A Modified Bregman Proximal Scheme To Minimize The Difference Of Two Convex Functions*

Abdelouahed Hamdi†

Abstract

We introduce a family of new transforms based on the imitation of the Bregman proximal mapping to minimize the difference of two convex functions. It can be seen as a descent scheme which takes into consideration the convex properties of the two convex functions separately. A direct application of the proposed scheme to variational inclusion is given.

1 Preliminaries

In non-convex programming, the fundamental property of convex problems that local solutions are global ones is no longer true. Therefore, methods using only local information are insufficient to locate global minima. Thus, optimality conditions for non-convex optimization problems have to take into account the form and the structure of the model. Here in this work, we are interested in a certain class of models called d.c. problems. These problems deal with a minimization or maximization of the difference of two or more convex functions. It is well known that with two convex functions g and h the sum $g + h$ is again a convex function, as is the maximum $\max\{g, h\}$ and the multiple λg for any positive λ. The difference $g - h$, however, is not a convex function. This is why d.c. problems are difficult as they are non-convex problems (see [11, 8, 4] and references therein).

In this work, we are presenting a modified Bregman proximal scheme to find out the minimum of the difference of two convex functions, using the convexity of both functions involved in the d.c model. The presented scheme is a type of descent method.

Let E be a finite-dimensional vector space and let $\langle \cdot, \cdot \rangle$ denote the inner product in E. Let $\Theta(E)$ denote the set of convex proper and lower semi-continuous functions on E. Let f be a d.c function on E, that means that there exist two functions g and h both in $\Theta(E)$ such that $f(x) = g(x) - h(x), \ \forall \ x \in E$. Moreover, suppose that $h(x) \neq +\infty$ for all $x \in E$ and that $\mathrm{ridom}(g) \cap \mathrm{ridom}(h) \neq \emptyset$.

In this paper we are concerned with minimizing a d.c. function f on E:

$$\min_{x \in E} f(x). \tag{1}$$

The next theorem gathers several global optimality conditions from d.c. problems given in the literature [1]

THEOREM 1. Let $g, h : E \to \Re$ be lower semi-continuous proper convex functions and let $x^* \in \operatorname{dom} g \cap \operatorname{dom} h$. Then the following conditions are equivalent[2] :

(G) x^* is a global minimizer of $g - h$ on E,

(FB) $\partial^\gamma h(x^*) \subset \partial^\gamma g(x^*)$,

(HU) $\partial_\epsilon h(x^*) \subset \partial_\epsilon g(x^*), \quad \forall\, \epsilon \geq 0$,

(ST) $g(x^*) - h(x^*) = \inf_{z \in E^*} \{h^*(z) - g^*(z)\}$,

(HPT) $\max \{h(x) - \tau : g(x) \leq \sigma,\ \sigma - \tau \leq g(x^*) - h(x^*),\ x \in E,\ \sigma, \tau \in \Re\} = 0$,

with

$$\partial^\gamma f(x^*) = \{\phi \in \Omega : f(x) \geq f(x^*) + \phi(x) - \phi(x^*),\ \forall x \in E\},\ x^* \in \operatorname{dom}(f),$$

where Ω denotes the space of continuous, real-valued functions of E and with

$$z \in \partial_\epsilon f(x^*) \Leftrightarrow f^*(z) + f(x^*) - \langle x^*, z \rangle \leq \epsilon.$$

For the proof see [6] Theorem 1 and/or [4] Theorem 2.3.1 pp. 8-9.

As usual in optimization, the necessary optimality condition consists, in general of the variational problem

$$\text{Find } x^* \in E \ \text{ such that } \ 0 \in \partial f(x^*) = \partial\, (g - h)\, (x^*). \tag{2}$$

According to the assumption $\operatorname{ridom}(g) \cap \operatorname{ridom}(h) \neq \emptyset$, the above equation can be rewritten as

$$\text{Find } x^* \in E \ \text{ such that } \ \partial h(x^*) \subset \partial g(x^*). \tag{3}$$

The condition (3) is not a simple sub-differential inclusion in general, and this why we may content ourselves by solving the relaxed variational problem

$$\text{Find } x^* \in E \ \text{ such that } \ \partial h(x^*) \cap \partial g(x^*) \neq \emptyset. \tag{4}$$

---

[1]This theorem is taken from the PhD dissertation of Dr. Mirjam Dür ([4], pp. 8-9). For more references, see [8].

