APPROXIMATE GRADIENT PROJECTED CONDITION IN MULTIOBJECTIVE OPTIMIZATION

Ramos, Alberto[†], Sánchez, María Daniela[‡] and Schuverdt, María Laura[‡]

[†]*Department of Mathematics, Federal University of Parana, Curitiba-PR, Brazil,albertoramos@ufpr.br*
[‡]*Faculty of Engineering, University of La Plata, CP 172, 1900 La Plata Bs. As., Argentina,*
dsanchez@mate.unlp.edu.ar
[‡]*CONICET, University of La Plata, CP 172, 1900 La Plata Bs. As., Argentina, schuverd@mate.unlp.edu.ar*

Abstract: In this work we present an extention of the well-known Approximated Gradient Projection (AGP) [8] property from the scalar problem with equality and inequality constraints to multiobjective problems. We prove that the condition called Multiobjective Approximate Gradient Projection (MAGP), is necessary for a point to be a local weak Pareto point and we study, under convex assumptions, sufficient conditions.

1   INTRODUCTION

We will consider the multiobjective optimization problem (MOP) of the form:

$$\text{Minimize } F(x) \text{ subject to } x \in \Omega \tag{1}$$

where $\Omega = \{x \in \mathbb{R}^n : h(x) = 0, g(x) \leq 0\}$ and $F : \mathbb{R}^n \to \mathbb{R}^r, h : \mathbb{R}^n \to \mathbb{R}^m, g : \mathbb{R}^n \to \mathbb{R}^p$ are continuously differentiable functions and $F(x) = (f_1(x), \ldots, f_r(x))$.

In [8] the authors introduce a sequential optimality condition, for the scalar optimization problem ($r = 1$), called Approximate Gradient Projection (AGP). It is proved that the AGP property is satisfied by local minimizers of constrained optimization problems independently of constraint qualification.

One of the attractiveness of AGP is that it does not involve Lagrange multipliers estimates. AGP is also the natural optimality condition that fits a stopping criteria for algorithms based on inexact restoration, and it is strictly stronger than the usual AKKT condition [7]. Consequently, the stopping criteria based on AGP is more reliable that those based on AKKT.

In this paper, we will extend the AGP condition to multiobjective problems, in the same way as the AKKT condition [7] has been extended recently in [5] for multiobjective problems.

Given the multiobjective optimization problem in the form (1), a point $x^* \in \Omega$ is said to be a Pareto optimal solution [9] or an efficient solution if there is no $x \in \Omega$ such that $F(x) \leq F(x^*)$ and $F(x) \neq F(x^*)$. A point $x^* \in \Omega$ is said to be a weak Pareto optimal solution of (1) if there is no $x \in \Omega$ such that $F(x) < F(x^*)$. A point x^* is a local weak Pareto optimal solution if there exists a neighbourhood $N \cap \Omega$ of x^* such that there is no $x \in N$ with $F(x) < F(x^*)$. In this paper we are interested in the Fritz-John necessary condition for weak Pareto optimality stated in the following theorem:

Theorem 1 *[4] A necessary condition for $x^* \in \Omega$ to be a local weak Pareto optimal solution is that there exist vectors $\theta \in \mathbb{R}^r, \lambda \in \mathbb{R}^m, \mu \in \mathbb{R}^p$ such that $\theta \geqslant 0$, $\mu \geqslant 0$ and*

$$\sum_{l=1}^{r} \theta_l \nabla f_l(x^*) + \sum_{i=1}^{m} \lambda_i \nabla h_i(x^*) + \sum_{j=1}^{p} \mu_j \nabla g_j(x^*) = 0, \tag{2}$$

$$\mu_j g_j(x^*) = 0, j = 1, \ldots, p, \tag{3}$$

$$(\theta, \lambda, \mu) \neq (0, 0, 0). \tag{4}$$

The non-zero vector (θ, λ, μ) satisfying (2)-(4) is known as the Fritz-John multiplier vector.

Following the classification in [2], given $x^* \in \Omega$ we say that x^* is weak- regular if there exists (θ, λ, μ) satisfying (2)-(4) with $\theta \neq 0$. Observe that a weak-regular point is a Karush-Kuhn-Tucker point when $r = 1$.

