left| d_i'(x_i^k, x_i^{k-1}) \right| + \beta_k.
\end{aligned}$$

Let ϵ be as in Assumption 2.1.4. If there is an infinite set $\mathcal{K}'' \subseteq \mathcal{K}'$ such that $x_i^{k-1} \leq x_i^k \leq a_i + \epsilon$ for all $k \in \mathcal{K}''$, we can conclude from the assumption that

$$\begin{aligned}
\zeta &\leq \frac{2}{c d_i''(x_i^{k-1}, x_i^{k-1})} \left| d_i'(x_i^k, x_i^{k-1}) \right| + \beta_k \\
&\leq \frac{2 d_i''(x_i^{k-1}, x_i^{k-1})}{\rho c d_i''(x_i^{k-1}, x_i^{k-1})} \left| x_i^k - x_i^{k-1} \right| + \beta_k \\
&= \frac{2}{\rho c} \left| x_i^k - x_i^{k-1} \right| + \beta_k \\
&\underset{\mathcal{K}''}{\to} 0,
\end{aligned}$$

since $x^{k-1} \to_{\mathcal{K}'} \bar{x}_i$ and $\beta_k \to 0$; but this conclusion contradicts $\zeta > 0$. Therefore, $x_i^k \leq x_i^{k-1}$ for sufficiently large $k \in \mathcal{K}'$.

As $d_i(\cdot, x_i^{k-1})$ achieves its minimum at x_i^{k-1}, having $x_i^k \leq x_i^{k-1}$ implies that $d_i'(x_i^k, x_i^{k-1}) \leq 0$. Hence

$$\begin{aligned}
\gamma_i^k &= e_i^k - \frac{1}{\alpha_i^{k-1}} d_i'(x_i^k, x_i^{k-1}) \\
&\geq \frac{1}{\alpha_i^{k-1}} \left| d_i'(x_i^k, x_i^{k-1}) \right| - \left| e_i^k \right| \\
&\geq -\beta_k \\
&> -\zeta
\end{aligned}$$

for sufficiently large $k \in \mathcal{K}'$, a contradiction with $\gamma_i^k \leq -\zeta < 0$, $k \in \mathcal{K}'$.

Finally, the case of $\bar{x}_i = b_i$ is analogous to the case of $\bar{x}_i = a_i$. ∎

LEMMA 2.5. *Let $\bar{x}$ be a limit point of $\{x^k\}$, i.e., $x^k \to_{\mathcal{K}} \bar{x}$ for some infinite set $\mathcal{K} \subseteq \mathbb{N}$. Then, $\{\gamma^k\}_{\mathcal{K}}$ is bounded.*

Proof. By Assumption 2.2, there must exist some $\widetilde{x} \in \operatorname{dom} T \cap \operatorname{int} B$. Let $\widetilde{\gamma} \in T(\widetilde{x})$. The monotonicity of T implies that, for all $k \geq 0$,

$$0 \leq \langle x^k - \widetilde{x}, \gamma^k - \widetilde{\gamma} \rangle = \sum_{i=1}^{n} (x_i^k - \widetilde{x}_i)(\gamma_i^k - \widetilde{\gamma}_i). \tag{2.4}$$

We will show that unboundedness of $\{\gamma^k\}_{\mathcal{K}}$ would contradict this inequality for some sufficiently large k.

If $\{\gamma^k\}_{\mathcal{K}}$ is unbounded, there must exist an infinite $\mathcal{K}' \subseteq \mathcal{K}$ such that $\{\gamma^k\}_{\mathcal{K}'}$ converges in $[-\infty, \infty]^n$, with at least one $\{\gamma_i^k\}_{\mathcal{K}'}$ unbounded. Lemma 2.4 implies that for each unbounded coordinate i, either

$$\gamma_i^k \to_{\mathcal{K}'} +\infty \text{ and } \bar{x}_i = a_i$$
$$\text{or}$$
$$\gamma_i^k \to_{\mathcal{K}'} -\infty \text{ and } \bar{x}_i = b_i.$$

Therefore, for each unbounded coordinate of $\{\gamma^k\}_{\mathcal{K}'}$, we have

$$(x_i^k - \widetilde{x}_i)(\gamma_i^k - \widetilde{\gamma}_i) \to_{\mathcal{K}'} (a_i - \widetilde{x}_i)(+\infty) = -\infty$$
$$\text{or}$$
$$(x_i^k - \widetilde{x}_i)(\gamma_i^k - \widetilde{\gamma}_i) \to_{\mathcal{K}'} (b_i - \widetilde{x}_i)(-\infty) = -\infty.$$

On the other hand, for coordinates such that $\{\gamma_i^k\}_{\mathcal{K}'}$ is bounded, $(x_i^k - \widetilde{x}_i)(\gamma_i^k - \widetilde{\gamma}_i)$ is also bounded. Thus, for sufficiently large $k \in \mathcal{K}' \subseteq \mathcal{K}$, $\langle x^k - \widetilde{x}, \gamma^k - \widetilde{\gamma}\rangle$ must be negative, contradicting (2.4). ☐

Finally, the main convergence theorem for the RPMVI follows.

THEOREM 2.6. *If $\{x^k\}$ is a sequence generated by the RPMVI algorithm with Assumptions* 2.1, 2.2, *and* 2.3 *holding, then all the limit points of $\{x^k\}$ are solutions to the variational inequality problem* (1.2).

Proof. Let $\bar{x}$ be any limit point of $\{x^k\}$, i.e., $x^k \to_{\mathcal{K}} \bar{x}$, for some infinite set $\mathcal{K} \subseteq \mathbb{N}$. From Lemma 2.5, we know that the corresponding sequence $\gamma^k \in T(x^k)$ is bounded. Then, there must exist some $\mathcal{K}' \subseteq \mathcal{K}$ with $\gamma^k \to_{\mathcal{K}'} \bar{\gamma} \in \mathbb{R}^n$. Since T must be outer semicontinuous [27, Exercise 12.8(b)], it follows that $\bar{\gamma} \in T(\bar{x})$. Lemma 2.4 implies that

$$\begin{aligned} \bar{\gamma}_i &= 0 && \text{if } \bar{x}_i \in (a_i, b_i), \\ \bar{\gamma}_i &\geq 0 && \text{if } \bar{x}_i = a_i, \\ \bar{\gamma}_i &\leq 0 && \text{if } \bar{x}_i = b_i, \end{aligned}$$

and these conditions are equivalent to $0 \in T(\bar{x}) + N_B(\bar{x})$. ☐

Incidentally, it is possible to eliminate the requirement of twice-differentiability of $d_i(\cdot, y_i)$, at the cost of some additional complexity in the description of the method. Specifically, consider replacing Assumption 2.1.4 with the condition that there exist $\delta, \epsilon > 0$ and functions $L_i : (a_i, b_i) \to (\delta, +\infty)$ such that if either $-\infty < a_i < y_i \leq x_i < a_i + \epsilon$ or $b_i - \epsilon < x_i \leq y_i < b_i < +\infty$, then

$$|d_i'(x_i, y_i)| \leq L_i(y_i)\, |x_i - y_i|\,.$$

If the stepsizes are now selected so that for some scalar $c > 0$, we have for all $i = 1, \ldots, n$ and $k \geq 0$ that $\alpha_i^k \geq c L_i(x_i^k)$, then the conclusions of Theorem 2.6 continue to hold. We may examine this variation of the analysis in subsequent research. The present approach is equivalent to taking $L_i(y_i) = (1/\rho) d''(y_i, y_i)$, a natural choice since $d''(y_i, y_i)$ measures the rate of change of $d'(\cdot, y_i)$ around y_i.

2.2. Some examples of d_i functions. We present some examples of d_i functions that conform with Assumption 2.1. In particular, we show that two classes of regularizations widely studied in the literature, Bregman distances [11, 13] and φ-divergences [19], conform to the assumption under very mild restrictions.

2.2.1. Bregman distances. Bregman distances were introduced in [8] and have been studied in the context of proximal methods in [11, 12, 13], as well as many subsequent works. To construct each regularization $d_i(\cdot,\cdot)$, one uses an auxiliary convex function h_i and defines $d_i(x_i, y_i) = h_i(x_i) - h_i(y_i) - h_i'(y_i)(x_i - y_i)$. Nonseparable distances can also be constructed in a similar way, but the separable case is the most common.

The following properties guarantee that Assumption 2.1 holds for such d_i.

ASSUMPTION 2.7. *For $i = 1, \ldots, n$, the function $h_i : \mathbb{R} \to (-\infty, \infty]$ has the following properties:*

2.7.1. *h_i is closed,* $\operatorname{int}\operatorname{dom} h = (a_i, b_i)$*, and h_i is twice continuously differentiable, with a strictly positive second derivative throughout (a_i, b_i).*

2.7.2. *h_i is essentially smooth.*

2.7.3. *There exist $\rho > 0$ and $\epsilon > 0$ such that if either $-\infty < a_i < y_i \le x_i < a_i + \epsilon$ or $b_i - \epsilon < x_i \le y_i < b_i < +\infty$, then $\rho\,|h_i'(x_i) - h_i'(y_i)| \le h_i''(y_i)\,|x_i - y_i|$.*

Note that Assumption 2.7.1 implies that each h_i is strictly convex. Assumption 2.7.3 corresponds to Assumption 2.1.4, since $d_i''(x_i, y_i) = h_i''(x_i)$. Fortunately, it is not very restrictive. Consider the case of finite a_i. Since $\lim_{x_i \searrow a_i} h_i'(x_i) = -\infty$, we know that $h_i''(x_i)$ must be unbounded above as $x_i \searrow a_i$. To violate the assumption, $h_i''(x_i)$ would have to oscillate unboundedly as $x_i \searrow a_i$. As far as we are aware, every separable Bregman function proposed so far conforms not only to Assumption 2.7.3 but to a more stringent, easier-to-verify condition, as follows.

