SIAM J. OPTIM.
Vol. 20, No. 5, pp. 2653–2678

INEXACT PROXIMAL POINT METHODS FOR VARIATIONAL INEQUALITY PROBLEMS*

REGINA BURACHIK† AND JOYDEEP DUTTA‡

Abstract. We present a new family of proximal point methods for solving monotone variational inequalities. Our algorithm has a relative error tolerance criterion in solving the proximal subproblems. Our convergence analysis covers a wide family of regularization functions, including double regularizations recently introduced by Silva, Eckstein, and Humes, Jr. [*SIAM J. Optim.*, 12 (2001), pp. 238–261] and the Bregman distance induced by $h(x) = \sum_{i=1}^{n} x_i \log x_i$. We do not use in our analysis the assumption of paramonotonicity, which is standard in proving convergence of Bregman-based proximal methods.

Key words. maximal monotone operators, variational inequalities, generalized proximal point algorithms, double regularizations, Bregman distances, second order homogeneous kernels

AMS subject classifications. 65J20, 65K10, 90C25, 90C48

DOI. 10.1137/080733437

1. Introduction.

Let $T : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ be a set-valued map, and let C be a closed convex subset of $\mathbb{R}^n$. The *variational inequality problem* associated with T and C, denoted as $VIP(T, C)$, consists of finding $x^* \in C$ and $w^* \in Tx^*$ such that

$$\langle w^*, y - x^* \rangle \geq 0 \quad \forall y \in C. \tag{1.1}$$

We assume that T is maximal monotone, and we further assume that $D(T) \cap \operatorname{int} C$ is nonempty. Under these assumptions, it is well known that the above problem is equivalent to the following:

$$\text{Find} \quad x^* \quad \text{such that} \quad 0 \in (T + N_C)(x^*),$$

where the set-valued map $N_C : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ is the normal cone map associated with C, i.e., it returns us the normal cone to the set C at x if $x \in C$ and the empty set otherwise. Under the assumption $D(T) \cap \operatorname{int} C \neq \emptyset$, the map $\hat{T} := T + N_C$ is a maximal monotone operator. Thus problem $VIP(T, C)$ boils down to one of finding the zero of $\hat{T} = T + N_C$. A classical approach for finding a zero of a maximal monotone operator $\hat{T}$ is given by the proximal point method (see, e.g., [24]), which is defined as follows. Given $x^{k-1} \in \mathbb{R}^n$ and a positive scalar λ_k, find x^k such that

$$0 \in \lambda_k \hat{T}(x^k) + (x^k - x^{k-1}).$$

When $\hat{T} = T + N_C$, then the above iteration has the additional restriction $x^k \in C$, which has to be dealt with separately. In order to remove this drawback, one can replace the term $\rho(\cdot) = (\cdot - x^{k-1})$, which is the gradient of the quadratic distance $d_0(\cdot, x^{k-1}) := (1/2)\| \cdot - x^{k-1}\|^2$, with the gradient of a generalized distance function

*Received by the editors August 25, 2008; accepted for publication (in revised form) May 26, 2010; published electronically July 20, 2010.
http://www.siam.org/journals/siopt/20-5/73343.html

†School of Mathematics and Statistics, University of South Australia, Mawson Lakes, SA 5095, Australia (regina.burachik@unisa.edu.au). This author acknowledges support by the Australian Research Council Discovery Project grant DP0556685 for this study.

‡Department of Mathematics and Statistics, Indian Institute of Technology, Kanpur-208106, India (jdutta@iitk.ac.in).

$d(\cdot, x^{k-1})$, specifically chosen so that x^k is forced to be in the interior of C. Using the fact that $\hat{T} = T$ in the interior of C, the proximal iteration becomes

$$0 \in \lambda_k T(x^k) + \nabla_1 d(x^k, x^{k-1}),$$

where $\nabla_1 d$ stands for the gradient of $d(\cdot, \cdot)$ with respect to the first variable. The so-obtained methods are called *generalized proximal methods* (GPPMs). Well-known examples of these regularizing functionals are the Bregman distances (see, e.g., [1, 9, 14, 16, 22, 26]), φ-divergences (see [30, 7, 19, 20, 21, 31, 32]), log-quadratic distances (also known as second order homogeneous kernels) [4, 5], and double regularizations, which extend the latter ones, and were recently introduced in [17].

An implementable scheme of generalized proximal methods must be able to accept inexact solutions of the subproblems. Many such inexact proximal point schemes have been devised for variational inequalities. There are mainly two ways in the literature for devising inexact schemes of GPPMs. One of them uses *summable error analysis* (see, e.g., [3, 4, 5, 8, 18]), which means that the infinite sum of all errors is finite. The drawback of this approach is that the summability assumption may force high accuracy for accepting iterates at early stages of the algorithm. The other, more recent, approach is based on *relative error analysis* (see, e.g., [28, 29, 26, 13]). This latter approach may be more convenient from the computational point of view because it allows iterations which are more tolerant to errors.

Our aim is to provide a convergence analysis for inexact generalized proximal methods, which includes (i) the Bregman distance induced by $h(x) = \sum_{i=1}^n x_i \ln x_i$, (ii) log-quadratic distances, and (iii) double regularizations. Moreover, our analysis does not require paramonotonicity of T, which is a standard assumption for convergence of Bregman-based proximal methods (see, e.g., [26, 9]). Our analysis considers two kinds of relative error criteria. The first one (see Definition 3.1) is for a proximal methods with double regularizations, and it extends the one used in [13] for second order kernels. The second criterion (see Definition 3.2) is for the proximal method with the Bregman distance induced by the function $h(x) = \sum_{i=1}^n x_i \ln x_i$. We show in section 6 that the use of the enlargement of T and the relative error analysis allows a simple implementation of the step given by Definition 3.2. To our knowledge, this specific implementation is not available for related algorithms in the literature such as the one in [27]. We also show that when the Bregman distance is quadratic, then our algorithm is equivalent to the hybrid inexact proximal method as given in [28]. In this way, the same inexact step can be used to analyze very different schemes, such as the ones induced by log-quadratic, double regularizations, and pure quadratic regularizations. Note also that we do not require paramonotonicity of T. The latter assumption is restrictive in the sense that important kinds of maximal monotone maps are not paramonotone. The subdifferential of a convex function is paramonotone, but the subdifferential of a saddle function is not. We describe now the main differences between our methods and related ones in the literature. A convergence analysis for proximal methods without paramonotonicity was presented by Silva, Eckstein, and Humes, Jr. in [27]. The latter analysis covers the case of the Bregman distance induced by the function $h(x) = \sum_{i=1}^n x_i \ln x_i$. The differences between the method in [27] and the one in the present paper are as follows. First, the error analysis used in [27] is summable (see [27, Assumption 4.4]). Second, the step in [27] uses the given operator T, while ours uses the enlargement of T (see (2.1)), which gives more freedom in the choice of the iterates. Third, we use in the present paper an extragradient step, while in [27] the authors use a "pure" proximal step (see [27, BIPPA]).

In [17], an inexact proximal method is presented for double regularizations. As in [27], summable error analysis is used, the iterations do not involve an extragradient step, and the inexact step does not make use of the enlargement of T.

We also note that both our method and our error criteria reduce to the one in [13] when the distance is given by a second order kernel.

In our analysis, we will use the concept of *proximal distance*, recently introduced in [3, 2]. Our error criteria are given in terms of the proximal distance and are closely inspired by [13].

The paper is organized as follows. In section 2 we recall basic notions and properties on set-valued maps, proximal distances, and induced proximal distances. Section 3 introduces our generalized proximal point method and the formal definition of approximate solution. In section 4 we provide examples of proximal distances and their associated induced proximal distances. We pay particular attention to the Bregman distance induced by the function $h(x) = \sum_{i=1}^{n} x_i \ln x_i$. We also explain how our error criterion is related to the one introduced in [28] in the context of an inexact proximal point scheme with quadratic regularization. We show that our relative error scheme is equivalent to the relative error scheme in [28] when, in particular, we consider the quadratic distance. In section 5 we develop the convergence analysis of our inexact proximal scheme, and we show that the convergence can be achieved without paramonotonicity, including the case of the Bregman distance induced by $h(x) = \sum_{i=1}^{n} x_i \ln x_i$. In section 6 we show how our inexact proximal scheme can be easily implemented, even when T is point-to-set, for the constraint $C = \mathbb{R}^n_+$ and the proximal distance is the Bregman distance induced by the function h above.

2. Basic definitions. Given a subset $C \subset \mathbb{R}^n$, we denote by $\operatorname{int} C$ its interior and by $\operatorname{bdry} C$ its boundary. We use the notation $\mathbb{R}_{+\infty} := \mathbb{R} \cup +\infty$. We collect next a few definitions related to point-to-set operators and generalized distances. A point-to-set valued map $T : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ is an operator which associates to each point $x \in \mathbb{R}^n$ a (possibly empty) set $T(x) \subset \mathbb{R}^n$. The domain and the graph of a point-to-set valued map T are defined as

$$D(T) := \{x \in \mathbb{R}^n \mid T(x) \neq \emptyset\},$$

$$G(T) := \{(x, v) \in \mathbb{R}^n \times \mathbb{R}^n \mid x \in D(T),\ v \in T(x)\}.$$

A point-to-set operator T is said to be *monotone* if

$$\langle v' - v, x' - x\rangle \geq 0 \ \ \forall\, v \in T(x),\, v' \in T(x').$$

A monotone operator is said to be *maximal* when its graph is not properly contained in the graph of any other monotone operator.

To deal with inexact proximal iterations, we use the notion of enlargement of a maximal monotone operator $T : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ (see [8, 12, 11]). Given $\varepsilon \geq 0$ and $x \in \mathbb{R}^n$, the *ε-enlargement of* T at the point x is the set

$$T^{\varepsilon}(x) = \{u \in \mathbb{R}^n \mid \langle v - u, y - x\rangle \geq -\varepsilon\ \forall y \in \mathbb{R}^n, v \in T(y)\}\,. \tag{2.1}$$

The above enlargement of T has some useful theoretical and algorithmic applications, thanks to the fact that it shares many properties with the ε-subdifferential of a convex function. In the same way as the ε-subdifferential is used for devising and analyzing inexact minimization schemes in convex optimization (see, e.g., [4, 2]), the T^{ε} can be used for the analysis of inexact proximal schemes for variational inequalities. (See, e.g., [8, 13] and [10, Chapter 5] for a comprehensive review of these enlargements.)

We shall now list the desirable properties of the generalized distance d. Our assumptions for d, listed below, are taken from the ones in [2, Definition 2.1].

DEFINITION 2.1. *A function* $d : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}_+ \cup \{+\infty\}$ *is said to be a* proximal distance *with respect to a closed nonempty and convex set* $C \subset \mathbb{R}^n$ *if, for every fixed* $y \in \operatorname{int} C$*, the following properties hold:*

($d1$) $d(\cdot, y)$ *is a proper, lsc convex function and* C^1 *on* $\operatorname{int} C$*.*

($d2$) *dom* $d(\cdot, y) \subseteq C$*, and* *dom*$\nabla_1 d(\cdot, y) = \operatorname{int} C$*.*

We write $d \in \mathcal{D}(C)$ *when a function* d *satisfies conditions* $(d1)-(d2)$*.*

Following the analysis given in [2], we associate with every $d \in \mathcal{D}(C)$ an induced proximal distance H_d which we define below.

