A new bound on the multipliers given by Carathéodory's theorem and a result on the internal penalty method *

Gabriel Haeser †

June 30, 2009.

Abstract

Carathéodory's theorem for cones states that if we have a linear combination of vectors in $\mathbb{R}^n$, we can rewrite this combination using a linearly independent subset. This theorem has been successfully applied in nonlinear optimization in many contexts. In this work we present a new version of this celebrated theorem, in which we prove a bound for the size of the scalars in the linear combination and we provide examples where this bound is useful. We also prove that the convergence property of the internal penalty method cannot be improved.

Key words: Nonlinear Programming, Constraint Qualifications, Internal Penalty Method.
AMS Subject Classification: 90C30, 49K99, 65K05.

1 Introduction

In 1911 Carathéodory proved that if a point $x \in \mathbb{R}^n$ lies on the convex hull of a compact set P, then x lies on the convex hull of a subset P' of P with no more than $n+1$ points [6]. In 1914 Steinitz generalized this result for a general set P [17].

Here we will see a different version of Carathéodory's theorem, which appears in [5], and we will provide bounds on the size of the multipliers given by the theorem. We address the following nonlinear optimization problem:

$$\text{Minimize } f(x) \quad \text{Subject to} \quad h(x) = 0, \ \ g(x) \leq 0, \tag{1}$$

where $f : \mathbb{R}^n \to \mathbb{R}, h : \mathbb{R}^n \to \mathbb{R}^m$ and $g : \mathbb{R}^n \to \mathbb{R}^p$ are continuously differentiable functions. Under a given constraint qualification, the solution x^* satisfies the KKT condition, that is, x^* is feasible with respect to equality and inequality constraints and there exist $\lambda \in \mathbb{R}^m$ and $\mu_j \geq 0$ for every $j \in I(x^*) = \{i \in \{1, \dots, p\} | g_i(x^*) = 0\}$ such that

$$\nabla f(x^*) + \sum_{i=1}^{m} \lambda_i \nabla h_i(x^*) + \sum_{j \in I(x^*)} \mu_j \nabla g_j(x^*) = 0.$$

A common constraint qualification usually employed is the Linear Independence constraint qualification, which states that $\{\nabla h_i(x^*)\}_{i=1}^m \cup \{\nabla g_j(x^*)\}_{j \in I(x^*)}$ is linearly independent. We will refer to this multi-set as the *active set of gradients at* x^*. The weaker Mangasarian-Fromovitz

*This work was supported by FAPESP Grant 05/02163-8.
†Department of Applied Mathematics, Institute of Mathematics, Statistics and Scientific Computing, University of Campinas, Campinas SP, Brazil. E-mail: ghaeser@ime.unicamp.br

constraint qualification (MFCQ) [13] states that the active set of gradients is positive-linearly independent, which means that there is no $\alpha \in \mathbb{R}^m$, $\beta_j \geq 0$ for every $j \in I(x^*)$ such that

$$\sum_{i=1}^{m} \alpha_i \nabla h_i(x^*) + \sum_{j \in I(x^*)} \beta_j \nabla g_j(x^*) = 0,$$

except if we take all α_i and β_j equal to zero. In fact, this is a reformulation of the original definition, given in [15].

Recently, a weaker constraint qualification appeared in the literature: the Constant Positive Linear Dependence constraint qualification (CPLD) [14, 4], which has been successfully applied to obtain new practical algorithms [1, 2, 9]. We say that the CPLD condition holds for a feasible x^* if for every $I \subset \{1, \ldots, m\}$, $J \subset I(x^*)$ such that the set of gradients $\{\nabla h_i(x^*)\}_{i \in I} \cup \{\nabla g_j(x^*)\}_{j \in J}$ is positive-linearly dependent, there exists a neighborhood $V(x^*)$ of x^* such that the set of gradients $\{\nabla h_i(y)\}_{i \in I} \cup \{\nabla g_j(y)\}_{j \in J}$ remains positive-linearly dependent for every $y \in V(x^*)$. The CPLD condition is a natural generalization of the Constant Rank constraint qualification of Janin [12], which states the same as above, replacing "positive-linearly dependent" by "linearly dependent". The CPLD condition is weaker than the Constant Rank condition [16].

