Approximate KKT Points and a Proximity Measure for Termination

Joydeep Dutta[1], Kalyanmoy Deb[2], Rupesh Tulshyan[1] and Ramnik Arora[1]
[1]Department of Mathematics and Statistics, [2]Department of Mechanical Engineering
Indian Institute of Technology Kanpur
Kanpur, PIN 208016, India
Web: http://www.iitk.ac.in/kangal/

KanGAL Report No. 2010007

Abstract

Karush-Kuhn-Tucker (KKT) optimality conditions are often checked for investigating whether a solution obtained by an optimization algorithm is a likely candidate for the optimum. In this study, we report that although the KKT conditions must all be satisfied at the KKT point, the extent of violation of KKT conditions at points arbitrarily close to the KKT point is not smooth, thereby making the KKT conditions difficult to use directly to evaluate the performance of an optimization algorithm. This happens due to the requirement of complimentary slackness condition associated with KKT optimality conditions. To overcome this difficulty, we define modified ϵ-KKT points by relaxing the complimentary slackness and equilibrium equations of KKT conditions and suggest a *KKT-proximity measure*, that reduces sequentially to zero as the iterates approach the KKT point. Besides the theoretical development defining the modified ϵ-KKT point, we present extensive computer simulations of the proposed methodology on a set of iterates obtained through an evolutionary optimization optimization algorithm to illustrate the working of our proposed procedure on smooth and non-smooth problems. The results indicate that the proposed KKT-proximity measure can be used a termination condition to optimization algorithms. We also provide a comparison of our KKT-proximity measure with the stopping criterion used in the commercial softwares like Knitro.

1 Introduction

The Karush-Kuhn-Tucker (KKT) conditions are necessary for a solution in a non-linear programming problem to be optimal (provided some regularity conditions are satisfied) and hence, they play an important part in optimization theory [14]. However, the KKT conditions are not adequately investigated for their regularity in the neighborhood of the KKT point, nor have they been widely used in optimization algorithm design, primarily owing to the fact that these are point conditions. The KKT conditions we know are precisely satisfied at an exact optimal point, local or global, provided a suitable constraint qualification condition holds. Thus, it would be important to know to what extent the KKT conditions are violated in the neighborhood of an exact optimal point. Thus, when approaching the KKT point through a series of iterates obtained by an optimization algorithm, the extent of violation of KKT conditions may not reduce in any smooth manner. This makes any derivation of the extent of satisfaction of KKT conditions as a check for termination difficult, even for smooth problems. Commercial softwares, like Knitro [3] and MATLAB Optimization toolbox [12], base their stopping criterion on KKT conditions, but they use complicated modifications of the KKT conditions. However, it is also true that the KKT conditions are never used in the algorithm design. If one could judge the proximity and direction of the optimum from a given point, using some metric derived from the KKT condition violations, this could be very helpful for devising a theoretically motivated termination condition which could be used in the actual design of algorithms. This study is an aim to satisfy the former of the requirements.

In this paper, a simplistic KKT-proximity measure has been derived from KKT conditions to indicate the closeness of a given iterate to the optimum of an optimization problem. We begin the paper by considering a minimization problem with smooth data. We present here the notion of an ϵ-KKT point. In fact we show that if $\{\mathbf{x}_k\}$ is a sequence of feasible points converging to $\bar{\mathbf{x}}$ and if a suitable qualification condition holds at $\bar{\mathbf{x}}$ then $\bar{\mathbf{x}}$ is a KKT point provided each $\mathbf{x}_k$ is an ϵ_k-KKT point where $\epsilon_k > 0$ for all k and $\epsilon_k \downarrow 0$.

Thereafter, we suggest a *modified* ϵ-KKT point which we do in the general setting of a nonsmooth optimization problem with locally Lipschitz data. The fundamental feature of this definition is that we relax the complementary slackness condition and further in the subgradient conditions instead of taking the exact reference point we move away from it consider a point from the neighborhood. Both these steps are a departure from the approach used to define an ϵ-KKT point in the smooth case. The move away from the exact reference point is significant since for a locally Lipschitz problem one can very likely move to a point around which the function is continuously differentiable and thus allowing us to work with derivatives rather than subgradients. Of course this is made possible by the fact that a locally Lipschitz function is densely differentiable in any neighborhood of a given point. Further,for any iterate, we suggest an optimization procedure to find a minimum ϵ parameter that would satisfy the proposed modified ϵ-KKT conditions. Additionally, we define the minimum ϵ value as a KKT-proximity measure of the iterate from the KKT point. Due to the flexibility associated in satisfying all the approximate KKT conditions, the KKT-proximity measure is found to behave smoothly in the vicinity of the KKT point. We demonstrate this aspect on both smooth and non-smooth problems.

The remainder of the paper is organized as follows. Section 2 states the KKT conditions separately for smooth and non-smooth problems. Section 3 defines an ϵ-KKT point by relaxing the conditions involving the gradients. The inadequacy of the resulting KKT-error measure in capturing the closeness of a point to the KKT point is then illustrated on a simple two-variable constrained problem. Thereafter, in Section 3.3, we define a modified ϵ-KKT point which further relaxes the complimentary slackness equation of the KKT condition set. Associated theorems outlining the conditions for a series of modified ϵ-KKT points to approach the KKT point are proven next. Based on the theorems, we propose a *KKT-proximity* measure for smooth and non-smooth cases. We also try to answer (at least partially) the following question. Given a local minimum, can we find a sequence of points converging to it which satisfies one of the type of approximate optimality notions that we define here? Results are shown in Section 4 where the KKT-proximity measure for a sequence of iterates is plotted for a number of standard constrained test problems. The iterates are taken as the generation-wise best solutions of a real-coded genetic algorithm (RGA) [5]. We also include the best-reported solution in the iterate set so that the convergence property of the RGA can be assessed from a theoretical point of view. Section 5 compares the variation of KKT-proximity measure with the termination parameter of a commercial optimization software – Knitro [3]. Conclusions and further research work are suggested in Section 6.

2 KKT Conditions: Smooth and Non-smooth Cases

In this section, we present the KKT optimality conditions for smooth and non-smooth problems.

2.1 Smooth Case

For the given single-objective, constrained smooth optimization problem (P):

$$\begin{array}{rl} \text{Minimize} & f(\mathbf{x}), \\ \text{Subject to} & g_i(\mathbf{x}) \leq 0, \quad i = 1, 2, \ldots, m, \end{array} \tag{1}$$

the Karush-Kuhn-Tucker (KKT) optimality conditions are given as follows:

$$\nabla f(\bar{\mathbf{x}}) + \sum_{i=1}^{m} u_i \nabla g_i(\bar{\mathbf{x}}) = \mathbf{0}, \tag{2}$$

$$g_i(\bar{\mathbf{x}}) \leq 0, \quad \forall i, \tag{3}$$

$$u_i g_i(\bar{\mathbf{x}}) = 0, \quad \forall i, \tag{4}$$

$$u_i \geq 0, \quad \forall i. \tag{5}$$

The parameter u_i is called the Lagrange multiplier for the i-th constraint. Any solution $\bar{\mathbf{x}}$ that satisfies all the above conditions is called a KKT point [14]. The Equation 2 is known as the *equilibrium equation* or the *gradient condition*. If we take the norm of the vector on the right hand side of Equation 2 and the norm is non-zero then the value of the norm is called the KKT error at that the reference point. Equation 4 is known as the *complimentary slackness* equation. Note that the conditions given in equation 3 ensure feasibility for $\bar{\mathbf{x}}$ while the equation 5 tells us that the Lagrange multipliers are non-negative.

The complimentary slackness condition implies that if a KKT point $\bar{\mathbf{x}}$ makes a constraint inactive (meaning $g_i(\bar{\mathbf{x}}) < 0$ for the i-th constraint), the corresponding Lagrange multiplier u_i must be zero. On the other hand, along with equation 5, we conclude that if the KKT point makes the i-th constraint active (meaning $g_i(\bar{\mathbf{x}}) = 0$), u_i may take either zero or positive values. Also, the equilibrium equation requires that the negative of the gradient vector of the objective function at the KKT point be a positive linear combination of the gradient vectors of the active constraints.

It is important to note that a KKT point is not necessarily a minima of the original problem. Further conditions (in general, involving constraint qualification conditions or second-order derivatives) are necessary to establish the optimality of a point. However, in this paper, we keep our discussions to KKT points which are candidates for the minimum point.

The KKT conditions clearly state that the KKT-error ($\|\nabla f(\bar{\mathbf{x}}) + \sum_{i=1}^{m} u_i \nabla g_i(\bar{\mathbf{x}})\|$) is zero at the KKT point. However, it is not clear and is not adequately mentioned in textbooks on mathematical optimization as to how the KKT-error varies for points in the proximity of the KKT point. If the KKT-error reduces monotonically for points as we get closer to the KKT point, the KKT-error can be reliably used as a *termination criterion* for any constrained optimization algorithm, including an evolutionary algorithm. We shall investigate this aspect through the definition of a ϵ-KKT point a little later, but before that let us discuss the KKT conditions for a non-smooth problem.

2.2 Non-Smooth Case

In optimization we often come across situations where the minimum or rather the optimum of a function is achieved precisely at the point where the function is not differentiable. As an example consider the function, $f(x) = |x|,\ x \in \Re$. It is clear that the minimum of f is achieved at $x = 0$ where it is non-differentiable. The question is how to overcome this bane of non-differentiability. The class of functions for which the issue of non-differentiability was first studied was the class of convex functions. Note that the function $f(x) = |x|,\ x \in \Re$, is also a convex function. For details on convex analysis see for example [14]. The tool that is used to replace the derivative at a point of non-differentiability is called the subdifferential, which we now formally define.

Definition 2.1 *Let $f : \Re^n \to \Re$ be a convex function. Then $\boldsymbol{\xi} \in \Re^n$ is called a subgradient of f at $\mathbf{x} \in \Re^n$, if*

$$f(\mathbf{y}) \ \geq \ f(\mathbf{x}) + \langle \ \boldsymbol{\xi}, \ \mathbf{y} - \mathbf{x} \ \rangle \quad \forall \ \mathbf{y} \in \Re^n.$$

The collection of all subgradients of f at $\mathbf{x}$ is called the subdifferential of f at $\mathbf{x}$ and is denoted by $\partial f(\mathbf{x})$. The set $\partial f(\mathbf{x})$ is non-empty, convex and compact for any $\mathbf{x} \in \Re^n$. A point $\bar{\mathbf{x}} \in \Re^n$ is a global minimum of f if and only if $0 \in \partial f(\bar{\mathbf{x}})$. For a convex function f, the one-sided directional derivative at $\mathbf{x} \in \Re^n$ in the direction $\mathbf{d} \in \Re^n$ is given as

$$f'(\mathbf{x}, \mathbf{d}) \ = \ \lim_{\lambda \downarrow 0} \frac{f(\mathbf{x} + \lambda \mathbf{d}) - f(\mathbf{x})}{\lambda},$$

where $\lambda \downarrow 0$ means $\lambda > 0$ & $\lambda \to 0$. For a convex function the directional derivative exists for each $\mathbf{x}$ and in each direction $\mathbf{d}$. Further, we have

$$\partial f(\mathbf{x}) \ = \ \left\{ \ \boldsymbol{\xi} \in \Re^n : \ f'(\mathbf{x}, \mathbf{d}) \geq \langle \boldsymbol{\xi}, d \rangle \quad \forall \ \mathbf{d} \in \Re^n \ \right\}.$$

Also,

$$f'(\mathbf{x}, \mathbf{d}) \ = \ \max_{\boldsymbol{\xi} \in \partial f(\mathbf{x})} \langle \boldsymbol{\xi}, d \rangle.$$

Let us consider the convex problem (P). Assume that the Slater constraint qualification holds true, i.e. there exists $\hat{\mathbf{x}} \in \Re^n$ such that $g_i(\hat{\mathbf{x}}) < 0, \ \forall \ i = 1, \ldots, m$. Then $\bar{\mathbf{x}} \in \Re^n$ is a minimum of the above problem if and only if there exists $\mathbf{u} \in \Re^m_+$ such that

1. $0 \ \in \ \partial f(\bar{\mathbf{x}}) \ + \ \sum_{i=1}^m u_i \partial g_i(\bar{\mathbf{x}})$,
2. $g_i(\bar{\mathbf{x}}) \ \leq \ 0, \quad i = 1, \ldots, m$,
3. $u_i g_i(\bar{\mathbf{x}}) \ = \ 0, \quad i = 1, \ldots, m.$

For more details on convex analysis and optimization, readers are referred to [14].

An important property of a convex function is that it is locally Lipschitz. A function $f : \Re^n \to \Re$ is locally Lipschitz at $\bar{\mathbf{x}} \in \Re^n$ if there exists a neighborhood $U(\bar{\mathbf{x}})$ of $\bar{\mathbf{x}}$ such that there exists a $K \geq 0$ (depending on $\bar{\mathbf{x}}$) for which

$$|f(\mathbf{y}) - f(\mathbf{x})| \ \leq \ K \ ||\mathbf{y} - \mathbf{x}||, \quad \forall \ \mathbf{x}, \mathbf{y} \in U(\bar{\mathbf{x}}).$$

Moreover the famous Rademacher's theorem tells us that a locally Lipschitz function is differentiable almost everywhere. This means that the points of non-differentiability form a set of measure zero. The issue of non-differentiability for locally Lipschitz function has been tackled in the following way. For a locally Lipschitz function the Clarke generalized directional derivative at $\mathbf{x}$ in the direction $\mathbf{v} \in \Re^n$, is defined as:

$$f^o(\mathbf{x}, \mathbf{v}) \ = \ \limsup_{\mathbf{y} \to \mathbf{x}, \lambda \downarrow 0} \frac{f(\mathbf{y} + \lambda \mathbf{v}) - f(\mathbf{y})}{\lambda}.$$

The Clarke subdifferential of a locally Lipschitz function f at $\mathbf{x}$ is given by

$$\partial^o f(\mathbf{x}) \ = \ \left\{ \ \boldsymbol{\xi} \in \Re^n : \ f^o(\mathbf{x}, \mathbf{v}) \geq \langle \boldsymbol{\xi}, v \rangle, \quad \forall \ \mathbf{v} \in \Re^n \ \right\}.$$

The set $\partial^o f(\mathbf{x})$ is a non-empty, convex and compact set for each $\mathbf{x} \in \Re^n$. Further as a set valued map $\partial^o f$ is locally-bounded and graph closed and hence upper semi-continuous. For more details on the Clarke subdifferential or the Clarke derivative see for example [4]. If $\bar{\mathbf{x}}$ is a local minimum of f over $\Re^n$, then $0 \in \partial^o f(\bar{\mathbf{x}})$. This condition is only necessary and not sufficient in general.

It is simple to observe that every locally Lipschitz function is not convex. Further if f is convex then

$$f'(\mathbf{x}, \mathbf{v}) \;=\; f^o(\mathbf{x}, \mathbf{v}), \quad \forall\, \mathbf{x}, \mathbf{v} \in \Re^n,$$

and

$$\partial^o f(\mathbf{x}) \;=\; \partial f(\mathbf{x}), \quad \forall\, \mathbf{x} \in \Re^n.$$

A locally Lipschitz function is called regular at $\mathbf{x} \in \Re^n$ if the one-sided directional derivative exists and

$$f'(\mathbf{x}, \mathbf{v}) \;=\; f^o(\mathbf{x}, \mathbf{v}), \quad \forall\, \mathbf{v} \in \Re^n.$$

Of course every convex function is regular.

Let us consider two locally Lipschitz functions f_1 & f_2, then we have

$$\partial^o (f_1 + f_2)(\mathbf{x}) \;\subset\; \partial^o f_1(\mathbf{x}) + \partial^o f_2(\mathbf{x}).$$

Equality holds if both f_1 and f_2 are regular. Thus if f_1 and f_2 are convex, we have

$$\partial (f_1 + f_2)(\mathbf{x}) \;=\; \partial f_1(\mathbf{x}) + \partial f_2(\mathbf{x}).$$

Furthermore, $\partial^o(\lambda f)(\mathbf{x}) \;=\; \lambda \partial^o f(\mathbf{x})$ for all $\lambda \in \Re$. However, if f is convex then

$$\partial(\lambda f)(\mathbf{x}) \;=\; \lambda \partial f(\mathbf{x}) \quad \forall\, \lambda \geq 0.$$

Let us now consider the optimization problem (P) where f and each g_i are now locally Lipschitz. Assume that $\bar{\mathbf{x}}$ is a local minimum of the above problem. Denote by $I(\bar{\mathbf{x}}) = \{\ i : g_i(\bar{\mathbf{x}}) = 0\ \}$, the set of all active constraints at $\bar{\mathbf{x}}$. We say that the Mangasarian-Fromovitz type constraint qualification is satisfied at $\bar{\mathbf{x}}$ if there exists $\mathbf{d} \in \Re^n$ such that

$$g_i^o(\bar{\mathbf{x}}, \mathbf{d}) \;<\; 0, \quad \forall\, i \in I(\bar{\mathbf{x}}). \tag{6}$$

Hence if the Mangasarian-Fromovitz type constraint qualification holds at a local minimum $\bar{\mathbf{x}}$, then there exists $\mathbf{u} \in \Re^m_+$ such that

1. $0 \;\in\; \partial^o f(\bar{\mathbf{x}}) + \sum_{i=1}^m u_i \partial^o g_i(\bar{\mathbf{x}})$,
2. $g_i(\bar{\mathbf{x}}) \;\leq\; 0, \quad i = 1, \ldots, m$,
3. $u_i g_i(\bar{\mathbf{x}}) \;=\; 0, \quad i = 1, \ldots, m$.

