Math. Program., Ser. A 86: 65–90 (1999)

Digital Object Identifier (DOI) 10.1007/s101079900076

Jonathan Eckstein · Michael C. Ferris

Smooth methods of multipliers for complementarity problems

Received February 21, 1997 / Revised version received December 11, 1998
Published online May 12, 1999

Abstract. This paper describes several methods for solving nonlinear complementarity problems. A general duality framework for pairs of monotone operators is developed and then applied to the monotone complementarity problem, obtaining primal, dual, and primal-dual formulations. We derive Bregman-function-based generalized proximal algorithms for each of these formulations, generating three classes of complementarity algorithms. The primal class is well-known. The dual class is new and constitutes a general collection of methods of multipliers, or augmented Lagrangian methods, for complementarity problems. In a special case, it corresponds to a class of variational inequality algorithms proposed by Gabay. By appropriate choice of Bregman function, the augmented Lagrangian subproblem in these methods can be made continuously differentiable. The primal-dual class of methods is entirely new and combines the best theoretical features of the primal and dual methods. Some preliminary computation shows that this class of algorithms is effective at solving many of the standard complementarity test problems.

Key words. complementarity problems – smoothing – proximal algorithms – augmented Lagrangians

1. Introduction

This paper concerns the solution of the *nonlinear complementarity problem* (NCP). Let $l \in [-\infty, \infty)^n$ and $u \in (-\infty, \infty]^n$, with $l \le u$. Suppose $\{x \in \Re^n \mid l \le x \le u\} \subseteq D \subseteq \Re^n$, and let $F : D \to \Re^n$ be continuous. Then, the NCP is to find some $x \in \Re^n$ satisfying the conditions

$$l \le x \le u \qquad \operatorname{mid}(l, x - F(x), u) = x, \tag{1}$$

where $\operatorname{mid}(a, b, c)$ denotes the componentwise median of the vectors a, b, and c. This problem is a special case of the standard *variational inequality* problem: given F and a set $C \subseteq \Re^n$, find some x such that

$$x \in C \qquad \langle F(x), y - x \rangle \ge 0 \quad \forall\, y \in C \quad . \tag{2}$$

If we take $C = \{x \in \Re^n \mid l \le x \le u\}$, then (2) is identical to (1).

The special case of $l = 0$ and $u = \infty$ reduces (1) to

$$x \ge 0 \qquad \max(x - F(x), 0) = x,$$

J. Eckstein: Faculty of Management and RUTCOR, Rutgers University, 640 Bartholomew Road, Piscataway, NJ 08854, e-mail: `jeckstei@rutcor.rutgers.edu`.
The research was partially supported by a Rutgers University Faculty of Management Research Resources Fellowship.

M.C. Ferris: University of Wisconsin, Computer Sciences Department, 1210 West Dayton Street, Madison, WI 53706, e-mail: `ferris@cs.wisc.edu`.
The research was partially supported by National Science Foundation Grant CCR-9619765.

or equivalently

$$x \geq 0 \qquad F(x) \geq 0 \qquad \langle x, F(x) \rangle = 0. \tag{3}$$

If the mapping F is affine, then (3) is the classical *linear complementarity problem*, or LCP.

In the theoretical portion of this paper, we will restrict our attention to the *monotone* case in which F satisfies

$$\langle F(x) - F(y), x - y \rangle \geq 0 \quad \forall\, x, y \in \Re^n. \tag{4}$$

This assumption will allow us to model (1) as the problem of finding a root of the sum of two monotone operators (see *e.g.* [3]), as will be explained in Section 2. To find such a root, we then apply generalized proximal algorithms based on Bregman functions [6, 7, 9, 12, 17, 18, 33].

A number of recent papers [5, 6, 8] have stressed the ability of proximal terms arising from appropriately-formulated Bregman functions to act like barrier functions, giving rise to "interior point" proximal methods for variational inequality problems. Such methods are derived by applying Bregman proximal methods to a primal formulation of (1) or (2).

In contrast, we emphasize dual and primal-dual formulations. Applying Bregman proximal methods to such formulations yields augmented-Lagrangian-like algorithms, or "methods of multipliers." In the dual case, we obtain a class of methods generalizing [21, "ALG1"]. By careful choice of Bregman function, we generate methods which involve solving (provided that F is differentiable) a once-differentiable system of equations at each iteration, as opposed to a nonsmooth system, as in [21]. Therefore, we can use a standard algorithm such as Newton's method to solve these subproblems. A similar phenomenon has already been pointed out for smooth convex programming problems in [24]. That paper notes that one of the augmented Lagrangian methods proposed in [17] yields a twice-differentiable augmented Lagrangian, as opposed to the classical once-differentiable augmented Lagrangian for inequality constraints (*e.g.* [30]).

In producing sequences of subproblems consisting of differentiable nonlinear equations, our algorithms bear some resemblance to recently proposed smoothing methods for the LCP and NCP [10, 11, 22]. However, such methods are akin to pure penalty methods in constrained optimization — they have a penalty parameter that must be driven to infinity to obtain convergence. By contrast, our algorithms are generalized versions of augmented Lagrangian methods: there is a Lagrange multiplier adjustment at the end of each iteration, and we obtain convergence even if the penalty parameter does not approach infinity.

In the course of our derivation, Section 2 develops a simple duality framework for pairs of set-valued operators. The framework resembles [1], but allows the two mappings in the pair to operate on different spaces. A similar duality structure for pairs of monotone operators appears in [20]. The main distinction of our approach, as opposed to [1, 20], is to introduce a primal-dual, "saddle-point" formulation, in addition to the standard primal and dual formulations. Towards the end of Section 2, we show how to apply the duality framework to variational inequalities and complementarity problems, refining the framework for variational inequalities that appears in [21, 27].

Section 3 combines the duality framework of Section 2 with Bregman function proximal algorithms and shows how to produce new, smooth methods of multipliers for (1). The primal-dual formulation yields a new *proximal* method of multipliers for (1), along the lines of the proximal method of multipliers for convex programming (*e.g.* [30]). This primal-dual method combines the best theoretical features of primal methods in the spirit of [5,6,8] with the best features of the new dual method. Some preliminary computational results on the MCPLIB [14] suite of test problems are given in Section 4. These results show that proximal method of multipliers is effective even when the underlying problem is not monotone.

2. A simple duality framework for pairs of monotone operators

In this paper, an *operator* T on a real Hilbert space X is a subset of $X \times Y$, where Y is also a Hilbert space. We call Y the *range space* of T; typically, but not always, we will have $X = Y$.

For every such $T \subseteq X \times Y$ and $x \in X$, $T(x) \doteq \{y \in Y \mid (x, y) \in T\}$ defines a point-to-set mapping from X to Y; in fact, we make no distinction between this point-to-set mapping and its graph T. Thus, the statements $y \in T(x)$ and $(x, y) \in T$ are completely equivalent. The *inverse* of any operator T is $T^{-1} = \{(y, x) \in Y \times X \mid (x, y) \in T\}$, which will always exist. Trivially, $(T^{-1})^{-1} = T$. We define

$$\operatorname{dom} T \doteq \{x \mid T(x) \neq \emptyset\} = \{x \in X \mid \exists\, y \in Y : (x, y) \in T\},$$

and similarly $\operatorname{im} T \doteq \operatorname{dom}(T^{-1}) = \{y \in Y \mid \exists\, x \in X : (x, y) \in T\}$. When $T(x)$ is a singleton set $\{y\}$ for all x, that is, T is the graph of some function $\operatorname{dom} T \to Y$, we say that T is *single-valued*, and we may write, in a slight abuse of notation, $T(x) = y$ instead of $T(x) = \{y\}$.

Given two operators T and U on X with the same range space Y, their sum $T + U$ is defined via $(T + U)(x) = T(x) + U(x) = \{t + u \mid t \in T(x), u \in U(x)\}$. If T is any operator on X and U an operator on Z, we define their *direct product* $T \otimes U$ on $X \times Z$ via $(T \otimes U)(x, z) = T(x) \times U(z)$.

An operator T on X is said to be *monotone* if its range space is X and

$$\left\langle x - x', y - y' \right\rangle \geq 0 \quad \forall\, (x, y), (x', y') \in T. \tag{5}$$

Note that (5) is a natural generalization of (4): if one takes $X = \Re^n$ and T to be the graph of the function F, (5) reduces to (4). Note also that monotonicity of T and T^{-1} are equivalent, and that it is straightforward to show that if two operators T and U are both monotone, then so is $T + U$.

A monotone operator T is *maximal* if no strict superset of T is monotone, that is,

$$(x, y) \in X \times X, \ \left\langle x - x', y - y' \right\rangle \geq 0 \ \forall\, (x', y') \in T \quad \Rightarrow \quad (x, y) \in T.$$

Maximality of an operator and maximality of its inverse are equivalent.

The fundamental problem customarily associated with a monotone operator T is that of finding a *zero* or *root*, that is, some $x \in X$ such that $0 \in T(x)$ (see *e.g.* [3,31]).

2.1. The duality framework

Suppose we are given an operator A on a Hilbert space X, an operator B on a Hilbert space Y, and a linear mapping $M : X \to Y$. We will denote such a triple by $\mathcal{P}(A, B, M)$. For the development in Section 3, we will require only the special case $X = Y = \Re^n$ and $M = I$, but we consider the general $\mathcal{P}(A, B, M)$ in order to make connections to [16, 20] and other previous work.

We associate with $\mathcal{P}(A, B, M)$ a *primal formulation* of finding $x \in X$ such that

$$0 \in A(x) + M^\top B(Mx), \tag{6}$$

or equivalently $0 \in T_{\mathrm{P}}(x) \doteq \left[A + M^\top BM\right](x)$, where $M^\top$ denotes the adjoint of M.

Similarly, we associate with each $\mathcal{P}(A, B, M)$ a *dual formulation* of finding $y \in Y$ such that

$$0 \in -MA^{-1}(-M^\top y) + B^{-1}(y), \tag{7}$$

or equivalently $0 \in T_{\mathrm{D}}(y) \doteq \left[-MA^{-1}(-M^\top) + B^{-1}\right](y)$. Note that (7) is the primal formulation of $\mathcal{P}(B^{-1}, A^{-1}, -M^\top)$, and that twice applying the transformation

$$\mathcal{P}(A, B, M) \mapsto \mathcal{P}(B^{-1}, A^{-1}, -M^\top)$$

produces the original triple $\mathcal{P}(A, B, M)$; that is, the dual of $\mathcal{P}(B^{-1}, A^{-1}, -M^\top)$ is the original primal formulation (6). The duality scheme of [1] is similar, with the restrictions $X = Y$ and $M = I$.

