JOURNAL OF OPTIMIZATION THEORY AND APPLICATIONS: Vol. 73, No. 3, JUNE 1992

Proximal Minimization Algorithm with D-Functions[1,2]

Y. CENSOR[3] AND S. A. ZENIOS[4]

Communicated by O. L. Mangasarian

Abstract. The original proximal minimization algorithm employs quadratic additive terms in the objectives of the subproblems. In this paper, we replace these quadratic additive terms by more general D-functions which resemble (but are not strictly) distance functions. We characterize the properties of such D-functions which, when used in the proximal minimization algorithm, preserve its overall convergence. The quadratic case as well as an entropy-oriented proximal minimization algorithm are obtained as special cases.

Key Words. Proximal minimization algorithms, Bregman functions, D-functions, entropy optimization.

1. Introduction

The proximal minimization algorithm is designed to solve the optimization problem

$$\min \quad F(x), \tag{1a}$$

$$\text{s.t.} \quad x \in X, \tag{1b}$$

[1]The work of the first author was supported by NSF Grant CCR-8811135 and NIH Grant HL-28438, while visiting the Decision Sciences Department of the Wharton School and the Medical Image Processing Group at the Department of Radiology, both at the University of Pennsylvania, Philadelphia, Pennsylvania. The work of the second author was supported by NSF Grant CCR-91-04042 and AFOSR Grant 91-0168.

[2]The authors received valuable comments on the December 1989 and June 1990 versions of this paper, which led to considerable improvements of the results and their presentation. For this, they are indebted to D. Bertsekas, A. De Pierro, J. Eckstein, P. Eggermont, A. Iusem, M. Ferris, M. Teboulle, and three anonymous referees.

[3]Professor of Mathematics and Computer Science, University of Haifa, Mt. Carmel, Haifa, 31905 Israel.

[4]Associate Professor of Decision Sciences, Wharton School, University of Pennsylvania, Philadelphia, Pennsylvania 19104-6366.

where $F:\mathbb{R}^n \to \mathbb{R}$ is a given convex function and $X \subseteq \mathbb{R}^n$ is a nonempty closed convex subset of the n-dimensional Euclidean space $\mathbb{R}^n$. The approach is based on converting (1) into a sequence of optimization problems with strictly convex objective functions obtained by adding quadratic terms to $F(x)$.

The method is as follows. There is a (given or constructed) sequence $\{c(t)\}$ of positive numbers for all $t \in \mathbb{N}_0$, $\mathbb{N}_0 = \{0, 1, 2, 3, \ldots\}$, with

$$\liminf_{t\to\infty} c(t) = c > 0. \tag{2}$$

A sequence $\{x(t)\}$ is generated, starting from an arbitrary initial vector $x(0) \in \mathbb{R}^n$, by

$$x(t+1) = \arg\min_{x\in X}\{F(x) + [1/2c(t)]\|x - y(t)\|^2\}, \tag{3a}$$

$$y(t+1) = x(t+1), \tag{3b}$$

where $\|\cdot\|$ is the standard Euclidean norm in $\mathbb{R}^n$. Equivalently, the algorithm is written as

$$x(t+1) = \arg\min_{x\in X}\{F(x) + [1/2c(t)]\|x - x(t)\|^2\}. \tag{4}$$

The origins of this algorithm go back to Minty (Ref. 1), Moreau (Ref. 2), and Rockafellar (Refs. 3–4). In addition to considerable theoretical interest in the family of proximal point algorithms, of which it is a member, this algorithm is also an important computational tool. This is so because the dual problem of a strictly convex optimization problem is differentiable and can be solved by simple iterative procedures like dual coordinate ascent. For several important problem classes, these dual algorithms can be decomposed for parallel computations; the results of this investigation are reported elsewhere in Nielsen and Zenios (Refs. 28–29).

In this paper, we generalize the proximal minimization algorithm by replacing the quadratic term in (3a) by a function $D:\mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ and specifying the structure and properties of some such D-functions for which convergence of the algorithm can be preserved.

These D-functions were introduced by Bregman (Ref. 5) and subsequently studied further in conjunction with primal–dual methods for linearly constrained convex programming by Censor and Lent (Ref. 6) and by De Pierro and Iusem (Ref. 7). The original proximal minimization algorithm (3) is obtained from our scheme by one special choice of a D-function. A different choice leads to a proximal minimization algorithm with entropy additive terms. In the case of linear programming (F and

$x \in X$ are all linear), the latter leads to pure entropy optimization problems for which several good special-purpose algorithms exist; see, e.g., Refs. 8–10. Such an approach of replacing a linear programming problem by a sequence of entropy problems was heuristically suggested by Eriksson (Ref. 11). He discusses also a specific strategy for choosing the parameters $\{c(t)\}$ and a solution algorithm. However, no overall convergence analysis is given there. The practical question of whether any efficient useful algorithm result from this new look at things has been addressed by Nielsen and Zenios (Refs. 28–29), where encouraging computational results are reported.