3 Approximate KKT Optimality Conditions

Let us begin by considering the smooth case. In this section we study the approximate KKT optimality conditions and their relationship with the exact KKT optimality conditions (equation 2–5). Our main aim in this section is to define certain notions of approximate KKT points and show that, if a sequence of such points converges to a point where some constraint qualification is also satisfied, then the limit point is a KKT point.

Very recently Andreani et al.[1] introduced some notions of approximate KKT conditions. They studied only the smooth case while we shall consider both the smooth and non-smooth cases. We would like to point out that our approach to approximate KKT points is quite different from [1]. We discuss further the differences in subsection 3.6.

3.1 An ϵ-KKT Point

Concentrating on the optimization problem (P) mentioned in section 2.1, we have the following definition:

Definition 3.1 *A point* $\mathbf{x}$ *which is feasible to (P) is said to be an* ϵ*-KKT point if given* $\epsilon > 0$*, there exist scalars* $u_i \geq 0,\ i = 1, 2, \ldots, m$ *such that*

1. $\left\|\nabla f(\mathbf{x}) + \sum_{i=1}^{m} u_i \nabla g_i(\mathbf{x})\right\| \ \leq \ \epsilon,$

2. $u_i g_i(\mathbf{x}) \ = \ 0,$ *for* $i = 1, 2, \ldots, m.$

The Mangasarian-Fromovitz constraint qualification (MFCQ for short) for the problem (P) with smooth data is given as follows : Let $\mathbf{x}$ be a feasible point of (P). Then MFCQ holds at $\mathbf{x}$ if there exists $d \in \Re^n$ such that $\langle \nabla g_i(\mathbf{x}), d\rangle < 0$ for all $i \in I(\mathbf{x})$ where $I(\mathbf{x})$ denotes the index of active constraints at x, i.e. $I(\mathbf{x}) = \{i : g_i(x) = 0\}$.
The Mangasarian-Fromovitz constraint qualification (MFCQ) discussed above (equation 6) can be alternatively stated in the following equivalent form, which can be deduced using separation theorem for convex sets.
The constraints of (P) satisfy the MFCQ at a feasible $\mathbf{x}$ if there exists no vector $\mathbf{0} \neq \mathbf{u} \in \Re^m_+$ ($u_i \geq 0$ for $i \in I(\mathbf{x})$ and $u_i = 0$ for $i \notin I(\mathbf{x})$) such that

$$\sum_{i=1}^{m} u_i \nabla g_i(\mathbf{x}) \ = \ 0.$$

Now we will state our main result in the smooth case.

Theorem 3.2 *Let* $\{\mathbf{x}_k\}$ *be a sequence of feasible points of (P) such that* $\mathbf{x}_k \to \bar{\mathbf{x}}$ *as* $k \to \infty$*. Let* $\{\epsilon_k\}$ *be a sequence of positive real numbers such that* $\epsilon_k \downarrow 0,$ *as* $k \to \infty$*. Further assume that for each* k*,* $\mathbf{x}_k$ *is an* ϵ_k*-KKT point of (P). If MFCQ holds at* $\bar{\mathbf{x}}$*, then* $\bar{\mathbf{x}}$ *is a KKT point.*

Proof: Since $\mathbf{x}_k$ is an ϵ_k-KKT point for (P), it is clear from the definition that $\mathbf{x}_k$ is feasible for each k and as each g_i is continuous and $\{\mathbf{x}_k\} \to \bar{\mathbf{x}}$ it is clear that $\bar{\mathbf{x}}$ is a feasible point for (P). Now from the definition of ϵ_k-KKT points there exists a vector $\mathbf{u}^k \in \Re^m_+$ for each k such that

1. $\left\|\nabla f(\mathbf{x}_k) + \sum_{i=1}^{m} u_i^k \nabla g_i(\mathbf{x}_k)\right\| \ \leq \ \epsilon_k,$

2. $u_i^k g_i(\mathbf{x}_k) \ = \ 0$ for $i = 1, 2, \ldots, m.$

Our claim is that the sequence $\mathbf{u}^k$ is bounded. On the contrary assume that $\mathbf{u}^k$ is not bounded. Thus without loss of generality we can sat that $\|\mathbf{u}^k\| \to \infty$ as $k \to \infty$. Now consider the sequence $\{\mathbf{w}^k\}$, with

$$\mathbf{w}^k \ = \ \frac{\mathbf{u}^k}{\|\mathbf{u}^k\|}, \quad \forall\ k.$$

It is clear that $\mathbf{w}^k$ is bounded and hence without loss of generality we can conclude that $\mathbf{w}^k \to \bar{\mathbf{w}}$ and $\|\bar{\mathbf{w}}\| = 1$. Now we have,

$$\left\|\nabla f(\mathbf{x}_k) + \nabla g(\mathbf{x}_k)^T \mathbf{u}^k\right\| \ \leq \ \epsilon_k, \tag{7}$$

where $\nabla g(\mathbf{x})$ denotes the Jacobian matrix at the point $\mathbf{x}$ of the vector function $g : \Re^n \to \Re^m$, given as $g(\mathbf{x}) = [g_1(\mathbf{x}), \ldots, g_m(\mathbf{x})]$.

Now by dividing both sides of equation 7 by $\|\mathbf{u}^k\|$ we have

$$\left\| \frac{1}{\|\mathbf{u}^k\|} \nabla f(\mathbf{x}_k) + \nabla g(\mathbf{x}_k)^T \frac{\mathbf{u}^k}{\|\mathbf{u}^k\|} \right\| \;\leq\; \frac{1}{\|\mathbf{u}^k\|} \epsilon_k.$$

That is,

$$\left\| \frac{1}{\|\mathbf{u}^k\|} \nabla f(\mathbf{x}_k) + \nabla g(\mathbf{x}_k)^T \mathbf{w}^k \right\| \;\leq\; \frac{1}{\|\mathbf{u}^k\|} \epsilon_k. \tag{8}$$

Since f is a smooth function as $\mathbf{x}_k \to \bar{\mathbf{x}}$ we have $\nabla f(\mathbf{x}_k) \to \nabla f(\bar{\mathbf{x}})$ and thus the sequence $\{\nabla f(\mathbf{x}_k)\}$ is bounded and further as $\epsilon_k \to 0$, the sequence $\{\epsilon_k\}$ is bounded. This shows that

$$\frac{1}{\|\mathbf{u}^k\|} \nabla f(\mathbf{x}_k) \to 0 \text{ as } k \to \infty,$$

and

$$\frac{1}{\|\mathbf{u}^k\|} \epsilon_k \to 0 \text{ as } k \to \infty.$$

Thus, passing to the limit in equation 8 as $k \to \infty$, we have $\|\nabla g(\bar{\mathbf{x}})^T \bar{\mathbf{w}}\| \to 0$ (note that since g is smooth $\nabla g(\mathbf{x}_k) \to \nabla g(\bar{\mathbf{x}})$). That is, $\sum_{i=1}^{m} \bar{w}_i \nabla g_i(\bar{\mathbf{x}}) = 0$, where $\bar{\mathbf{w}} = [\bar{w}_1, \bar{w}_2, \ldots, \bar{w}_m]$. Since $\|\bar{\mathbf{w}}\| = 1$, it is clear that MFCQ is violated at $\mathbf{x} = \bar{\mathbf{x}}$. This is a contradiction. Hence, the sequence $\{\mathbf{u}^k\}$ is indeed bounded. Thus, we can assume without loss of generality that $\mathbf{u}^k \to \bar{\mathbf{u}} \in \Re^m_+$ (since $\Re^m_+$ is a closed set). Hence as $k \to \infty$ from items (1) and (2), we have

1. $\left\| \nabla f(\bar{\mathbf{x}}) + \sum_{i=1}^{m} \bar{u}_i \nabla g_i(\bar{\mathbf{x}}) \right\| \;=\; 0$, and
2. $\bar{u}_i g_i(\bar{\mathbf{x}}) \;=\; 0$, for $i = 1, 2, \ldots, m$.

Hence, $\bar{\mathbf{x}}$ is a KKT point. □

Figure 1 illustrates the outcome of the above theorem. At any iterate $\mathbf{x}_k$, the parameter ϵ_k satisfying the two conditions given in theorem 3.2 is obtained. If for a sequence of iterates $\mathbf{x}_k$ (as $k \to \infty$) the corresponding ϵ_k reduces to zero and if at the limit point ($= \bar{\mathbf{x}}$) the MFCQ holds, then the point $\bar{\mathbf{x}}$ is the KKT point.

Remark 1 *It is clear from the above theorem that if f and g_i, $i = 1, 2, \ldots, m$ are differentiable convex functions, then $\bar{\mathbf{x}}$ as in the above theorem is a solution of the problem. However, an important question is whether the sequence $\{\mathbf{x}_k\}$ will converge at all. Of course, if the set*

$$C \;=\; \left\{\, \mathbf{x} : g_i(\mathbf{x}) \leq 0, \quad i = 1, 2, \ldots, m \right\}$$

is compact then $\{\mathbf{x}_k\}$ will have a subsequence which will converge and that would be enough for our purposes. Further, in many simple situations C is actually compact. Consider for example,

$$C \;=\; \left\{(x, y) \in \Re^2 :\; x^2 + y^2 \leq 1,\; x + y \leq 1\right\}.$$

It is simple to observe that C is compact.

3.2 KKT-Error Measure using ϵ-KKT Points

We now suggest a KKT-error measure based on the above definition of an ϵ-KKT point. To compute the KKT-error measure, we only consider constraints which are active with the following active index set: $I(\mathbf{x}_k) = \{\ i \mid g_i(\mathbf{x}_k) = 0\}$. Then, we solve the following optimization problem to find the Lagrange multipliers for all active constraints:

$$\begin{array}{rl} \text{Minimize} & \left\|\ \nabla f(\mathbf{x}_k) + \sum_{i \in I(\mathbf{x}_k)} u_i \nabla g_i(\mathbf{x}_k)\ \right\|, \\ \text{Subject to} & u_i \geq 0 \quad \forall\ i \in I(\mathbf{x}_k). \end{array} \tag{9}$$

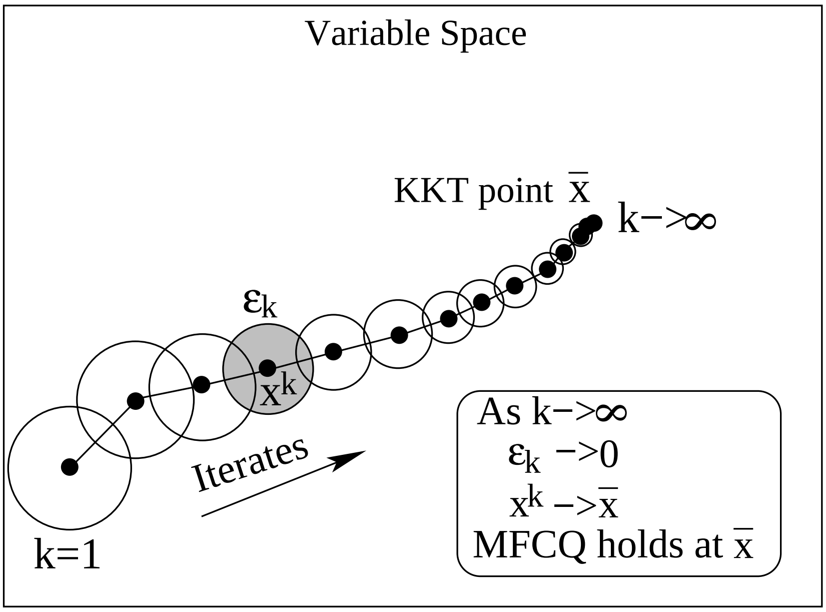


Figure 1: A sketch explaining Theorem 3.2.

Note that only u_i's for active constraints are variables to the above problem. The optimal solution of the above problem is the KKT-error (ϵ_k) at $\mathbf{x}_k$, or

$$\epsilon_k = \| \nabla f(\mathbf{x}_k) + \sum_{i \in I} u_i^* \nabla g_i(\mathbf{x}_k) \|. \tag{10}$$

We now consider a two-variable problem and illustrate how the KKT error (ϵ_k) changes as the iterates $\mathbf{x}_k$ approach the optimal point (also a KKT point).

Consider the problem (Figure 2):

$$\begin{array}{rrcl} \text{Minimize} & f(\mathbf{x}) & = & x^2 + y^2 - 10x + 4y + 2, \\ \text{Subject to} & g_1(\mathbf{x}) & = & x^2 + y - 6 \leq 0, \\ & g_2(\mathbf{x}) & = & x - y \leq 0, \\ & g_3(\mathbf{x}) & = & -x \leq 0. \end{array} \tag{11}$$

It can be verified that the point $\mathbf{x}^* = (1.5, 1.5)^T$ is the global optimum (also a KKT point since f and g_i are convex and the Slater constraint qualification holds). At $\mathbf{x}^*$, the second constraint is active. Now, consider a point close to the optimum, say $(1.495, 1.505)^T$. At this feasible point, none of the constraints are active, and the KKT-error is simply $\|\nabla f\|$ which is not equal to zero. Further, consider sequences of points approaching $\mathbf{x}^*$ along two paths, (i) $x = y$ and (ii) $x + y = 3$, with their KKT-error plots shown in Figures 3(a) and 3(b), respectively.

On the path $x = y$, the constraint g_2 (equation 11) is active and the error smoothly reduces to zero at the KKT point. However, on the path $x + y = 3$, all constraints are inactive along this line (except at the KKT point itself) and we can observe the discontinuous behavior of the error.

The above results show that the sequence of KKT-error values computed keeping the complementary slackness condition depends on the manner the iterates approach the KKT point. For certain sequence of points, the KKT-error can remain quite high in the vicinity of a KKT point before suddenly dropping to zero at it. Also, no particular relationship can be obtained between the KKT-error and the proximity of an iterate to the optimum. Thus, the KKT-error computed

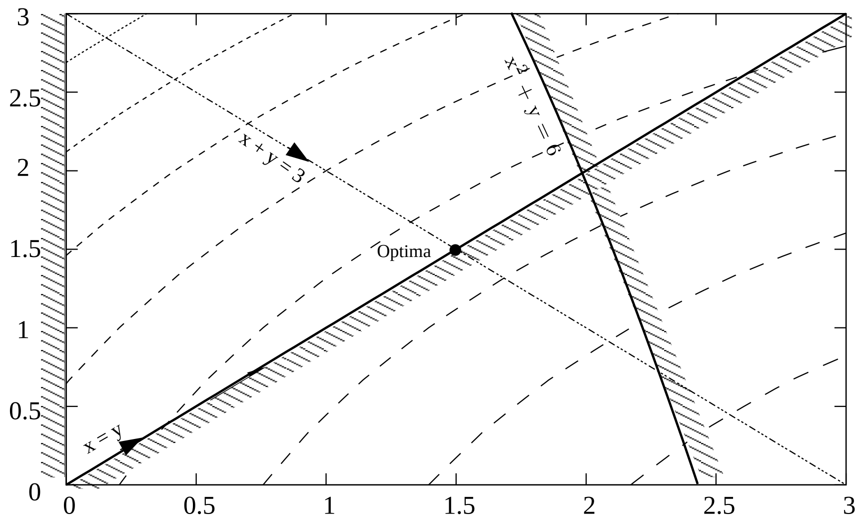
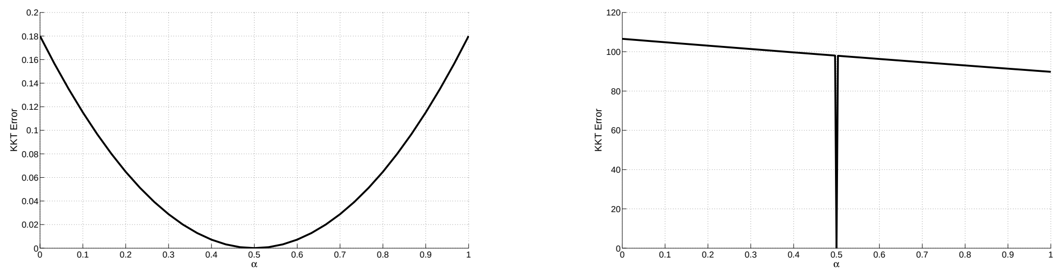


Figure 2: Contour plot of the objective and feasible region.



