# Approximate KKT Points for Iterates of an Optimizer

Joydeep Dutta, Kalyanmoy Deb, Ramnik Arora, Rupesh Tulshyan
Kanpur Genetic Algorithms Laboratory (KanGAL)
Indian Institute of Technology Kanpur
PIN 208016, India

**KanGAL Report No. 2009009**

December, 2009

**Abstract**

In this technical note, we suggest a new definition for an approximate KKT point. The concept of approximate KKT point can then be used on iterates (points found by an optimization algorithm) to check whether the iterates lead to a KKT point.

We will concentrate on the following simple optimization problem (P):

$$\begin{array}{lll} \text{Min} & f(\mathbf{x}) & \\ \text{subject to} & g_i(\mathbf{x}) \leq 0, & i = 1, 2, \ldots, m. \end{array} \tag{1}$$

We will assume that $f$ and each $g_i$ $(i = 1, 2, \ldots, m)$ are smooth functions (see, for example, [1] for more details on problem (P)). Our main aim in this section is to define certain notions of approximate KKT points and show that if a sequence of such points converge to a point where the constraints satisfy some constraint qualification then such a point is a KKT point.

We shall introduce two such notions of approximate KKT points and our first one is given below.

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

1. $\|\nabla f(\bar{\mathbf{x}}) + \sum_{i=1}^{m} \lambda_i \nabla g_i(\bar{\mathbf{x}})\| \leq \epsilon$,
2. $\lambda_i g_i(\bar{\mathbf{x}}) = 0$, *for* $i = 1, 2, \ldots, m$

In order to state our main result we need the notion of Mangasarian-Fromovitz constraint qualification (MFCQ) [4] which is given as follows: The constraints of the problem (P) is said to satisfy MFCQ at $\bar{x}$ which is feasible if there exists a vector $d \in \mathbf{R}^n$ such that $\langle \nabla g_i(\bar{\mathbf{x}}), d \rangle < 0$, $\forall i \in I(\bar{\mathbf{x}})$, where $I(\bar{\mathbf{x}}) = \{i : g_i(\bar{\mathbf{x}}) = 0\}$ is the set of active constraints.

The MFCQ can be alternatively stated in the following equivalent form which can be deduced using separation theorem for convex sets.

The constraints of (P) satisfies MFCQ at a feasible $\bar{\mathbf{x}}$ if there exists no vector $\lambda \neq 0$, $\lambda \in \mathbf{R}^m_+$, with $\lambda_i \geq 0$, $i \in I(\bar{\mathbf{x}})$ and $\lambda_i = 0$ for $i \neq I(\bar{\mathbf{x}})$ with

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

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

**Theorem 1** *Let $\{\mathbf{x}^\nu\}$ be a sequence of feasible points of (P) such that $\mathbf{x}^\nu \to \bar{\mathbf{x}}$ as $\nu \to \infty$. Let $\{\epsilon_\nu\}$ be a sequence of positive real numbers such that $\epsilon_\nu \downarrow 0$, as $\nu \to \infty$. Further assume that for each $\nu$, $\mathbf{x}^\nu$ is an $\epsilon_\nu$-KKT point of (P). If MFCQ holds at $\bar{\mathbf{x}}$, then $\bar{\mathbf{x}}$ is a KKT point.*

**Proof** Since $\mathbf{x}^\nu$ is an $\epsilon_\nu$-KKT point for (P), it is clear from the definition that $\mathbf{x}^\nu \in C$ for each $\nu$ and as each $g_i$ is continuous and $\{\mathbf{x}^\nu\} \to \bar{\mathbf{x}}$ it is clear that $\bar{\mathbf{x}}$ is a feasible point for (P). Now from the definition of $\epsilon_\nu$-KKT points we have for each $\nu$, there exists a vector $\lambda^\nu \in \mathbf{R}^m_+$ such that

1. $\|\nabla f(\mathbf{x}^\nu) + \sum_{i=1}^m \lambda_i^\nu \nabla g_i(\mathbf{x}^\nu)\| \leq \epsilon_\nu$,
2. $\lambda_i^\nu g_i(\bar{\mathbf{x}^\nu}) = 0$, for $i = 1, 2, \dots, m$.