The fundamental proximal point algorithm for solving the problem $0 \in T(z)$ for an arbitrary maximal monotone operator T and its specialization for $T = \partial F$ (the subdifferential of F) make it clear why quadratic additive terms in (3a) are mandatory; see, e.g., Ref. 3. Therefore, we do not resort to the operator theory, but rather follow the more direct method of Ref. 12. It is quite conceivable that the idea of incorporating D-functions could propagate in other directions within the theory of proximal point and related methods.

The idea of replacing quadratic penalty terms by nonquadratic ones exists already with respect to other algorithms; see, e.g., Ref. 13, Chapter 5. Bertsekas (Ref. 14) kindly pointed out that the special entropy case of our PMD algorithm (see Section 4 below) is the Fenchel dual to the primal augmented Lagrangian minimization with exponential penalty. In a similar vein, Teboulle (Ref. 15) has recently derived what he calls "entropic proximal maps" and used them to construct generalized augmented Lagrangian methods. Although his paper can be considered a close companion to ours, his results do not include the PMD algorithm that we propose here. See also Ref. 16. An important work on monotone operators and the proximal point algorithm is Eckstein's thesis (Ref. 17). Moreover, in his recent paper (Ref. 18), Eckstein showed how to construct proximal point algorithms with Bregman functions, thereby further extending the scope of the connection between Bregman functions and proximal minimization presented here.

Another recent related study is Eggermont's (Ref. 19). Nonquadratic additive terms are used there, but with only nonnegativity constraints. Of particular interest is the connection revealed there between multiplicative iterative algorithms and the well-known EM-algorithm for maximum likelihood estimation in emission tomography; see Shepp and Vardi (Ref. 20) and other references in Ref. 19. The algorithms of Ref. 19, however, are not special instances of our proximal minimization algorithm with D-functions. Finally, we mention the work of Tseng and Bertsekas (Ref. 21), where they use the entropy proximal term in the proximal minimization algorithm to study the exponential multiplier method.

2. Proximal Minimization Algorithm with D-Functions

Let S be a nonempty open convex set in $\mathbb{R}^n$ such that $\bar{S} \subseteq \Lambda$, where $\bar{S}$ is the closure of S and Λ is the domain of a function $f: \Lambda \subseteq \mathbb{R}^n \to \mathbb{R}$. Assume that $f(x)$ is twice continuously differentiable at every $x \in S$, and denote by $\nabla f(x)$ and $\nabla^2 f(x)$ its gradient and its Hessian matrix at x, respectively. Furthermore, assume that $f(x)$ is continuous and strictly convex on $\bar{S}$.

The set S is called the zone of f, and f obeying the assumptions made above will be referred to as an auxiliary function.

From $f(x)$, construct the D-function $D_f(x, y)$, $D_f: \bar{S} \times S \subseteq \mathbb{R}^{2n} \to \mathbb{R}$, by

$$D_f(x, y) = f(x) - f(y) - \langle \nabla f(y), x - y \rangle, \tag{5}$$

where $\langle \cdot, \cdot \rangle$ denotes the usual inner product in $\mathbb{R}^n$. Such D_f-functions are a specific realization of the more general D-functions defined in Ref. 5 and will be clearly distinguishable from the latter by the subscript f. D_f-functions appear in Ref. 5, p. 206, and in Refs. 6, 22. They are instrumental in defining D_f-projections onto convex sets and play a key role in the primal-dual optimization algorithms in Refs. 5, 6, 22.

We will need the following additional properties to be postulated for the auxiliary functions, their zones, and the D_f-functions constructed from them. For any $\alpha \in \mathbb{R}$, denote by

$$L_f^1(\alpha, y) = \{x \in \bar{S} \mid D_f(x, y) \leq \alpha\}, \tag{6a}$$

$$L_f^2(x, \alpha) = \{y \in S \mid D_f(x, y) \leq \alpha\}, \tag{6b}$$

the partial level sets of $D_f(x, y)$.

Assumption A1. For every $\alpha \in \mathbb{R}$, the partial level sets $L_f^1(\alpha, y)$ and $L_f^2(x, \alpha)$ are bounded for every $y \in S$, for every $x \in \bar{S}$, respectively.

Assumption A2. If $y^k \xrightarrow[k\to\infty]{} y^* \in \bar{S}$, then $D_f(y^*, y^k) \xrightarrow[k\to\infty]{} 0$.

