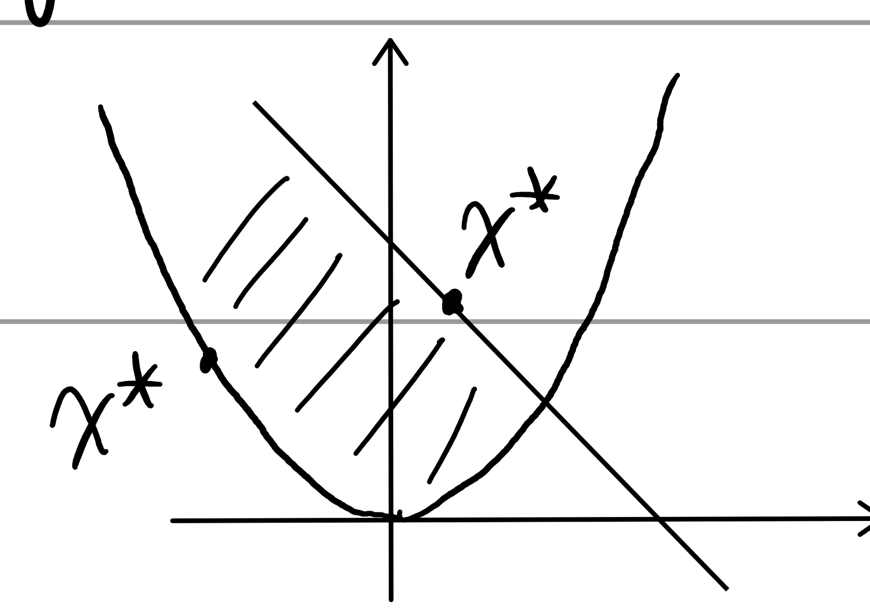
Lecture 20. Lagrange duality.

In KKT conditions, why can we ignore inactive constraints?

Both Lagrange multiplier conditions and KKT conditions hold for local minimum. Inactive constraints are always satisfied in $B(x^*, \varepsilon)$

Example subject to $g_1(x) = x_1^2 - x_2 \leq 0$
$g_2(x) = x_1 + x_2 - 2 \leq 0$.



if $x^* = \begin{pmatrix} 1/2 \\ 3/2 \end{pmatrix}$ $g_1(x)$ holds in neighborhood.

Remark. KKT is only necessary for regular points. Critical optimal exists.

For LP. KKT condition $\Longleftrightarrow$ duality + complementary slackness.

KKT multipliers are optimal to another optimization. NOT a coincidence.

Lagrangian function: $L(x, \lambda, \mu) = f(x) + \sum_{i=1}^{m} \lambda_i g_i(x) + \sum_{j=1}^{\ell} \mu_j h_j(x)$.

Domain: $x \in D = \operatorname{dom} f \cap \operatorname{dom} g_1 \cap \cdots \cap \operatorname{dom} h_\ell$. $\lambda \in \mathbb{R}^m$. $\mu \in \mathbb{R}^m_{\geq 0}$.

Recall Lagrange multiplier method: $\nabla L(x, \lambda) = 0 \Rightarrow (x, \lambda)$ saddle point.

for convex problems. $f^* = \min_{Ax=b} f(x) = \max_{\lambda} \min_{x} L(x, \lambda)$

Lagrange dual function: $\phi(\lambda, \mu) = \inf_{x \in D} L(x, \lambda, \mu)$.

Lagrange dual problem: $\max \phi(\lambda, \mu)$. subject to $\mu \geq 0$. $\begin{pmatrix} \lambda \in \mathbb{R}^m \\ \mu \in \mathbb{R}^\ell_{\geq 0} \end{pmatrix}$

Example. $\min x_1^2 + x_2^2$ subject to $x_1 + x_2 \leq -1$

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

dual function: $\phi(\mu) = \min_{x_1, x_2} x_1^2 + x_2^2 + \mu(x_1 + x_2 + 1) = -\frac{1}{2}\mu^2 + \mu$.

dual problem: $\max \phi(\mu)$ s.t. $\mu \geq 0 = \max_{\mu \geq 0} -\frac{1}{2}\mu^2 + \mu = \frac{1}{2}$.

