Noname manuscript No.
(will be inserted by the editor)

Approximate-KKT stopping criterion when Lagrange multipliers are not available

Gabriel Haeser · Vinícius V. de Melo

Received: date / Accepted: date

Abstract In this paper we investigate how to efficiently apply Approximate-Karush-Kuhn-Tucker (AKKT) proximity measures as stopping criteria for optimization algorithms that do not generate approximations to Lagrange multipliers, in particular, Genetic Algorithms. We prove that for a wide range of constrained optimization problems the KKT error measurement tends to zero. We also develop a simple model to compute the KKT error measure requiring only the solution of a non-negative linear least squares problem. Our numerical experiments show the efficiency of the strategy.

Keywords Optimality Conditions · Genetic Algorithms · Stopping Criteria

1 Introduction

Many iterative solvers in nonlinear optimization provide natural Lagrange multipliers approximations at each iteration (Sequential Quadratic Programming, Augmented Lagrangian Methods, Interior Point Methods [8]). These approximations are often used to stop the execution of the algorithm when the Karush-Kuhn-Tucker (KKT) optimality conditions are approximately satisfied. In this paper we will review the basic theory that justifies this procedure and develop a strategy to apply the same reasoning for algorithms that do not generate Lagrange multipliers

This work was supported by FAPESP Grant 2010/19720-5.

G. Haeser (✉)
Federal University of São Paulo, Institute of Science and Technology. São José dos Campos, SP, Brazil.
Tel.: +55-12-3309-9585
Fax: +55-12-3921-8857
E-mail: gabriel.haeser@unifesp.br

V.V. de Melo
Federal University of São Paulo, Institute of Science and Technology. São José dos Campos, SP, Brazil.
Tel.: +55-12-3309-9576
Fax: +55-12-3921-8857
E-mail: vinicius.melo@unifesp.br

approximations. Our attention will be focused on Genetic Algorithms (GA), but the approach is sufficiently general to be applied in any iterative method.

Genetic Algorithms are one of the most employed metaheuristics for global optimization [17,10,9,22]. It is a populational evolutionary algorithm in which two stop conditions are usually employed: value-to-reach or maximum number of generations. The first one can only be used when the user knows the optimal or desired objective function value. However, in several problems this is not the case. That is why the second condition is commonly employed.

If the optimal value is unknown, the user may set the GA to run for more generations than required to reach the global optimum because no verification is performed. In this case, the GA will keep running after having found a solution or get stuck in a local optimum. In this paper, a more adequate strategy to stop the GA by using KKT optimality conditions [8] is proposed. This criterion can also be used to detect if the GA is near to a local optimum, indicating that the algorithm should continue searching in a different region.

Recently, the idea of Approximate-KKT (AKKT, [5]) has been used to define some stopping criteria to a general GA [31]. The idea is to set, at each generation of the GA, a parameter ε that measures how distant the best individual of the population is from fulfilling the KKT optimality conditions. The authors verified in practice that a GA implementation generates individuals arbitrarily close to satisfying the KKT conditions. This was reported with surprise since KKT conditions involve the derivatives of the objective and constraint functions, while GA does not use derivative information. In this paper we will show that this behavior is expected whenever the GA converges to a non-degenerate optimal solution. We prove that this must be true for a fairly broad class of optimization problems. Hence, one can safely stop the execution of the GA when the parameter ε, updated at each generation, is sufficiently close to zero.

This gives rise to the important question of how one can efficiently calculate the parameter ε at each iteration. This paper describes a strategy that requires the solution of a non-negative linear least squares problem. In Section 2, we present a review of AKKT conditions. In Section 3, we prove our result that explains the numerical results from [31] and we describe an idea to efficiently compute a KKT error measure at each iteration. In Section 4, we describe our numerical experiments. In Section 5, we draw some conclusions. All norms in this paper are euclidean norms and $\mathbb{R}_+$ denotes the set of non-negative real numbers.

2 Approximate KKT conditions

Let us consider the general nonlinear optimization problem:

$$\text{Minimize } f(x), \quad \text{subject to} \quad x \in \Omega, \tag{1}$$

where $\Omega = \{x \in \mathbb{R}^n \mid h(x) = 0, g(x) \leq 0\}$, $f : \mathbb{R}^n \to \mathbb{R}$, $h : \mathbb{R}^n \to \mathbb{R}^m$ and $g : \mathbb{R}^n \to \mathbb{R}^p$ are continuously differentiable functions. For each feasible point $x \in \Omega$, we define the set of active inequality constraints $A(x) = \{i \in \{1, \ldots, p\} \mid g_i(x) = 0\}$.

