Convergence of the projected gradient method for quasiconvex multiobjective optimization

J.Y. Bello Cruz* L.R. Lucambio Pérez† J.G. Melo‡

January 12, 2011

Abstract

We consider the projected gradient method for solving the problem of finding a Pareto optimum of a quasiconvex multiobjective function. We show convergence of the sequence generated by the algorithm to a stationary point. Furthermore, when the components of the multiobjective function are pseudoconvex, we obtain that the generated sequence converges to a weakly efficient solution.

Keywords: Armijo-type search; Multiobjective optimization; Pareto optimality; Projected gradient methods; Quasiconvex multiobjective functions.

Mathematical Subject Classification (2010): 90C29, 90C30.

*Instituto de Matemática e Estatística, Universidade Federal de Goiás, Campus Samambaia, CEP 74001-970 GO, Goiânia, Brazil, e-mail: yunier@impa.br

†e-mail: lrlp@mat.ufg.br

‡e-mail: jefferson@mat.ufg.br

1 Introduction

In multicriteria optimization, several objective functions have to be minimized simultaneously. Usually, no single point will minimize all given objective functions at once, and so the concept of optimality has to be replaced by the concept of Pareto-optimality or efficiency.

Finding efficient points for the preference order induced by the Paretian cone $\mathbb{R}^m_+$ is a very relevant problem on many areas, such as engineering, statistics, design and others, see [1, 2, 3, 4, 5, 6, 7].

A popular strategy for solving multiobjective optimization problems is the scalarization approach. The most widely used scalarization technique is the weighting method. Basically, one minimizes a linear combination of the objectives, where the vector of "weights" is not known a priori and, so, this procedure may lead to unbounded numerical problems, which, therefore, may lack minimizers, see [8, 9, 10]. Another disadvantage of this approach is that the choice of the parameters is not known in advance, leaving the modeler and the decision-maker with the burden of choosing them.

The class of quasiconvex multiobjective functions has many applications in the real life problems, for example in economy. For this kind of problems the weighting method has another weakness: linear combinations of quasiconvex functions may not be quasiconvex. This fact is showed in the following example:

Given $m,\, n \in \mathbb{N}$, m odd and n even, $m > n$, consider the quasiconvex multiobjective function $f\colon \mathbb{R} \to \mathbb{R}^2$ defined by

$$f(x) = \begin{pmatrix} f_1(x) \\ f_2(x) \end{pmatrix} = \begin{pmatrix} -\frac{x^m}{m} \\ \frac{x^n}{n} \end{pmatrix}.$$

Take the parameters $w = (\alpha\,,\, \beta) \in \mathbb{R}^2_{++}$. The scalar functions

$$F^{(\alpha,\beta)}(x) = \langle w\,,\, f\rangle(x) = \alpha f_1(x) + \beta f_2(x) = -\alpha\frac{x^m}{m} + \beta\frac{x^n}{n},$$

are not quasiconvex.

In this paper we are interesting in the study of the projected gradient algorithm for solving quasiconvex multiobjective optimization problem. An advantage of this method over the weighting method is that it explores the structure of quasiconvexity of the problem, as we will show.

Recently, the gradient method was proposed for multiobjective optimization problem in [11]. Since then, it has been considered in more general setting, for instance, for vector optimization problem, see [12], and for constrained vector optimization, see [13, 14]. Classical methods for solving scalar optimization problem have been extended for the setting of vector-valued optimization problem, see [11, 13, 15, 16, 17, 18].

We analyze the projected gradient method for quasiconvex multicriteria optimization. Under suitable assumption, we show that the sequence generated by the algorithm converges to a stationary point. Moreover, when the components of the multiobjective function are pseudoconvex we obtain that the sequence converges to a weakly efficient point.

The outline of this article is as follows. In Section 2 we present some basic definitions, assumptions and some preliminary materials. In Section 3 we present the projected gradient method for multiobjective optimization. Section 4 contains the convergence analysis of the algorithm.