In practical algorithms, weaker constraint qualifications are preferred, since convergence results are stronger.

In section 2 we will state Carathéodory's theorem and obtain new bounds on the size of the multipliers. Examples of applications of the new result will be given. In section 3 we will prove the impossibility of generalizing the convergence result of the internal penalty method using the CPLD condition.

2 Generalized Carathéodory's theorem for cones

The main tool which enables us to prove convergence results under the CPLD condition is Carathéodory's theorem for cones which appears in [5]. We state it here with the new bounds given by item *4*.

Theorem 1. *If* $x = \sum_{i=1}^{m} \alpha_i v_i$ *with* $v_i \in \mathbb{R}^n$ *and* $\alpha_i \neq 0$ *for every* i*, then there exist* $I \subset \{1, \ldots, m\}$ *and scalars* $\bar{\alpha}_i$ *for every* $i \in I$ *such that*

1. $x = \sum_{i \in I} \bar{\alpha}_i v_i$*;*
2. $\alpha_i \bar{\alpha}_i > 0$ *for every* $i \in I$*;*
3. $\{v_i\}_{i \in I}$ *is linearly independent;*
4. $|\bar{\alpha}_i| \leq 2^{m-1} |\alpha_i|$ *for every* $i \in I$*.*

Proof. We assume that $\{v_i\}_{i=1}^m$ is linearly dependent, otherwise the result follows trivially. Then, there exists $\beta \in \mathbb{R}^m$, $\beta \neq 0$ such that $\sum_{i=1}^m \beta_i v_i = 0$. Thus, we may write

$$x = \sum_{i=1}^{m} (\alpha_i - \gamma \beta_i) v_i,$$

for every $\gamma \in \mathbb{R}$. Let $i^* = \operatorname{argmin}_i \left| \frac{\alpha_i}{\beta_i} \right|$ and $\bar{\gamma} = \frac{\alpha_{i^*}}{\beta_{i^*}}$, then $\bar{\gamma}$ is the least modulus coefficient $\frac{\alpha_i}{\beta_i}$. Note that $\bar{\gamma}$ is such that $\alpha_i - \bar{\gamma}\beta_i = 0$ for at least one index $i = i^*$. If $\alpha_i(\alpha_i - \bar{\gamma}\beta_i) < 0$, then

$|\alpha_i|^2 = \alpha_i^2 < \alpha_i \bar{\gamma} \beta_i = |\alpha_i||\bar{\gamma}||\beta_i|$, with $\alpha_i \neq 0, \beta_i \neq 0$, thus $|\bar{\gamma}| > \left|\frac{\alpha_i}{\beta_i}\right|$ which contradicts the definition of $\bar{\gamma}$. Therefore we conclude that $\alpha_i(\alpha_i - \bar{\gamma}\beta_i) \geq 0$. Also, $|\alpha_i - \bar{\gamma}\beta_i| \leq |\alpha_i| + |\bar{\gamma}||\beta_i| \leq 2|\alpha_i|$, since $\bar{\gamma} \leq \left|\frac{\alpha_i}{\beta_i}\right|$ for every i. Including in the sum only the indexes such that $\bar{\alpha}_i = \alpha_i - \bar{\gamma}\beta_i \neq 0$ we are able to write the linear combination x with at least one less vector. We can repeat this procedure until $\{v_i\}_{i \in I}$ is linearly independent with $\alpha_i \bar{\alpha}_i > 0$ and $|\bar{\alpha}_i| \leq 2^{m-1}|\alpha_i|$ for every $i \in I$. □

We usually apply Carathéodory's theorem when we have a sequence $\{x^k\}$ converging to a feasible point x^* that satisfies a quasi-KKT condition of the form

$$\nabla f(x^k) + \sum_{i=1}^{m} \lambda_i^k \nabla h_i(x^k) + \sum_{j \in I(x^*)} \mu_j^k \nabla g_j(x^k) = \varepsilon_k, \tag{2}$$

where $\lambda^k \in \mathbb{R}^m$, $\mu_j^k \geq 0$ for every $j \in I(x^*)$ and every $k \in \mathbb{N}$, with $\|\varepsilon_k\| \to 0$.