DEFINITION 2.2. *The* induced proximal distance *associated with* $d \in \mathcal{D}(C)$ *is a function* $H_d : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}_+ \cup \{+\infty\}$ *such that* $\operatorname{int} C \times \operatorname{int} C \subseteq$ *dom* $H_d := \{(x, y) \in \mathbb{R}^n \times \mathbb{R}^n : H_d(x, y) < +\infty\}$ *and satisfies the following properties:*

($H1a$) *For every* $a \in \operatorname{int} C$*,* $H_d(a, \cdot)$ *is continuous on* $\operatorname{int} C$*.*

($H1b$) $H_d(a, a) = 0$ *for all* $a \in \operatorname{int} C$*.*

($H2$) *For all* $a \in C$ *and* $\alpha \in \mathbb{R}$*, the set* $\{y \in \operatorname{int} C : H_d(a, y) \leq \alpha\}$ *is bounded.*

($H3$) *For every* $a, b \in \operatorname{int} C$*, it holds that*

$$\langle c - b, \nabla_1 d(b, a) \rangle \leq H_d(c, a) - H_d(c, b) - \gamma H_d(b, a)$$

for all $c \in C$ *and some fixed* $\gamma > 0$*.*

($H4$) *If* $\{y^k\} \subset \operatorname{int} C$ *and* $y^k \to y \in C$*, then* $H_d(y, y^k) \to 0$*.*

($H5$) *Let* $z \in C$ *and* $y \in \operatorname{int} C$*, and take* $w := \alpha z + (1 - \alpha) y$*, with* $\alpha \in (0, 1)$*. Then*

$$H_d(z, w) + H_d(w, y) \leq H_d(z, y).$$

($H6$) *If* $\{x^k\}, \{y^k\} \subset \operatorname{int} C$ *are sequences such that* $\{x^k\}$ *converges to* x *and* $\{y^k\}$ *converges to* y*, with* $x \neq y$*, then*

$$\liminf_k H_d(x^k, y^k) > 0.$$

We write $(d, H_d) \in \mathcal{F}(C)$ *when a triple* $[C, d, H_d]$ *verifies conditions* $(H1)-(H6)$*. In this case, we say that* (d, H_d) *is a* proximal pair *associated with* C*.*

Assumption ($H1b$) is used in [3, 2] in the definition of induced proximal distance. Requirements ($H2$) and ($H4$) are classical assumptions in the context of Bregman distances. Assumption ($H3$) is used in [3, Definition 2.2 (Equation 2.14)]. Conditions $(H5)-(H6)$ are inspired from [26]. In [26, Lemmas 2.2 and 2.3] they are proved to hold for Bregman distances.

The fact that conditions $(H5)-(H6)$ also hold for second order kernels and double regularizations is pointed out below in Remarks 4.2 and 4.4.

Before stating the method, we recall two important facts regarding proximal distances verifying ($H5$) and ($H6$). The proofs of the following two lemmas are similar to those in [26, Lemmas 2.2 and 2.3], but we include the proofs here for completeness.

LEMMA 2.3. *Assume that* H_d *verifies* ($H5$) *and* ($H6$)*. If* $\{x^k\} \subset C$ *and* $\{y^k\} \subset \operatorname{int} C$ *are sequences such that*

$$\lim_k H_d(x^k, y^k) = 0$$

and one of the sequences ($\{x^k\}$ *or* $\{y^k\}$) *converges, then the other one also converges to the same limit.*

Proof. Suppose, by contradiction, that one of the sequences, say, $\{y^k\}$, converges to y and that $\{x^k\}$ does not converge or converges to a different limit. In that case, there exists a set of indexes $\{k_j\}$ and $\varepsilon > 0$ such that $\|x^{k_j} - y^{k_j}\| > \varepsilon$. Define $z^j := y^{k_j} + \frac{\varepsilon}{\|x^{k_j}-y^{k_j}\|}(x^{k_j} - y^{k_j})$. From (H6) we know that

$$H_d(z^j, y^{k_j}) \leq H_d(x^{k_j}, y^{k_j}).$$

Using now our assumption, we have that $H_d(z^j, y^{k_j})$ converges to 0. Noting that $\|z^j - y^{k_j}\| = \varepsilon$ and $\{y^{k_j}\}$ converges, we conclude that $\{z^j\}$ is bounded, and hence it has a convergent subsequence. Call z the limit of that subsequence. Without loss of generality, we also denote the subsequence as $\{z^j\}$. Altogether, we have

$$\begin{gathered} \|z - y\| = \varepsilon, \\ \liminf_j H_d(z^j, y^{k_j}) = 0, \\ \lim_j z^j = z, \\ \lim_j y^{k_j} = y. \end{gathered}$$

The above list of facts contradicts (H6). If we assume that $\{x^k\}$ converges (and $\{y^k\}$ does not converge or converges to a different limit), then a similar argument as above leads to a contradiction. Therefore, the conclusion is true. ◻

The following result is a direct consequence of the lemma above.

LEMMA 2.4. *Assume that H_d verifies (H6) and (H7). Suppose that*

$$\lim_k H_d(x^k, y^k) = 0$$

and that one of the sequences ($\{x^k\}$ or $\{y^k\}$) is bounded. Then the following hold:

(a) *The other sequence is also bounded.*

(b) $(x^k - y^k) \to 0$.

Proof. Assume (a) is not true. For instance, assume that $\{y^k\}$ is bounded and that the sequence $\{x^k\}$ is unbounded. So there exists an infinite set of indices J such that $||x^k|| \to \infty$ as $k \to \infty$ with $k \in J$. Further, since $\{y^k\}$ is bounded, there exists an infinite set of indices J' with $J' \subset J$ such that y^k converges to a limit $\bar{y}$ as $k \to \infty$ with $k \in J'$. We also have $H_d(x^k, y^k) \to 0$ as $k \to \infty$ and $k \in J'$. Now by Lemma 2.3 we have that $x^k \to \bar{y}$ as $k \to \infty$ and $k \in J'$. This contradicts the fact that $||x^k|| \to \infty$ as $k \to \infty$ and $k \in J'$.

For (b), assume that $(x^k - y^k)$ does not converge to zero. Then there exists an infinite set of indices J and $\varepsilon > 0$ such that

$$||x^k - y^k|| \geq \varepsilon, \quad k \in J. \tag{2.2}$$

Now using (a), we have that both sequences $\{x^k\}$ and $\{y^k\}$ are bounded, and thus we can find an infinite set of indices $J' \subset J$ such that $x^k \to x$ and $y^k \to y$ as $k \to \infty$ and $k \in J'$. Now from Lemma 2.3 we have that $x = y$. This clearly contradicts (2.2). Hence we have the result. ◻

The following simple lemma will be used in our convergence analysis.

LEMMA 2.5. *Assume that the real sequences $\{\alpha_j\}$ and $\{\beta_j\}$ are such that*

(a) $\liminf_j \alpha_j \geq 0$,

(b) $\beta_j \geq \bar{\beta} > 0$.

Then we must have $\liminf_j \frac{\alpha_j}{\beta_j} \geq 0$.

Proof. If the conclusion of the lemma is not true, then there exists $\delta > 0$ such that $\liminf_j \frac{\alpha_j}{\beta_j} < -\delta$. Therefore, there exists a subsequence n_j such that

$$\frac{\alpha_{n_j}}{\beta_{n_j}} < -\delta \ \forall j.$$

Combine the above fact with (b) to obtain

$$\inf_{n \geq j} \alpha_n \leq \alpha_{n_j} < -\beta_{n_j}\delta \leq -\bar{\beta}\delta < 0$$

for every $j \in \mathbb{N}$. Taking supremum over n in the above expression yields

$$\liminf_j \alpha_j \leq -\bar{\beta}\delta < 0,$$

contradicting (a). ▯

3. Inexact proximal point method. We work under the assumption that $D(T) \cap \operatorname{int} C \neq \emptyset$. The iterative steps associated with the inexact proximal point methods are as follows. Assume d is a fixed proximal distance with respect to C. Given $x^{k-1} \in \operatorname{int} C$ and $\lambda_k > 0$, find a triplet $(\tilde{x}^k, \tilde{v}^k, \varepsilon_k)$ such that the following holds:

$$(B11) \qquad \tilde{v}^k \in T^{\varepsilon_k}(\tilde{x}^k),$$

$$(B12) \qquad e_k = \lambda_k \tilde{v}^k + \nabla_1 d(\tilde{x}^k, x^{k-1}).$$

As is standard in the analysis of proximal-like methods, the regularization parameter λ_k is bounded away from zero, i.e., $\lambda_k \geq \hat{\lambda} > 0$ for all k.

The inexact proximal point method also requires what is known as an extragradient step which is given as

$$(B13) \qquad x^k \in (\nabla_1 d(\cdot, x^{k-1}))^{-1}(-\lambda_k \tilde{v}^k), \quad \text{i.e.,} \quad \nabla_1 d(x^k, x^{k-1}) = -\lambda_k \tilde{v}^k.$$

In the case in which $d = D_h$ is the Bregman distance induced by $h(x) = \sum_{i=1}^n x_i \ln x_i$ and $d(x,y) = D_h(x,y) = h(x) - h(y) - \langle \nabla h(y), x - y \rangle = \sum_1^n x_i \ln(x_i/y_i) + y_i - x_i$ (see section 4.1), we consider a rescaled extragradient step as

$$(B13)' \qquad r_{k-1} \nabla_1 d(x^k, x^{k-1}) = -\lambda_k \tilde{v}^k,$$

where $r_{k-1} := \min_{i=1,\dots,n}\{x_i^{k-1}\}$.

We give next a motivation for our inexact step. The *exact* proximal iteration can be described as follows:

$$(3.1) \qquad v^k \in T(x^k),$$

$$(3.2) \qquad \lambda_k v^k + \nabla_1 d(x^k, x^{k-1}) = 0.$$

With this in mind, we see that $(B11)$ relaxes the inclusion in (3.1) and that $(B12)$ relaxes the equality in (3.2).

DEFINITION 3.1. *Let d be the proximal distance used in* $(B11)$–$(B13)$*, and take H_d a proximal distance induced by d. A vector $(\tilde{x}^k, x^k, \tilde{v}^k, \varepsilon_k)$ verifying* $(B11)$–$(B13)$ *is called an* approximate solution *of* (3.1)–(3.2) *if the following estimates are satisfied:*

$$(E1) \qquad H_d(\tilde{x}^k, x^k) \leq c_1 H_d(x^k, x^{k-1}),$$

$$(E2) \qquad \lambda_k(\varepsilon_k + \langle \tilde{v}^k, \tilde{x}^k - x^k \rangle) \leq c_2 H_d(x^k, x^{k-1}),$$

where c_1 and c_2 are positive constants with $c_2 = \sigma\gamma$, where $\gamma > 0$ as in $(H3)$ and $\sigma \in [0, 1)$.

As mentioned above, the error criteria $(E1)-(E2)$ reduces to the one in [13] for the case in which the distance d is induced by a second order kernel. Indeed, in this case, $H_d(x, y)$ is a multiple of $\frac{\|x-y\|^2}{2}$ (see Lemma 4.3). So our inexact scheme is an extension of the one presented in [13, Equations (17)–(18)].

In order to establish convergence for the Bregman-based proximal method induced by $h(x) = \sum_{i=1}^n x_i \ln x_i$, we need to rescale $(E2)$ as follows.

DEFINITION 3.2. *Let $d = D_h$ in $(B11)-(B12)$ and $(B13)'$ be the Bregman distance induced by $h(x) = \sum_{i=1}^n x_i \ln x_i$. A vector $(\tilde{x}^k, x^k, \tilde{v}^k, \varepsilon_k)$ verifying $(B11)-(B12)$ and $(B13)'$ is called an* approximate solution *of* (3.1)–(3.2) *if the following estimates are satisfied:*

$$(E1) \qquad H_d(\tilde{x}^k, x^k) \le c_1 H_d(x^k, x^{k-1}),$$

$$(E2)' \qquad \lambda_k(\varepsilon_k + \langle \tilde{v}^k, \tilde{x}^k - x^k\rangle) \le \sigma H_d(x^k, x^{k-1}) r_{k-1},$$

where c_1 is a positive constant, $\sigma \in [0, 1)$, and r_{k-1} is as in $(B13)'$.

4. Examples of proximal pairs (d, H_d). We construct here specific examples of proximal pairs (d, H_d) for the choices of d as a Bregman distance, a double regularization, or a second order kernel.

4.1. Bregman distances. We start by recalling the definition and well-known properties of Bregman distances. Let $h : \mathbb{R}^n \to \mathbb{R}_{+\infty}$ be a proper, lsc function with $\operatorname{dom} h \subset C$ and $\operatorname{dom} \nabla h = \operatorname{int} C$, strictly convex and continuous on $\operatorname{dom} h$, and continuously differentiable on $\operatorname{int}\operatorname{dom} h = \operatorname{int} C$. Define

$$D_h(x, y) := \begin{cases} h(x) - h(y) - \langle \nabla h(y), x - y\rangle & \text{if } x \in \mathbb{R}^n,\ y \in \operatorname{dom} \nabla h, \\ +\infty & \text{otherwise.} \end{cases}$$

Let $d = D_h$, a Bregman distance induced by h. It is proved in [2] that the proximal distance induced by d is $H_d := D_h$, and hence the proximal methods generated by these distances are called in [2] *self-proximal*.