2 Basic definitions and preliminary material

For $u, v \in \mathbb{R}^m$, $u \preceq v$ ($u \prec v$) means $u_i \leq v_i$ ($u_i < v_i$) for $i = 1, \ldots, m$. The usual inner product in $\mathbb{R}^n$ is denoted by $\langle \cdot, \cdot \rangle$. The norm determined by the inner product is $\| \cdot \|$. In that follows $f : \mathbb{R}^n \to \mathbb{R}^m$ is a continuously differentiable function, $f_i \colon \mathbb{R}^n \to \mathbb{R}$, $i = 1, \ldots, m$, are its coordinate functions, $C \subseteq \mathbb{R}^n$ is a nonempty closed convex set and for $x \in \mathbb{R}^n$, the orthogonal projection of x onto C is denoted by $P_C(x)$.

An element $x \in C$ is a Pareto optimum or efficient for f restricted to C if and only if there does not exist $y \in C$ such that $f(y) \preceq f(x)$ and $f(y) \neq f(x)$. The problem of finding an efficient point for f restricted to C is the following

$$\min_{\mathbb{R}^m_+} f(x) \qquad \text{s.t.} \qquad x \in C. \tag{1}$$

Observe that Problem (1) becomes a scalar-valued optimization problem when $m = 1$ and $\preceq$ is the usual order in $\mathbb{R}$.

A weakly efficient solution for Problem (1) is a point $x \in C$ such that there does not exist $y \in C$ satisfying $f(y) \prec f(x)$. A necessary condition for a point $x \in C$ to be weakly efficient is that

$$-\mathbb{R}^m_{++} \cap J_f(x)(C - x) = \emptyset, \tag{2}$$

where $C - x = \{y - x : \ y \in C\}$. This condition, in general, is not sufficient for a point to be weakly efficient, see [19]. The points of C satisfying (2) are called stationary points.

A scalar function $g : \mathbb{R}^n \to \mathbb{R}$ is called quasiconvex if and only if $g(\alpha x + (1 - \alpha)y) \leq \max\{g(x), g(y)\}$ for every $\alpha \in [0, 1]$ and $x, y \in \mathbb{R}^n$. When g is differentiable, g is quasiconvex if and only if for each $x, y \in \mathbb{R}^n$, the inequality $g(x) \leq g(y)$ implies that $\langle \nabla g(y) \, , \, x - y \rangle \leq 0$, see [20].

A multiobjective function $f : \mathbb{R}^n \to \mathbb{R}^m$ is quasiconvex if and only if every coordinate function f_i, $i = 1, \ldots, m$, is quasiconvex, see [21].

A differentiable function $g : \mathbb{R}^n \to \mathbb{R}$ is said to be pseudoconvex if and only if $\langle \nabla g(x), y - x \rangle \geq 0$ implies $g(y) \geq g(x)$. It is well known that pseudoconvex functions are quasiconvex.

We will consider a special class of multiobjective quasiconvex functions, the ones such that each component is pseudoconvex. When the components of f are pseudoconvex then (2) is a necessary and sufficient condition for a point to be weakly efficient. Indeed, suppose to the contrary, that there exists $y \in C$ such that $f(y) \prec f(x)$. Since f_i is pseudoconvex for each $i = 1, \ldots, m$, it follows that $\langle \nabla f_i(x), (y-x) \rangle < 0$ and therefore $J_f(x)(y-x) \in -\mathbb{R}^m_{++}$, contradicting (2). This last result, in general, does not hold for quasiconvex functions, see [22].

The convergence analysis of the proposed algorithm is based on quasi-Fejér convergence. We recall that a sequence $\{z^k\} \subset \mathbb{R}^n$ is said to be quasi-Fejér convergent to a set V, $V \neq \emptyset$, if and only if for each $z \in V$ there exists a sequence $\{\epsilon_k\} \subset \mathbb{R}_+$ such that $\sum_{k=1}^{+\infty} \epsilon_k < +\infty$ and

$$\|z^{k+1} - z\|^2 \leq \|z^k - z\|^2 + \epsilon_k.$$

The following result on quasi-Fejér convergence is well known.

Lemma 1. *If* $\{z^k\} \subset \mathbb{R}^n$ *is* quasi-Fejér *convergent to some set* $V \neq \emptyset$*, then:*

i) The sequence $\{z^k\}$ *is bounded;*

ii) if an accumulation point of $\{z^k\}$ *belongs to* V*, then* $\{z^k\}$ *is convergent to some* $z \in V$*.*

Proof. See Theorem 1 in [23]. □

3 Projected gradient algorithm

Consider two constants: $\beta > 0$ and $\sigma \in (0,1)$. The projected gradient algorithm is as follows.

Initialization: Take $x^0 \in C$.