Our claim is that the sequence $\lambda^\nu$ is bounded. On the contrary assume that $\lambda^\nu$ is not bounded. Hence, $\|\lambda^\nu\| \to \infty$ as $\nu \to \infty$. Now consider the sequence $\{\mathbf{w}^\nu\}$, with

$$\mathbf{w}^\nu = \frac{\lambda^\nu}{\|\lambda^\nu\|}, \text{ for all } \nu.$$

It is clear that $\mathbf{w}^\nu$ is bounded and hence without loss of generality we can conclude that $\mathbf{w}^\lambda \to \bar{\mathbf{w}}$ and $\|\mathbf{w}\| = 1$. Now observe that item (i) can be written as

$$\|\nabla f(\mathbf{x}^\nu) + \nabla g(\mathbf{x}^\nu)^T \lambda^\nu\| \leq \epsilon_\nu, \tag{2}$$

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

Now by dividing both sides of eqn. 2 by $\|\lambda^\nu\|$ we have

$$\left\| \frac{1}{\|\lambda^\nu\|} \nabla f(\mathbf{x}^\nu) + \nabla g(\mathbf{x}^\nu)^T \frac{\lambda^\nu}{\|\lambda^\nu\|} \right\| \leq \frac{1}{\|\lambda^\nu\|} \epsilon_\nu$$

That is,

$$\left\| \frac{1}{\|\lambda^\nu\|} \nabla f(\mathbf{x}^\nu) + \nabla g(\mathbf{x}^\nu)^T \mathbf{w}^\nu \right\| \leq \frac{1}{\|\lambda^\nu\|} \epsilon_\nu. \tag{3}$$

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

$$\frac{1}{\|\lambda^\nu\|} \nabla f(\mathbf{x}^\nu) \to 0 \text{ as } \nu \to \infty$$

and

$$\frac{1}{\|\lambda^\nu\|} \epsilon_\nu \to 0 \text{ as } \nu \to \infty$$

Thus, passing to the limit in eqn. 3 as $\nu \to 0$, we have $\|\nabla g(\bar{\mathbf{x}})^T \bar{\mathbf{w}}\| \leq 0$ (note that since $g$ is smooth $\nabla g(\mathbf{x}^\nu) \to \nabla g(\bar{\mathbf{x}})$). That is, $\sum_{i=1}^m \bar{w}_i \nabla g_i(\bar{\mathbf{x}}) = 0$, where $\bar{\mathbf{w}} = (\bar{w}_1, \bar{w}_2, \dots, \bar{w}_m)$. Since $\|\mathbf{w}\| = 1$, it is clear that MFCQ is violated at $\mathbf{x} = \bar{\mathbf{x}}$. This is a contradiction. Hence, the sequence $\{\lambda^\nu\}$ is indeed bounded. Thus, we can assume without loss of generality that $\lambda^\nu \to \bar{\lambda} \in \mathbf{R}^m_+$ (since $\mathbf{R}^m_+$ is a closed set). Hence as $\nu \to \infty$ from items (1) and (2), we have

$$\|\nabla f(\bar{\mathbf{x}}) + \sum_{i=1}^m \bar{\lambda}_i \nabla g_i(\bar{\mathbf{x}})\| = 0,$$

and $\lambda_i g_i(\bar{\mathbf{x}}) = 0$, for $i = 1, 2, \dots, m$. Thus, we have

1. $\nabla f(\bar{\mathbf{x}}) + \sum_{i=1}^m \bar{\lambda}_i \nabla g_i(\bar{\mathbf{x}}) = 0$ and

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

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

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

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

is compact then $\{\mathbf{x}^\nu\}$ 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,

$$C = \{(x, y) \in \mathbf{R}^2 : x^2 + y^2 \leq 1, x + y \leq 1\}.$$

It is simple to observe that $C$ is compact.

Our next notion of an approximate KKT point will be called modified $\epsilon$-KKT point which we will introduce in the more general context of a non-smooth problem (P). In fact we will now assume that (P) is a locally Lipschitz problem in the sense that both $f$ and $g_i$'s are locally Lipschitz function.