Remark 4.1. The assumptions on h imply that $H_d = D_h$ verifies $(H1.a)-(H1.b)$ as well as $(H3)$ because of the three point identity [15, Lemma 3.1], which is stated as

$$D_g(c, a) = D_g(c, b) + D_g(b, a) + \langle c - b, \nabla_1 D_g(b, a)\rangle \tag{4.1}$$

for every $a, b \in \operatorname{int} C$ and every $c \in C$. As mentioned before (see [26, Lemmas 2.2 and 2.3]), every Bregman distance verifies both $(H5)$ and $(H6)$. Therefore, conditions $(H1)$, $(H3)$, and $(H5)-(H6)$ hold automatically for the pair $(d, H_d) = (D_g, D_g)$. On the other hand, conditions $(H2)$ and $(H4)$ are standard assumptions in the context of Bregman distances, so it is natural to assume them to hold in our analysis.

We consider in our analysis two kinds of Bregman distances. One of them is the case in which $C = \mathbb{R}^n_+$ and the Bregman distance is induced by the function $h(x) = \sum_1^n x_i \ln x_i$. For this kind of regularization, we will use Definition 3.2, which allows for a natural and simple implementation (see section 6).

The other Bregman distance we consider is the one induced by $h(x) = \frac{1}{2}\sum_{i=1}^n {x_i}^2$ with $C = \mathbb{R}^n$. Note that in this case, we have $H_d(x, y) = \frac{1}{2}\|x - y\|^2$, which is not included in section 4.2 nor section 4.3. For this kind of regularization, we will show that the algorithm produced through Definition 3.1 is equivalent to the hybrid inexact proximal method of [28]. To describe the latter method, we start by recalling the *exact* proximal step as follows. Given a current iterate x^{k-1}, find x^{*k} such that

$$\begin{cases} v^k \in T(x^{*k}), \\ 0 = \lambda_k v^k + x^{*k} - x^{k-1}. \end{cases} \tag{4.2}$$

DEFINITION 4.1. *Given a fixed constant* $\theta \in [0,1)$ *and a current iterate* x^{k-1}, *a vector* $(\tilde{x}^k, x^k, \tilde{v}^k, \varepsilon_k)$ *verifying*

$$\begin{cases} \tilde{v}^k \in T^{\varepsilon_k}(\tilde{x}^k), \\ e^k := \lambda_k \tilde{v}^k + \tilde{x}^k - x^{k-1}, \\ x^k := x^{k-1} - \lambda_k \tilde{v}^k \quad \textit{(extragradient step)} \end{cases} \tag{4.3}$$

is called a hybrid approximate solution *of* (4.2) *if the following estimate is satisfied:*

$$2\lambda_k \varepsilon_k + \|e^k\|^2 \le \theta^2 \left(\|\lambda_k \tilde{v}^k\|^2 + \|\tilde{x}^k - x^{k-1}\|^2\right). \tag{H}$$

Note that (4.3) coincides with $(B11)-(B13)$. Moreover, when $H_d(x,y) = (1/2)\|x - y\|^2$, then the parameter c_2 in (E2) reduces to $c_2 = \sigma \in [0,1)$ because $\gamma = 1$.

We say that two given methods are equivalent when every instance of one method can be obtained by a specific instance of the other and vice versa. This statement is made precise in the next proposition.

PROPOSITION 4.2. *With the notation of Definitions* 3.1 *and* 4.1, *assume that* $(\tilde{x}^k, x^k, \tilde{v}^k, \varepsilon_k)$ *verifies* (4.3). *The following two statements hold:*

(i) *For every* $c_1 > 0$ *and* $\sigma \in [0,1)$, *there exists* $\theta \in [0,1)$ *such that if* (H) *holds with parameter* θ, *then* $(E1)-(E2)$ *hold with parameters* c_1, σ *for every* k.

(ii) *For every* $\theta \in [0,1)$, *there exists* $c_1 > 0$ *and* $\sigma \in [0,1)$ *such that if* $(E1)-(E2)$ *hold with parameters* c_1, σ, *then* (H) *holds with parameter* θ.

Proof. To prove (i), take $c_1 > 0$ and $\sigma \in [0,1)$ arbitrary. Take θ in condition (H) such that

$$\theta < \min\left\{\frac{1}{2}, \frac{c_1}{4}, \frac{\sigma^2}{6}\right\}. \tag{4.4}$$

Given $c_1 > 0$, we can always find θ small enough such that the above requirement holds. Note also that

$$\begin{aligned} \frac{1}{2} H_d(\tilde{x}^k, x^{k-1}) = \frac{1}{4}\left\|\tilde{x}^k - x^{k-1}\right\|^2 &= \left\|\frac{(\tilde{x}^k - x^k)}{2} + \frac{(x^k - x^{k-1})}{2}\right\|^2 \\ &\le \frac{1}{2}\left[\left\|\tilde{x}^k - x^k\right\|^2 + \left\|x^k - x^{k-1}\right\|^2\right] \\ &= H_d(\tilde{x}^k, x^k) + H_d(x^k, x^{k-1}). \end{aligned} \tag{4.5}$$

On the other hand, condition (H) and (4.5) imply

$$\begin{aligned} H_d(\tilde{x}^k, x^k) = \frac{\left\|\tilde{x}^k - x^k\right\|^2}{2} = \frac{\left\|e^k\right\|^2}{2} &\le \frac{\theta^2}{2}\left[\|\lambda_k \tilde{v}^k\|^2 + \|\tilde{x}^k - x^{k-1}\|^2\right] \\ &= \theta^2\left[H_d(x^k, x^{k-1}) + H_d(\tilde{x}^k, x^{k-1})\right] \\ &\le \theta^2\left[H_d(x^k, x^{k-1}) + 2(H_d(\tilde{x}^k, x^k) + H_d(x^k, x^{k-1}))\right] \\ &= \theta^2\left[3\, H_d(x^k, x^{k-1}) + 2H_d(\tilde{x}^k, x^k))\right] \end{aligned}$$

which can be rearranged to

$$H_d(\tilde{x}^k, x^k) \le \left(\frac{3\theta^2}{1-2\theta^2}\right) H_d(x^k, x^{k-1}). \tag{4.6}$$

It is direct to check that for θ as in (4.4), we have

$$\left(\frac{3\theta^2}{1-2\theta^2}\right) \le 3\theta < 4\theta < c_1.$$

The above expression and (4.6) yield $(E1)$ with parameter c_1. In particular, we can combine the expression above with (4.6) to obtain

$$\left\|\tilde{x}^k - x^k\right\| \le 2\sqrt{\theta}\, \left\|x^k - x^{k-1}\right\|. \tag{4.7}$$

To complete the proof of (i), we must check that our choice of θ implies that $(E2)$ holds with parameter σ. We have that

$$\begin{aligned} \lambda_k\varepsilon_k + \left\langle \lambda_k\tilde{v}^k, \tilde{x}^k - x^k\right\rangle &= \lambda_k\varepsilon_k + \left\langle x^{k-1} - x^k, e^k\right\rangle \\ &= \lambda_k\varepsilon_k + \langle x^{k-1} - \tilde{x}^k, e^k\rangle + \langle \tilde{x}^k - x^k, e^k\rangle \\ &= \lambda_k\varepsilon_k + \langle x^{k-1} - \tilde{x}^k, e^k\rangle + \|e^k\|^2 \\ &= \frac{1}{2}\left[2\lambda_k\varepsilon_k + \|e^k\|^2\right] + \langle x^{k-1} - \tilde{x}^k, e^k\rangle + \frac{1}{2}\|e^k\|^2 \\ &= \frac{1}{2}\left[2\lambda_k\varepsilon_k + \|e^k\|^2\right] + \frac{\|x^k - x^{k-1}\|^2}{2} - \frac{\|\tilde{x}^k - x^{k-1}\|^2}{2} \\ &\le \theta^2\left[H_d(x^k, x^{k-1}) + H_d(\tilde{x}^k, x^{k-1})\right] + H_d(x^k, x^{k-1}) - H_d(\tilde{x}^k, x^{k-1}) \\ &= (\theta^2+1)H_d(x^k, x^{k-1}) - (1-\theta^2)H_d(\tilde{x}^k, x^{k-1}). \end{aligned} \tag{4.8}$$

To obtain the fourth equality above, we used the fact that $\tilde{x}^k - x^k = e^k$ in the expression

$$\frac{1}{2}\left\|x^k - x^{k-1}\right\|^2 = \frac{1}{2}\left\|x^k - \tilde{x}^k\right\|^2 + \frac{1}{2}\left\|\tilde{x}^k - x^{k-1}\right\|^2 + \left\langle x^k - \tilde{x}^k, \tilde{x}^k - x^{k-1}\right\rangle$$

to conclude that

$$\frac{1}{2}\left\|x^k - x^{k-1}\right\|^2 = \frac{1}{2}\left\|e^k\right\|^2 + \frac{1}{2}\left\|\tilde{x}^k - x^{k-1}\right\|^2 + \langle e^k, x^{k-1} - \tilde{x}^k\rangle.$$

Therefore, we obtain

$$\langle e^k, x^{k-1} - \tilde{x}^k\rangle + \frac{1}{2}\left\|e^k\right\|^2 = \frac{1}{2}\left\|x^k - x^{k-1}\right\|^2 - \frac{1}{2}\left\|\tilde{x}^k - x^{k-1}\right\|^2,$$

which is used in the fourth equality in (4.8). By (4.1) and (4.7), we can write

$$\begin{aligned} H_d(\tilde{x}^k, x^{k-1}) &= H_d(\tilde{x}^k, x^k) + H_d(x^k, x^{k-1}) + \langle \tilde{x}^k - x^k, x^k - x^{k-1}\rangle \\ &\ge H_d(x^k, x^{k-1}) + \langle \tilde{x}^k - x^k, x^k - x^{k-1}\rangle \\ &\ge H_d(x^k, x^{k-1}) - \left\|\tilde{x}^k - x^k\right\|\, \left\|x^k - x^{k-1}\right\| \\ &\ge H_d(x^k, x^{k-1}) - 4\sqrt{\theta}\, H_d(x^k, x^{k-1}) \\ &= (1-4\sqrt{\theta})H_d(x^k, x^{k-1}). \end{aligned}$$