Iterative step: Given x^k, compute

$$v^k := \arg\min_{v \in C - x^k} \left\{ \frac{\|v\|^2}{2} + \beta \max_{1 \leq i \leq m} \langle \nabla f_i(x^k)\,,\, v \rangle \right\}. \tag{3}$$

If $v^k = 0$, then stop. Otherwise compute

$$j(k) := \min\{j \in \mathbb{Z}_+ \colon f(x^k + 2^{-j} v^k) \preceq f(x^k) + \sigma 2^{-j} J_f(x^k) v_k\}. \tag{4}$$

Set

$$x^{k+1} = x^k + \gamma_k v^k, \tag{5}$$

where $\gamma_k = 2^{-j(k)}$.

This algorithm was proposed by Fliege and Svaiter in [11] for unconstrained multiobjective problem. Its extension for the constrained case was studied by Graña and Iusem in [13]. Other variants of this algorithm can be found in [14] and [12]. It is an extension of the classical (scalar) projected gradient algorithm to the constrained multiobjective problem. Indeed, taking $m = 1$, (3) becomes

$$\begin{aligned} v^k &= \arg\min_{v \in C - x^k} \left\{ \frac{\|v\|^2}{2} + \beta \langle \nabla f(x^k) \, , \, v \rangle \right\} \\ &= \arg\min_{v \in C - x^k} \left\{ \frac{\|v\|^2}{2} + \beta \langle \nabla f(x^k) \, , \, v \rangle + \frac{\beta^2}{2} \|\nabla f(x^k)\|^2 \right\} \\ &= \arg\min_{v \in C - x^k} \left\{ \frac{1}{2} \|v + \beta \, \nabla f(x^k)\|^2 \right\} = P_{C - x^k} \left(-\beta \, \nabla f(x^k) \right). \end{aligned}$$

Combining the last equation with (5) and using the definition of P_C, we obtain

$$x^{k+1} = (1 - \gamma_k) x^k + \gamma_k P_C \left(x^k - \beta \, \nabla f(x^k) \right).$$

The convergence analysis of variants of this algorithm ($m = 1$), for quasiconvex objectives, can be seen in [24], [25] and [26].

4 Convergence analysis

¿From now on, $\{x^k\}$ is the sequence generated by the projected gradient algorithm. Observe that if x^k is not a stationary point then $J_f(x^k) v^k \prec 0$. Thus, Armijo rule (4) is well defined, see Proposition 1 of [13].

Proposition 1. *The sequence generated by the projected gradient algorithm is feasible and* $\{f_i(x^k)\}_{k=0}^{\infty}$, $i = 1, \ldots, m$, *are monotone decreasing.*

Proof. The feasibility of the sequence $\{x^k\}$ is a consequence of the definition of the algorithm, see [13, Proposition 5], and the decreasing property follows from (4). $\square$

Assuming only differentiability of f the main result on the convergence of the projected gradient algorithm is the following.

Proposition 2. *Every accumulation point, if any, of* $\{x^k\}$ *is a stationary point of Problem* (1).

Proof. See Theorem 1 in [13]. $\square$

The novelty of this paper occurs in what follows.

Lemma 2. *For all* $x \in C$ *and each* k, *there exists* $\{\lambda_j^k\}_{j=1}^m \subset [0,1]$ *satisfying* $\sum_{j=1}^m \lambda_j^k = 1$, *and*

$$\|x^{k+1} - x\|^2 \leq \|x^k - x\|^2 + 2\gamma_k \beta \langle \sum_{j=1}^m \lambda_j^k \nabla f_j(x^k) \, , \, x - x^k \rangle + \frac{2\beta}{\sigma} \sum_{j=1}^m \big(f_j(x^k) - f_j(x^{k+1}) \big) \, .$$

Proof. Fix $k \geq 0$. The function

$$\varphi_k(v) := \frac{\|v\|^2}{2} + \beta \max_{1 \leq i \leq m} \left\{ \langle \nabla f_i(x^k) \, , \, v \rangle \right\} ,$$

is strongly convex. Therefore the first order optimality condition for the problem $\min\limits_{v \in C - x^k} \varphi_k(v)$ is necessary and sufficient. So, there exist a feasible direction v^k and $u^k \in \partial \varphi_k(v_k)$ such that

$$\langle u^k \, , \, v - v^k \rangle \geq 0, \qquad \forall \, v \in C - x^k. \tag{6}$$