A reasonable criterion for optimality of a feasible point x^* is the fulfillment of the so–called KKT conditions. This means that there exist Lagrange multipliers

$\lambda \in \mathbb{R}^m$ and $\mu \in \mathbb{R}^p_+$ such that

$$\nabla f(x^*) + \sum_{i=1}^{m} \lambda_i \nabla h_i(x^*) + \sum_{i \in A(x^*)} \mu_i \nabla g_i(x^*) = 0.$$

In order to guarantee that the KKT conditions hold at a (local) minimizer one must admit that the description of the feasible set satisfies a constraint qualification [8]. For example, this is the case when the gradients of the equality and active inequality constraints are linearly independent at the minimizer. Many less restrictive constraint qualifications can be found in the literature ([23,29,19], or, more recently [27,6,7]). Finding new constraint qualifications is of great interest for many theoretical and practical reasons [7].

When dealing with implementations of an iterative solver one must choose a reasonable stopping criterion. Even when the sequence of iterates, say $\{x^k\} \subset \mathbb{R}^n$, converges to a point x^* one can not check exactly the KKT conditions since one does not know exactly the limit point x^*. In fact, the limit point can even be non-representable in floating-point arithmetic. Also, it is not usual (and not trivial) to check if a constraint qualification holds. Hence, one must verify some kind of approximate optimality condition that relies on calculations at the readily available point x^k at each iteration.

We say that a *sequential optimality condition* [5] associated to the mathematical proposition $\mathcal{P}$ holds at a feasible point $x^* \in \Omega$ if there exists a sequence $\{x^k\} \subset \mathbb{R}^n$ that converges to x^* and satisfies $\mathcal{P}(\{x^k\})$. The mathematical proposition $\mathcal{P}$ must be such that the sequential optimality condition associated to $\mathcal{P}$ is satisfied at any solution x^* of (1). This approach does not need constraint qualification assumptions.

One trivial example of sequential optimality condition is the one that assures the existence of a sequence $x^k \to x^*$ such that x^k is a solution of the problem for every k. This is clearly a sequential optimality condition, since if x^* is a solution one can define the sequence $x^k = x^*$ for every k to verify it. This is a very strong optimality condition in the sense that if one could develop an algorithm in which the iterates x^k are the elements of a sequence that verifies the condition, then one would have very strong indicatives that the algorithm converges to a solution of the problem (in fact, in this case, one would be sure of that). Unfortunately, it is not possible to build an algorithm that generates this kind of sequence. Hence, we are interested in strong optimality conditions that can be verified in practice. This trivial sequential optimality condition has another serious drawback: the set of sequences that verifies it is too small. Thus, it is important to develop conditions in which many sequences could be used to verify it, maximizing the chances of finding one of them.

The idea of sequential optimality conditions is not new. It has been implicitly used as a technique to prove optimality conditions in [8] and to analyze convergence of a Sequential Quadratic Programming algorithm in [29]. A sequential optimality condition called Approximate Gradient Projection (AGP, [26]) has been successfully applied in the context of Inexact Restoration methods ([24,25,15]). The augmented Lagrangian method from [3,4] also verifies a sequential optimality condition (AKKT or a stronger one called Complementarity-AKKT [2]). Sequential optimality conditions have also been extended to the more general framework of variational inequalities [16,18] and to non-smooth problems [13,14]. Recently,

the AKKT condition has been used as a theoretical tool to obtain new constraint qualifications and to obtain convergence results for nonlinear programming algorithms under these weaker assumptions [6,7].

Next, we define the Approximate-KKT condition from [5].

Definition 1 (AKKT) We say that a feasible point $x^* \in \Omega$ satisfies the AKKT condition if there exist sequences $\{x^k\} \subset \mathbb{R}^n$ converging to x^*, $\{\lambda^k\} \subset \mathbb{R}^m$ and $\{\mu^k\} \subset \mathbb{R}^p_+$ such that

$$\nabla f(x^k) + \sum_{i=1}^{m} \lambda_i^k \nabla h_i(x^k) + \sum_{i \in A(x^*)} \mu_i^k \nabla g_i(x^k) \to 0.$$

Note that this definition depends on information regarding the intangible limit point x^*. In principle, one would need to know which are the active constraints at the limit point, which is a fairly nontrivial issue in nonlinear optimization. We will deal with this issue in the next session, using the result below from [5]:

Lemma 1 *A point $x^* \in \mathbb{R}^n$ satisfies the AKKT condition if, and only if, there exist sequences $\{x^k\} \subset \mathbb{R}^n$ converging to x^*, $\{\lambda^k\} \subset \mathbb{R}^m, \{\mu^k\} \subset \mathbb{R}^p_+$, $\{\varepsilon^k_{feas}\} \subset \mathbb{R}_+, \{\varepsilon^k_{opt}\} \subset \mathbb{R}_+$ and $\{\varepsilon^k_{compl}\} \subset \mathbb{R}_+$ such that*

$$\varepsilon^k_{feas} \to 0, \varepsilon^k_{opt} \to 0, \varepsilon^k_{compl} \to 0, \tag{2}$$

$$|h_i(x^k)| \leq \varepsilon^k_{feas}, i = 1, \ldots, m, \max\{0, g_i(x^k)\} \leq \varepsilon^k_{feas}, i = 1, \ldots, p, \tag{3}$$

$$\|\nabla f(x^k) + \sum_{i=1}^{m} \lambda_i^k \nabla h_i(x^k) + \sum_{i=1}^{p} \mu_i^k \nabla g_i(x^k)\| \leq \varepsilon^k_{opt}, \tag{4}$$

$$g_i(x^k) < -\varepsilon^k_{compl} \Rightarrow \mu_i^k = 0, i = 1, \ldots, p. \tag{5}$$

Remark: In [5], the authors considered only sequences such that $\varepsilon^k_{opt} = \varepsilon^k_{compl}$ for every k, but it is easy to check that their proof applies to the general case. Note that x^* is necessarily feasible since $\varepsilon^k_{feas} \to 0$ and the functions are continuous.

What is interesting about the AKKT condition is that it is a sequential optimality condition, that is, if x^* is a local minimizer, then there must exist a sequence $\{x^k\}$ converging to x^* that verifies AKKT. This is true even when the problem is somewhat degenerate, in the sense that the minimizer x^* is not a KKT point. Hence, if for a given iteration k, the iterand x^k satisfies (3), (4), (5) for some $\lambda^k \in \mathbb{R}^m, \mu^k \in \mathbb{R}^p_+$ and for sufficiently small $\varepsilon^k_{feas}, \varepsilon^k_{opt}, \varepsilon^k_{compl}$ then the algorithm may stop since x^k is a good candidate for a local solution.

Another interesting fact about the AKKT condition is that it is a strong optimality condition, in the sense that it implies the usual KKT condition under mild constraint qualifications [7]. It is well-known that the converse is true without the assumption of constraint qualifications, that is, when x^* is a KKT point, it is clearly an AKKT point since one can choose $x^k = x^*$ for every k. In the next theorem we observe that, in this case, one can use *any* sequence $\{x^k\} \subset \mathbb{R}^n$ converging to x^* to verify AKKT.

Theorem 1 *Assume that x^* is a KKT point of Problem (1) with associated Lagrange multipliers $\lambda^* \in \mathbb{R}^m$ and $\mu^* \in \mathbb{R}^p_+$. Then any sequence $\{x^k\} \subset \mathbb{R}^n$ converging to x^* and any sequences $\{\lambda^k\} \subset \mathbb{R}^m$ converging to λ^*, $\{\mu^k\} \subset \mathbb{R}^p_+$ converging to μ^* can be used to check that x^* verifies AKKT.*

Proof: From the continuity of the gradients of the functions, we have that

$$\nabla f(x^k) + \sum_{i=1}^{m} \lambda_i^k \nabla h_i(x^k) + \sum_{i \in A(x^*)} \mu_i^k \nabla g_i(x^k)$$

converges to

$$\nabla f(x^*) + \sum_{i=1}^{m} \lambda_i^* \nabla h_i(x^*) + \sum_{i \in A(x^*)} \mu_i^* \nabla g_i(x^*),$$

which is equal to zero from the KKT conditions. This concludes the proof. □

Hence, if the problem is non-degenerate, any sequence that converges to a solution (such as a successful run of a GA) will satisfy an Approximate-KKT condition. The problem is that in practice, a true Lagrange multiplier (or a sequence converging to it) is not known, hence we can not check for near optimality of the iterands x^k. We will deal with this subject in the next section.

We note that if x^* is not a KKT point, then the conclusion of Theorem 1 is not true. Consider the problem of minimizing x, subject to $x^2 \leq 0$. It is clear that the global solution $x^* = 0$ is not a KKT point. Assume that x^k is a sequence that converges to x^*. We can see that if $x^k \geq 0$, there is no sequence of multipliers $\mu^k \geq 0$ such that $1 + 2\mu^k x^k \to 0$. Although the conclusion fails for non-negative sequences, the result is true for every sequence x^k such that $x^k < 0$ for sufficiently large k (take $\mu^k = -\frac{1}{2x^k} > 0$). The geometry of sequences that fulfill AKKT is currently unknown for degenerate problems.