Assumption A3. If $D_f(x^k, y^k) \xrightarrow[k\to\infty]{} 0$, $y^k \xrightarrow[k\to\infty]{} y^* \in \bar{S}$, and $\{x^k\}$ is bounded, then $x^k \xrightarrow[k\to\infty]{} y^*$.

Assumption A4. $D_f(x, y)$ is jointly convex w.r.t. both x and y, i.e., as a function on $\mathbb{R}^{2n}$.

With milder differentiability assumptions, auxiliary functions f which obey Assumptions A1-A3 are called Bregman functions in Ref. 6. In particular, all results of Section 2 and 3 of Ref. 6 hold. $D_f(x, y)$ is not a distance function, but

$$D_f(x, y) \geq 0 \quad \text{and} \quad D_f(x, y) = 0, \text{ iff } x = y; \tag{7}$$

see Ref. 6, Lemma 2.1.

Definition 2.1. Given $\Omega \subseteq \mathbb{R}^n$ and $y \in S$, a point $x^* \in \Omega \cap \bar{S}$ for which

$$x^* = \arg\min_{z \in \Omega \cap \bar{S}} D_f(z, y) \tag{8}$$

is a D_f-projection of y onto Ω, denoted by $P_\Omega y = x^*$. The existence and uniqueness of D_f-projections onto closed convex sets are guaranteed by Lemma 2.2 of Ref. 6.

If $f(x) = (1/2)\|x\|^2$ and $\Lambda = S = \bar{S} = \mathbb{R}^n$, then $D_f(x, y) = (1/2)\|x - y\|^2$ and D_f-projections are ordinary orthogonal projections.

The proximal minimization algorithm with D_f-functions, henceforth abbreviated PMD, is as follows. Given are an auxiliary function f with zone S, satisfying Assumptions A1-A4, and a positive sequence $\{c(t)\}$ for which (2) holds.

Initialization. $x(0) \in S$ is arbitrary.

Iterative Step. $$x(t+1) = \arg\min_{x \in X \cap \bar{S}} \{F(x) + [1/c(t)] D_f(x, x(t))\}. \tag{9}$$

In order that this algorithm be well defined, we make the next assumption.

Assumption A5. The PMD algorithm (9) generates a sequence $\{x(t)\}$ such that $x(t) \in S$, $\forall t$.

This assumption is needed, because D_f is defined on $\bar{S} \times S$. It actually tells us that, given F and X of (1), we are free to choose only such f and S that Assumption A5 would hold. If $X \subseteq S$, then Assumption A5 trivially holds, which is true for the quadratic case $f = (1/2)\|x\|^2$, where $S = \mathbb{R}^n$. We show later that it holds also for the entropy case.

3. Convergence Analysis of the PMD Algorithm

The analysis given here follows the one given in Ref. 12, Chapter 3.4.3. Proposition 3.1 secures the existence and uniqueness of the minimum of $\{F(x) + (1/c) D_f(x, y)\}$. Propositions 3.2, 3.3, and 3.4 are not directly necessary for the proof of convergence, but they extend to the D_f-function setting some closely related results from Ref. 12. One referee has correctly observed that a few steps in the proofs of Propositions 3.2 and 3.3 can be deduced from existing results. We prefer, however, to supply full proofs which better demonstrate how D_f-functions fit into the analysis. Proposition 3.5 is a generalization of a classical lemma about projection operators which is crucial for the final convergence result given in Proposition 3.6.

Denote by X^* the solution set of problem (1),

$$X^* = \{x^* \in X \mid F(x^*) \leq F(x), \forall x \in X\}. \tag{10}$$

Proposition 3.1. Let f be an auxiliary function with zone S, and let Assumption A1 hold. For every $y \in S$ and $c > 0$, the minimum of $\{F(x) + (1/c)D_f(x, y)\}$ over $X \cap \bar{S}$ is attained at a unique point, denoted by $x_f(y, c)$, provided that $F(x)$ is bounded below over X.

Proof. For all $c > 0$ and $y \in S$, the level sets

$$\{x \in X \cap \bar{S} \mid F(x) + (1/c)D_f(x, y) \le \alpha\}, \qquad \alpha \in \mathbb{R}, \tag{11}$$

are bounded. This is true because otherwise, for some $c > 0$ and $y \in \mathbb{R}^n$, there would exist an unbounded sequence $\{x^k\} \subseteq X \cap \bar{S}$ for which

$$D_f(x^k, y) \le c(\alpha - L), \tag{12}$$

where L is the lower bound for $F(x)$ over X. But this would contradict Assumption A1; thus, the level sets (11) must be bounded.