Combining the inequality above with (4.8), we conclude that

$$\begin{aligned}
\lambda_k\varepsilon_k \quad &+ \quad \langle \lambda_k \tilde{v}^k, \tilde{x}^k - x^k\rangle \\
&\leq \quad (\theta^2+1)H_d(x^k,x^{k-1}) - (1-\theta^2)(1-4\sqrt{\theta})H_d(x^k,x^{k-1}) \\
&= \quad ((\theta^2+1) - (1-\theta^2)(1-4\sqrt{\theta}))H_d(x^k,x^{k-1}) \\
&= \quad 2(\theta^2 + 2\sqrt{\theta} - 2\sqrt{\theta^3})H_d(x^k,x^{k-1}) \leq 6\sqrt{\theta}H_d(x^k,x^{k-1}) \\
&\leq \quad \sigma H_d(x^k,x^{k-1}),
\end{aligned}$$

where we used (4.4) in the last two inequalities. Hence $(E2)$ also holds for our choice of θ. This proves (i). To prove (ii), let $\theta \in [0,1)$ be arbitrary. We must find σ, c_1 such that (H) holds with parameter θ. Using $(E1)$, $(E2)$, the Cauchy–Schwartz inequality, and some algebra, we can write

$$\begin{aligned}
2\lambda_k\varepsilon_k + \left\|e^k\right\|^2 &= 2\left[\lambda_k\varepsilon_k + \frac{\left\|\lambda_k\tilde{v}^k + \tilde{x}^k - x^{k-1}\right\|^2}{2}\right] \\
&= 2\left[\lambda_k\varepsilon_k + \langle\lambda_k\tilde{v}^k, \tilde{x}^k - x^k\rangle - \langle\lambda_k\tilde{v}^k, \tilde{x}^k - x^k\rangle + \frac{\left\|\lambda_k\tilde{v}^k + \tilde{x}^k - x^{k-1}\right\|^2}{2}\right] \\
&= 2\left[\lambda_k\varepsilon_k + \langle\lambda_k\tilde{v}^k, \tilde{x}^k - x^k\rangle + \frac{\left\|\tilde{x}^k - x^{k-1}\right\|^2}{2} - \frac{\left\|\lambda_k\tilde{v}^k\right\|^2}{2}\right] \\
&\leq 2\left[(\sigma-1)\frac{\left\|x^k - x^{k-1}\right\|^2}{2} + \frac{\left\|\tilde{x}^k - x^{k-1}\right\|^2}{2}\right] \\
&= \left[(\sigma-1)\left\|x^k - x^{k-1}\right\|^2 + (1-\sigma)\left\|\tilde{x}^k - x^{k-1}\right\|^2 + \sigma\left\|\tilde{x}^k - x^{k-1}\right\|^2\right] \\
&= (1-\sigma)\left[\left\|\tilde{x}^k - x^{k-1}\right\|^2 - \left\|x^k - x^{k-1}\right\|^2\right] + \sigma\left\|\tilde{x}^k - x^{k-1}\right\|^2 \\
&= (1-\sigma)\left[\left\|\tilde{x}^k - x^k\right\|^2 + 2\langle\tilde{x}^k - x^k, x^k - x^{k-1}\rangle\right] + \sigma\left\|\tilde{x}^k - x^{k-1}\right\|^2 \\
&\leq (1-\sigma)\left[c_1\left\|x^k - x^{k-1}\right\|^2 + 2\left\|\tilde{x}^k - x^k\right\| \ \left\|x^k - x^{k-1}\right\|\right] + \sigma\left\|\tilde{x}^k - x^{k-1}\right\|^2 \\
&\leq (1-\sigma)(c_1 + 2\sqrt{c_1})\left\|x^k - x^{k-1}\right\|^2 + \sigma\left\|\tilde{x}^k - x^{k-1}\right\|^2.
\end{aligned}$$

If $\theta = 0$, then $c_1 = \sigma = 0$ gives (H) for this value of θ. Assume now that $\theta > 0$. In this case, take $\sigma < \theta^2$ and $c_1 < \min\{1, \frac{\theta^4}{9}\}$. With these choices of σ, c_1, we have $(1-\sigma)(c_1 + 2\sqrt{c_1}) < (c_1 + 2\sqrt{c_1}) < 3\sqrt{c_1} < \theta^2$, and hence

$$2\lambda_k\varepsilon_k + \left\|e^k\right\|^2 \leq \theta^2\left[\left\|x^k - x^{k-1}\right\|^2 + \left\|\tilde{x}^k - x^{k-1}\right\|^2\right]$$

as required. ∎

4.2. Double regularizations. Double regularizations, introduced in [17], extend the notion of second order homogeneous proximal distances introduced in [4, 5]. Consider the constraint set

$$B := [a_1, b_1] \times \cdots \times [a_n, b_n],$$

where the intervals $[a_i, b_i]$ may be finite or infinite. For $i = 1, \dots, n$, consider a function $d_i : \mathbb{R} \times (a_i, b_i) \to R_{+\infty}$. Let now $\tilde{d}_i : \mathbb{R} \times (a_i, b_i) \to R_{+\infty}$ be a function of the form

$$\tilde{d}_i(x_i, y_i) = d_i(x_i, y_i) + \frac{\mu}{2}(x_i - y_i)^2, \tag{4.9}$$

where $\mu \geq 1$. Assume that $\tilde{d}(x, y) := \sum_{i=1}^n \tilde{d}_i(x_i, y_i)$ verifies the following conditions:

(DR1) For all $y \in \Pi_{i=1}^n (a_i, b_i)$, $\tilde{d}(\cdot, y)$ is closed and strictly convex, with its minimum attained at y. Moreover, $\operatorname{intDom}(\tilde{d}_i(\cdot, y_i)) = (a_i, b_i)$.

(DR2) $\tilde{d}_i$ is differentiable with respect to its first argument on $(a_i, b_i) \times (a_i, b_i)$, and this partial derivative is continuous at all points of the form $(x_i, x_i) \in (a_i, b_i) \times (a_i, b_i)$.

(DR3) For all $y \in \Pi_{i=1}^n (a_i, b_i)$, $\tilde{d}(\cdot, y)$ is essentially smooth [25, Chapter 26].

(DR4) There exist $L, \epsilon > 0$ 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

$$|\tilde{d}_i'(x_i, y_i)| \leq L|x_i - y_i|.$$

If $\tilde{d}_i$ as in (4.9) verifies (DR1)–(DR4), then the distance

$$\tilde{d}(x, y) = \sum_{i=1}^n d_i(x, y) + \frac{\mu}{2} \|x - y\|^2$$

is called a *double regularization* for the constraint set B.

Remark 4.2. Assumptions (DR1)–(DR4) imply that $\tilde{d}$ verifies (d1)–(d2). Moreover, if $\mu > 1$, then it is proved in [17, Lemmas 3.3 and 3.4] that condition $(H3)$ holds with $\tilde{H}(x, y) := \frac{\mu+1}{2}\|x - y\|^2$ and $\gamma := \frac{\mu-1}{2}$. This pair $(\tilde{d}, \tilde{H})$ is a proximal pair because all other conditions (H1)–(H2) and (H4)–(H6) hold for $\tilde{H}(x, y) := \frac{\mu+1}{2}\|x - y\|^2$ when $\mu > 1$.

Remark 4.3. We point out that the exact solution of $(B11)$–$(B12)$, i.e., where $e^k = 0$ and $\varepsilon_k = 0$, exists for every double regularization. Indeed, it can be shown by using [9, Proposition 3] that $T + \beta \nabla_1 d(\cdot, y)$ is onto for every $y \in \operatorname{int} C$ and every $\beta > 0$.

4.3. Second order homogeneous proximal distances. It has been observed in [17] that second order homogeneous kernels are particular cases of double regularizations. Because of its importance, we present this example in detail. In this section we recall the definition and properties of the proximal method with second order homogeneous proximal distances. Let $\varphi : \mathbb{R} \to (-\infty, \infty]$ be given by

$$\varphi(t) := \mu h(t) + (\nu/2)(t - 1)^2,$$

where h is a closed and proper convex function satisfying the following additional properties:

1. h is twice continuously differentiable on $\operatorname{int}(\operatorname{dom} h) = (0, +\infty)$.
2. h is strictly convex on its domain.
3. $\lim_{t \to 0^+} h'(t) = -\infty$.
4. $h(1) = h'(1) = 0$ and $h''(1) > 0$.
5. For any $t > 0$,

$$h''(1)\left(1 - \frac{1}{t}\right) \leq h'(t) \leq h''(1)(1 - t).$$

Items (1–4) and items (1–5) were used in [6] to define the families Φ and Φ_2, respectively. The positive parameters ν, μ shall satisfy the following inequality:

$$\nu > \mu\, h''(1) > 0. \tag{4.10}$$

Recall that the generalized distance $d_\varphi(x, y)$, defined for $x, y \in \mathbb{R}^n_{++}$, is given by

$$d_\varphi(x, y) := \sum_{i=1}^{n} y_i^2 \varphi(x_i/y_i). \tag{4.11}$$

The following lemma, which has a crucial role in the convergence analysis, has been established in [4, Lemma 3.4].

LEMMA 4.3. *For any $x, y \in \mathbb{R}^n_{++}$ and $z \in \mathbb{R}^n_{+}$,*

$$\langle \nabla_1 d_\varphi(x, y), x - z\rangle \geq \left(\frac{\nu + \alpha\mu}{2}\right)(\|x - z\|^2 - \|y - z\|^2) + \left(\frac{\nu - \alpha\mu}{2}\right)\|x - y\|^2, \tag{4.12}$$

where $\alpha := h''(1)$.

Distances d defined as in (4.11) were studied, for instance, in [4, 5, 7, 2].

Remark 4.4. From Lemma 4.3, we see that $H_d(a, b) = \frac{\nu+\alpha\mu}{2}\|a - b\|^2$ verifies condition $(H3)$. All other conditions $(H1)$, $(H2)$, and $(H4)-(H6)$ trivially hold for this H_d.

The following result was proved in [6, Proposition 2.1], and guarantees the existence of an exact solution of (B11)–(B13).

PROPOSITION 4.4. *Assume that $D(T) \cap \mathbb{R}^n_{++} \neq \emptyset$. For any fixed $\lambda_k > 0$, $e^k \in \mathbb{R}^n$, and $x^{k-1} \in \mathbb{R}^n_{++}$, there exists a unique $x^k \in \mathbb{R}^n_{++}$ satisfying* (3.1)–(3.2) *for $d = d_\varphi$ defined as in* (4.11).

5. Convergence analysis.

In this section we establish convergence of the sequence generated by the inexact proximal point algorithm presented in section 3.

PROPOSITION 5.1. *Let d be a proximal distance with respect to the set C, and let H_d be the proximal distance induced by d. Assume that H_d verifies condition $(H3)$. Assume either of the following:*

(a) *$\{x_k\}$ is generated by $(B11), (B12)-(B13)$, with error criteria $(E2)$.*

(b) *$d = D_h$ Bregman distance, with $h(x) = \sum_{i=1}^n x_i \ln x_i$, and $\{x_k\}$ is generated by $(B11), (B12)$, and $(B13)'$, with error criteria $(E2)'$.*

Suppose also that z is a solution of the variational inequality problem $VIP(T, C)$. Then, for all k, we have

$$H_d(z, x^k) \leq H_d(z, x^{k-1}) - (1 - \sigma)\gamma H_d(x^k, x^{k-1}), \tag{5.1}$$

where γ is as in $(H3)$. Under assumption (b), $\gamma = 1$.

Proof. From $(B11)$ we have that $\tilde{v}^k \in T^{\varepsilon_k}(\tilde{x}^k)$, and from $(B12)$ we have that $\tilde{x}^k \in \operatorname{int} C$. Since z solves $VIP(T, C)$, there exists $w \in Tz$ such that

$$\langle w, \tilde{x}^k - z\rangle \geq 0. \tag{5.2}$$

Since $\tilde{v}^k \in T^{\varepsilon_k}(\tilde{x}^k)$, we have

$$\langle \tilde{v}^k - w, \tilde{x}^k - z\rangle \geq -\varepsilon_k. \tag{5.3}$$

Now by adding (5.2) and (5.3), we have

$$\langle \tilde{v}^k, \tilde{x}^k - z\rangle \geq -\varepsilon_k.$$

By adding and subtracting x^k in the above expression, we get

$$\langle \tilde{v}^k, \tilde{x}^k - x^k \rangle + \varepsilon_k \geq -\langle \tilde{v}^k, x^k - z \rangle.$$

Hence we have

$$\lambda_k \left[\varepsilon_k + \langle \tilde{v}^k, \tilde{x}^k - x^k \rangle \right] \geq \langle -\lambda_k \tilde{v}^k, x^k - z \rangle. \tag{5.4}$$

Assume now that (a) holds. From $(E2)$, we have

$$\lambda_k \left[\varepsilon_k + \langle \tilde{v}^k, \tilde{x}^k - x^k \rangle \right] \leq \sigma\gamma H_d(x^k, x^{k-1}),$$

where we used $c_2 = \sigma\gamma$. Further, from $(B13)$, we have

$$-\lambda_k \tilde{v}^k = \nabla_1 d(x^k, x^{k-1}).$$

Thus, by using (5.4), we have

$$\sigma\gamma H_d(x^k, x^{k-1}) \geq \langle \nabla_1 d(x^k, x^{k-1}), x^k - z \rangle.$$

Now, by $(H3)$, we have

$$\langle \nabla_1 d(x^k, x^{k-1}), x^k - z \rangle \geq H_d(z, x^k) - H_d(z, x^{k-1}) + \gamma H_d(x^k, x^{k-1}).$$

Thus, combining the above two expressions, we get

$$H_d(z, x^k) \leq H_d(z, x^{k-1}) - (1 - \sigma)\gamma H_d(x^k, x^{k-1}).$$

Hence we get the result under assumption (a). If (b) holds, using $(E2)'$ we have

$$\lambda_k \left[\varepsilon_k + \langle \tilde{v}^k, \tilde{x}^k - x^k \rangle \right] \leq \sigma r_{k-1} H_d(x^k, x^{k-1}) = \sigma r_{k-1} D_h(x^k, x^{k-1}).$$

From $(B13)'$, we have

$$-\lambda_k \tilde{v}^k = r_{k-1} \nabla_1 d(x^k, x^{k-1}).$$

Using (5.4) again, we have

$$\sigma r_{k-1} D_h(x^k, x^{k-1}) \geq r_{k-1} \langle \nabla_1 d(x^k, x^{k-1}), x^k - z \rangle.$$

Simplifying the expression and using the three point property (4.1), we have

$$\sigma D_h(x^k, x^{k-1}) \geq D_h(z, x^k) - D_h(z, x^{k-1}) + D_h(x^k, x^{k-1}).$$

Hence we get the result under assumption (b). □

LEMMA 5.2. *Let d be a proximal distance with respect to the set C, and let H_d be the proximal distance induced by d. Assume that H_d verifies conditions $(H3)-(H6)$. Assume either of the following:*

(a) *$\{x_k\}$ is generated by $(B11)$, $(B12)-(B13)$, with error criteria $(E2)$.*
(b) *$d = D_h$ with $h(x) = \sum_{i=1}^n x_i \ln x_i$, and $\{x_k\}$ is generated by $(B11)$, $(B12)$, and $(B13)'$ with error criteria $(E2)'$.*

If there exists $\bar{x}$, an accumulation point of $\{x^k\}$ which solves $VIP(T, C)$, then the whole sequence $\{x^k\}$ converges to $\bar{x}$.