3 AKKT stopping criterion for genetic algorithms

For many GAs, stopping criteria are based on the stagnation of functional values or on a maximum generation number. This may be a good strategy when the optimum value is known, but this is rarely the case in real-world applications. Also, relying only on a maximum generation number may be insufficient or result in an excessive amount of function evaluations. AKKT conditions can be used as a pure stopping criterion or together with other criteria. If functional values seem to be converging, instead of declaring convergence, one can check the KKT error measurement, and if it is large, this may indicate stagnation and the algorithm can be changed to search for an optimum at a different region. Also, a small KKT error measurement indicates that the GA is approaching the neighborhood of a local minimum. Hence, if a better point is known one can stop searching for a global solution at this region, otherwise, one can activate a local-search algorithm to reach the minimizer more effectively.

If an iterative solver generates at each iteration k an approximation x^k of the solution and approximations λ^k and μ^k for the Lagrange multipliers, then the AKKT condition can be used to define a stopping criterion in a straightforward way. By Lemma 1, one can stop the algorithm when

$$|h_i(x^k)| \leq \varepsilon_{feas}, i = 1, \ldots, m, \max\{0, g_i(x^k)\} \leq \varepsilon_{feas}, i = 1, \ldots, p, \qquad (6)$$

$$\|\nabla f(x^k) + \sum_{i=1}^{m} \lambda_i^k \nabla h_i(x^k) + \sum_{i=1}^{p} \mu_i^k \nabla g_i(x^k)\| \leq \varepsilon_{opt} \tag{7}$$

and

$$\mu_i^k = 0 \text{ whenever } g_i(x^k) < -\varepsilon_{compl}, \tag{8}$$

where $\varepsilon_{feas} > 0, \varepsilon_{opt} > 0, \varepsilon_{compl} > 0$ are small tolerances associated respectively to feasibility, optimality and complementarity. This is the stopping criterion used in many optimization solvers such as [3,11].

In [31], the AKKT condition of [5] and others were used as stopping criteria for a GA. They did this computing approximate Lagrange multipliers in such a way that a KKT error measure is minimized. That is, being x^k the best individual in the population at generation k of a GA, they computed the multiplier approximations $\lambda^k \in \mathbb{R}^m$, $\mu^k \in \mathbb{R}^p_+$ and the KKT error measure $\varepsilon_k \geq 0$ as the solution of the following optimization problem in λ, μ and ε:

$$\begin{array}{l} \text{Minimize } \varepsilon, \\ \\ \text{Subject to} \\ \|\nabla f(x^k) + \sum_{i=1}^{m} \lambda_i \nabla h_i(x^k) + \sum_{i=1}^{p} \mu_i \nabla g_i(x^k)\| \leq \varepsilon, \\ \mu_i \geq 0, i = 1, \ldots, p, \\ \mu_i = 0 \text{ if } g_i(x^k) < -\varepsilon, i = 1, \ldots, p. \end{array} \tag{9}$$

The authors in [31] observed in practice that when x^k is generated by a GA and the KKT error measure ε_k is computed using Problem (9), then ε_k tends to zero. Hence it can be used to stop the algorithm. Since the test problems in [31] are six non-degenerate problems from [21], the result is expected in view of Theorem 1. But it is not always the case that the approximate Lagrange multipliers obtained solving Problem (9) converge to true Lagrange multipliers associated to the limit point x^*. The next theorem clarifies that even in this case, for every sequence $\{x^k\}$ that converges to a non-degenerate solution, the KKT error measurement ε_k always converges to zero.

Theorem 2 *Assume that x^* is a KKT point of Problem (1) and that $\{x^k\} \subset \mathbb{R}^n$ is any sequence converging to x^*. Then, the sequence of solutions $\{(\lambda^k, \mu^k, \varepsilon_k)\}$ of Problem (9) is such that $\varepsilon_k \to 0$. Moreover, the limit points of $\{(\lambda^k, \mu^k)\}$ are true Lagrange multipliers of Problem (1), and if x^* fulfills the Mangasarian-Fromovitz constraint qualification, then the sequence $\{(\lambda^k, \mu^k)\}$ is bounded.*

