Applied Mathematics E-Notes, 5(2005), 164-170 © ISSN 1607-2510
Available free at mirror sites of http://www.math.nthu.edu.tw/∼amen/

A MOREAU-YOSIDA REGULARIZATION OF A DIFFERENCE OF TWO CONVEX FUNCTIONS*

Abdelouahed Hamdi†

Received 26 September 2004

Abstract

We present a scheme to minimize a difference of two convex functions by solving a variational problem. The proposed scheme uses a proximal regularization step (see [8]) to construct a translated fixed point iteration. 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 algorithm to variational inclusion is given.

1 Preliminaries

In non-convex programming problems, the fundamental property of convex problems concerning the fact that local solutions are global ones is not true anymore. Therefore, methods using only local information are insufficient to locate global minima. Thus, optimality conditions for nonconvex optimization problems have to take into account the form and the structure of the model. Here in this work, we are interested on a certain class of models called d.c. problems. These problems deal with a minimization or maximization of a 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 any more. This is why d.c. problems are difficult as they are nonconvex problems.

In this work, we are presenting a regularization approach to find out the minimum using the convexity of the both functions involved in the d.c. model. The presented scheme is a type of descent method to locate the minimum.

Let E be a finite-dimensional vector space and let $\langle\cdot,\cdot\rangle$ denotes the inner product. $\Theta(E)$ denotes the set of convex proper and lower semi-continuous functions on E. Let f be a d.c. function on E that means there exist two functions g and h both in $\Theta(E)$ such that:

$$f(x) = g(x) - h(x), \quad \forall x \in E.$$

*Mathematics Subject Classifications: 65K10, 90C26, 90C30.

†Department of Mathematics, College of Science, King Saud University, P.O. Box 2455, Riyadh 11451. Saudi Arabia.

Moreover, suppose that $h(x) \neq +\infty$ for $x \in E$ and that $ridom(g) \cap ridom(h) \neq \emptyset$.

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

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

The next theorem gathers several global optimality conditions for 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 $x^* \in dom\ g\ \cap\ dom\ h$. 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) $\partial_\epsilon h(x^*) \subset \partial_\epsilon g(x^*), \ \ \forall\ \epsilon \geq 0$,

(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

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

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

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

PROOF. We sketch here the proof of the above theorem, and for a more complete proof see [3] Theorem 2.3.1 pp. 8-9. Suppose (G) is satisfied. Let us consider a function $\phi \in \Omega$. If $h(x) - h(x^*) \geq \phi(x) - \phi(x^*)$ for any value of $x \in E$, according to the global optimality of x^*, we conclude that $g(x) - g(x^*) \geq \phi(x) - \phi(x^*)$ and (FB) is satisfied.

Now let us show that (FB) implies (HU). To this goal, let $\epsilon \geq 0$ and let $z \in \partial_\epsilon h(x^*)$ and $z^* \in \partial h(x^*)$ be given and let us define for any $x \in E$ the function $\phi(x) = \sup\{\langle x - x^*, z\rangle - \epsilon,\ \langle x - x^*, z^*\rangle\}$. It is easy to see that $\phi(x^*) = 0$ which implies $\phi(x) - \phi(x^*) \geq \langle x - x^*, z\rangle - \epsilon$ and $\phi \in\ \partial^\gamma h(x^*)$. Hence (HU) is fulfilled.

To prove that (HU) implies (ST) it suffices to notice that if $z \in \partial h(x^*)$, then (ST) is obtained directly, otherwise, we only need to take $\epsilon = h^*(x^*) + h(x^*) - \langle x^*, z\rangle$.

From (ST), we have $g(x^*) - g(^*(z) \leq h(x^*) - h^*(z)$ for all $z \in E^*$ and for $x \in dom(g) \bigcap dom(h)$ we can write

$$\langle x, z\rangle - g(x^*) - g(^*(z) \geq \langle x, z\rangle - h(x^*) - h(^*(z), \quad \forall\ z \in E^*.$$

[1]This theorem is taken from the PhD dissertation of Dr Mirjam D¨ur pp.8-9. in [3] and for more references see [5].

[2]The authors of these conditions are Flores-Bazan (FB), Hirriart-Urruty (HU), Singer and Toland (ST) and Horst , Pardalos and Thoai (HPT) see [3].