Proof. Call $\{x^{k_j}\}$ a subsequence converging to $\bar{x}$. From $(H4)$ we know that

$$\lim_j H_d(\bar{x}, x^{k_j}) = 0. \tag{5.5}$$

Proposition 5.1 with $z := \bar{x}$ becomes

$$H_d(\bar{x}, x^k) \le H_d(\bar{x}, x^{k-1}) - (1-\sigma)\gamma H_d(x^k, x^{k-1}),$$

where $\gamma = 1$ under assumption (b). Hence the sequence $\{H_d(\bar{x}, x^k)\}$ is decreasing, with a subsequence converging to zero by (5.5). This yields

$$\lim_k H_d(\bar{x}, x^k) = 0. \tag{5.6}$$

Using now Lemma 2.3, we conclude that the whole sequence $\{x^k\}$ converges to $\bar{x}$. ∎

Let us now present a corollary which will be helpful in proving the main convergence result.

COROLLARY 5.3. *Let d be a proximal distance with respect to the set C, and let H_d be the proximal distance induced by d. Assume that H_d verifies conditions $(H2)$–$(H6)$ and that the solution set of $VIP(T, C)$ is nonempty. Assume either of the following:*

(a) *$\{x_k\}$ is generated by $(B11)$, $(B12)$–$(B13)$, with error criteria $(E1)$ and $(E2)$.*

(b) *$d = D_h$ with $h(x) = \sum_{i=1}^n x_i \ln x_i$, and $\{x_k\}$ is generated by $(B11)$, $(B12)$, and $(B13)'$ with error criteria $(E1)$ and $(E2)'$.*

Then the following hold:

(i) *The sequence $\{x^k\}$ is bounded.*

(ii) $\sum_{k=1}^\infty H_d(x^k, x^{k-1}) < \infty$.

(iii) $\lim_{k\to\infty} \|x^k - \tilde{x}^k\| = \lim_{k\to\infty} H_d(\tilde{x}^k, x^k) = 0$.

(iv) *The sequence $\{\tilde{x}^k\}$ is bounded.*

Proof. Let us fix $\bar{z} \in (T + N_C)^{-1}(0)$. Then from (5.1) it is easy to deduce that

$$0 \le H_d(\bar{z}, x^k) \le H_d(\bar{z}, x^{k-1}) \le H_d(\bar{z}, x^0),$$

where x^0 is the first iterate. Further, we also know from $(H2)$ that the set

$$\{y \in \operatorname{int} C : H_d(\bar{z}, y) \le H_d(\bar{z}, x^0)\}$$

is bounded. Hence (i) holds. Again from (5.1), we deduce that

$$H_d(x^k, x^{k-1}) \le \frac{1}{(1-\sigma)\gamma}\left[H_d(\bar{z}, x^{k-1}) - H_d(\bar{z}, x^k)\right].$$

Now for any natural number m, we have

$$\sum_{k=1}^m H_d(x^k, x^{k-1}) \le \frac{1}{(1-\sigma)\gamma}\left[H_d(\bar{z}, x^0) - H_d(\bar{z}, x^m)\right].$$

Since $H_d(\bar{z}, x^m) \ge 0$, we deduce that

$$\sum_{k=1}^m H_d(x^k, x^{k-1}) \le \frac{H_d(\bar{z}, x^0)}{(1-\sigma)\gamma} < \infty.$$

Hence (ii) holds. Using (E1) we conclude that $\lim_{k\to\infty} H_d(\tilde{x}^k, x^k) = 0$. From Lemma 2.4(a), we have that $\{\tilde{x}^k\}$ is bounded. So (iv) holds. Moreover, using now (iv), the fact that $\lim_{k\to\infty} H_d(\tilde{x}^k, x^k) = 0$, and Lemma 2.4(b), we get $\lim_{k\to\infty} \|\tilde{x}^k - x^k\| = 0$. ∎

Our main convergence results are a consequence of the following two results.

THEOREM 5.4. *Let d be a proximal distance with respect to the set C, and let H_d be the proximal distance induced by d such that H_d verifies $(H1)$–$(H3)$ and $(H5)$–$(H6)$. Assume either of the following:*

(a) *d is a double regularization with $\mu > 1$, and $\{x^k\}$ is a sequence generated by $(B11)$–$(B13)$ with error criteria $(E1)$ and $(E2)$.*

(b) *$d = D_h$ is the Bregman distance induced by $h(x) = \sum_1^n x_i \ln x_i$, and $\{x^k\}$ is a sequence generated by $(B11), (B12)$, and $(B13)'$ with error criteria $(E1)$ and $(E2)'$.*

If the solution set of $VIP(T, C)$ is nonempty, then there exists an infinite set $J \subset \mathbb{N}$ such that the subsequence $\{x^j\}_{j\in J}$ converges and

$$\liminf_{j\in J} \langle y - x^j, \tilde{v}^j \rangle \geq 0 \tag{5.7}$$

for every $y \in C$.

Proof. From Corollary 5.3, we have that $\{x^k\}$ is bounded, so there exists an infinite set K and a subsequence $\{x^k\}_{k\in K}$ converging to some $\bar{x}$. Note that

$$\langle \tilde{v}^k, y - x^k \rangle = \frac{1}{\lambda_k} \langle -\lambda_k \tilde{v}^k, x^k - y \rangle.$$

Assume that (a) holds. Then by $(B13)$, we can write

$$\langle \tilde{v}^k, y - x^k \rangle = \frac{1}{\lambda_k} \langle \nabla_1 d(x^k, x^{k-1}), x^k - y \rangle. \tag{5.8}$$

Condition $(H3)$ yields

$$\begin{aligned} \frac{1}{\lambda_k} \langle \nabla_1 d(x^k, x^{k-1}), x^k - y \rangle \geq \frac{1}{\lambda_k} &\left[H_d(y, x^k) - H_d(y, x^{k-1}) \right] \\ &+ \frac{\gamma}{\lambda_k} H_d(x^k, x^{k-1}). \end{aligned} \tag{5.9}$$

Because d is a double regularization, we know by Remark 4.2 that $H_d(x, y) = \frac{\mu+1}{2}\|x - y\|^2$ and $\gamma = \frac{\mu-1}{2}$. Considering (5.9) for $k \in K$ yields

$$\begin{aligned} \frac{2}{\lambda_k} \langle \nabla_1 d(x^k, x^{k-1}), x^k - y \rangle \geq \frac{\mu+1}{\lambda_k} &\left[\left\|y - x^k\right\|^2 - \left\|y - x^{k-1}\right\|^2 \right] \\ &+ \frac{\mu^2 - 1}{\lambda_k} \left\|x^k - x^{k-1}\right\|^2. \end{aligned} \tag{5.10}$$

The right-hand side tends to zero because $\{x^k\}_{k\in K}$ converges to $\bar{x}$. This proves (5.7) for $J := K$ in case (a). Assume now that (b) holds, so $h(x) = \sum_{i=1}^n x_i \ln x_i$. Let us write, for simplicity, $h_i(x_i) := x_i \ln x_i$. Therefore $h_i'(x_i) := \ln x_i + 1$ and $h_i''(x_i) := \frac{1}{x_i}$. Then by $(B13)'$, we can write for $k \in K$

$$\begin{aligned} \langle \tilde{v}^k, y - x^k \rangle &= \frac{r_{k-1}}{\lambda_k} \langle \nabla h(x^k) - \nabla h(x^{k-1}), x^k - y \rangle \\ &= \frac{r_{k-1}}{\lambda_k} \sum_{i=1}^n (h_i'(x_i^k) - h_i'(x_i^{k-1}))(x_i^k - y_i) \\ &= \frac{1}{\lambda_k} \sum_{i=1}^n r_{k-1} \ln \left(\frac{x_i^k}{x_i^{k-1}} \right) (x_i^k - y_i). \end{aligned} \tag{5.11}$$

Call $q_i^k(y_i) := r_{k-1}\ln(\frac{x_i^k}{x_i^{k-1}})(x_i^k - y_i)$ and $I_0 := \{i \in \{1,\dots,n\} \;:\; \bar{x}_i = 0\}$. We have that

$$\begin{aligned}\liminf_{k\in K}\langle \tilde{v}^k, y - x^k\rangle &= \liminf_{k\in K}\frac{1}{\lambda_k}\sum_{i=1}^n q_i^k(y_i)\\ &\geq \sum_{i=1}^n \liminf_{k\in K}\frac{1}{\lambda_k}q_i^k(y_i)\\ &= \sum_{i\in I_0}\liminf_{k\in K}\frac{1}{\lambda_k}q_i^k(y_i) + \sum_{i\notin I_0}\liminf_{k\in K}\frac{1}{\lambda_k}q_i^k(y_i)\\ &= S_0 + S_1. \end{aligned} \tag{5.12}$$

Note that $S_1 = 0$ by continuity of h_i' in $\mathbb{R}_{++}$ and the facts that $\lambda_k \geq \hat{\lambda} > 0$ and $\{r_{k-1}\}$ is a bounded sequence. We must show that $S_0 \geq 0$. To prove that $S_0 \geq 0$, we will prove the following claim.

Claim. There exists an infinite index set $K_0 \subset K$, K_0 not depending on y, such that

$$\liminf_{j\in K_0} q_i^j(y_i) \geq 0 \tag{5.13}$$

for every $i \in I_0$. Because $\lambda_j \geq \bar{\lambda} > 0$, the above expression, Lemma 2.5, and (5.12) yield (5.7) for the choice $J := K_0$. Let $p := |I_0|$ be the number of elements in the set I_0, and assume without loss of generality that $I_0 := \{1,\dots,p\}$. Define the set

$$K_1 := \{k \in K \;:\; x_1^k < x_1^{k-1}\}.$$

We have two possible cases to consider.
Case 1. The set K_1 is infinite.
Case 2. The set K_1 is finite, so there exists an index $k_0 \in K$ such that $x_1^k \geq x_1^{k-1}$ for all $k \in K$ such that $k \geq k_0$. In this situation, call $\bar{K}_1 := \{k \in K \;:\; k \geq k_0\}$.