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

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

$\phi(\mu_1, \mu_2) = \begin{cases} -\infty & \text{if } \mu_1 + \mu_2 \leq 0 \\ \frac{-2(\mu_1 - \mu_2)^2 + 2(\mu_1 - \mu_2) - 1}{2(\mu_1 + \mu_2)} & \end{cases}$ dual problem: $\sup_{\mu_1, \mu_2 \geq 0} \phi = 0$.

Example. $\min_{x_2 > 0} e^{-x_1}$ s.t. $x_1^2 / x_2 \leq 0$. $D = \{(x_1, x_2) : x_2 > 0\}$.

$\phi(\mu) = \inf_{x_2 > 0} (e^{-x_1} + \mu x_1^2 / x_2) = \begin{cases} 0 & \mu \geq 0 \\ -\infty & \text{otherwise.} \end{cases}$

Example. $\max c^T x$. s.t. $Ax_1 = b_1$. $Ax_2 \leq b_2$.

$\phi(\lambda, \mu) = \inf_x (\lambda^T A_1 + \mu^T A_2 - c^T) x - \lambda^T b_1 - \mu^T b_2 = \begin{cases} -\infty & \lambda^T A_1 + \mu^T A_2 \neq c^T \\ -\lambda^T b_1 - \mu^T b_2 & \end{cases}$

dual problem. $\sup_{\mu \geq 0} \phi(\lambda, \mu)$: $\min \lambda^T b_1 + \mu^T b_2$. s.t. $\begin{array}{c} \lambda^T A_1 + \mu^T A_2 = c^T \\ \mu \geq 0. \end{array}$

Remark. Lagrange dual problem is always a convex optimization.

for any fixed x. $L(x, \lambda, \mu)$ is an affine function of λ, μ

So $\phi(\lambda, \mu)$ is a pointwise minimum of a family of affine functions.

Question: If x^* has KKT multipliers (λ^*, μ^*). is it optimal for dual ?

To answer this question, we consider the relationship between optimal values.

Definition. Ω: the feasible set of x. $f^* = \inf_{x\in\Omega} f(x)$. $\phi^* = \sup_{\mu\geq 0} \phi(\lambda, \mu)$.

Lemma. (Weak duality theorem) $f^* \geq \phi^*$.

Remark. why not "=" (recall equality constrained problems).

KKT condition rewritten (using Lagrangian function).

1. $\nabla f(x^*) + \sum_{i=1}^{m} \lambda_i^* \nabla g_i(x^*) + \sum_{j=1}^{\ell} \mu_j^* \nabla h_j(x^*) = 0$. $\quad \nabla_x L(x^*, \lambda^*, \mu^*) = 0$.

2. x^* is feasible. $\quad \nabla_\lambda L(x^*, \lambda^*, \mu^*) = 0$, $\nabla_\mu L(x^*, \lambda^*, \mu^*) \leq 0$

3. $\mu_j^* \geq 0$. $\mu_j^* h_j(x^*) = 0$. $\quad \nabla_\mu L(x^*, \lambda^*, \mu^*)^T \mu^* = 0$.

We only have $\nabla_{x,\lambda} L(x^*, \lambda^*, \mu^*) = 0$. $\nabla_\mu L(x^*, \lambda^*, \mu^*) \leq 0$ now.

Proof. For any feasible solution $x \in \Omega$. $g_i(x) = 0$. $h_j(x) \leq 0$.

So $f(x) = L(x, \lambda, 0) \geq L(x, \lambda, \mu)$ for all $\mu \geq 0$.

Thus $f(x) = \sup_{\mu\geq 0} L(x, \lambda, \mu)$. $f^* = \inf_{x\in\Omega} \sup_{\mu\geq 0} L(x, \lambda, \mu)$

For any λ and $\mu \geq 0$. $\phi(\lambda, \mu) = \inf_{x\in D} L(x, \lambda, \mu) \leq \inf_{x\in\Omega} L(x, \lambda, \mu)$

So $\phi(\lambda, \mu) \leq \inf_{x\in\Omega} \sup_{\mu\geq 0} L(x, \lambda, \mu) = f^*$. $\square$

Remark: Recall minimax theorem. $\sup_{y\in Y} \inf_{x\in X} F(x, y) \leq \inf_{x\in X} \sup_{y\in Y} F(x, y)$.