We also associate with $\mathcal{P}(A, B, M)$ a *primal-dual formulation*, which is to find $(x, y) \in X \times Y$ such that

$$0 \in A(x) + M^\top y \qquad\qquad 0 \in -Mx + B^{-1}(y), \tag{8}$$

or equivalently $0 \in T_{\mathrm{PD}}(x, y) \doteq K[A, B, M](x, y)$, where $K[A, B, M]$ is defined by

$$K[A, B, M]\begin{pmatrix} x \\ y \end{pmatrix} = \left(A(x) \times B^{-1}(y)\right) + \begin{bmatrix} 0 & M^\top \\ -M & 0 \end{bmatrix}\begin{pmatrix} x \\ y \end{pmatrix}. \tag{9}$$

In the special case of convex optimization, we can take $A = \partial f$, the subdifferential map of some closed proper convex function $f : X \to (-\infty, +\infty]$, and $B = \partial g$ for some closed proper convex $g : Y \to (-\infty, +\infty]$. Then the primal formulation is equivalent to the optimization problem

$$\min_{x \in X} \; f(x) + g(Mx). \tag{10}$$

Similarly, the dual formulation is equivalent to

$$\min_{y \in Y} \; f^*(-M^\top y) + g^*(y), \tag{11}$$

where "*" denotes the convex conjugacy operation [28, Section 12]. Furthermore, the subdifferential of the generalized Lagrangian $L : X \times Y \to [-\infty, +\infty]$ defined by

$$L(x, y) = f(x) + \langle y, Mx\rangle - g^*(y)$$

is precisely $K[A, B, M] = K[\partial f, \partial g, M]$. Therefore, the primal-dual formulation is equivalent to finding a saddle point of L, that is, to the problem

$$\min_{x\in X} \max_{y\in Y} f(x) + y^\top Mx - g^*(y). \tag{12}$$

The standard convex programming duality relations between (10), (11), and (12) may be viewed as a consequence of the higher-level, more abstract duality embodied in the following elementary proposition, whose proof is omitted.

Proposition 1. *The following statements are equivalent:*

(i) (x, y) *solves the primal-dual formulation (8).*
(ii) $x \in X, \quad y \in Y, \quad (x, -M^\top y) \in A, \quad (Mx, y) \in B.$

Furthermore, x solves the primal formulation (6) if and only if there exists $y \in Y$ such that (i)-(ii) hold, and y solves the dual formulation (7) if and only if there exists $x \in X$ such that (i)-(ii) hold.

Note that for general choices of A, B, and M, this duality framework is slightly weaker than, for example, linear programming, in that x being a primal solution and y being a dual solution are *not* sufficient for (x, y) to be an solution of the primal-dual ("saddle point") formulation, even if A and B are maximal monotone. For an example of this phenomenon, consider the case $X = Y = \Re^2$, $M = I$, $A(x_1, x_2) = \{(-x_2, x_1)\}$, and $B(x_1, x_2) = \{(x_2, -x_1)\}$.

We now turn to the issue of solving (6), (7) or (8), under the assumption that A and B are maximal monotone.

Consider first the primal formulation (6). Given that B is monotone, it is straightforward to show that $M^\top BM$ is also monotone. The monotonicity of A then gives the monotonicity of $T_P = A + M^\top BM$. Therefore, the primal formulation is a problem of locating a root of the monotone operator T_P on X. The convergence analyses of root-finding methods for monotone operators typically require that the operator be not only monotone, but also maximal. While T_P will typically be maximal if A and B are, such maximality cannot be guaranteed without imposing additional regularity conditions. Some typical sufficient conditions for T_P to be maximal are that A and B be maximal, that $MM^\top$ be an isomorphism of Y, thus guaranteeing maximality of $M^\top BM$ (see [21, Proposition 4.1] or [20, Proposition 3.2]), and a condition such as $\operatorname{dom} A \cap \operatorname{int} \operatorname{dom}(BM) \neq \emptyset$, in order to ensure maximality of the sum $T_P = A + M^\top BM$ [29]. This last condition can be weakened somewhat if X is finite-dimensional.

The analysis of the dual formulation is similar. The formulation involves locating the root of the operator $T_D = -MA^{-1}(-M^\top) + B^{-1}$ on Y, which is necessarily monotone by the monotonicity of A and B, but is not guaranteed to be maximal solely by maximality of A and B. One must impose similar conditions to the primal case, such as $M^\top M$ being an isomorphism of X, and $\operatorname{dom}(A^{-1}(-M^\top)) \cap \operatorname{int} \operatorname{im} B \neq \emptyset$.

The primal-dual formulation also involves finding the root of a monotone operator: we establish in Proposition 2 below that the operator $T_{PD} = K[A, B, M]$ on $X \times Y$ (with the canonical inner product induced by X and Y) is monotone if A and B are. The proposition also shows that the primal-dual is in some sense the "best behaved" of our three formulations, in the sense that $K[A, B, M]$ is maximal whenever A and B are both maximal.

Proposition 2. *If A and B are monotone operators on the Hilbert spaces X and Y, respectively, and M is any linear map $X \to Y$, then the operator $K[A, B, M]$ on $X \times Y$ defined by (9) is monotone. Furthermore, if A and B are both maximal, $K[A, B, M]$ is maximal.*

Proof. Set

$$T_1 = A \otimes B^{-1} \qquad\qquad T_2(x, y) = \begin{bmatrix} 0 & M^\top \\ -M & 0 \end{bmatrix} \begin{pmatrix} x \\ y \end{pmatrix}.$$

Note that T_1 and T_2 are both monotone, and $K[A, B, M] = T_1 + T_2$. If A and B are maximal, T_1 is also maximal. The linear map T_2 is also maximal [26], and maximality of $T_1 + T_2$ then follows from [29, Theorem 1(a)].

We remark that it is also straightforward (but more lengthy) to prove Proposition 2 from first principles, without invoking the deep analytical machinery of [26,29].

In summary, given a linear M and monotone A and B, we can formulate the same problem in three essentially equivalent ways: finding a root of the primal monotone operator $T_P = A + M^\top BM$ on X, finding a root of the dual monotone operator $T_D = -MA^{-1}(-M^\top) + B^{-1}$ on Y, or finding a root of the primal-dual monotone operator $T_{PD} = K[A, B, M]$ on $X \times Y$. Of these operators, T_{PD} is the only one *guaranteed* to be maximal, given the maximality of A and B.

2.2. Dual and primal-dual formulations of variational inequality and complementarity problems

We now return to the variational inequality problem (2), where $F : D \to \Re^n$ satisfies the monotonicity condition (4), $D \supseteq C$, and C is a closed convex set. Define the operator $N_C \subseteq C \times \Re^n \subseteq \Re^n \times \Re^n$ via

$$N_C(x) = \begin{cases} \left\{ d \in \Re^n \;\middle|\; \langle d, y - x \rangle \le 0 \;\; \forall\, y \in C \right\}, & x \in C \\ \emptyset, & x \notin C. \end{cases} \tag{13}$$

It is well-known that N_C is maximal monotone on $\Re^n$. Furthermore, the variational inequality (2) is equivalent to the problem

$$0 \in F(x) + N_C(x). \tag{14}$$

We take (14) as our primal formulation in the duality framework of (6), (7), and (8). Consequently, we let $A = F$, $B = N_C$, $X = Y = \Re^n$, and $M = I$, whence $T_P = F + N_C$. We then have $T_D = -F^{-1}(-I) + {N_C}^{-1}$, and the problem dual to (14) is thus

$$0 \in -F^{-1}(-y) + {N_C}^{-1}(y), \tag{15}$$

where “$^{-1}$” denotes the operator-theoretic inverse. F^{-1} and ${N_C}^{-1}$ may both be general set-valued operators on $\Re^n$, in the sense of Section 2. Although the notation is different, this dual problem is essentially the same dual proposed in [21,27]. The formulation (15) may appear somewhat awkward, but we will not have to work with it directly in a computational setting. It will, however, prove very useful in deriving algorithms.

It is a simple consequence of Proposition 1 that y solves (15) if and only if $y = -F(x)$ for some solution x of (14), or equivalently of the variational inequality (2).

The primal-dual formulation, in this setting, is to find a zero of the operator $T_{\mathrm{PD}} = K[F, N_C, I]$ defined via

$$T_{\mathrm{PD}}(x, y) = \Big(F(x) \times {N_C}^{-1}(y)\Big) + \begin{pmatrix} y \\ -x \end{pmatrix}.$$

Equivalently, x and y solve the system

$$F(x) = -y \qquad {N_C}^{-1}(y) \ni x, \tag{16}$$

that is, x solves the variational inequality (2), and $y = -F(x)$.

We now investigate the structure of N_C and ${N_C}^{-1}$ in the case of the NCP (1), where $C = \{x \in \Re^n \mid l \le x \le u\}$. In this case, N_C is the direct product of n simple operators on $\Re$ of the form

$$N_i = \Big[\Big(\{l_i\} \times (-\infty, 0)\Big) \cup \Big([l_i, u_i] \times \{0\}\Big) \cup \Big(\{u_i\} \times (0, +\infty)\Big)\Big] \cap \Re^2,$$

as depicted on the left side of Figure 1. It then follows that ${N_C}^{-1}$ is the direct product of the n operators

$${N_i}^{-1} = \Big[\Big((-\infty, 0) \times \{l_i\}\Big) \cup \Big(\{0\} \times [l_i, u_i]\Big) \cup \Big((0, +\infty) \times \{u_i\}\Big)\Big] \cap \Re^2, \tag{17}$$

as depicted on the right side of Figure 1.

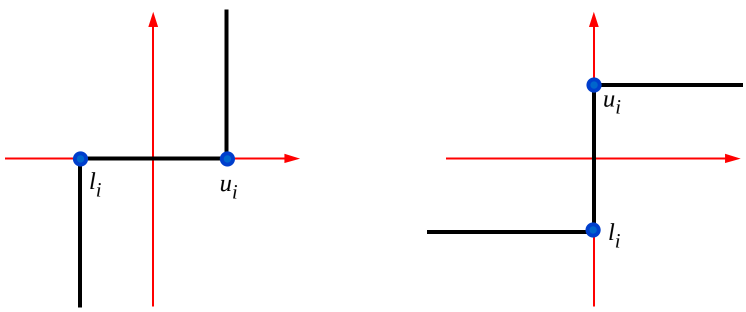

Fig. 1. The operator N_i on $\Re$ (left), and its inverse ${N_i}^{-1}$ (right)

Since maximality is needed to prove convergence of the solution methods we propose in Section 3, we now address the question of maximality of F, $T_{\mathrm{P}} = F + N_C$, $T_{\mathrm{D}} = -F^{-1}(-I) + {N_C}^{-1}$, and $T_{\mathrm{PD}} = K[F, N_C, I]$.

Proposition 3. *Let F be a continuous monotone function on $\Re^n$ with open domain $D \supset C = \{x \in \Re^n \mid l \le x \le u\}$. Then $T_{\mathrm{P}} = F + N_C$ is maximal monotone.*

Proof. Let $\widehat{F}$ be some maximal extension of F into a monotone operator [32, Proposition 12.6]. Then we have $\operatorname{dom} \widehat{F} \supseteq D \supset C = \operatorname{dom} N_C \neq \emptyset$, and therefore $\operatorname{ri} \operatorname{dom} \widehat{F} \cap \operatorname{ri} \operatorname{dom} N_C \neq \emptyset$, where "ri" denotes relative interior [28, Section 6]. From [29] we have that $\widehat{F} + N_C$ must be maximal. Now, the openness of D and the analysis of [26, Theorem 4] imply that $\widehat{F}$ agrees in value with F on $D \supset C = \operatorname{dom} N_C = \operatorname{dom} T_{\mathrm{P}}$, so it follows that $\widehat{F} + N_C = T_{\mathrm{P}}$.