LEMMA 2.8. *If there is an $\epsilon > 0$ such that for all $x_i \in (a_i, a_i + \epsilon) \cap \mathbb{R}$, h_i'' is nonincreasing, and for all $x \in (b_i - \epsilon, b_i) \cap \mathbb{R}$, h_i'' is nondecreasing, then Assumption* 2.7.3 *holds.*

Proof. Suppose that $a_i > -\infty$ and let $x_i, y_i \in (a_i, a_i + \epsilon)$ and $y_i < x_i$. Then

$$|h_i'(x_i) - h_i'(y_i)| \;=\; \int_{y_i}^{x_i} h_i''(z)\, dz \;\le\; h''(y_i)\,|x_i - y_i|\,.$$

Therefore, Assumption 2.7.3 holds with $\rho = 1$. The case $b_i < \infty$ is analogous. $\square$

Examples of functions h_i for which all of these assumptions hold are

- $h_i(x) = \frac{1}{2}x^2$, with $a_i = -\infty$, $b_i = +\infty$,
- $h_i(x) = -\log x$, with $a_i = 0$, $b_i = +\infty$,
- $h_i(x) = x \log x$, with $a_i = 0$, $b_i = +\infty$,
- $h_i(x) = x \log(e^x - 1)$, with $a_i = 0$, $b_i = +\infty$,
- $h_i(x) = x^\alpha - x^\beta$, for $\alpha \in [1, 2]$ and $\beta \in (0, 1)$, with $a_i = 0$, $b_i = +\infty$.

Finally, we note that for finite a_i we do *not* yet assume that $h_i(x_i)$ must approach a finite limit as $x_i \searrow a_i$, nor similarly for $x_i \nearrow b_i < +\infty$. Such an assumption is quite common in the theory of Bregman distances [11, 13, 9, 29], but, similarly to [21], it is not needed for the results of section 3 below. We will use it, however, in the variational inequality analysis of section 4.

2.2.2. φ-divergences. The φ-divergence regularizations have been studied in the context of proximal methods, for example, in [19], and more recently in [5, 3]. In these works, the box considered is the positive orthant, i.e., $B = \mathbb{R}^n_+$. An auxiliary strictly convex scalar function φ is used to define the distance d_i, but this time by

$$d_i(x_i, y_i) = y_i \varphi\!\left(\frac{x_i}{y_i}\right). \tag{2.5}$$

The following hypotheses can be used to guarantee Assumption 2.1 when $B = \mathbb{R}^n_+$.

ASSUMPTION 2.9. *The function $\varphi : \mathbb{R} \to (-\infty, +\infty]$ is such that*

2.9.1. *φ is closed and convex, with* $\operatorname{int}\operatorname{dom}\varphi = (0,+\infty)$*;*
2.9.2. *φ is twice differentiable on* $(0,+\infty)$*, with $\varphi''(t) > 0$ for all $t > 0$;*
2.9.3. $\varphi(1) = \varphi'(1) = 0$*;*
2.9.4. *φ is essentially smooth;*
2.9.5. *There exists a $\rho > 0$ such that $\rho\varphi'(t) \le \varphi''(1)(t-1)$ for all $t \ge 1$.*

Slight variations on these assumptions appear, for example, in [5, 3], together with the following examples:

- $\varphi(t) = t\log t - t + 1$;
- $\varphi(t) = -\log t + t - 1$;
- $\varphi(t) = 2(\sqrt{t} - 1)^2$.

The next lemma states that Assumption 2.9.5 above implies Assumption 2.1.4.

LEMMA 2.10. *Let* $(a_i, b_i) = (0,+\infty)$ *and d_i be defined as in* (2.5)*. Then Assumption* 2.1.4 *is equivalent to the existence of a $\rho > 0$ such that $\rho\varphi'(t) \le \varphi''(1)(t-1)$ for all $t \ge 1$.*

Proof. First we observe that

$$d_i'(x_i, y_i) = \varphi'\left(\frac{x_i}{y_i}\right),$$
$$d_i''(x_i, y_i) = \frac{1}{y_i}\varphi''\left(\frac{x_i}{y_i}\right),$$

and thus

$$d_i''(y_i, y_i) = \frac{1}{y_i}\varphi''(1).$$

Therefore, Assumption 2.1.4 reduces to

$$\exists\rho,\epsilon > 0: \quad 0 < y_i \le x_i < \epsilon \quad \Rightarrow \quad \rho\varphi'\left(\frac{x_i}{y_i}\right) \le \frac{1}{y_i}\varphi''(1)(x_i - y_i). \tag{2.6}$$

Taking $x_i \in (0,\epsilon)$, letting y_i range over $(0, x_i]$, and setting $t = x_i/y_i$, we obtain

$$\exists\rho > 0: \quad \rho\varphi'(t) \le \varphi''(1)(t-1)\ \forall t \ge 1. \tag{2.7}$$

Conversely, if (2.7) is true, then (2.6) holds for an arbitrary choice of $\epsilon > 0$. ∎

We note that in [5], one assumes that the iterations are of the form

$$0 \in \partial f(x^{k+1}) + \operatorname{diag}(\alpha^k)^{-1}\nabla_1 D(x^{k+1}, x^k),$$

for which each α_i^k is greater than c/x_i^k, c being a positive constant. In [2, 3], this property is guaranteed by redefining the distance measure to be

$$\tilde{d}_i(x_i, y_i) = y_i d_i(x_i, y_i) = y_i^2\varphi\left(\frac{x_i}{y_i}\right), \qquad \tilde{D}(x, y) = \sum_{i=1}^n \tilde{d}_i(x_i, y_i)$$

and assuming stepsizes bounded away from zero. In this case, the iteration is

$$0 \in \partial f(x^{k+1}) + \operatorname{diag}(\tilde{\alpha}^k)^{-1}\nabla_1\tilde{D}(x^{k+1}, x^k),$$

with $\liminf_{k\to\infty}\tilde{\alpha}_i^k > 0$ for all i. Defining $\alpha_i^k = \tilde{\alpha}_i^k/x_i^k$ and rewriting the iteration with respect to D instead of $\tilde{D}$, we recover the rule from [5].

It turns out that these techniques are a special case of our stepsize choice rule, which gives in the case of a φ-divergence that

$$\alpha_i^k \geq c d_i''(x_i^k, x_i^k) = \frac{c\varphi''(1)}{x_i^k},$$

which is identical if one redefines the constant factor c.

Thus, the reader should note that the class of φ-divergences described by Assumption 2.9 encompasses the regularizations studied in [5, 2, 3]. In particular, it includes the classes Φ_1 and Φ_2 described in [3].

However, the stepsize rule in the RPMVI is more stringent than the one in [5, 2, 3], as it also assumes that the stepsize is bounded away from zero. To overcome this slight restriction, we point out that the assumption $\alpha_i^k > c$ is used here only in the first part of the proof of Lemma 2.4, and it can be replaced by the assumption that $d_i''(y_i, y_i)$ is continuous and strictly positive over (a_i, b_i). This condition holds for φ-divergences, since $d_i''(y_i, y_i) = (1/y_i)\varphi''(1) > 0$ for all $y_i > 0$.

In this sense, the results here can be seen as extensions of those in [5, 2, 3].

3. Proximal minimization methods with rescaling. This section applies the analysis of the RPMVI method to the minimization problem (1.1). We leave Assumption 2.1 as a standing assumption; we also make the following standard regularity assumption, which in view of the barrier function properties of D, is required for any sensible application of (1.3).

ASSUMPTION 3.1. $\operatorname{dom} f \cap \operatorname{int} B \neq \emptyset$.

Note that, since $\operatorname{int} B$ is open, this assumption implies that $\operatorname{ri}\operatorname{dom} f \cap \operatorname{int} B \neq \emptyset$, which implies that $\operatorname{dom} \partial f \cap \operatorname{int} B \neq \emptyset$. Then, using [24, Theorem 23.8], one can show that the minimization problem (1.1) is equivalent to the variational inequality problem (1.2) with $T = \partial f$. Moreover, Assumption 2.2 holds.

Then, we specialize the RPMVI to the following algorithm.

RESCALING PROXIMAL MINIMIZATION METHOD (RPMM).

