Appendix Proofs for the Note

(https://bobondemon.github.io/2017/11/14/Notes-for-KKT-Conditions/)

Chih-Sheng Chen

Thm5: The following two statements are true.

(1) If d is a limiting direction of a feasible sequence at a feasible point x^*, then

$d^T \nabla c_i(x^*) = 0, \forall i \in \mathrm{E}$ and $d^T \nabla c_i(x^*) \geq 0, \forall i \in \mathrm{A}(x^*) \cap \mathrm{I}$.

(2) $\forall d \in \left\{ d \mid \begin{array}{c} d^T \nabla c_i(x^*) = 0, i \in \mathrm{E} \\ d^T \nabla c_i(x^*) \geq 0, i \in \mathrm{A}(x^*) \cap \mathrm{I} \\ \|d\| = 1 \end{array} \right\}$ and ***LICQ condition*** is satisfied, then d is

a limiting direction for some feasible sequence at x^*.

<u>In summary, suppose LICQ holds, then</u> $\left\{ d \mid \begin{array}{c} d^T \nabla c_i(x^*) = 0, i \in \mathrm{E} \\ d^T \nabla c_i(x^*) \geq 0, i \in \mathrm{A}(x^*) \cap \mathrm{I} \\ \|d\| = 1 \end{array} \right\}$ ***<u>= the set of</u>***

<u>all limiting directions at</u> x^****<u>.</u>***

Pf:

(1) W.l.o.g. Let $\{z_k\}$ be such feasible sequence that $\exists 1$ limiting direction d. Then by the definition of derivative, we have

$$\lim_{k \to \infty} \frac{c_i(z_k) - c_i(x^*) - \nabla c_i(x^*)^T (z_k - x^*)}{\|z_k - x^*\|} = 0$$

$$\Rightarrow \lim_{k \to \infty} \frac{c_i(z_k) - c_i(x^*)}{\|z_k - x^*\|} = \lim_{k \to \infty} \frac{\nabla c_i(x^*)^T (z_k - x^*)}{\|z_k - x^*\|}$$

$$\Rightarrow \lim_{k \to \infty} \frac{c_i(z_k) - c_i(x^*)}{\|z_k - x^*\|} = \nabla c_i(x^*)^T \lim_{k \to \infty} \frac{(z_k - x^*)}{\|z_k - x^*\|} = \nabla c_i(x^*)^T d$$

$$\therefore \nabla c_i(x^*)^T d = \lim_{k \to \infty} \frac{c_i(z_k) - c_i(x^*)}{\|z_k - x^*\|} = \lim_{k \to \infty} \frac{0 - 0}{\|z_k - x^*\|} = 0, \forall i \in \mathrm{E}$$

$$\text{and } \nabla c_i(x^*)^T d = \lim_{k \to \infty} \frac{c_i(z_k) - c_i(x^*)}{\|z_k - x^*\|} = \lim_{k \to \infty} \frac{c_i(z_k)}{\|z_k - x^*\|} \geq 0, \forall i \in \mathrm{A}(x^*) \cap \mathrm{I}$$

(2) The proof of this part is quite tricky. First, since LICQ holds, we have a $m \times n$

matrix $A = \begin{bmatrix} \nabla c_1(x^*)^T \\ \vdots \\ \nabla c_m(x^*)^T \end{bmatrix}_{m \times n}$ of active constraint gradients with full row rank m.

Let Z be a matrix whose columns are a basis for the null space of A; that is, $Z \in \Re^{n\times(n-m)}, AZ = 0$ and Z has full column rank.

Let $\{t_k\}$ be any sequence of positive scalars such $\lim_{k\to\infty} t_k = 0$. Define the parameterized system of equations $R : \Re^n \times \Re \to \Re^n$ by

$$R(z,t) = \begin{bmatrix} c(z) - tAd \\ Z^T(z - x^* - td) \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}.$$

We claim that for each $t = t_k$, the solutions $z = z_k$ of this system for small $t > 0$ give a feasible sequence $\{z_k\}$ that approaches x^* and has a unique limiting direction that equals to d.

Clearly, $R(x^*,0) = 0$, and $\nabla_z R(x^*,0) = \begin{bmatrix} \nabla c_1(x^*)^T \\ \vdots \\ \nabla c_m(x^*)^T \\ Z^T \end{bmatrix} = \begin{bmatrix} A \\ Z^T \end{bmatrix}$ which is nonsingular.