Proposition 4. *Suppose F is a continuous monotone function on $\Re^n$ that is maximal as a monotone operator (some sufficient conditions are* $\operatorname{im}(I + F) = \Re^n$ *or that F has maximal open domain). Suppose* $\operatorname{ri} \operatorname{im} F$ *contains some point $y \in \Re^n$ with the property that*

$$\begin{array}{ll} y_i = 0 & \forall\, i : l_i = -\infty,\ u_i = +\infty \\ y_i < 0 & \forall\, i : l_i = -\infty,\ u_i < +\infty \\ y_i > 0 & \forall\, i : l_i > -\infty,\ u_i = +\infty \quad . \end{array} \tag{18}$$

Then $T_{\mathrm{D}} = -F^{-1}(-I) + N_C{}^{-1}$ is maximal, where $C = \{x \in \Re^n \mid l \leq x \leq u\}$.

Proof. Given that F constitutes a maximal monotone operator, it is straightforward to show that $-F^{-1}(-I)$ is also maximal. Now, $\operatorname{dom}(-F^{-1}(-I)) = -\operatorname{im} F$. By appealing to (17), it is clear that the conditions (18) on y are equivalent to $-y \in \operatorname{ri} \operatorname{dom}(N_C{}^{-1})$. Therefore, we have $\operatorname{ri} \operatorname{dom}(-F^{-1}(-I)) \cap \operatorname{ri} \operatorname{dom}(N_C{}^{-1}) \neq \emptyset$. The maximality of N_C and [29] then imply the maximality of $T_{\mathrm{D}} = -F^{-1}(-I) + N_C{}^{-1}$.

Note that if $l > -\infty$ and $u < +\infty$, the conditions (18) are void, and Proposition 4 requires only maximality of F. Finally, we address the maximality of T_{PD} with the following proposition, which follows immediately from Proposition 2 and the maximality of F and N_C.

Proposition 5. *Suppose F is a monotone function on $\Re^n$ that is maximal as a monotone operator. Then, for any closed convex set $C \supseteq \Re^n$, the operator $T_{\mathrm{PD}} = K[F, N_C, I]$ is maximal.*

3. Bregman proximal algorithms for complementarity problems

For the remainder of this paper, we let $C = \{x \in \Re^n \mid l \leq x \leq u\}$. We now have three formulations of the monotone complementarity problem (1): finding a root of the primal monotone operator $T_{\mathrm{P}} = F + N_C$, finding a root of the dual monotone operator $T_{\mathrm{D}} = -F^{-1}(-I) + N_C{}^{-1}$, and finding a root of the primal-dual monotone operator $T_{\mathrm{PD}} = K[F, N_C, I]$. We can attempt to solve (1) by applying any method for finding the root of a monotone operator to either T_{P}, T_{D}, or T_{PD}. In this paper, we employ only the Bregman-function-based proximal algorithm of [18], and study the algorithms for (1) that result when it is applied to T_{P}, T_{D}, and T_{PD}.

We now describe the algorithm of [18] for solving the inclusion $0 \in T(x)$, where T is a maximal monotone operator on $\Re^n$. Earlier treatments of closely related algorithms may be found in [6,7,9,12,17,23,33]

The algorithm in [18] requires two auxiliary constructs, a function h and a set S. Given two points $x, y \in \Re^n$ and a function h differentiable at y, we define

$$D_h(x, y) \doteq h(x) - h(y) - \langle \nabla h(y), x - y \rangle. \tag{19}$$

We then say that h is a *Bregman function with zone S* if the following conditions hold:

B1. $S \subseteq \Re^n$ is a convex open set.
B2. $h : \Re^n \to \Re \cup \{+\infty\}$ is finite and continuous on $\overline{S}$.
B3. h is strictly convex on $\overline{S}$.
B4. h is continuously differentiable on S.
B5. Given any $x \in \overline{S}$ and scalar α, the *right partial level set*

$$L(x, \alpha) \doteq \{y \mid D_h(x, y) \leq \alpha\}$$

is bounded.
B6. If $\{y^k\} \subset S$ is a convergent sequence with limit y^∞, then $D_h(y^\infty, y^k) \to 0$.
B7. If $\{v^k\} \subset \overline{S}$, $\{w^k\} \subset S$ are sequences such that $w^k \to w^\infty$ and $\{v^k\}$ is bounded, and furthermore $D_h(v^k, w^k) \to 0$, then one has $v^k \to w^\infty$.

Examples of pairs (h, S) meeting these conditions may be found in [9,13,17,33], and many references therein. In particular, [13] gives some general sufficient conditions for (h, S) to satisfy B1-B7. We now state the main result of [18].

Proposition 6. *Let T be a maximal monotone operator on $\Re^n$, and let h be a Bregman function with zone S, where $S \cap \operatorname{ri} \operatorname{dom} T \neq \emptyset$. Let any* one *of the following assumptions A1-A3 hold:*

A1. $S \supseteq \overline{\operatorname{dom} T}$.
A2. $T = \partial f$, *the subdifferential mapping of some closed proper convex function* $f : \Re^n \to \Re \cup \{+\infty\}$.
A3. T *has the following two properties (see,* e.g. *[6–8]):*
(i) If $\{(x^k, y^k)\} \subset T$, $\{x^k\} \subset S$, and $\{x^k\}$ is convergent, then $\{y^k\}$ has a limit point;
(ii) T is paramonotone *[4,8], that is, $(x, y), (x', y') \in T$ and*

$$\langle x - x', y - y' \rangle = 0$$

collectively imply that $(x, y') \in T$.

Suppose the sequences $\{z^k\}_{k=0}^\infty \subset S$ and $\{e^k\}_{k=0}^\infty \subset \Re^n$ conform to the recursion

$$T(z^{k+1}) + \frac{1}{c_k}\Big(\nabla h(z^{k+1}) - \nabla h(z^k)\Big) \ni e^k, \tag{20}$$

where $\{c^k\}_{k=0}^\infty$ is a sequence of positive scalars bounded away from zero. Further suppose that

$$\sum_{k=0}^{\infty} c_k \|e^k\| < \infty \tag{21}$$

and

$$\sum_{k=1}^{\infty} c_k \langle e^k, z^k \rangle \text{ exists and is finite.} \tag{22}$$

Then if $\widehat{T} \doteq T + N_{\overline{S}}$ *has any roots,* $\{z^k\}$ *converges to some* z^∞ *with* $\widehat{T}(z^\infty) \ni 0$.

Proof. By minor reformulation of [18, Theorem 1].

Similar forms for the error sequence can be found for example in [25]. Note that the condition (22) is implied by the more easily-verified condition

$$\sum_{k=0}^{\infty} c_k \|e^k\| \, \|z^k\| < \infty. \tag{23}$$

Furthermore, when S or $\operatorname{dom} T$ is bounded, $\{z^k\}$ is necessarily bounded, and (21) implies (23) and (22).

One question not addressed in Proposition 6 is whether sequences $\{z^k\}_{k=0}^{\infty} \subset S$ and $\{e^k\}_{k=1}^{\infty} \subset \Re^n$ conforming to (20) are guaranteed to exist. The following proposition gives sufficient conditions for the purposes of this paper.

Proposition 7. *Let* T *be a maximal monotone operator on* $\Re^n$*, let* $\{c_k\}_{k=0}^{\infty}$ *be a sequence of positive scalars, and let* h *be a Bregman function with zone* $S \supseteq \operatorname{dom} T$*. Then if* $\operatorname{im} \nabla h = \Re^n$*, sequences* $\{z^k\}_{k=0}^{\infty} \subset S$ *and* $\{e^k\}_{k=0}^{\infty} \subset \Re^n$ *jointly conforming to (20) exist.*

Proof. Set $e^k = 0$ for all k, and consult case (i) of [17, Theorem 4].

We now consider applying Proposition 6 with either $T = T_{\text{P}}$, $T = T_{\text{D}}$, or $T = T_{\text{PD}}$. Each choice will yield a different algorithm for solving the complementarity problem (1).

3.1. Primal application to complementarity

The most straightforward application of Proposition 6 to the complementarity problem (1) is to set $T = T_{\text{P}} = F + N_C$. Substituting $T = F + N_C$ and $z^k = x^k$ into the fundamental recursion (20) and rearranging, we obtain the recursion:

$$\left[F\big(x^{k+1}\big) + \frac{1}{c_k}\Big(\nabla h\big(x^{k+1}\big) - \nabla h\big(x^k\big)\Big)\right] + N_C\big(x^{k+1}\big) \ni e^k. \tag{24}$$

In other words, x^{k+1} is an $\|e^k\|$-accurate approximate solution of the complementarity problem

$$l \le x \le u \qquad \operatorname{mid}\Big(l, x - \tilde{F}_k(x), u\Big) = x,$$

where $\tilde{F}_k(x) = F(x) + c_k{}^{-1}(\nabla h(x) - \nabla h(x^k))$. For general choices of h, there appears to be little point to such a procedure: to solve a single nonlinear complementarity problem, we must now (approximately) solve an infinite sequence of similar nonlinear complementarity problems. However, the situation is more promising in the special case that $l < u$, the zone S of h is $\operatorname{int} C$, and $\|\nabla h(x)\| \to \infty$ as x approaches any $\overline{x} \in \operatorname{bd} C$. In this case, we must have $x^{k+1} \in \operatorname{int} C$ for all $k \geq 0$. Since $N_C(x) = \{0\}$ for all $x \in \operatorname{int} C$, we can drop the $N_C(x^{k+1})$ term from the recursion (24), reducing it to the equation

$$F\big(x^{k+1}\big) + \frac{1}{c_k}\Big(\nabla h\big(x^{k+1}\big) - \nabla h\big(x^k\big)\Big) = e^k. \tag{25}$$

So, each iteration must solve $F(x) + c_k^{-1}\nabla h(x) = c_k^{-1}\nabla h(x^k)$ for x within accuracy $\|e^k\|$. If F is differentiable, then $F + c_k{}^{-1}\nabla h$ is differentiable on $\operatorname{int} C$. Thus, we can solve a nonlinear complementarity problem by approximately solving a sequence of differentiable nonlinear systems of equations. Since ∇h approaches infinity on the boundary of C, it acts as a barrier function that simplifies the subproblems by removing boundary effects. This phenomenon has already been noted in numerous prior works, including [5,8].

However, setting $S = \operatorname{int} C$ also has drawbacks. First, in attempting to apply Proposition 6, $S = \operatorname{int} C$ rules out invoking Assumption A1, forcing one to appeal to Assumptions A2 or A3, each of which places restrictions on the maximal monotone operator T. In applying Proposition 6 to the primal formulation, these restrictions on T imply restrictions on the monotone function F. The following result summarizes what we can say about the convergence of method (25) for complementarity problems:

Theorem 1. *Suppose the complementarity problem (1) has some solution, and also that $l < u$, F is monotone and continuous on some open set $D \supset C = \{x \in \Re^n \mid l \leq x \leq u\}$, and F satisfies at least one of the following restrictions:*

P1. $F(x) = \nabla f(x)$ for all $x \in C$, where f is convex and continuously differentiable on C.
P2. For all $x, x' \in C$, $\big\langle x - x', F(x) - F(x')\big\rangle = 0$ implies $F(x) = F(x')$.