Definition: Duality gap $\triangleq f^* - \phi^*$. Strong duality: $f^* = \phi^*$.

Remark. Strong duality may not hold. ($\min_{x_2 > 0} e^{-x_1}$. s.t. $x_1^2/x_2 \leq 0$).

Theorem. If x^* has KKT multipliers λ^*, μ^*, then strong duality holds.

Moreover, (λ^*, μ^*) is an optimal solution to the dual problem.

Proof. It suffices to show $\phi(\lambda^*, \mu^*) = f(x^*)$, since for convex

problems, x^* has KKT multipliers $\lambda^*, \mu^* \Rightarrow f(x^*) = f^*$.

① $L(x^*, \lambda^*, \mu^*) = f(x^*)$, since $g_i(x^*)=0$, $\mu_j^* h_j(x^*) = 0$.

② fix λ^*, μ^*, let $\hat{L}(x) = L(x, \lambda^*, \mu^*)$. $(\lambda^*, \mu^*) = \min \hat{L}(x)$

$\hat{L}(x)$ is a convex function and $\nabla \hat{L}(x^*) = \nabla_x L(x^*, \lambda^*, \mu^*) = 0$. $\square$

Question: If x^* optimal to primal, (λ^*, μ^*) optimal to dual, does

x^* have KKT multipliers (λ^*, μ^*)? No.

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

Example. $\min\limits_{x_2>0}\ e^{-x_1}$ s.t. $x_1^2/x_2 \le 0$. $f^* = 1$. $\phi^* = 0$.

KKT conditions: $\begin{cases} -e^{x_1} + \mu \cdot 2x_1/x_2 = 0 \\ -\mu x_1^2/x_2^2 = 0 \end{cases}$ $\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$

Theorem. If x^* is a finite optimal solution to the primal, (λ^*, μ^*)

is a finite optimal solution to the dual, and strong duality holds,

then (λ^*, μ^*) is KKT multipliers of x^*.

Proof: By strong duality, $f^* = \phi^* = \phi(\lambda^*, \mu^*) = \inf\limits_x L(x, \lambda^*, \mu^*)$.

Note that x^* is feasible so $g_i(x^*) = 0$, $h_j(x^*) \le 0$. Thus

$f^* = f(x^*) \ge L(x^*, \lambda^*, \mu^*)$ (since $\mu^* \ge 0$) $\ge \inf\limits_x L(x, \lambda^*, \mu^*)$.

It yields that $f^* = f(x^*) = L(x^*, \lambda^*, \mu^*) = \inf\limits_x L(x, \lambda^*, \mu^*)$.

(λ^*, μ^*) is feasible for dual $\Rightarrow \mu^* \ge 0$.

$f(x^*) = L(x^*, \lambda^*, \mu^*) \Rightarrow \forall j,\ \mu_j^* h_j(x^*) = 0$.

$L(x^*, \lambda^*, \mu^*) = \inf\limits_x L(x, \lambda^*, \mu^*) \Rightarrow \nabla_x L(x^*, \lambda^*, \mu^*) = 0$. $\square$

Summary. 1. x^* optimal + regularity $\Rightarrow$ KKT multipliers (all).