Hence, by Implicit Function Theorem, we can locally parameterize z by t around the point $(x^*,0)$. Formally speaking, $\exists$ n.b.d. V of 0 and W of x^*, and $\exists 1\, f \in C^p(V,W)$, such that $R(f(t),t) = 0, \forall t \in V$ and $f(0) = x^*$. Hence for all k sufficiently large, we have $t_k \in V$, and we define $z_k = f(t_k)$. Now consider the sequence $\{z_k\}$. We first prove it is a feasible sequence and then prove it has a unique limiting direction that equals to d, and hence conclude the proof.

Since $c(z_k) - t_k Ad = 0$, we have

$$\begin{bmatrix} c_1(z_k) \\ \vdots \\ c_m(z_k) \end{bmatrix} = t_k \begin{bmatrix} \nabla c_1(x^*)^T d \\ \vdots \\ \nabla c_m(x^*)^T d \end{bmatrix} \Rightarrow \begin{matrix} c_i(z_k) = 0, \forall i \in \mathrm{E} \\ c_i(z_k) \geq 0, \forall i \in \mathrm{A}(x^*) \cap \mathrm{I} \end{matrix} \Rightarrow z_k \text{ is feasible} \ldots\ldots\ldots(\mathrm{a}).$$

Also, $\because f \in C^p$ and $f(0) = x^*$, we have

$$\lim_{k\to\infty} z_k = \lim_{k\to\infty} f(t_k) = f\left(\lim_{k\to\infty} t_k\right) = f(0) = x^* \ldots\ldots\ldots(\mathrm{b}).$$

Moreover, $z_k \neq x^*, \forall k \ldots\ldots\ldots(\mathrm{c})$.

Since if it is true for some $t_k > 0$, we have

$R(z_k,t_k) = R(x^*,t_k) = 0 = \begin{bmatrix} c(x^*) - t_k Ad \\ Z^T(x^* - x^* - t_k d) \end{bmatrix} = -t_k \begin{bmatrix} A \\ Z^T \end{bmatrix} d$. Since $\begin{bmatrix} A \\ Z^T \end{bmatrix}$ is nonsingular and $d \neq 0$, we have a contradiction.

From $(a),(b)$ and (c), $\{z_k\}$ is a feasible sequence.
It remains to show that d is a limiting direction of $\{z_k\}$. Using the fact that $R(z_k, t_k) = 0$ for all k together with Taylor's theorem, we find that

$$0 = R(z_k, t_k) = \begin{bmatrix} c(z_k) - t_k Ad \\ Z^T(z_k - x^* - t_k d) \end{bmatrix} = \begin{bmatrix} c(x^*) + A(z_k - x^*) + o(\|z_k - x^*\|) - t_k Ad \\ Z^T(z_k - x^* - t_k d) \end{bmatrix}$$

$$= \begin{bmatrix} A(z_k - x^* - t_k d) + o(\|z_k - x^*\|) \\ Z^T(z_k - x^* - t_k d) \end{bmatrix} == \begin{bmatrix} A \\ Z^T \end{bmatrix}(z_k - x^* - t_k d) + o(\|z_k - x^*\|)$$

By dividing this expression by $\|z_k - x^*\|$ and using nonsingularity, we obtain

$$\lim_{k\to\infty} \frac{z_k - x^* - t_k d}{\|z_k - x^*\|} = 0 \Rightarrow \lim_{k\to\infty} d_k = \left(\lim_{k\to\infty} \frac{t_k}{\|z_k - x^*\|}\right) d, \text{where } d_k = \frac{z_k - x^*}{\|z_k - x^*\|}.$$

Since $\|d\| = \|d_k\| = 1, \forall k$, we must have $\lim_{k\to\infty} \frac{t_k}{\|z_k - x^*\|} = 1$. Hence

$\lim_{k\to\infty} d_k = \left(\lim_{k\to\infty} \frac{t_k}{\|z_k - x^*\|}\right) d = d$. Therefore, $\{z_k\}$ has a limiting direction d. (The uniqueness is clear, since every convergent subsequences of a convergent sequence converge to the same point.)