Let h be a Bregman function with zone $S = \operatorname{int} C$, with $\lim_{w \to \overline{w}} \|\nabla h(w)\| = \infty$ for any $\overline{w} \in \operatorname{bd} S = \operatorname{bd} C$. Suppose the sequences $\{x^k\}_{k=0}^\infty \subset S$, $\{e^k\}_{k=0}^\infty \subset \Re^n$, and $\{c_k\}_{k=0}^\infty \subset [\underline{c}, \infty) \subset (0, \infty)$ satisfy the recursion (25) and that $\sum_{k=0}^\infty c_k\|e^k\| < \infty$, while $\sum_{k=0}^\infty c_k\langle e^k, x^k\rangle$ exists and is finite. Then $\{x^k\}$ converges to a solution of the NCP (1).

Proof. (25) is equivalent to the fundamental recursion (20) of Proposition 6 with $T = T_{\mathrm{P}} = F + N_C$ and $z^k = x^k$. The conditions on $\{e^k\}$ are identical to the error conditions (21) and (22) of Proposition 6. The condition that F be continuous on D ensures that T_{P} will be maximal, via Proposition 3. Therefore, we may invoke Proposition 6 if we can show at least one of its alternative Assumptions A1-A3 hold.

Now consider Assumption P1. In this case, we have $T_{\mathrm{P}} = \nabla f + N_C = \nabla f + \partial\delta(\cdot\,|C) = \partial(f + \delta(\cdot\,|C))$, where the last equality follows from [28, Theorem 23.8] and $\operatorname{dom} f \supseteq C = \operatorname{dom} \delta(\cdot\,|C) \neq \emptyset$. Therefore, Assumption A2 of Proposition 6 is satisfied.

Alternatively, assume that P2 holds. Since F is continuous on $D \supset C = \overline{S}$ and $N_C(x) = \{0\}$ for all $x \in S = \operatorname{int} C$, Assumption A3(i) holds for $T = F + N_C$. P2 implies that A3(ii) holds for $T = F$. It is also easily confirmed that A3(ii) holds for

$T = N_C$. Finally, it is straightforward to show that paramonotonicity is preserved under the addition of operators, so A3(ii) also holds for $T = F + N_C$.

We may therefore invoke Proposition 6 and conclude that $\{x^k\}$ must converge to a root of $T_P + N_{\overline{S}} = T_P + N_C = T_P$, that is, a solution of (1).

This result represents a minor advance in the theory of primal complementarity methods, in that most prior results have required exact computation of each iteration, that is, $e^k \equiv 0$, the exception being [7]. The approximation condition (25) is much more practical to check than the corresponding condition in [7].

We cannot apply Proposition 7 to show existence of $\{x^k\}$ in this setting, because $S \not\supseteq \operatorname{dom} T$. However, suitable existence results may be found in [5–8].

Note that in the case $l > -\infty$ and $u < +\infty$, the condition on $\sum_{k=0}^{\infty} c_k \langle e^k, x^k \rangle$ is an immediate consequence of $\sum_{k=0}^{\infty} c_k \|e^k\| < \infty$, and becomes redundant. It only comes into play when there is a possibility of $\{x^k\}$ being unbounded.

While the restriction that F be continuous on $D \supset C$ seems reasonable, the alternative Hypotheses P1 and P2 impose extra restrictions on F. Furthermore, while it is not necessary to drive c_k to infinity to obtain convergence, as in a true barrier method, the procedure does inherit some numerical difficulties typical of barrier algorithms. The nonlinear system to be approximately solved in (25) becomes progressively more ill-conditioned as x approaches $\operatorname{bd} C$, where the solution is likely to lie. This ill-conditioning constrains the numerical methods that may be used. Furthermore, the function on the left-hand side of (25) is not defined for x outside $\operatorname{int} C$; to apply a standard numerical procedure such as Newton's method, one needs to install appropriate safeguards to avoid stepping to or evaluating points outside $\operatorname{int} C$.

3.2. Dual application to complementarity

In situations where the above drawbacks of the primal method are significant, we suggest dual or primal-dual algorithms, as described below. In these approaches, the Bregman function acts through the duality framework to provide a smooth, augmented-Lagrangian-like penalty function, rather than the barrier function one obtains from a primal approach. We first consider a purely dual approach, applying Proposition 6 to $T = T_D$.

The fundamental Bregman proximal recursion (20) for $T = T_D$ and iterates $z^k = y^k$ takes the form

$$-F^{-1}\big(y^{k+1}\big) + {N_C}^{-1}\big(y^{k+1}\big) + \frac{1}{c_k}\Big(\nabla h\big(y^{k+1}\big) - \nabla h\big(y^k\big)\Big) \ni e^k. \qquad (26)$$

Since the domain of T_D will in general be unknown, we will choose the Bregman-function/zone pair (h, S) so that $S = \Re^n$. This choice ensures that $\widehat{T} = T + N_{\overline{S}} = T_D + N_{\Re^n} = T_D$, and thus that the recursion will locate roots of T_D.

In general, it will not be possible to express the inverse operator F^{-1} in a manner convenient for computation, so we cannot work directly with the formula (26). Instead, we "dualize" the recursion using Proposition 1. For simplicity, temporarily assume that

$e^k \equiv 0$, so that (26) becomes

$$-F^{-1}\big(y^{k+1}\big) + {N_C}^{-1}\big(y^{k+1}\big) + \frac{1}{c_k}\left(\nabla h\big(y^{k+1}\big) - \nabla h\big(y^k\big)\right) \ni 0 \tag{27}$$

We now take (27) to be the primal problem in the framework of Section 2.1, setting $X = Y = \Re^n$ and $M = I$. We take $A = A_k$ and $B = B_k$, where A_k and B_k are defined by

$$A_k(y) = -F^{-1}(-y) \tag{28}$$
$$B_k(y) = {N_C}^{-1}(y) + \frac{1}{c_k}\left(\nabla h(y) - \nabla h\big(y^k\big)\right) \quad . \tag{29}$$

Note that if F constitutes a maximal monotone operator, $A = A_k$ will be maximal, and ${N_C}^{-1}$ is maximal by the maximality of N_C. ∇h is maximal monotone since it is the subgradient map of the function h, continuous on $\Re^n$. The operations of subtracting the constant $\nabla h(y^k)$ and scaling by $1/c_k$ preserve this maximality. Finally, since $\operatorname{dom} \nabla h = \Re^n$, we also have maximality of $B = B_k$ from [29].

Invoking Proposition 1, the problem dual to (27), or equivalently $A_k(y) + B_k(y) \ni 0$, is of the form $-{A_k}^{-1}(-x) + {B_k}^{-1}(x) \ni 0$, where we are interchanging the notational roles of "x" and "y". It is immediate that $-{A_k}^{-1}(-x) = -[-F^{-1}(-I)]^{-1}(-x) = -(-F(-(-x))) = F(x)$, so $-{A_k}^{-1}(-I) = F$.

We now consider ${B_k}^{-1}$. We know that ${N_C}^{-1}$ has the separable structure ${N_C}^{-1} = {N_1}^{-1} \otimes \ldots \otimes {N_n}^{-1}$, where ${N_i}^{-1}$ is given by (17). Further assume that h has the separable structure $h(y) = \sum_{i=1}^n h_i(y_i)$, whence (as an operator) $\nabla h = \nabla h_1 \otimes \ldots \otimes \nabla h_n$. Assume temporarily that $l > -\infty$ and $u < +\infty$. Then $B_k = B_{k1} \otimes \ldots \otimes B_{kn}$, where each B_{ki} is an operator on $\Re$ given by

$$B_{ki}(\gamma) = \begin{cases} \left\{ l_i + \frac{1}{c_k}\left(\nabla h_i(\gamma) - \nabla h_i\big(y_i^k\big)\right) \right\} & \gamma < 0 \\ \left[l_i + \frac{1}{c_k}\left(\nabla h_i(0) - \nabla h_i\big(y_i^k\big)\right), u_i + \frac{1}{c_k}\left(\nabla h_i(0) - \nabla h_i\big(y_i^k\big)\right) \right] & \gamma = 0 \\ \left\{ u_i + \frac{1}{c_k}\left(\nabla h_i(\gamma) - \nabla h_i\big(y_i^k\big)\right) \right\} & \gamma > 0. \end{cases}$$

Since ${B_k}^{-1} = {B_{k1}}^{-1} \otimes \ldots \otimes {B_{kn}}^{-1}$, it suffices to invert B_{ki}, $k = 1, \ldots, n$. For each B_{ki}, we have $B_{ki} = B_{ki}^- \cup B_{ki}^0 \cup B_{ki}^+$, where

$$B_{ki}^- = \left\{ \left(\gamma, l_i + \frac{1}{c_k}\left(\nabla h_i(\gamma) - \nabla h_i\big(y_i^k\big)\right)\right) \;\middle|\; \gamma < 0 \right\}$$
$$B_{ki}^0 = \{0\} \times \left[l_i + \frac{1}{c_k}\left(\nabla h_i(0) - \nabla h_i\big(y_i^k\big)\right), u_i + \frac{1}{c_k}\left(\nabla h_i(0) - \nabla h_i\big(y_i^k\big)\right) \right]$$
$$B_{ki}^+ = \left\{ \left(\gamma, u_i + \frac{1}{c_k}\left(\nabla h_i(\gamma) - \nabla h_i\big(y_i^k\big)\right)\right) \;\middle|\; \gamma > 0 \right\}.$$

It follows directly from the definition of the operator-theoretic inverse that ${B_{ki}}^{-1} = (B_{ki}^-)^{-1} \cup (B_{ki}^0)^{-1} \cup (B_{ki}^+)^{-1}$. Now,

$$
\begin{aligned}
\left(B_{ki}^-\right)^{-1} &= \left\{ \left(l_i + \frac{1}{c_k}\left(\nabla h_i(\gamma) - \nabla h_i\left(y_i^k\right)\right), \gamma \right) \;\middle|\; \gamma < 0 \right\} \\
&= \left\{ \left(\xi, (\nabla h_i)^{-1}\left(\nabla h_i\left(y_i^k\right) + c_k(\xi - l_i)\right)\right) \;\middle|\; (\nabla h_i)^{-1}\left(\nabla h_i\left(y_i^k\right) + c_k(\xi - l_i)\right) < 0 \right\} \\
&= \left\{ \left(\xi, (\nabla h_i)^{-1}\left(\nabla h_i\left(y_i^k\right) + c_k(\xi - l_i)\right)\right) \;\middle|\; \xi < l_i + \frac{1}{c_k}\left(\nabla h_i(0) - \nabla h_i\left(y_i^k\right)\right) \right\},
\end{aligned}
$$

where the first equality is obtained by solving for γ in terms of ξ in

$$\xi = l_i + \frac{1}{c_k}\left(\nabla h_i(\gamma) - \nabla h_i\left(y_i^k\right)\right),$$

and the second by solving $(\nabla h_i)^{-1}(\nabla h_i\left(y_i^k\right) + c_k(\xi - l_i)) < 0$ for ξ.