(a) The KKT-Error reduces smoothly to zero at x^* while traversing along $x = y = 1.35 + 0.3\alpha$ keeping the constraint active.

(b) The KKT-Error has a discontinuity at x^*, where it is zero. Plot of points along the line $x = 1.35 + 0.3\alpha, y = 1.65 - 0.3\alpha$

Figure 3: Behavior of KKT-Error.

keeping strict complementary slackness condition does not give us any information about the proximity from the optimum and hence, cannot be used as a reliable termination criterion for an optimization algorithm.

A careful observation reveals that this happens mainly due to the discontinuity enforced by the complimentary slackness condition (equation 4). For a feasible point very near to a constraint boundary, but not on the boundary, u_i must be zero whereas as soon as the point is on the constraint boundary, u_i is allowed to take any non-negative value. Hence, the participation of the gradient vector of the corresponding constraint in the equilibrium equation (equation 2) may be sudden, thereby causing a fluctuation in the KKT-error value.

Thus, in order to use the KKT-error as a KKT-proximity measure or as a termination criterion for an algorithm so that the progress of the algorithm towards the optimum solution can be measured by its magnitude, we need to relax the complimentary slackness condition. In the following section, we discuss a couple of such recent efforts.

Please note that Deb et al. [6] have suggested a KKT based technique for establishing KKT-optimality conditions for multi-objective optimization problems. Since KKT conditions for multi-objective optimization involve additional multipliers related to objective functions [11], the optimization task involved in computing the KKT-error has a greater flexibility in reducing its value. However, in handling single-objective optimization, there is no additional parameter for the objective and the flexibility only comes from the active constraints. Although we do not consider multi-objective optimization problems in this paper, we are currently extending the ideas of this paper to such problems.

3.3 Modified ϵ-KKT Point

In order to modify the ϵ-KKT point defined in the above subsection, we now consider the problem (P) with locally Lipschitz data. This means that both the objective f and the constraints g_i are locally Lipschitz and not necessarily differentiable. We will begin by introducing the notion of a modified ϵ-KKT point.

Definition 3.3 *A point* $\mathbf{x}$ *which is feasible for (P) is said to be a modified* ϵ*-KKT point for a given* $\epsilon > 0$ *if there exists* $\hat{\mathbf{x}} \in \Re^n$ *such that* $\|\mathbf{x} - \hat{\mathbf{x}}\| \leq \sqrt{\epsilon}$ *and there exists* $\boldsymbol{\zeta} \in \partial^o f(\hat{\mathbf{x}})$ *and* $\boldsymbol{\psi}_i \in \partial^o g_i(\hat{\mathbf{x}})$ *and scalars* $u_i \geq 0$ *for* $i = 1, \ldots, m$ *such that*

1. $\|\boldsymbol{\zeta} + \sum_{i=1}^m u_i \boldsymbol{\psi}_i\| \ \leq \ \sqrt{\epsilon},$
2. $\displaystyle\sum_{i=1}^m u_i g_i(\mathbf{x}) \ \geq \ -\epsilon.$

Interestingly, there is no restriction for $\hat{\mathbf{x}}$ to be feasible. Although the first condition is defined for $\hat{\mathbf{x}}$, the second condition must be true for the original point $\mathbf{x}$.

The above definition is given for non-smooth objective and constraint functions. The benefit of being able to choose $\hat{\mathbf{x}}$ is that, for a point $\mathbf{x}$ of non-differentiability, we can avoid computing the subdifferentials by choosing a point $\hat{\mathbf{x}}$ where the functions are differentiable. If the objective function and constraints are continuously differentiable at $\hat{\mathbf{x}}$ or for smooth cases, where every function is smooth, the subdifferentials can be replaced by the gradient vectors putting, $\boldsymbol{\zeta} = \nabla f(\hat{\mathbf{x}})$ and $\boldsymbol{\psi}_i = \nabla g_i(\hat{\mathbf{x}})$. Also for smooth functions, the condition $\|\mathbf{x} - \hat{\mathbf{x}}\| \leq \sqrt{\epsilon}$ is trivially satisfied for all non-negative value of ϵ, if we take $\hat{\mathbf{x}} = \mathbf{x}$. Hence, for smooth cases, the difference between the Definitions 3.1 and 3.3 lie only in the relaxation of the complimentary slackness condition.

It is interesting to note that in general given a feasible point of the problem (P) it is possible find an $\epsilon > 0$ with respect to which the given point is an modified ϵ-KKT point. Let us consider for simplicity the problem (P) with smooth data and let us consider the definition of the modified ϵ-KKT point as considered above for the smooth case. Then we can simply evaluate ϵ by solving the following problem

$$
\begin{array}{rl}
\text{Minimize} & \epsilon, \\
\text{Subject to} & \left\| \nabla f(\mathbf{x}) + \sum_{i=1}^m u_i \nabla g_i(\mathbf{x}) \right\| \ \leq \ \sqrt{\epsilon}, \\
& \sum_{i=1}^m u_i g_i(\mathbf{x}) \ \geq \ -\epsilon, \\
& u_i \ \geq \ 0 \quad \forall\, i.
\end{array}
$$

The value of ϵ which solves this problem will be referred to as the KKT proximity measure. We will discuss this in more details in subsections 3.4 and 3.5.

The main results below shall be considered in the non-smooth setting. The famous Ekeland's variational principle (EVP) will play a pivotal role and we state it below.

Theorem 3.4 *(Ekeland's Variational Principle) Let* $f : \Re^n \to \Re \cup \{+\infty\}$ *be a proper lower-semicontinuous function which is bounded below on* $\Re^n$*. Let* $\epsilon > 0$ *be given and let* $\mathbf{x} \in \Re^n$ *is such a point for which we have*

$$
f(\mathbf{x}) \ \leq \ \inf_{\mathbf{y} \in \Re^n} f(\mathbf{y}) + \epsilon.
$$

Then for any $\gamma > 0$ there exists $\hat{\mathbf{x}} \in \Re^n$ such that

1. $\|\mathbf{x} - \hat{\mathbf{x}}\| \ \leq \ \gamma$,
2. $f(\hat{\mathbf{x}}) \ \leq \ f(\mathbf{x}) \ \leq \inf_{\mathbf{y} \in C} f(\mathbf{y}) + \epsilon$, *and*
3. $\hat{\mathbf{x}}$ *is the solution of the problem*

$$\min_{\mathbf{y} \in C} f(\mathbf{y}) + \frac{\epsilon}{\gamma}\|\mathbf{y} - \hat{\mathbf{x}}\|.$$

It is important to note that Ekeland's variational principle was originally given for an optimization problem over a complete metric space. Further any closed subset of a complete metric space is a also a complete metric space and thus we could have have presented the Ekeland's variational principle in terms of a closed subset of $\Re^n$ and considering f to be just finite valued.

The natural question to ask is whether the modified ϵ-KKT point arises in a natural way. We show that at least in the case when (P) is a convex problem, it is indeed the case. We show this fact through the following theorem.

Theorem 3.5 *Let us consider the problem (P) where f and each g_i, $i = 1,,\ldots,m$ is a convex function. Let $\mathbf{x}$ be a feasible point which is an ϵ-minimum of (P). That is,*

$$f(\mathbf{x}) \ \leq \ \inf_{\mathbf{y} \in C} f(\mathbf{y}) + \epsilon.$$

Assume further that the Slater's constraint qualification holds, that is, there exists a vector $\mathbf{x}^* \in \Re^n$ such that $g_i(\mathbf{x}^*) < 0$, for all $i = 1,\ldots,m$. Then $\mathbf{x}$ is a modified ϵ-KKT point.

Proof: Since $\mathbf{x}$ is an ϵ-minimum of the convex problem it is clear that there is no $\mathbf{x}' \in \Re^n$ which satisfies the system

$$\begin{aligned} f(\mathbf{x}') - f(\mathbf{x}) + \epsilon &< 0, \\ g_i(\mathbf{x}') &< 0, \quad i = 1,\ldots,m. \end{aligned}$$

Now using standard separation arguments (or the Gordan's theorem of the alternative) we conclude that there exists a vector $\mathbf{0} \neq (u_0, \mathbf{u}) \in \Re_+ \times \Re^m_+$ such that for all $\mathbf{x}' \in \Re^n$

$$u_0(f(\mathbf{x}') - f(\mathbf{x})) + u_0\epsilon + \sum_{i=1}^m u_i g_i(\mathbf{x}') \ \geq \ 0. \tag{12}$$

Suppose $u_0 = 0$. Then, from equation 12 we have

$$\sum_{i=1}^m u_i g_i(\mathbf{x}') \ \geq \ 0, \quad \forall \ \mathbf{x}' \in \Re^n. \tag{13}$$

Since Slater's constraint qualification holds, we have $\sum_{i=1}^m u_i g_i(\mathbf{x}^*) < 0$. This contradicts equation 13. Hence, $u_0 > 0$ and without loss of generality, we can set $u_0 = 1$. Equation 12 becomes

$$f(\mathbf{x}') - f(\mathbf{x}) + \epsilon + \sum_{i=1}^m u_i g_i(\mathbf{x}') \ \geq \ 0, \quad \forall \ \mathbf{x}' \in \Re^n. \tag{14}$$

Now putting $\mathbf{x}' = \mathbf{x}$, we have

$$\sum_{i=1}^m u_i g_i(\mathbf{x}) \ \geq \ -\epsilon.$$

This establishes item 2 in the definition of a modified ϵ-KKT point. Now setting,

$$L(\mathbf{x}', \mathbf{u}) \; = \; f(\mathbf{x}') + \sum_{i=1}^{m} u_i g_i(\mathbf{x}'),$$

we have from equation 14

$$L(\mathbf{x}', \mathbf{u}) \; \geq \; L(\mathbf{x}, \mathbf{u}) - \epsilon, \quad \forall \; \mathbf{x}' \in \Re^n. \tag{15}$$

Thus, $\mathbf{x}$ is the ϵ-minimum of $L(\cdot, \mathbf{u})$ over $\Re^n$. Now applying the Ekeland's variational principle we have by setting $\gamma = \sqrt{\epsilon}$, that there exists $\hat{\mathbf{x}} \in \Re^n$ such that $\|\mathbf{x} - \hat{\mathbf{x}}\| \leq \sqrt{\epsilon}$ and $\hat{\mathbf{x}}$ solves the convex problem

$$\min_{\mathbf{x}' \in \Re^n} L(\mathbf{x}', \mathbf{u}) + \sqrt{\epsilon} \, \|\mathbf{x}' - \hat{\mathbf{x}}\|.$$

Hence we have

$$0 \in \partial_x (L(., \mathbf{u}) + \sqrt{\epsilon} \, \|. - \hat{\mathbf{x}}\|)(\hat{\mathbf{x}}).$$

Now using the sum rule for the subdifferentials of a convex function and further noting that subdifferential of the norm function at the origin is the unit ball we have

$$\mathbf{0} \; \in \; \partial_{\mathbf{x}} L(\hat{\mathbf{x}}, u) + \sqrt{\epsilon} \, \mathbf{B}_{\Re^n},$$

where $\mathbf{B}_{\Re^n}$ denotes the unit ball in $\Re^n$ and $\partial - \mathbf{x}$ denotes subdifferentiation with respect to the first variable. . Hence, using again usual rules of convex analysis, we have

$$\mathbf{0} \; \in \; \partial f(\hat{\mathbf{x}}) + \sum_{i=1}^{m} u_i \partial g_i(\hat{\mathbf{x}}) + \sqrt{\epsilon} \, \mathbf{B}_{\Re^n}.$$

Thus, there exists $\boldsymbol{\zeta} \in \partial f(\hat{\mathbf{x}})$ and $\boldsymbol{\psi}_i \in \partial g_i(\hat{\mathbf{x}})$ and $\mathbf{b} \in \mathbf{B}_{\Re^n}$ such that

$$\mathbf{0} \; = \; \boldsymbol{\zeta} + \sum_{i=1}^{m} u_i \boldsymbol{\psi}_i + \sqrt{\epsilon} \, \mathbf{b}.$$

Hence, $\left\| \boldsymbol{\zeta} + \sum_{i=1}^{m} u_i \boldsymbol{\psi}_i \right\| \leq \sqrt{\epsilon}$. This establishes the result. □

Before stating the next result let us mention the non-smooth version of MFCQ that we need in the sequel. We shall call this the *Basic constraint qualification (BCQ)*. The problem (P) satisfies BCQ at $\mathbf{x}$ if there exists no $\mathbf{u} \in \Re^m_+$ with $\mathbf{u} \neq 0$ and $u_i \geq 0$, for all $i \in I(\mathbf{x})$ and $u_i = 0$ for $i \notin I(\mathbf{x})$ such that

$$\mathbf{0} \; \in \; \sum_{i=1}^{m} u_i \partial^o g_i(\mathbf{x}).$$

Theorem 3.6 *Let us consider the problem (P) with locally Lipschitz objective function and constraints. Let $\{\mathbf{x}_k\}$ be a sequence of vectors feasible to (P) and let $\mathbf{x}_k \to \bar{\mathbf{x}}$ as $k \to \infty$. Consider $\{\epsilon_k\}$ to be a sequence of positive real numbers such that $\epsilon_k \downarrow 0$ as $k \to \infty$. Further assume that for each k, $\mathbf{x}_k$ is a modified ϵ_k-KKT point of (P). Let the Basic constraint qualification (BCQ) hold at $\bar{\mathbf{x}}$. Then $\bar{\mathbf{x}}$ is a KKT point of (P).*

Proof: Since each $\mathbf{x}_k$ is a modified ϵ_k-KKT point, for each k there exists $\mathbf{y}_k$ with $\|\mathbf{x}_k - \mathbf{y}_k\| \leq \sqrt{\epsilon_k}$ and there exists $\boldsymbol{\zeta}^k \in \partial^o f(\mathbf{y}_k)$, $\boldsymbol{\psi}_i^k \in \partial^o g_i(\mathbf{y}_k)$, $i = 1, 2, \ldots m$ and scalars $u_i^k \geq 0$, $i = 1, 2 \ldots m$ such that

(i) $\left\| \boldsymbol{\zeta}^k + \sum_{i=1}^{m} u_i^k \boldsymbol{\psi}_i^k \right\| \; \leq \; \sqrt{\epsilon_k}$,

(ii) $\sum_{i=1}^{m} u_i^k g_i(\mathbf{x}_k) \;\geq\; -\epsilon_k$.

Let us first show that $\{\mathbf{u}^k\}$ is bounded. We assume on the contrary that $\{\mathbf{u}^k\}$ is unbounded. Thus, without loss of generality, let us assume that $\|\mathbf{u}^k\| \to \infty$ as $k \to \infty$. Now consider the sequence $\mathbf{w}^k = \frac{\mathbf{u}^k}{\|\mathbf{u}^k\|}$. Then $\{\mathbf{w}^k\}$ is a bounded sequence and hence has a convergent subsequence. Thus, without loss of generality we can assume that $\mathbf{w}^k \to \bar{\mathbf{w}}$. Further it is clear that $\|\bar{\mathbf{w}}\| = 1$. Now observe the following:

$$\|\mathbf{y}_k - \bar{\mathbf{x}}\| \;\leq\; \|\mathbf{y}_k - \mathbf{x}_k\| + \|\mathbf{x}_k - \bar{\mathbf{x}}\|.$$

Hence,

$$\|\mathbf{y}_k - \bar{\mathbf{x}}\| \;\leq\; \sqrt{\epsilon_k} \;+\; \|\mathbf{x}_k - \bar{\mathbf{x}}\|.$$

Now as $k \to \infty$, $\epsilon_k \downarrow 0$ and $\mathbf{x}_k \to \bar{\mathbf{x}}$. This shows that $\mathbf{y}_k \to \bar{\mathbf{x}}$. Since, the Clarke subdifferential is locally bounded, the sequences $\{\boldsymbol{\zeta}^k\}$ and $\{\boldsymbol{\psi}_{ik}\}$, $i = 1, 2, \ldots, m$ are bounded. Thus, without loss of generality we can conclude that $\boldsymbol{\psi}_i^k \to \bar{\boldsymbol{\psi}}_i$ for all $i = 1, 2, \ldots m$. From (i) we have:

$$\frac{1}{\|\mathbf{u}^k\|}\;\|\boldsymbol{\zeta}^k + \sum_{i=1}^{m} u_i^k \boldsymbol{\psi}_i^k\| \;\leq\; \frac{1}{\|\mathbf{u}^k\|}\sqrt{\epsilon_k}.$$

Thus,

$$\|\;\frac{\boldsymbol{\zeta}^k}{\|\mathbf{u}^k\|} + \sum_{i=1}^{m} w_i^k \boldsymbol{\psi}_i^k\;\| \;\leq\; \frac{1}{\|\mathbf{u}^k\|}\sqrt{\epsilon_k}. \tag{16}$$

where $\mathbf{w}^k = [w_1^k, w_2^k, \ldots w_m^k]$, and $\mathbf{w}^k = \frac{\mathbf{u}^k}{\|\mathbf{u}^k\|}$. Now as $\epsilon_k \downarrow 0$, $\sqrt{\epsilon_k} \to 0$ and hence, $\{\sqrt{\epsilon_k}\}$ is a bounded sequence and

$$\frac{1}{\|\mathbf{u}^k\|}\sqrt{\epsilon_k} \to 0 \quad as \quad k \to \infty.$$

