Karush-Kuhn-Tucker conditions*

María L. Schuverdt† Leonardo D. Secchin‡

1 Introduction

Let us consider the constrained nonlinear programming problem (NLP)

$$\begin{aligned} &\text{minimize } f(x) \\ &\text{subject to } x \in \Omega \subset \mathbb{R}^n \end{aligned} \tag{P}$$

where

$$\Omega = \{x \in \mathbb{R}^n \mid h(x) = 0,\ g(x) \leq 0\} \tag{1}$$

is the feasible set. We assume that $f : \mathbb{R}^n \to \mathbb{R}$, $h_i : \mathbb{R}^n \to \mathbb{R}$, $i = 1, \ldots, m$, $g_j : \mathbb{R}^n \to \mathbb{R}$, $j = 1, \ldots, p$ are continuously differentiable functions in an open set containing Ω. For each $x \in \Omega$, $A(x) = \{j \mid g_j(x) = 0\}$ is the set of indexes of active inequality constraints.

A local minimizer of (P) is a feasible point x^* such that $f(x^*) \leq f(x)$ for all feasible x closer to x^*. Solving (P) consists in find such points, but this is not possible in practice since it involves the evaluation of f in an uncountable set. Thus, it is suitable to deal with computationally verifiable conditions that describe some property satisfied by every local minimizer x^*. These are the *(necessary) optimality conditions*.

A good optimality condition should be easy to verify and strong as possible. In this sense, "$x^* \in \mathbb{R}^n$" and "x^* is a local minimizer" are not adequate: the first does not describe well local minimizers, while the second is too hard to check. To obtain useful conditions, we should consider the behavior of f and the structure of the feasible set Ω whenever possible. In this context, a key fact is that for every local minimizer x^* of (P), the objective function f does not decrease locally by moving along feasible directions. This geometry can be characterized by using the rich theory of cones.

For a general NLP with inequality and equality constraints, we consider the cone of *tangent directions* of Ω at x

$$\mathcal{T}(x) = \left\{ d \in \mathbb{R}^n \mid \exists (x^k) \subset \Omega : x^k \to x,\ x^k \neq x,\ \frac{x^k - x}{\|x^k - x\|} \to \frac{d}{\|d\|} \right\} \cup \{0\}.$$

$\mathcal{T}(x)$ includes all directions for which small movements from x result in feasible points (*feasible directions*). Another cone is that of *descent directions*, which consists of directions in which f

*This text was published in IFORS News, 16(2):16–18, June 2021, ISSN 2223-4373. URL `https://www.ifors.org/newsletter/ifors-news-june-2021.pdf`.

†CONICET, Department of Mathematics, FCE, University of La Plata, CP 172, 1900 La Plata Bs. As., Argentina. Email: `schuverd@mate.unlp.edu.ar`.

‡Department of Applied Mathematics, Federal University of Espírito Santo, São Mateus, Espírito Santo, Brazil. Email: `leonardo.secchin@ufes.br`.

decreases locally, $\mathcal{V}(x) = \{d \in \mathbb{R}^n \mid \exists \varepsilon > 0 : f(x+td) < f(x), \forall t \in (0,\varepsilon]\}$. However, due to the difficult in dealing with $\mathcal{V}(x)$, it is common to consider the restricted simple set

$$\mathcal{D}(x) = \{d \in \mathbb{R}^n \mid \nabla f(x)^T d < 0\} \subset \mathcal{V}(x).$$

Therefore, the following necessary condition for local optimality is intuitive: if x^* is a local minimizer of (P) then

$$\nabla f(x^*)^T d \geq 0 \text{ for all } d \in \mathcal{T}(x^*), \tag{2}$$

which is equivalent to $\mathcal{D}(x^*) \cap \mathcal{T}(x^*) = \emptyset$. Another convenient way to write (2) is

$$-\nabla f(x^*) \in \mathcal{T}^\circ(x^*), \tag{3}$$

where $C^\circ = \{y \in \mathbb{R}^n \mid y^T d \leq 0, \ \forall d \in C\}$ is the *polar* of set $C \subset \mathbb{R}^n$. See Figure 1(a).