Similarly, we obtain

$$
\begin{aligned}
&\left(B_{ki}^+\right)^{-1} = \\
&\left\{ \left(\xi, (\nabla h_i)^{-1}\left(\nabla h_i\left(y_i^k\right) + c_k(\xi - u_i)\right)\right) \;\middle|\; \xi > u_i + \frac{1}{c_k}\left(\nabla h_i(0) - \nabla h_i\left(y_i^k\right)\right) \right\}.
\end{aligned}
$$

$(B_{ki}^0)^{-1}$ is simply the function that yields 0 on the interval

$$\Phi_{ki} \doteq \left[l_i + \frac{1}{c_k}\left(\nabla h_i(0) - \nabla h_i\left(y_i^k\right)\right), u_i + \frac{1}{c_k}\left(\nabla h_i(0) - \nabla h_i\left(y_i^k\right)\right) \right]. \tag{30}$$

Combining these three results and using the monotonicity of ∇h and $(\nabla h)^{-1}$, we obtain

$$
\begin{aligned}
{B_{ki}}^{-1}(\xi) &= \begin{cases}
(\nabla h_i)^{-1}\left(\nabla h_i\left(y_i^k\right) + c_k(\xi - l_i)\right) & \xi < l_i + \frac{1}{c_k}\left(\nabla h_i(0) - \nabla h_i\left(y_i^k\right)\right) \\
(\nabla h_i)^{-1}\left(\nabla h_i\left(y_i^k\right) + c_k(\xi - u_i)\right) & \xi > u_i + \frac{1}{c_k}\left(\nabla h_i(0) - \nabla h_i\left(y_i^k\right)\right) \\
0 & \text{otherwise}
\end{cases} \\
&= (\nabla h_i)^{-1}\left(\operatorname{mid}\left(\nabla h_i\left(y_i^k\right) + c_k(\xi - l_i), \nabla h_i(0), \nabla h_i\left(y_i^k\right) + c_k(\xi - u_i)\right)\right).
\end{aligned}
$$

Note that this operator is single-valued, so we have dropped extraneous braces.

We have not considered the possibility that $l_i = -\infty$ and/or $u_i = +\infty$. In these cases, B_{ki}^- and/or B_{ki}^+, respectively, are absent from the calculations. In all cases, however, it may be seen that the above relationship continues to hold.

Combining our results for $i = 1, \ldots, n$, we obtain that ${B_k}^{-1} = P_k$, where $P_k : \Re^n \to \Re^n$ is given by

$$P_k(x) = (\nabla h)^{-1}\left(\operatorname{mid}\left(\nabla h(y^k) + c_k(x - l), \nabla h(0), \nabla h(y^k) + c_k(x - u)\right)\right). \tag{31}$$

The dual problem $-{A_k}^{-1}(-x) + {B_k}^{-1}(x) \ni 0$ of the exact recursion formula (27) then simplifies to the equation

$$F(x) + P_k(x) = 0. \tag{32}$$

Let x^{k+1} be a solution to this equation. Invoking part (ii) of Proposition 1, the solution y^{k+1} of the original recursion (27) is simply given by

$$y^{k+1} = P_k\big(x^{k+1}\big). \tag{33}$$

Now, solving (32) for x is a considerably more familiar and tractable computation than its dual, the inclusion (27). We now address a number of issues relating to this computation: first, we would like $F + P_k$ to be differentiable, so that we can employ standard smooth numerical methods; second, we would like to solve (32) approximately, rather than exactly. We address differentiability of $F + P_k$ first.

For a start, it seems reasonable to require that F be differentiable. Therefore, the question reduces to that of the differentiability of P_k. Let us further suppose that $(\nabla h)^{-1}$ is everywhere differentiable. In this case, non-differentiabilities in P_k can only occur at "breakpoints" satisfying any of the equations

$$\begin{aligned} \nabla h_i\big(y_i^k\big) \ + \ c_k\,(x_i - l_i) \ &= \ \nabla h_i(0) \qquad i = 1, \ldots, n \\ \nabla h_i\big(y_i^k\big) \ + \ c_k\,(x_i - u_i) \ &= \ \nabla h_i(0) \qquad i = 1, \ldots, n \end{aligned}$$

that is, at $x \in \Re^n$ that have components x_i at the endpoints of any of the intervals Φ_{ki}, $i = 1, \ldots, n$. Now, $P_k(x)$ is constant as x_i moves within any of these intervals, all other coordinates being constant, that is, $[\nabla P_k(x)]_i = 0$ for $x_i \in \operatorname{int} \Phi_{ki}$. Thus, to have P_k be continuously differentiable, it must have zero derivative as x_i approaches Φ_{ki} from either above or below. Appealing to (31), this requirement is equivalent to the condition that $(\nabla h_i)^{-1}$ must have zero derivative at $\nabla h_i(0)$ for all i. Compactly, but somewhat opaquely, we require

$$\nabla \left((\nabla h)^{-1}\right)(\nabla h(0)) = 0. \tag{34}$$

To clarify this condition, we invoke the standard chain-rule based formula for the gradient of an inverse function, which in this case gives

$$\nabla \left((\nabla h_i)^{-1}\right)(x_i) = \frac{1}{\nabla^2 h((\nabla h_i)^{-1}(x_i))}$$

for all i. Therefore, we can restate the requirements that $(\nabla h)^{-1}$ be differentiable and that (34) hold as

$$\begin{array}{lll} \nabla^2 h_i(y_i) > 0 & \forall\, y_i \neq 0 & i = 1, \ldots, n \\ \lim\limits_{y_i \to 0} \nabla^2 h_i(y_i) = +\infty & & i = 1, \ldots, n \quad . \end{array} \tag{35}$$

One possible choice of a Bregman function meeting these conditions [17, Example 2] is

$$h(y) = \frac{1}{q} \sum_{i=1}^{n} |y_i|^q\ , \quad 1 < q < 2. \tag{36}$$

In this case, $\nabla h_i(y_i) = (\operatorname{sgn} y_i)|y_i|^{q-1}$, and $\nabla^2 h_i(y_i) = (q-1)|y_i|^{q-2}$ has the desired properties. We then obtain

$$P_k(x) = \operatorname{mid}\Big(\big(y^k\big)^{\langle q-1\rangle} + c_k(x-l),\, 0,\, \big(y^k\big)^{\langle q-1\rangle} + c_k(x-u)\Big)^{\left\langle \frac{1}{q-1}\right\rangle},$$

where $w^{\langle p\rangle} \doteq ((\operatorname{sgn} w_1)|w_1|^p \ \ \dots \ \ (\operatorname{sgn} w_n)|w_n|^p)$. The case $q = 3/2$ leads to an expression resembling the convex programming cubic augmented Lagrangian discussed in [24].

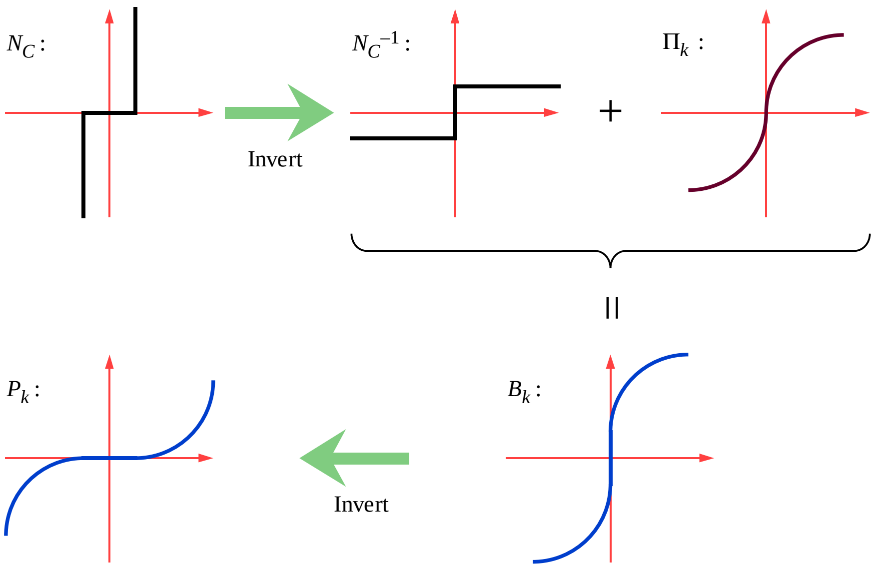

Fig. 2. Taking the inverse of N_C, adding a perturbation with an infinite slope at 0, and then inverting once again produces the smoothed exterior function P_k

Figure 2 illustrates, in the one-dimensional case $n = 1$, how dual application of the Bregman proximal method smoothes the set-valued, nonsmooth N_C term in the original problem $F(x) + N_C(x) \ni 0$ into the differentiable term P_k of the subproblem computation. First, we take N_C, and "dualize" it to obtain its inverse $N_C{}^{-1}$. To $N_C{}^{-1}$, we add the proximal perturbation function $\Pi_k : y \mapsto (1/c_k)(\nabla h(y) - \nabla h(y^k))$, which has infinite slope at 0, and finite positive slope elsewhere. This operation yields the operator B_k; because of the infinite slope of the perturbation Π_k at zero, the "corners" in the graph of $N_C{}^{-1}$ are now smoothly "rounded off." We now dualize once more by taking the inverse of B_k, obtaining the function P_k. Because of the rounded corners of B_k, P_k is a differentiable function. Note that the smoothing is applied to the exterior of C, whereas in the primal approach it is applied to the interior.

Summarizing, if we choose a separable h with zone $\Re^n$ and having the properties (35), then the system of nonlinear equations (32) to be solved at each iteration will be differentiable. Note that the domain of definition of this system will be the same as F's, since P^k is finite and defined everywhere. Therefore, unlike the primal method, there is no need for stepsize guards, except for those required for F.

To make our dual procedure practical, we need only allow for approximate solution of (32). In the following two theorems, we summarize the above development, incorporating analysis of approximate forms of the iteration; however, the approximation criteria take a somewhat strange form due to the subtleties of working in the dual. We let $\operatorname{dist}(x, Y) \doteq \inf_{y\in Y} \|x - y\|$.

Theorem 2. *Let F, l, and u describe a monotone NCP of the form (1), conforming to the hypothesis of Proposition 4, and possessing some solution. For $i = 1, \ldots, n$, let h_i be a Bregman function with zone $\Re$, and let $\{c_k\}_{k=0}^{\infty} \subset (0, \infty)$ be bounded away from zero. Suppose that the sequences $\{y^k\}_{k=0}^{\infty}$,$\{x_{[1]}^k\}_{k=1}^{\infty}$, $\{x_{[2]}^k\}_{k=1}^{\infty} \subset \Re^n$ and $\{\delta_k\}_{k=0}^{\infty} \subset [0, \infty)$ meet the conditions*

$$\sum_{k=0}^{\infty} c_k \delta_k \max\bigl(1, \|y^k\|\bigr) < \infty \tag{37}$$

$$\left\|x_{[1]}^{k+1} - x_{[2]}^{k+1}\right\| \le \delta_k \qquad \forall\, k \ge 0 \tag{38}$$

$$-F\bigl(x_{[1]}^{k+1}\bigr) = y^{k+1} = P_k\bigl(x_{[2]}^{k+1}\bigr) \quad \forall\, k \ge 0, \tag{39}$$

where P_k is defined as in (31). Then $y^k \to y^* = -F(x^*)$, where x^* is some solution to (1). All limit points x^∞ of $\{x_{[1]}^k\}$ and $\{x_{[2]}^k\}$ are also solutions of (1), with $F(x^\infty) = -y^* = F(x^*)$. If $\operatorname{im}\nabla h_i = \Re$ for all i, then such sequences are guaranteed to exist.

