KARUSH-KUHN-TUCKER CONDITIONS

Wednesday, April 17, 2019 11:06 AM

INTRODUCTION

IN THIS SECTION WE INTRODUCE THE FAMOUS KARUSH-KUHN-TUCKER (K.K.T.) CONDITIONS FOR THE CONSTRAINED NON-LINEAR OPTIMIZATION PROBLEM. WE WILL SEE:

1. UNDER SOME ASSUMPTIONS, THE KKT CONDITIONS IS NECESSARY FOR THE (LOCAL) OPTIMALITY.
2. IF OUR PROBLEM IS CONVEX, THE KKT CONDITIONS IS SUFFICIENT AND NECESSARY FOR THE GLOBAL OPTIMALITY.

PROBLEMS WITH ONLY INEQUALITY CONSTRAINTS

CONSIDER PROBLEM

$$\min f(\underline{x})$$

$$\text{s.t.} \quad g_i(\underline{x}) \leq 0 \qquad i \in I = \{1, \dots, m\}$$

SPACE OF CONTINUOUSLY DIFFERENTIABLE FUNCTIONS

WHERE $f, g \in \mathcal{C}^1(\mathbb{R}^n)$, $\underline{x} \in \mathbb{R}^n$, AND WITH THE ASSUMPTION: THE FEASIBLE REGION IS NOT EMPTY, BUT ITS INTERIOR CAN BE EMPTY.

MATHEMATICALLY SPEAKING,

FEASIBLE REGION → $S = \{\underline{x} \in \mathbb{R}^n : g_i(\underline{x}) \leq 0, \ \forall i \in I\} \neq \phi$

DEF 1 FOR EACH $\underline{x} \in S$, WE DEFINE:

- CONE OF THE FEASIBLE DIRECTIONS AS

$$D(\underline{x}) := \{ \underline{d} \in \mathbb{R}^n : \exists \tilde{\alpha} > 0 \text{ s.t. } \underline{x} + \alpha \underline{d} \in S, \ \forall \alpha \in [0, \tilde{\alpha}] \}$$

WHY DO WE DEFINE $D(\underline{x})$? BECAUSE GRADIENT IS THE DIRECTION WITH WHICH THE FUNCTION

- SET OF INDICES OF ACTIVE CONSTRAINTS

- SET OF INDICES OF ACTIVE CONSTRAINTS

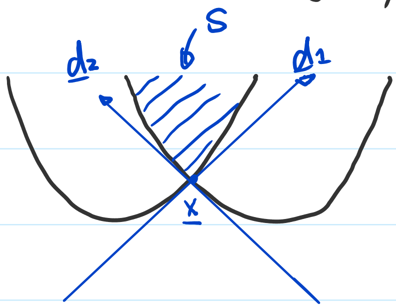
$$I(\underline{x}) := \{ i \in I : g_i(\underline{x}) = 0 \} \subseteq I$$

BECAUSE GRADIENT IS THE DIRECTION WITH WHICH THE FUNCTION GROWS MOST FASTLY, SO WE WANT TO CHOOSE THE DIRECT S.T. THE FUNCTION POSSIBLY DECREASES, IN THIS WAY $g_i(\underline{x}) \leq 0$ STILL HOLDS

- CONE OF THE DIRECTIONS CONSTRAINED BY THE GRADIENTS OF THE ACTIVE CONSTRAINTS

$$D(\underline{x}) := \{ d \in \mathbb{R}^n : \nabla^T g_i(\underline{x}) \cdot \underline{d} \leq 0, \ \forall i \in I(\underline{x}) \}$$

REMARK THE CONE $\Phi(\underline{x})$ CAN BE (TOPOLOGICALLY) OPEN, FOR EXAMPLE



OBVIOUSLY $\underline{d}_1$ AND $\underline{d}_2$ ARE NEITHER FEASIBLE DIRECTIONS FOR $\underline{x}$.

VECTOR IN $\mathbb{R}^n$, THEN WE DISCUSS OPEN AND CLOSED WITH RESPECT TO CANONICAL TOPOLOGY OF $\mathbb{R}^n$.