Further as $\{\boldsymbol{\zeta}^k\}$ is a bounded sequence, we have $\frac{\boldsymbol{\zeta}^k}{\|\mathbf{u}^k\|} \to 0$ as $k \to \infty$. Hence from 16, we have

$$\|\sum_{i=1}^{m} \bar{w}_i \bar{\boldsymbol{\psi}}_i\| \;\leq\; 0.$$

i.e. $\sum_{i=1}^{m} \bar{w}_i \bar{\boldsymbol{\psi}}_i = 0$ where $\bar{\mathbf{w}} = [\bar{w}_1, \bar{w}_2, \ldots \bar{w}_m]$ and $\mathbf{w}^k \to \bar{\mathbf{w}}$. As $\|\bar{\mathbf{w}}\| = 1$, it shows that the Basic constraint qualification is violated. Hence, $\{\mathbf{u}^k\}$ is a bounded sequence and thus we can assume that $\mathbf{u}^k \to \bar{\mathbf{u}}$, $\bar{\mathbf{u}} \in \Re_+^m$. Further as $\{\boldsymbol{\zeta}^k\}$ is bounded we can assume that $\boldsymbol{\zeta}^k \to \bar{\boldsymbol{\zeta}}$ and since $\partial^o f$ is graph closed, $\bar{\boldsymbol{\zeta}} \in \partial^o f(\bar{\mathbf{x}})$. Moreover since each ∂g_i is graph closed we have $\bar{\boldsymbol{\psi}}_i \in \partial g_i(bo\bar{l}dx)$ for all $i = 1, \ldots, m$. Hence, from (i) we can have $\|\bar{\boldsymbol{\zeta}} + \sum_{i=1}^{m} \bar{u}_i \bar{\boldsymbol{\psi}}_i\| \leq 0$, where $\bar{\mathbf{u}} = [\bar{u}_1, \bar{u}_2, \ldots, \bar{u}_m]$. Thus, $\bar{\boldsymbol{\zeta}} + \sum_{i=1}^{m} \bar{u}_i \bar{\boldsymbol{\psi}}_i = 0$. This shows that

$$\mathbf{0} \;\in\; \partial^o f(\bar{\mathbf{x}}) + \sum_{i=1}^{m} \bar{u}_i \partial^o g_i(\bar{\mathbf{x}}).$$

From (ii) we have $\sum_{i=1}^{m} \bar{u}_i g_i(\bar{\mathbf{x}}) \geq 0$. Now since $\mathbf{x}_k$ is feasible we have $g_i(\mathbf{x}_k) \leq 0$, for all $i = 1, 2, \ldots m$. Hence, $g_i(\bar{\mathbf{x}}) \leq 0$ for all $i = 1, 2, \ldots m$. Hence, $\bar{\mathbf{x}}$ is feasible to (P). Since $\bar{u}_i \geq 0$, we have $\sum_{i=1}^{m} \bar{u}_i g_i(\bar{\mathbf{x}}) \leq 0$. Therefore,

$$\sum_{i=1}^{m} \bar{u}_i g_i(\bar{\mathbf{x}}) \;=\; 0.$$

Hence $\bar{\mathbf{x}}$ is a KKT point of (P). □

With the revised conditions stated in the above theorem, Figure 1 can be used to understand the principle stated in the theorem with the modified ϵ-KKT points.

Theorem 3.7 *Let us consider the convex problem (P). Let the Slater's constraint qualification holds, that is, there exists* $\mathbf{x}^* \in \Re^n$ *such that* $g_i(\mathbf{x}^*) < 0$ *for all* $i = 1, \dots, m$. *Let* $\{\mathbf{x}_k\}$ *be a sequence of* ϵ_k*-minimum points and let* $\mathbf{x}_k \to \bar{\mathbf{x}}$ *as* $\epsilon_k \downarrow 0$. *Then,* $\bar{\mathbf{x}}$ *is a minimum point of (P).*

Proof: Since $\mathbf{x}_k$ is an ϵ_k-minimum point, by applying Theorem 3.5 we have $\mathbf{x}_k$ is modified ϵ_k-KKT point for (P) for each k. Now applying Theorem 3.6 we have $\bar{\mathbf{x}}$ is a KKT point for (P) and hence $\bar{\mathbf{x}}$ is a solution of (P). Note that the Slater's Constraint Qualification implies the Basic Constraint qualification in the case of a convex optimization problem. □

In all the above results in this section we consider a sequence $\{\mathbf{x}_k\}$ satisfying the modified ε-KKT conditions converging to $\bar{\mathbf{x}}$ which turns out to be a KKT point if a regularity condition is satisfied. It is also interesting to ask the reverse question. Given a KKT point or a local minimum of the problem (P) it is interesting to ask whether there exists a sequence of points converging to it such that each element of the sequence satisfies certain approximate optimality condition. The following results tries to answer this question at least partially.

Theorem 3.8 *Let us consider the problem (P) with locally Lipschitz data. We assume that the constraint functions* g_i , $i = 1, \dots, m$ *are convex and the Slater constraint qualification hold. Let* $\bar{x}$ *be a local minimum for the problem (P). Further consider a sequence of positive numbers* $\{\epsilon_k\}$ *with* $\epsilon_k \downarrow 0$. *Then there exists a feasible sequence* $\{\mathbf{x}_k\}$ *with* $\mathbf{x}_k \to \bar{\mathbf{x}}$ *such that for* k *sufficiently large there exists an element* $\mathbf{x}_{j_k}$ *of* $\{\mathbf{x}_k\}$ *corresponding to which there exists another element* $\hat{\mathbf{x}}_{j_k}$ *such that* $||\mathbf{x}_{j_k} - \hat{\mathbf{x}}_{j_k}|| \leq \sqrt{\epsilon_k}$ *and there exists* $\boldsymbol{\zeta}_0^k \in \partial^\circ f(\hat{\mathbf{x}}_{j_k}), \boldsymbol{\psi}_i^k \in \partial g_i(\hat{\mathbf{x}}_{j_k})$ *and* $\hat{\lambda}_i \geq 0$ *such that*

i) $||\boldsymbol{\zeta}_0^k + \sum_{i=1}^m \hat{\lambda}_i \boldsymbol{\psi}_i^k|| \leq \sqrt{\epsilon_k}$,

ii) $\hat{\lambda}_i g_i(\hat{\mathbf{x}}_{j_k}) = 0$, *for all* $i = 1, \dots, m$.

Proof: Since $\bar{\mathbf{x}}$ is a local minimum there exists $\delta > 0$ such that $f(\mathbf{x}) \geq f(\bar{\mathbf{x}})$ for all $x \in \overline{B_\delta(\bar{\mathbf{x}})} \cap C)$, where $\overline{B_\delta(\bar{\mathbf{x}})}$ denotes the closure of the open ball of radius $\delta > 0$ centered at $\bar{x}$ which is denoted at $B_\delta(\bar{\mathbf{x}})$ and C the feasible set of the problem (P). For simplicity let us denote $\overline{B_\delta(\bar{\mathbf{x}})} \cap C$ by X and it is simple to observe that closed and bounded convex set. Thus there exists a sequence $\{\mathbf{x}_k\}$ in X with $\mathbf{x}_k \neq \bar{\mathbf{x}}$ and $\mathbf{x}_k \to \bar{\mathbf{x}}$. Since f is locally Lipschitz we have $f(\mathbf{x}_k) \to f(\overline{\mathbf{x}})$. Then, for any natural number k and $\epsilon_k > 0$ there exists an element $\mathbf{x}_{j_k}$ such that

$$|f(\mathbf{x}_{j_k}) - f(\bar{\mathbf{x}})| < \epsilon_k.$$

It is simple to observe that when $k_2 > k_1$ we have $\epsilon_{k_2} \leq \epsilon_{k_1}$ and one can choose $\mathbf{x}_{j_{k_1}}$ and $\mathbf{x}_{j_{k_2}}$ such that $j_{k_2} > j_{k_1}$. Hence from the above inequality we see that

$$f(\mathbf{x}_{j_k}) < f(\bar{\mathbf{x}}) + \epsilon_k.$$

Thus $\mathbf{x}_{j_k}$ is an ϵ_k-minimum of f over X. Thus using the Ekeland's variational principle and using the arguments in Theorem 2.1 in Hamel [8] we conclude that there exists $\hat{\mathbf{x}}_{j_k} \in X$ such that $||\mathbf{x}_{j_k} - \hat{\mathbf{x}}_{j_k}|| \leq \sqrt{\epsilon_k}$ along with the fact that $f(\hat{\mathbf{x}}_{j_k}) \leq f(\mathbf{x}_{j_k})$. Further from the Ekeland's variational principle we conclude that $\hat{\mathbf{x}}_{j_k}$ solves the problem

$$\min_{x \in X} f(\mathbf{x}) + \sqrt{\epsilon_k}||\mathbf{x} - \hat{\mathbf{x}}_{j_k}||.$$

Now using the standard necessary optimality condition for locally Lipschitz minimization from Clarke [4] we have

$$0 \in \partial^\circ(f + \sqrt{\epsilon_k}||. - \hat{\mathbf{x}}_{j_k}||)(\hat{\mathbf{x}}_{j_k}) + N_X(\hat{\mathbf{x}}_{j_k}).$$

Now using the sum rule for Clarke subdifferential and noting that subdifferential of the norm function at the origin is the unit ball and putting in the expression for X we have that

$$0 \in \partial^\circ f(\hat{\mathbf{x}}_{j_k}) + \sqrt{\epsilon_k}\mathbf{B}_{\Re^n} + N_{\overline{B_\delta(\bar{\mathbf{x}})} \cap C}(\hat{\mathbf{x}}_{j_k}). \tag{17}$$

Note that when k is sufficiently $\mathbf{x}_{j_k} \in B_\delta(\bar{x}) \cup C$. Again for k sufficiently large we $B_{\sqrt{\epsilon_k}}(\mathbf{x}_{j_k}) \subset B_\delta(\bar{\mathbf{x}})$. This shows that $\hat{\mathbf{x}}_{j_k} \in B_\delta(\bar{\mathbf{x}})$. Hence $N_{\overline{B_\delta(\bar{\mathbf{x}})}}(\hat{\mathbf{x}}_{j_k}) = \{0\}$. Thus using the standard results from convex analysis (see Rockafellar [14]) we have that

$$N_{\overline{B_\delta(\bar{\mathbf{x}})} \cup C}(\hat{\mathbf{x}}_{j_k}) = N_C(\hat{\mathbf{x}}_{j_k}).$$

Since the Slater constraint qualification holds it is well known result in convex analysis (see Rockafellar [14]) that

$$N_C(\hat{\mathbf{x}}_{j_k}) = \left\{ \sum_{i=1}^{m} \lambda_i \partial g_i(\hat{\mathbf{x}}_{j_k}) : \lambda_i \geq 0, \quad \lambda_i g_i(\hat{\mathbf{x}}_{j_k}) = 0, \quad \forall i = 1, \ldots, m \right\}.$$

Hence using (17) we conclude that there exists $\boldsymbol{\zeta}_0^k \in \partial^\circ f(\hat{\mathbf{x}}_{j_k}), \boldsymbol{\psi}_i^k \in \partial g_i(\hat{\mathbf{x}}_{j_k})$ and $\hat{\lambda}_i \geq 0$ such that

i) $||\boldsymbol{\zeta}_0^k + \sum_{i=1}^m \hat{\lambda}_i \boldsymbol{\psi}_i^k|| \leq \sqrt{\epsilon_k}$,

ii) $\hat{\lambda}_i g_i(\hat{\mathbf{x}}_{j_k}) = 0$, for all $i = 1, \ldots, m$.

This establishes the result. □

Using the above definition of modified ϵ-KKT point and associated theorems, and an overall understanding of the interplay between satisfaction of complimentary slackness condition and equilibrium equation, we now define a KKT-proximity measure as follows.

3.4 KKT-Proximity Measure for Smooth Problems

For a feasible iterate $\mathbf{x}_k$, we solve the following optimization problem with $(\epsilon_k, \mathbf{u}^k)$ as the variable vector:

$$\begin{array}{rl} \text{Minimize} & \epsilon_k, \\ \text{Subject to} & \left\| \nabla f(\mathbf{x}_k) + \sum_{i=1}^m u_i \nabla g_i(\mathbf{x}_k) \right\| \leq \sqrt{\epsilon_k}, \\ & \sum_{i=1}^m u_i g_i(\mathbf{x}_k) \geq -\epsilon_k, \\ & u_i \geq 0 \quad \forall i. \end{array} \tag{18}$$

Let us say the optimal solution to the above problem is $(\epsilon_k^*, \mathbf{u}^{k^*})$. Then the KKT-proximity measure is simply ϵ_k^*. As a by-product of this optimization task, we also get the corresponding Lagrange multiplier vector $(\mathbf{u}^{k^*})$.

3.5 KKT-Proximity Measure for Non-Smooth Problems

For a feasible iterate $\mathbf{x}_k$, we solve the following optimization problem with $(\epsilon_k, \hat{\mathbf{x}}, \boldsymbol{\zeta}, \boldsymbol{\psi}, \mathbf{u}^k)$ as the variable vector:

$$\begin{array}{rl} \text{Minimize} & \epsilon_k, \\ \text{Subject to} & \left\| \boldsymbol{\zeta} + \sum_{i=1}^m u_i \boldsymbol{\psi}_i \right\| \leq \sqrt{\epsilon_k}, \\ & \sum_{i=1}^m u_i g_i(\mathbf{x}_k) \geq -\epsilon_k, \\ & ||\hat{\mathbf{x}}_k - \mathbf{x}_k|| \leq \sqrt{\epsilon_k}, \\ & u_i \geq 0, \quad \forall i, \\ & \boldsymbol{\zeta} \in \partial^o f(\hat{\mathbf{x}}_k), \\ & \boldsymbol{\psi}_i \in \partial^o g_i(\hat{\mathbf{x}}_k), \quad \forall i. \end{array} \tag{19}$$

This is the generic formulation for finding the KKT-proximity measure for any $\mathbf{x}_k$. However, the Rademacher's theorem states that a locally Lipschitz function is differentiable almost everywhere. This means that for a point $\mathbf{x}_k$, the functions f and g_i's are densely differentiable on any neighborhood of $\mathbf{x}_k$. Thus, for a locally Lipschitz case, we should be able find $\hat{\mathbf{x}}_k$ where the objective and constraint functions are differentiable. However, this simply means that the gradients belong to the respective Clarke subdifferentials. But, in most cases the functions may be *continuously* differentiable at $\hat{\mathbf{x}}_k$ in which case, the Clarke subdifferentials contain only the gradients. Thus, moving away from the reference point $\mathbf{x}_k$ in most cases will allow us to work with derivatives, thereby eliminating the need of having variables $\boldsymbol{\zeta}$ and $\boldsymbol{\psi}$ in the above optimization problem.

To reiterate, we mention that there is no restriction on $\hat{\mathbf{x}}_k$ to be feasible. Also, the optimum value of problem 19 is defined as the KKT-proximity measure for the iterate $\mathbf{x}_k$.

3.6 Andreani et al.'s Definition

Andreani [1] defined the notion of approximate KKT (AKKT) condition as follows:

Definition 3.9 *A feasible point $\bar{\mathbf{x}}$ satisfies AKKT condition if and only if, there exists sequence of feasible solutions $\{\mathbf{x}_k\} \subset R^n$, $\{\mathbf{u}^k\} \subset R^m_+$ and $\epsilon_k \geq 0$ such that $\mathbf{x}_k \to \bar{\mathbf{x}}$, $\epsilon_k \to 0$ and for all $k \in I\!N$*

1. $\| \nabla f(\mathbf{x}_k) + \sum_{i=1}^{m} u_i^k \nabla g_i(\mathbf{x}_k) \| \leq \epsilon_k$,
2. $u_i^k = 0$, *for all i such that $g_i(\mathbf{x}_k) < -\epsilon_k$.*

These conditions differ from our approximate KKT conditions in the previous section in the sense that they relax the multiplier u_i to be nonzero for some feasible points, lying only in an ϵ_k-proximity to the i-th constraint boundary. This is a more stringent condition than that in our definition and this may not be enough to have adequate number of constraints with non-zero Lagrange multipliers to make the KKT-error close to zero. Our approach relaxes the complimentary slackness conditions more, but only to an extent where there is a balance between the KKT-error and violation of complimentary slackness conditions.