Proof. Invoking Proposition 4, $T_{\mathrm{D}} = -F(-I) + N_C{}^{-1}$ is maximal monotone. Also $h(x) \doteq \sum_{i=0}^n h_i(x_i)$ is a Bregman function with zone $\Re^n$. We claim that $\{y^k\}$ confirms to the recursion (26), where $\{e^k\}_{k=0}^{\infty} \subset \Re^n$ is such that $\|e^k\| \le \delta_k$ for all $k \ge 0$. The recursion can be rewritten $A_k(y^{k+1}) + B_k(y^{k+1}) \ni e^k$, where A_k and B_k are defined by (28)-(29). From (39), we have $(x_{[1]}^{k+1}, -y^{k+1}) \in F$ and $(x_{[2]}^{k+1}, y^{k+1}) \in P_k$, which yield $(y^{k+1}, -x_{[1]}^{k+1}) \in A_k$ and $(y^{k+1}, x_{[2]}^{k+1}) \in B_k$, courtesy of (28) and $P_k = B_k{}^{-1}$, as established above. Setting $e^k \doteq x_{[1]}^{k+1} - x_{[2]}^{k+1}$ for all $k \ge 1$, whence $\|e^k\| \le \delta_k$ by (38), we have $A_k(y^{k+1}) + B_k(y^{k+1}) \ni e^k$, and the claim is established.

Appealing to (37), (21) must hold with our choice of $\{e^k\}$, and also (23). All the hypotheses of Proposition 6 are thus satisfied, and so $\{y^k\}$ converges to a root of $T_{\mathrm{D}} + N_{\Re^n} = T_{\mathrm{D}}$. The final statement follows from Proposition 7, even if we were to require $\delta_k \equiv 0$, so it only remains to show that all limit points of $\{x_{[1]}^k\}$ and $\{x_{[2]}^k\}$ are primal solutions.

From (37) and $\{c_k\}$ being bounded away from zero, $\delta_k \to 0$ and $e^k \to 0$. Therefore, $\{x_{[1]}^k\}$ and $\{x_{[2]}^k\}$ have the same limit points. Let x^∞ be such that

$$x_{[1]}^k, x_{[2]}^k \underset{k\in K}{\to} x^\infty$$

for some infinite set $K \subseteq \{0, 1, 2, \ldots\}$. Since F is continuous and $y^k = -F(x_{[1]}^k)$ for all $k \ge 1$, taking limits over $k \in K$ yields $y^* = -F(x^\infty)$. From $y^{k+1} = P_k(x_{[2]}^{k+1})$, we also have $x_{[2]}^{k+1} \in B_k(y^{k+1})$, and hence

$$\left(x_{[2]}^k + \frac{1}{c_k}\left(\nabla h\big(y^k\big) - \nabla h\big(y^{k+1}\big)\right), y^{k+1}\right) \in N_C$$

for all $k \geq 0$. N_C, being maximal monotone, is a closed set in $\Re^n \times \Re^n$, while ∇h must be continuous at y^*, and $\{c_k\}$ is bounded away from zero. So, taking limits over $k \in K$ yields $(x^\infty, y^*) \in N_C$. Proposition 1 then gives that x^∞ must solve the primal problem $F(x) + N_C(x) \ni 0$.

Theorem 3. *In Theorem 2, sufficient conditions assuring (38)-(39) are*

$$F\big(x^{k+1}\big) + P_k\big(x^{k+1}\big) = 0 \tag{40}$$

$$y^{k+1} = P_k\big(x^{k+1}\big) \tag{41}$$

or

$$\mathrm{dist}\Big(x^{k+1}, F^{-1}\big(-P_k\big(x^{k+1}\big)\big)\Big) \leq \delta_k \tag{42}$$

$$y^{k+1} = P_k\big(x^{k+1}\big) \tag{43}$$

or

$$\mathrm{dist}\Big(x^{k+1}, B_k\big(-F\big(x^{k+1}\big)\big)\Big) \leq \delta_k \tag{44}$$

$$y^{k+1} = -F\big(x^{k+1}\big), \tag{45}$$

where B^k and P_k are defined as in (29) and (31), respectively. If one of these alternatives holds at each $k \geq 0$, all limit points of $\{x^k\}$ solve the complementarity problem (1). If F is continuously differentiable, $\nabla^2 h_i(y_i)$ exists and is positive for all $y_i \neq 0$, while $\lim_{y_i \to 0} \nabla^2 h_i(y_i) = +\infty$, then the function $F + P_k$ on the left-hand side of (40) is continuously differentiable.

Proof. First consider the exact iteration (40)-(41). Then we can set $x_{[1]}^{k+1} = x_{[2]}^{k+1} = x^{k+1}$, and (38)-(39) will hold for any $\delta_k \geq 0$. The continuous differentiability of $F + P_k$ follows from the discussion above.

Now consider (42)-(43). In this case, we let $x_{[2]}^{k+1} = x^{k+1}$. Since F and hence F^{-1} constitute maximal monotone operators, the set $F^{-1}(y)$ must be closed and convex for every $y \in \Re^n$ (see *e.g.* [3]). Thus, (42) guarantees the existence of some $x_{[1]}^{k+1} \in F^{-1}(-P_k(x^{k+1}))$ such that $\|x_{[1]}^{k+1} - x_{[2]}^{k+1}\| \leq \delta_k$. Thus, (38)-(39) can be satisfied.

The analysis of (44)-(45) is similar, except that we have $x_{[1]}^{k+1} = x^{k+1}$, and (44) guarantees the existence of $x_{[2]}^{k+1}$.

Since either $x^k = x_{[1]}^k$ or $x^k = x_{[2]}^k$ for every k, the assertion about limit points of $\{x^k\}$ follows from the limit point properties of $\{x_{[1]}^k\}$ and $\{x_{[2]}^k\}$

(40)-(41) constitute a generalized method of multipliers iteration for the complementarity problem (1), and by appropriate choice of h, the subproblem function $F + P_k$ of (40) can be made differentiable, if F is differentiable. Of course, such an exact procedure may not be practical. (44)-(45) is implementable in the general case and is likely to be the most useful inexact version of (40)-(41). However, in special cases where F^{-1} may be easily computed, (42)-(43) might also find application. To attempt to meet either set of approximate conditions, one would apply a standard iterative numerical method to (40) until (42) or (44) holds.

The dual method set forth in Theorems 2 and 3 has several advantages over the primal method of Section 3.1. Most crucially, the supplementary requirements P1 or P2 imposed on F in Theorem 1 may be dropped in place of the far weaker hypotheses of Proposition 4. Furthermore, the stepsize limit and ill-conditioning issues associated with the primal subproblem $F(x^{k+1}) + c_k{}^{-1}(\nabla h(x^{k+1}) - \nabla h(x^k)) \approx 0$ do not arise in the dual subproblem $F(x^{k+1}) + P_k(x^{k+1}) \approx 0$.

On the other hand, the dual method also has some disadvantages. First, the Jacobian of the primal subproblem takes the form $\nabla F + c_k{}^{-1}\nabla^2 h$, and can be forced to be positive definite by requiring that $\nabla^2 h$ be everywhere positive definite. The Jacobian $\nabla F + \nabla P_k$ of the dual subproblem, however, is only guaranteed to be positive *semi*definite, unless one requires ∇F to be positive definite. Second, the primal method has the simple, residual-based approximation rule (32), whereas the dual method requires formulas such as (42) or (44). Depending on the problem, these conditions might be difficult to verify. Finally, the dual method's theory does not guarantee convergence of the primal iterates $\{x^k\}$, $\{x^k_{[1]}\}$, or $\{x^k_{[2]}\}$, but only makes assertions about limit points.

3.3. Primal-dual application to complementarity

The primal-dual method obtained by applying Proposition 6 to $T = T_{\text{PD}} = K[F, N_C, I]$ combines and improves upon the best theoretical features of the primal and dual methods. We now consider the basic recursion (20), as applied to $T = T_{\text{PD}}$. First, we need a Bregman function $\hat{h}$ on $\Re^n \times \Re^n$, which we construct via

$$\hat{h}(x, y) = \tilde{h}(x) + \sum_{i=1}^{n} h_i(y_i), \tag{46}$$

where the h_i are as in the dual method, and $\tilde{h}$ is a Bregman function with zone $\tilde{S} \supseteq \overline{\text{dom}\, F}$. We partition the error vector e^k of (20), which in this case lies in $\Re^n \times \Re^n$, into subvectors $e^k_{[1]}, e^k_{[2]} \in \Re^n$. Then the fundamental recursion (20), with iterates $z^k = (x^k, y^k)$, Bregman function $\hat{h}$, and operator T_{PD}, takes the form

$$F\big(x^{k+1}\big) + y^{k+1} + \frac{1}{c_k}\Big(\nabla\tilde{h}\big(x^{k+1}\big) - \nabla\tilde{h}\big(x^k\big)\Big) = e^k_{[1]} \tag{47}$$

$$-x^{k+1} + N_C{}^{-1}\big(y^{k+1}\big) + \frac{1}{c_k}\Big(\nabla h\big(y^{k+1}\big) - \nabla h\big(y^k\big)\Big) \ni e^k_{[2]}, \tag{48}$$

where $h(x) = \sum_{i=1}^n h_i(x_i)$, as before. If we set $e^k_{[2]} \equiv 0$, then (48) is equivalent to $B_k(y^{k+1}) \ni x^{k+1}$, where B_k is defined as in (29) for the dual method. Using the prior definition of P_k, this condition is in turn equivalent to $y^{k+1} = P_k(x^{k+1})$, with P_k as in (31). Substituting this simple formula into (47), we obtain

$$F\big(x^{k+1}\big) + P_k\big(x^{k+1}\big) + \frac{1}{c_k}\Big(\nabla\tilde{h}\big(x^{k+1}\big) - \nabla\tilde{h}\big(x^k\big)\Big) = e^k_{[1]}.$$

At this point, application of Proposition 6 is straightforward.