PROP 2 $\overline{\Phi(\underline{x})} \subseteq D(\underline{x})$.

PROOF GIVEN $\underline{d} \in \Phi(\underline{x})$, FOR SUFFICIENTLY SMALL VALUES OF α WE HAVE

DEFINED AS $\lim_{\alpha \to 0} \frac{o(\alpha)}{\alpha} = 0$

$$0 \geq g_i(\underline{x} + \alpha \underline{d}) = g_i(\underline{x}) + \alpha \nabla^T g_i(\underline{x}) \cdot \underline{d} + \underline{o(\alpha)}$$

FOR ALL $i \in I(\underline{x})$ WITH $g_i(\underline{x}) = 0$. THEREFORE

$$\nabla^T g_i(\underline{x}) \underline{d} \leq 0 \leftarrow \text{LET } \alpha \to 0^+: \ 0 \geq \lim_{\alpha \to 0^+} \frac{g_i(\underline{x} + \alpha \underline{d})}{\alpha} = \nabla^T g_i(\underline{x}) \underline{d}$$

FOR ALL $i \in I(\underline{x})$, THAT IS $\underline{d} \in D(\underline{x})$. SO $\Phi(\underline{x}) \subseteq D(\underline{x})$.

MOREOVER, ACCORDING TO THE DEFINITION OF $D(\underline{x})$, $D(\underline{x})$ MUST BE CLOSED SET. HENCE $\overline{\Phi(\underline{x})} \subseteq D(\underline{x})$. ∎

THEOREM 3 [EXTENSION OF THE FIRST ORDER NECESSARY OPT. COND.S]

IF $f \in \mathcal{C}^1(S)$ AND $\underline{\tilde{x}} \in S$ IS A LOCAL MINIMUM OF f ON S,

THEN $\nabla^T f(\tilde{x}) d \geq 0$ FOR ALL $d \in \overline{\Phi(\tilde{x})}$, THAT IS ... THE

IF $f \in \mathcal{C}^1(S)$ AND $\tilde{\underline{x}} \in S$ IS A LOCAL MINIMUM OF f ON S, THEN $\nabla^T f(\tilde{\underline{x}})\, \underline{d} \geq 0$, FOR ALL $\underline{d} \in \overline{\mathcal{D}(\tilde{\underline{x}})}$, THAT IS, ALL THE FEASIBLE DIRECTIONS ARE ASCENT DIRECTIONS.

PROOF WE ALREADY KNOW THE RESULT HOLDS FOR ALL $\underline{d} \in \mathcal{D}(\tilde{\underline{x}})$, IT SUFFICES TO EXTEND IT TO THE CLOSURE. FOR ALL $\underline{d} \in \overline{\mathcal{D}(\tilde{\underline{x}})}$, THERE EXISTS $\{\underline{d}_k\}_{k \in \mathbb{N}}$ SUCH THAT $\lim_{k \to \infty} \underline{d}_k = \underline{d}$. SINCE $\nabla f^T(\tilde{\underline{x}})\, \underline{d}^k \geq 0$ FOR ALL $k \in \mathbb{N}$, THEN

$$\lim_{k \to +\infty} \nabla f^T(\tilde{\underline{x}})\, \underline{d}_k = \nabla f^T(\tilde{\underline{x}})\, \underline{d} \geq 0$$

BECAUSE $\nabla f \in \mathcal{C}^0(S)$, RECALL THAT $f \in \mathcal{C}^1(S)$. ∎

NOTICE THAT THESE NECESSARY CONDITIONS CAN BE HARDLY USED BECAUSE $\overline{\mathcal{D}(\underline{x})}$ IS DIFFICULT TO CHARACTERIZE. SINCE $D(\underline{x})$ IS WELL-CHARACTERIZED, WE INTRODUCE FURTHER CONDITIONS ON THE SET OF CONSTRAINTS.

DEF 4 [ZANGWILL CONSTRAINT QUALIFICATION, C.Q.]
THE CONSTRAINT QUALIFICATION ASSUMPTION HOLDS AT $\underline{x} \in S$ IF $\overline{\mathcal{D}(\underline{x})} = D(\underline{x})$.