We now consider non-differentiable problems. In the event of non-differentiable objective functions and constraints, the concept of Clarke subdifferentials [2, 3] can be utilized. To define the Clarke subdifferential, we first define Clarke directional derivative at $\mathbf{x} \in R^n$ and in the direction $\mathbf{v} \in R^n$ for a locally Lipschitz function $f$ as follows:

$$f^o(\mathbf{x}, \mathbf{v}) = \lim_{\mathbf{y} \to \mathbf{x}} \sup_{t \to 0+} \frac{f(\mathbf{y} + t\mathbf{v}) - f(\mathbf{y})}{t}, \tag{4}$$

where $\|\mathbf{v}\| = 1$, $\mathbf{y} \in R^n$ and $t > 0$. For locally Lipschitz function, the right-hand side is bounded and the limit is finite. The Clarke subdifferential is then defined as follows:

$$\partial^o f(\mathbf{x}) = \{\zeta \in R^n : f^o(\mathbf{x}, \mathbf{v}) \geq \langle \zeta, \mathbf{v} \rangle, \forall \mathbf{v} \in R^n\}. \tag{5}$$

In other words, the Clarke subdifferential is a set of all vectors whose component along any direction $\mathbf{v}$ is smaller than or equal to the Clarke directional derivative defined above. For locally Lipschitz functions, an important result is that the Clarke subdifferential is a convex and compact set made up with limiting derivatives $\lim_{i \to \infty} \nabla f(\mathbf{x}^{(i)})$ at neighboring points $\mathbf{x}^{(i)} \to \mathbf{x}$. For example, the function $f(x) = |x|$ is not differentiable at $x = 0$. However, the limiting derivatives for neighboring solutions $x^{(i)} > 0$ is 1 and for $x^{(i)} < 0$ is $-1$. Thus, any real-value in $[-1, 1]$ is a Clarke subdifferential of $f(x)$ at $x = 0$. It is interesting to note that Clarke subdifferential contains the $\nabla f(\mathbf{x})$ at a point $\mathbf{x}$ if the function is continuously differentiable.

**Definition 2** *A point $\bar{\mathbf{x}}$ which is feasible for (P) is said to be a modified $\epsilon$-KKT point for a given $\epsilon > 0$ if there exists $\hat{\mathbf{x}} \in \mathbf{R}^n$ such that $\|\bar{\mathbf{x}} - \hat{\mathbf{x}}\| \leq \sqrt{\epsilon}$ 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 $\lambda_i \geq 0$ for $i = 1, \ldots, m$ such that*

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

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 point $\bar{\mathbf{x}}$.

The above definition is motivated from the famous Ekeland's variational principle (EVP) which we now state below.

**Theorem 2** *(Ekeland's Variational Principle) Let $f : \mathbf{R}^n \to \mathbf{R}$ be a lower-semicontinuous function which is bounded below. Let $\epsilon > 0$ be given and let $\bar{\mathbf{x}} \in \mathbf{R}^n$ is such a point for which we have*

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

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

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

$$\min_{\mathbf{x} \in \mathbf{R}^n} f(\mathbf{x}) + \frac{\epsilon}{\lambda} \|\mathbf{x} - \hat{\mathbf{x}}\|$$

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

Another definition of a subdifferential is given below: [5].

**Definition 3** *Let $f : \mathbf{R}^n \to \mathbf{R}$ be a convex function. Then the subdifferential of $f$ at an $x \in \mathbf{R}^n$ is given as*

$$\partial f(\bar{\mathbf{x}}) = \{v \in \mathbf{R}^n : f(\mathbf{x}) \geq f(\bar{\mathbf{x}}) + < v, \mathbf{x} - \bar{\mathbf{x}} >, for\ all\ x \in \mathbf{R}^n\}$$

*It is important to note that $\partial f(\bar{\mathbf{x}})$ is non-empty, convex and compact set for any $\bar{\mathbf{x}} \in \mathbf{R}^n$. If $f$ is differentiable at $\bar{\mathbf{x}}$ then $\partial f(\bar{\mathbf{x}}) = \{\nabla f(\bar{\mathbf{x}})\}$. If $\bar{\mathbf{x}}$ is a global minimum of $f$ then $\mathbf{0} \in \partial f(\bar{\mathbf{x}})$, then $\bar{\mathbf{x}}$ is a global minimum of $f$. When $f$ is convex then $\partial f(\bar{\mathbf{x}}) = \partial^o f(\bar{\mathbf{x}})$, for all $\bar{\mathbf{x}} \in \mathbf{R}^n$.*