This allows us to equivalently search for the minimum of $F(x) + (1/c)D_f(x, y)$ over a compact subset of $X \cap \bar{S}$ instead of $X \cap \bar{S}$. The Weierstrass theorem (e.g., Ref. 12, Proposition A.8) then implies that the above-mentioned minimum is attained. The strict convexity of $D_f(x, y)$ with respect to x for fixed y, which follows from (5), and the strict convexity of the auxiliary function f, ensure the uniqueness. □

Proposition 3.2. If Assumption A4 holds, then the function $\Phi_c : S \to \mathbb{R}$, defined by

$$\Phi_c(y) = \min_{x \in X \cap \bar{S}} \{F(x) + (1/c)D_f(x, y)\}, \tag{13}$$

is convex over S.

Proof. Let $y^1, y^2 \in S$ and $\alpha \in [0, 1]$. Denote $x_f^i = x_f(y^i, c)$, for $i = 1, 2$. Using the convexity of F and Assumption A4, we have

$$\begin{aligned}
&\alpha\Phi_c(y^1) + (1-\alpha)\Phi_c(y^2) \\
&= \alpha[F(x_f^1) + (1/c)D_f(x_f^1, y^1)] + (1-\alpha)[F(x_f^2) + (1/c)D_f(x_f^2, y^2)] \\
&\ge F(\alpha x_f^1 + (1-\alpha)x_f^2) + (1/c)D_f(\alpha x_f^1 + (1-\alpha)x_f^2, \alpha y^1 + (1-\alpha)y^2) \\
&\ge \min_{x \in X \cap \bar{S}} \{F(x) + (1/c)D_f(x, \alpha y^1 + (1-\alpha)y^2)\} \\
&= \Phi_c(\alpha y^1 + (1-\alpha)y^2).
\end{aligned}$$

□

Proposition 3.3. Let Assumptions A1 and A4 hold. The function $\Phi_c(y)$ is continuously differentiable on S and its gradient is given by

$$\nabla\Phi_c(y) = \nabla^2 f(y)^T[[y - x_f(y, c)]/c], \tag{14}$$

where T denotes matrix transposition.

Proof. Consider any $y \in S$, $d \in \mathbb{R}^n$, and $\alpha > 0$ such that $y + \alpha d \in S$. Using the directional derivative $\Phi'_c(y; d)$, we have

$$\begin{aligned} &F(x_f(y, c)) + (1/c) D_f(x_f(y, c), y + \alpha d) \\ &\geq \Phi_c(y + \alpha d) \geq \Phi_c(y) + \alpha \Phi'_c(y; d) \\ &= F(x_f(y, c)) + (1/c) D_f(x_f(y, c), y) + \alpha \Phi'_c(y; d), \end{aligned} \tag{15}$$

where the second inequality in (15) follows from the convexity of Φ_c (Proposition 3.2). Therefore, using (5), we get from (15)

$$\begin{aligned} &(1/c)[f(y) - f(y + \alpha d) + \langle \nabla f(y) - \nabla f(y + \alpha d), x_f(y, c) - y \rangle \\ &\qquad + \langle \nabla f(y + \alpha d), \alpha d \rangle] \geq \alpha \Phi'_c(y; d). \end{aligned} \tag{16}$$

Since

$$\lim_{\alpha \to 0} [[f(y) - f(y + \alpha d)]/\alpha + (1/\alpha)\langle \nabla f(y + \alpha d), \alpha d \rangle] = 0, \tag{17}$$

$$\lim_{\alpha \to 0} [[\nabla f(y) - \nabla f(y + \alpha d)]/\alpha] = -\nabla^2 f(y) d, \tag{18}$$

we obtain from (16), by dividing by α and letting $\alpha \to 0$,

$$\langle \nabla^2 f(y) d, [y - x_f(y, c)]/c \rangle \geq \Phi'_c(y; d), \qquad \forall d \in \mathbb{R}^n. \tag{19}$$

Replacing d by $-d$ in (19), we get

$$-\langle \nabla^2 f(y) d, [y - x_f(y, c)]/c \rangle \geq \Phi'_c(y; -d) \geq -\Phi'_c(y; d), \tag{20}$$

where the second inequality is a standard relation for directional derivatives of convex functions; see, e.g., Ref. 12, p. 648.

The relations (19) and (20) imply that

$$\Phi'_c(y; d) = \langle \nabla^2 f(y) d, [y - x_f(y, c)]/c \rangle, \qquad \forall d \in \mathbb{R}^n, \tag{21}$$

or equivalently that Φ_c is differentiable and that its gradient is given by (14). Since Φ_c is convex (Proposition 3.2), its gradient is continuous; see, e.g., Ref. 12, Proposition A.42. □

The next proposition gives a relation between S^*, the minimum set of $\Phi_c(y)$, the zone S of the auxiliary function f, and the solution set X^*. For a function f with zone $S = \mathbb{R}^n$, we get, as a special case, that $X^* = S^*$, which was given in Ref. 12, p. 234.