Proof: Let $\lambda^* \in \mathbb{R}^m$ and $\mu^* \in \mathbb{R}^p_+$ be Lagrange multipliers associated to x^*. Define $\tilde{\lambda}^k = \lambda^*$, $\tilde{\mu}^k = \mu^*$ and $\tilde{\varepsilon}_k = \max\{\|\nabla f(x^k) + \sum_{i=1}^{m} \tilde{\lambda}_i^k \nabla h_i(x^k) + \sum_{i=1}^{p} \tilde{\mu}_i^k \nabla g_i(x^k)\|, -g_i(x^k), \text{ for } i \in A(x^*)\}$. We will prove that $(\tilde{\lambda}^k, \tilde{\mu}^k, \tilde{\varepsilon}_k)$ is feasible for Problem (9) for sufficiently large k. It is clear from the construction that the first constraint in Problem (9) is satisfied. The non-negativity of $\tilde{\mu}^k$ is also clear. Now, observe that the continuity of the gradients and the fact that λ^* and μ^* are Lagrange multipliers associated to x^* imply that $\tilde{\varepsilon}_k \to 0$. If $i \notin A(x^*)$, that is, $g_i(x^*) < 0$, the continuity of g_i implies that $g_i(x^k) < -\tilde{\varepsilon}_k$ for sufficiently large k and also, from the KKT conditions, we have $\tilde{\mu}^k = \mu^* = 0$. If $i \in A(x^*)$, we have from the definition of $\tilde{\varepsilon}_k$ that $-g_i(x^k) \leq \tilde{\varepsilon}_k$, that is, $g_i(x^k) \geq -\tilde{\varepsilon}_k$. Hence, for every i and sufficiently large k, $\tilde{\mu}_i^k = 0$ whenever $g_i(x^k) < -\tilde{\varepsilon}_k$, which proves feasibility. The proof that $\varepsilon_k \to 0$

is completed observing that the optimality of ε_k implies that $0 \leq \varepsilon_k \leq \tilde{\varepsilon}_k \to 0$. Now, if the Mangasarian-Fromovitz constraint qualification [23,30] holds, assuming that $\{(\lambda^k, \mu^k)\}$ has an unbounded subsequence, we define $M_k = \max\{|\lambda_i^k|, \mu_j^k, i = 1, \ldots, m, j = 1, \ldots, p\}$. Since $f(x^k)+\sum_{i=1}^{m} \lambda_i^k \nabla h_i(x^k)+\sum_{i\in A(x^*)}^{p} \mu_i^k \nabla g_i(x^k) \to 0$, dividing by M_k and taking a suitable subsequence, we get a contradiction with the definition of the Mangasarian-Fromovitz constraint qualification. It is clear that any convergent subsequence of $\{(\lambda^k, \mu^k)\}$ converges to true Lagrange multipliers. □

The above theorem explains the results observed in [31], where they found that the above ε_k KKT error measurement tends to zero when the iterates of a GA converges to a solution. In fact, if one chooses non-degenerate test problems, any algorithm that converges to a solution will generate sequences that fulfill Approximate-KKT conditions for suitable Lagrange multipliers. This motivates the use of AKKT conditions as stopping criteria for general iterative solvers, where the optimum value is unknown. Also, the boundedness of the sequence of approximate multipliers implies that it converges to a true Lagrange multiplier, at least in an appropriate subsequence. When the gradients of active constraints are linearly independent at the solution, the whole sequence of computed Lagrange multipliers approximation will converge to the unique Lagrange multipliers at the solution.

An important issue is how one can efficiently compute the KKT error measurement ε_k. The authors in [31] modeled the last inequality of Problem (9) as:

$$\mu_i \min\{0, g_i(x^k) + \varepsilon\} = 0, i = 1, \ldots, p,$$

which is non-differentiable. This motivated them to define new AKKT conditions without this drawback. We address this issue using Lemma 1, which states that the error measurements for feasibility, optimality and complementarity may be different from each other. Our goal is, at each iteration k, to define approximate Lagrange multipliers $\lambda^k \in \mathbb{R}^m$ and $\mu^k \in \mathbb{R}^p_+$ such that equations (2), (3), (4), and (5) are satisfied. Our idea is to consider the sequences $\{\varepsilon_{feas}^k\}$ and $\{\varepsilon_{compl}^k\}$ as exogenous sequences, that is, before computing the approximate Lagrange multipliers at iteration k, the parameters ε_{feas}^k and ε_{compl}^k are already known. The parameter ε_{opt}^k and the Lagrange multipliers λ^k and μ^k are defined as the solution of the problem:

$$\begin{aligned} &\text{Minimize } \varepsilon_{opt}, \\ &\\ &\text{Subject to} \\ &\|\nabla f(x^k) + \textstyle\sum_{i=1}^{m} \lambda_i \nabla h_i(x^k) + \sum_{i=1}^{p} \mu_i \nabla g_i(x^k)\| \leq \varepsilon_{opt}, \\ &\mu_i \geq 0, i = 1, \ldots, p, \\ &\mu_i = 0 \text{ if } g_i(x^k) < -\varepsilon_{compl}^k. \end{aligned} \tag{10}$$