Such a sequence may be obtained, for example, when we apply the external or internal penalty method (see [3]). To prove that the limit point x^* is a KKT point, two cases are usually considered. Define $M_k = \max\{|\lambda_i|, \mu_j, \forall i \in \{1, \ldots, m\}, j \in I(x^*)\}$. If there is a subsequence such that $\{M_k\}$ is bounded, we can obtain a convergent subsequence of $\{\lambda_i^k\}$ and $\{\mu_j^k\}$, then, taking limits and using the continuity of the gradients we obtain that x^* is a KKT point. If, on the other hand, $M_k \to +\infty$, we may divide (2) by M_k, then, the infinity norm of the new multipliers is equal to 1, thus we get a bounded sequence with non-null limit points (by the continuity of the norm). Taking a convergent subsequence we get a non-null linear combination of the active set of gradients, which proves that x^* fails to satisfy the Mangasarian-Fromovitz constraint qualification.

The idea to generalize this kind of result under the CPLD condition is to apply Carathéodory's theorem to equation (2). This gives us, for every k, two sets $I_k \subset \{1, \ldots, m\}, J_k \subset I(x^*)$, new multipliers $\bar{\lambda}_i^k$ for every $i \in I_k$ and $\bar{\mu}_j^k \geq 0$ for every $j \in J_k$ such that

$$\{\nabla h_i(x^*)\}_{i \in I_k} \cup \{\nabla g_j(x^*)\}_{j \in J_k} \text{ is linearly independent}$$

and

$$\nabla f(x^k) + \sum_{i \in I_k} \bar{\lambda}_i^k \nabla h_i(x^k) + \sum_{j \in J_k} \bar{\mu}_j^k \nabla g_j(x^k) = \varepsilon_k. \tag{3}$$

The key point is to observe that we can take a subsequence such that I_k is the same set I for every k and J_k is the same set J for every k. This comes from the finiteness of the possible sets I_k, J_k. Proceeding in the same fashion, defining $\bar{M}_k = \max\{|\bar{\lambda}_i^k|, \bar{\mu}_j^k, \forall i \in I, j \in J\}$ we obtain the KKT condition if there is a bounded subsequence of $\{\bar{M}_k\}$. Otherwise, if $\bar{M}_k \to +\infty$, dividing (3) by $\bar{M}_k$ and taking limits, we have as before that the gradients at x^* are positive-linearly dependent. This proves that x^* fails to satisfy the CPLD condition since the gradients at x^k are linearly independent, with x^k arbitrarily close to x^*. Thus, positive-linear dependence is not maintained in a neighborhood of x^*, for this particular choice of $I \subset \{1, \ldots, m\}, J \subset I(x^*)$. These natural ideas appeared for the first time in the first applications of the CPLD condition [14, 1, 2].

The new bounds $|\bar{\lambda}_i^k| \leq 2^{m-1}|\lambda_i^k|$ for every $i \in I$ and $|\bar{\mu}_j^k| \leq 2^{p-1}|\lambda_j^k|$ for every $j \in J$ may be useful in many ways. For example, if we have that $\{(\lambda^k, \mu^k)\}$ is bounded, then the same is true for the sequence of new multipliers $\{(\bar{\lambda}^k, \bar{\mu}^k)\}$. The converse is not always true. Consider for instance $x^k = \alpha_1^k v_1^k + \alpha_2^k v_2^k$, $v_1^k \neq 0$ with $\beta_1^k v_1^k + \beta_2^k v_2^k = 0$ for $\beta_1^k = \beta_2^k = 1$, $\alpha_1^k = 1 + 10^k, \alpha_2^k = 10^k$. We have $\left|\frac{\alpha_1^k}{\beta_1^k}\right| > \left|\frac{\alpha_2^k}{\beta_2^k}\right|$ for every k, then $\bar{\alpha}_1^k = \alpha_1^k - \left(\frac{\alpha_2^k}{\beta_2^k}\right)\beta_1^k = 1$ and $x^k = \bar{\alpha}_1^k v_1^k$ for every k.