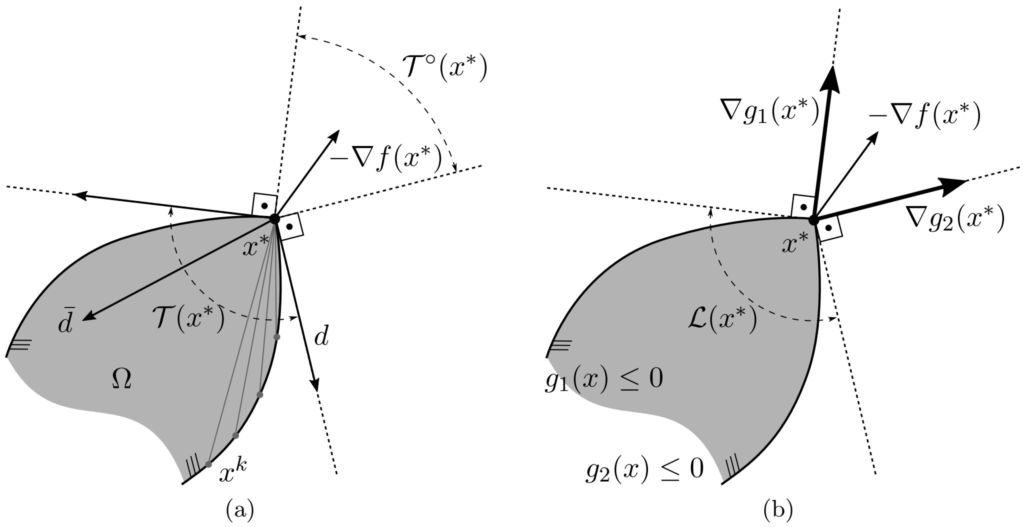

Figure 1: (a) $d, \bar{d} \in \mathcal{T}(x^*)$. The direction d is tangent, while, in addition, $\bar{d}$ is feasible. We have $-\nabla f(x^*) \in \mathcal{T}^\circ(x^*)$, and it is not possible to decrease f along $d \in \mathcal{T}(x^*)$ maintaining feasibility; (b) $\mathcal{L}(x^*)$ is the cone of directions that do not make obtuse angles with $\nabla g_1(x^*)$ or $\nabla g_2(x^*)$.

Condition (2) gives an authentic criterion to check if a feasible point is not a local minimizer of (P), but $\mathcal{T}(x^*)$ is still difficult, perhaps impossible, to compute. As we have already mentioned, we want to obtain a practical and computable condition. Taking advantage of the algebraic description (1) of Ω, we consider the linearization of Ω at x

$$\mathcal{L}(x) = \{d \in \mathbb{R}^n \mid \nabla h_i(x)^T d = 0, \, i = 1, \ldots, m; \ \nabla g_j(x)^T d \leq 0, \, j \in A(x)\}.$$

See Figure 1(b). Note that only active constraints are present, since the inactive ones do not affect Ω locally around x. Also, note that the linearized cone $\mathcal{L}(x)$ is computationally tractable since it involves only the point x and a system of linear equalities/inequalities.

It is straightforward to verify that $\mathcal{T}(x) \subset \mathcal{L}(x)$ for all $x \in \Omega$, but the contrary inclusion may not happen (for example, if $\Omega = \{x \in \mathbb{R} \mid x^2 = 0\}$ then $\mathcal{L}(0) = \mathbb{R} \not\subset \{0\} = \mathcal{T}(0)$). So, we can not change in principle $\mathcal{T}(x^*)$ by $\mathcal{L}(x^*)$ in (2) without an additional assumption, otherwise there may exist local minimizers that do not fulfil the resulting property. Nevertheless, besides its simplicity, $\mathcal{L}(x)$ gives the following useful property:

Theorem 1. *If $x^* \in \Omega$ satisfies $\nabla f(x^*)^T d \geq 0$ for all $d \in \mathcal{L}(x^*)$ then there exist scalars $\lambda_1, \ldots, \lambda_m, \mu_1, \ldots, \mu_p$ such that*

$$\nabla f(x^*) + \sum_{i=1}^{m} \lambda_i \nabla h_i(x^*) + \sum_{i=1}^{p} \mu_i \nabla g_i(x^*) = 0, \tag{4a}$$

$$h(x^*) = 0, \; g(x^*) \leq 0, \tag{4b}$$

$$\mu_i \geq 0, \; \mu_i g_i(x^*) = 0, \quad i = 1, \ldots, p. \tag{4c}$$

Theorem 1 can be proved by the Farkas' lemma. In fact, it can be stated that $\mathcal{L}(x^*)$ is equal to the polar of the cone generated by the gradients of active constraints [3], or equivalently, that

$$\mathcal{L}^\circ(x^*) = \left\{ \sum_{i=1}^{m} \lambda_i \nabla h_i(x^*) + \sum_{i=1}^{p} \mu_i \nabla g_i(x^*) \;\middle|\; \mu_i \geq 0, \; \mu_i g_i(x^*) = 0, \; i = 1, \ldots, p \right\}.$$

Expression (4) is known as *Karush-Kuhn-Tucker (KKT) conditions.* The vectors $\lambda \in \mathbb{R}^m$ and $\mu \in \mathbb{R}^p_+$ are called *Lagrange multipliers.* The origin of KKT conditions dates back to 1939 with the unpublished master's dissertation written by Karush. They were rediscovered in 1951 in an independent work of Kuhn and Tucker. Curiously, Karush's work remained forgotten for several years. See [5]. KKT is one of the most important concepts in non-linear continuous optimization; it also has been specialized for many particular important classes of problems, such as multiobjective, Nash equilibrium problems and even for non-smooth optimization.

We stress that (2) or (3) may not occur at a local minimizer with $\mathcal{L}(x^*)$ instead of $\mathcal{T}(x^*)$. In turn, conditions (4) also may not occur at local minimizers. For example, $x^* = 0$ is the (unique) minimizer of $f(x) = x$ subject to $x^2 = 0$, but in this case (4a) requires that $1 + 0 \cdot \lambda = 0$, which is impossible. Nevertheless, Theorem 1 (i.e., KKT conditions) gives a suitable way to decide if a point x^* is a good candidate for local minimizer of (P). In fact, several practical algorithms (perhaps all) for constrained optimization aim to reach points satisfying (4). So, it is reasonable to consider hypotheses under which changing $\mathcal{T}(x^*)$ by $\mathcal{L}(x^*)$ in (2) or (3) still provide a necessary optimality condition. Due to the nature of $\mathcal{L}(x)$, such assumptions act on the constraints, and then they are called *constraint qualifications* (CQs). Two obvious CQs are $\mathcal{L}(x^*) = \mathcal{T}(x^*)$ and $\mathcal{L}^\circ(x^*) = \mathcal{T}^\circ(x^*)$, which are known as Abadie's CQ and Guignard's CQ, respectively. In time, Guignard's CQ is the weakest hypothesis in the sense of it ensures that x^* fulfils (4) for all smooth functions f that has x^* as local minimizer [4]. Many others CQs were proposed in the literature. Among them, we highlighted the most common in classic books such as [3, 6], the *linear independence of the gradients of the active constraints* (LICQ):

$$\nabla h_i(x^*), \; i = 1, \ldots, m, \quad \nabla g_j(x^*), \; j \in A(x^*) \quad \text{are linearly independent.} \tag{5}$$

It is well known that LICQ implies the uniqueness of Lagrange multipliers. Some other CQs in the literature are: linear/affine constraints (all h_i's and g_j's affine); Mangasarian-Fromovitz CQ – MFCQ (the only linear combination of (5) equal to zero with nonnegative ∇g_i's coefficients is the trivial); Quasinormality (a hypothesis associated with external penalty methods); Constant Rank CQ – CRCQ (each subset of gradients in (5) linearly dependent at x^* remains linearly dependent around x^*); and Constant Positive Linear Dependence CQ – CPLD (a generalization of CRCQ to nonnegative linear combinations). See [1, 2]. Figure 2 summarizes the mentioned CQs. It is worth mentioning that they are not exhaustive. See [1] and references therein.

