KKT Condition and Its Dual Form

橋本浩一 (koichi@k2.t.u-tokyo.ac.jp)

1 Constrained Problems

KKT 条件に関する証明，および KKT 条件と Dual Form との関係を簡潔に示す．本資料に関して不明な点があればメールにて質問してほしい．

簡単のため，等式条件の入らない問題について説明する．

$$\begin{array}{lll} \text{minimize} & f(\boldsymbol{w}), & \boldsymbol{w} \in \Omega \\ \text{subject to} & g_i(\boldsymbol{w}) \le 0, & i = 1, \dots, k \end{array} \tag{1}$$

2 Saddle Point Condition

定理 1: (1) 式の最適化において，$f, g_i : R^n \to R$ は任意の関数とし，Lagrangian

$$L(\boldsymbol{w}, \boldsymbol{\alpha}) = f(\boldsymbol{w}) + \sum_{i=1}^{k} \boldsymbol{\alpha}_i g_i(\boldsymbol{w}) \quad \text{where} \quad \alpha_i \ge 0 \tag{2}$$

を考える．もし，任意の $\boldsymbol{w} \in R^n$ と任意の $\boldsymbol{\alpha} \in [0, \infty)^n$ に対して

$$L(\bar{\boldsymbol{w}}, \boldsymbol{\alpha}) \le L(\bar{\boldsymbol{w}}, \bar{\boldsymbol{\alpha}}) \le L(\boldsymbol{w}, \bar{\boldsymbol{\alpha}}) \qquad \text{(Saddle Point)} \tag{3}$$

が成り立つような $\bar{\boldsymbol{w}} \in \Omega$ と $\bar{\alpha}_i \ge 0 \ \ (i = 1, \dots k)$ が存在するならば，$\bar{\boldsymbol{w}}$ は (1) のひとつの解である．

(証明) (3) をみたす $\bar{\boldsymbol{w}}, \bar{\boldsymbol{\alpha}}$ を考える．このとき第 1 の不等式より

$$\sum_{i=1}^{k} (\alpha_i - \bar{\alpha}_i) g_i(\bar{\boldsymbol{w}}) \le 0 \tag{4}$$

を得る．α_i は任意であるから $\alpha_i = \bar{\alpha}_i + 1, \alpha_j = \bar{\alpha}_j \ (j \ne i)$ とおくと $g_i(\boldsymbol{w}) \le 0$ であるから，最適化の条件をみたす．一方，$\alpha_i = 0, \alpha_j = \bar{\alpha}_j \ (j \ne i)$ とおくと $\bar{\alpha}_i g_i(\bar{\boldsymbol{w}}) \ge 0$ となる．このふたつの条件を同時にみたすためには

$$\bar{\alpha}_i g_i(\bar{\boldsymbol{w}}) = 0 \quad i = 1, \dots k \tag{5}$$

でなければならない．最後に，$g_i(x) \ge 0$ を考慮すると (3) の第 2 の不等式より $f(\bar{\boldsymbol{w}}) \le f(\boldsymbol{w})$ が任意の $\boldsymbol{w}$ に対して成り立つ．したがって $\bar{\boldsymbol{w}}$ は最適である． ∎

なお，証明は省略するが，以下の条件 (i) または (ii) が成り立つとき，上記の $\bar{\boldsymbol{w}}$ は唯一の最適解となる．

(i) すべての $i = 1, \dots, k$ に対して $g_i(\boldsymbol{w}) < 0$ となるような $\boldsymbol{w} \in \Omega$ が存在する．

(ii) 任意の $\boldsymbol{\alpha} \ne \boldsymbol{0}$ に対し，$\sum_{i=1}^{k} \alpha_i g_i(\boldsymbol{w}) \le 0$ となるような $\boldsymbol{w} \in \Omega$ が存在する．

3 KKT-Gap

(3) は任意の関数 f, g_i に対して $L(\bar{\boldsymbol{w}}, \bar{\boldsymbol{\alpha}})$ が鞍点 (Saddle point) となる条件であったが，f, g_i が微分可能な場合にはより使いやすい形で表現することが可能である．