There is another fundamental difference between the two approaches. In Andreani et al.'s definition, the AKKT point is not defined for any arbitrary feasible iterate $\mathbf{x}_k$, rather an AKKT point is the limit point $\bar{\mathbf{x}}$ of a sequence of iterates approaching with the condition that the associated ϵ_k approaches to zero at the AKKT point. On the other hand, in our definition, for every iterate $\mathbf{x}_k$ there is a modified ϵ-KKT point with a ϵ_k that need not be even close to zero. The parameter ϵ_k is such that the violation of the equilibrium equation and complimentary slackness conditions are smaller than it. Moreover, we solve an optimization problem to find a Lagrange multiplier vector $\mathbf{u}$ and ϵ_k such that ϵ_k takes the minimum possible value, thereby making a balance between both equilibrium-error and violation of complimentary slackness conditions. Our theorems suggest that if for a sequence of iterates the corresponding ϵ_k values approach zero and certain other conditions hold, the resulting limiting iterate is a KKT point.

4 Simulation Results

The application of the ideas discussed in the earlier sections on the approximate KKT conditions is that if an algorithm produces a sequence of points $\{\mathbf{x}_k\}$ and if it is possible to find corresponding ϵ_k values exhibiting a reducing sequence to zero, in accordance with either theorems 3.2 or 3.6 then the limit of the sequence is a KKT point.

Here, we consider iterates obtained from a real-coded evolutionary optimization algorithm (RGA). A C-code code is available at `http://www.iitk.ac.in/kangal/codes.shtml`. The RGA uses a population of points in each iteration (called a 'generation' in the evolutionary algorithm literature). At every generation, we consider all the feasible solutions and choose the one having the best objective value as an iterate. If at any generation, no population member is feasible, we skip our KKT-proximity measure computation for that generation.

In every problem, we add the best-known solution in the set of iterates at the end and compute the KKT-proximity measure for that solution as well. A comparison of KKT-proximity measure between the final RGA solution and the best-known solution will indicate the specific RGA's ability to approach the best-known solution to a problem and importantly will reveal whether the best-known solution itself is a KKT point.

4.1 A Demonstration Problem

First, we investigate the results of the above-mentioned scheme for computing the KKT-proximity measure on a demonstration problem.

We consider a simple two-variable, two-constraint problem to illustrate the working of our scheme:

$$\begin{aligned} \text{Minimize} \quad & f(x_1, x_2) = x_1^2 + x_2^2, \\ \text{subject to} \quad & g_1(x_1, x_2) \equiv 3x_1 - x_2 + 1 \ \leq \ 0, \\ & g_2(x_1, x_2) \equiv x_1^2 + (x_2 - 2)^2 - 1 \ \leq \ 0. \end{aligned} \tag{20}$$

Figure 4 shows the feasible space and the optimum point ($\bar{\mathbf{x}} = [0, 1]^T$). We consider a series of

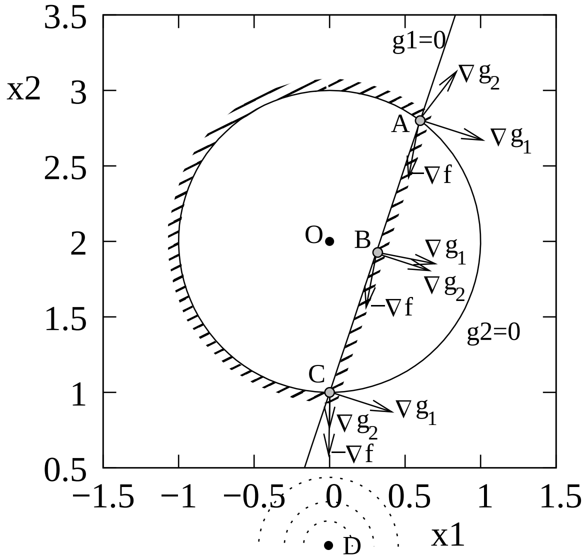


Figure 4: Feasible search space of the demonstration problem.

iterates (point number 0 at A to point number 49 at C) along the linear constraint boundary from A to C and investigate the behavior of KKT-proximity measure estimate scheme (by solving problem 18) and compare it with the KKT-error estimate scheme based on the strict complementary slackness scheme (by solving problem 9). It is clear from the figure that point A will have a large KKT-error value, as no linear combination of ∇g_1 and ∇g_2 vectors will construct $-\nabla f$ at this point. However, as points towards C and inside the circle are considered, the second constraint is inactive and it will have no role to play in the KKT conditions. Thus, for points inside the circle and on the first constraint, only the first constraint participates in the KKT-error calculation for the complementary slackness scheme. It is clear that on none of these points, $\nabla g_1 = -\nabla f$ in order to make a zero error. In fact, the KKT-error reduces from near point A to near point C, as shown in Figure 5.

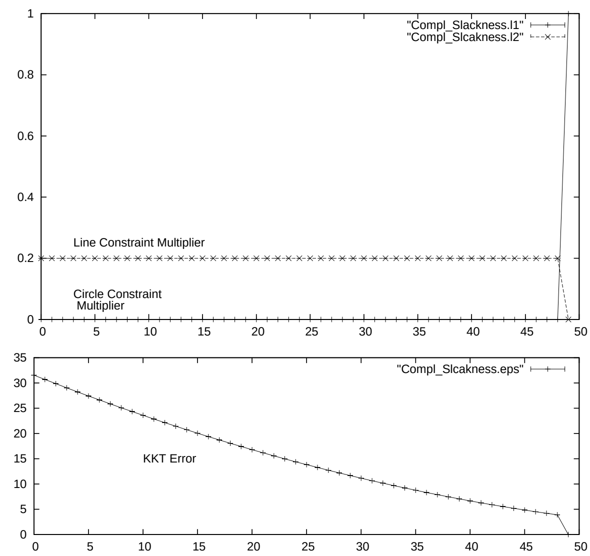


Figure 5: Lagrange multipliers and KKT-error using the complementary slackness scheme.

At point C, the second constraint is active again and ∇g_2 is equal to $-\nabla f$ at C. Thus, the KKT-error will be zero. As it is clear from the figure, in the neighborhood of C, although KKT-error reduces as the point gets closer to C along the line AC, there is a discontinuity of the error at C. Lagrange multipliers for g_1 and g_2 are also shown in Figure 5. Since the second constraint is inactive for points along AC (except A and C), u_2 is zero and at C it is a positive value ($u_2 = 1$). Interestingly, u_1 is constant ($= 0.2$) throughout, except at C, at which it is zero. The KKT-error actually varies as $40x_1^2 + 24x_1 + 3.6$ in the range $0 < x_1 < 0.6$ and at a point near C ($x_1 = 0$) the error is near 3.6. Then exactly at C, the error is zero, making a jump in the error value from near 3.6 to zero, as shown in the figure.

Figure 6 shows the KKT-proximity measure computed using the proposed scheme (equation 18).

Our proposed approach seems to maintain a continuity in the KKT-proximity measure as it reduces to zero. Due to this property, this modified KKT-proximity measure can be used as the termination criterion of an optimization algorithm. Corresponding Lagrange multipliers u_1 and u_2 are also shown in Figure 6. Interestingly, u_2 is zero from A till an intermediate point B. As shown in Figure 4 at points before B, the gradient of constraint g_1 is more directed towards $-\nabla f$ and contributes in minimizing the KKT-error. At around B g_2 is more directed towards $-\nabla f$ and so its Lagrange multiplier becomes nonzero. The roles of g_1 and g_2 essentially are interchanged. Since the scheme allows larger values of u_2, it grows to the extent needed to reduce the KKT-proximity measure. Note that unless $u_2 \nabla g_2$ is equal to $-\nabla f$, the KKT-proximity measure can never be exactly zero, but due to the flexibility in choosing a large enough u_2, the KKT-proximity measure smoothly reduces to zero.

4.2 Numerical Results: Smooth Problems

The procedure proposed in subsection 3.3 is tested on a variety of test problems borrowed from the constrained optimization literature [10].

We take the sequence of best individual of the population, for different generations of a real-

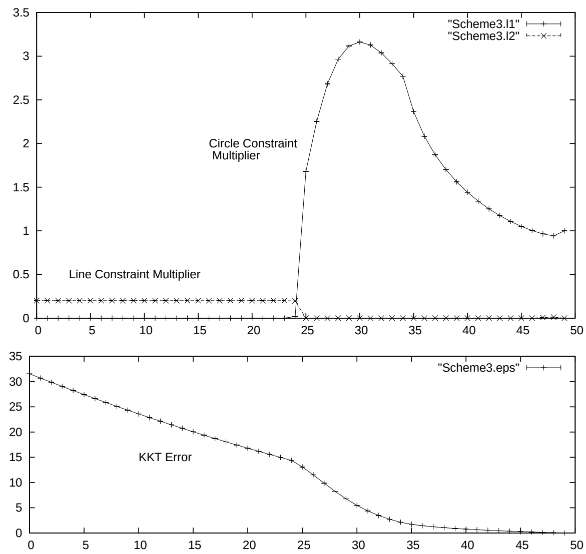


Figure 6: Lagrange multipliers and KKT-error using proposed scheme.

coded genetic algorithm, adapted for handling constraints using a penalty-parameter-less strategy [5]. For the RGA solutions, we first delete contiguously duplicate solutions and then solutions which are infeasible. Thereafter, to each remaining solution, we apply the proposed scheme (equation 18) to compute the discussed KKT-proximity measure. For problems in which the RGA does not converge to the optimum, the reported optimal solution from [10] is manually appended at the end of the RGA's sequence of solutions $\{\mathbf{x}_k\}$. This is done to mainly demonstrate the accuracy of the computation scheme, in checking whether the KKT-proximity measure goes to zero at the optimum or not. Also, since our scheme produces the Lagrange multipliers, we are able to tabulate the multipliers at the optimum for all problems. Lagrange multipliers obtained for the best-known solutions (here, for the first time, confirmed as KKT points) are tabulated in the appendix.

4.2.1 Problem *g01*

The following is Problem g01 from [10] containing 35 constraints and 13 variables.

$$\begin{array}{rl}
\text{Minimize} & f(\mathbf{x}) = 5\sum_{i=1}^{4} x_i - 5\sum_{i=1}^{4} {x_i}^2 - \sum_{i=5}^{13} x_i, \\
\text{subject to} & g_1(\mathbf{x}) \equiv 2x_1 + 2x_2 + x_{10} + x_{11} - 10 \le 0, \\
& g_2(\mathbf{x}) \equiv 2x_1 + 2x_3 + x_{10} + x_{12} - 10 \le 0, \\
& g_3(\mathbf{x}) \equiv 2x_2 + 2x_3 + x_{11} + x_{12} - 10 \le 0, \\
& g_4(\mathbf{x}) \equiv -8x_1 + x_{10} \le 0, \\
& g_5(\mathbf{x}) \equiv -8x_2 + x_{11} \le 0, \\
& g_6(\mathbf{x}) \equiv -8x_3 + x_{12} \le 0, \\
& g_7(\mathbf{x}) \equiv -2x_4 - x_5 + x_{10} \le 0, \\
& g_8(\mathbf{x}) \equiv -2x_6 - x_7 + x_{11} \le 0, \\
& g_9(\mathbf{x}) \equiv -2x_8 - x_9 + x_{12} \le 0, \\
& g_{9+i}(\mathbf{x}) \equiv -x_i \le 0, \quad (\mathrm{i} = 1, \ldots, 13), \\
& g_{22+i}(\mathbf{x}) \equiv x_i - 1 \le 0, \quad (\mathrm{i} = 1, \ldots, 9, 13), \\
& g_{22+i}(\mathbf{x}) \equiv x_i - 100 \le 0, \quad (\mathrm{i} = 10, 11, 12).
\end{array} \tag{21}$$

Solution sequences from 10 RGA runs are taken and the median, best and worst KKT-proximity measures are plotted in Figure 7. In all the runs, the RGA converges to the best-known solution for this problem ($\bar{f} = -15$). Despite some initial fluctuations in the KKT-proximity measure, it finally reduces to zero, indicating that the final RGA point is a KKT point. For this problem, we observe that six constraints $\{g_1,\ g_2,\ g_3,\ g_7,\ g_8,\ g_9\}$ including 10 upper-limit constraints $\{x_i \leq 1,\ \text{for}\ i = 1, \ldots, 9, 13\}$ are active at the KKT point.

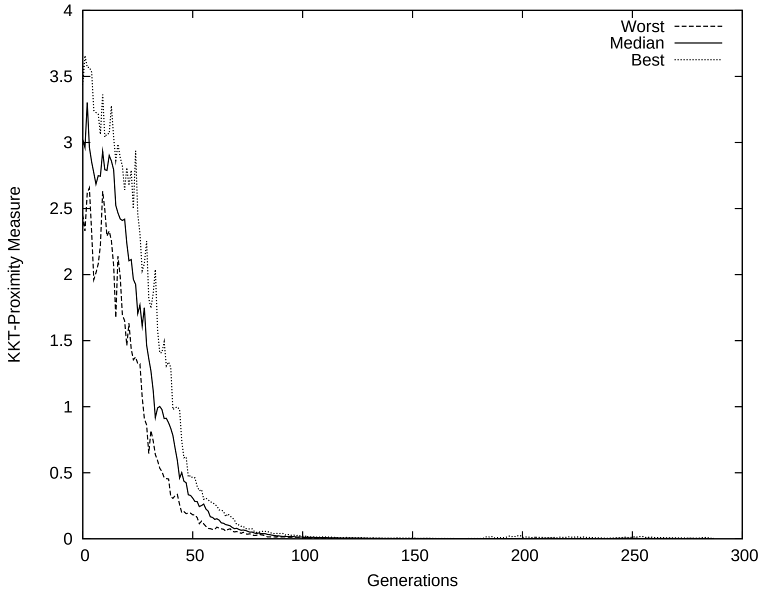


Figure 7: KKT-proximity measure for problem g01.

The real-coded genetic algorithm used in this study to generate the iterates was not theoretical proven to converge to the minimum point, even for a smooth problem. The RGA does not use any gradient information in any of its operators. The RGA compares population members and emphasizes better solutions. Its recombination and mutation operators utilize the selected solutions to create new solutions by using probabilistic operators. The RGA used in this study does not even explicitly preserves the population-best solution, nor does it copy the current best solution to the next generation. But the overall RGA that uses selection, recombination and mutation operators in tandem to update a randomly created population of solutions iteratively seems to take its generation-wise population-best solution towards the KKT point which is theoretically defined through gradients of objectives and constraints and is a likely candidate for the minimum point. Although our conclusion about the specific RGA-obtained solution being a KKT point is correct, it does not prove that the specific RGA we have used here has a convergence proof for any arbitrary problem. But this is the first time we can report that a direct search algorithm, such as the specific RGA we have used here, finds a theoretically significant KKT point for a standard constrained test problem.

4.2.2 Problem *hs23*

The following problem is taken from [9]. It is a two-variable, nine-constraint problem with a quadratic objective function and a number of smooth constraints.

$$
\begin{aligned}
\text{Minimize} \quad & f(\mathbf{x}) = x_1^2 + x_2^2, \\
\text{subject to} \quad & g_1(\mathbf{x}) \equiv 1 - x_1 - x_2 \le 0, \\
& g_2(\mathbf{x}) \equiv 1 - x_1^2 - x_2^2 \le 0, \\
& g_3(\mathbf{x}) \equiv 9 - 9x_1^2 - x_2^2 \le 0, \\
& g_4(\mathbf{x}) \equiv x_2 - x_1^2 \le 0, \\
& g_5(\mathbf{x}) \equiv x_1 - x_2^2 \le 0, \\
& g_{5+i}(\mathbf{x}) \equiv -50 - x_i \le 0, \quad (\mathrm{i} = 1, 2), \\
& g_{7+i}(\mathbf{x}) \equiv x_i - 50 \le 0, \quad (\mathrm{i} = 1, 2).
\end{aligned} \tag{22}
$$

Again, solution sequences from 10 RGA runs are taken and the best, worst and median KKT-proximity measure values are plotted in Figure 8. In all the runs, the RGA converges to the KKT point $(1.0, 1.0)^T$ and the KKT-proximity measure converges to zero between $30 - 70$ generations. At the optimum, g_4 and g_5 are active. Importantly, the specific RGA seems to find the KKT point for this problem as well.

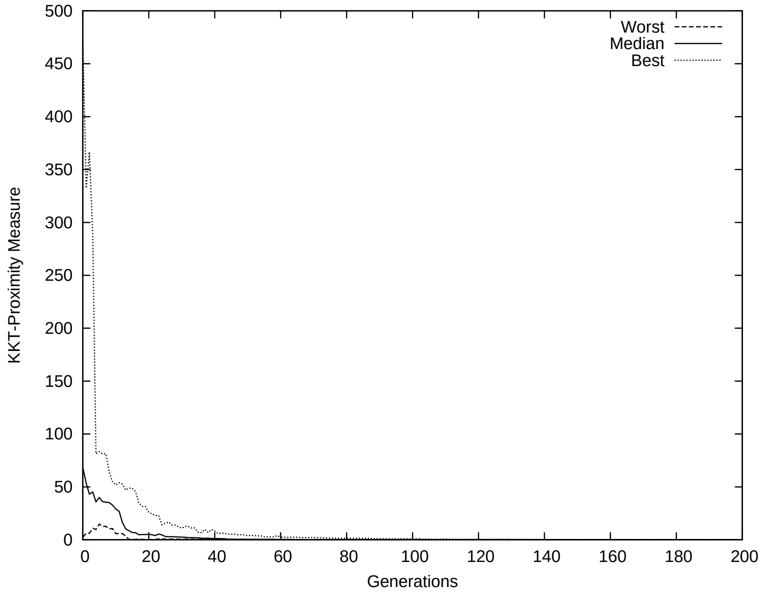


Figure 8: KKT-proximity measure for problem hs23.