□

Thm6: W.l.o.g. Let $\mathrm{A}(x^*) = \{1 \ldots m\}$, $A^T = [\nabla c_1(x^*) \cdots \nabla c_m(x^*)]$, and $x^* \in$ feasible point

$\nabla f(x^*)^T d \geq 0$ for all $d \in F_1$ $\Leftrightarrow$

$\exists \lambda \in \Re^m$ where $\lambda_i \geq 0 \ \forall i \in \mathrm{A}(x^*) \cap \mathrm{I}$, such that $\nabla f(x^*) = \sum_{i=1}^m \lambda_i \nabla c_i(x^*) = A^T \lambda$

Pf:
$(\Leftarrow)$ For any $d \in F_1$

$$\nabla f(x^*)^T d = \sum_{i=1}^m \lambda_i \nabla c_i(x^*)^T d = \sum_{i\in \mathrm{E}} \lambda_i \nabla c_i(x^*)^T d + \sum_{i \in \mathrm{A}(x^*)\cap \mathrm{I}} \lambda_i \nabla c_i(x^*)^T d$$

$$= \sum_{i \in \mathrm{A}(x^*)\cap \mathrm{I}} \lambda_i \nabla c_i(x^*)^T d \geq 0$$

$(\Rightarrow)$ We define the cone N by

$$N = \left\{ s \mid s = \sum_{i\in \mathrm{A}(x^*)} \lambda_i \nabla c_i(x^*), \text{where } \lambda_i \geq 0, \forall i \in \mathrm{A}(x^*) \cap \mathrm{I} \right\},$$

the condition is equivalent to $\nabla f(x^*) \in N$. We first note that N is closed (every accumulation points are still in N). Then if $\nabla f(x^*) \notin N$, we are to prove $\exists d \in F_1$ such that $\nabla f(x^*)^T d < 0$.

Since N is closed, we have a $\hat{s} \in N$ such that $\hat{s}$ is the closest point in N to $\nabla f(x^*)$. In other words, $\hat{s}$ is the only one minimizer ($\because N$ is a convex set, check!) of the following convex problem,

$$\min_{s \in N} \left\| s - \nabla f(x^*) \right\|_2^2$$

Since $\hat{s} \in N, \therefore t\hat{s} \in N$ for $t > 0$, and $\left\| t\hat{s} - \nabla f(x^*) \right\|_2^2$ is minimized at $t = 1$. Hence,

$$\frac{d}{dt} \left\| t\hat{s} - \nabla f(x^*) \right\|_2^2 \Big|_{t=1} = 0 \Rightarrow \hat{s}^T \left(\hat{s} - \nabla f(x^*) \right) = 0 \ldots\ldots\ldots (a).$$

Now, let s be any other vector in N. Since N is convex, we have by the minimizing property of $\hat{s}$ that

$$\left\| (1-\theta)\hat{s} + \theta s - \nabla f(x^*) \right\|_2^2 \geq \left\| \hat{s} - \nabla f(x^*) \right\|_2^2, \text{ for all } \theta \in [0,1]$$
$$\Rightarrow 2\theta (s - \hat{s})^T \left(\hat{s} - \nabla f(x^*) \right) + \theta^2 \left\| s - \hat{s} \right\|_2^2 \geq 0, \text{ for all } \theta \in [0,1]$$

Let $\theta \to 0^+$, $\Rightarrow (s - \hat{s})^T \left(\hat{s} - \nabla f(x^*) \right) \geq 0$, and by (a), we have

$$s^T \left(\hat{s} - \nabla f(x^*) \right) \geq 0, \text{ for all } s \in N \ldots\ldots\ldots (b).$$

We claim that if $d = \hat{s} - \nabla f(x^*)$, then $d \in F_1$ and $\nabla f(x^*)^T d < 0$.

First, $\nabla f(x^*)^T d = (\hat{s} - d)^T d = \hat{s}^T \left(\hat{s} - \nabla f(x^*) \right) - \|d\|_2^2 = -\|d\|_2^2 < 0$.

We are to prove $d \in F_1$, i.e.

$$\begin{cases} d^T \nabla c_i(x^*) = 0, \forall i \in \mathrm{E} \\ d^T \nabla c_i(x^*) \geq 0, \forall i \in \mathrm{A}(x^*) \cap \mathrm{I} \end{cases}$$

Note that, $\nabla c_i(x^*), \text{and} -\nabla c_i(x^*) \in N, \text{for all } i \in \mathrm{E}$, and

$\nabla c_i(x^*) \in N, \text{for all } i \in \mathrm{A}(x^*) \cap \mathrm{I}$. By (b), we have

($\nabla c_i(x^*)^T d \geq 0$ and $-\nabla c_i(x^*)^T d \geq 0, \forall i \in \mathrm{E}$) $\Rightarrow \nabla c_i(x^*)^T d = 0, \forall i \in \mathrm{E}$, and

$\nabla c_i(x^*)^T d \geq 0$ for all $i \in \mathrm{A}(x^*) \cap \mathrm{I}$.

□