THEOREM 5 [KARUSH-KUHN-TUCKER NECESSARY OPTIMALITY CONDS]
SUPPOSE $f, g \in \mathcal{C}^1(\mathbb{R}^n)$ AND THE C.Q. ASSUMPTION HOLDS AT $\tilde{\underline{x}} \in S = \{\underline{x} \in \mathbb{R}^n : g_i(\underline{x}) \leq 0,\ \forall i \in I\}$. IF $\tilde{\underline{x}}$ IS A LOCAL MINIMUM OF f OVER S, THEN $\exists\, u_1, \dots, u_m \geq 0$ (REFERRED AS KKT-MULTIPLIERS) SUCH THAT:

$$\nabla f(\tilde{\underline{x}}) + \sum u_i \nabla g_i(\tilde{\underline{x}}) = 0$$

AS KKT-MULTIPLIERS) SUCH THAT:

$$\nabla f(\tilde{\underline{x}}) + \sum_{i\in I(\tilde{\underline{x}})} u_i \nabla g_i(\tilde{\underline{x}}) = \underline{0}$$

OR EQUIVALENTLY,

$$\begin{cases} \nabla f(\tilde{\underline{x}}) + \sum_{i=1}^{m} u_i \nabla g_i(\tilde{\underline{x}}) = \underline{0} \\ u_i g_i(\tilde{\underline{x}}) = 0, \quad \forall i \in I \end{cases}$$

PROOF ASSUMING C.Q. ASSUMPTIONS HOLDS AT $\tilde{\underline{x}}$, WE HAVE $\overline{\mathcal{D}}(\tilde{\underline{x}}) = D(\tilde{\underline{x}})$.

A NECESSARY CONDITION FOR $\tilde{\underline{x}}$ TO BE A LOCAL MINIMUM OF f OVER S IS: ORIGINALLY WE HAVE (*) HOLDS FOR $d \in \overline{\mathcal{D}}(\tilde{\underline{x}})$, SINCE WE HAVE C.Q., THEN $\overline{\mathcal{D}}(\tilde{\underline{x}}) = D(\tilde{\underline{x}})$, WE GET THIS:

$$(*) \quad \nabla f^T(\tilde{\underline{x}})\,\underline{d} \geq 0, \quad \forall \underline{d} \text{ s.t. } \nabla g_i^T(\tilde{\underline{x}})\,\underline{d} \leq 0, \ \forall i \in I(\tilde{\underline{x}})$$

NOW RECALL THE **FARKAS LEMMA**:

$$\begin{cases} A\underline{u} = \underline{b} \\ \underline{u} \geq 0 \end{cases} \text{ HAS A SOLUTION} \iff \begin{cases} \underline{b}^T\underline{d} \geq 0 \\ \forall \underline{d} \text{ s.t. } \underline{d}^T A \geq \underline{0} \end{cases} \text{ HAS A SOLUTION}$$

TAKING $\underline{b} = \nabla f(\tilde{\underline{x}})$ AND $A = (-\nabla g_{i_1}(\tilde{\underline{x}}), \ldots, -\nabla g_{i_\ell}(\tilde{\underline{x}}))$ MATRIX WITH n ROWS AND $\ell := |I(\tilde{\underline{x}})|$ COLUMNS, AND ACCORDING TO FARKAS LEMMA, (*) IS EQUIVALENT TO

$$\begin{cases} A\underline{u} = \underline{b} \\ \underline{u} \geq 0 \end{cases} \iff \begin{cases} \sum_{i\in I(\tilde{\underline{x}})} u_i(-\nabla g_i(\tilde{\underline{x}})) = \nabla f(\tilde{\underline{x}}) \\ \underline{u} \geq 0 \end{cases}$$

WHICH COMPLETES THE PROOF. $\blacksquare$

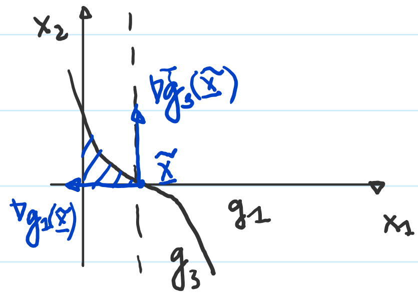
REMARK [GEOMETRIC INTERPRETATION]