It follows from the expression of φ_k that there exist $\lambda_j^k > 0$, $j \in J_k \subseteq \{1, 2, \ldots, m\}$, such that

$$\sum_{j \in J_k} \lambda_j^k = 1,$$

and

$$u^k = v^k + \beta \sum_{j \in J_k} \lambda_j^k \nabla f_j(x^k). \tag{7}$$

Defining $\lambda_j^k = 0$ for $j \notin J_k$, (7) can be rewriting as

$$u^k = v^k + \beta \sum_{j=1}^{m} \lambda_j^k \nabla f_j(x^k), \tag{8}$$

with

$$\sum_{j=1}^{m} \lambda_j^k = 1. \tag{9}$$

After some simple algebra, for each $x \in C$, we obtain

$$d_k := \|x^k - x\|^2 - \|x^{k+1} - x\|^2 + \|x^{k+1} - x^k\|^2 = 2\langle x^{k+1} - x^k, x - x^k \rangle.$$

Using the update formula $x^{k+1} = x^k + \gamma_k v^k$, it follows that

$$\begin{aligned} d_k &= 2\gamma_k \langle v^k \, , \, x - x^k \rangle \\ &= 2\gamma_k \langle v^k \, , \, x - x^k - v^k + v^k \rangle \\ &= 2\gamma_k \langle v^k \, , \, x - x^k - v^k \rangle + 2\gamma_k \|v^k\|^2. \end{aligned} \tag{10}$$

Using (8) and (10), we get

$$\begin{aligned} d_k - 2\gamma_k \|v^k\|^2 &= 2\gamma_k \langle u^k - \beta \sum_{j=1}^{m} \lambda_j^k \nabla f_j(x^k) \, , \, x - x^k - v^k \rangle \\ &= 2\gamma_k \Big(\langle u^k \, , \, x - x^k - v^k \rangle - \beta \langle \sum_{j=1}^{m} \lambda_j^k \nabla f_j(x^k) \, , \, x - x^k - v^k \rangle \Big) \\ &\geq -2\gamma_k \Big(\beta \langle \sum_{j=1}^{m} \lambda_j^k \nabla f_j(x^k) \, , \, x - x^k \rangle - \beta \langle \sum_{j=1}^{m} \lambda_j^k \nabla f_j(x^k) \, , \, v^k \rangle \Big), \end{aligned}$$

where the inequality above follows from (6). Therefore,

$$\begin{aligned} \|x^k - x\|^2 - \|x^{k+1} - x\|^2 &+ \gamma_k(\gamma_k - 2)\|v^k\|^2 = d_k - 2\gamma_k \|v^k\|^2 \\ &\geq 2\gamma_k \beta \Big(\langle \sum_{j=1}^{m} \lambda_j^k \nabla f_j(x^k) \, , \, x^k - x \rangle + \langle \sum_{j=1}^{m} \lambda_j^k \nabla f_j(x^k) \, , \, v^k \rangle \Big). \end{aligned}$$

Rewriting this last inequality and using that $f(x^{k+1}) \preceq f(x^k) + \sigma\gamma_k J_f(x^k)v_k$ with $\gamma_k = 2^{-j(k)} < 1$, it follows that

$$\|x^{k+1} - x\|^2 \leq \|x^k - x\|^2 + 2\gamma_k\beta\langle\sum_{j=1}^{m} \lambda_j^k \nabla f_j(x^k)\,,\, x - x^k\rangle + \frac{2\beta}{\sigma}\sum_{j=1}^{m} \lambda_j^k \left(f_j(x^k) - f_j(x^{k+1})\right).$$

Using that $f(x^{k+1}) \preceq f(x^k)$ and $1 \geq \lambda_j^k \geq 0$ for all $j = 1, ..., m$, we obtain

$$\sum_{j=1}^{m} \left(f_j(x^k) - f_j(x^{k+1})\right) \geq \sum_{j=1}^{m} \lambda_j^k \left(f_j(x^k) - f_j(x^{k+1})\right),$$

establishing the result.