Theorem 4. *Let F be a continuous monotone function that is maximal when considered as a monotone operator, with maximal open domain $D \subseteq \Re^n$. Suppose (F, l, u) describes a complementarity problem of the form (1), and that this problem has some solution. Let $\tilde{h}$ be a Bregman function with (open) zone $\tilde{S} \supseteq \overline{D}$, and let the h_i, $i = 1, \ldots, n$ be Bregman functions with zone $\Re$. Let $\{c_k\}_{k=0}^{\infty} \subset (0, \infty)$ be a sequence of positive scalars bounded away from zero, and suppose that the sequences $\{x^k\}_{k=0}^{\infty} \subset \tilde{S}$, $\{y^k\}_{k=0}^{\infty} \subset \Re^n$, and $\{d^k\}_{k=0}^{\infty} \subset \Re^n$ conform to the recursion formulae*

$$F\big(x^{k+1}\big) + \frac{1}{c_k}\Big(\nabla\tilde{h}\big(x^{k+1}\big) - \nabla\tilde{h}\big(x^k\big)\Big) + P_k\big(x^{k+1}\big) = d^k \tag{49}$$

$$y^{k+1} = P_k\big(x^{k+1}\big) \tag{50}$$

for all $k \geq 0$, where P_k is defined by (31). Suppose also that $\sum_{k=0}^{\infty} c_k\|d^k\| < \infty$, while $\sum_{k=0}^{\infty} c_k\langle d^k, x^k\rangle$ exists and is finite. Then $\{x^k\}$ converges to a solution x^* of the the complementarity problem (1), and $y^k \to -F(x^*)$. If $\operatorname{im} h_i = \Re$ for all i and $\operatorname{im}\tilde{h} = \Re^n$, such sequences are guaranteed to exist. If F is continuously differentiable and $\nabla^2 h_i(y_i)$ exists and is positive for all $y_i \neq 0$, while $lim_{y_i \to 0}\nabla^2 h_i(y_i) = +\infty$, then the function $F + c_k{}^{-1}\nabla\tilde{h} + P_k$ in the equation system (49) is continuously differentiable. If, in addition, $\nabla^2\tilde{h}$ is everywhere positive definite, then the Jacobian $\nabla F + c_k{}^{-1}\nabla^2\tilde{h} + \nabla P_k$ of this function is everywhere positive definite.

Proof. Proposition 5 asserts that T_{PD} is maximal monotone. Let $e^k = (d^k, 0) \in \Re^n \times \Re^n$ for all $k \geq 1$. Then, similarly to the above discussion, (49)-(50) are equivalent to the Bregman proximal recursion (20) with iterates $z^k = (x^k, y^k)$ and the Bregman function $\hat{h}$, which has zone $\tilde{S} \times \Re^n$. Now, $\sum_{k=0}^{\infty} c_k\|d^k\| < \infty$ is equivalent to $\sum_{k=0}^{\infty} c_k\|e^k\| < \infty$, and $\langle d^k, x^k\rangle = \langle e^k, (x^k, y^k)\rangle = \langle e^k, z^k\rangle$, so $\sum_{k=1}^{\infty} c_k\langle e^k, z^k\rangle$ exists and is finite.

We can then apply Proposition 6 to give that $\{z^k\} = \{(x^k, y^k)\}$ converges to a root $z^* = (x^*, y^*)$ of

$$T_{\text{PD}} + N_{\bar{\tilde{S}} \times \Re^n} = T_{\text{PD}}.$$

So, x^* solves (1) and $y^* = -F(x^*)$ by the analysis of Section 2.2. The claim of existence follows directly from Proposition 7. The remaining statements follow from arguments like those of Section 3.2.

Note that the primal-dual method given as (49)-(50) requires neither the primal method's restrictions P1 or P2 of Theorem 1, nor the dual method's regularity conditions of Proposition 4. The stepsize limit and ill-conditioning issues of the primal approach are also absent, because we choose the primal-space Bregman function $\tilde{h}$ to have zone containing the domain of F, as opposed to having zone int C. At the same time, the approximation criterion of (49) is based on simple measurement of a residual, as in the primal method. The Jacobian $\nabla F + {c_k}^{-1}\nabla^2\tilde{h} + \nabla P_k$ of the primal-dual subproblem function $F + {c_k}^{-1}\nabla\tilde{h} + P_k$ combines the desirable existence/continuity and positive definiteness features of the primal and dual methods. Unlike the dual method, convergence of the primal iterates $\{x^k\}$ is fully guaranteed.

Thus, the iteration (49)-(50) has all the theoretical advantages of the primal and dual approaches, and the disadvantages of neither. The three methods bear much the same relationship as the proximal minimization algorithms, methods of multipliers, and proximal methods of multipliers presented for convex optimization in [30] (for the special case $h(x) = (1/2)\|x\|^2$) and later in [17] (for general h). We therefore refer to the dual method as a "method of multipliers," and the primal-dual method as a "proximal method of multipliers."

4. Computational results on the MCPLIB test suite

We conclude with some preliminary computational results for the proximal method of multipliers. We coded a version of the algorithm (49)-(50) in MATLAB, and used it to solve the problems in the MCPLIB collection [14], exploiting the interface developed in [19]. We note that most of the problems in the collection do not satisfy the monotonicity condition (5) postulated in our theory. In fact, only the problems `cycle` and `optcont31` are definitely known to be monotone. However, for the method to be practical, we believe it must robustly solve a large number of the problems from this standard test suite.

In our initial implementation, we set $\|d^k\| < 10^{-6}$ for all k, that is, we solved (49) essentially exactly at all iterations. With later work, we intend to refine this approach, starting from a larger tolerance and gradually decreasing it. We chose h as in (36) with $q = 3/2$, and set $\tilde{h}(x) = (1/2)x^\top Dx$, D being a diagonal matrix determined via

$$D_{ii} = \frac{1.0}{\max\left(0.1\left\|\nabla F_{ii}(x^0)\right\|, 10.0\right)}.$$

This choice corresponds to standard problem scaling mechanisms that have proven successful in [10, 15]. In the interest of further improving scaling, we also define the function P_k slightly differently from (31). Instead, we use $P_k(x) = P(x, y^k; c_k)$ where

$$P(x, y; c) \doteq (\nabla h)^{-1}\left(\operatorname{mid}\begin{pmatrix}\nabla h(y) + cD^{-1}(x - l)\\ \nabla h(0)\\ \nabla h(y) + cD^{-1}(x - u)\end{pmatrix}\right), \tag{51}$$

D being the diagonal matrix defined above. This change corresponds to a simple rescaling of the overall Bregman function $\hat{h}$ of (46).

By way of illustration, consider the special case of minimization over the nonnegative orthant, where we have $F = \nabla f$ for some differentiable convex function f, $l = 0$, and

$u = +\infty$. Then the version of (49)-(50) we implemented would correspond to the following cubic augmented Lagrangian method, with a quadratic proximal term:

$$x^{k+1} = \arg\min_{x \in \Re^n} \left\{ f(x) + \frac{1}{2c_k}\left(x - x^k\right)^\top D\left(x - x^k\right) + \frac{1}{3}\sum_{j=1}^{n} \max\left\{ \left[\sqrt{y_j^k} + \frac{c_k x_j}{D_{jj}}\right]^3, 0 \right\} \right\}$$
$$y_j^{k+1} = \max\left(\sqrt{y_j^k} + \frac{c_k x_j^{k+1}}{D_{jj}}, 0\right)^2.$$

The initial values x^0 of the primal variables are specified in the MCPLIB test suite [14]. For the initial multipliers, we used the formula

$$y^0 = \begin{cases} P\left(x^0, -F\left(x^0\right); c_0\right), & \left\| P\left(x^0, -F\left(x^0\right); c_0\right) \right\| \geq 10^{-6} \\ -F\left(x^0\right), & \text{otherwise,} \end{cases}$$

where P is defined by (51).

The major work involved in each step of the algorithm is in solving the system of nonlinear equations (49), for which we use a simple backtracking variant of Newton's method. We start by computing a "pure" Newton step for (49), with d^k replaced by zero. If this step does not yield a reduction in the residual of (49), we repeatedly halve the step size until a reduction is obtained, or the step is less than $1/1000$th of its original magnitude. In the former case, we then attempt another Newton step, repeating the process until the residual of (49) falls below 10^{-6}. We then update the multiplier vector via (50), and check the global residual $r_k \doteq \|F(x^k) + y^k\|$. If $r_k < 10^{-6}$, we successfully terminate. Otherwise, if $k < 100$, we loop, increment k, and execute another "outer" iteration. If $k \geq 100$ we quit and declare failure.

When the Newton line search fails, that is, a reduction of the step by a factor of $1/1024$ fails to yield any improvement in the residual of (49), we update the proximal stepsize parameter c_k. In fact, we separately maintain a primal c_k ("pc_k") and a dual c_k ("dc_k"), corresponding to the usage of c_k in the equations (49) and (31)/(51), respectively. Allowing for additional rescaling of $\tilde{h}$, the convergence theory above stipulates that pc_k and dc_k be held in a fixed ratio to one another throughout the algorithm. In practice, we allow a limited number of independent adjustments of these two parameters. Assuming monotonicity of F, our convergence theory applies after the last such independent adjustment.

We start by setting p$c_0 = \max\{10, \|x^0\|\}$ and d$c_0 = 10$. Upon failure of the line search, pc_k is reduced by a factor of 10 and dc_k is set to 1. After successful solution of (49) to the tolerance of 10^{-6}, both pc_k and dc_k are multiplied by 1.05; this adjustment is consistent with our theory and also with standard techniques for accelerating convergence of proximal methods. We then calculate y^{k+1}, and if

$$\left\|x^{k+1} - x^k\right\| > 100\left\|y^{k+1} - y^k\right\|,$$

dc_k is doubled, whereas if

$$100\left\|x^{k+1} - x^k\right\| < \left\|y^{k+1} - y^k\right\|,$$

then d$c_k = \|y^k\|$.