1. **Initialization:** *Choose* $c > 0$ *and* $\sigma \in [0, 1]$. *Choose nonnegative scalar sequences* $\{s_k\}$ *and* $\{z_k\}$ *with* $\sum_{k=1}^{\infty} s_k < \infty$ *and* $z_k \to 0$. *Let* $k = 0$ *and* $x^0 \in \operatorname{int} B$.
2. **Iteration:**
   (a) *Choose* $\alpha^k \in \mathbb{R}^n_{++}$ *such that* $\alpha_i^k \geq c \max\left\{1, d_i''(x_i^k, x_i^k)\right\}$ *for* $i = 1, \ldots, n$.
   (b) *Find* $x^{k+1}, e^{k+1} \in \mathbb{R}^n$ *such that*

$$e^{k+1} \in \partial f(x^{k+1}) + \operatorname{diag}(\alpha^k)^{-1} \nabla_1 D(x^{k+1}, x^k), \tag{3.1}$$

$$\left|e_i^{k+1}\right| \leq \frac{\sigma}{\alpha_i^k} \left|d_i'(x_i^{k+1}, x_i^k)\right| + \min\left\{\frac{s_{k+1}}{\|x^{k+1} - x^k\|}, z_{k+1}\right\}, \quad i = 1, \ldots, n, \tag{3.2}$$

   with the standing convention that $\min\left\{s_{k+1}/\|x^{k+1} - x^k\|, z_{k+1}\right\}$ *is* z_{k+1} *whenever* $x^{k+1} = x^k$.
   (c) *Let* $k = k + 1$*, and repeat the iteration.* □

Note that if one chooses $s_k, z_k = 0$ for all k, then (3.2) reduces to the "constructive" criterion

$$\left|e_i^{k+1}\right| \leq \frac{\sigma}{\alpha_i^k} \left|d_i'(x_i^{k+1}, x_i^k)\right|,$$

reminiscent of [29].

3.1. Convergence analysis. We start by showing that the iteration step is well defined if f is bounded below on B.

LEMMA 3.2. *If f is bounded below on B, then there is a unique point that solves the iteration step of the RPMM with $e^{k+1} = 0$. Thus, a solution to* (3.1)–(3.2) *exists if f is bounded below on B.*

Proof. Let ℓ be a lower bound of f on B. Given $\zeta \in \mathbb{R}$, the level set

$$\left\{ x \in B \;\middle|\; f(x) + \sum_{i=1}^{n} \frac{1}{\alpha_i^k} d_i(x_i, x_i^k) \le \zeta \right\} \subseteq \left\{ x \in B \;\middle|\; \sum_{i=1}^{n} \frac{1}{\alpha_i^k} d_i(x_i, x_i^k) \le \zeta - \ell \right\}.$$

This last set is a level set of $\sum_{i=1}^{n}(1/\alpha_i^k) d_i(\cdot, x_i^k)$ on B, which must be bounded, since by Assumption 2.1.1 this function attains its minimum at the unique point x^k [24, Corollary 8.7.1]. Therefore, $f(\cdot) + \sum_{i=1}^{n}(1/\alpha_i^k) d_i(\cdot, x_i^k)$ attains a minimum on B. The uniqueness of the minimum follows from the strict convexity of $D(\cdot, x^k)$. ∎

To apply the convergence analysis of the previous section to the sequence $\{x^k\}$ computed by the RPMM, it suffices to show that Assumption 2.3 holds. Verification of Assumption 2.3.1 is straightforward.

LEMMA 3.3. *With the definition*

$$\beta_k \stackrel{\text{def}}{=} \min\left\{ \frac{s_k}{\|x^k - x^{k-1}\|}, z_k \right\}$$

for all $k \ge 1$, Assumption 2.3.1 *holds for the RPMM.*

Proof. From the nonnegativity of $\{s_k\}$ and $\{z_k\}$, it follows that $\{\beta_k\}$ is also nonnegative. Since $z_k \to 0$, one also has $\beta_k \to 0$. Moreover, since $\sigma \in [0, 1]$,

$$\left|e_i^k\right| \le \frac{\sigma}{\alpha_i^{k-1}} \left|d_i'(x_i^k, x_i^{k-1})\right| + \beta_k \le \frac{1}{\alpha_i^{k-1}} \left|d_i'(x_i^k, x_i^{k-1})\right| + \beta_k$$

for all k, so Assumption 2.3.1 holds. ∎

As in (2.2), we define for all $k \ge 0$ and $i = 1, \ldots, n$,

$$\gamma_i^k = e_i^k - \frac{1}{\alpha_i^{k-1}} d_i'(x_i^k, x_i^{k-1}),$$

and let $\gamma^k \in \mathbb{R}^n$ be the vector with elements γ_i^k.

LEMMA 3.4. *$\gamma^k \in \partial f(x^k)$ and $\gamma_i^k(x_i^{k-1} - x_i^k) \ge -s_k$ for all $k \ge 0$ and $i = 1, \ldots, n$.*

Proof. The claim that $\gamma^k \in \partial f(x^k)$ follows from the definition of γ^k. For the second claim, we have, using the convexity of $d_i(\cdot, x_i^{k-1})$,

$$\begin{aligned}
\gamma_i^k(x_i^{k-1} - x_i^k) &= \left(e_i^k - \frac{1}{\alpha_i^{k-1}} d_i'(x_i^k, x_i^{k-1}) \right)(x_i^{k-1} - x_i^k) \\
&\ge -\frac{1}{\alpha_i^{k-1}} \underbrace{d_i'(x_i^k, x_i^{k-1})(x_i^{k-1} - x_i^k)}_{\le 0} - \left|e_i^k\right| \left|x_i^{k-1} - x_i^k\right| \\
&= \left(\frac{1}{\alpha_i^{k-1}} \left|d_i'(x_i^k, x_i^{k-1})\right| - \left|e_i^k\right| \right) \left|x_i^{k-1} - x_i^k\right|.
\end{aligned}$$

Using (3.2), it then follows that

$$\begin{aligned}\gamma_i^k(x_i^{k-1} - x_i^k) &\geq \left(\frac{1-\sigma}{\alpha_i^{k-1}}\left|d_i'(x_i^k, x_i^{k-1})\right| - \min\left\{\frac{s_k}{\|x^k - x^{k-1}\|}, z_k\right\}\right)\left|x_i^{k-1} - x_i^k\right| \\ &\geq -\min\left\{\frac{s_k}{\|x^k - x^{k-1}\|}, z_k\right\}\left|x_i^{k-1} - x_i^k\right| \\ &\geq -\min\left\{\frac{s_k}{\|x^k - x^{k-1}\|}, z_k\right\}\left\|x^{k-1} - x^k\right\| \\ &\geq -s_k. \qquad \square\end{aligned}$$

Before proving the next result, we state a helpful technical lemma.

LEMMA 3.5 (see [22, section 2.2]). *Suppose that* $\{a_k\}$, $\{\gamma_k\} \subset \mathbb{R}$ *are sequences such that* $\{a^k\}$ *is bounded below,* $\sum_{i=1}^{\infty} \gamma_k$ *exists and is finite, and the recursion* $a_{k+1} \leq a_k + \gamma_k$ *holds for all* k*. Then,* $\{a_k\}$ *is convergent.*

It is now possible to establish that Assumption 2.3.2 also holds.

LEMMA 3.6. *If* f *is bounded below on* B*, then* $\{f(x^k)\}$ *is convergent and*

$$\left|\gamma_i^k\right|\left|x_i^{k-1} - x_i^k\right| \to 0 \quad \forall i = 1, \ldots, n.$$

Hence Assumption 2.3.2 *holds for the RPMM.*

Proof. Using Lemma 3.4,

$$\begin{aligned}f(x^{k-1}) &\geq f(x^k) + \langle \gamma^k, x^{k-1} - x^k \rangle \\ &\geq f(x^k) - ns_k.\end{aligned}$$

Then, recalling that $\{s_k\}$ is summable, Lemma 3.5 implies that $\{f(x^k)\}$ is a convergent sequence. For $i = 1, \ldots, n$, we also have

$$\begin{aligned}f(x^{k-1}) &\geq f(x^k) + \langle \gamma^k, x^{k-1} - x^k \rangle \\ &\geq f(x^k) - (n-1)s_k + \gamma_i^k(x_i^{k-1} - x_i^k).\end{aligned}$$

Using Lemma 3.4 once again, it follows that

$$f(x^{k-1}) - f(x^k) + (n-1)s^k \geq \gamma_i^k(x_i^{k-1} - x_i^k) \geq -s_k.$$

Taking limits, we conclude that $\gamma_i^k(x_i^{k-1} - x_i^k) \to 0$. $\square$

Thus, Theorem 2.6 implies the optimality of all accumulation points of the sequence $\{x^k\}$. We strengthen this observation below.

THEOREM 3.7. *Suppose that Assumptions* 2.1 *and* 3.1 *hold and that* f *is bounded below on* B*. If* $\{x^k\}$ *has a limit point, then* $\{f(x^k)\}$ *converges to the infimum of* f *on* B*, and all limit points of* $\{x^k\}$ *will be minimizers of* f *on* B*. A condition that guarantees the existence of limit points of* $\{x^k\}$ *is the boundedness of the solution set, or any other level set of* f*.*

Proof. As just noted, Lemma 3.6 implies that Assumption 2.3.2 holds, and so Assumption 2.3 holds in its entirety. Assumption 2.1 holds by hypothesis, and, setting $T = \partial f$, Assumption 3.1 implies Assumption 2.2. Thus, the conclusions of Theorem 2.6 apply. Let $\bar{x}$ be a limit point of $\{x^k\}$, i.e., $x^k \to_{\mathcal{K}} \bar{x}$, for some infinite set $\mathcal{K} \subseteq \mathbb{N}$. Theorem 2.6 asserts that $0 \in \partial f(\bar{x}) + N_B(\bar{x})$; by Assumption 3.1, $\bar{x}$ is a minimizer of f on B. Moreover, since Lemma 2.5 states that $\{\gamma^k\}_{\mathcal{K}}$ is bounded, and since $\{f(x^k)\}$ is convergent by Lemma 3.6,

$$\min_{x \in B} f(x) \;=\; f(\bar{x}) \;\geq\; f(x^k) + \sum_{i=1}^{n} \gamma_i^k (\bar{x}_i - x_i^k) \underset{\mathcal{K}}{\rightarrow} \lim_{k\to\infty} f(x^k) \;\geq\; f(\bar{x}).$$

Therefore, $\lim_{k\to\infty} f(x^k) = f(\bar{x})$.