Since $\{\varepsilon_{compl}^k\}$ is given, one may check *a priori* which constraints satisfy $g_i(x^k) < -\varepsilon_{compl}^k$, defining the corresponding approximate Lagrange multiplier to be $\mu_i^k = 0$. Let A_k be the index set of inequality constraints such that μ_i^k was not defined to be zero, that is, $A_k = \{i \in \{1, \ldots, p\} \mid g_i(x^k) \geq -\varepsilon_{compl}^k\}$. Hence, we may transform Problem (10) to the equivalent Non-Negative linear Least Squares problem (NNLS) in the variables λ and μ:

$$\begin{aligned} &\text{Minimize} \\ &\|\nabla f(x^k) + \textstyle\sum_{i=1}^{m} \lambda_i \nabla h_i(x^k) + \sum_{i \in A_k} \mu_i \nabla g_i(x^k)\|^2 \\ &\text{subject to } \ \mu_i \geq 0, i \in A_k \end{aligned} \tag{11}$$

The parameter ε_{opt}^k is defined as the square root of the objective function of (11) at the solution. We stop the execution of the algorithm when the KKT error measurement $\varepsilon_k = \max\{\varepsilon_{feas}^k, \varepsilon_{opt}^k, \varepsilon_{compl}^k\}$ is sufficiently small. We note that in order to ε_{opt}^k to converge to zero, the exogenous sequence ε_{compl}^k must be such that the true active constraints at the solution are detected, that is, for sufficiently large k, it must hold that $A_k = A(x^*)$. It is easy to see that a sufficient condition is that ε_{compl}^k is a sufficiently small constant sequence ($\varepsilon_{compl}^k := \varepsilon_{compl} < \min\{-g_i(x^*) \mid g_i(x^*) < 0\}$). That is how we define ε_{compl}^k in our numerical experiments. Since the limit point x^* is unknown, we cannot be sure that ε_{compl} was chosen sufficiently small. But this should not be an issue, since it is unlikely that the inactive inequality constraints are too close to being active at a solution. If this is the case, then the problem is badly scaled.

We also defined the exogenous sequence ε_{feas}^k as a small positive constant tolerance ε_{feas}. We note that one must solve the NNLS problem (11) only for iterates that are close to feasibility, that is, (6) holds. The class of NNLS problems is well-known in the literature, with fast tailored algorithms available. See for example [28,20].

4 Numerical experiments

The purpose of our numerical experiments is twofold: to test if the KKT error measurement goes to zero even for degenerate problems, and to verify how soon the GA stops using the AKKT error measure as stopping criterion.

The optimization process is performed by a GA implemented in Scilab. The employed constraint handling is the penalty function approach proposed by Deb [12]. The configuration employed was: population size=200; maximum generations=1000; crossover probability=50%; crossover operator type=line; crossover line-extension=0.5; mutation operator=gaussian convolution with a standard deviation of 0.0001; mutation probability=1%; tournament selection size=3; elitism=1.

To compute the KKT error measure, we defined the exogenous sequences to be constant: $\varepsilon_{feas}^k = \varepsilon_{feas}$ and $\varepsilon_{compl}^k = \varepsilon_{compl}$, where $\varepsilon_{compl} = \varepsilon_{feas} = 10^{-1}$. When the feasibility test (6) is not fulfilled, we define the KKT error measurement ε_k to be the maximum infeasibility, otherwise, the set A_k is computed and Problem (11) is solved accordingly. The Mathworks implementation (lsqnonneg.m) of the algorithm described in [20], chapter 23, page 161, was employed with its default options. The square root of the objective function is then defined to be the error measurement ε_{opt}^k with respect to optimality, at each iteration. The KKT error measurement ε_k is then defined to be the maximum of ε_{compl}, ε_{feas} and ε_{opt}^k, and we stop the GA when $\varepsilon_{opt}^k < 10^{-2}$. We note that the Lagrange multipliers approximation $(\lambda^{k-1}, \mu^{k-1})$ of the previous iteration is a natural good starting point to solve Problem (11) at iteration k, which accelerates convergence.

Problem id	Stop iteration	Relative optimal error
00201	16.1	0.30508%
00302	112.5	0.02189%
00501	1028.6	3.36801%
10201	32.1	4.58896%
10202	39.0	0.51698%
10205	28.1	1.18954%
10206	30.6	4.06320%
10207	30.5	3.88757%
10301	86.5	4.17318%
10302	84.9	4.14185%
10401	521.3	0.01163%
20218	13.5	4.92705%
20219	36.5	6.50355%
20220	47.6	7.49352%
20221	39.6	7.14952%
30202	15.9	0.00222%
30207	17.8	2.57150%
30208	16.6	2.06597%
30302	39.7	3.05818%
40201	48.7	0.60967%
40202	50.7	0.23441%
40203	21.3	0.00054%
40205	14.9	0.04261%
40206	19.5	0.00082%
40208	20.3	0.00153%
40401	55.2	0.05271%
40402	349.8	4.74934%

Table 1 Average results on 30 runs of GA at the DEGEN collection.