Suppose we are in Case 1. Assume first that $y_1 = 0$. For every $k \in K_1$, we have $x_1^k < x_1^{k-1}$, so $\ln(\frac{x_1^k}{x_1^{k-1}}) < 0$, and by the mean value theorem there exists $\theta_k \in (x_1^k, x_1^{k-1})$ such that $\ln(\frac{x_1^k}{x_1^{k-1}}) = (1/\theta_k)(x_1^k - x_1^{k-1}) < 0$. Hence we can write

$$\begin{aligned}|q_1^k(y_1)| = |q_1^k(0)| &= r_{k-1}\ln\left(\frac{x_1^{k-1}}{x_1^k}\right)x_1^k\\ &= r_{k-1}\frac{(x_1^{k-1}-x_1^k)}{\theta_k}x_1^k \leq \frac{x_1^{k-1}(x_1^{k-1}-x_1^k)}{\theta_k}x_1^k\\ &\leq (x_1^{k-1}-x_1^k)\,x_1^{k-1},\end{aligned}$$

where we used the definition of r_{k-1} in the first inequality and the fact that $x_1^k < \theta_k$ in the second one. Note that the right-hand side of the expression above tends to zero because the sequences $\{x_1^{k-1} - x_1^k\}$ and $\{x_1^{k-1}\}$ tend to zero. Hence, in Case 1 when $y_1 = 0$, we have $\liminf_{k\in K_1} q_1^k(0) = 0$. Assume now that we are in Case 1 with $y_1 > 0$. Because x_1^k tends to zero, there exists $k_1 = k_1(y_1) \in K_1$ such that $(x_1^k - y_1) < 0$ for every $k \in K_1$ such that $k \geq k_1$. Note also that $\ln(x_1^k/x_1^{k-1}) < 0$ for every $k \in K_1$. Hence, we can write

$$\begin{aligned}\liminf_{k\in K_1} q_1^k(y_1) &= \sup_{j\in\mathbb{N}}\inf_{k\geq j,\,k\in K_1} r_{k-1}\ln(x_1^k/x_1^{k-1})(x_1^k - y_1)\\ &\geq \inf_{k\geq k_1,\,k\in K_1} r_{k-1}\ln(x_1^k/x_1^{k-1})(x_1^k - y_1) \geq 0,\end{aligned}$$

where we used the definition of k_1. Note that even though k_1 depends on y_1, the overall $\liminf_{k\in K_1} q_1^k(y_1) \geq 0$ for every choice of y_1.

Suppose we are in Case 2. By definition of $\bar{K}_1$, we have that $x_1^k \geq x_1^{k-1}$ for all $k \in \bar{K}_1$. We will show that $\liminf_{k\in\bar{K}_1} q_1^k(y_1) = 0$ in this case. For every $k \in \bar{K}_1$, we have $\ln(x_1^k/x_1^{k-1}) \geq 0$. Using the mean value theorem again, we can write

$$\begin{aligned} |q_1^k(y_1)| &= r_{k-1}\ln(x_1^k/x_1^{k-1})|x_1^k - y_1| \leq x_1^{k-1}\ln(x_1^k/x_1^{k-1})|x_1^k - y_1| \\ &= (x_1^{k-1}/\theta_k)(x_1^k - x_1^{k-1})|x_1^k - y_1| \leq (x_1^k - x_1^{k-1})|x_1^k - y_1| \end{aligned}$$

for every $k \in \bar{K}_1$. Because $(x_1^k - x_1^{k-1})$ tends to zero when $k \in \bar{K}_1$ goes to infinity and the term $|x_1^k - y_1|$ is bounded, we conclude that $\liminf_{k\in\bar{K}_1} q_1^k(y_1) = 0$ as wanted. We thus proved that there exists $J_1 \subset K$ such that J_1 is infinite, J_1 does not depend on y_1, and

$$\liminf_{j\in J_1} q_1^j(y_1) \geq 0. \tag{5.14}$$

Define now

$$K_2 := \{k \in J_1 \, : \, x_2^k < x_2^{k-1}\}.$$

Again we have the two cases above to consider, and in the same way as above, we can obtain an infinite set $J_2 \subset J_1$ such that

$$\liminf_{j\in J_2} q_2^j(y_2) \geq 0, \tag{5.15}$$

where J_2 does not depend on y_2. Iterating this process p times, we obtain p infinite sets $\{J_1, \ldots, J_p\}$ such that $J_p \subset J_{p-1} \subset \cdots \subset J_1 \subset K$ satisfying

$$\liminf_{j\in J_l} q_l^j(y_l) \geq 0 \ \forall\, l = 1, \ldots, p \tag{5.16}$$

with J_l independent of y_l for all $l = 1, \ldots, p$. Take now $J_0 := J_p$. Because $J_p \subset J_i$ for all $i \in I_0$, we have that

$$\liminf_{j\in J_0} q_i^j(y_i) \geq 0 \tag{5.17}$$

for all $i \in I_0$. This proves the claim for $K_0 = J_0$. Hence (5.7) holds under assumption (b) for the infinite set $J = J_0$. ◻

LEMMA 5.5. *Consider the problem $VIP(T, C)$, and let the sequence $\{(z^k, u^k)\}$ be an arbitrary sequence such that $\{z^k\}$ is bounded and there exists a subsequence $\{(z^{k_j}, u^{k_j})\}$ of $\{(z^k, u^k)\}$ which verifies*

$$\liminf_{j\to\infty} \langle y - z^{k_j}, u^{k_j}\rangle \geq 0 \tag{5.18}$$

for all $y \in C$. Assume that the subsequence $\{(z^{k_j}, u^{k_j})\}$ also verifies

$$\liminf_{j} \langle w - u^{k_j}, y - z^{k_j}\rangle \geq 0 \tag{5.19}$$

for all $y \in C$ and that all $w \in (T + N_C)y$. Then every accumulation point of $\{z^{k_j}\}$ is a solution of $VIP(T, C)$.

Proof. Because $\{z^{k_j}\}$ is bounded, there exists a subsequence (which we still denote $\{z^{k_j}\}$ for simplicity) converging to some $\bar{z}$. We can write

$$\langle y - z^{k_j}, w\rangle = \langle y - z^{k_j}, w - u^{k_j}\rangle + \langle y - z^{k_j}, u^{k_j}\rangle.$$

Using the properties of lim inf and the definition of $\bar{z}$ in the above expression, we obtain

$$\langle y - \bar{z}, w\rangle = \liminf_j \langle y - z^{k_j}, w\rangle \geq \liminf_j \langle y - z^{k_j}, w - u^{k_j}\rangle + \liminf_j \langle y - z^{k_j}, u^{k_j}\rangle \geq 0,$$

where we used (5.19) and the fact that the subsequence $\{(z^{k_j}, u^{k_j})\}$ verifies (5.18). Because $y \in C$ and $w \in (T + N_C)y$ are arbitrary, this implies that $\bar{z}$ solves $VIP(T, C)$. □

THEOREM 5.6. *Let d be a proximal distance with respect to the set C, and let H_d be the proximal distance induced by d, such that H_d verifies $(H1)-(H3)$ and $(H5)-(H6)$. Assume either of the following:*

(a) *d is a double regularization with $\mu > 1$, and $\{x^k\}$ is a sequence generated by $(B11)-(B13)$ with error criteria $(E1)$ and $(E2)$.*

(b) *$d = D_h$ is the Bregman distance induced by $h(x) = \sum_{i=1}^n x_i \ln x_i$, and $\{x^k\}$ is a sequence generated by $(B11)$, $(B12)$, and $(B13)'$ with error criteria $(E1)$ and $(E2)'$.*

If the solution set of $VIP(T, C)$ is nonempty, then every accumulation point of the sequence $\{x^k\}$ is a solution of variational inequality $VIP(T, C)$.

Proof. Using Theorem 5.4, under either assumption (a) or (b), there exists an infinite set $K \subset \mathbb{N}$ such that $\{x^k\}_{k\in K}$ converges to, say, $\bar{x}$, and for every $y \in C$,

$$\liminf_{k\in K} \langle y - x^k, \tilde{v}^k\rangle \geq 0. \tag{5.20}$$

Our next step is to prove that for every $y \in C$ and every $w \in (T + N_C)y$, we have

$$\liminf_{k\in K} \langle y - x^k, w - \tilde{v}^k\rangle \geq 0, \tag{5.21}$$

where $\{x^k\}_{k\in K}$ and $\{\tilde{v}^k\}_{k\in K}$ are as in assumption (a) or (b). Take $y \in C$ and $w \in (T + N_C)y$. Because $D(T) \cap \operatorname{int} C \neq \emptyset$, we can write

$$w \in (T + N_C)(y) = T(y) + N_C(y).$$

Thus we have $w = u + r$, where $u \in T(y)$ and $r \in N_C(y)$. For every $k \in K$, we have

$$\langle y - x^k, w - \tilde{v}^k\rangle = \langle y - x^k, u - \tilde{v}^k\rangle + \langle y - x^k, r\rangle.$$

Since $r \in N_C(y)$, we have

$$\langle y - x^k, r\rangle \geq 0.$$

This shows that

$$\langle y - x^k, w - \tilde{v}^k\rangle \geq \langle y - x^k, u - \tilde{v}^k\rangle. \tag{5.22}$$

Further, we can write

$$\langle y - x^k, u - \tilde{v}^k\rangle = \langle y - \tilde{x}^k, u - \tilde{v}^k\rangle + \langle \tilde{x}^k - x^k, u - \tilde{v}^k\rangle. \tag{5.23}$$

Since $u \in T(y)$ and $\tilde{v}^k \in T^{\varepsilon_k}(\tilde{x}^k)$, we have

$$\langle y - \tilde{x}^k, u - \tilde{v}^k \rangle \geq -\varepsilon_k.$$

Thus from (5.23), we have

$$\langle y - x^k, u - \tilde{v}^k \rangle \geq -\varepsilon_k + \langle \tilde{x}^k - x^k, u - \tilde{v}^k \rangle.$$

Using now (5.22) yields

$$\langle y - x^k, w - \tilde{v}^k \rangle \geq -\varepsilon_k + \langle \tilde{x}^k - x^k, u - \tilde{v}^k \rangle.$$

The above expression can be further rearranged as

$$\langle y - x^k, w - \tilde{v}^k \rangle \geq - \left[\varepsilon_k + \langle \tilde{x}^k - x^k, \tilde{v}^k \rangle \right] + \langle \tilde{x}^k - x^k, u \rangle.$$

Under assumption (a), we use $(E2)$ and the Cauchy–Schwartz inequality to obtain

$$\langle y - x^k, w - \tilde{v}^k \rangle \geq \frac{-\sigma\gamma}{\lambda_k} H_d(x^k, x^{k-1}) - ||u|| \, ||\tilde{x}^k - x^k||. \tag{5.24}$$

The right-hand side of the expression above tends to zero by Corollary 5.3, items (ii) and (iii). Therefore, we have

$$\liminf_{k \in K} \langle y - x^k, w - \tilde{v}^k \rangle \geq 0, \tag{5.25}$$

where we used again the fact that λ_k is bounded away from zero for all k. From (5.20), (5.21), and Lemma 5.5, we have that there exists an accumulation point of $\{x^k\}$ which solves $VIP(T, C)$. Using now Lemma 5.2, we conclude that the whole sequence $\{x^k\}$ converges to that solution.

Under assumption (b), we use $(E2)'$ and the Cauchy–Schwartz inequality to obtain

$$\langle y - x^k, w - \tilde{v}^k \rangle \geq \frac{-\sigma r_{k-1}}{\lambda_k} D_h(x^k, x^{k-1}) - ||u|| \, ||\tilde{x}^k - x^k||. \tag{5.26}$$

Recall that $r_{k-1} = \min_{i=1,\ldots,n}\{x_i^{k-1}\}$, so there exists $L > 0$ such that $r_{k-1} \leq L$ for every k. Using this fact and taking limits in the above expression, we conclude that

$$\liminf_{k \in K} \langle y - x^k, w - \tilde{v}^k \rangle \geq 0,$$

where we used also the fact that λ_k is bounded away from zero for all k. From (5.20), (5.21), and Lemma 5.5, we have that there exists an accumulation point of $\{x^k\}$ which solves $VIP(T, C)$. Using now Lemma 5.2, we conclude that the whole sequence $\{x^k\}$ converges to that solution. ∎

6. A specific implementation. We discuss in this section the computational feasibility of the algorithm (B11)–(B12)(B13)′, with error criteria (E1)(E2)′. Our analysis shows the usefulness of the ε-enlargement of T in the implementation of the relative error criteria. To our knowledge, these features cannot be exploited by using a summable error criterion as the one on [27].