Finally, the boundedness of any level set of a proper closed convex function implies boundedness of all level sets [24, Corollary 8.7.1], and Lemma 3.6 states that $\{f(x^k)\}$ is convergent; consequently it is bounded. Thus, $\{x^k\}$ is also bounded and has limit points. □

3.2. Multiplier methods. We now discuss applying the RPMM to the dual of the convex program (1.5) to obtain multiplier methods. The use of proximal methods to derive multiplier methods for constrained convex optimization is a now-classical subject and may be traced to the seminal paper [26]. In the context of generalized proximal methods, applications can be found, for example, in [30, 13, 19, 21, 31, 3, 17]. In this section, we consider only the case in which the proximal step is done exactly, i.e., we will let $e^k = 0$ for all k, as in [30, 13, 19, 17]. Unfortunately, our approximate-step acceptance rule for the RPMM does not translate directly to an easily verifiable acceptance criterion for an approximate solution of the penalized problem (3.5) below. However, partial results in this direction may be obtained under stringent assumptions on the original problem (1.5); see Appendix B. A criterion in the spirit of (3.2) that does not depend on such assumptions is the subject of ongoing research [15]. We further observe that the approximation criteria of [17, 29] also do not translate readily to a multiplier method setting. On the other hand, under the assumption that the primal objective function g_0 is strongly convex, [26, 21, 3] present some inexact multiplier methods based on a rather different acceptance rule involving optimizing the augmented Lagrangian function to within some tolerance ϵ of its minimum value.

Consider the convex problem (1.5), and let δ_C denote the indicator function of a convex set C. Then we define f to be minus the dual function associated with (1.5), plus $\delta_{\mathbb{R}^n_+}$:

$$f(x) \overset{\text{def}}{=} -\inf_{y \in \mathbb{R}^m} \left\{ g_0(y) + \sum_{i=1}^{n} x_i g_i(y) \right\} + \delta_{\mathbb{R}^n_+}(x). \tag{3.3}$$

The dual problem to (1.5) is then equivalent to the minimization of f. Furthermore, we assume the following.

Assumption 3.8.

3.8.1. *The primal problem* (1.5) *has a finite optimal value, and it conforms to the Slater condition.*

3.8.2. *For all* $i = 1, \ldots, n$, *the generalized distances* d_i *conform to Assumption* 2.1 *for* $a_i \leq 0,\ b_i = +\infty$.[2]

3.8.3. *There is an* $\bar{x} > 0$ *such that* $\bar{x} \in \operatorname{dom} f$, *where* f *is as defined in* (3.3).

This assumption has the following consequences: Assumption 3.8.1 implies that the dual solution set is nonempty and bounded [16] and that there is no duality gap. Assumption 3.8.3 implies that Assumption 3.1 holds for f as defined by (3.3).

[2] The case $a_i = -\infty$ is of interest because it includes the classical method of multipliers for problems with inequality constraints [26], along with various extensions described in [13, 20].

Under Assumption 3.8, if we fix $e^k = 0$ for all k, then each iterate x^{k+1} of the RPMM applied to the negative dual functional f may be calculated by the following multiplier method whenever the unconstrained problems (3.5) have solutions:

$$\alpha_i^k \geq c \max\left\{1, d_i''(x_i^k, x_i^k)\right\}, \quad i = 1, \ldots, n, \tag{3.4}$$

$$y^{k+1} \in \operatorname*{Arg\,min}_{y \in \mathbb{R}^m} \left\{ g_0(y) + \sum_{i=1}^{n} \frac{1}{\alpha_i^k} d_i^{\oplus}\left(\alpha_i^k g_i(y), x_i^k\right) \right\}, \tag{3.5}$$

$$x_i^{k+1} = \nabla_1 d_i^{\oplus}\left(\alpha_i^k g_i(y^{k+1}), x_i^k\right), \quad i = 1, \ldots, n. \tag{3.6}$$

Here, "$\oplus$" denotes the monotone conjugate [24, p. 111] with respect to the first argument, that is, $d_i^{\oplus}(u_i, w_i) = \sup_{x_i \geq 0}\{u_i x_i - d_i(x_i, w_i)\}$.[3] Theorem 3.10 below gives conditions guaranteeing that a y^{k+1} satisfying (3.5) exists.

We relegate to Appendix A the technical aspects of the proof of the equivalence of (3.4)–(3.6) to the RPMM applied to the f defined in (3.3), since they are very similar to earlier proofs for various special cases of (3.5)–(3.6), for example in [30, 13, 19, 21, 17]. In particular, Corollary A.4 establishes the equivalence of the two calculations.

Given this equivalence, Theorem 3.7 asserts the subsequential convergence of the sequence $\{x^k\}$ to a dual solution of (1.5). For the primal sequence, however, it has historically been harder to prove good behavior. For example, in the case of Bregman distances, a guarantee of feasibility of primal accumulation points has relied on stringent assumptions like $\mathbb{R}^n_+ \subset \operatorname{int} B$, as in [13], or strict complementarity [18].

In the case of the RPMM, with its strong stepsize restrictions, the feasibility, and therefore optimality, of accumulation points of $\{y^k\}$ is easily demonstrated.

THEOREM 3.9. *Suppose that Assumption* 3.8 *holds. Pick a scalar $c > 0$, let $x^0 \in \mathbb{R}^n_{++}$, and suppose that it is possible to obtain a sequence $\{(\alpha^k, x^k, y^k)\}$ that obeys the recursions* (3.4)–(3.6)*. Then, $\{x^k\}$ is bounded and all its accumulation points are solutions of the dual of* (1.5)*. Moreover,*

$$\limsup_{k \to \infty} g_i(y^k) \leq 0, \qquad i = 1, \ldots, n, \tag{3.7}$$

$$\lim_{k \to \infty} \sum_{i=1}^{n} x_i^k g_i(y^k) = 0, \tag{3.8}$$

and $\{g_0(y^k)\}$ converges to the optimal value of the primal problem (1.5)*. Therefore, any accumulation point of $\{y^k\}$ solves the primal problem.*

Proof. As shown in Corollary A.4, the sequence $\{x^k\}$ is the same as would be computed by using the RPMM to solve the dual problem, that is, to minimize f. In particular, $\{x^k\}$ and all its limit points must be nonnegative. Moreover, the Slater condition implies that the dual function has bounded level sets. Then, the boundedness of $\{x^k\}$ and the optimality of its limit points follow from Theorem 3.7.

Let us analyze the primal sequence. For each $i = 1, \ldots, n$, (3.6) implies that

$$g_i(y^k) = \frac{1}{\alpha_i^{k-1}} d_i'(x_i^k, x_i^{k-1}) + \zeta_i^k,$$

where $\zeta_i^k \in \partial \delta_{\mathbb{R}_+}(x_i^k)$. Hence, $\zeta_i^k - g_i(y^k)$ plays the same role as γ_i^k in (2.2), with $e_i^k = 0$.

[3]The classical conjugate ψ^* of a function ψ is defined [24, Chapter 12] via $\psi^*(y) = \sup_{x \in \mathbb{R}^n} \{\langle x, y\rangle - \psi(x)\}$ for any $\psi : \mathbb{R}^n \to (\infty, +\infty]$. The monotone conjugate of ψ is then the classical conjugate of $\psi + \delta_{\mathbb{R}^n_+}$, that is, $\psi^{\oplus}(y) = \sup_{x \geq 0} \{\langle x, y\rangle - \psi(x)\}$.

Let $\{x^k\}_{\mathcal{K}}$ be any convergent subsequence of $\{x^k\}$, and $\bar{x}$ the respective accumulation point, $x^k \to_{\mathcal{K}} \bar{x}$. Lemma 2.4 implies that

$$\begin{array}{rclclr} 0 & = & \lim\limits_{k\to_{\mathcal{K}}\infty} g_i(y^k) - \zeta_i^k & = & \lim\limits_{k\to_{\mathcal{K}}\infty} g_i(y^k) & \text{if } \bar{x}_i > 0, \\ 0 & \geq & \limsup\limits_{k\to_{\mathcal{K}}\infty} g_i(y^k) - \zeta_i^k & \geq & \limsup\limits_{k\to_{\mathcal{K}}\infty} g_i(y^k) & \text{if } \bar{x}_i = 0. \end{array} \tag{3.9}$$

As $\{x^k\}$ is bounded, the above relations imply that

$$0 \geq \limsup_{k\to\infty} g_i(y^k), \qquad i = 1, \dots, n. \tag{3.10}$$

Now, suppose for the purposes of contradiction that (3.8) does not hold. Then, for some $i = 1, \dots, n$, there must be an infinite set $\mathcal{K} \subset \mathbb{N}$ and an $\epsilon > 0$ such that

$$\forall k \in \mathcal{K}, \quad \left| x_i^k g_i(y^k) \right| \geq \epsilon. \tag{3.11}$$

Since $\{x^k\}$ is bounded, there exists a refined subsequence $\mathcal{K}' \subseteq \mathcal{K}$ such that $\{x^k\}_{\mathcal{K}'}$ is convergent, with limit $\bar{x} \geq 0$. If $\bar{x}_i > 0$, then (3.11) contradicts (3.9). If $\bar{x}_i = 0$, then (3.11) and (3.9) imply that $g_i(y^k) \to_{\mathcal{K}'} -\infty$. Since Lemma 2.5 asserts that $\{\zeta_i^k - g_i(y^k)\}$ is bounded, we can conclude that $\zeta_i^k \to_{\mathcal{K}'} -\infty$. However, this divergence would imply that x_i^k should be 0 for infinitely many $k \in \mathcal{K}' \subseteq \mathcal{K}$, once again a contradiction of (3.11). Therefore,

$$\lim_{k\to\infty} x_i^k g_i(y^k) = 0, \qquad i = 1, \dots, n, \tag{3.12}$$

and (3.8) holds.