[2]The authors of these conditions are Flores-Bazan (FB), Hirriart-Urruty (HU), Singer and Toland (ST) and Horst, Pardalos and Thoai (HPT) see [4].

2   Bregman Functions

Before describing the modified Bregman regularization scheme, we need to recall briefly what we mean by a Bregman function. To this end let $S \subset E$ be a closed and convex set and let S^o be its interior. Consider $\phi : S \to \Re$ a continuously differentiable function on S^o. For $w \in S$ and $z \in S^o$, define $D_\phi : S \times S^o \to \Re$ by

$$D_\phi(w, z) = \phi(w) - \phi(z) - \langle \nabla \phi(z), w - z \rangle. \tag{5}$$

D_ϕ is said to be the Bregman distance associated to ϕ [2] and it was developed and used in the proximal theory by ([2, 3, 5]). We say that ϕ is a *Bregman function with zone* S^o if the following conditions hold:

B1. ϕ is strictly convex on S.

B2. The level sets of $D_\phi(w, \cdot)$ and $D_\phi(\cdot, z)$ are bounded, for all $w \in S$, $z \in S^o$, respectively.

B3. If $\{z^k\} \subset S^o$ converges to z^*, then $\lim_{k \to \infty} D_\phi(z^*, z^k) = 0$.

These special functions have been widely used in convex optimization, e.g. [5, 2] and references therein.

REMARK 1   *1. $S = \Re^n$, $\phi_1(x) = \frac{1}{2}\|x\|^2$, then $D_{\phi_1}(x, y) = \frac{1}{2}\|x - y\|^2$*

   2. $S = \Re^n_{++}$, $\phi_2(x) = \sum_{i=1}^n x_i \log(x_i) - x_i$, then for $(x, y) \in \bar{\Re}^n \times \Re^n_*$ and with $0 \log 0 = 0$. $D_{\phi_2}(x, y) = \sum_{i=1}^n x_i \log(x_i/y_i) + y_i - x_i$, D_{ϕ_2} is called the Kullback-Leibler *relative entropy distance.*

   3. $S =]-1, 1[^n$, $\phi_3(x) = -\sum_{i=1}^n \sqrt{1 - x_i^2}$, then $D_{\phi_3}(x, y) = \sum_{i=1}^n \frac{1 - x_i y_i}{\sqrt{1 - y_i^2}} - \sum_{i=1}^n \sqrt{1 - x_i^2}$.

   4. $S = \Re^n_{++}$, $\phi_4(x) = \sum_{i=1}^n x_i - \sqrt{x_i}$, then $D_{\phi_4}(x, y) = \sum_{i=1}^n \frac{1}{2\sqrt{y_i}}(\sqrt{x_i} - \sqrt{y_i})^2$.

   5. $S = \Re^n$, $\phi_5(x) = \frac{1}{2}\langle x, Ax \rangle$, then $D_{\phi_5}(x, y) = \frac{1}{2}\|x - y\|_A^2 = \frac{1}{2}\langle x - y, A(x - y)\rangle$ where $A \in \Re^{n \times n}$ is symmetric positive definite.

   6. Let $\phi : \Re^n \to \Re$ be a strictly convex function such that:

$$\phi \in C^2(\Re^n) \quad and \quad \lim_{\|x\| \to \infty} \frac{\phi(x)}{\|x\|} = +\infty,$$

   then ϕ is a Bregman function.

   7. If ϕ is a Bregman function, then $\phi(x) + c^\top x + d$ for any $c \in \Re^n$, $d \in \Re$, also is a Bregman function.

3 Bregman Proximal Scheme

In this section, we propose the construction of a scheme to approximate the critical point x^* of f satisfying (4). To this end, we use the nice and useful property of convexity of the functions g and h. The idea can be stated as follows. Starting by an initial point $x_{old} \in S$, we can select a point $z \in \partial h(x_{old})$ and using a proximal scheme based on a Bregman function ϕ, we construct a translated fixed point iterative scheme by setting $x_{new} = (\nabla\phi + \lambda\partial g)^{-1}(\nabla\phi(x_{old}) + \lambda z)$ yielding a fixed point $\tilde{x}$ that coincides with a critical point of f. These steps are possible under the convexity of both functions f and g. Now, we state the algorithm and then we show the well-definedness of all the steps.