For a first test, we employed the collection of 111 degenerate test problems from [1]. We note that when equality constraints are present, the associated Lagrange multipliers may assume negative values, and this is not supported by lsqnonneg.m. A change of variables would require the software to deal with rank deficient matrices, which is also not supported. Hence, we removed from our test set the problems that include some equality constraint. Other problems that returned a rank deficiency error were also discarded. We also did not consider problems with unlimited global solution. This reduces our test set to 56 problems with a number of variables ranging from 1 to 4. We run the GA for each problem 30 times, and we declare a success when the average distance to the reported global optimum value is less than 10^{-1}. In this way, 27 problems where considered successful. This is not a drawback since, given the degenerate nature of the problems, we do not expect the GA to always find the solution without a fine tuning of the parameters for each problem. Our goal is to analyze the behavior of the ε_k parameter when the GA converges to a degenerate solution and to check how soon the iterations can be stopped. In Table 1 we describe our results. In the first column we indicate the Problem's name as in [1], the second column is the average number of iterations needed to achieve the prescribed precision of $\varepsilon_{opt}^k < 10^{-2}$ and the third column is the relative optimal value error, comparing with the reported optimal value.

We note that we were able to detect convergence to the optimum solution using the ε_k parameter, and to stop the execution of the GA at a near optimum solution

Problem identification	ε	Stop iteration	Relative optimal error
g01	10^{-1}	121.9	2.51%
g01	10^{-2}	177.2	0.256%
g01	10^{-4}	297.5	0.00278%
g06	10^{-1}	11.1	1.19%
g06	10^{-2}	76.1	0.181%
g06	10^{-4}	533.9	0.00221%

Table 2 Average results on 30 runs of the GA for different tolerances.

using a tolerance of 10^{-1} for feasibility and complementarity, and of 10^{-2} for optimality, even for potentially degenerated problems. Although further research is needed to understand the behavior of the KKT error measure for degenerate problems, this test shows that our approach can be successfully applied.

For a second test we used problems g01 and g06 of the CEC2006 collection [21]. These are non-degenerate problems, which implies that the KKT error measurement will converge to zero. We want to compare the behavior of the algorithm for different precisions $\varepsilon = 10^{-1}, 10^{-2}$ and 10^{-4}. We fixed $\varepsilon^k_{compl} = \varepsilon^k_{feas} = \varepsilon$ and we defined the KKT error measurement ε_k in the same way as in the previous test. We stop the execution of the GA when $\varepsilon_k \leq \varepsilon$. The average results for 30 successful runs of the GA are described in Table 2.

We note that in [31], new measures of near optimality are proposed to stop a GA, but calculations of the proposed ε_k KKT error measurements are considerably more costly. We can infer from [31] that for a simple GA run reported in that paper, one could stop the execution of the algorithm approximately at iterations 200 and 45, respectively for g01 and g06. When comparing the number of iterations at tolerances 10^{-1} and 10^{-2}, we see that our strategy is compatible with the ones in [31], where tolerances are not reported.

5 Conclusions

In this paper we showed that even for algorithms that do not generate Lagrange multipliers approximations, one may use Approximate-KKT conditions to develop stopping criteria. We proved that when Lagrange multipliers exist at the solution, the KKT error measurement goes to zero when using a standard Approximate-KKT condition. We also provided a strategy to define Lagrange multipliers approximations in such a way that a non-negative linear least squares problem is solved at each iteration. We proved that under standard constraint qualifications, the sequence of approximated Lagrange multipliers converges to the true multipliers at the solution. We tested our strategy on a collection of degenerated test problems and we checked that the KKT error may still be used to stop a GA in this case. Further research is needed to understand this behavior. In a test in two non-degenerate problems, we compared the average number of iterations needed to stop the algorithm with a similar strategy from the literature. The results are similar, but the subproblems in our approach are much simpler.