Define

$$S^* = \{y^* \in S \mid \Phi_c(y^*) \leq \Phi_c(y), \forall y \in S\}. \tag{22}$$

Proposition 3.4. Let $\nabla^2 f(z)$ be nonsingular for all $z \in S^*$. Then,

$$X^* \cap S = S^*. \tag{23}$$

Proof. The function $F(x)+(1/c)D_f(x,y)$ takes the value $F(y)$ for $x=y$ because $D_f(x,y)=0$, iff $x=y$ (Ref. 6, Lemma 2.1). It follows that

$$\Phi_c(y)\leq F(y), \qquad \forall y\in X\cap S. \tag{24}$$

If $z^*\in X^*$, then (24) holds and we have

$$\Phi_c(z^*)\leq F(z^*)\leq F(x_f(y,c))\leq F(x_f(y,c))+(1/c)D_f(x_f(y,c),y) \\ =\Phi_c(y), \qquad \forall y\in S, \tag{25}$$

because always $D_f(x,y)\geq 0$. Thus, z^* minimizes $\Phi_c(y)$ over S, i.e., $z^*\in S^*$. Conversely, if $z^*\in S^*$, then we have, from (14),

$$c\nabla\Phi_c(z^*)=\nabla^2 f(z^*)^T[z^*-x_f(z^*,c)]=0, \tag{26}$$

which implies that $z^*=x_f(z^*,c)\in X\cap S$. Using again (24), we have

$$F(z^*)=\Phi_c(z^*)\leq\Phi_c(y)\leq F(y), \qquad \forall y\in X\cap S, \tag{27}$$

and therefore $z^*\in X\cap S$. □

The following D_f-function version of a classical lemma (Ref. 23, p. 76) about projection operators plays an important role in the proof of convergence.

Proposition 3.5. Let f be an auxiliary function with zone S, and let $\Omega\subseteq\mathbb{R}^n$ be some given closed convex set. Denote by Py the D_f-projection of any $y\in S$ onto Ω, and assume that $Py\in S$. Let $z\in\Omega\cap\bar{S}$. Then, for any $y\in S$, the inequality

$$D_f(Py,y)\leq D_f(z,y)-D_f(z,Py) \tag{28}$$

holds.

Proof. This is a specialization of Ref. 5, Lemma 1, for the case of D_f-functions; see Ref. 5, p. 206, Example 2. □

Finally, we present the convergence proof of the PMD algorithm.

Proposition 3.6. Assume that $X^*\cap\bar{S}\neq\emptyset$. Any sequence $\{x(t)\}$ generated by a PMD algorithm, where $c(t)>0$ and $\lim_{t\to\infty}\inf c(t)=c>0$, converges to an element of X^*.

Proof. The proof consists of three steps. First, we prove that $\{x(t)\}$ is bounded; then, we show that all its accumulation points belong to X^*; and finally, we prove that there is a unique limit point.

Step 1. We have

$$F(x(t+1))+[1/c(t)]D_f(x(t+1), x(t)) \leq F(x)+[1/c(t)]D_f(x, x(t)), \qquad \forall x \in X \cap \bar{S}, \tag{29}$$

from which follows that, for all $x \in X \cap \bar{S}$ with

$$F(x) \leq F(x(t+1)), \tag{30}$$

it is true that

$$D_f(x(t+1), x(t)) \leq D_f(x, x(t)). \tag{31}$$

Therefore, $x(t+1)$ is the unique D_f-projection of $x(t)$ onto the convex set

$$\Omega = \{x \in X \mid F(x) \leq F(x(t+1))\}. \tag{32}$$

Using Proposition 3.5 and the fact that $X^* \subseteq \Omega$, we have

$$0 \leq D_f(x(t+1), x(t)) \leq D_f(x^*, x(t)) - D_f(x^*, x(t+1)), \tag{33}$$

for every $x^* \in X^* \cap \bar{S}$. Thus,

$$D_f(x^*, x(t+1)) \leq D_f(x^*, x(t)), \qquad \forall x^* \in X^* \cap \bar{S}, \quad \forall t. \tag{34}$$

This last inequality amounts to saying that $\{x(t)\}$ is D_f-Fèjer-monotone with respect to the set $X^* \cap \bar{S}$, and it implies that $\{x(t)\}$ is bounded because it means that

$$x(t) \in L_f^2(x^*, \alpha), \qquad \forall t, \tag{35}$$

with $\alpha = D_f(x^*, x(0))$, and Assumption A1 applies.