4.2.3 Problem *hs45*

The following is problem hs45 from [9]. It is a five-variable, 10-constraint problem.

$$
\begin{aligned}
\text{Minimize} \quad & f(\mathbf{x}) = 2 - \tfrac{1}{120} x_1 x_2 x_3 x_4 x_5, \\
\text{subject to} \quad & g_i(\mathbf{x}) \equiv -x_i \le 0, \quad (\mathrm{i} = 1, \ldots, 5), \\
& g_{5+i}(\mathbf{x}) \equiv x_i - i \le 0, \quad (\mathrm{i} = 1, \ldots, 5).
\end{aligned} \tag{23}
$$

The results in Figure 9 indicate that the best, median and worst KKT-proximity measure values converge to zero close to the 30-th generation, after a large initial fluctuations. All the runs converge to the optimum at $(1.0,\ 2.0,\ 3.0,\ 4.0,\ 5.0)^T$ where g_6, g_7, g_8, g_9, and g_{10} are active. Since, it is evident from the diagram that the worst and the best KKT-proximity measure values display a similar behavior, it suffices to consider only a singular RGA run and plot the KKT-proximity measure. From next problem on, we only show the behavior of one run, however in all cases a similar behavior for 10 runs are observed.

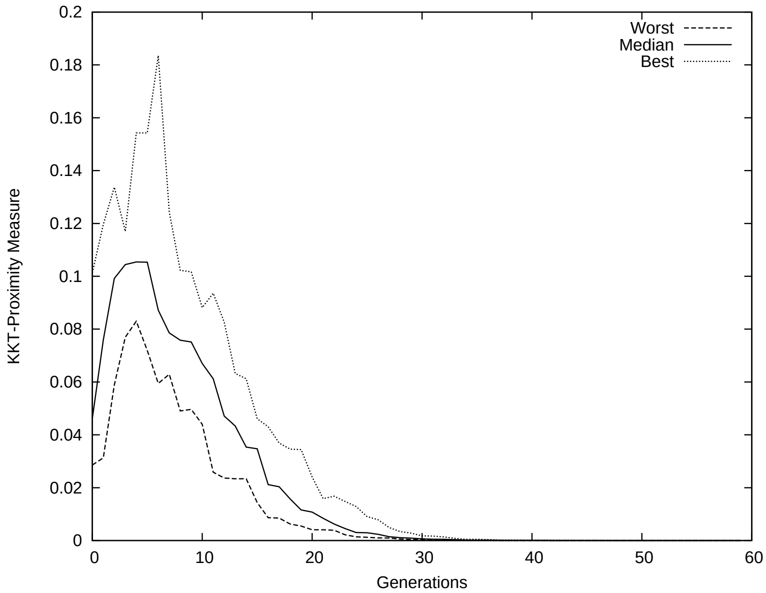


Figure 9: KKT-proximity measure for problem hs45.

4.2.4 Problem *g02*

The following is problem g02 from [10]. It is a 20 variable, two-constraint problem, besides variable bounds.

$$\begin{array}{rl}
\text{Minimize} & f(\mathbf{x}) = -\left|\frac{\sum_{i=1}^{20} cos^4(x_i) - 2\prod_{i=1}^{20} cos^2(x_i)}{\sqrt{\sum_{i=1}^{20} i{x_i}^2}}\right|, \\
\text{subject to} & g_1(\mathbf{x}) \equiv 0.75 - \prod_{i=1}^{20} x_i \leq 0, \\
& g_2(\mathbf{x}) \equiv \sum_{i=1}^{20} x_i - 7.5 * 20 \leq 0, \\
& g_{1+2i}(\mathbf{x}) \equiv -x_i \leq 0, \quad (\text{i} = 1, \ldots, 20), \\
& g_{2+2i}(\mathbf{x}) \equiv x_i - 10 \leq 0, \quad (\text{i} = 1, \ldots, 20).
\end{array} \tag{24}$$

For this problem, the RGA converges to a point at an Euclidean distance of 2.3568 from the best-reported solution [10]. At the best-reported solution, the KKT-proximity measure is found to be exactly zero (Figure 10), meaning that the reported solution (having $\bar{f} = -0.80362$) is a likely candidate for the minimum. Only the first constraint g_1 is found to be active at this point. It is interesting to note that for this problem the specific RGA cannot locate the KKT point, although the best-reported solution is a KKT point. The objective function is highly multi-modal for this problem, the KKT-proximity analysis indicates that the specific RGA used here needs further improvement for solving such multi-modal problems.

4.2.5 Problem *g04*

The following is problem g04 from [10], containing 5 variables and 6 constraints besides variable bounds.

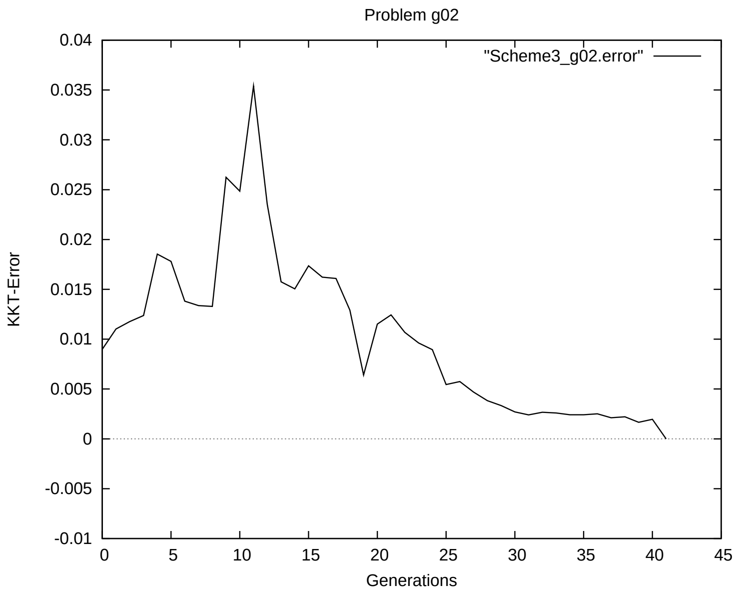


Figure 10: KKT-proximity measure for problem g02.

$$
\begin{aligned}
\text{Minimize} \quad & f(\mathbf{x}) = 5.3578547{x_3}^2 + 0.8356891 x_1 x_5 + \\
& 37.293239 x_1 - 40792.141, \\
\text{subject to} \quad & g_1(\mathbf{x}) \equiv 85.334407 + 0.0056858 x_2 x_5 + 0.0006262 x_1 x_4 - \\
& \quad 0.0022053 x_3 x_5 - 92 \leq 0, \\
& g_2(\mathbf{x}) \equiv -85.334407 - 0.0056858 x_2 x_5 - 0.0006262 x_1 x_4 + \\
& \quad 0.0022053 x_3 x_5 \leq 0, \\
& g_3(\mathbf{x}) \equiv 80.51249 + 0.0071317 x_2 x_5 + 0.0029955 x_1 x_2 + \\
& \quad 0.0021813 {x_3}^2 - 110 \leq 0, \\
& g_4(\mathbf{x}) \equiv -80.51249 - 0.0071317 x_2 x_5 - 0.0029955 x_1 x_2 - \\
& \quad 0.0021813 {x_3}^2 + 90 \leq 0, \\
& g_5(\mathbf{x}) \equiv 9.300961 + 0.0047026 x_3 x_5 + 0.0012547 x_1 x_3 + \\
& \quad 0.0019085 x_3 x_4 - 25 \leq 0, \\
& g_6(\mathbf{x}) \equiv -9.300961 - 0.0047026 x_3 x_5 - 0.0012547 x_1 x_3 - \\
& \quad 0.0019085 x_3 x_4 + 20 \leq 0, \\
& g_7(\mathbf{x}) \equiv 78 - x_1 \leq 0, \\
& g_8(\mathbf{x}) \equiv 33 - x_2 \leq 0, \\
& g_{6+i}(\mathbf{x}) \equiv 27 - x_i \leq 0, \quad (\mathrm{i} = 3, 4, 5), \\
& g_{12}(\mathbf{x}) \equiv x_1 - 102 \leq 0, \\
& g_{11+i}(\mathbf{x}) \equiv x_i - 45 \leq 0, \quad (\mathrm{i} = 2, 3, 4, 5).
\end{aligned}
\tag{25}
$$

The KKT-proximity measure is plotted in Figure 11 with a single RGA run. The KKT-proximity measure smoothly reduces, eventually converging to zero close to the 550-th generation, thereby concluding that the RGA is able to converge to the KKT point in this problem. The KKT point is $(78.0, 33.0, 29.995, 45.0, 36.775)^T$ where constraints g_1, g_6, g_7, g_9, and g_{14} are active.

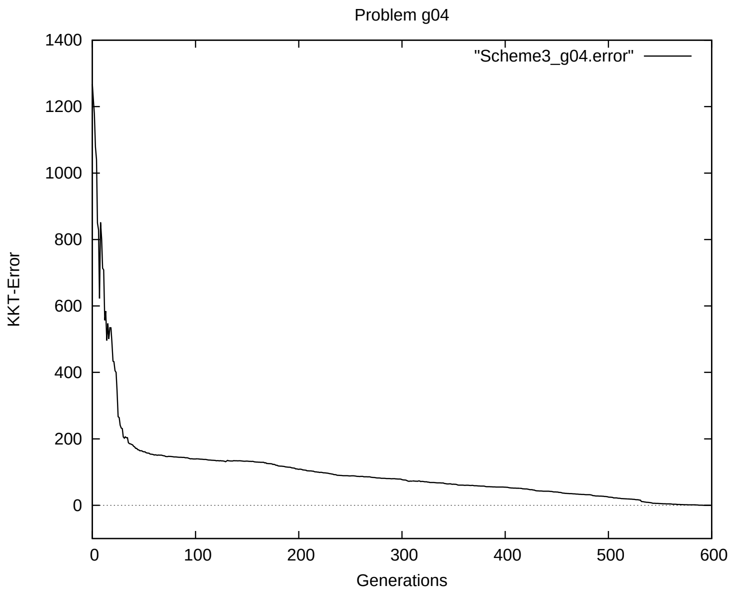


Figure 11: KKT-proximity measure for problem g04.

4.2.6 Problem *g06*

The following is the problem g06 from [10] containing two variables, and six constraints, including the variable bounds.

$$\begin{array}{rl} \text{Minimize} & f(\mathbf{x}) = (x_1 - 10)^3 + (x_2 - 20)^3, \\ \text{subject to} & g_1(\mathbf{x}) \equiv -(x_1 - 5)^2 - (x_2 - 5)^2 + 100 \leq 0, \\ & g_2(\mathbf{x}) \equiv (x_1 - 6)^2 + (x_2 - 5)^2 - 82.81 \leq 0, \\ & g_3(\mathbf{x}) \equiv 13 - x_1 \leq 0, \\ & g_4(\mathbf{x}) \equiv -x_2 \leq 0, \\ & g_5(\mathbf{x}) \equiv x_1 - 100 \leq 0, \\ & g_6(\mathbf{x}) \equiv x_2 - 100 \leq 0. \end{array} \tag{26}$$

The RGA converges to the best-known solution in this problem ($\bar{f} = -6961.8139$), where we obtain a zero KKT-proximity measure (Figure 12), confirming that the RGA-obtained solution is a KKT point. Constraints g_1 and g_2 are found to be active at the optimum.

4.2.7 Problem *g07*

The following is a 10 variable, 28 constraint problem g07 from [10].

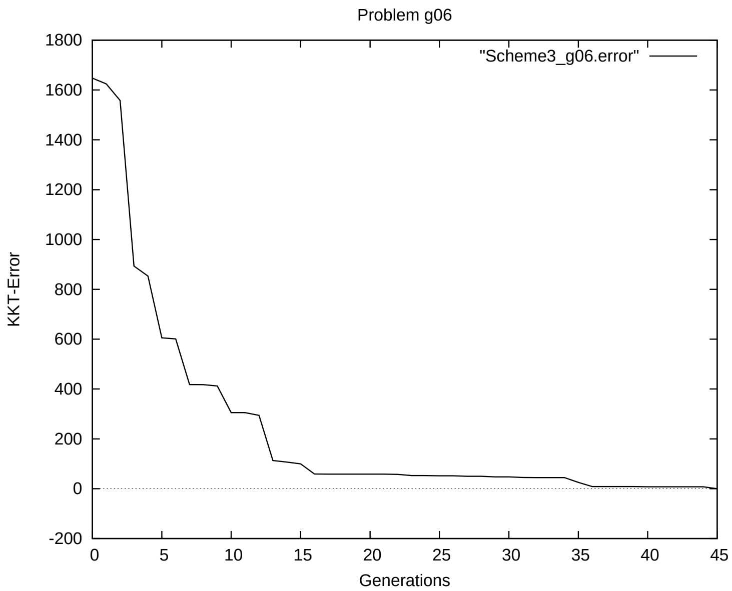


Figure 12: KKT-proximity measure for problem g06.

$$
\begin{array}{rl}
\text{Minimize} & f(\mathbf{x}) = x_1^2 + x_2^2 + x_1x_2 - 14x_1 - 16x_2 + (x_3 - 10)^2 + \\
& \quad 4(x_4 - 5)^2 + (x_5 - 3)^2 + 2(x_6 - 1)^2 + 5x_7^2 + \\
& \quad 7(x_8 - 11)^2 + 2(x_9 - 10)^2 + (x_{10} - 7)^2 + 45, \\
\text{subject to} & g_1(\mathbf{x}) \equiv -105 + 4x_1 + 5x_2 - 3x_7 + 9x_8 \leq 0, \\
& g_2(\mathbf{x}) \equiv 10x_1 - 8x_2 - 17x_7 + 2x_8 \leq 0, \\
& g_3(\mathbf{x}) \equiv -8x_1 + 2x_2 + 5x_9 - 2x_{10} - 12 \leq 0, \\
& g_4(\mathbf{x}) \equiv 3(x_1 - 2)^2 + 4(x_2 - 3)^2 + 2x_3^2 - 7x_4 - 120 \leq 0, \\
& g_5(\mathbf{x}) \equiv 5x_1^2 + 8x_2 + (x_3 - 6)^2 - 2x_4 - 40 \leq 0, \\
& g_6(\mathbf{x}) \equiv x_1^2 + 2(x_2 - 2)^2 - 2x_1x_2 + 14x_5 - 6x_6 \leq 0, \\
& g_7(\mathbf{x}) \equiv 0.5(x_1 - 8)^2 + 2(x_2 - 4)^2 + 3x_5^2 - x_6 - 30 \leq 0, \\
& g_8(\mathbf{x}) \equiv -3x_1 + 6x_2 + 12(x_9 - 8)^2 - 7x_{10} \leq 0, \\
& g_{8+i}(\mathbf{x}) \equiv -10 - x_i \leq 0, \quad (\text{i} = 1, \ldots, 10), \\
& g_{18+i}(\mathbf{x}) \equiv x_i - 10 \leq 0, \quad (\text{i} = 1, \ldots, 10).
\end{array} \tag{27}
$$

The RGA run converges within a Euclidean distance of 0.0298 from the reported minima. At the manually appended best-known solution (with $\bar{f} = 24.3062$), the KKT-proximity measure is obtained to be zero (Figure 13), thereby confirming that the best-known solution is a KKT point. Six constraints g_1, g_2, g_3, g_4, g_5, and g_6 are active at the KKT point. This problems remains to be another problem in which the specific RGA cannot locate the KKT point.

4.2.8 Problem *g09*

The following problem, g09 from [10], is a seven-variable, 18-constraint problem.