1. **Choose** $x^0 \in S^o$ and $\lambda > 0$. Set $t = 0$

2. **Evaluate** $z^t \in \partial h(x^t)$

3. **Evaluate**

$$x^{t+1} = (\nabla\phi + \lambda\partial g)^{-1}\left(\nabla\phi(x^t) + \lambda z^t\right). \tag{6}$$

4. **If** $x^{t+1} = x^t$. Stop , the solution is x^t. **Else** set $t = t + 1$ and go back to step 2.

4 Convergence Analysis

We will focus on the well-definedness and the convergence of the sequence $\{x^t\}_t$.

PROPOSITION 1. ([10]) Let $h \in \Theta(E)$, then $\partial h(x) \neq \emptyset$ for all $x \in \text{ridom}(h)$.

To show the well-definedness of the iterative scheme (6), we need the following results and these conditions:

B4. (Boundary coerciveness [2]). If $\{y^t\} \subset S^o$ is such that $\lim_{t\to\infty} y^t = y \in \partial S^o$, then $\lim_{t\to\infty}\langle\nabla\phi(y^t), x - y\rangle = -\infty$ for all $x \in S^o$.

B5. $Im(\nabla\phi) = \Re^n$.

LEMMA 1. ([3]) Let ϕ a Bregman function with zone S^o. If ϕ satisfies B4, then $\text{dom}(\nabla\phi) = S^o$.

LEMMA 2. Let T a monotone operator on E, and ϕ a Bregman function with zone S^o. Then the mapping $(\nabla\phi + T)^{-1} o\nabla\phi$ is at most single-valued.

PROOF[3]. From the strict convexity of ϕ, $\nabla\phi$ is a strictly monotone operator. It follows that $\nabla\phi + T$ is strictly monotone, and therefore that $(\nabla\phi + T)(x)$ and $(\nabla\phi + T)(y)$ do not intersect for any $x \neq y$. Thus, $(\nabla\phi + T)^{-1}$ is at most single-valued, and as $\nabla\phi$ is at most single-valued, $(\nabla\phi + T)^{-1} o\nabla\phi$ is at most single-valued.

PROPOSITION 2. Let $T = \partial g$ be a maximal monotone operator on $\Re^n$, and let ϕ be a Bregman function with zone $S^o \supset Dom(T)$. If B5 holds, then for all $x^0 \in Dom(\nabla\phi)$, an infinite sequence $\{x^t\}_t$ conforming to the recursion (6) exists.

---

[3]This proof can be found in [5]

PROOF. For a complete proof see the seminal paper of Eckstein ([5] Theorem 4 page 209).

Since the set S will appear in the algorithm only through the fact that its interior is the zone of ϕ, so that S should be recoverable from ϕ. It is clear that if ϕ is a Bregman function of zone S^o and U is an open convex subset of S^o, then ϕ is Bregman function of zone U. This implies that we can not recover S^o from ϕ. On the other hand, we want to use D_ϕ for penalization purposes on the closed set S. The information about the set S in our algorithm will be encapsulated in D_ϕ, so that S will have to be recoverable from ϕ. The conditions B4 and B5 fit this situation because the divergence of $\nabla\phi$ at the border of S^o makes it univocally determined by ϕ. Furthermore, the condition B4 is equivalent to Rockafellar's concept of essential smoothness ([10], page 251) used by Censor *et al.* [3] to obtain the interior point characteristic[4] of their iterative proximal scheme. Thus, this last characteristic is obtained in our case through the above conditions.

PROPOSITION 3. If ϕ is a Bregman function with zone $S^o \supset \mathrm{dom}(\partial g)$, satisfying conditions B4-B5, then the sequence $\{x^t\}_t$ generated in (6) is well-defined and x^{t+1} is in S.