Finally, we prove that $\{g_0(y^k)\}$ converges to the optimal value. We may use (3.5), (3.6), and the chain rule to see that y^k minimizes the Lagrangian corresponding to the primal problem with the fixed multiplier x^k. Hence,

$$g_0(y^k) + \sum_{i=1}^{n} x_i^k g_i(y^k) = -f(x^k). \tag{3.13}$$

Let $-f^*$ denote the dual optimal value, which is equal to the primal optimal value since there is no duality gap. Theorem 3.7 states that $f(x^k) \to f^*$. Taking limits in (3.13) and using (3.12), it follows that

$$\lim_{k\to\infty} g_0(y^k) = -f^*. \tag{3.14}$$

The feasibility and optimality of the accumulation points of $\{y^k\}$ are then consequences of the continuity of g_i, $i = 0, \dots, n$, (3.10), and (3.14). $\square$

Finally, it is natural to seek conditions under which the penalized subproblems (3.5) must have solutions and the primal sequence $\{y^k\}$ is bounded. The following result addresses these questions under the standard assumption of a bounded solution set.

THEOREM 3.10. *Suppose that the primal solution set is bounded. Given any* $\alpha^k > 0$ *and* (x^k, y^k), *there exist* (x^{k+1}, y^{k+1}) *satisfying the recursions* (3.5)–(3.6). *Moreover, the primal sequence* $\{y^k\}$ *is bounded.*

Proof. For the first assertion, it suffices to show that the penalized problems (3.5) have solutions. Given any closed proper convex function ψ, we define its *recession* *function* ψ_∞ via $\psi_\infty(d) = \lim_{\lambda\to\infty} (\psi(z+\lambda d) - \psi(z))/\lambda$, where $z \in \operatorname{dom}\psi$ may be chosen arbitrarily [24, Theorem 8.5]. The boundedness of the primal solution set is equivalent [7, section 5.3] to

$$(g_i)_\infty(d) \le 0 \quad \forall\, i = 1,\dots,n \quad \Rightarrow \quad (g_0)_\infty(d) > 0. \tag{3.15}$$

Thus, the existence of a solution to (3.5) is a corollary of Lemma A.5 in the appendix, along with the sum rule for recession functions [24, Theorem 9.3].

We now prove that $\{y^k\}$ is bounded. Theorem 3.9 shows that the sequences $\{g_i(y^k)\}$, $i = 1,\dots,n$, are bounded above. From (3.15), unboundedness of $\{y^k\}$ would imply that $g_0(y^k) \to_{\mathcal{K}} \infty$ for some infinite $\mathcal{K} \subseteq \mathbb{N}$. But such unboundedness would contradict $g_0(y^k)$'s convergence to the optimal value. ∎

We remark that the penalty parameter adjustment rule (3.4), as discussed in section 2.2.2, essentially subsumes, in a context broader than φ-divergences, the corresponding rules described in [32] for the exponential method of multipliers and in [5, 3, 4] for a general φ-divergence setting.

We end this section by giving some examples of $d_i^\oplus$ functions that may be derived from separable Bregman distances (see section 2.2.1). Further examples may be obtained from [21, 18, 28]. For a Bregman-derived distance, we have $d_i(x_i, w_i) = h_i(x_i) - h_i(w_i) - h'(w_i)(x_i - w_i)$, whence

$$\begin{aligned} d_i^\oplus(u_i, w_i) &= \sup_{x_i \ge 0} \{u_i x_i - (h_i(x_i) - h_i(w_i) - h'(w_i)(x_i - w_i))\} \\ &= \sup_{x_i \ge 0} \{(u_i + h'(w_i))\, x_i - h_i(x_i)\} + h_i(w_i) - w_i h'(w_i) \\ &= h^\oplus(h'(w_i) + u_i) + h_i(w_i) - w_i h'(w_i), \end{aligned}$$

where $h^\oplus$ denotes the standard monotone conjugate of h. Note that when such a $d_i^\oplus(u_i, w_i)$ is used in the minimization operation in (3.5), the additive terms $h_i(w_i) - w_i h'(w_i)$ are constant and may be discarded. The following examples may now be easily verified:

- If $h_i(x_i) = \frac{1}{2}x_i^2$, then $d_i^\oplus(u_i, w_i) = \frac{1}{2}(\max\{u_i + w_i, 0\}^2 - w_i^2)$, where the $-w_i^2$ term may be disregarded; this choice gives the classical quadratic method of multipliers for inequality constraints.
- If $h_i(x_i) = x_i \log x_i - x_i$, then $d_i^\oplus(u_i, w_i) = w_i e^{u_i} - w_i$, where the $-w_i$ term may be disregarded, yielding the exponentional method of multipliers.
- If $h_i(x_i) = -\log x_i$, then $d_i^\oplus(u_i, w_i) = -\log(1 - w_i u_i)$.

4. Bregman interior point proximal methods for variational inequalities. We now turn our attention to the box-constrained variational inequality problem (1.2), where $T : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ is a (possibly set-valued) maximal monotone operator. In this section, we confine ourselves to Bregman distances, as defined in section 2.2.

We augment Assumption 2.2 as follows.

ASSUMPTION 4.1. *T is maximal monotone, the solution set of* (1.2) *is nonempty, and there exists some $\widetilde{x} \in \operatorname{dom} T \cap \operatorname{int} B$.*

Our goal is to show convergence of an approximate version of the iteration (1.4), without further conditions on T. We modify and extend Assumption 2.7 as follows.

ASSUMPTION 4.2. *For $i = 1,\dots,n$, the functions $h_i : \mathbb{R} \to (-\infty,\infty]$ have the same properties specified in Assumption* 2.7*, and furthermore, h_i is continuous on $[a_i, b_i] \cap \mathbb{R}$. Moreover, defining $h(x) = \sum_{i=1}^n h_i(x_i)$ and $D_h(x, y) = \sum_{i=1}^n h_i(x_i) - h_i(y_i) + h_i'(y_i)(x_i - y_i)$,*

4.2.1. *for all* $x \in B$ *and* $\alpha \in \mathbb{R}$, *the level set* $\{y \in \operatorname{int} B \mid D_h(x,y) \le \alpha\}$ *is bounded;*
4.2.2. *if* $\{x^k\} \subset \operatorname{int} B$ *converges to* $x \in \mathbb{R}^n$, *then* $\lim_{k\to\infty} D_h(x,x^k) = 0$;
4.2.3. $\operatorname{rge} h' = \mathbb{R}$.

Note that at finite a_i's and b_i's, the corresponding h_i is now required to take a finite value. The algorithm can now be stated.

Box Interior Proximal Point Algorithm (BIPPA).

1. **Initialization:** *Let* $k = 0$, *and fix some scalar* $c > 0$. *Let* $x^0 \in \operatorname{int} B$.
2. **Iteration:** *Choose* α_k *such that* $\alpha_k \ge c \max\{1, h_1''(x_1^k), \ldots, h_n''(x_n^k)\}$. *Find vectors* $x^{k+1}, e^{k+1} \in \mathbb{R}^n$ *such that*

$$\begin{aligned} e^{k+1} &\in T(x^{k+1}) + \frac{1}{\alpha_k}\nabla_1 D_h(x^{k+1}, x^k) \\ &= T(x^{k+1}) + \frac{1}{\alpha_k}\big(\nabla h(x^{k+1}) - \nabla h(x^k)\big). \end{aligned} \tag{4.1}$$

Let $k = k + 1$ *and repeat the iteration.* □

4.1. Convergence analysis. First, we cite a result showing that the iteration step of BIPPA is well defined.

Lemma 4.3 (See [13, Theorem 4(i)]). *Under Assumption* 4.2, *there is a unique point* x^{k+1} *that solves the iteration step* (4.1) *of the BIPPA with* $e^{k+1} = 0$.

We note that it is shown in the unpublished dissertation [28] that (4.1) has a unique exact solution even if Assumption 4.2.3 does not hold. This result permits one to dispense completely with Assumption 4.2.3. However, the proof, while essentially a minor modificiation of that of [1, Theorem A.1], is quite involved, so we do not include it here.

To guarantee the convergence of the BIPPA, we must assume some vanishing behavior for $\{e^k\}$; we will use the assumptions of [14]. Although not as general as the criterion used in RPMM, these conditions are better suited to our analysis, since they will permit us to use properties associated with Fejér monotonicity, and are still feasible to enforce computationally.

Assumption 4.4 (See [14]). *The error sequence* $\{e^k\}$ *conforms to*

$$\sum_{k=0}^{\infty} \alpha_k \left\|e^{k+1}\right\| < +\infty;$$

$$\sum_{k=0}^{\infty} \alpha_k \langle e^{k+1}, x^{k+1}\rangle \text{ exists and is finite.}$$

Note that this assumption implies that $\|e^k\| \to 0$, and therefore Assumption 2.3.1 holds with $\beta_k = \|e^k\|_\infty$. We now state some necessary lemmas.