Taking the supremum over all $z \in E^*$ it is easy to see that x^* is a global minimizer, then (G) holds.

To complete the proof, it remains to show that (G) implies (HPT) and vice-versa. Let us assume that (HPT) is not satisfied, i.e., there exists (y, σ_1, τ_1) such that $g(y) \leq \sigma_1$ and $\sigma_1 - \tau_1 \leq g(x^*) - h(x^*)$ with $h(y) - \tau_1 > 0$. Then it follows obviously that y is a global minimizer which contradicts G. Now, let us assume that x^* is not a global minimizer of $g - h$, i.e., there exists y such that $g(y) - h(y) < g(x^*) - h(x^*)$. To obtain the contradiction it suffices to set $\sigma_1 = g(y)$ and $\tau = g(y) - g(x^*) + h(x^*)$. Finally (HPT) implies (G). The proof is complete.

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

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

According to the assumption $ridom(g) \cap ridom(h) \neq \emptyset$, 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 subdifferential inclusion in general, and this is why we may content ourselves by solving the relaxed variational problem

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

2 Moreau-Yosida Regularization Scheme

2.1 A Minimization Scheme

In this section , we propose the construction of a scheme to approximate the critical point x^* of f satisfying (4). To this goal, we aim to use the nice and useful property of convexity of the functions g and h. The idea can be stated as follows. Starting from an initial point $x \in E$, we can select a point $z \in \partial h(x)$ and using a proximal regularization scheme (see [8]), we construct a translated fixed point iterative scheme by $x_{new} = (I + \lambda\partial g)^{-1}(x + \lambda z)$ yielding a fixed point x that coincides with a critical point of f. These steps are possible under the convexity of both functions f and g. Next we state the algorithm and then we show the well-definedness of all the steps.

Algorithm 1.

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

3. Proximal Step. $x^{t+1} = J_\lambda^{\partial g}(x^t + \lambda z^t)$.

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

In the above, $J_\lambda^{\partial g}$ denotes the Moreau-Yosida resolvent corresponding to the operator ∂g defined by $J_\lambda^{\partial g} = (I + \lambda\partial g)^{-1}$.

2.2 Convergence Analysis

In this subsection, we will focus on the well-definedness and the convergence of the sequences involved in Algorithm 1.

PROPOSITION 1. Let $h \in \Theta(E)$, then $\partial h(x) \neq \emptyset, \quad \forall\, x \in ridom(h)$.

For proof, see Rockafellar [8].

PROPOSITION 2. Let $g \in \Theta(E)$, then ∂g is a maximal monotone operator and the corresponding resolvent operator $J_\lambda^{\partial g}$ is univoque (single-valued).

For proof, see Bresis [1].

PROPOSITION 3 . A vector $x \in E$ satisfies the necessary optimality condition (2) if and only if

$$x = J_\lambda^{\partial g}\left(x + \lambda z\right), \quad \forall \lambda > 0, \quad z \in \partial h(x).$$

PROOF. Let $x \in E, \quad z \in \partial h(x)$ and λ a positive parameter. According to proposition 3, $x = J_\lambda^{\partial g}\left(x + \lambda z\right)$ is well-defined and it gives

$$x + \lambda z \in \left(I + \lambda \partial g\right)(x),$$

i.e.,

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

By 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,

$$\begin{aligned} z \in \partial g(x) &\implies \lambda z \in \lambda \partial g(x) \\ &\implies x + \lambda z \in x + \lambda \partial g(x) \\ &\implies (x + \lambda z) \in \left(I + \lambda \partial g\right)(x) \\ &\implies x = \left(I + \lambda \partial g\right)^{-1}(x + \lambda z). \end{aligned}$$

The proof is complete since $z \in \partial h(x)$.

PROPOSITION 4. Let $\{x^t\}_t$ be the sequence generated by the proximal step in Algorithm 1:

$$x^{t+1} = J_\lambda^\partial g\left(x^t + \lambda z^t\right). \tag{5}$$

It converges to a critical point of f.