PROOF. The iterative scheme $x^{t+1} = (\nabla\phi + \lambda\partial g)^{-1} (\nabla\phi(x^t) + \lambda z^t)$, can be seen as the following variational problem

$$0 \in \partial g(x^{t+1}) - z^t + \frac{1}{\lambda} \left(\nabla\phi(x^{t+1}) - \nabla\phi(x^t) \right),$$

which implies $x^{t+1} := arg\min_x \left\{ g(x) - \langle z^t, x \rangle + \frac{1}{\lambda} D_\phi(x, x^t) \right\}$. Finally, according to the strict convexity of ϕ and the convexity of g, we conclude the existence of x^{t+1} which complete the proof.

PROPOSITION 4. Assume that ϕ is a Bregman function with zone $S^o \supset \mathrm{dom}(\partial g)$, satisfying conditions B4-B5, A vector $x \in S^o$ satisfies the necessary optimality condition (4) if and only if $x = (\nabla\phi + \lambda\partial g)^{-1} (\nabla\phi(x) + \lambda z)$, $\forall \lambda > 0$, $z \in \partial h(x)$.

PROOF. Let $x \in S^o$, $z \in \partial h(x)$ and λ a positive parameter. According to propositions 2-3, $x = (\nabla\phi + \lambda\partial g)^{-1} (\nabla\phi(x) + \lambda z)$ is well-defined and it gives

$$\nabla\phi(x) + \lambda z \in (\nabla\phi + \lambda\partial g)(x).$$

By simplifying the $\nabla\phi(x)$ and dividing both sides by $\lambda > 0$, we obtain that $z \in \partial g(x)$ and this implies that $z \in \partial h(x) \cap \partial g(x)$ which means that x is a critical point of f.

Conversely, let $0 \in \partial f(x)$ and let $\lambda > 0$, then there exists a vector $z \in \partial h(x) \cap \partial g(x)$. Thus,

$$
\begin{aligned}
z \in \partial g(x) \implies{} & \lambda z \in \lambda\partial g(x) \\
\implies{} & \nabla\phi(x) + \lambda z \in \nabla\phi(x) + \lambda\partial g(x) \\
\implies{} & (\nabla\phi(x) + \lambda z) \in (\nabla\phi + \lambda\partial g)(x) \\
\implies{} & x = (\nabla\phi + \lambda\partial g)^{-1} (\nabla\phi(x) + \lambda z),
\end{aligned}
$$

---

[4]See the paper of Yair Censor *et al.* ([3], 1995) to go deeply into this part. The author would like to thank the anonymous referee for pointing out the necessity of this remark to complete the analysis.

and since $z \in \partial h(x)$, the proof is complete.

PROPOSITION 5. The sequence $\{x^t\}_t$ generated in the algorithm by $x^{t+1} = (\nabla\phi + \lambda\partial g)^{-1} (\nabla\phi(x^t) + \lambda z^t)$ converges to a critical point of the objective f.

PROOF. Assume that the convergence is not reached yet, then from

$$x^{t+1} = (\nabla\phi + \lambda\partial g)^{-1} (\nabla\phi(x^t) + \lambda z^t),$$

we get

$$\nabla\phi(x^t) + \lambda z^t \in \nabla\phi(x^{t+1}) + \lambda\partial g(x^{t+1}),$$

which can be rewritten obviously in the following form

$$\frac{\nabla\phi(x^t) - \nabla\phi(x^{t+1})}{\lambda} + z^k \in \partial g(x^{t+1}).$$

According to the convexity of g and the definition of sub-differentials, we obtain the following inequality:

$$g(x) \geq g(x^{t+1}) + \left\langle x - x^{t+1}, \frac{\nabla\phi(x^t) - \nabla\phi(x^{t+1})}{\lambda} + z^t \right\rangle, \quad \forall\, x. \tag{7}$$

Particularly for $x = x^t$, we have

$$g(x^t) \geq g(x^{t+1}) + \left\langle x^t - x^{t+1}, \frac{\nabla\phi(x^t) - \nabla\phi(x^{t+1})}{\lambda} + z^t \right\rangle. \tag{8}$$

