

Lecture 20. KKT condition cont'd; Lagrange dual.

Proof of $\mu_j \geq 0$ in KKT condition: Prove by contradiction.

Assume $\exists j \in \bar{J}(x^*)$ s.t. $\mu_j^* < 0$. consider all other constraints.

$$\hat{M} = \{x: g(x^*)=0, \forall i \neq j \in \bar{J}(x^*), h_i(x^*)=0\}. \quad \hat{T}(x^*) = T_{x^*} \hat{M}.$$

By regularity of x^* . $\hat{M}$ is locally a submanifold. $\hat{T}(x^*)$ a tangent space.

Note that $\hat{T}(x^*) = \ker \begin{pmatrix} \nabla g_i \\ \nabla h_l \end{pmatrix}_{\substack{1 \leq i \leq m \\ j \neq l \in \bar{J}(x^*)}}. \quad x^* \text{ is regular w.r.t. } \begin{pmatrix} \nabla g_i \\ \nabla h_l \end{pmatrix}_{l \in \bar{J}(x^*)}.$

so $\exists v \in \hat{T}(x^*)$. $\nabla h_j(x^*)^T v \neq 0$. (o.w. $\nabla h_j(x^*) \in \text{span}\{\nabla g_i, \nabla h_l\}_{l \neq j \in \bar{J}})$.

Without loss of generality. further assume $\nabla h_j(x^*)^T v < 0$.

Consider the Lagrange condition: $\nabla f(x^*) + Dg(x^*)^T \lambda^* + D_h(x^*) \mu^* = 0$.

$$\Rightarrow (\nabla f(x^*) + Dg(x^*)^T \lambda^* + \mu_j^* \nabla h_j(x^*) + \sum_{j \neq l \in \bar{J}(x^*)} \mu_l^* \nabla h_l(x^*))^T v = 0.$$

$$(v \in \hat{T}(x^*) = \ker \begin{pmatrix} \nabla g_i \\ \nabla h_l \end{pmatrix}_{\substack{1 \leq i \leq m \\ j \neq l \in \bar{J}}}) \Rightarrow \nabla f(x^*)^T v + \mu_j^* \nabla h_j(x^*)^T v = 0.$$

$$(\mu_j^* < 0 \text{ by assumption}) \Rightarrow \nabla f(x^*)^T v < 0.$$

$$v \in \hat{T}(x^*) \quad \exists \gamma: (-\varepsilon, \varepsilon) \rightarrow \hat{M} \text{ s.t. } \gamma(0) = x^*, \gamma'(0) = v.$$

$$\frac{d}{dt} (f \circ \gamma) \Big|_{t=0} = \nabla f(\gamma(0))^T \gamma'(0) = \nabla f(x^*)^T v. \quad \frac{d}{dt} (h_j \circ \gamma) \Big|_{t=0} = \nabla h_j(x^*)^T v.$$

$$\Rightarrow \exists \delta > 0. (f \circ \gamma)(\delta) < (f \circ \gamma)(0) = f(x^*). \quad \} \Rightarrow \exists x' \text{ feasible.}$$

$$\exists \xi > 0. (h_j \circ \gamma)(\xi) < (h_j \circ \gamma)(0) = h_j(x^*) = 0 \quad \} \quad f(x') < f(x^*) \quad \square.$$

By continuity. $(h_j \circ \gamma)(0) \leq 0 \quad \forall j \notin \bar{J}(x^*)$. $|\theta|$ sufficiently small.

Sufficiency of KKT conditions for convex problems.

Recall that Lagrange condition is sufficient for $\bar{\text{ECP}} \begin{matrix} \min f(x) \\ \text{s.t. } Ax=b. \end{matrix}$

Also recall that complementary slackness is sufficient for LP.

Proof. It suffices to show that $\nabla f(x^*)^T (x - x^*) \geq 0 \quad \forall x \in M$ ^{feasible solution.}

According to KKT. $\nabla f(x^*)^T (x - x^*) = - \sum_{i=1}^m \lambda_i^* \nabla g_i^T (x - x^*) - \sum_{j=1}^k \mu_j^* \nabla h_j^T (x - x^*)$

We claim that $\nabla g_i(x^*)^T (x - x^*) = 0$ and $\nabla h_j(x^*)^T (x - x^*) \leq 0 \quad j \in J(x^*)$.

g_i is affine. $g_i(x^*) = g_i(x) = 0 \Rightarrow \nabla g_i(x^*)^T (x - x^*) = 0 \quad (\mu \in \ker(\nabla g_i))$.