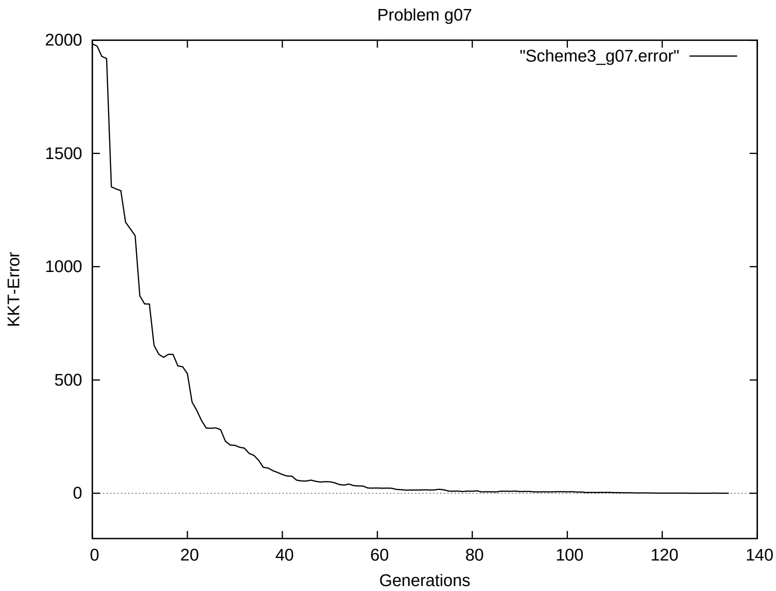


Figure 13: KKT-proximity measure for problem g07.

$$
\begin{array}{rl}
\text{Minimize} & f(\mathbf{x}) = (x_1-10)^2 + 5(x_2-12)^2 + x_3^4 + 3(x_4-11)^2 + \\
& \quad 10x_5^6 + 7x_6^2 + x_7^4 - 4x_6x_7 - 10x_6 - 8x_7, \\
\text{subject to} & g_1(\mathbf{x}) \equiv -127 + 2x_1^2 + 3x_2^4 + x_3 + 4x_4^2 + 5x_5 \le 0, \\
& g_2(\mathbf{x}) \equiv -282 + 7x_1 + 3x_2 + 10x_3^2 + x_4 - x_5 \le 0, \\
& g_3(\mathbf{x}) \equiv -196 + 23x_1 + x_2^2 + 6x_6^2 - 8x_7 \le 0, \\
& g_4(\mathbf{x}) \equiv 4x_1^2 + x_2^2 - 3x_1x_2 + 2x_3^2 + 5x_6 - 11x_7 \le 0, \\
& g_{4+i}(\mathbf{x}) \equiv -10 - x_i \le 0, \quad (\text{i} = 1, \ldots, 7), \\
& g_{11+i}(\mathbf{x}) \equiv x_i - 10 \le 0, \quad (\text{i} = 1, \ldots, 7).
\end{array} \tag{28}
$$

The RGA converges within a Euclidean distance of 0.0641 from the best-reported solution. The KKT-proximity measure reduces with iterates, as shown in Figure 14 and is zero at the best-reported solution. Two constraints g_1 and g_4 are active at the reported solution. Again, the specific RGA could not locate the KKT point for this problem.

4.2.9 Problem *g10*

The following problem, g10 from [10], is an eight-variable, 22-constraint problem.

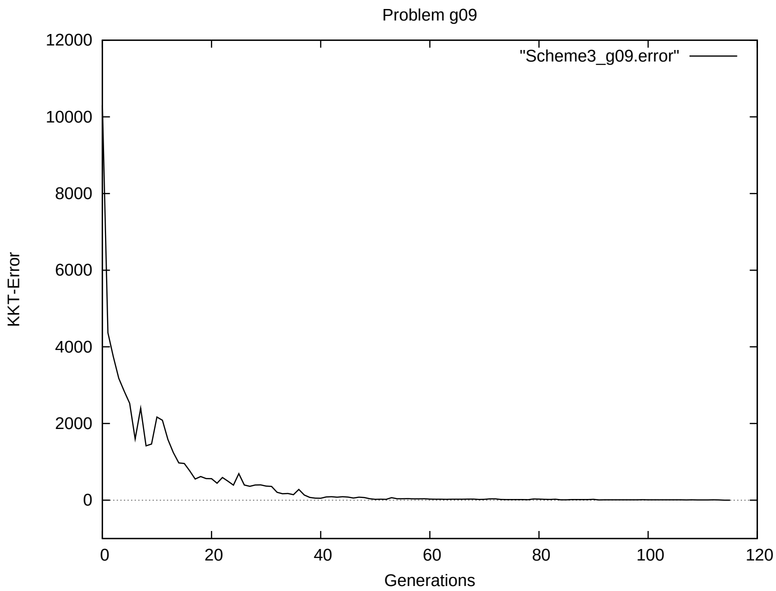


Figure 14: KKT-proximity measure for problem g09.

$$
\begin{aligned}
\text{Minimize} \quad & f(\mathbf{x}) = x_1 + x_2 + x_3, \\
\text{subject to} \quad & g_1(\mathbf{x}) \equiv -1 + 0.0025(x_4 + x_6) \leq 0, \\
& g_2(\mathbf{x}) \equiv -1 + 0.0025(x_5 + x_7 - x_4) \leq 0, \\
& g_3(\mathbf{x}) \equiv -1 + 0.01(x_8 - x_5) \leq 0, \\
& g_4(\mathbf{x}) \equiv -x_1 x_6 + 833.33252 x_4 \\
& \qquad +100 x_1 - 83333.333 \leq 0, \\
& g_5(\mathbf{x}) \equiv -x_2 x_7 + 1250 x_5 + x_2 x_4 - 1250 x_4 \leq 0, \\
& g_6(\mathbf{x}) \equiv -x_3 x_8 + 1250000 + x_3 x_5 - 2500 x_5 \leq 0, \\
& g_7(\mathbf{x}) \equiv 100 - x_1 \leq 0, \\
& g_8(\mathbf{x}) \equiv 1000 - x_2 \leq 0, \\
& g_9(\mathbf{x}) \equiv 1000 - x_3 \leq 0, \\
& g_{6+i}(\mathbf{x}) \equiv 10 - x_i \leq 0, \quad (\mathrm{i} = 4, \ldots, 8), \\
& g_{14+i}(\mathbf{x}) \equiv x_i - 10000 \leq 0, \quad (\mathrm{i} = 1, 2, 3), \\
& g_{14+i}(\mathbf{x}) \equiv x_i - 1000 \leq 0, \quad (\mathrm{i} = 4, \ldots, 8).
\end{aligned}
\tag{29}
$$

The RGA doesn't converge to the optimum in this problem as well. The best solution comes within an Euclidean distance of 1405.0 from the reported best solution for this problem. The KKT-proximity measure at the reported solution (having $\bar{f} = 7049.24802$) is found to be exactly zero (Figure 15), thereby indicating that the reported solution is truly a KKT point. All six constraints are found to be active, contrary to that in the previous study [10] which reported constraints g_1, g_2 and g_3 as active.

We have not emphasized enough here that one of the advantages of finding the KKT-proximity measure is that the process also finds the Lagrange multiplier values of all active constraints at the KKT point. As shown in Table 2, for this problem, the first three constraints have Lagrange multipliers that are about six orders of magnitude higher than that of the next three constraints. This indicates that there is a scaling issue among the constraint values, which can be somewhat alleviated by re-scaling the constraint functions by the Lagrange multiplier values. We belabor this task for another study.

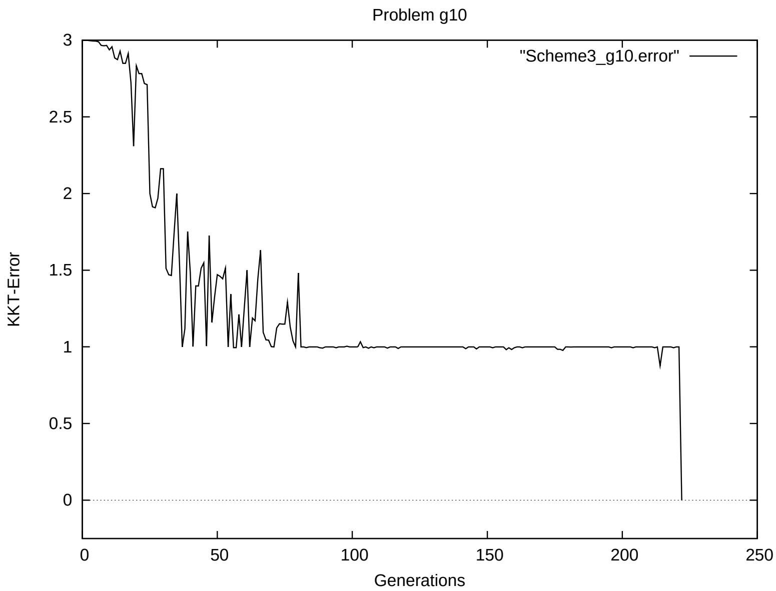


Figure 15: KKT-proximity measure for problem g10.

4.2.10 Problem *g18*

The following problem, g18 from [10], is a 9 variable, 31 constraint problem.

$$
\begin{array}{rl}
\text{Minimize} & f(\mathbf{x}) = -0.5(x_1x_4 - x_2x_3 + x_3x_9 - x_5x_9 + x_5x_8 - x_6x_7), \\
\text{subject to} & g_1(\mathbf{x}) \equiv x_3^2 + x_4^2 - 1 \le 0, \\
& g_2(\mathbf{x}) \equiv x_9^2 - 1 \le 0, \\
& g_3(\mathbf{x}) \equiv x_5^2 + x_6^2 - 1 \le 0, \\
& g_4(\mathbf{x}) \equiv x_1^2 + (x_2 - x_9)^2 - 1 \le 0, \\
& g_5(\mathbf{x}) \equiv (x_1 - x_5)^2 + (x_2 - x_6)^2 - 1 \le 0, \\
& g_6(\mathbf{x}) \equiv (x_1 - x_7)^2 + (x_2 - x_8)^2 - 1 \le 0, \\
& g_7(\mathbf{x}) \equiv (x_3 - x_5)^2 + (x_4 - x_6)^2 - 1 \le 0, \\
& g_8(\mathbf{x}) \equiv (x_3 - x_7)^2 + (x_4 - x_8)^2 - 1 \le 0, \\
& g_9(\mathbf{x}) \equiv x_7^2 + (x_8 - x_9)^2 - 1 \le 0, \\
& g_{10}(\mathbf{x}) \equiv x_2x_3 - x_1x_4 \le 0, \\
& g_{11}(\mathbf{x}) \equiv -x_3x_9 \le 0, \\
& g_{12}(\mathbf{x}) \equiv x_5x_9 \le 0, \\
& g_{13}(\mathbf{x}) \equiv x_6x_7 - x_5x_8 \le 0, \\
& g_{13+i}(\mathbf{x}) \equiv -10 - x_i \le 0 \quad (\mathrm{i} = 1, \ldots, 8), \\
& g_{22+i}(\mathbf{x}) \equiv x_i - 10 \le 0 \quad (\mathrm{i} = 1, \ldots, 8), \\
& g_{22}(\mathbf{x}) \equiv -x_9 \le 0, \\
& g_{31}(\mathbf{x}) \equiv x_9 - 20 \le 0,
\end{array}
\tag{30}
$$

The RGA run converges to the KKT point for this problem, as the KKT-proximity measure plotted in Figure 16, shows that it reduces to zero after certain generations. After initial fluctuations, the KKT-proximity measure converges to zero. At this solution, the constraints, g_1, g_3, g_4, g_6, g_7 and g_9 are active.

A summary of the problems, the corresponding KKT points, and the associated Lagrange multipliers of the active constraints are tabulated in Table 2 in the appendix. Having shown the working of our proposed KKT-proximity measure on standard smooth problems, we now consider

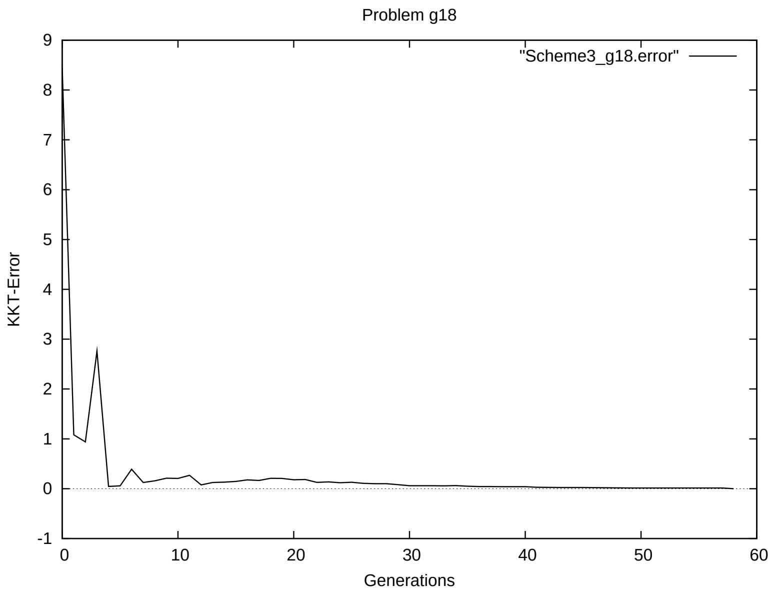


Figure 16: KKT-proximity measure for problem g18.

one non-smooth problem.

4.3 Numerical Results: Non-smooth Problems

In this subsection, we consider a modified Freudenstein-Roth function [7].

$$\begin{aligned} \text{Minimize} \quad & f(\mathbf{x}) = |c_1(\mathbf{x})| + |c_2(\mathbf{x})|, \\ \text{subject to} \quad & g_1(\mathbf{x}) \equiv (9 - x_1)^2 + (-0.5 - x_2)^2 - 6.516 \leq 0, \\ \text{where} \quad & c_1(\mathbf{x}) = x_1 - x_2^3 + 5x_2^2 - 2x_2 - 13, \\ & c_2(\mathbf{x}) = x_1 + x_2^3 + x_2^2 - 14x_2 - 29. \end{aligned} \tag{31}$$

We take the iterates along $c_1(\mathbf{x}) = 0$, where the objective function at every point is non-differentiable (due to the modulus function). The KKT-proximity measure is computed by solving the following optimization problem in which $(\epsilon_k, \hat{\mathbf{x}}_k, \rho, \mathbf{u})$ is the variable vector for the iterate $\mathbf{x}_k$ at every generation and $\hat{\mathbf{x}}_k \in \Re^2$, $\rho \in \partial^o f(\hat{\mathbf{x}}^k)$:

$$\begin{aligned} \text{Minimize} \quad & \epsilon_k \\ \text{Subject to} \quad & \left\|\rho + \textstyle\sum_{i=1}^m u_i \nabla g_i(\hat{\mathbf{x}}_k)\right\| \leq \sqrt{\epsilon_k}, \\ & \textstyle\sum_{i=1}^m u_i g_i(\mathbf{x}_k) \geq -\epsilon_k, \\ & \|\hat{\mathbf{x}}_k - \mathbf{x}_k\| \leq \sqrt{\epsilon_k}, \\ & u_i \geq 0, \quad \forall\, i, \\ & \rho \in \partial^o f(\hat{\mathbf{x}}_k). \end{aligned} \tag{32}$$

Figure 17 shows the contour plot of the modified Freudenstein-Roth function. The iterates $\mathbf{x}_k$ are placed on $c_1(\mathbf{x}) = 0$ curve, as shown in the figure. The corresponding $\hat{\mathbf{x}}_k$ points are also marked in the figure. Iterate A corresponds to the the approximate point A' shown in the figure. As the iterates approach the minimum point (point O), the approximate points $(\hat{\mathbf{x}}_k)$ approach the minimum point as well and when the minimum point is chosen as the iterate, the approximate point is identical to the minimum point.

As iterates go past the optimum point O, the approximate points are identical to the iterates. Figure 18 shows the KKT-proximity measure (ϵ_k) from point A to B via O. It is clear that when

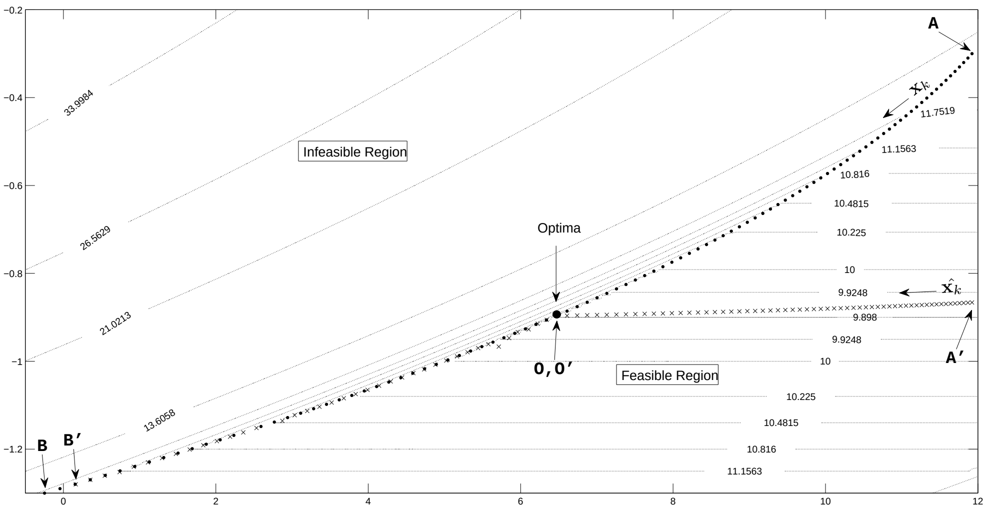


Figure 17: Freudenstein-Roth Function

iterates approach the minimum point O, the KKT-proximity measure reduces to zero and as it crosses past the minimum point, the KKT-proximity measure increases.