In [13, Section 5] it is shown how the inexact scheme induced by the log-quadratic distance can take full advantage of the enlargement of T and avoid the extragradient step, thus resulting in a simple implementation of the method. Key parts of the analysis in the latter reference rely on the strong convexity of the log-quadratic regularization. Furthermore, the operator T is assumed to be point-to-point in [13, section 5]. In the case we study below, we no longer have the strong convexity assumption, and we analyze the general point-to-set case. Hence, our analysis is significantly more involved than that in [13, section 5].

In what follows, we consider the case in which the constraint set is $C = \mathbb{R}^n_+$, and our generalized distance is the Bregman distance D_h induced by $h(x) = \sum_{i=1}^n x_i \log x_i$. In this situation, problem $VIP(T, \mathbb{R}^n_+)$ is equivalent to $VIP(T + N_{\mathbb{R}^n_+}, \mathbb{R}^n_+)$:

$$\begin{array}{c}\text{Find } \bar{x} \text{ such that} \\ \exists\, \bar{v} \in (T + N_{\mathbb{R}^n_+})\bar{x},\ \langle \bar{v}, y - \bar{x}\rangle \geq 0\ \forall\, y \in \mathbb{R}^n_+\end{array} \tag{6.1}$$

because they have the same solution set. So we may apply our algorithm to the original $VIP(T, \mathbb{R}^n_+)$ or to (6.1) in order to get a solution.

If the algorithm is applied to (6.1), the k-th proximal system consists of finding x such that

$$\left\{\begin{array}{l} v \in (T + N_{\mathbb{R}^n_+})(x), \\ \lambda_k v + \nabla_1 D_h(x, x^{k-1}) = 0. \end{array}\right. \tag{6.2}$$

Because the domain of $\nabla_1 D_h(\cdot, x^{k-1})$ is $\mathbb{R}^n_{++}$, the exact solutions of (3.1)–(3.2) and (6.2) coincide. For each $x \in \mathbb{R}^n_{++}$,

$$T^\varepsilon(x) \subset (T + N_{\mathbb{R}^n_+})^\varepsilon(x).$$

Then the set of approximated solutions (in the sense of Definition 3.2) of (6.2) is "potentially" bigger than the corresponding set for (3.1)–(3.2). Moreover, the following result, which is [13, Lemma 2.2], provides a practical method for generating elements in $(T + N_{\mathbb{R}^n_+})^\varepsilon(x)$.

LEMMA 6.1. *Let T be a maximal monotone operator on $\mathbb{R}^n$, with $dom\, T \subset \mathbb{R}^n_+$. If $v \in T(x)$, then for each $w \in \mathbb{R}^n_+$,*

$$(v - w) \in T^\varepsilon(x)$$

for $\varepsilon \geq \langle w, x\rangle$.

The following analysis focuses on the computation of an approximated solution of (6.2), in the sense of Definition 3.2. For brevity, an approximated solution in the sense of Definition 3.2 will be called in this section an approximated solution.

At each iteration, two computations must be performed. The first one is to find $\tilde{x} \in \mathbb{R}^n_{++}$, $\tilde{v}$, ε an approximated solution of the system (3.1)–(3.2) or (6.2). Given $\tilde{v}$ obtained in the first computation, the second computation is the extragradient-like step (B13)$'$, which for our choice of D_h can be restated as requiring x^k to satisfy

$$r_{k-1} \nabla_1 D_h(x^k, x^{k-1}) = r_{k-1} \sum_{i=1}^n \log\left(\frac{x_i^k}{x_i^{k-1}}\right) e^i = -\lambda_k \tilde{v}, \tag{6.3}$$

where $r_{k-1} := \min_{i=1,\dots,n}\{x_i^{k-1}\}$ and $\{e^1, \dots, e^n\}$ is the canonical basis of $\mathbb{R}^n$. These two computations are interconnected because x^k is needed to verify whether $\tilde{x}$, $\tilde{v}$, ε satisfy (E1)(E2)$'$. So from the practical point of view, it is advantageous to devise a procedure that avoids this second computation. This is achieved in the implementation below, in which the extragradient step is automatically satisfied.

First let us observe that if $0 \in T(x^{k-1})$, then $(\tilde{x}^k, x^k, \tilde{v}^k, \varepsilon_k) = (x^{k-1}, x^{k-1}, 0, 0)$ solves (B11)(B12)(B13)′ and trivially verifies (E1)(E2)′. In this case, x^{k-1} is a solution of (6.1), and the algorithm may stop here. So from now on, we assume that x^{k-1} is not a solution of problem (6.1), i.e.,

$$0 \notin T(x^{k-1}).$$

In what follows, we denote $\gamma_k := \lambda_k / r_{k-1}$, where λ_k, r_{k-1} are as in (6.3). Fix $\beta > 0$, and for the current iterate $x^{k-1} \in \mathbb{R}^n_{++}$, define the point-to-set mapping

$$F_{\beta,k}(x) := \frac{\lambda_k}{r_{k-1}} T(x) + \nabla_1 D_h(x, x^{k-1}) - \beta \sum_{i=1}^{n} \frac{x_i^{k-1}}{x_i} e^i. \tag{6.4}$$

Our assumptions on T and γ_k imply that $F_{\beta,k}$ is strictly monotone. Using [9, Theorem 1] we have that $F_{\beta,k}$ is onto. The following result shows that the approximate solution $\tilde{x}$ can be chosen in such a way that the extragradient step is unnecessary.

THEOREM 6.2. *Fix $\theta \in (0,1)$, and consider $F_{\beta,k}$ as in* (6.4). *Suppose that $0 \notin T(x^{k-1})$. Let $\tilde{x} \in \mathbb{R}^n_{++}$ and $\tilde{w} \in F_{\beta,k}(\tilde{x})$ be such that*

$$|\,[\tilde{w}]_i\,| \le \theta\beta \frac{x_i^{k-1}}{\tilde{x}_i} \quad \forall i = 1, \dots, n, \tag{6.5}$$

where $[z]_i$ stands for the i-th coordinate of $z \in \mathbb{R}^n$. For all $i = 1, \dots, n$, define

$$\rho := \tilde{w} + \beta \sum_{i=1}^{n} \frac{x_i^{k-1}}{\tilde{x}_i} e^i,$$

$$[\hat{v}]_i := -\frac{1}{\gamma_k} \ln\left(\frac{\tilde{x}_i}{x_i^{k-1}}\right),$$

$$\varepsilon := \frac{1}{\gamma_k} \langle \tilde{x}, \rho \rangle.$$

Then we have the following:

(i) $\rho \ge 0$.
(ii) $\hat{v} \in (T + N_{\mathbb{R}^n_+})^{\varepsilon}(\tilde{x})$.
(iii) *The extragradient equation* (6.3) *holds for the choice $x^k = \tilde{x}$ and $\tilde{v} = \hat{v}$.*
(iv) *For small enough β, the choice $x^k := \tilde{x}$ verifies* (E1) *and* (E2)′.

Proof. We start by noting that for every $\theta, \beta > 0$ and every $x^{k-1} \in \mathbb{R}^n_{++}$, there exists $\tilde{x}$ verifying (6.5). Indeed, this is a consequence of the fact that $F_{(1+\theta)\beta,k}$ is onto. Let $\tilde{x}, \tilde{w}$ be as in (6.5). From the definitions of ρ and $\tilde{w}$, we can write

$$\rho_i - \beta \frac{x_i^{k-1}}{\tilde{x}_i} = [\tilde{w}]_i\,. \tag{6.6}$$

Combine the above expression with (6.5) to obtain

$$-\theta\beta \frac{x_i^{k-1}}{\tilde{x}_i} \le \rho_i - \beta \frac{x_i^{k-1}}{\tilde{x}_i} \le \theta\beta \frac{x_i^{k-1}}{\tilde{x}_i}.$$

This shows that

$$0 < (1-\theta)\beta \frac{x_i^{k-1}}{\tilde{x}_i} \le \rho_i \le \beta(1+\theta) \frac{x_i^{k-1}}{\tilde{x}_i},$$

which proves (i). To prove (ii), we use Lemma 6.1 and the fact that

$$T(\tilde{x}) = (T + N_{\mathbb{R}^n_+})(\tilde{x})$$

because $\tilde{x} \in \mathbb{R}^n_{++}$. Recall that $\tilde{w} \in F_{\beta,k}(\tilde{x})$, so there exists $\tilde{v} \in T(\tilde{x})$ such that

$$\tilde{w} = \gamma_k \tilde{v} + \nabla_1 D_h(\tilde{x}, x^{k-1}) - \beta \sum_{i=1}^{n} \frac{x_i^{k-1}}{\tilde{x}_i} e^i.$$

Therefore, $\tilde{v} \in (T + N_{\mathbb{R}^n_+})(\tilde{x})$. Combining this fact with the definitions of $\rho, \hat{v}, \varepsilon$ and Lemma 6.1 yields

$$\hat{v} = \tilde{v} - \frac{\rho}{\gamma_k} \in (T + N_{\mathbb{R}^n_+})^{\varepsilon}(\tilde{x}).$$

Let us now establish (iii). Take $\varepsilon_k := \varepsilon$, $\tilde{v}^k := \hat{v} \in (T + N_{\mathbb{R}^n_+})^{\varepsilon}(\tilde{x})$, and $\tilde{x}^k := \tilde{x}$ in $(B11)$. Also take $x^k := \tilde{x}$ in $(B13)'$. With these choices, we have

$$\nabla_1 D_h(x^k, x^{k-1}) = \nabla_1 D_h(\tilde{x}, x^{k-1}) = \gamma_k \frac{\nabla_1 D(\tilde{x}, x^{k-1})}{\gamma_k} = -\gamma_k \hat{v} = -\gamma_k \tilde{v}^k,$$

which is precisely $(B13)'$. We proceed to prove (iv). Note that our choice of $x^k = \tilde{x}$ implies that condition $(E1)$ is automatically satisfied. So it is enough to prove that $(E2)'$ will be satisfied for small enough β. Assume this is not true. This means that there exists a sequence $\beta_m \to 0$ and sequences $\{z^m\}, \{w^m\}$ satisfying

$$|\,[w^m]_i\,| \le \theta \beta_m \frac{x_i^{k-1}}{z_i^m} \quad \forall i = 1, \dots, n$$

such that there exists a subsequence $\{z^{m_j}\}_{j\in\mathbb{N}}$ of $\{z^m\}$ for which $(E2)'$ (with the choice $\tilde{x}^k = x^k := z^m$) is violated. For simplicity of exposition and without loss of generality, we still denote the subsequence $\{z^{m_j}\}_{j\in\mathbb{N}}$ as $\{z^m\}$. Because we propose the next iterate as $x^k := z^m$, the violation of condition $(E2)'$ for this choice can be written as

$$\gamma_k \varepsilon_m > \sigma D_h(z^m, x^{k-1}) \, \forall m \in \mathbb{N},$$

where we are using z^m instead of $\tilde{x}$ in all the above definitions, i.e., $\varepsilon_m = \frac{1}{\gamma_k}\langle z^m, \rho^m \rangle$. The fact that $w^m \in F_{\beta,k}(z^m)$ implies that

$$\rho^m = w^m + \beta \sum_{i=1}^{n} \frac{x_i^{k-1}}{z_i^m} e^i = \gamma_k v^m + \nabla_1 D_h(z^m, x^{k-1})$$

for some $v^m \in T(z^m)$. Using the definition of ε_m, we have

$$\langle z^m, \rho^m \rangle > \sigma D_h(z^m, x^{k-1}). \tag{6.7}$$

By the same arguments as those used for case (i), we have that

$$0 < \beta_m (1 - \theta) \frac{x_i^{k-1}}{z_i^m} \le \rho_i^m \le \beta_m (1 + \theta) \frac{x_i^{k-1}}{z_i^m}. \tag{6.8}$$

Multiplying the above inequality $z_i^m > 0$, we obtain $z_i^m \rho_i^m \leq \beta_m(1+\theta)x_i^{k-1}$. Also using (6.7), we can conclude that

$$\sigma D_h(z^m, x^{k-1}) < \langle z^m, \rho^m \rangle \leq \beta_m(1+\theta)||x^{k-1}||_1,$$

where $\|u\|_1 = \sum_{i=1}^n |u_i|$ is the l_1-norm in $\mathbb{R}^n$. Now letting $m \to \infty$, i.e., when $\beta_m \to 0$ in the above inequality, we conclude using Lemma 2.4 that $z^m \to x^{k-1}$. Now using this fact, we can conclude from (6.8) that $\rho^m \to 0$. Now from the definition of ρ^m and the last two facts, we conclude that $v^m \to 0$. Altogether, we have that $z^m \to x^{k-1}$, $v^m \in T(z^m)$, and $v^m \to 0$. By maximal monotonicity, the graph of T is closed, so we conclude that $0 \in T(x^{k-1})$, a contradiction. This proves that for $\beta > 0$ small enough, we must satisfy condition $(E2)'$. This completes the proof. ☐

We have thus established that (6.5) provides suitable approximated solutions for our method. The following result proves that these approximate solutions x can be updated in a natural way. More precisely, if the norm of some $v \in F_{\beta,k}(x)$ is small enough, then x verifies (6.5).