PROOF. Assume that the convergence is not reached yet, then from (5), we get $x^{t+1} = \left(I + \lambda \partial g\right)^{-1}(x^t + \lambda z^t)$, which can be rewritten in the following form

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

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

$$g(x) \geq g(x^{t+1}) + \left\langle x - x^{t+1}, \frac{x^t - x^{t+1}}{\lambda} + z^t \right\rangle, \quad \forall\, x. \tag{6}$$

In particular, for $x = x^t$, we have

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

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

$$h(x) \geq h(x^t) + \langle x - x^t, z^t \rangle, \quad \forall\, x, \tag{8}$$

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

$$h(x^{t+1}) \geq h(x^t) + \langle x^{t+1} - x^t, z^t \rangle. \tag{9}$$

(7) can be rewritten as

$$g(x^{t+1}) \leq g(x^t) - \frac{1}{\lambda}\|x^t - x^{t+1}\|^2 - \langle x^t - x^{t+1}, z^t \rangle \tag{10}$$

and (9) becomes

$$-h(x^{t+1}) \leq -h(x^t) + \langle x^t - x^{t+1}, z^t \rangle. \tag{11}$$

By adding (10) and (11), we get $f(x^{t+1}) \leq f(x^t) - \frac{1}{\lambda}\|x^t - x^{t+1}\|^2$, i.e., $f(x^{t+1}) \leq f(x^t)$ which proves that the considered scheme is a descent one.

THEOREM 2. If we assume the boundedness of the sequence $\{x^t\}_t$ generated by the proximal step in Algorithm 1, then every accumulation point of $\{x^t\}_t$ is a critical point of f.

3 Application

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

$$\text{Find } x^* \in C,\ y^* \in F(x^*) \ \text{ such that } \ \left\langle y^*, x - x^* \right\rangle \geq 0, \quad \forall\, x \in C, \tag{12}$$

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 by $\langle \cdot, \cdot \rangle$ and $\| \cdot \|$ respectively.

This problem has many important applications, e.g., in economics, operations research, industry, 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 the seminal surveys of Harker and Pang [4] and A.M. Noor [6].

We are interested in the case where the mapping $F = -\partial h$ where h is a convex function. Thus, (12) becomes

$$\text{Find } x^* \in C,\ y^* \in -\partial h(x^*) \ \text{ such that } \ \left\langle y^*, x - x^* \right\rangle \geq 0, \quad \forall\, x \in C. \tag{13}$$

Let the indicator function ψ_C of the set C defined by $\psi_C(x) = 0$ if $x \in C$ and ∞ if $x \notin C$. Since it is well known that

$$\partial\psi_C(x^*) = N_C(x^*) = \begin{cases} \{w \in Re^n : \langle w, x - x^*\rangle \leq 0,\ \forall x \in C\} \text{ if } x^* \in C \\ \emptyset, \quad \text{otherwise.} \end{cases},$$

it is easy to see that (13) is equivalent to

$$y^* \in -\partial h(x^*) \bigcap -\partial\psi_C(x^*), \tag{14}$$

i.e., if we set $w^* = -y^*$, then solving (12) is equivalent to finding w^* such that

$$w^* \in \partial h(x^*) \bigcap \partial\psi_C(x^*). \tag{15}$$

Our Algorithm can be applied to solve (15) 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. $x^{t+1} = (I + \lambda\partial\psi_C)^{-1}(x^t + \lambda w^t)$.

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

Note that step 3. is equivalent to $x^{t+1} = P_C(x^t + \lambda w^t)$, where P_C denotes the projection operator on C.

Acknowledgments. The author would like to express his sincere thanks to Professor Aslam Noor for his helpful suggestions and Dr. M. Bounkhel for his valuable remarks on the final version of this work. This work was supported by the research center of the college of science, King Saud University, under the grant (1426/–).

References

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

[2] R. S. Burachik, Generalized proximal point methods for the variational inequality problem, PhD thesis, Instituto de Matemàtica Pura e Aplcada, Rio de Janeiro, Brazil, 1995.

[3] M. Dür, Duality in Global Optimization: Optimality Conditions and Algorithmical Aspects., PhD Dissertation, Trier University, 1999, Germany.

[4] 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.

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

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

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

[8] R. T. Rockafellar, Monotone operators and the proximal point algorithm in convex programming, SIAM J. on Control and Optim., 14(1976), 877–898.