□

We are interested in the study of the convergence properties of the projected gradient algorithm when the objective function is quasiconvex. Define

$$T = \left\{x \in C\colon f(x) \preceq f(x^k),\ \forall\, k\right\}.$$

Proposition 3. *Assume that f is quasiconvex and $x \in T$. Then, $\langle\nabla f_i(x^k)\,,\, x - x^k\rangle \leq 0$ for all k and all $i = 1, ..., m$.*

Proof. Since f is quasiconvex, the f_i for $i = 1, ..., m$ are quasiconvex. Then, the result is only the classical characterization of the differentiable quasiconvex functions, see [20]. □

The next theorem establishes a sufficient condition for the convergence of the sequence $\{x^k\}$. This result is the main convergence result in the quasiconvex case.

Theorem 1. *Assume that f is a quasiconvex function. If $T \neq \emptyset$ then $\{x^k\}$ converges to a stationary point.*

Proof. By Proposition 1, the sequence $\{f(x^k)\}$ is strictly monotone decreasing, i.e., $f_i(x^k) > f_i(x^{k+1})$, for all k and $i = 1, \dots, m$. Therefore $\varepsilon_k := \sum_{i=1}^{m} \left[f_i(x^k) - f_i(x^{k+1})\right]$ is positive for all k. Take $x \in T$. By Lemma 2, we get

$$\|x^{k+1} - x\|^2 \leq \|x^k - x\|^2 + 2\gamma_k \beta \langle \sum_{i=1}^{m} \lambda_i^k \nabla f_i(x^k)\, ,\, x - x^k \rangle + \frac{2\beta}{\sigma}\varepsilon_k.$$

It follows from Proposition 3 that

$$\|x^{k+1} - x\|^2 \leq \|x^k - x\|^2 + \frac{2\beta}{\sigma}\varepsilon_k.$$

Observe that the series $\sum_{k=1}^{\infty} \varepsilon_k$ is convergent. Indeed, $\varepsilon_k > 0$ for all k and

$$\begin{aligned}\sum_{k=0}^{\ell} \varepsilon_k &= \sum_{k=0}^{\ell} \left(\sum_{i=1}^{m} (f_i(x^k) - f_i(x^{k+1})) \right) = \sum_{i=1}^{m} \left(\sum_{k=0}^{\ell} (f_i(x^k) - f_i(x^{k+1})) \right) \\ &= \sum_{i=1}^{m} (f_i(x^0) - f_i(x^{\ell+1})) \leq \sum_{i=1}^{m} (f_i(x^0) - f_i(x)).\end{aligned}$$

Therefore, $\{x^k\}$ is quasi-Fejér convergent to T. Using Lemma 1(i) the sequence $\{x^k\}$ is bounded. Let $\bar{x}$ be an accumulation point of $\{x^k\}$. By Proposition 1, the sequence $\{f(x^k)\}$ is monotone decreasing, which implies that $\bar{x} \in T$. It follows from Lemma 1(ii) that $\{x^k\}$ is convergent to $\bar{x}$, which is stationary by Proposition 2. $\square$

It is well known that convex functions are quasiconvex, but the converse is not true in general. Therefore, the above theorem generalizes the results in [13] and [14] to the case where the multiobjective function is quasiconvex. Also it generalizes the scalar case in [26]. Now, we present the main convergence result in the pseudoconvex case.

Corollary 1. *Assume that the components of f are pseudoconvex. If $T \neq \emptyset$ then $\{x^k\}$ converges to a weakly efficient point.*

Proof. By Theorem 5, Section 3, Chapter 9 of [20], pseudoconvexity implies quasiconvexity. Then using Theorem 1, we obtain that $\{x^k\}$ converges to a stationary point, which is weakly efficient point. □

Note that when f is quasiconvex, $\{x^k\}$ has cluster points if and only if T is a nonempty set. Indeed, if there exists an accumulation point x of $\{x^k\}$ then, $\{f_i(x^k)\}_{k=0}^{\infty}$ converge to $f_i(x)$ for all $i = 1, \ldots, m$. Since $f_i(x^k)$, $i = 1, \ldots, m$, are monotone decreasing, x belongs to T. The converse result follows from Theorem 1.

The assumption that $T \neq \emptyset$ was used in [13] and [14] for proving the convergence of the algorithm in the convex case. All these assumptions have relation with the completeness of the image of f, namely that all non-increasing sequences in the image of f have lower bound. It is important to say that completeness is a standard assumption for ensuring existence of efficient points [10].

Acknowledgment

The authors were partially supported by Procad-NF/Capes and Pronex/Faperj/CNPq.