**Theorem 3** *Let us consider the problem (P) where $f$ and each $g_i$, $i = 1, , \ldots, m$ is a convex function. Let $\bar{\mathbf{x}}$ be a feasible point which is an $\epsilon$-minimum of (P). That is,*

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

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

**Proof** Since $\bar{\mathbf{x}}$ is an $\epsilon$-minimum of the convex problem it is clear that there is no $\mathbf{x} \in \mathbf{R}^n$ which satisfies the system

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

Now using standard separation arguments (or the Gordon's theorem of the alternative) we conclude that there exists a vector $\mathbf{0}$ $(= (\lambda_0, \boldsymbol{\lambda}) \in \mathbf{R}_+ \times \mathbf{R}_+^m$ such that for all $\mathbf{x} \in \mathbf{R}^n$

$$\lambda_0(f(\mathbf{x}) - f(\bar{\mathbf{x}})) + \lambda_0 \epsilon + \sum_{i=1}^{n} \lambda_i g_i(\mathbf{x}) \geq 0. \tag{6}$$

We claim that $\lambda_0 = 0$ and hence from eqn. 6 we have

$$\sum_{i=1}^{m} \lambda_i g_i(\mathbf{x}) \geq 0, \quad \forall \mathbf{x} \in \mathbf{R}^n.$$

However, it is clear that $\sum_{i=1}^{m} \lambda_i g_i(\mathbf{x}^*) < 0$. Hence, $\lambda_0 > 0$ and one can take $\lambda_0 = 1$ without loss of generality. This shows that

$$(f(\mathbf{x}) - f(\bar{\mathbf{x}})) + \epsilon + \sum_{i=1}^{m} \lambda_i g_i(\mathbf{x}) \geq 0, \quad \forall \mathbf{x} \in \mathbf{R}^n. \tag{7}$$

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

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

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

$$L(\mathbf{x}, \boldsymbol{\lambda}) = f(\mathbf{x}) + \sum_{i=1}^{m} \lambda_i g_i(\mathbf{x}),$$

we have from eqn. 7

$$L(\mathbf{x}, \boldsymbol{\lambda}) \geq L(\bar{\mathbf{x}}, \boldsymbol{\lambda}) - \epsilon, \quad \forall \mathbf{x} \in \mathbf{R}^n. \tag{8}$$

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

$$\min_{\mathbf{x} \in \mathbf{R}^n} \; L(\mathbf{x}, \boldsymbol{\lambda}) + \sqrt{\epsilon}\|\mathbf{x} - \hat{\mathbf{x}}\|.$$

Thus from standard rules of convex analysis [5], we have

$$\mathbf{0} \in \partial L(\hat{\mathbf{x}}, \lambda) + \sqrt{\epsilon}\mathbf{B}_{\mathbf{R}^n},$$

where $\mathbf{B}_{\mathbf{R}^n}$ denotes the unit ball in $\mathbf{R}^n$. Hence, using again usual rules of convex analysis, we have

$$\mathbf{0} \in \partial f(\hat{\mathbf{x}}) + \sum_{i=1}^{m} \lambda_i \partial g_i(\hat{\mathbf{x}}) + \sqrt{\epsilon}\mathbf{B}.$$

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

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

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

If we consider $\hat{\mathbf{x}} = \bar{\mathbf{x}}$ in the above definition, any positive $\epsilon$ value will satisfy $\|\bar{\mathbf{x}} - \hat{\mathbf{x}}\| \leq \sqrt{\epsilon}$. Since $\boldsymbol{\zeta}$ and $\boldsymbol{\psi}_i$ are now to be chosen from sub-gradients of $f$ and $g_i$ at $\bar{\mathbf{x}}$, respectively, we can use the first condition to set $\epsilon$ such that it is the maximum of two quantities: (i) the minimum squared value of $\|\boldsymbol{\zeta} + \sum_{i=1}^{m} \lambda_i \boldsymbol{\psi}_i\|$ and (ii) $\sum_{i=1}^{m} \lambda_i g_i(\bar{\mathbf{x}})$, over $\boldsymbol{\zeta} \in \partial^o f(\bar{\mathbf{x}})$, $\boldsymbol{\psi}_i \in \partial^o g_i(\bar{\mathbf{x}})$ and $\lambda_i \geq 0$ for $i = 1, \ldots, m$.