FOR A LOCAL MINIMUM $\tilde{\underline{x}}$, $-\nabla f(\tilde{\underline{x}})$ MUST BE EXPRESSIBLE AS A LINEAR COMBINATION WITH COEFFICIENTS $u_i \geq 0$ OF THE $\nabla g_i(\tilde{\underline{x}})$, FOR ALL $i \in I(\tilde{\underline{x}})$, I.E., MUST FORM AN **OBTUSE ANGLE** WITH ALL THE FEASIBLE DIRECTIONS.

OF THE $\nabla g_i(\underline{x})$, FOR ALL $i \in I(\underline{x})$, I.E., MUST FORM AN OBTUSE ANGLE WITH ALL THE FEASIBLE DIRECTIONS.

REMARK IF C.Q. ASSUMPTION DOES NOT HOLD AT $\underline{\tilde{x}}$, THEN THE KKT CONDITIONS NEED NOT BE NECESSARY FOR $\underline{\tilde{x}}$ TO BE A LOCAL OPTIMAL SOLUTION. FOR EXAMPLE, CONSIDER

$$\min f(\underline{x}) = -x_1$$
$$\text{s.t.} \quad g_1(\underline{x}) = -x_1 \leq 0$$
$$g_2(\underline{x}) = -x_2 \leq 0$$
$$g_3(\underline{x}) = -(1-x_1)^3 + x_2 \leq 0.$$



WE CAN OBSERVE THAT GLOBAL MINIMAL IS OBTAINED AT $\underline{\tilde{x}}$, THEN $\overline{\mathcal{D}(\underline{\tilde{x}})} = \{(a,0) : a \leq 0\}$. ON THE OTHER HAND

$$\nabla^T g_1(\underline{\tilde{x}}) = (-1, 0), \quad \nabla^T g_3(\underline{\tilde{x}}) = (0, 1)$$

THEN OBVIOUSLY $D(\underline{\tilde{x}}) \neq \overline{\mathcal{D}(\underline{\tilde{x}})}$.

NOW, f OBTAINS THE GLOBAL MINIMUM AT $\underline{\tilde{x}} = (1,0)$ BUT NO MULTIPLIERS $u_1, u_2, u_3 \geq 0$ SUCH THAT $\nabla f(\underline{\tilde{x}}) + \sum_{i \in I(\underline{\tilde{x}})} u_i \nabla g_i(\underline{\tilde{x}}) = \underline{0}$.

INDEED $u_1 = 0$, AND

$$\begin{pmatrix} -1 \\ 0 \end{pmatrix} + u_2 \begin{pmatrix} 0 \\ -1 \end{pmatrix} + u_3 \begin{pmatrix} 0 \\ 1 \end{pmatrix} \neq \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \quad \forall\, u_2, u_3 \geq 0$$

THAT IS, $\underline{\tilde{x}}$ DOES NOT SATISFY THE KKT CONDITIONS.

SINCE C.Q. ASSUMPTION IS DIFFICULT TO VERIFY IN PRATICE, THERE EXIST SUFFICIENT CONDITIONS FOR C.Q. TO HOLD.

PROP 6 [SUFFICIENT CONDITIONS FOR CONSTRAINT QUALIFICATION]

IF 1) ALL g_i ARE LINEAR FUNCTIONS. OR

2) ALL g_i'S ARE CONVEX AND THERE EXISTS $\underline{a} \in \mathbb{R}^n$

IF 1) ALL g_i ARE LINEAR FUNCTIONS. OR

2) ALL g_i'S ARE CONVEX AND THERE EXISTS $\underline{a} \in \mathbb{R}^n$ SUCH THAT $g_i(\underline{a}) < 0$ FOR ALL $i \in I$.

THEN THE C.Q. ASSUMPTION HOLDS AT EVERY FEASIBLE POINT $\underline{x} \in S$.

IF 3) $\nabla g_i(\underline{x})$ ARE LINEARLY INDEPENDENT FOR ALL $i \in I(\underline{x})$.