Also, according to the convexity of h and the fact that $z^t \in \partial h(x^t)$, then

$$h(x) \geq h(x^t) + \left\langle x - x^t, z^t \right\rangle, \quad \forall\, x, \tag{9}$$

and for $x = x^{t+1}$, we get

$$h(x^{t+1}) \geq h(x^t) + \langle x^{t+1} - x^t, z^t \rangle. \tag{10}$$

Inequality (8) can be rewritten as

$$g(x^{t+1}) \leq g(x^t) - \frac{1}{\lambda} \left\langle x^t - x^{t+1}, \nabla\phi(x^t) - \nabla\phi(x^{t+1}) \right\rangle - \langle x^t - x^{t+1}, z^t \rangle, \tag{11}$$

and (10) becomes

$$-h(x^{t+1}) \leq -h(x^t) + \langle x^t - x^{t+1}, z^t \rangle. \tag{12}$$

By adding (11) and (12), we get $f(x^{t+1}) \leq f(x^t) - \frac{1}{\lambda} \langle x^t - x^{t+1}, \nabla\phi(x^t) - \nabla\phi(x^{t+1}) \rangle$. Finally, according to the convexity of ϕ, we obtain $f(x^{t+1}) \leq f(x^t)$ which proves that the considered scheme is a descent one.

We finish this section by the following theorem.

THEOREM 2. Every accumulation point of the sequence $\{x^t\}_t$ generated by the modified Bregman regularization step is a critical point of f.

PROOF. This is obvious using Propositions 4 and 5 after proving that the sequence $\{x^t\}_t$ is bounded. To this goal, let us consider the iteration (6), which is equivalent to the variational inclusion $\frac{1}{\lambda}\left(\nabla\phi(x^t) - \nabla\phi(x^{t+1})\right) + z^t \in \partial g(x^{t+1})$. We denote x^* an optimal solution to (4) and let $z^* \in \partial h(x^*) \cap \partial g(x^*)$, thus, z^* belongs to $\partial g(x^*)$. Using the monotonicity of the operator ∂g, we obtain $\left\langle \frac{1}{\lambda}\left(\nabla\phi(x^t) - \nabla\phi(x^{t+1})\right), x^{t+1} - x^*\right\rangle \geq \langle z^t - z^*, x^* - x^{t+1}\rangle$ i.e.,

$$\left\langle \frac{1}{\lambda}\left(\nabla\phi(x^t) - \nabla\phi(x^{t+1})\right), x^* - x^{t+1}\right\rangle \leq \langle z^t - z^*, x^{t+1} - x^*\rangle.$$

It is sufficient to see that at optimality z^t coincides with z^*, which implies that the left hand side of the above inequality is always less or equal than zero. i.e.,

$$\left\langle \frac{1}{\lambda}\left(\nabla\phi(x^t) - \nabla\phi(x^{t+1})\right), x^* - x^{t+1}\right\rangle \leq 0.$$

Now, according to the definition of the Bregman distance D_ϕ, and after direct calculations we obtain

$$\left\langle \left(\nabla\phi(x^t) - \nabla\phi(x^{t+1})\right), x^* - x^{t+1}\right\rangle = D_\phi(x^*, x^{t+1}) - D_\phi(x^*, x^t) + D_\phi(x^t, x^{t+1}) \leq 0,$$

and since $D_\phi(x^*, x^{t+1}) \geq 0$, we have $D_\phi(x^*, x^{t+1}) \leq D_\phi(x^*, x^t)$, which implies that $D_\phi(x^*, x^{t+1})$ is bounded above by $D_\phi(x^*, x^0)$ and thus, the term x^{t+1} belongs to the level set of $D_\phi(x^*, .)$. This set is bounded by the Bregman assumption B2, so the sequence $\{x^t\}_t$ is bounded and the proof is complete. We finish the paper by a direct application of the proposed algorithm to solve a variational inequality problem.

5   Application: Variational Inequality Problems

We consider in this section the following variational inequality problem (VIP(F,C) for short).

$$\text{Find } x^* \in C, \ y^* \in F(x^*) \ \text{ such that } \ \langle y^*, x - x^*\rangle \geq 0, \quad \forall\, x \in C, \qquad (13)$$

where $F : H \to \mathcal{P}(H)$ is a multi-valued mapping, C is a closed convex set of H and H is a real Hilbert space whose inner product and norm are denoted $<.,.>$ and $\|.\|$ respectively.