With one of these additional hypotheses, conditions (4) provided the important practical characterization of local minimizers of (P):

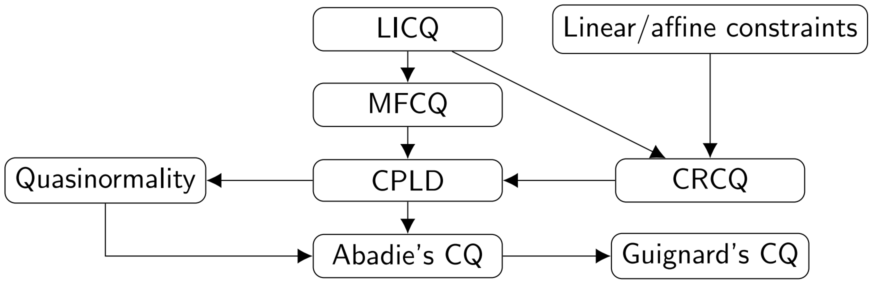

Figure 2: Some CQs and their relationship. An arrow indicates an implication.

Theorem 2. *If x^* is a local minimizer of* (P) *and satisfies some CQ of Figure 2, then x^* is a KKT point, that is, x^* fulfils* (4)*.*

Supposing *a priori* the validity of some CQ, the practical implication of Theorem 2 is that if an algorithm converges to a point x^* for which there are vectors λ and μ satisfying (4), then x^* is a good candidate for local minimizer. Based on this, several methods has been proposed such as external penalty strategies, augmented Lagrangian schemes, sequential quadratic programming and interior point methods. See [6].

We finish with a brief comment on the recently introduced sequential optimality conditions, which are characterizations of minimizers by means of limits of sequences. The goal of such conditions is two-fold: on the one hand, they are satisfied by minimizers of (P) without the necessity of any constraint qualification, such as LICQ; on the other hand, they are associated with the sequences generated by practical algorithms. This means that the behaviour of an algorithm can be treated regardless of the fulfilment of any CQ. Furthermore, when compared to KKT, sequential optimality conditions provide very general CQs based on the continuity of associated set-valued maps. See [1] and references there in.

References

[1] R. Andreani, N. Fazzio, M. Schuverdt, and L. Secchin. A sequential optimality condition related to the quasi-normality constraint qualification and its algorithmic consequences. *SIAM Journal on Optimization*, 29(1):743–766, 2019.

[2] R. Andreani, J. M. Martinez, and M. L. Schuverdt. On the relation between constant positive linear dependence condition and quasinormality constraint qualification. *Journal of Optimization Theory and Applications*, 125(2):473–483, 2005.

[3] M. S. Bazaraa, H. D. Sherali, and C. M. Shetty. *Nonlinear Programming. Theory and Algorithms.* John Wiley & Sons, New Jersey, 3 edition, 2006.

[4] F. J. Gould and J. W. Tolle. A necessary and sufficient qualification for constrained optimization. *SIAM Journal on Applied Mathematics*, 20(2):164–172, 1971.

[5] H. W. Kuhn. Nonlinear programming: A historical view. In C. E. Lemke and R. W. Cottle, editors, *Nonlinear Programming*, volume 9 of *SIAM-AMS Proceedings*, New York, 1976.

[6] J. Nocedal and S. Wright. *Numerical Optimization.* Springer Series in Operations Research and Financial Engineering. Springer-Verlag, New York, 2 edition, 2006.