References

[1] H. Eschenauer, J. Koski, A. Osyczka, Multicriteria Design Optimization, Springer, Berlin (1990).

[2] E. Carrizosa, J.B.G. Frenk, Dominating sets for convex functions with some applications, Journal of Optimization Theory and Applications 96 (1998) 281-295.

[3] J. Fliege, OLAF - A general modeling system to evaluate and optimize the location of an air polluting facility, OR Spektrum 23 (2001) 117-136.

[4] M. Gravel, J.M. Martel, R. Nadeau, W. Price, R. Tremblay, A multicriterion view of optimal ressource allocation in job-shop production, European Journal of Operational Research 61 (1992) 230-244.

[5] D. Prabuddha, J.B. Ghosh, C.E. Wells, On the minimization of completion time variance with a bicriteria extension, Operations Research 40 (1992) 1148-1155.

[6] D.J. White, Epsilon-dominating solutions in mean-variance portfolio analysis. European Journal of Operations Research 105 (1998) 457-466.

[7] Y. Fu, U.M. Diwekar, An efficient sampling approach to multiobjective optimization, Annals of Operations Research 132 (2004) 109-134.

[8] L.M. Graña Drummond, N. Maculan, B.F. Svaiter, On the choice of parameters for the weighting method in vector optimization. Mathematical Programming 111 (2008) 201-216.

[9] J. Jahn, Scalarization in vector optimization, Mathematical Programming 29 (1984) 203-218.

[10] D.T Luc, Theory of Vector Optimization. Lecture Notes in Economics and Mathematical Systems, Springer, Berlin (1989).

[11] J. Fliege, B.F. Svaiter, Steepest descent methods for multicriteria optimimization, Mathematical Methods of Operations Research 51 (2000) 479-494.

[12] L.M. Graña Drummond, B.F. Svaiter, A steepest descent method for vector optimization, Journal of Computational and Applied Mathematics 175 (2005) 395-414.

[13] L.M. Graña Drummond, A.N. Iusem, A projected gradient method for vector optimization problems, Computational Optimization and Applications 28 (2004) 5-30.

[14] E.H. Fukoda, L.M. Graña Drummond, On the convergence of the projected gradient method for vector optimization, (to be published in Optimization).

[15] H. Bonnel, A.N. Iusem, Svaiter, B.F. Proximal methods in vector optimization, SIAM Journal on Optimization 15 (2005) 953-970.

[16] Z. Chen, C. Xiang, K. Zhao, Liu, X. Convergence analysis of Tikhonov-type regularization algorithms for multiobjective optimization problems, Applied Mathematics and Computation 211 (2009) 167-172.

[17] L.C. Ceng, B.S. Mordukhovich, Yao, J.C. Hybrid Approximate Proximal Method with Auxiliary Variational Inequality for Vector Optimization, Journal of Optimization Theory and Applications 146 (2010) 267-303.

[18] J. Fliege, L.M. Graña Drummond, B.F. Svaiter, Newton's method for multiobjective optimization, SIAM Journal on Optimization 20 (2009) 602-662.

[19] G.Y. Chen, J. Jahn, Optimality conditions for set-valued optimization problems. Set-valued optimization, Mathematical Methods of Operations Research 48 (1998) 187-200.

[20] O.L. Mangasarian, Nonlinear Programming, McGraw-Hill, New York (1969).

[21] J. Benoist, J.M. Borwein, N. Popovici, A characterization of quasi-convex vector valued-functions, Proceedings of the American Mathematical Society 4 (2003) 1109-1113.

[22] S. Boyd, L. Vandenberghe, Convex Optimization, Cambridge University Press, New York (2007).

[23] R. Burachik, L.M. Graña Drummond, A.N. Iusem, B.F. Svaiter, Full convergence of the steepest descent method with inexact line searches, Optimization 32 (1995) 137-146.

[24] K. Kiwiel, K.G. Murty, Convergence of the steepest descent method for minimizing quasi convex functions, Journal of Optimization Theory and Applications 89 (1996) 221-226.

[25] C. Wang, N. Xiu, Convergence of the gradient projection method for generalized convex minimization, Computational Optimization and Applications 16 (2000) 111-120.

[26] J.Y. Bello Cruz, L.R. Lucambio Pérez, Convergence of a projected gradient method variant for quasiconvex objectives, Nonlinear Analysis 73 (2010) 2917-2922.