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

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

**Theorem 4** *Let us consider the problem (P) with locally Lipschitz objective function and constraints. Let $\{\mathbf{x}^\nu\}$ be a sequence of vectors feasible to (P) and let $\mathbf{x}^\nu \to \bar{\mathbf{x}}$ as $\nu \to \infty$. Consider $\{\epsilon_\nu\}$ to be a sequence of positive real numbers such that $\epsilon_\nu \downarrow 0$ as $\nu \to \infty$. Further assume that for each $\nu$, $\mathbf{x}^\nu$ is a modified $\epsilon_\nu$-KKT point of (P). Let the Basic Constraint qualification hold at $\bar{\mathbf{x}}$. Then $\bar{\mathbf{x}}$ is a KKT point of (P).*

**Proof** Since each $\mathbf{x}^\nu$ is a modified $\epsilon_\nu$-KKT point, for each $\nu$ there exists $\mathbf{y}^\nu$ with $\|\mathbf{x}^\nu - \mathbf{y}^\nu\| \leq \sqrt{\epsilon_\nu}$ and there exists $\boldsymbol{\zeta}^\nu \in \partial^o f(\mathbf{y}^\nu)$, $\boldsymbol{\psi}_i^\nu \in \partial^o g_i(\mathbf{y}^\nu)$, $i = 1, 2, \dots m$ and scalars $\lambda_i^\nu \geq 0$, $i = 1, 2 \dots m$ such that

1. $\|\boldsymbol{\zeta}^\nu + \sum_{i=1}^m \lambda_i^\nu \boldsymbol{\psi}_i^\nu\| \leq \sqrt{\epsilon_\nu}$
2. $\sum_{i=1}^m \lambda_i^\nu g_i(\mathbf{x}^\nu) \geq -\epsilon_\nu$

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

$$\|\mathbf{y}^\nu - \bar{\mathbf{x}}\| \leq \|\mathbf{y}^\nu - \mathbf{x}^\nu\| + \|\mathbf{x}^\nu - \bar{\mathbf{x}}\|$$

Hence,

$$\|\mathbf{y}^\nu - \bar{\mathbf{x}}\| \leq \sqrt{\epsilon_\nu} + \|\mathbf{x}^\nu - \bar{\mathbf{x}}\|$$

Now as $\nu \to \infty$, $\epsilon_\nu \downarrow 0$ and $\mathbf{x}^\nu \to \bar{\mathbf{x}}$. This shows that $\mathbf{y}^\nu \to \bar{\mathbf{x}}$. Since, the Clarke subdifferential is locally bounded, the sequence $\{\boldsymbol{\zeta}\}$ and $\{\boldsymbol{\psi}_i^\nu\}$, for $i = 1, 2, \dots m$ are bounded. Thus, without loss of generality we can conclude that $\boldsymbol{\psi}_i^\nu \to \bar{\boldsymbol{\psi}}_i$ for all $i = 1, 2, \dots m$. Further, as $\partial^o g_i(\bar{\mathbf{x}})$. From 1 we have:

$$\frac{1}{\|\lambda^\nu\|}\|\boldsymbol{\zeta}^\nu + \sum_{i=1}^m \lambda_i^\nu \boldsymbol{\psi}_i^\nu\| \leq \sqrt{\epsilon_\nu}$$

Thus,

$$\|\frac{1}{\|\lambda^\nu\|}\boldsymbol{\zeta}^\nu + \sum_{i=1}^m w_i^\nu \boldsymbol{\psi}_i^\nu\| \leq \frac{1}{\|\lambda^\nu\|}\sqrt{\epsilon_\nu}$$

where $w^\nu = (w_1^\nu, w_2^\nu, \dots w_m^\nu)$, and $w^\nu = \frac{\lambda^\nu}{\|\lambda^\nu\|}$. Now as $\epsilon_\nu \downarrow 0$, $\sqrt{\epsilon_\nu} \to 0$ and hence, $\{\sqrt{\epsilon_\nu}\}$ is a bounded sequence and

$$\frac{1}{\|\lambda^\nu\|}\sqrt{\epsilon_\nu} \to 0 \ as \ \nu \to \infty$$

Further as $\{\boldsymbol{\zeta}^\nu\}$ is a bounded sequence, we have $\frac{1}{\|\lambda^\nu\|}\boldsymbol{\zeta}^\nu \to 0$ as $\nu \to \infty$. Hence from we have