2 APPROXIMATE GRADIENT PROPERTY FOR MULTIOBJECTIVE PROBLEMS

In this section we will extend rigorously the AGP condition for multiobjective problems and we will analyze its properties.

Definition 1 *Given $\gamma \in [-\infty, 0]$. We say that $x^* \in \Omega$ satisfies the Multiobjective Approximate Gradient Projection $(MAGP(\gamma))$ condition, if and only if, there exist sequences $\{\theta^k\} \subset \mathbb{R}^r$, $\{x^k\}$ such that $\theta^k \geqslant 0$, $\sum_{l=1}^{r} \theta_l^k = 1$, $x^k \to x^*$ and*

$$d(x^k, \gamma) = P_{\Omega(x^k, \gamma)}(x^k - \sum_{l=1}^{r} \theta_l^k \nabla f_l(x^k)) - x^k \to 0$$

where $P_{\Omega(x^k, \gamma)}$ is the orthogonal projection and $\Omega(x, \gamma)$ is the polyhedron defined by the points $z \in \mathbb{R}^n$ such that:

1. $g_j(x) + \nabla g_j(x)^T(z - x) \leqslant 0$ *if* $\gamma < g_j(x) < 0$,

2. $\nabla g_j(x)^T(z - x) \leqslant 0$ *if* $g_j(x) \geqslant 0$,

3. $\nabla h_i(x)^T(z - x) = 0$ *for* $i = 1, \ldots, m$.

Observe that $\Omega(x, 0) = \{z \in \mathbb{R}^n | \nabla h(x)^T(z - x) = 0, \nabla g_j(x)^T(z - x) \leqslant 0 \text{ if } g_j(x) \geqslant 0, j = 1, \ldots, p\}$. In order to establish necessary optimality conditions for problem (1) we are going to scalarize it. For this aim, we consider the non-smooth function $\phi : \mathbb{R}^r \to \mathbb{R}$ defined by

$$\phi(y) = \max_{1 \leq l \leq r}\{y_l\}.$$

As ϕ is a non-smooth function we introduce the so-called Clarke subdifferential of ψ (see [3]): The upper Clarke directional derivative of a locally Lipschitz function $\psi : \mathbb{R}^n \to \mathbb{R}$ at $x \in \mathbb{R}^n$ in the direction $d \in \mathbb{R}^n$ is

$$\psi^o(x, d) = \limsup_{y \to x, t \downarrow 0} \frac{\psi(y + td) - \psi(y)}{t}$$

and the Clarke subdifferential of ψ at x is given by

$$\partial_C \psi(x) = \{\xi \in \mathbb{R}^n : \langle \xi, d \rangle \leqslant \psi^o(x, d) \forall d \in \mathbb{R}^n\},$$

where $\langle \cdot, \cdot \rangle$ stands for the Euclidean scalar product.

Lemma 1 *If x^* is a local weak Pareto optimal point of (1), then x^* is a local minimizer of the problem Minimize $\phi(F(x) - F(x^*))$ subject to $x \in \Omega$.*

The following result says that, every local weak Pareto solution of (1) satisfies the $MAGP(\gamma)$ condition.

Theorem 2 *If $x^* \in \Omega$ is a local weak Pareto solution of problem (1), then x^* satisfies the $MAGP(\gamma)$ condition.*

Proof. Let $\varepsilon, \delta > 0$, and $\rho \in (0, \delta)$. By assumption and Lemma 1 we can conclude that x^* is a global minimizer of $\phi(F(x) - F(x^*))$ on $\Omega \cap B(x^*, \rho)$. In consequence, we may suppose that for δ small enough, x^* is the unique solution of problem

$$\text{Minimize } \phi(F(x) - F(x^*)) + \frac{\varepsilon}{2\rho}\|x - x^*\|^2, \quad \text{subject to } x \in \Omega, \|x - x^*\| \leq \rho.$$

We define, for each $k \in \mathbb{N}$,

$$\psi_k(x) = \phi(F(x) - F(x^*)) + \frac{\varepsilon}{2\rho}\|x - x^*\|^2 + \sum_{i=1}^m kh_i(x)^2 + \sum_{j=1}^p kg_j(x)_+^2,$$

and let x^k be the solution of the problem

$$\text{Minimize } \psi_k(x), \text{ subject to } \|x - x^*\| \leq \rho. \tag{5}$$