THEN THE C.Q. ASSUMPTION HOLDS AT $\underline{x} \in S$.

REMARK IN 3), WHEN THE GRADIENTS OF THE ACTIVE CONSTRAINTS ARE LINEAR INDEPENDENT, THE KKT MULTIPLIER VECTOR IS UNIQUE.

THEN WE REVEAL THE IMPORTANCE OF KKT CONDITIONS IN CONVEX PROBLEMS:

THEOREM 7 [NECESSARY AND SUFFICIENT CONDITIONS FOR GLOBAL OPTIMALITY IN CONVEX PROBLEMS]

IF $f \in \mathcal{C}^1(\mathbb{R}^n)$ IS CONVEX, $g_i \in \mathcal{C}^1(\mathbb{R}^n)$ IS CONVEX FOR ALL $i \in I$, AND THERE EXISTS $\underline{a}$ SUCH THAT $g_i(\underline{a}) < 0$ FOR ALL $i \in I$, THEN $\underline{x}^* \in S$ IS A GLOBAL MINIMUM IF AND ONLY IF THERE EXIST $u_1, \dots, u_m \geq 0$ SUCH THAT

$$\begin{cases} \nabla f(\underline{x}^*) + \sum_{i=1}^{m} u_i \nabla g_i(\underline{x}^*) = \underline{0} \\ u_i g_i(\underline{x}^*) = 0, \quad \forall i \in I. \end{cases}$$

PROOF "$\Rightarrow$": $\underline{x}^*$ (LOCAL) MINIMUM $\Rightarrow$ $\underline{x}^*$ SATISFIES KKT CONDITIONS (NOTICE WE HAVE C.Q. CONDITION, THEN KKT ARE NECESSARY)

"$\Leftarrow$": KKT CONDITIONS $\Rightarrow$ $\nabla f^T(\underline{x}^*)\underline{d} \geq 0$, $\forall \underline{d} \in \overline{\mathcal{D}(\underline{x}^*)} = D(\underline{x}^*)$ (BY C.Q. CONDITION AND FARKAS LEMMA)

$\Rightarrow$ $\underline{x}^*$ (GLOBAL) MINIMUM (THM FOR CONVEX PROBLEMS)

(w/ C.Q. CONDITION AND FARKAS LEMMA)

$\Rightarrow \underline{x}^*$ (GLOBAL) MINIMUM (THM FOR CONVEX PROBLEMS)

$\blacksquare$

PROBLEM IN GENERAL CASE

CONSIDER:

$$\begin{aligned} &\min f(\underline{x}) \\ \text{s.t.} \quad & g_i(\underline{x}) \leq 0 \quad i \in I = \{1, \dots, m\} \\ & h_\ell(\underline{x}) = 0 \quad \ell \in L = \{1, \dots, p\} \\ & \underline{x} \in X \subseteq \mathbb{R}^n \end{aligned}$$

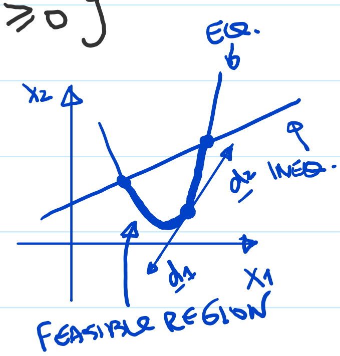
WHERE $f, g_i, h_\ell \in \mathcal{C}^1$ FOR ALL $i \in I$, ALL $\ell \in L$.

OBSERVE THAT: IN THE PRESENCE OF NON-LINEAR EQUALITY CONSTRAINTS, USUALLY $\mathcal{D}(\underline{x}) = \{\underline{0}\}$. TO EXTEND THE PREVIOUS RESULTS, WE DEFINE A CONE OF DIRECTIONS THAT ALSO TAKES INTO ACCOUNT THE EQUALITY CONSTRAINTS.

DEF 8 CLOSED CONE OF THE TANGENTS AT $\underline{x}$ IS DEFINED AS:

$$T(\underline{x}) := \left\{ \underline{d} \in \mathbb{R}^n : \underline{d} = \lambda \lim_{k \to \infty} \frac{\underline{x}_k - \underline{x}}{\|\underline{x}_k - \underline{x}\|}, \lambda \geq 0 \right\}$$

WHERE $\{x_k\}_{k \in \mathbb{N}}$ ARE FEASIBLE SOLUTIONS.