Another common situation in which bounds may be useful is when we have $\mu_j^k \to 0$. This appears for example in the internal penalty method, in which quasi-KKT points are defined as

$$\nabla f(x^k) + \sum_{i=1}^{m} \lambda_i^k \nabla h_i(x^k) + \sum_{j=1}^{p} \mu_j^k \nabla g_j(x^k) = \varepsilon_k, \tag{4}$$

with $\mu_j^k \to 0$ when $g_j(x^*) < 0$. With the new bounds, we have that $\bar{\mu}_j^k \to 0$ whenever $\mu_j^k \to 0$, and this is crucial to obtain the complementarity condition $\mu_j g_j(x^*) = 0$ of the KKT condition. The reciprocal is also not true. This can be observed by taking the previous counter-example with α_1^k and α_2^k divided by 10^k.

3 Internal Penalty

In this section we will consider problem (1) with only inequality constraints:

$$\text{Minimize } f(x) \quad \text{Subject to} \quad g(x) \leq 0. \tag{5}$$

The internal penalty method consists of solving the following subproblem:

$$\text{Minimize } f(x) + r_k \sum_{j=1}^{p} \frac{1}{g_j(x)} \quad \text{Subject to} \quad g(x) < 0, \tag{6}$$

for a sequence of positive scalars $r_k \to 0$. If there are additional constraints $x \in \Omega$, they are added to the constraints of the subproblem.

It is a well known fact that if x^* is a limit point of the sequence x^k generated by the interior penalty method, such that x^* satisfies the sufficient interior property, that is, x^* can be approximated by a sequence of strictly feasible points $y^k \to x^*$ ($g(y^k) < 0$), then x^* is a solution to problem (5) [7].

If x^* is a local solution of problem (5) and we apply the internal penalty method to:

$$\text{Minimize } f(x) + \frac{1}{2}\|x - x^*\|_2^2 \quad \text{Subject to} \quad ||x - x^*|| \leq \delta, \;\; g(x) \leq 0, \tag{7}$$

for a sufficiently small δ. The corresponding subproblem is:

$$\text{Minimize } f(x) + \frac{1}{2}\|x - x^*\|_2^2 + r_k \sum_{j=1}^{p} \frac{1}{g_j(x)} \quad \text{Subject to} \quad ||x - x^*||_2 \leq \delta, \;\; g(x) < 0. \tag{8}$$

We can write the KKT condition for the subproblem in the solution x^k, and we will arrive at the quasi-KKT condition given by (4), but without the terms depending on h. See details in [3].

According to the proof given in the previous section, we have that under the CPLD condition and the sufficient interior property, limit points of the interior penalty method are KKT point. We will prove that these hypotheses are equivalent to the Mangasarian-Fromovitz condition.

For this purpose we shall define the quasi-normality constraint qualification proposed by [11].

Definition. *We say that a feasible point x^* to problem (5) satisfies the quasi-normality constraint qualification if x^* satisfies MFCQ, or if $\mu_j \geq 0$ exists for every $j \in I(x^*)$, not all zero, with $\sum_{j \in I(x^*)} \mu_j \nabla g_j(x^*) = 0$ then a sequence $z^k \to x^*$ does not exist, such that $\mu_j > 0 \Rightarrow g_j(z^k) > 0$ for every $j \in I(x^*)$.*

We will use the result proved in [4] that CPLD implies quasi-normality.

Theorem 2. *A feasible point x^* satisfies CPLD and the sufficient interior property if, and only if, x^* satisfies MFCQ.*

Proof. Suppose a feasible x^* satisfies the CPLD condition and the sufficient interior property. Then x^* satisfies the CPLD condition for the problem:

$$\text{Minimize } f(x) \quad \text{Subject to} \quad -g_i(x) \leq 0, \quad \forall i \in I(x^*), \tag{9}$$

therefore x^* satisfies the quasi-normality condition for problem (9). If MFCQ does not hold, then not all zero scalars $\mu_j \geq 0$ exist such that $\sum_{j \in I(x^*)} \mu_j \nabla g_j(x^*) = 0$, multiplying by -1 we get that MFCQ does not hold for problem (9). Thus, by the quasi-normality for this problem we obtain that there is no sequence $z^k \to x^*$ such that $\mu_j > 0 \Rightarrow -g_j(z^k) > 0$ for every $j \in I(x^*)$. Since at least one index $j \in I(x^*)$ exists such that $\mu_j > 0$, we conclude that there is no sequence $z^k \to x^*$ such that $g_j(z^k) < 0$, which contradicts the sufficient interior property.