定理 2: $f(\boldsymbol{w}), g_i(\boldsymbol{w})$ は $\boldsymbol{w} = \bar{\boldsymbol{w}}$ において (下に) 凸であり，微分可能とする．以下の条件をみたす $\bar{\alpha}_i \ (i = 1, \dots, k)$ が存在するならば，$\bar{\boldsymbol{w}}$ は最適化問題 (1) の解である．

$$\frac{\partial L(\bar{\boldsymbol{w}}, \bar{\boldsymbol{\alpha}})}{\partial \boldsymbol{w}} = \frac{\partial f(\bar{\boldsymbol{w}})}{\partial \boldsymbol{w}} + \sum_{i=1}^{k} \bar{\alpha}_i \frac{\partial g_i(\bar{\boldsymbol{w}})}{\partial \boldsymbol{w}} = \boldsymbol{0}, \quad \text{(Saddle Point in } \bar{\boldsymbol{w}}\text{)}, \tag{6}$$

$$\frac{\partial L(\bar{\boldsymbol{w}}, \bar{\boldsymbol{\alpha}})}{\partial \alpha_i} = g_i(\bar{\boldsymbol{w}}) \le 0, \text{(Saddle Point in } \bar{\boldsymbol{\alpha}}\text{)}, \tag{7}$$

$$\sum_{i=1}^{k} \bar{\alpha}_i g_i(\bar{\boldsymbol{w}}) = 0, \quad \text{(Vanishing KKT-Gap)}. \tag{8}$$

(証明) 凸関数の性質[1]を用いて $f(\boldsymbol{w}) - f(\bar{\boldsymbol{w}}) \ge 0$ が任意の $\boldsymbol{w} \in \Omega$ について成立することを示す．

$$\begin{aligned} f(\boldsymbol{w}) - f(\bar{\boldsymbol{w}}) &\ge \frac{\partial f(\bar{\boldsymbol{w}})}{\partial \boldsymbol{w}}^T (\boldsymbol{w} - \bar{\boldsymbol{w}}) \quad \text{(Convexity)} \\ &= -\sum_{i=1}^{k} \bar{\alpha}_i \frac{\partial g_i(\bar{\boldsymbol{w}})}{\partial \boldsymbol{w}}^T (\boldsymbol{w} - \bar{\boldsymbol{w}}) \quad \text{(from (6))} \\ &\ge -\sum_{i=1}^{k} \bar{\alpha}_i (g_i(\boldsymbol{w}) - g_i(\bar{\boldsymbol{w}})) \quad \text{(Convexity)} \\ &= -\sum_{i=1}^{k} \bar{\alpha}_i g_i(\boldsymbol{w}) \quad \text{(from (8))} \\ &\ge 0 \quad \text{(from (7))} \end{aligned} \tag{9}$$

したがって $\bar{\boldsymbol{w}}$ は f の最小値を達成する． ∎

4 Dual Form

これまでは $\boldsymbol{w}$ に関する最小化を考えてきたが，Lagrangian $L(x, \boldsymbol{\alpha})$ に関する鞍点条件を考えると，これは $\boldsymbol{\alpha}$ に関する $L(x, \boldsymbol{\alpha})$ の最大化問題ともとらえることができる．つまり

$$\begin{array}{ll} \text{maximize} & L(\boldsymbol{w}, \boldsymbol{\alpha}) = f(\boldsymbol{w}) + \sum_{i=1}^{k} \alpha_i g_i(\boldsymbol{w}), \\ \text{where} & (\boldsymbol{w}, \boldsymbol{\alpha}) \in W \end{array} \tag{10}$$

ただし W は次式で定義される集合である．

$$W := \left\{ (\boldsymbol{w}, \boldsymbol{\alpha}) \;\middle|\; \begin{array}{l} \boldsymbol{w} \in \Omega, \alpha_i \ge 0 \quad (i = 1, \dots k) \\ \text{and} \quad \frac{\partial L(\boldsymbol{w}, \boldsymbol{\alpha})}{\partial \boldsymbol{w}} = 0 \end{array} \right\} \tag{11}$$

[1] $f(\boldsymbol{w})$ が下に凸であるから Taylor 展開して 2 階微分が正定行列．$g_i(\boldsymbol{w})$ が $\boldsymbol{w}$ に関して線形ならばふたつめの不等号は等号．