Lemma 4.5 (See [14, Lemma 2]). *Let* $z \in (T + N_B)^{-1}(0)$. *Then, for all* $k \ge 0$,

$$D_h(z, x^{k+1}) \le D_h(z, x^k) - D_h(x^{k+1}, x^k) + \alpha_k \langle e^{k+1}, x^{k+1} - z\rangle. \tag{4.2}$$

Lemma 4.6. *If Assumption* 4.4 *holds, then the sequence* $\{x^k\}$ *is bounded and* $D_h(x^{k+1}, x^k) \to 0$.

Proof. The result will follow from [14, Lemma 3] once we show that, for $z \in (T + N_B)^{-1}(0)$,

$$E(z) \stackrel{\text{def}}{=} \sum_{i=0}^{\infty} \alpha_k \langle e^{k+1}, x^{k+1} - z\rangle$$

exists and is finite. But

$$\sum_{i=0}^{\infty} \left|\alpha_k \langle e^{k+1}, z\rangle\right| \le \sum_{i=0}^{\infty} \alpha_k \|e^{k+1}\|\|z\|,$$

and Assumption 4.4 implies that the right-hand side of this relation is finite. Hence, $\sum_{i=0}^{\infty} \alpha_k \langle e^{k+1}, z\rangle$ exists and is finite. Using Assumption 4.4 once more, we conclude that $E(z)$ exists and is finite. $\square$

We also use a key result from Solodov and Svaiter [29].

THEOREM 4.7 (See [29, Theorem 2.4]). *Let h_i satisfy Assumption 4.2. Given two sequences $\{x^k\} \subset B$ and $\{y^k\} \subset \operatorname{int} B$, either one of which is convergent, with $\lim_{k\to\infty} D_h(x^k, y^k) = 0$, then the other sequence also converges to the same limit.*

This theorem implies that $h(x) = \sum_{i=1}^{n} h_i(x_i)$ is a *Bregman function* in the classical sense [8, 10]. Using Theorem 4.7 and Lemma 4.6, we derive the following.

COROLLARY 4.8. *Under Assumptions 4.1, 4.2, and 4.4, $\{x^k\}$ has at least one limit point. Moreover, if for some infinite set $\mathcal{K} \subseteq \mathbb{N}$ we have $x^k \to_{\mathcal{K}} \bar{x}$, then $x^{k-1} \to_{\mathcal{K}} \bar{x}$. Therefore, Assumption 2.3.2 holds.*

Before presenting the main convergence theorem for the BIPPA, we present a final technical lemma that will help us to prove the uniqueness of the accumulation points of $\{x^k\}$.

LEMMA 4.9. *Under Assumption 4.4, for all $z \in (T + N_B)^{-1}(0)$, $D_h(z, x^k)$ converges to a value in $[0, +\infty)$ which we will denote by $d(z)$.*

Proof. Consider any $z \in (T + N_B)^{-1}(0)$. Then Lemma 4.5 implies that (4.2) holds. Using Assumption 4.4 and $D_h(x^{k+1}, x^k) \ge 0$, the hypotheses of Lemma 3.5 are satisfied with $a_k = D_h(z, x^k)$ and $\gamma_k = \alpha_k\langle e^{k+1}, x^{k+1} - z\rangle$. Therefore, $\{D_h(z, x^k)\}$ converges, necessarily to a nonnegative value. $\square$

Now, the main convergence theorem follows.

THEOREM 4.10. *Under Assumptions 4.1, 4.2, and 4.4, $\{x^k\}$ converges to a solution of $0 \in T(x) + N_B(x)$.*

Proof. Let $\bar{x}$ be an accumulation point of $\{x^k\}$, i.e., $x^k \to_{\mathcal{K}} \bar{x}$, for some infinite set $\mathcal{K} \subseteq \mathbb{N}$. Such a point exists by Lemma 4.6. From Theorem 2.6, $0 \in T(\bar{x}) + N_B(\bar{x})$.

We now prove the uniqueness of the limit point: from Assumption 4.2.2, we know that $D_h(\bar{x}, x^k) \to_{\mathcal{K}} 0$. Then, $d(\bar{x})$, as defined in Lemma 4.9, is zero. Suppose that $\{x^k\}$ has another accumulation point $x^k \to_{\mathcal{K}'} x'$ for some infinite set $\mathcal{K}' \subseteq \mathbb{N}$. We then have that $D_h(\bar{x}, x^k) \to_{\mathcal{K}'} d(\bar{x}) = 0$, and it follows from Theorem 4.7 that $x' = \bar{x}$. $\square$

Another possible application of our fundamental analysis is to try to generalize to solutions of (1.2) the idea of adding the square of the Euclidean norm and an arbitrary generalized distance to obtain Fejér monotonicity, as in [2, 3] for the special case of φ-divergences. The difficulty here is to generalize the condition that defines the class Φ_2 in [3]. This topic is the subject of ongoing research.

Appendix A. Relationship between multiplier and proximal methods. This appendix proves that the RPMM may be applied to minus the dual functional associated with (1.5) via the multiplier method (3.5)–(3.6).

The proof is very similar to the derivation of a special case presented in [17, section 4.2]. Therefore, we will follow the steps in [17], changing notation whenever necessary to suit the present setting.

In particular, as in (1.5), $g_0 : \mathbb{R}^m \to \mathbb{R}$ is the primal objective and $g : \mathbb{R}^m \to \mathbb{R}^n$ is the constraint function, with components g_i, $i = 1, \ldots, n$. We assume that the

g_i, $i = 0, \ldots, n$, are differentiable convex functions and that (1.5) is feasible. Let f be the negative dual function defined in (3.3), which we assume to be somewhere finite. Note that, since f is the pointwise supremum of a nonempty collection of affine functions, it cannot take the value $-\infty$, and is therefore proper. Let $v(\cdot)$ denote the right-hand-side perturbation function associated with the optimization problem (1.5):

$$\forall u \in \mathbb{R}^n,\ v(u) \overset{\text{def}}{=} \inf\{g_0(y) \mid y \in \mathbb{R}^m,\ g(y) \leq u\}.$$

It is well known that for all $x \in \mathbb{R}^n$, $f(x) = v^*(-x)$; see [25, Example 1 and Theorem 7].

We also assume the following throughout this section.

Assumption A.1. *$D : \mathbb{R}^n \to (-\infty, +\infty]$ is a closed, proper, and strictly convex function such that* $\operatorname{ri}(\operatorname{dom} D \cap \mathbb{R}^n_{++}) \cap \operatorname{ri}\operatorname{dom} f \neq \emptyset$.

$D^\oplus$ denotes the monotone conjugate of D, that is, the convex conjugate of $D + \delta_{\mathbb{R}^n_+}$, and $D^\oplus(g(\cdot))$ denotes the usual composition of $D^\oplus$ and g.

We start by proving a slight modification of [6, equation (4.41)] which plays a fundamental role in our analysis.

Lemma A.2. *If Assumption* A.1 *holds, then*

$$\inf_{y \in \mathbb{R}^m} \{g_0(y) + D^\oplus(g(y))\} = \inf_{u \in \mathbb{R}^n} \{v(u) + D^\oplus(u)\} = \sup_{x \geq 0} \{-f(x) - D(x)\}.$$

Proof. The definition of $D^\oplus$ implies that if $a \geq b$, then $D^\oplus(a) \geq D^\oplus(b)$, i.e., it is nondecreasing.[4] Therefore,

$$\begin{aligned}
\inf_{y \in \mathbb{R}^m} \{g_0(y) + D^\oplus(g(y))\} &= \inf_{y \in \mathbb{R}^m} \{g_0(y) + D^\oplus(g(y))\} \\
&= \inf_{u \in \mathbb{R}^n} \inf_{\substack{y \in \mathbb{R}^m \\ g(y) \leq u}} \{g_0(y) + D^\oplus(g(y))\} \\
&\leq \inf_{u \in \mathbb{R}^n} \inf_{\substack{y \in \mathbb{R}^m \\ g(y) \leq u}} \{g_0(y) + D^\oplus(u)\} \\
&= \inf_{u \in \mathbb{R}^n} \{v(u) + D^\oplus(u)\}.
\end{aligned}$$

On the other hand,

$$\begin{aligned}
\inf_{y \in \mathbb{R}^m} \{g_0(y) + D^\oplus(g(y))\} &\geq \inf_{y \in \mathbb{R}^m} \{v(g(y)) + D^\oplus(g(y))\} \\
&\geq \inf_{u \in \mathbb{R}^n} \{v(u) + D^\oplus(u)\}.
\end{aligned}$$

Hence, the first equality is proved.

Finally, we use Fenchel's duality theorem [24, Theorem 31.1] and the fact that for all $x \in \mathbb{R}^n$, $-f(x) = -v^*(-x)$, to assert that

$$\inf_{u \in \mathbb{R}^n} \{v(u) + D^\oplus(u)\} = \sup_{x \in \mathbb{R}^n} \{-f(x) - D(x) - \delta_{\mathbb{R}^n_+}(x)\} = \sup_{x \geq 0} \{-f(x) - D(x)\}. \qquad \square$$

Theorem A.3. *Suppose that Assumption* A.1 *holds. Suppose that the (strictly convex) function $f + D$ has the minimizer $\bar{x}$ over $\mathbb{R}^n$, and that there is $\bar{y}$ such that*

$$\bar{y} \in \operatorname*{Arg\,min}_{y \in \mathbb{R}^m} \{g_0(y) + D^\oplus(g(y))\}. \tag{A.1}$$

[4] This inequality is a simple consequence of the definition of the convex conjugate; see [17, Proposition 3].