The converse holds trivially since one can easily prove that the sufficient interior property holds using the direction given by the original MFCQ definition, see details in [8, 10]. Clearly, MFCQ also implies the CPLD condition. □

This shows that the internal penalty method converges to a KKT point under MFCQ, and relaxing this condition to CPLD does not provide a stronger result. This is clear since we cannot expect convergence of the internal penalty method if the sufficient interior property does not hold.

We conclude with a counter-example showing that a stronger form of Theorem 2, in which CPLD is replaced by quasi-normality, does not hold. Consider the problem:

$$\text{Minimize } x \quad \text{Subject to} \quad -x^2 \leq 0,$$

at the point $x^* = 0$. It is clear that MFCQ does not hold and the sufficient interior property holds. Also the quasi-normality condition holds since there is no infeasible point.

References

[1] R. Andreani, E.G. Birgin, J.M. Martínez, and M.L. Schuverdt. On augmented lagrangian methods with general lower–level constraints. *SIAM Journal on Optimization*, 18:1286–1309, 2007.

[2] R. Andreani, E.G. Birgin, J.M. Martínez, and M.L. Schuverdt. Augmented lagrangian methods under the constant positive linear dependence constraint qualification. *Mathematical Programming*, 112:5–32, 2008.

[3] R. Andreani, G. Haeser, and J.M. Martínez. On sequential optimality conditions. *submitted*, 2009. Available in Optimization Online http://www.optimization-online.org/DB_HTML/2009/06/2326.html.

[4] R. Andreani, J.M. Martínez, and M.L. Schuverdt. On the relation between constant positive linear dependence condition and quasinormality constraint qualification. *Journal of Optimization Theory and Applications*, 125:473–485, 2005.

[5] D.P. Bertsekas. *Nonlinear Programming: 2nd Edition.* Athena Scientific, 1999.

[6] C. Carathéodory. über den variabilitätsbereich der fourierschen konstanten von positiven harmonischen funktionen. *Rend. Circ. Mat. Palermo*, 32:193–217, 1911.

[7] A.V. Fiacco and G.P. McCormick. *Nonlinear Programming: Sequential Unconstrained Minimization Techniques.* Wiley, 1968. (reprinted, SIAM, 1990).

[8] A. Forsgren, P.E. Gill, and M.H. Wright. Interior methods for nonlinear optimization. *SIAM Review*, 44(4):525–597, 2002.

[9] M.A. Gomes-Ruggiero, J.M. Martínez, and S.A. Santos. Spectral projected gradient method with inexact restoration for minimization with nonconvex constraints. *SIAM Journal on Scientific Computing*, 31(3):1628–1652, 2009.

[10] G. Haeser. PhD thesis, IMECC-UNICAMP, Departamento de Matemática Aplicada, Campinas-SP, Brazil, 2009.

[11] M.R. Hestenes. *Optimization theory: the finite dimensional case.* John Wiley, 1975.

[12] R. Janin. Directional derivative of the marginal function in nonlinear programming. *Mathematical Programming Study*, 21:110–126, 1984.

[13] O.L. Mangasarian and S. Fromovitz. The Fritz John optimality conditions in the presence of equality and inequality constraints. *Journal of Mathematical Analysis and Applications*, 17:37–47, 1967.

[14] L. Qi and Z. Wei. On the constant positive linear dependence condition and its applications to SQP methods. *SIAM Journal on Optimization*, 10(4):963–981, 2000.

[15] R.T. Rockafellar. Lagrange multipliers and optimality. *SIAM Review*, 35(2):183–238, 1993.

[16] M.L. Schuverdt. *Métodos de Lagrangiano aumentado com convergência utilizando a condição de dependência linear positiva constante.* PhD thesis, IMECC-UNICAMP, Departamento de Matemática Aplicada, Campinas-SP, Brazil, 2006. http://libdigi.unicamp.br/document/?code=vtls000375801.

[17] E. Steinitz. Bedingt konvergente reihen und konvexe systeme i-iv. *J. Reine Angew. Math.*, 143:128–175, 1913.