For $r \geq 0$, we denote by $B[0,r]$ the closed ball of $\mathbb{R}^n$, with radius r and center $0 \in \mathbb{R}^n$.

THEOREM 6.3. *Assume that* $0 \notin T(x^{k-1})$. *Fix* $\theta > 0$ *and* $\beta > 0$. *There exists* $\delta > 0$ *small enough for which the following is true.*

$$If \quad F_{\beta,k}(x) \cap B[0,\delta] \neq \emptyset, \textit{ then } (6.5) \textit{ holds for } x \textit{ with parameters } \theta, \beta.$$

Proof. For simplicity of exposition, denote $F_{\beta,k}$ by F. To prove the claim, we will show that if the sequence $\{(x^j, v^j)\} \subset G(F)$ satisfies $v^j \to 0$, then for j large enough we must have

$$|\left[v^j\right]_i x_i^j| < \theta\beta x_i^{k-1}. \tag{6.9}$$

Fix therefore a sequence $\{(x^j, v^j)\} \subset G(F)$ such that $v^j \to 0$, and let x^* be such that $0 \in F(x^*)$. Such x^* exists by surjectivity of $F = F_{\beta,k}$. It is enough to prove that the sequence $\{x^j\}$ is bounded because in this case, (6.9) will be true for j large enough. Indeed, if $\{x^j\}$ is bounded, then the left-hand side of the inequality (6.9) tends to zero and the right-hand side is a fixed positive number, so for j large enough, the inequality in (6.9) will hold. Therefore, let us prove that $\{x^j\}$ is bounded. We can write

$$\begin{aligned} L_j &:= \langle x^j - x^*, v^j - 0 \rangle \\ &\geq \sum_{i=1}^n \left(x_i^j - x_i^*\right) \left[\ln\left(\frac{x_i^j}{x_i^*}\right) + \beta x_i^{k-1}\left(\frac{1}{x_i^*} - \frac{1}{x_i^j}\right)\right] =: R_j, \end{aligned} \tag{6.10}$$

where we used the definition of F, the fact that $0 \in F(x^*)$, and the monotonicity of T. For all $i \in \{1,\dots,n\}$, define $\varphi_i : \mathbb{R}_{++} \to \mathbb{R}$ as $\varphi_i(t) := \ln(t) - \frac{\beta x_i^{k-1}}{t}$. Then φ_i is strictly increasing for all $i \in \{1,\dots,n\}$. Hence, each term of the sum R_j is nonnegative, and therefore it is bounded below by any of its individual terms:

$$\begin{aligned} R_j &\geq (x_i^j - x_i^*)\left[\ln\left(\frac{x_i^j}{x_i^*}\right) + \beta x_i^{k-1}\left(\frac{1}{x_i^*} - \frac{1}{x_i^j}\right)\right] \\ &= (x_i^j - x_i^*)\left[\varphi_i(x_i^j) - \varphi_i(x_i^*)\right] \\ &= |x_i^j - x_i^*| \; |\varphi_i(x_i^j) - \varphi_i(x_i^*)| \geq 0. \end{aligned} \tag{6.11}$$

Using also the Cauchy–Schwartz inequality and the fact that the l_2-norm is smaller than n times the l_∞-norm, we can write

$$L_j \quad \leq \quad \|v^j\| \, \|x^j - x^*\| \leq n\|v^j\| \, \|x^j - x^*\|_\infty. \tag{6.12}$$

If the sequence $\{x^j\}$ is unbounded, then there exists an infinite set $J \subset \mathbb{N}$ and an index $i_0 \in \{1, \dots, n\}$ such that $\lim_{j\in J,\, j\to\infty} \|x^j - x^*\|_\infty = \lim_{j\in J,\, j\to\infty} |x^j_{i_0} - x^*_{i_0}| = \infty$. Such an index i_0 must exist because we have a finite number of coordinates, and hence one of these coordinates must verify the equality $\|x^j - x^*\|_\infty = |x^j_{i_0} - x^*_{i_0}|$ an infinite number of times. Combining these facts with (6.10), (6.12), and (6.12) for $i = i_0$, we can write for $j \in J$

$$\begin{aligned} n\|v^j\| \, \|x^j - x^*\|_\infty \geq R_j \geq |x^j_{i_0} - x^*_{i_0}| \; |\varphi_{i_0}(x^j_{i_0}) - \varphi_{i_0}(x^*_{i_0})| \\ = \|x^j - x^*\|_\infty |\varphi_{i_0}(x^j_{i_0}) - \varphi_{i_0}(x^*_{i_0})|, \end{aligned}$$

which simplifies to

$$\|v^j\| \geq \frac{|\varphi_{i_0}(x^j_{i_0}) - \varphi_{i_0}(x^*_{i_0})|}{n}.$$

The definition of the coordinate i_0 implies, in particular, that $\lim_{j\in J,\, j\to\infty} x^j_{i_0} = +\infty$. The latter fact yields $\lim_{t\to\infty} \varphi_{i_0}(t) = +\infty$. Altogether, we have that the right-hand side of the expression above tends to infinity for $j \in J$, $j \to \infty$, while the left-hand side tends to zero. This entails a contradiction, and therefore the sequence $\{x^j\}$ is bounded. As mentioned before, this proves the statement of the theorem. ∎

7. Concluding remarks. Using the concept of proximal distances introduced in [2], we have presented here two families of inexact proximal point methods. One of them (Definition 3.1) uses generalized distances which have a quadratic-like behavior, such as double regularizations and log-quadratic distances. In particular, we show that Definition 3.1 includes the hybrid proximal method of [28] as a particular case. More precisely, when $H_d(x, y) = \frac{\|x-y\|^2}{2}$, then the method we propose is equivalent to the hybrid proximal method of [28]. Therefore, Definition 3.1 provides a unifying setting for the analysis of different kinds of regularizations with relative errors, such as quadratic, log-quadratic, and double regularizations.

The second inexact criterion, which is given in Definition 3.2, is used for the Bregman distance induced by the function $h(x) = \sum_{i=1}^n x_i \ln(x_i)$. As far as we know, there is no convergence analysis for this kind of regularization when the operator is not paramonotone. With the setting of Definition 3.2, we proved convergence of the proximal method in section 5. We believe that our general setting is precisely what allows for the proof of convergence in this case, and we could say that this is a theoretical advantage of our scheme. Finally, we show in section 6 that the latter theoretical advantage can also become practical because the relative error and the enlargement of T can be of help in specific implementations. More precisely, Definition 3.2 allows for a specific useful algorithm based on the Bregman distance D_h. In our specific implementation, we avoid the extragradient step in the computations by using the enlargement of T. These advantageous features, which are new for this generalized distance, cannot be obtained by using a summable error framework nor by using exact values of T in the iterations.

Acknowledgments. The authors are very thankful to two anonymous referees and to the associate editor Heinz Bauschke for several helpful and constructive comments, which greatly contributed to the improvement of the manuscript. The authors are also thankful to Shoham Sabach for his careful reading of the manuscript.

REFERENCES

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

[2] A. Auslender and M. Teboulle, *Interior gradient and proximal methods for convex and conic optimization*, SIAM J. Optim., 16 (2006), pp. 697–725.

[3] A. Auslender and M. Teboulle, *Interior projection-like methods for monotone variational inequalities*, Math. Progr., 104 (2005), pp. 39–68.

[4] 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. 645–669.

[5] A. Auslender, M. Teboulle, and S. Ben-Tiba, *Coupling the logarithmic-quadratic proximal method and the block nonlinear Gauss–Seidel algorithm for linearly constrained convex minimization*, in Ill-posed Variational Problems and Regularization Techniques, M. Thera and R. Tichatschke, eds., Springer, Berlin, 1999, pp. 35–47.

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

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

[8] R. S. Burachik and A. N. Iusem, *Enlargement of monotone operators with applications to variational inequalities*, Set-Valued Anal., 5 (1997), pp. 159–180.

[9] R. S. Burachik and A. N. Iusem, *A generalized proximal point algorithm for the variational inequality problem in a Hilbert space*, SIAM J. Opt., 8 (1998), pp. 197–216.

[10] R. S. Burachik and A. N. Iusem, *Set-Valued Mappings and Enlargements of Monotone Operators*, Springer Optimization and Its Applications, Vol. 8, Springer, New York, 2007.

[11] R. S. Burachik, C. A. Sagastizábal, and B. F. Svaiter, *ϵ-enlargements of maximal monotone operators: theory and applications*, in Reformulation: Nonsmooth, Piecewise Smooth, Semismooth and Smoothing Methods, M. Fukushima and L. Qi, eds., Appl. Optim., Vol 22, Kluwer Academic Publishers, Dordrecht, 1999, pp. 25–43.

[12] R. S. Burachik and B. F. Svaiter, *ε-enlargements of maximal monotone operators in Banach spaces*, Set-Valued Anal., 7 (1999), pp. 117–132.

[13] R. S. Burachik and B. F. Svaiter, *A relative error tolerance for a family of generalized proximal point methods*, Math. Oper. Res., 26 (2001), pp. 816–831.

[14] 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. Program., 81 (1998), pp. 373–400.

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

[16] J. Eckstein, *Approximate iterations in Bregman-function-based proximal algorithms*, Math. Program., 83 (1998), pp. 113–123.

[17] J. Eckstein and P. J. S. Silva, *Double-regularization proximal methods, with complementarity applications*, Comput. Optim. Appl., 33 (2006), pp. 115–156.

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

[19] P. Eggermont, *Multiplicative iterative algorithms for convex programming*, Linear Algebra Appl., 130 (1990), pp. 25–42.

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

[21] A. N. Iusem and M. Teboulle, *Convergence rate analysis of nonquadratic proximal and augmented Lagrangian methods for convex and linear programming*, Math. Oper. Res., 20 (1995), pp. 657–677.

[22] K. C. Kiwiel, *Proximal minimization methods with generalized Bregman functions*, SIAM J. Control Optim., 35 (1997), pp. 1142–1168.

[23] L. Lewin, ed., *Structural Properties of Polilogarithms*, Math. Surveys Monogr., Vol 37, Amer. Math. Soc., Providence, RI, 1991.

[24] T. R. Rockafellar, *Monotone operators and the proximal point algorithm*, SIAM J. Control Optim., 14 (1976), pp. 877–898.

[25] T. R. Rockafellar, *Convex Analysis*, Princeton University Press, Princeton, NJ, 1970.

[26] 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.

[27] P. J. da Silva, J. Eckstein, and C. Humes, Jr., *Rescaling and stepsize selection in proximal methods using separable generalized distances*, SIAM J. Optim., 12 (2001), pp. 238–261.

[28] M. V. Solodov and B. F. Svaiter, *A unified framework for some inexact proximal point algorithms*, Numer. Funct. Anal. Optim., 22 (2001), pp. 1013–1035.
[29] M. V. Solodov and B. F. Svaiter, *Error bounds for proximal point subproblems and associated inexact proximal point algorithms*, Math. Program. Ser. B., 88 (2000), pp. 371–389.
[30] M. Teboulle, *Entropic proximal mappings with application 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. Bertsekas, *On the convergence of exponential multiplier method for convex programming*, Math. Program., 60 (1993), pp. 1–19.