This problem has many important applications e.g., in economics, operations research, industry, physics, the obstacle problem and engineering sciences. Many research papers have been written lately both on the theory and applications of this field. Important connections with main areas of pure and applied sciences have been made, see for example [12] and the seminal surveys of Harker and Pang [7] and Noor [9]. Specially in this application, we are interested in the case where the mapping $F = -\partial h$ where h is a convex function. Thus, (13) becomes

$$\text{Find } x^* \in C, \ y^* \in -\partial h(x^*) \ \text{ such that } \ \langle y^*, x - x^*\rangle \geq 0, \quad \forall\, x \in C. \qquad (14)$$

Using the indicator function ϕ_C of the set C defined by $\phi_C(x) = 0$ if $x \in C$ and ∞ if $x \notin C$, it is easy to see that (14) is equivalent to finding[5]

$$y^* \in -\partial h(x^*) \bigcap -\partial \phi_C(x^*) \qquad (15)$$

i.e., if we set $w^* = -y^*$, then solving (13) is equivalent to find w^* such that

$$w^* \in \partial h(x^*) \bigcap \partial \phi_C(x^*). \qquad (16)$$

Finally our Algorithm can be applied to solve (16) in the following manner.

1. **Choose** $x^0 \in \Re^n$ and $\lambda > 0$. Set $t = 0$

2. **Evaluate** $w^t \in \partial h(x^t)$

3. **Evaluate** $x^{t+1} = (\nabla \phi + \lambda \partial \phi_C)^{-1} (\nabla \phi(x^t) + \lambda w^t)$

4. **If** $x^{t+1} = x^t$. Stop , the solution is x^t. **Else** set $t = t + 1$ and go back to step 2.

Furthermore, if the Bregman function $\phi(x) = 0.5\|x\|^2$, the above step 3 becomes $x^{t+1} = P_C(x^t + \lambda w^t)$, where P_C denotes the projection operator on C.

References

[1] H. Bresis, Opérateurs maximaux monotones, Mathematics Studies, N. 5, North-Holland, 1973.

[2] R. S. Burachik and A. N. Iusem , A Generalized proximal point Algorithm for the variational inequality problem in a Hilbert space, SIAM J. Opt., 8(1998), 197-216.

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

[4] M. Dür, Duality in global optimization: optimality conditions and algorithmical aspects, PhD Dissertation, Trier University, Trier, Germany 1999.

[5] J. E. Eckstein, Nonlinear proximal point algorithms using Bregman functions with applications to convex programming, Mathematics of O.R., 18(1)(1993), 203–226.

[6] A. Hamdi, A Moreau-Yosida regularization of a D-C Functions. Application to Variational Inequality Problems, Applied Mathematics E-Notes, 5(2005), 164–170.

---

[5]it is well known that $\partial \phi_C(x^*) = N_C(x^*) = \begin{cases} \{w \in Re^n : \langle w, x - x^* \rangle \leq 0, \ \forall x \in C\} \text{ if } x^* \in C \\ \emptyset, \ \text{otherwise.} \end{cases}$

[7] P. T. Harker and J. S. Pang, Finite-dimensional variational inequality and nonlinear complementarity problems: A Survey of Theory, Algorithms and Applications, Mathematical Programming, 48(1990), 161–220.

[8] J. B. Hiriart-Urruty, From Convex Optimization to Nonconvex Optimization, Necesary and Sufficient Conditions for Global Optimization, Nonsmooth Optimization and Related Topics, Plenum Press, 1989, 219–240.

[9] M. A. Noor, Some developments in general variational inequalities, Appl. Math. Computation, 152(1)(2004), 199–277.

[10] R. T. Rockafellar, Convex Analysis, Princeton University Press, 1970.

[11] H. Tuy, Global minimization of a difference of two convex functions, Mathematical Programming Study, 30(1987), 150–182.

[12] J. C. Yao, Multi-valued variational inequalities with K-pseudomonotone operators, J. Optimization Theory and Applications., 83(1994), 391–403.