$\forall j \in J(x^*) \quad h_j(x^*) = 0 \quad h_j(x) \leq 0 \Rightarrow \nabla h_j(x^*)^T (x - x^*) \leq 0$ by convexity. $\square$

Another proof. Consider the Lagrangian function $L(x, \lambda, \mu)$.

$L(x, \lambda, \mu) = f(x) + \lambda^T g(x) + \mu^T h(x)$. $\text{KKT} \Rightarrow D_x L(x^*, \lambda^*, \mu^*) = 0$.

Define $\tilde{L}(x) = L(x, \lambda^*, \mu^*) = f(x) + (\lambda^*)^T g(x) + (\mu^*)^T h(x)$. $\nabla \tilde{L}(x^*) = 0$.

Since g_i affine, h_j convex and $\mu_j^* \geq 0 \quad \tilde{L}(x)$ is convex $\Rightarrow x^* = \arg\min \tilde{L}$

Since $\forall j \notin J(x^*) \quad \mu_j^* = 0 \Rightarrow f(x^*) = \tilde{L}(x^*) \leq \tilde{L}(x) \quad \forall x \in M$.

Note that $\forall x \in M \quad \tilde{L}(x) = f(x) + (\mu^*)^T h(x) \leq f(x)$. $\square$

Example. (before proof). $\min x_1^2 + x_2^2 \quad \text{s.t. } x_1 + x_2 = 1 \quad x_2 \leq \alpha$.

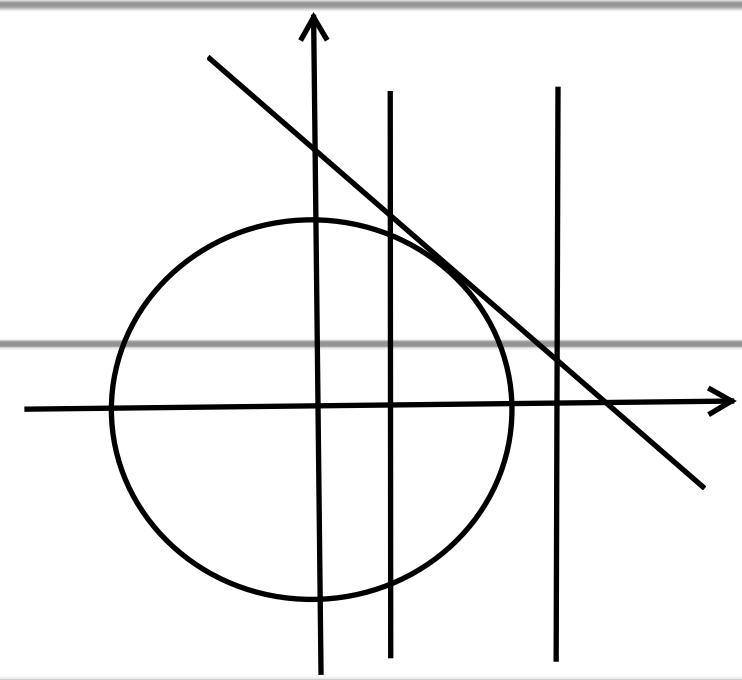
Lagrangian. $L(x_1, x_2, \lambda, \mu) = x_1^2 + x_2^2 + \lambda(1 - x_1 - x_2) + \mu(x_2 - \alpha)$.

KKT condition. $\nabla_x L = 0$ $x_1 + x_2 = 1$ $\mu \geq 0$
 $x_2 - \alpha \leq 0$ $\mu(x_2 - \alpha) = 0$.

$\nabla_x L = 0 \Rightarrow \begin{cases} 2x_1 - \lambda = 0 \\ 2x_2 - \lambda + \mu = 0 \end{cases} \Rightarrow \begin{cases} x_1 = \mu/4 + 1/2 \\ x_2 = -\mu/4 + 1/2 \end{cases} \quad \mu \geq 2 - 4\alpha.$