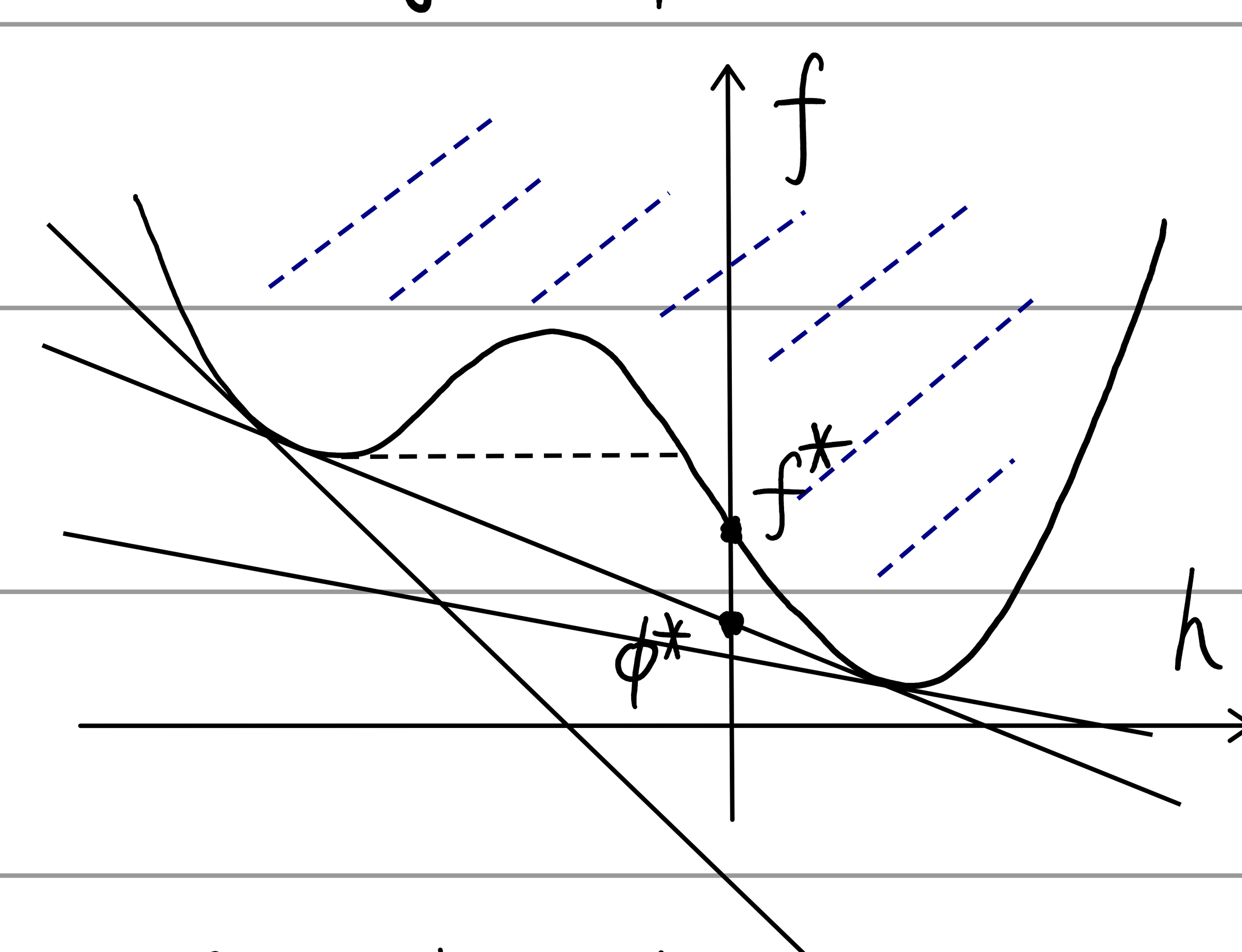
2. weak duality $f^* \ge \phi^*$ (all).

3. x^* has multipliers λ^*, μ^* $\Rightarrow \begin{cases} x^* \text{ optimal for primal} \\ (\lambda^*, \mu^*) \text{ optimal for dual} \\ \text{strong duality holds} \end{cases}$ (convex)

4. $\left.\begin{array}{l} x^*, (\lambda^*, \mu^*) \text{ finite optimal} \\ \text{strong duality holds} \end{array}\right\} \Rightarrow x^*$ has multipliers λ^*, μ^*. (all).

Question: Solving KKT multipliers is complicated, any simple conditions?

Geometric interpretation of duality.



$\min \quad x^4 - 50x^2 + 100x$.

$\text{s.t.} \quad x \ge -5/2$. $(h(x) = -5/2 - x)$.

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

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

l is a supporting hyperplane of C.

f^*: C intersects axis

ϕ^*: l intersects axis

Slater's condition for convex optimization:

If $x \in \text{relint}(D)$ such that $\begin{cases} g_i(x) = 0 \\ h_j(x) < 0 \end{cases}$. then strong duality holds.