Table 1. Primal-dual smooth multiplier method applied to MCPLIB problems (part 1)

Problem (Starting Point)	Iterations	Newton Steps	Updates of pc_k	Updates of $\mathrm{d}c_k$	Primal Residual
bertsekas (1)	15	40	0	0	5.4×10^{-7}
bertsekas (2)	15	47	0	0	6.3×10^{-7}
bertsekas (3)	6	59	0	0	1.2×10^{-8}
billups (1)	47	350	3	21	4.9×10^{-7}
choi (1)	5	8	0	1	9.3×10^{-7}
colvdual (1)	9	29	1	1	7.2×10^{-8}
colvdual (2)	7	36	0	0	2.4×10^{-7}
colvnlp (1)	9	28	1	1	7.4×10^{-8}
colvnlp (2)	7	25	0	0	2.3×10^{-7}
cycle (1)	4	11	0	0	8.3×10^{-7}
ehl_kost (1)	4	15	0	0	5.5×10^{-7}
ehl_kost (2)	4	15	0	0	5.5×10^{-7}
ehl_kost (3)	4	15	0	0	5.5×10^{-7}
explcp (1)	6	21	0	0	5.6×10^{-7}
freebert (1)	15	39	0	0	4.0×10^{-7}
freebert (2)	9	24	0	0	8.4×10^{-7}
freebert (3)	15	39	0	0	3.7×10^{-7}
freebert (4)	15	40	0	0	5.4×10^{-7}
freebert (5)	9	24	0	0	8.4×10^{-7}
freebert (6)	15	40	0	0	5.0×10^{-7}
gafni (1)	9	23	0	0	2.7×10^{-7}
gafni (2)	9	26	0	0	3.0×10^{-7}
gafni (3)	9	28	0	0	3.3×10^{-7}
hanskoop (1)	5	30	0	0	1.4×10^{-7}
hanskoop (2)	11	108	1	1	8.0×10^{-7}
hanskoop (3)	5	17	0	0	1.1×10^{-7}
hanskoop (4)	5	26	0	0	1.4×10^{-7}
hanskoop (5)	11	78	1	1	7.0×10^{-7}
hydroc06 (1)	5	9	0	0	5.8×10^{-7}
hydroc20 (1)					failed
josephy (1)	13	105	2	2	6.3×10^{-7}
josephy (2)	8	90	1	1	5.6×10^{-8}
josephy (3)	7	138	1	1	8.7×10^{-7}
josephy (4)	5	14	0	0	8.9×10^{-9}
josephy (5)	4	10	0	0	4.4×10^{-7}
josephy (6)	8	166	1	1	1.9×10^{-7}
kojshin (1)	56	248	3	3	5.3×10^{-7}
kojshin (2)	9	151	1	1	8.4×10^{-8}
kojshin (3)	43	357	4	4	8.6×10^{-7}
kojshin (4)	19	214	2	2	8.4×10^{-7}
kojshin (5)	20	227	2	2	5.5×10^{-7}
kojshin (6)	52	391	3	3	7.6×10^{-7}
mathinum (1)	5	9	0	0	4.1×10^{-8}
mathinum (2)	5	8	0	0	1.3×10^{-8}
mathinum (3)	5	13	0	0	2.4×10^{-8}
mathinum (4)	5	9	0	0	4.5×10^{-8}

Table 2. Primal-dual smooth multiplier method applied to MCPLIB problems (part 2)

Problem (Starting Point)	Iterations	Newton Steps	Updates of pc_k	Updates of dc_k	Primal Residual
`mathisum` (1)	4	9	0	0	2.5×10^{-7}
`mathisum` (2)	5	11	0	0	1.9×10^{-8}
`mathisum` (3)	5	19	0	0	3.9×10^{-8}
`mathisum` (4)	4	8	0	0	9.8×10^{-7}
`methan08` (1)	4	7	0	1	2.9×10^{-7}
`nash` (1)	5	10	0	0	1.2×10^{-8}
`nash` (2)	4	9	0	0	5.2×10^{-8}
`opt_cont31` (1)	6	85	0	0	4.7×10^{-7}
`pies` (1)	7	29	1	1	6.5×10^{-7}
`pgvon105` (1)					failed
`pgvon106` (1)					failed
`powell` (1)	4	12	0	0	4.2×10^{-7}
`powell` (2)	6	21	0	0	1.0×10^{-7}
`powell` (3)	14	176	2	2	2.8×10^{-7}
`powell` (4)	6	21	0	0	8.2×10^{-8}
`powell_mcp` (1)	5	10	0	0	2.2×10^{-7}
`powell_mcp` (2)	5	10	0	0	3.8×10^{-7}
`powell_mcp` (3)	5	14	0	1	1.6×10^{-7}
`powell_mcp` (4)	5	13	0	0	6.7×10^{-7}
`scarfanum` (1)	6	24	0	0	3.0×10^{-7}
`scarfanum` (2)	6	28	0	0	3.0×10^{-7}
`scarfanum` (3)	7	28	0	0	1.5×10^{-7}
`scarfasum` (1)	6	25	0	0	1.4×10^{-7}
`scarfasum` (2)	6	21	0	0	1.4×10^{-7}
`scarfasum` (3)	10	36	0	0	2.9×10^{-7}
`scarfbnum` (1)	43	133	0	0	5.9×10^{-7}
`scarfbnum` (2)	89	393	0	21	9.0×10^{-7}
`scarfbsum` (1)	18	81	0	0	5.7×10^{-7}
`scarfbsum` (2)	18	66	0	0	5.8×10^{-7}
`sppe` (1)	6	21	0	0	5.8×10^{-8}
`sppe` (2)	5	22	0	0	6.9×10^{-9}
`tobin` (1)	6	30	0	0	3.5×10^{-8}
`tobin` (2)	6	47	0	0	3.3×10^{-8}

Tables 1 and 2 summarize our computational results. “Iterations” is the total number of “outer” iterations, that is, the value of k necessary to obtain $r_k < 10^{-6}$. “Newton steps” is the total number of Newton steps taken, accumulated over all outer iterations. We also report the number of times that pc_k and dc_k are updated independently of one another; these counts do not include the simultaneous multiplications by 1.05. Note that there were no independent updates required for the two guaranteed monotone problems, as our convergence theory would suggest. For the remaining problems, independent updates were infrequent. Since our implementation is preliminary and MATLAB is an interpreted language, we do not list run times. The “primal residual” column gives the final value of $\|x^k - \text{mid}\big(l, x^k - F(x^k), u\big)\|$.

As can be seen from the tables, and by comparison with the results in [2], the algorithm is fairly robust. For all but 3 of the 79 instance/starting point combinations attempted, it terminates within 100 iterations with a primal residual of 10^{-6} or less, indicating convergence to a solution. Two of the failures were for the `pgvon10*` problems; since these problems are known to be poorly defined at the solution, we do not consider these failures to be a serious liability. The other failure, on `hydroc20`, seems to be due to convergence difficulties in the multiplier space. `Hydroc20` contains a large number of nonlinear equations, and we speculate that (36) with $q = 3/2$ may not be an ideal penalty kernel to use in such cases.

Acknowledgements. The authors also wish to thank an anonymous referee for suggestions on streamlining some of the analysis.

References

1. Attouch, H., Théra, M. (1996): A general duality principle for the sum of two operators. J. Convex Anal. **3**, 1–24
2. Billups, S.C., Dirkse, S.P., Ferris, M.C. (1997): A comparison of large scale mixed complementarity problem solvers. Comput. Optim. Appl. **7**, 3–25
3. Brézis, H. (1973): Opérateurs Maximaux Monotones et Semi-Groupes de Contractions dans les Espaces de Hilbert. North-Holland, Amsterdam
4. Bruck, R.E. (1975): An iterative solution of a variational inequality for certain monotone operators in Hilbert space. Bull. Am. Math. Soc. **81**, 890–892
5. Burachik, R.S., Iusem, A.N. (1995): A generalized proximal point algorithm for the nonlinear complementarity problem. Working paper, Instituto de Matemática Pura e Aplicada, Rio de Janeiro
6. Burachik, R.S., Iusem, A.N. (1998): A generalized proximal point algorithm for the variational inequality problem in a Hilbert space. SIAM J. Optim. **8**, 197–216
7. Burachik, R.S., Iusem, A.N., Svaiter, B.F. (1997): Enlargement of monotone operators with applications to variational inequalities. Set-Valued Anal. **5**, 159–180
8. Censor, Y., Iusem, A.N., Zenios, S.A. (1998): An interior-point method with Bregman functions for the variational inequality problem with paramonotone operators. Math. Program. **81**, 373–400
9. Censor, Y., Zenios, S.A. (1992): The proximal minimization algorithm with *D*-functions. J. Optim. Theory Appl. **73**, 451–464
10. Chen, C., Mangasarian, O.L. (1995): Smoothing methods for convex inequalities and linear complementarity problems. Math. Program. **78**, 51–70
11. Chen, C., Mangasarian, O.L. (1996): A class of smoothing functions for nonlinear and mixed complementarity problems. Comput. Optim. Appl. **5**, 97–138
12. Chen, G., Teboulle, M. (1993): A convergence analysis of proximal-like minimization algorithms using Bregman functions. SIAM J. Optim. **3**, 538–543
13. De Pierro, A.R., Iusem, A.N. (1986): A relaxed version of Bregman's method for convex programming. J. Optim. Theory Appl. **51**, 421–440
14. Dirkse, S.P., Ferris, M.C. (1995): MCPLIB: A collection of nonlinear mixed complementarity problems. Optim. Methods Software **5**, 319–345
15. Dirkse, S.P., Ferris, M.C. (1995): The PATH solver: A non-monotone stabilization scheme for mixed complementarity problems. Optim. Methods Software **5**, 123–156
16. Eckstein, J. (1989): Splitting Methods for Monotone Operators, with Applications to Parallel Optimization. PhD thesis. Massachusetts Institute of Technology, Cambridge, MA, 1989. Also available as Report LIDS-TH-1877, Laboratory for Information and Decision Systems, Massachusetts Institute of Technology, Cambridge, MA, 1989
17. Eckstein, J. (1993): Nonlinear proximal point algorithms using Bregman functions, with applications to convex programming. Math. Oper. Res. **18**, 202–226
18. Eckstein, J. (1998): Approximate iterations in Bregman-function-based proximal algorithms. Math. Program. **83**, 113–124

19. Ferris, M.C., Rutherford, T.F. (1996): Accessing realistic complementarity problems within Matlab. In: Di Pillo, G., Giannessi, F., eds., Proceedings of Nonlinear Optimization and Applications Workshop, Erice, June 1995. Plenum Press, New York
20. Fukushima, M. (1996): The primal Douglas-Rachford splitting algorithm for a class of monotone mappings with application to the traffic equilibrium problem. Math. Program. **72**, 1–15
21. Gabay, D. (1983): Applications of the method of multipliers to variational inequalities. In: Fortin, M., Glowinski, R., eds., Augmented Lagrangian Methods: Applications to the Solution of Boundary Value Problems. North-Holland, Amsterdam
22. Gabriel, S.A., Moré, J.J. (1997): Smoothing of mixed complementarity problems. In: Ferris, M.C., Pang, J.S., eds., Complementarity and Variational Problems: State of the Art. SIAM Publications, Philadelphia
23. Kabbadj, S. (1994): Methodes Proximales Entropiques. Doctoral thesis. Université de Montpellier II – Sciences et Techniques du Languedoc, Montpellier, France
24. Kiwiel, K.C. (1996): On the twice differentiable cubic augmented Lagrangian. J. Optim. Theory Appl. **88**, 233–236
25. Luo, Z.-Q., Tseng, P. (1993): Error bounds and convergence analysis of feasible descent methods: A general approach. Ann. Oper. Res. **46**, 157–178
26. Minty, G.J. (1962): Monotone (nonlinear) operators in Hilbert space. Duke Math. J. **29**, 341–346
27. Mosco, U. (1972): Dual variational inequalities. J. Math. Anal. Appl. **40**, 202–206
28. Rockafellar, R.T., (1970): Convex Analysis. Princeton University Press, Princeton, NJ
29. Rockafellar, R.T. (1970): On the maximality of sums of nonlinear monotone operators. Trans. Am. Math. Soc. **149**, 75–88
30. Rockafellar, R.T. (1976): Augmented Lagrangians and applications of the proximal point algorithm in convex programming. Math. Oper. Res. **1**, 97–116
31. Rockafellar, R.T. (1976): Monotone operators and the proximal point algorithm. SIAM J. Control Optim. **14**, 877–898
32. Rockafellar, R.T., Wets, R.J.-B. (1998): Variational Analysis. Springer
33. Teboulle, M. (1992): Entropic proximal mappings with applications to nonlinear programming. Math. Oper. Res. **17**, 670–690