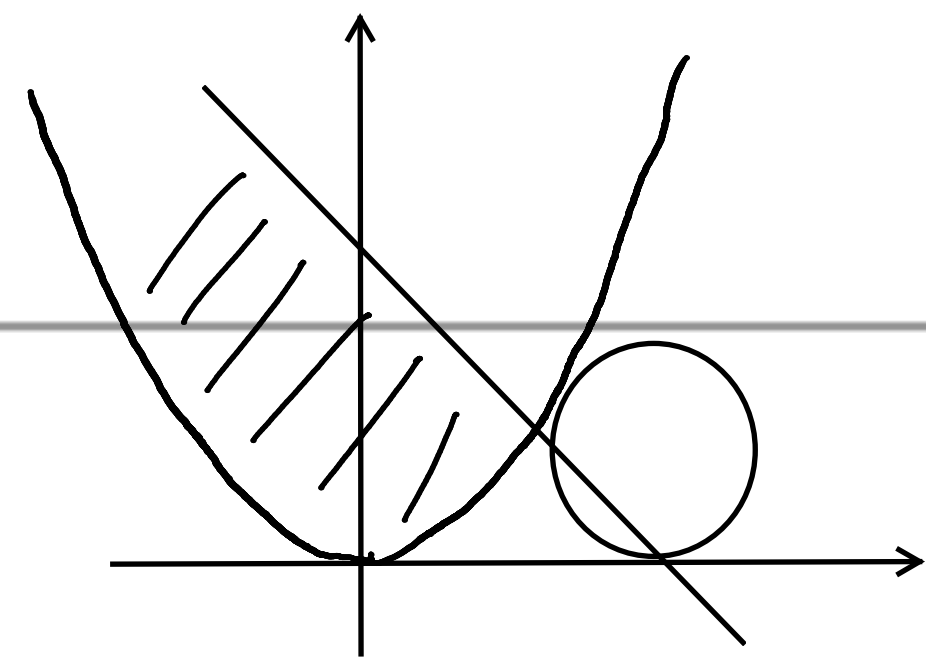
Case 1. $\alpha > \frac{1}{2} \Rightarrow \mu = 0$. $x_1^* = x_2^* = \frac{1}{2}$.

Case 2. $\alpha \leq \frac{1}{2} \Rightarrow \mu = 2 - 4\alpha$. $x_1^* = 1 - \alpha$. $x_2^* = \alpha$.



Example. $\min f(x) = (x_1 - 2)^2 + (x_2 - 1)^2$

s.t. $g_1(x) = x_1^2 - x_2 \leq 0$
 $g_2(x) = x_1 + x_2 - 2 \leq 0$



Lagrangian $L(x, \mu) = (x_1 - 2)^2 + (x_2 - 1)^2 + \mu_1(x_1^2 - x_2) + \mu_2(x_1 + x_2 - 2)$.

$\nabla_x L(x, \mu) = (2(x_1 - 2) + 2\mu_1 x_1 + \mu_2, 2(x_2 - 1) - \mu_1 + \mu_2)^T = (0, 0)^T$

Case 1. both g_1, g_2 are inactive. $\mu_1 = \mu_2 = 0 \Rightarrow \begin{cases} x_1 = 2 \\ x_2 = 1 \end{cases}$ infeasible.

Case 2. g_1 active. g_2 inactive $\Rightarrow \begin{cases} x_2 = x_1^2 \\ \mu_1 = 2(x_1^2 - 1) \end{cases}$ infeasible.

Case 3. g_1 inactive. g_2 active $\Rightarrow \begin{cases} x_1 = 3/2 \\ x_2 = 1/2 \end{cases} \mu_2 = 1$ infeasible.

Case 4. both g_1, g_2 are active. $\Rightarrow \begin{cases} x_1 = 1 \\ x_2 = 1 \end{cases}$ or $\begin{cases} x_1 = -2 \\ x_2 = -2 \end{cases}$ need $\mu \geq 0$ so $x^* = (1, 1)^T$.

Example. (nonregularity). $\min x_1 + x_2$ s.t. $(x_1 - 1)^2 + x_2^2 \leq 1$
 $(x_1 + 1)^2 + x_2^2 \leq 1$.

KKT condition. $\begin{cases} 1 + 2\mu_1(x_1 - 1) + 2\mu_2(x_1 + 1) = 0 \\ 1 + 2\mu_1 x_2 + 2\mu_2 x_2 = 0 \end{cases}$ or verify $\begin{pmatrix} 0 \\ 0 \end{pmatrix}$ directly.

In summary. KKT solvable $\Rightarrow \exists$ feasible optimal solution.

unsolvable $\Leftarrow$ infeasible or unbounded below.

possibly unsolvable but $\exists$ critical optimal solution

critical
optimal (KKT)

Note in the proof of sufficiency. we fix λ^*, μ^* and $\min L(x, \lambda^*, \mu^*)$.