DEF 9 [ABADIE CONSTRAINT QUALIFICATION]

THE C.Q. ASSUMPTION HOLDS AT $\underline{x} \in \mathcal{S}$ IF

$$T(\underline{x}) = \mathcal{D}(\underline{x}) \cap H(\underline{x})$$

WHERE $D(\underline{x}) = \{ \underline{d} \in \mathbb{R}^n : \nabla g_i^T(\underline{x})\, \underline{d} \leq 0, \ \forall i \in I(\underline{x}) \}$

AND $H(\underline{x}) = \{ \underline{d} \in \mathbb{R}^n : \nabla h_\ell^T(\underline{x})\, \underline{d} = 0, \ \forall \ell \in L \}$.

AND $H(\underline{x}) = \{ \underline{d} \in \mathbb{R}^n : \nabla h_\ell^T(\underline{x})\, \underline{d} = 0,\ \forall \ell \in L \}$.

THEOREM 10 [GENERAL KKT NECESSARY OPTIMALITY CONDITIONS]

SUPPOSE $f \in \mathcal{C}^1$, $g_i \in \mathcal{C}^1$ FOR ALL $i \in I$, $h_\ell \in \mathcal{C}^1$ FOR ALL $\ell \in L$ AND THE (ABADIE) C.Q. ASSUMPTION HOLDS AT $\tilde{\underline{x}} \in S$. IF $\tilde{\underline{x}}$ IS A LOCAL MINIMUM OF f OVER S, THEN THERE EXIST $u_i \geq 0$ FOR ALL $i \in I(\tilde{\underline{x}})$ AND $v_\ell \in \mathbb{R}$ FOR ALL $\ell \in L$ SUCH THAT

$$\nabla f(\tilde{\underline{x}}) + \sum_{i \in I(\tilde{\underline{x}})} u_i \nabla g_i(\tilde{\underline{x}}) + \sum_{\ell \in L} v_\ell \nabla h_\ell(\tilde{\underline{x}}) = \underline{0}$$

Proof PLEASE REFER TO BOOK LINEAR AND NON-LINEAR PROGRAMMING, D.G. LUENBERGER, Y. YE, SPRINGER, THIRD EDITION, PAGE 342. NOTICE ON THE BOOK THE PROOF IS UNDER A SUFFICIENT ASSUMPTION FOR OBTAINING (ABAIDE) C.Q. CONDITION.

REMARK IN THE CASE THAT ALL CONSTRAINTS ARE EQUALITY CONSTRAINTS THE KKT CONDITIONS COINCIDE WITH THE CLASSICAL LAGRANGE OPTIMILITY CONDITION.

STILL, (ABADIE) C.Q. CONDITION IS DIFFICULT TO CHARACTERIZE, HERE, AT THE END, WE GIVE A SUFFICIENT CONDITION.

PROP 11 [SUFFICIENT CONDITION FOR C.Q.]

1) C.Q. ASSUMPTION HOLDS AT EVERY $\tilde{\underline{x}} \in S$ IF g_i ARE CONVEX, FOR ALL $i \in I$; AND h_ℓ ARE LINEAR, FOR ALL $\ell \in L$; AND

FOR ALL $i \in I$; AND h_ℓ ARE LINEAR, FOR ALL $\ell \in L$; AND THERE EXISTS $\underline{a} \in X \subset \mathbb{R}^n$ SUCH THAT $g_i(\underline{a}) < 0$ FOR ALL $i \in I$ AND ALL $h_\ell(\underline{a}) = 0$ FOR ALL $\ell \in L$.

2) C.Q. ASSUMPTION HOLDS AT $\underline{\tilde{x}} \in S$ IF $\nabla g_i(\underline{\tilde{x}}), \forall i \in I(\underline{\tilde{x}})$ AND $\nabla h_\ell(\underline{\tilde{x}}), \forall \ell \in L$ ARE LINEAR INDEPENDENT.