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

i.e. $\sum_{i=1}^m w_i \bar{\boldsymbol{\psi}}_i = 0$ where $w = (w_1, w_2, \dots w_m)$ and $w^\nu \to w$. As $\|w\| = 1$, it shows the Basic constraint qualification is violated. Hence, $\{\lambda^\nu\}$ is a bounded sequence and thus we can assume that $\lambda^\nu \to \bar{\lambda}$, $\bar{\lambda} \in \mathbf{R}_+^m$. Further as $\{\boldsymbol{\zeta}^\nu\}$ is bounded we can assume that $\boldsymbol{\zeta}^\nu \to \bar{\boldsymbol{\zeta}}$ and since $\partial^o f$ is graph closed, $\bar{\boldsymbol{\zeta}} \in \partial^o f(\bar{\mathbf{x}})$. Hence, from 1 we can have $\|\bar{\boldsymbol{\zeta}} + \sum_{i=1}^m \bar{\lambda}_i \bar{\boldsymbol{\psi}}_i\| \leq 0$, where $\bar{\lambda} = (\bar{\lambda}_1, \bar{\lambda}_2, \dots, \bar{\lambda}_m)$. Thus, $\bar{\boldsymbol{\zeta}} + \sum_{i=1}^m \bar{\lambda}_i \bar{\boldsymbol{\psi}}_i = 0$. This shows that $\mathbf{0} \in \partial^o f(\bar{\mathbf{x}}) + \sum_{i=1}^m \bar{\lambda}_i \partial^o g_i(\bar{x})$. From 2 we have $\sum_{i=1}^m \bar{\lambda}_i g_i(\bar{\mathbf{x}}) \geq 0$. Now since $\mathbf{x}^\nu$ is feasible we have $g_i(\mathbf{x}^\nu) \leq 0$, for all $i = 1, 2, \dots m$. Hence, $g_i(\bar{x}) \leq 0$ for all $i = 1, 2, \dots m$. Hence, $\bar{x}$ is feasible to (P). Since $\bar{\lambda}_i \geq 0$, we have $\sum_{i=1}^m \bar{\lambda}_i g_i \bar{\mathbf{x}} \leq 0$. This shows that,

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

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

**Theorem 5** *Let us consider the convex problem (P). Let the Slater's constraint qualification holds, that is, there exists* $\mathbf{x}^* \in \mathbf{R}^n$ *such that* $g_i(\mathbf{x}^*) < 0$ *for all* $i = 1, \ldots, m$*. Let* $\{\mathbf{x}^\nu\}$ *be a sequence of* $\epsilon_\nu$*-minimum points and let* $\mathbf{x}^\nu \to \bar{\mathbf{x}}$ *as* $\epsilon_\nu \downarrow 0$*. Then,* $\bar{\mathbf{x}}$ *is a minimum point of (P).*

**Proof** Since $\mathbf{x}^\nu$ is an $\epsilon_\nu$-minimum point by applying Theorem 3 we have $\mathbf{x}^\nu$ is modified $\epsilon_\nu$-KKT point for (P) for each $\nu$. Now applying Theorem 4 we have $\bar{\mathbf{x}}$ is a KKT point for (P) and hence $\bar{\mathbf{x}}$ is a solution of (P). Note that that Slater's Constraint Qualification implies that Basic Constraint qualification is the case of a convex optimization problem.

## Conclusions

In this note, we have defined an approximate KKT point for differentiable and non-differential problems. The definitions can be used to determine whether iterates obtained from an optimization algorithm will eventually converge to a KKT point.

## References


[1] M. S. Bazaraa, H. D. Sherali, and C. M. Shetty. *Nonlinear programming: Theory and algorithms*. Singapore: Wiley, 2004.

[2] C. R. Bector, S. Chandra, and J. Dutta. *Principles of Optimization Theory*. New Delhi: Narosa, 2005.

[3] F. H. Clarke. *Optimization and nonsmooth analysis*. Wiley-Interscience, 1983.

[4] O. L. Mangasarian and S. Fromovitz. The fritz john necessary optimality conditions in the presence of equality and inequality constrtaints. *Journal of Mathematical Analysis and Applications*, 17:37–47, 1967.

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