By using the theory of external penalty, see for example [5, 6, 8], we have that $\|x^k - x^*\| < \delta$ for all k sufficiently large, $x^k \to x^*$ and x^k is a solution of problem (5) and it is an interior point of the feasible set, for k large enough. From Proposition 2.1 (ii) of [5] it follows that $0 \in \partial_C \psi_k(x^k)$. By applying Proposition 2.1 of [5], parts (i) and (iii), we have

$$0 \in co\left(\bigcup_{1 \leq l \leq r}\{\nabla f_l(x^k)\}\right) + \frac{\varepsilon}{\rho}(x^k - x^*) + \sum_{i=1}^m kh_i(x^k)\nabla h_i(x^k) + \sum_{j=1}^p kg_j(x^k)_+\nabla g_j(x^k),$$

where $co\left(\bigcup_{1 \leq l \leq r}\{\nabla f_l(x^k)\}\right)$ denote the convex hull. Hence, there exist $\theta_l^k \geq 0$, $l = 1, \ldots, r$ such that $\sum_{l=1}^r \theta_l^k = 1$ and

$$\sum_{l=1}^r \theta_l^k \nabla f_l(x^k) + k\left(\sum_{i=1}^m h_i(x^k)\nabla h_i(x^k) + \sum_{j=1}^p g_j(x^k)_+\nabla g_j(x^k)\right) = \frac{\varepsilon}{\rho}(x^* - x^k).$$

Then, since $\|x^* - x^k\| < \rho < \delta$,

$$\left\|\sum_{l=1}^r \theta_l^k \nabla f_l(x^k) + k\left(\sum_{i=1}^m h_i(x^k)\nabla h_i(x^k) + \sum_{j=1}^p g_j(x^k)_+\nabla g_j(x^k)\right)\right\| \leq \varepsilon.$$

So,

$$\left\|\left[x^k - \sum_{l=1}^r \theta_l \nabla f_l(x^k)\right] - \left[x^k + k\left(\sum_{i=1}^m h_i(x^k)\nabla h_i(x^k) + \sum_{j=1}^p g_j(x^k)_+\nabla g_j(x^k)\right)\right]\right\| \leq \varepsilon.$$

This implies, taking projections onto $\Omega(x^k, \gamma)$, that

$$\left\|P_{\Omega(x^k,\gamma)}\left(x^k - \sum_{l=1}^r \theta_l \nabla f_l(x^k)\right) - P_{\Omega(x^k,\gamma)}\left(x^k + k\left(\sum_{i=1}^m h_i(x^k)\nabla h_i(x^k) + \sum_{j=1}^p g_j(x^k)_+\nabla g_j(x^k)\right)\right)\right\| \leq \varepsilon.$$

It remains to prove that $P_{\Omega(x^k,\gamma)}\left(x^k + k\left(\sum_{i=1}^m h_i(x^k)\nabla h_i(x^k) + \sum_{j=1}^p g_j(x^k)_+\nabla g_j(x^k)\right)\right) = x^k$.

To see that , we consider the Karush-Kuhn-Tucker conditions of the subproblem with linear constraints

$$\text{Minimize }_y \left\|y - \left[x^k + k\left(\sum_{i=1}^m h_i(x^k)\nabla h_i(x^k) + \sum_{j=1}^p g_j(x^k)_+\nabla g_j(x^k)\right)\right]\right\|^2$$

$$\text{subject to } y \in \Omega(x^k, \gamma)$$

and observe that $y = x^k$ satisfies the KKT optimality conditions with multipliers $\lambda_i = kh_i(x^k)$ and $\mu_j = kg_j(x^k)$ with $\mu_j = 0$ when $g_j(x^k) \geq 0$. So, $\left\| P_{\Omega(x^k, \gamma)} \left(x^k - \sum_{l=1}^r \theta_l^k \nabla f_l(x^k) \right) - x^k \right\| \leq \varepsilon$ as we wanted to prove. $\square$

Note that, from Theorem 2, we can ensure that $MAGP(\gamma)$ is a necessary sequential optimality condition for any local weak Pareto solution.

In the following theorem we establish that, under the Cone-Continuity Property (CCP), the MAGP condition implies weak-regularity.