Then $\bar{x} = \nabla D^{\oplus}(g(\bar{y}))$*.*

Proof. From the definition of $\bar{y}$ and the nondecreasing property of $D^{\oplus}$, we have that $g_0(\bar{y}) = v(g(\bar{y}))$. Then, defining $\bar{u} = g(\bar{y})$, Lemma A.2 states that

$$v(\bar{u}) + D^{\oplus}(\bar{u}) = \inf_{u \in \mathbb{R}^n} \big\{ v(u) + D^{\oplus}(u) \big\} = \sup_{x \in \mathbb{R}^n} \big\{ -f(x) - D(x) - \delta_{\mathbb{R}^n_+}(x) \big\}.$$

Hence, we may use [23, Theorem 2] to verify that

$$\bar{x} = \arg\max_{x \in \mathbb{R}^n} \big\{ -f(x) - D(x) - \delta_{\mathbb{R}^n_+}(x) \big\} \in \partial D^{\oplus}(\bar{u}) = \big\{ \nabla D^{\oplus}\big(g(\bar{y})\big) \big\},$$

where the last equality is a consequence of $D^{\oplus}$ being the convex conjugate of a strictly convex function, meaning that $\partial D^{\oplus}$ is single-valued throughout its domain [24, Chapter 26]. ∎

COROLLARY A.4. *Let* $d_i,\ i = 1, \ldots, n$, *conform to Assumption* 2.1*, with* $B \supseteq \mathbb{R}^n_+$*. Suppose that there is an* $\bar{x} > 0$ *such that* $\bar{x} \in \operatorname{dom} f$*. Given* $x^k \in \operatorname{dom} f$*, there is a unique point* x^{k+1} *that satisfies* (1.3)*. Moreover, if there exists a point* y^{k+1} *satisfying* (3.5)*, then* (3.6) *holds.*

Proof. Since we assumed that primal problem (1.5) is feasible, the weak duality theorem asserts that the dual objective function is bounded above. Hence, f is bounded below and the existence and uniqueness of x^{k+1} is given by Lemma 3.2.

Finally, let $D(\cdot) = \sum_{i=1}^n \frac{1}{\alpha_i^k} d_i(\cdot, x_i^k)$. Then, for all $u \in \mathbb{R}^n$,

$$D^{\oplus}(u) = \sum_{i=1}^n \frac{1}{\alpha_i^k} d_i^{\oplus}(\alpha_i^k u, x_i^k),$$

as the convex conjugate of a separable function is just the sum of the convex conjugates of its components. Also, if we define $h_\alpha(x) = \alpha h(x)$ for some positive number and convex function h, we have

$$h_\alpha^*(x) = \alpha h^*\Big(\frac{x}{\alpha}\Big).$$

The result then follows from the previous theorem. ∎

Now, we analyze the existence of solutions to the penalized problem (A.1). In order to do so, we will use the notation

$$P(\cdot) \stackrel{\text{def}}{=} D^{\oplus}\big(g(\cdot)\big).$$

Note that P is closed because $D^{\oplus} \stackrel{\text{def}}{=} (D + \delta_{\mathbb{R}^n_+})^*$ must be closed [24, Theorem 12.2].

LEMMA A.5. *Suppose that Assumption* A.1 *holds,* $\operatorname{dom} D \supseteq \mathbb{R}^n_{++}$*, and* D *is bounded below. Let* $\mathcal{R} = \{d \mid (g_i)_\infty(d) \le 0,\ i = 1, \ldots, n\}$*. Then*

$$P_\infty(d) = \begin{cases} 0 & \text{if } d \in \mathcal{R} \\ +\infty & \text{otherwise.} \end{cases}$$

Proof. Let $\bar{y}$ be a feasible point for (1.5). From the definition of P and $g(\bar{y}) \le 0$,

$$P(\bar{y}) \;=\; \sup_{z \ge 0} \big\{ \langle z, g(\bar{y}) \rangle - D(z) \big\} \;\le\; \sup_{z \ge 0} \big\{ -D(z) \big\}.$$

Hence, as D is bounded below, $\bar{y} \in \operatorname{dom} P$. Therefore, since P is a closed convex function,

$$P_\infty(d) = \lim_{t \to \infty} \frac{P(\bar{y} + td) - P(\bar{y})}{t} \tag{A.2}$$

for all $d \in \mathbb{R}^n$.

As $\mathbb{R}^n_{++} \subseteq \operatorname{dom}(D+\delta_{\mathbb{R}^n_+})$ and is an open set, we have $\mathbb{R}^n_{++} \subseteq \operatorname{dom} \partial(D+\delta_{\mathbb{R}^n_+})$, and from $D^\oplus = (D+\delta_{\mathbb{R}^n_+})^*$ we then obtain $\mathbb{R}^n_{++} \subseteq \operatorname{rge} \partial D^\oplus$. Thus, for all $x > 0$, there exists some $\gamma \in \mathbb{R}^n$ with $x \in \partial D^\oplus(\gamma)$. So, for all $x > 0$, there exists some $\gamma \in \mathbb{R}^n$ such that

$$\forall\, t > 0,\, \forall\, d \in \mathbb{R}^m :\; D^\oplus(\gamma) + \langle x, g(\bar{y}+td) - \gamma\rangle - P(\bar{y}) \le P(\bar{y}+td) - P(\bar{y}).$$

Dividing both sides by t and taking limits as $t \to \infty$, (A.2) implies that for all $x \in \mathbb{R}^n_{++}$,

$$\sum_{i=1}^n x_i (g_i)_\infty(d) \le P_\infty(d). \tag{A.3}$$

This summation is well defined since the recession function of a closed proper convex function is also proper [27, Corollary 3.27]. Taking the limit as $x \to 0$, we may conclude that

$$\forall d \in \mathbb{R}^n,\; 0 \le P_\infty(d). \tag{A.4}$$

Now, we consider two cases:

1. $d \in \mathcal{R}$. Then
$$g(\bar{y}+td) \le g(\bar{y}) \quad \forall\, t \ge 0 \quad \Rightarrow \quad P(\bar{y}+td) - P(\bar{y}) \le 0 \quad \forall\, t \ge 0.$$
Dividing both sides by t and taking limits as $t \to \infty$, it follows that $P_\infty(d) \le 0$. Hence, using (A.4), $P_\infty(d) = 0$.
2. $d \notin \mathcal{R}$. Without loss of generality, let us assume that $(g_1)_\infty(d) > \zeta > 0$. Let $x = (M, 1, 1, \ldots, 1) \in \mathbb{R}^n$. From (A.3), it follows that
$$\forall M > 0,\; M\zeta + \sum_{i=2}^m (g_i)_\infty(d) \le P_\infty(d).$$
Since $(g_i)_\infty(d) > -\infty$, $i = 1, \ldots, m$, we can take the limit as $M \to \infty$ and conclude that $P_\infty(d) = +\infty$. ∎

Appendix B. Inexact multiplier methods. In this appendix, we present conditions that make it possible to use the RPMM acceptance criterion (3.2) to develop a verifiable test for accepting an approximate solution to the penalized problem (3.5). We retain the assumptions of section 3.2, in particular the differentiability assumptions and Assumption 3.8. Moreover, we assume that $a_i = 0$, $i = 1, \ldots, n$. Then $d_i^\oplus = d_i^*$ and, since d_i is essentially smooth, $\mathbb{R}_{++} = \operatorname{int}\operatorname{dom} d_i = \operatorname{dom} \nabla_1 d_i = \operatorname{rge} \nabla_1 d_i^*$ [24, Theorem 23.5].

Let $\sigma \in [0, 1]$, $\{s_k\}$ be a nonnegative summable sequence, and $\{z_k\}$ be a nonnegative vanishing sequence. Let y^{k+1} be an approximate solution of the unconstrained minimization (3.5) and let x^{k+1} be defined by (3.6). Note that $x^{k+1} > 0$. To obtain a subgradient of f at x^{k+1}, as required by (3.1), let

$$\widetilde{y} \in \operatorname*{Arg\,min}_{y \in \mathbb{R}^m} \left\{ g_0(y) + \sum_{i=1}^n x_i^{k+1} g_i(y) \right\}.$$

Then, for any $x \in \mathbb{R}^n_+$, we have from (3.3) that

$$\begin{aligned} f(x) &\geq -g_0(\widetilde{y}) - \langle x, g(\widetilde{y})\rangle \\ &= -g_0(\widetilde{y}) - \langle x^{k+1}, g(\widetilde{y})\rangle + \langle x^{k+1} - x, g(\widetilde{y})\rangle \\ &= f(x^{k+1}) + \langle x - x^{k+1}, -g(\widetilde{y})\rangle, \end{aligned}$$

whence $-g(\widetilde{y}) \in \partial f(x^{k+1})$. On the other hand, (3.6) and [24, Theorem 23.5] tell us that $g(y^{k+1}) \in \operatorname{diag}(\alpha^k)^{-1}\nabla_1 D(y^{k+1}, y^k)$. Letting $e^{k+1} = g(y^{k+1}) - g(\widetilde{y})$, we then conclude that the acceptance criterion (3.2) will hold if, for $i = 1, \ldots, n$,

$$\left|g_i(\widetilde{y}) - g_i(y^{k+1})\right| \leq \sigma \left|g_i(y^{k+1})\right| + \min\left\{\frac{s_{k+1}}{\|x^{k+1} - x^k\|}, z_{k+1}\right\}. \tag{B.1}$$