Step 2. Let $\{x(t)\}_{t \in T}$, $T \subseteq \mathbb{N}_0$, be a subsequence converging to $x^\infty \in X \cap \bar{S}$. Recall that, by (7) and (34), the sequence $\{D_f(x^*, x(t))\}_{t \in \mathbb{N}_0}$ is nonnegative and nonincreasing; thus, $\lim D_f(x^*, x(t))$ exists for any $x^* \in X^* \cap \bar{S}$. In view of (33),

$$\lim D_f(x(t+1), x(t)) = 0, \qquad \text{as } t \to \infty;$$

thus, also

$$\lim_{\substack{t \to \infty \\ t \in T}} D_f(x(t+1), x(t)) = 0, \tag{36}$$

which by Assumption A3 implies that $\{x(t+1)\}_{t \in T}$ also converges to x^∞.

Next, observe that (29) remains true with $x = x(t)$; thus,

$$F(x(t+1))+[1/c(t)]D_f(x(t+1), x(t)) \leq F(x(t)), \qquad \forall t, \tag{37}$$

because

$$D_f(x(t), x(t)) = 0.$$

This implies

$$F(x(t)) - F(x(t+1))$$
$$\geq [1/c(t)] D_f(x(t+1), x(t)) \geq 0, \qquad \forall t. \tag{38}$$

Therefore, $\{F(x(t))\}_{t\in\mathbb{N}_0}$ is nonincreasing and $\{F(x(t))\}_{t\in T}$ converges to $F(x^\infty)$.

Let $x^* \in X^* \cap \bar{S}$ and $\alpha \in (0, 1)$, and set

$$x = \alpha x^* + (1-\alpha)x(t+1)$$

in (29). From the convexity of $F(x)$, we get

$$F(x(t+1)) + [1/c(t)] D_f(x(t+1), x(t))$$
$$\leq F(\alpha x^* + (1-\alpha)x(t+1))$$
$$+ [1/c(t)] D_f(\alpha x^* + (1-\alpha)x(t+1), x(t))$$
$$\leq \alpha F(x^*) + (1-\alpha)F(x(t+1))$$
$$+ [1/c(t)] D_f(\alpha x^* + (1-\alpha)x(t+1), x(t)), \tag{39}$$

which, by (5), can be rewritten as

$$\alpha c(t)[F(x(t+1)) - F(x^*)]$$
$$\leq f(x(t+1) - \alpha(x^* - x(t+1))) - f(x(t+1))$$
$$- \alpha\langle \nabla f(x(t)), x^* - x(t+1)\rangle. \tag{40}$$

Dividing by α taking the limit as $\alpha \to 0^+$, and denoting by $f'(\cdot;\cdot)$ the directional derivative, we get

$$c(t)[F(x(t+1)) - F(x^*)]$$
$$\leq f'(x(t+1); x^* - x(t+1)) - \langle \nabla f(x(t)), x^* - x(t+1)\rangle$$
$$= \langle \nabla f(x(t+1)) - \nabla f(x(t)), x^* - x(t+1)\rangle$$
$$= D_f(x^*, x(t)) - D_f(x^*, x(t+1)) - D_f(x(t+1), x(t)). \tag{41}$$

The optimality of x^* and the fact that $c(t) \geq c > 0$, $\forall t$, guarantee the nonnegativity of the left-hand side of (41) for all t.

From the existence of $\lim_{t\to\infty} D_f(x^*, x(t))$, for any $x^* \in X^* \cap \bar{S}$, we obtain

$$\lim_{\substack{t\to\infty \\ t\in T}} [D_f(x^*, x(t)) - D_f(x^*, x(t+1))] = 0, \tag{42}$$

and the remaining term in the right-hand side of (41) also tends to zero by (36). Thus,

$$0 = \lim_{\substack{t \to \infty \\ t \in T}} [F(x(t+1)) - F(x^*)] = F(x^\infty) - F(x^*), \tag{43}$$

and so $x^\infty \in X^*$.

Step 3. Let

$$\{x(t)\}_{t \in T_1} \xrightarrow[t \to \infty]{} x^* \in X^* \cap \bar{S}, \tag{44}$$

$$\{x(t)\}_{t \in T_2} \xrightarrow[t \to \infty]{} x^{**} \in X^* \cap \bar{S}, \tag{45}$$

for some $T_1 \subseteq \mathbb{N}_0$ and $T_2 \subseteq \mathbb{N}_0$. Defining

$$H(x(t)) = D_f(x^*, x(t)) - D_f(x^{**}, x(t)), \tag{46}$$

it follows that the limit

$$\lim_{t \to \infty} H(x(t)) = H \tag{47}$$

exists. Therefore, we have, by (7),

$$H = \lim_{\substack{t \to \infty \\ t \in T_1}} H(x(t)) = -D_f(x^{**}, x^*) \leq 0, \tag{48}$$

$$H = \lim_{\substack{t \to \infty \\ t \in T_2}} H(x(t)) = D_f(x^*, x^{**}) \geq 0, \tag{49}$$

yielding

$$D_f(x^*, x^{**}) = D_f(x^{**}, x^*) = 0, \tag{50}$$

which implies that $x^* = x^{**}$. □

4. Quadratic Case, Entropy Case, and Other Cases

Choosing the auxiliary function $f(x) = (1/2)\|x\|^2$ with $\Lambda = S = \bar{S} = \mathbb{R}^n$ gives $D_f(x, y) = (1/2)\|x - y\|^2$ and immediately returns the PMD algorithm to its original form with quadratic additive terms; see Refs. 3–4.