5 Commercial Optimization Algorithms and Approximate KKT Error

KKT conditions are used as stopping criterion in commercially available softwares like MatLab [12] and Knitro [15]. Knitro, a widely used package integrates two powerful and complementary algorithmic approaches for non-linear optimization viz. the interior-point approach and the active-set approach, employing conjugate gradient iteration in the step computation. It uses the first-order KKT conditions to identify a locally optimal solution, and therefore as a terminating criterion [3].

For the problem (P), Knitro defines the *feasibility error* (`FeasErr`) at a point $\mathbf{x}_k$ as the maximum violation of the constraints:

$$\texttt{FeasErr} \;=\; \max_{i=1,\ldots,m} \big\{0,\; g_i(\mathbf{x}_k)\big\},$$

and the *optimality error* (`OptErr`) as,

$$\texttt{OptErr} \;=\; \max\Big\{\; \Big\|\nabla f(\mathbf{x}_k) + \sum_{j=1}^{m} u_j \nabla g_j(\mathbf{x}_k)\Big\|_\infty,\;\; -\max_{i=1,\ldots,m} u_i g_i(\mathbf{x}_k)\;\Big\},$$

with $u_i \geq 0$ enforced explicitly throughout the optimization. Lagrange multipliers u_is are computed by solving equilibrium equation alone in the least-square sense at every iterate $\mathbf{x}_k$. Furthermore, `FeasErr` and `OptErr` are scaled using

$$\begin{aligned}\tau_1 &= \max_{i=1,\ldots,m} \big\{1,\; g_i(\mathbf{x}_0)\big\},\\ \tau_2 &= \max\big\{1,\; \|\nabla f(\mathbf{x}_k)\|_\infty\big\},\end{aligned}$$

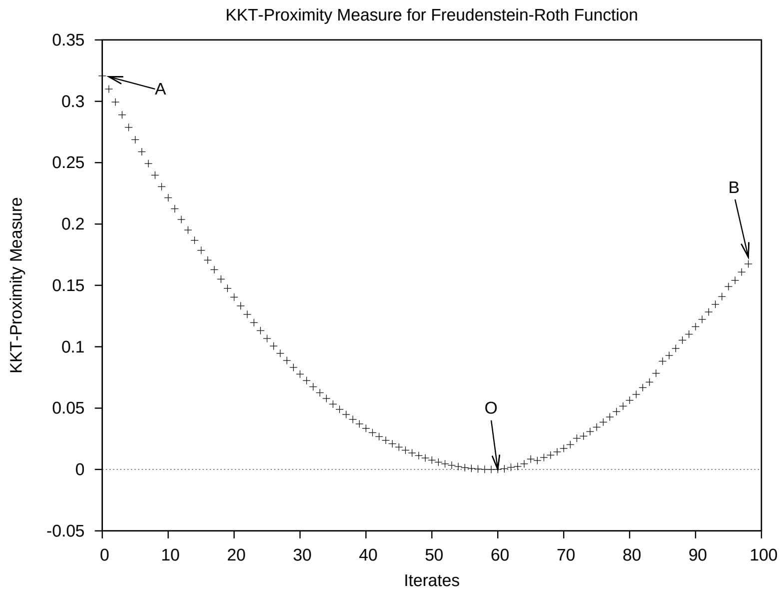


Figure 18: KKT-proximity measure for Freudenstein-Roth function.

where $\mathbf{x}_0$ represents the initial point. Based on constants defined by the user options `feastol` and `opttol`, Knitro declares a locally optimal solution if and only if the following conditions are true:

$$\begin{aligned} \texttt{FeasErr} &\leq \tau_1 * \texttt{feastol}, \\ \texttt{OptErr} &\leq \tau_2 * \texttt{opttol}. \end{aligned}$$

However, it is interesting to note that at no point in time in the algorithm, the Lagrange multipliers $\mathbf{u}$ are compute to estimate `OptErr`. These Lagrange multipliers are obtained by additional computation of the approximate solution to the quadratic model which is performed in every iteration of a Sequential Quadratic Programming (SQP) method with trust-region approach [2].

Matlab's fmincon() routine uses a similar approach for termination, Additionally, it checks on relative change in $\mathbf{x}$-vector, function value and constraint values between two consecutive iterations for termination. For more information on the first-order KKT optimality conditions, please see [13].

Our methodology is different from the approaches adopted in these commercial softwares. We form and solve an optimization problem balancing the violations in equilibrium equation and complementary slackness conditions to find the Lagrange multiplier vector at every iterate. The process provides us with a proximity measure which can be utilized as a termination condition of an optimization algorithm. Moreover, our proposed theorems support the concept that if the proximity measure reduces to zero for a sequence of iterates, with the satisfaction of certain constraint qualification conditions, the limiting iterate is a KKT point. Our approach is more direct and the simulation results on smooth as well as non-smooth problems suggest the simplicity and efficacy of the proposed methods.

5.1 Comparison of KKT-Error Between Knitro and Proposed Procedure

As discussed above, the Knitro software package computes the Lagrange multipliers in every iteration of its optimization routines, and computes a KKT-proximity measure of its own, which it uses as the terminating condition. Please note that like our method, the Lagrange multipliers in Knitro are a by-product of its internal routines and no explicit mechanism is provided for computing them at any arbitrary iterate of an optimization problem. Nonetheless, we conduct a comparative study on the problem given below.

$$\begin{aligned} \text{Minimize:} \quad f(\mathbf{x}) &= 100(x_2 - x_1^2)^2 + (1 - x_1)^2, \\ \text{Subject To:} \quad g_1(\mathbf{x}) &= 1 - x_1 x_2 \leq 0, \\ g_2(\mathbf{x}) &= -x_1 - x_2^2 \leq 0, \\ g_3(\mathbf{x}) &= x_1 - 0.5 \leq 0. \end{aligned}$$

First, we apply Knitro to the above problem with a starting point $\mathbf{x}_0 = (0.35, 10000)^T$. We record the intermediate iterates and corresponding `OptErr` values calculated by Knitro. These values are shown in Table 1 and are plotted in Figure 19.. We observe that `OptErr` reduces with iteration. Next, we use the same set of iterates and apply our approach to compute KKT-proximity measure. The table tabulates the values and the figure also plots these values. Although our KKT-proximity measure values are of different orders of magnitude, the measure reduces to zero with iteration smoothly. This confirms the accuracy of our approach vis-a-vis Knitro software.

Table 1: Comparison of Knitro's OptErr measure and the proposed KKT-proximity measure values. For the first iterates, the `fmincon` procedure used to solve the optimization problem did not converge after several minutes of simulation.

No.	$\mathbf{x}_k$			Our Scheme	KNitro
k	x_1	x_2	Feasible	Prox. Measure	`OptErr`
1	0.350000	10000.0	Yes	-	-
2	0.392786	5025.00	Yes	-	37.750
3	0.487528	2525.060	Yes	-	12.920
4	0.494462	1268.840	Yes	-	3.6280
5	0.498565	637.59400	Yes	3.8655e+07	0.936300
6	0.499620	320.39000	Yes	1.1828e+07	0.925500
7	0.499902	160.99500	Yes	2.4617e+06	0.961900
8	0.499975	80.898700	Yes	6.2062e+05	0.961900
9	0.499994	40.648500	Yes	1.5623e+05	0.986600
10	0.499998	20.419800	Yes	3.9212e+04	0.986600
11	0.499999	10.248700	Yes	9.7868e+03	0.517800
12	0.499997	5.125720	Yes	2.4183e+03	0.129300
13	0.499984	2.527150	Yes	232.0659	0.032160
14	0.483281	2.086940	Yes	38.7791	0.019330
15	0.499916	2.018170	Yes	6.4107	0.001846
16	0.500000	2.000080	Yes	0.0280	0.0000393
17	0.500000	2.000000	Yes	0.0000	0.0000

6 Conclusions

This work is one of the few studies aimed at exploiting KKT conditions in optimization algorithm design. It has been observed that the extent of violation of KKT conditions in the vicinity of the

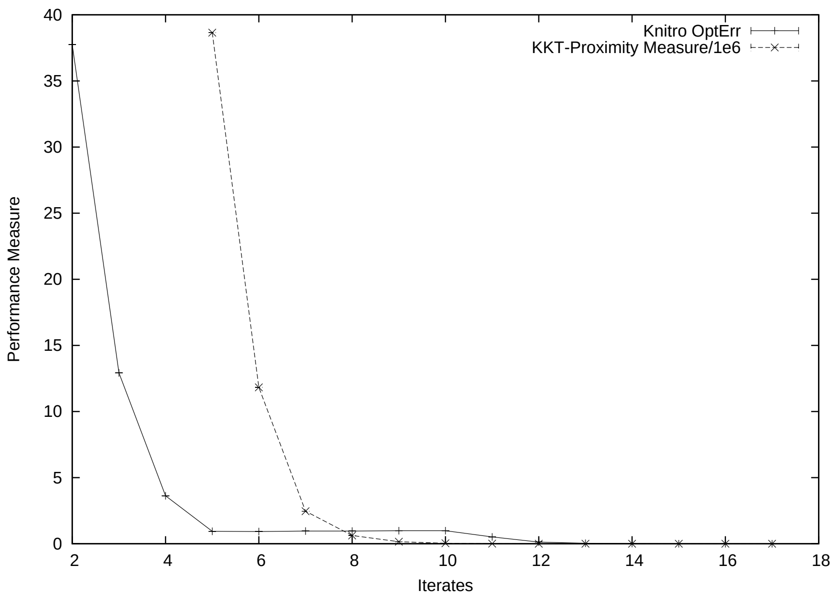


Figure 19: Proximity measure and Knitro's `OptErr` values are shown for a set of identical iterates.

KKT point is not smooth and hence a naive KKT-error measure is not suitable as a termination condition for an optimization algorithm. Based on the concept of an ϵ-KKT point, we have relaxed the complimentary slackness and equilibrium KKT conditions and proven theorems to guide as suggest a KKT-proximity measure that smoothly reduces to zero, as the iterates approach to the KKT point. We have considered both smooth and non-smooth problems for this purpose.

In addition to the theoretical developments on ϵ-KKT points, we have also done extensive simulations on a number of standard constrained test problems (smooth and non-smooth) to demonstrate the working of our procedures.

We have also compared our one-step optimization procedure of computing the KKT-proximity measure with a commercial software's (Knitro) internal `OptErr` on an identical set of iterates. The trend in reduction of both measures have been found to be similar.

The results presented in this study indicate that the proposed KKT-proximity measure can be effectively used as a termination condition in an optimization algorithm. For evolutionary optimization algorithms (EAs) which do not have a convergence proof, the evidence of its best generation-wise iterates approaching the KKT point on a number of optimization problems, as demonstrated in this paper, makes EAs worthy of more attention.

Further work should be aimed at exploiting the KKT-proximity measure in heuristic algorithms and integration of the same in designing better optimization algorithms.

Acknowledgments

The study is funded by Department of Science and Technology, Government of India under SERC-Engineering Sciences scheme (No. SR/S3/MERC/091/2009).

References

[1] R. Andreani, G. Haeser, and J. M. Martinez. On sequential optimality conditions for smooth constrained optimization. *Optimization Online*, 2009.

[2] R. H. Byrd, M. E. Hribar, and J. Nocedal. An interior point algorithm for large scale nonlinear programming. *SIAM Journal on Optimization*, 9:877–900, 1997.

[3] R. H. Byrd, J. Nocedal, and R. A. Waltz. *KNITRO: An integrated package for nonlinear optimization*, pages 35–59. Springer-Verlag, 2006.

[4] F. H. Clarke. *Optimization and nonsmooth anslysis.* Wiley-Interscience, 1983.

[5] K. Deb. An efficient constraint handling method for genetic algorithms. *Computer Methods in Applied Mechanics and Engineering*, 186(2–4):311–338, 2000.

[6] K. Deb, R. Tiwari, M. Dixit, and J. Dutta. Finding trade-off solutions close to kkt points using evolutionary multi-objective optimization. In *Proceedings of the Congress on Evolutionary Computation (CEC-2007)*, pages 2109–2116, 2007.

[7] R. Fletcher and C. M. Reeves. Function minimization by conjugate gradients. *Computer Journal*, 7:149–154, 1964.

[8] A. Hamel. An ε-lagrange multiplier rule for a mathematical programming problem on banach spaces. *Optimization*, 49:137–149, 2001.

[9] W. Hock and K. Schittkowski. *Test Examples for Nonlinear Programming Codes.* Springer-Verlag, New York, 1981.

[10] J. J. Liang, T. P. Runarsson, E. Mezura-Montes, M. Clerc, P. N. Suganthan, C. A. Coello Coello, and K. Deb. Special session on constrained real-parameter optimization (http://www.ntu.edu.sg/home/epnsugan/), 2006.

[11] K. Miettinen. *Nonlinear Multiobjective Optimization.* Kluwer, Boston, 1999.

[12] C. Moler. The mathworks: MATLAB requirements.

[13] J. Nocedal and S. J. Wright. *Numerical Optimization.* Springer Series in Operations Research, 2nd edition edition, 2006.

[14] R. T. Rockafellar. *Convex Analysis.* Princeton University Press, 1996.

[15] R. A. Watlz and T. D. Plantenga. *Knitro's User Manual Version 6.0.* Ziena Optimization, 2009.

A Obtained Lagrange Multipliers for Test Problems

Here, we tabulate the Lagrange multipliers obtained as a by-product of solving the optimization problem for finding the KKT-proximity measure for the KKT point. Table 2 tabulate the KKT solution and Lagrange multipliers for each problem. Lagrange multipliers for active constraints are shown.

Table 2: Test problems, their best-reported solutions, and obtained Lagrange multiplier values at the best-reported solutions.

No.	Problem	Optima	Lagrange Multipliers
1	g01	[1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 3.0 3.0 3.0 1.0]	$u_1 = 0.49986960$, $u_2 = 0.49986960$, $u_3 = 0.49995217$, $u_7 = 0.00062179$, $u_8 = 0.00036150$, $u_9 = 0.00036151$, $u_{23} = 3.00034711$, $u_{24} = 3.00025284$, $u_{25} = 3.00025284$, $u_{26} = 5.00107578$, $u_{27} = 1.00055454$, $u_{28} = 1.00063755$, $u_{29} = 1.00031322$, $u_{30} = 1.00063756$, $u_{31} = 1.00031323$, $u_{35} = 0.99998890$
2	hs23	[1.0 1.0]	$u_4 = 2.00046030$, $u_5 = 2.00060355$
3	hs45	[1.0 2.0 3.0 4.0 5.0]	$u_6 = 0.99951173$, $u_7 = 0.49975586$, $u_8 = 0.33317057$, $u_9 = 0.24987793$, $u_10 = 0.19990234$
4	g02	[3.16246061 3.128331428 3.09479212 3.061450595 3.0279292, 2.99382607, 2.95866871, 2.92184227, 0.494825115, 0.48835711, 0.4823164, 0.47664476, 0.4712955, 0.46623099, 0.46142, 0.4568366, 0.4524588, 0.448267622, 0.444247, 0.44038285]	$u_1 = 0.04689694$
5	g04	[78.0 33.0 29.995256 45.0 36.775813]	$u_1 = 403.27022000$, $u_6 = 809.42360424$, $u_7 = 48.92768703$, $u_8 = 84.32381214$, $u_{15} = 26.63967450$
6	g06	[14.095 0.842960789]	$u_1 = 1097.11096525$, $u_2 = 1229.53332532$, $u_4 = 0.00006220$
7	g07	[2.171996 2.363683 8.773926 5.095984 0.990655 1.430574 1.321644 9.828726 8.280092 8.375927]	$u_1 = 1.71648886$, $u_2 = 0.47450184$, $u_3 = 1.37590841$, $u_4 = 0.02054950$, $u_5 = 0.31202935$, $u_6 = 0.28707154$
8	g09	[2.3305084590 1.9513700830 -0.477418650 4.3657261380 -0.624486980 1.0381346830 1.5942188960]	$u_1 = 1.13972466$, $u_4 = 0.36850490$
9	g10	[579.306685 1359.970678 5109.970657 182.017699 295.601173 217.982300 286.416525 395.601173]	$u_1 = 1966.52920083$, $u_2 = 5217.30343e838$, $u_3 = 5116.48814974$, $u_4 = 0.00848649$, $u_5 = 0.00959083$, $u_6 = 0.01001275$
10	g18	[-0.657776192 -0.153418773 0.323413872 -0.946257612 -0.657776192 -0.753213435 0.323413874 -0.346462948 0.599794663]	$u_1 = 0.14409510$, $u_3 = 0.14409508$, $u_4 = 0.14462060$, $u_6 = 0.14425899$, $u_7 = 0.14445993$, $u_9 = 0.14408119$