Although $\widetilde{y}$ is unknown, the above inequality may be still be verified if we suppose that g_0 is strongly convex with modulus $\zeta > 0$, and the constraints $g_i,\ i = 1, \ldots, n$, are globally Lipschitz continuous with respective constants $L_i,\ i = 1, \ldots, n$.[5] Let

$$\phi_k(y) \stackrel{\text{def}}{=} g_0(y) + \sum_{i=1}^{n} \frac{1}{\alpha_i^k} d_i^*\big(\alpha_i^k g_i(y), x_i^k\big)$$

denote the augmented Lagrangian at step $k \geq 0$. Note that ϕ_k inherits the strong convexity of g_0. Then, since $\nabla\phi_k(\widetilde{y}) = 0$,

$$\zeta \left\|\widetilde{y} - y^{k+1}\right\| \leq \left\|\nabla\phi_k(y^{k+1})\right\|.$$

Using the Lipschitz continuity of the constraints, it follows that

$$\frac{\zeta}{L_i}\left|g_i(\widetilde{y}) - g_i(y^{k+1})\right| \leq \left\|\nabla\phi_k(y^{k+1})\right\|, \quad i = 1, \ldots, n.$$

Therefore, (B.1) holds whenever

$$\left\|\nabla\phi_k(y^{k+1})\right\| \leq \frac{\zeta}{L_i}\left[\sigma\left|g_i(y^{k+1})\right| + \min\left\{\frac{s_{k+1}}{\|x^{k+1} - x^k\|}, z_{k+1}\right\}\right],\ i = 1, \ldots, n. \tag{B.2}$$

This last relation may be readily tested in practice. Furthermore, our final lemma shows that if we choose $s_{k+1}, z_{k+1} > 0$ and use a convergent algorithm to solve the subproblem (3.5), then (B.2) must eventually be satisfied.

LEMMA B.1. *Suppose s_{k+1} and z_{k+1} are both positive, and let $\bar{y}$ be any solution of* (3.5). *There is a neighborhood $\mathcal{N}$ of $\bar{y}$ such that if $y^{k+1} \in \mathcal{N}$, then* (B.2) *holds.*

Proof. Define, for $i = 1, \ldots, n$,

$$\begin{aligned} x_i(y) &\stackrel{\text{def}}{=} \nabla_1 d_i^*\big(\alpha_i^k g_i(y), x_i^k\big), \\ w_i(y) &\stackrel{\text{def}}{=} \|\nabla\phi_k(y)\| - \frac{\zeta}{L_i}\left[\sigma\left|g_i(y)\right| + \min\left\{\frac{s_{k+1}}{\|x(y) - x^k\|}, z_{k+1}\right\}\right], \end{aligned}$$

where $x(y)$ denotes the n-vector of the $x_i(y)$, and the min is taken to be z_{k+1}, as in our standing convention, whenever the enclosed denominator is zero. With this

[5] The strong convexity assumption is usual in the literature; see [3, Remark 5.2] and [21, section 10]. However, these results do not require Lipschitz continuity of the constraints.

convention, the w_i, $i = 1, \dots, n$, are continuous functions. Moreover, at $\bar{y}$ we have

$$\begin{aligned} w_i(\bar{y}) &\le 0 - \frac{\zeta}{L_i}\left[\sigma\,|g_i(\bar{y})| + \min\left\{\frac{s_{k+1}}{\|x(\bar{y}) - x^k\|}, z_{k+1}\right\}\right] \\ &\le -\frac{\zeta}{L_i}\min\left\{\frac{s_{k+1}}{\|x(\bar{y}) - x^k\|}, z_{k+1}\right\} \\ &< 0, \end{aligned}$$

the last inequality following from the positivity of s_{k+1} and z_{k+1}. For each i, the continuity of w_i implies the existence of a neighborhood $\mathcal{N}_i$ of $\bar{y}$ over which w_i is negative. Let $\mathcal{N} = \bigcap_{i=1}^n \mathcal{N}_i$, which is also a neighborhood of $\bar{y}$. Recalling (3.6), we find that (B.2) holds if $y^{k+1} \in \mathcal{N}$. $\square$

REFERENCES

[1] A. Auslender and M. Haddou, *An interior-proximal method for convex linearly constrained problems and its extension to variational problems*, Math. Programming, 71 (1995), pp. 77–100.

[2] A. Auslender, M. Teboulle, and S. Ben-Tiba, *A logarithmic-quadratic proximal method for variational inequalities*, Comput. Optim. Appl., 12 (1999), pp. 31–40.

[3] A. Auslender, M. Teboulle, and S. Ben-Tiba, *Interior proximal and multiplier methods based on second order homogeneous kernels*, Math. Oper. Res., 24 (1999), pp. 645–668.

[4] A. Auslender and M. Teboulle, *Lagrangian duality and related multiplier methods for variational inequality problems*, SIAM J. Optim., 10 (2000), pp. 1097–1115.

[5] A. Ben-Tal and M. Zibulevsky, *Penalty/barrier multiplier methods for convex programming problems*, SIAM J. Optim., 7 (1997), pp. 347–366.

[6] D. Bertsekas, *Nonlinear Programming*, Athena Scientific, Belmont, MA, 1995.

[7] D. Bertsekas, *Constrained Optimization and Lagrange Multiplier Methods*, Athena Scientific, Belmont, MA, 1996.

[8] L. M. Bregman, *The relaxation method of finding the common point of convex sets and its application to the solution of problems in convex programming*, Comput. Math. Math. Phys., 7 (1967), pp. 200–217.

[9] Y. Censor, A. N. Iusem, and S. A. Zenios, *An interior-point method with Bregman functions for the variational inequality problem with paramonotone operators*, Math. Programming, 81 (1998), pp. 373–400.

[10] Y. Censor and A. Lent, *An iterative row-action method for interval convex programming*, J. Optim. Theory Appl., 34 (1981), pp. 321–353.

[11] Y. Censor and S. A. Zenios, *The proximal minimization algorithms with D-functions*, J. Optim. Theory Appl., 73 (1992), pp. 451–464.

[12] G. Chen and M. Teboulle, *Convergence analysis of a proximal-like minimization algorithm using Bregman functions*, SIAM J. Optim., 3 (1993), pp. 538–543.

[13] J. Eckstein, *Nonlinear proximal point algorithms using Bregman functions, with applications to convex programming*, Math. Oper. Res., 18 (1993), pp. 202–226.

[14] J. Eckstein, *Approximate iterations in Bregman-function-based proximal algorithms*, Math. Programming, 83 (1998), pp. 113–124.

[15] J. Eckstein, *A Practical General Approximation Criterion for Methods of Multipliers Based on Bregman Distances*, RUTCOR Research Report RRR 61-2000, Rutgers University, Piscataway, NJ, 2000.

[16] J. Gauvin, *A necessary and sufficient condition to have bounded multipliers in nonconvex programming*, Math. Programming, 12 (1977), pp. 136–138.

[17] C. Humes and P. Silva, *Strict convex regularizations, proximal point and augmented Lagrangians*, RAIRO Oper. Res., 34 (2000), pp. 283–303.

[18] A. N. Iusem, *Augmented Lagrangian methods and proximal points methods for convex optimization*, Investigación Operativa, 8 (1999), pp. 11–49.

[19] A. N. Iusem, B. F. Svaiter, and M. Teboulle, *Entropy-like proximal methods in convex programming*, Math. Oper. Res., 19 (1994), pp. 790–814.

[20] K. C. Kiwiel, *On the twice differentiable cubic augmented Lagrangian*, J. Optim. Theory Appl., 88 (1996), pp. 233–236.

[21] K. C. Kiwiel, *Proximal minimization methods with generalized Bregman functions*, SIAM J. Control Optim., 35 (1997), pp. 1142–1168.
[22] B. Polyak, *Introduction to Optimization*, Optimization Software, New York, 1987.
[23] R. T. Rockafellar, *Extension of Fenchel's duality theorem for convex functions*, Duke Math. J., 33 (1966), pp. 81–89.
[24] R. T. Rockafellar, *Convex Analysis*, Princeton University Press, Princeton, NJ, 1970.
[25] R. T. Rockafellar, *Conjugate Duality and Optimization*, SIAM, Philadelphia, 1974.
[26] R. T. Rockafellar, *Augmented Lagrangians and applications of the proximal point algorithm in convex programming*, SIAM J. Control Optim., (1976), pp. 97–116.
[27] R. T. Rockafellar and R. J.-B. Wets, *Variational Analysis*, Springer-Verlag, Berlin, 1998.
[28] P. Silva, *Tópicos em Métodos de Ponto Proximal*, Ph.D. thesis, Instituto de Matemática e Estatística—University of São Paulo, São Paulo, Brazil, 2000.
[29] M. V. Solodov and B. F. Svaiter, *An inexact hybrid generalized proximal point algorithm and some new results on the theory of Bregman functions*, Math. Oper. Res., 25 (2000), pp. 214–230.
[30] M. Teboulle, *Entropic proximal mappings with applications to nonlinear programming*, Math. Oper. Res., 17 (1992), pp. 670–690.
[31] M. Teboulle, *Convergence of proximal-like algorithms*, SIAM J. Optim., 7 (1997), pp. 1069–1083.
[32] P. Tseng and D. P. Bertsekas, *On the convergence of the exponential multiplier method for convex programming*, Math. Programming, 60 (1993), pp. 1–19.