Define the $x \log x$ (Shannon) entropy functional ent x, which maps $\mathbb{R}^n_+$ into $\mathbb{R}$ by

$$\text{ent}\, x = -\sum_{j=1}^{n} x_j \log(x_j/a_j), \tag{51}$$

where $a = (a_j) \in \mathbb{R}^n$ is a given positive vector, log is the natural logarithm, and $0 \log 0 = 0$ by convention.

The function $-\text{ent}\, x$ is a Bregman function, in the terminology of Ref. 6, with $\Lambda = \mathbb{R}^n_+$ and zone $S = \{x \in \mathbb{R}^n \mid x > 0\}$; see, e.g., Ref. 9, Lemma 5. Furthermore, it can be verified that Assumptions A1–A5 hold, and thus the analysis of the PMD algorithm presented here applies.

For $f(x) = -\text{ent}\, x$, we have

$$D_f(x, y) = \sum_{j=1}^{n} x_j[\log(x_j/y_j) - 1] + \sum_{j=1}^{n} y_j, \tag{52}$$

and the iterative step of the entropy-type PMD algorithm is obtained from (9).

In the case where problem (1) is linear programming, $F(x) = \langle b, x\rangle$ for some given $b = (b_j) \in \mathbb{R}^n$, and $x \in X$ are linear constraints, (9) becomes

$$x(t+1) = \arg\min_{x \in X \cap \mathbb{R}^n_+} \left\{ \sum_{j=1}^{n} x_j b_j + [1/c(t)] \sum_{j=1}^{n} x_j [\log(x_j/x_j(t)) - 1] + [1/c(t)] \sum_{j=1}^{n} x_j(t) \right\}. \tag{53}$$

This is essentially a linearly constrained pure entropy optimization problem obtained by subsuming all linear terms into the entropy functional. For such problems, several useful iterative algorithms exist (see Ref. 9), which lend themselves efficiently to parallel computation; see, e.g., Refs. 24–25. The advantages, in practice, of an entropy or otherwise oriented PMD algorithm over the original quadratic-additive-term proximal minimization algorithm, if any, depend to a great extent on two factors. One is the specific form of the original problem (1); the second is the availability of efficient special-purpose algorithms for performing the step (9). Various special-purpose algorithms exist (e.g., Refs. 26–27); others may be developed.

Such practical questions may be settled in further experimental research which is outside the scope of the present paper. See, e.g., Refs. 28–29 for recent experimental work with the algorithms of this paper. There seems to be also room for further study in the identification of additional specific usable and useful auxiliary functions; see, e.g., the characterization of

Bregman functions in Ref. 7 and the results in Ref. 22. See also Ref. 15, Examples 3.1.