Given $x^* \in \mathbb{R}^n$ such that $h(x^*) = 0, g(x^*) \leq 0$, we consider the closed convex cone defined as

$$ K(x) = \left\{ \sum_{i=1}^m \lambda_i \nabla h_i(x) + \sum_{j \in A(x^*)} \mu_j \nabla g_j(x) : \quad \lambda_i \in \mathbb{R}, \mu_j \in \mathbb{R}, \mu_j \geqslant 0 \right\} $$

where $A(x^*) = \{j = 1, \ldots, p : g_j(x^*) = 0\}$ is the set of active inequality constraints at x^*.

Given a set-valued mapping (multifunction) $F : \mathbb{R}^s \rightrightarrows \mathbb{R}^d$, the outer limit of $F(z)$ as $z \to z^*$ is denoted by

$$ \limsup_{z \to z^*} F(z) = \{w^* \in \mathbb{R}^d : \exists\, (z^k, w^k) \to (z^*, w^*) \text{ with } w^k \in F(z^k)\}. $$

Then, the multifunction F is said to be outer semicontinuous at z^* if $\limsup_{z \to z^*} F(z) \subset F(z^*)$.

Definition 2 *[1] We say that $x^* \in \mathbb{R}^n$ such that $h(x^*) = 0, g(x^*) \leq 0$ satisfies the Cone-Continuity Property (CCP) if the multifunction $x \rightrightarrows K(x)$ is outer semicontinuous at x^*, that is*

$$ \limsup_{x \to x^*} K(x) \subset K(x^*). $$

Theorem 3 *Let x^* be a feasible point and $\gamma \in [-\infty, 0]$. Assuming that the $MAGP(\gamma)$ condition holds at x^*, and furthermore, suppose that x^* satisfies CCP. Then, x^* is a weak-regular point of the problem (1).*

In [8] it is proved that AGP is a sufficient condition for the scalar problem. In the following theorem we extend the same result in the multiobjective case.

Theorem 4 *Suppose that, in the multiobjective optimization problem (1), f_l, $l = 1, \ldots, r$ and g_i $i = 1, \ldots, p$ are convex and h is an affine function. Let $\gamma \in [-\infty, 0]$. Suppose that $x^* \in \Omega$, $\{x^k\} \subset \mathbb{R}^n$ and $\{\theta^k\} \subset \mathbb{R}^r$ are such that $\sum_{l=1}^r \theta_l^k = 1$, $\theta_l^k \geqslant 0$, $x^k \to x^*$, $h(x^k) = 0$ for all $k = 0, 1, 2, \ldots$ and $d(x^k, \gamma) = 0$. Then, x^* is a weak Pareto solution of (1).*

REFERENCES

[1] R. ANDREANI, J. M. MARTÍNEZ, A. RAMOS, AND P. J. S. SILVA, *A cone-continuity constraint qualification and algorithmic consequences,* SIAM Journal on Optimization, 26(1):96-110, 2016.

[2] G. BIGI, M. PAPPALARDO, *Regularity conditions in vector optimization*, Journal of Optimization Theory and Applications, 102:83-96, 1999.

[3] F.H.CLARKE, *Optimization and Nonsmooth Analysis*, John Wiley & Sons, New York, 1983.

[4] N.O. DA CUNHA, E. POLAK, *Constraint minimization under vector-value criteria in finite dimensional spaces*, Journal of Mathematical Analysis and Applications, 19:103-124, 1967.

[5] G. GIORGI, B. JIMÉNEZ, V. NOVO, *Approximate Karush-Kuhn-Tucker Condition in Multiobjective Optimization*,Journal of Optimization Theory and Applications, 171 , no. 1, 70-89, 2016.

[6] G. HAESER, M. L. SCHUVERDT, *On approximate KKT condition and its extension to continuous variational inequalities*, Journal of Optimization Theory and Applications, 23:707-716, 2011.

[7] J.M. MARTÍNEZ, E.G. BIRGIN, *Practical Augmented Lagrangian Methods for Constrained Optimization*, SIAM Publications, 2014.

[8] J.M. MARTÍNEZ, B.F. SVAITER , *A practical optimality condition without constraint qualifications for nonlinear programming.* J. Optim. Theory Appl. 118, 117-133,2003.

[9] V. PARETO, *Cours d'Economie Politique, Rouge, Lausanne, Switxerland,* 1986.