How about other λ, μ ? Let $\phi(\lambda, \mu) = \inf_x L(x, \lambda, \mu)$. $\mu \geq 0$.

Of course $\forall \lambda, \mu$ and feasible x . $L(x, \lambda, \mu) \leq f(x)$. $D \triangleq \text{dom } f \cap \text{dom } g \cap \text{dom } h$.

let $X \triangleq$ feasible set of x . $\phi(\lambda, \mu) = \inf_{x \in D} L(x, \lambda, \mu) \leq \inf_{x \in X} f(x)$.

$\phi(\lambda, \mu)$ gives a lower bound of $\inf_{x \in X} f(x)$. called Lagrange dual function

Lagrange dual problem: $\max \phi(\lambda, \mu) = \inf_{x \in D} L(x, \lambda, \mu)$ s.t. $\mu \geq 0$

Example. $\begin{array}{ll} \max & c^T x \\ \text{s.t.} & A_1 x = b_1 \\ & A_2 x \leq b_2 \end{array}$ $\phi(\lambda, \mu) = \inf_x (\lambda^T A_1 + \mu^T A_2 - c^T)x - \lambda^T b_1 - \mu^T b_2$.

$$\phi(\lambda, \mu) = \begin{cases} -\infty & \text{if } \lambda^T A_1 + \mu^T A_2 - c^T \neq 0 \\ -\lambda^T b_1 - \mu^T b_2 & \text{o.w.} \end{cases} \quad \begin{array}{ll} \max & -\lambda^T b_1 - \mu^T b_2 \\ \text{s.t.} & \lambda^T A_1 + \mu^T A_2 = c^T, \mu \geq 0. \end{array}$$

Example. $\begin{array}{ll} \min & x^T x \\ \text{s.t.} & Ax = b, x \geq 0. \end{array}$ $\phi(\lambda, \mu) = \inf_x x^T x + (\lambda^T A - \mu^T)x - \lambda^T b$.

$$\nabla_x L(x, \lambda, \mu) = 0 \Rightarrow x = \frac{1}{2}(\mu - A^T \lambda). \quad \phi(\lambda, \mu) = -\frac{1}{4} \|\mu - A^T \lambda\|^2 - \lambda^T b.$$

- $\phi(\lambda, \mu)$ is convex (pointwise sup of affine). so dual is a COP.

Duality: $f^* = \min_{(\text{or inf})} f(x)$ s.t. $\begin{array}{l} g(x) = 0 \\ h(x) \leq 0 \end{array}$ $= \inf_{x \in X} f(x)$ convex optimization.

note that $\forall x \in X, \lambda, \mu \geq 0, f(x) \geq L(x, \lambda, \mu)$ and $f(x) = L(x, 0, 0)$

$$\text{so } f^* = \inf_{x \in X} f(x) = \inf_{x \in X} \sup_{\lambda, \mu \geq 0} L(x, \lambda, \mu) \geq \inf_x \sup_{\lambda, \mu \geq 0} L(x, \lambda, \mu).$$

(in fact. $\forall x \notin X, \sup_{\lambda, \mu \geq 0} L(x, \lambda, \mu) = \infty$, so equality always holds.)

$$\phi^* = \sup_{\lambda, \mu \geq 0} \phi(\lambda, \mu) = \sup_{\lambda, \mu \geq 0} \inf_x L(x, \mu, \lambda) \leq \sup_{\lambda, \mu \geq 0} L(x, \lambda, \mu), \forall x.$$

$$\Rightarrow \phi^* \leq \inf_x \sup_{\lambda, \mu \geq 0} L(x, \mu, \lambda) = f^* \quad (\text{weak duality}).$$

Max-min inequality: $\sup_{x \in X} \inf_{y \in Y} f(x, y) \leq \inf_{y \in Y} \sup_{x \in X} f(x, y).$

Geometric interpretation of duality.

$$\begin{aligned} \min \quad & x^4 - 50x^2 + 100x. \\ \text{s.t.} \quad & x \geq -5/2. \quad (h(x) = -5/2 - x). \end{aligned}$$

consider the epigraph of $(h(x), f(x))$.

$$C = \{(p, t) : \exists x, h(x) \leq p, f(x) \leq t\} \quad l: f + \mu h = \phi, \quad \mu \geq 0.$$

l is a supporting hyperplane of C .

f^* : C intersects axis
 ϕ^* : l intersects axis