References

1. Http://w3.impa.br/∼optim/DEGEN_collection.zip
2. Andreani R. Martínez, J., Svaiter, B.: A new sequential optimality condition for constrained optimization and algorithmic consequences. SIAM Journal on Optimization **20**(6), 3533–3554 (2010)
3. Andreani, R., Birgin, E., Martínez, J., Schuverdt, M.: On augmented lagrangian methods with general lower–level constraints. SIAM Journal on Optimization **18**, 1286–1309 (2007)
4. Andreani, R., Birgin, E., Martínez, J., Schuverdt, M.: Augmented lagrangian methods under the constant positive linear dependence constraint qualification. Mathematical Programming **112**, 5–32 (2008)
5. Andreani, R., Haeser, G., Martínez, J.: On sequential optimality conditions for smooth constrained optimization. Optimization **60**(5), 627–641 (2011)
6. Andreani, R., Haeser, G., Schuverdt, M., Silva, P.: A relaxed constant positive linear dependence constraint qualification and applications. Mathematical Programming **135**, 255–273 (2012)
7. Andreani, R., Haeser, G., Schuverdt, M., Silva, P.: Two new weak constraint qualifications and applications. SIAM Journal on Optimzation **22**(3), 1109–1135 (2012)
8. Bertsekas, D.: Nonlinear Programming: 2nd Edition. Athena Scientific (1999)
9. Chelouah, R., Siarry, P.: Genetic and nelder-mead algorithms hybridized for a more accurate global optimization of continuous multi-minima functions. European Journal of Operational Research. **148**(2), 335–348 (2003)
10. Chou, H.H., Premkumar, G., Chu, C.H.: Genetic algorithms for communications network design: An empirical study of factor that influence performance. In: IEEE Transactions on Evolutionary Computation., vol. 5, pp. 236–249 (2001)
11. Conn, A.R., Gould, N., Toint, P.: Trust Region Methods. SIAM (2000)
12. Deb, K.: An efficient constraint handling method for genetic algorithms. Computer Methods in Applied Mechanics and Engineering **186**(2-4), 311–338 (2000)
13. Dutta, J., Deb, K., Arora, R., Tulshyan, R.: Approximate KKT points for iterates of an optimizer. Tech. Rep. 2009009, Kanpur Genetic Algorithms Laboratory (2009)
14. Dutta, J., Deb, K., Tulshyan, R., Arora, R.: Approximate KKT points and a proximity measure for termination. Tech. Rep. 2010007, Kanpur Genetic Algorithms Laboratory (2010)
15. Fischer, A., Friedlander, A.: A new line search inexact restoration approach for nonlinear programming. Computational Optimization and Applications **46**(2), 333–346 (2010)
16. Gárciga-Otero, R., Svaiter, B.: A new condition characterizing solutions of variational inequality problems. Journal of Optimization Theory and Applications **137**, 89–98 (2008)
17. Goldberg, D.E.: Genetic Algorithms in Search, Optimization, and Machine Learning. Addison-Wesley (1989)
18. Haeser, G., Schuverdt, M.: Approximate KKT condition for continuous variational inequalities. Journal of Optimization Theory and Applications **149**(3), 528–539 (2011)
19. Janin, R.: Directional derivative of the marginal function in nonlinear programming, pp. 110–126. Springer Berlin Heidelberg (1984)
20. Lawson, C.L., Hanson, R.J.: Solving Least Squares Problems. Prentice-Hall (1974)
21. Liang, J., Runarsson, T., Mezura-Montes, E., Clerc, M., Suganthan, P., Coello, C., Deb, K.: Problem definitions and evaluation criteria for CEC-2006. Tech. rep., Nanyang Technological University (2006)
22. Lung, Y.W., Wang, Y.: An orthogonal genetic algorithm with quantization for global numerical oplimization. In: IEEE Transactions on Evolutionary Computation, vol. 5(1), pp. 41–53 (2001)
23. Mangasarian, O., Fromovitz, S.: The Fritz John optimality conditions in the presence of equality and inequality constraints. Journal of Mathematical Analysis and Applications **17**, 37–47 (1967)
24. Martínez, J.: Inexact restoration method with Lagrangian tangent decrease and new merit function for nonlinear programming. Journal of Optimization Theory and Applications **111**, 39–58 (2001)
25. Martínez, J., Pilotta, E.: Inexact restoration algorithms for constrained optimization. Journal of Optimization Theory and Applications **104**, 135–163 (2000)
26. Martínez, J., Svaiter, B.: A practical optimality condition without constraint qualifications for nonlinear programming. Journal of Optimization Theory and Applications **118**, 117–133 (2003)

27. Minchenko, L., Stakhovski, S.: On relaxed constant rank regularity condition in mathematical programming. Optimization **60**(4), 429–440 (2011)
28. Portugal, L., Judice, J., Vicente, L.: A comparison of block pivoting and interior-point algorithms for linear least-squares problems with nonnegative variables. Mathematics of Computation **63**, 625–643 (1994)
29. Qi, L., Wei, Z.: On the constant positive linear dependence condition and its application to SQP methods. SIAM Journal on Optimization **10**, 963–981 (2000)
30. Rockafellar, R.: Lagrange multipliers and optimality. SIAM Review **35**(2), 183–238 (1993)
31. Tulshyan, R., Arora, R., Deb, K., Dutta, J.: Investigation EA solutions for approximate KKT conditions in smooth problems. GECCO (2010)