References

1. MINTY, G. J., *Monotone (Nonlinear) Operators in Hilbert Space*, Duke Mathematics Journal, Vol. 29, pp. 341–346, 1962.
2. MOREAU, J. J., *Proximité et Dualité dans un Espace Hilbertien*, Bulletin de la Société Mathématique de France, Vol. 93, pp. 273–299, 1965.
3. ROCKAFELLAR, R. T., *Monotone Operators and the Proximal Point Algorithm*, SIAM Journal on Control and Optimization, Vol. 14, pp. 877–898, 1976.
4. ROCKAFELLAR, R. T., *Augmented Lagrangians and Applications of the Proximal Point Algorithm in Convex Programming*, Mathematics of Operations Research, Vol. 1, pp. 97–116, 1976.
5. BREGMAN, L. M., *The Relaxation Method of Finding the Common Point of Convex Sets and Its Application to the Solution of Problems in Convex Programming*, USSR Computational Mathematics and Mathematical Physics, Vol. 7, pp. 200–217, 1967.
6. CENSOR, Y., and LENT, A., *An Iterative Row-Action Method for Interval Convex Programming*, Journal of Optimization Theory and Applications, Vol. 34, pp. 321–353, 1981.
7. DE PIERRO, A. R., and IUSEM, A. N., *A Relaxed Version of Bregman's Method for Convex Programming*, Journal of Optimization Theory and Applications, Vol. 51, pp. 421–440, 1986.
8. CENSOR, Y., and SEGMAN, J., *On Block-Iterative Entropy Maximization*, Journal of Information and Optimization Sciences, Vol. 8, pp. 275–291, 1987.
9. CENSOR, Y., DE PIERRO, A. R., ELFVING, T., HERMAN, G. T., and IUSEM, A. N., *On Iterative Methods for Linearly Constrained Entropy Maximization*, Numerical Analysis and Mathematical Modelling, Edited by A. Wakulicz, Banach Center Publications, PWN-Polish Scientific Publishers, Warsaw, Poland, Vol. 24, pp. 145–163, 1990.
10. LENT, A., and CENSOR, Y., *The Primal-Dual Algorithm as a Constraint-Set Manipulation Device*, Mathematical Programming, Vol. 50, pp. 343–357, 1991.
11. ERIKSSON, J. R., *An Iterative Primal-Dual Algorithm for Linear Programming*, Report LITH-MAT-R-1985-10, Department of Mathematics, Linköping University, Linköping, Sweden.
12. BERTSEKAS, D. P., and TSITSIKLIS, J. N., *Parallel and Distributed Computation: Numerical Methods*, Prentice-Hall, Englewood Cliffs, New Jersey, 1989.
13. BERTSEKAS, D. P., *Constrained Optimization and Lagrange Multiplier Methods*, Academic Press, New York, New York, 1982.
14. BERTSEKAS, D. P., Private Communication, March 1990.
15. TEBOULLE, M., *Entropic Proximal Mappings with Applications to Nonlinear Programming*, Mathematics of Operations Research (to appear), also available as Technical Report, Department of Mathematics and Statistics, University of Maryland, Baltimore, Maryland, 1990.

16. CHEN, G., and TEBOULLE, M., *Convergence Analysis of Proximal-Like Minimization Algorithm Using Bregman Functions*, Research Report 90-23, Department of Mathematics, University of Maryland at Baltimore County, Catonsville, Maryland, 1990.
17. ECKSTEIN, J., *Splitting Methods for Monotone Operators with Applications to Parallel Optimization*, Technical Report CICS-TH-140, Center for Intelligent Control Systems, Massachusetts Institute of Technology, Cambridge, Massachusetts, 1989.
18. ECKSTEIN, J., *Nonlinear Proximal Point Algorithms Using Bregman Functions, with Applications to Convex Programming*, Mathematics of Operations Research (to appear).
19. EGGERMONT, P. P. B., *Multiplicative Iterative Algorithms for Convex Programming*, Linear Algebra and Its Applications, Vol. 130, pp. 25–42, 1990.
20. SHEPP, L. A., and VARDI, Y., *Maximum Likelihood Reconstruction in Emission Tomography*, IEEE Transactions on Medical Imaging, Vol. MI-1, pp. 113–122, 1982.
21. TSENG, P., and BERTSEKAS, D. P., *On the Convergence of the Exponential Multiplier Method for Convex Programming*, Technical Report LIDS-P-1995, Laboratory for Information and Decision Systems, Massachusetts Institute of Technology, Cambridge, Massachusetts, 1990.
22. CENSOR, Y., DE PIERRO, A. R., and IUSEM, A. N., *Optimization of Burg's Entropy over Linear Constraints*, Applied Numerical Mathematics, Vol. 7, pp. 151–165, 1991.
23. AUSLENDER, A., *Optimisation: Méthodes Numériques*, Masson, Paris, France, 1976.
24. ZENIOS, S. A., and CENSOR, Y., *Parallel Computing with Block-Iterative Image Reconstruction Algorithms*, Applied Numerical Mathematics, Vol. 7, pp. 399–415, 1991.
25. ZENIOS, S. A., and CENSOR, Y., *Massively Parallel Row-Action Algorithms for Some Nonlinear Transportation Problems*, SIAM Journal on Optimization, Vol. 1, pp. 373–400, 1991.
26. POWELL, M. J. D., *An Algorithm for Maximizing Entropy Subject to Simple Bounds*, Mathematical Programming, Vol. 42, pp. 171–180, 1988.
27. FANG, S. C., and RAJASEKERA, R. J., *Quadratically Constrained Minimum Cross-Entropy Analysis*, Mathematical Programming, Vol. 44, pp. 85–96, 1989.
28. NIELSEN, S., and ZENIOS, S. A., *Proximal Minimizations with D-Functions and the Massively Parallel Solutions of Linear Network Programs*, Report 91-06-05, Decision Sciences Department, The Wharton School, University of Pennsylvania, Philadelphia, 1991.
29. NIELSEN, S., and ZENIOS, S. A., *Solving Linear Stochastic Network Problems Using Massively Parallel Proximal Algorithms*, Report 92-01-05, Decision Sciences Department, The Wharton School, University of Pennsylvania, Philadelphia, 1992.