(Review)

KKT conditions provide a way to certify optimality.
Lagrange turns them into an optimization problem.

For convex opt. problem (p^* = value of primal)

$$p^* = \min_x f(x) \quad \text{s.t.} \quad g_i(x) \le 0 \quad 1 \le i \le k, \qquad (1)$$

KKT conditions are

1. $g_i(x^*) \le 0$ for $1 \le i \le k$
2. $\lambda^*(i) \ge 0$ for $1 \le i \le k$
3. $\lambda^*(i) g_i(x^*) = 0$ for $1 \le i \le k$
4. $\nabla f(x^*) + \sum_i \lambda^*(i) \nabla g_i(x^*) = 0$

These can help us solve (1):
trade the min, which is hard to check,
for conditions that can test.

<u>Thm</u>! If $f, g_1 \dots g_k$ are convex and differentiable and x^*, λ^* satisfy 1-4,
then x^* is optimal. If the g_i are linear or
if exists strictly feasible x_0 ($g_i(x_0) < 0 \ \forall i$)
then there exist x^* and λ^* that satisfy 1-4.

Question: could we replace $g_1 \dots g_k$
with $G(x) = \max_i g_i(x)$?

- $G(x) \le 0 \iff g_i(x) \le 0 \;\; \forall i$ ✓
- the max of convex functions is convex ✓
- But, max is NOT differentiable X

  consider $\max(x_1, x_2)$, at $(0,0)$

  and try to compute derivative in direction $t = (1,0)$

$$\max(x + \varepsilon t) = \begin{cases} \varepsilon & \varepsilon > 0 \\ 0 & \varepsilon < 0 \end{cases}$$

  so, can not approximate by a linear function.

- And, trading $\lambda_1, \dots \lambda_k$ for one parameter would be less informative anyway.

lesson: reformulation can help -
is good to replace G by $g_1, \dots, g_k$

Lagrange: create function in λ that gives a lower bound.

Define $L(x, \lambda) = f(x) + \sum_i \lambda(i) g_i(x)$

And $q(\lambda) = \inf_x L(x, \lambda)$

q is the Lagrange Dual, and has dual value

$d_* = \max_\lambda q(\lambda)$ s.t. $\lambda \geq 0$

Surprisingly, can compute $q(\lambda)$ for many nice problems

issue is "inf"

helped by lack of constraints. and, can be $-\infty$

<u>Lem 1</u> $q(\lambda) \leq f(x)$ for all $\lambda \geq 0$ and feasible x.

So, $d^* \leq p^*$. (weak duality)

<u>proof</u> For feasible x, $g_i(x) \leq 0$.

So, $\lambda \geq 0 \Rightarrow \sum_i \lambda(i) g_i(x) \leq 0$

This implies $L(x, \lambda) = f(x) + \sum_i \lambda(i) g_i(x) \leq f(x)$

On the other hand, $q(\lambda) = \inf_{x_0} L(x_0, \lambda) \leq L(x, \lambda)$,

because $\inf_{x_0}$ gives the minimum.

Also note for every x $L(x, \lambda)$ is linear in λ.
As $q(\lambda)$ is the inf of linear functions,
it is concave.
So, maximizing q is reasonable.

<u>Thm 2</u> (Strong duality)
If $f, g_1, \ldots, g_k$ are differentiable and convex,
and either $g_1 \ldots g_k$ are linear or strictly feasible
then $d^* = p^*$

<u>proof</u> By KKT theorem,
$\exists\ x^*$ and λ^* s.t. x^* is feasible, $\lambda^* \geq 0$ s.t. ...
let $h(x) = L(x, \lambda^*) = f(x) + \sum \lambda^*(i) g_i(x)$
as $\lambda^* \geq 0$, $h(x)$ is convex.
KKT 4 $\Rightarrow$ $\nabla h(x^*) = 0$
so x^* is a global minimizer.

And, $q(\lambda^*) = \inf\limits_x L(x, \lambda^*) = \inf\limits_x h(x) = h(x^*)$

KKT 3 $\Rightarrow$ $h(x^*) = f(x^*) + \sum\limits_i \lambda^*(i) g_i(x^*) = f(x^*)$

So, $q(\lambda^*) = f(x^*)$ $\Rightarrow$ $d^* \geq p^*$
Combined with lem 1 $\Rightarrow$ $d^* = p^*$

For general x, λ $f(x) - q(\lambda) = p - d$ is the <u>duality gap</u>

Before doing examples, let's generalize to add equality constraints.

Can write $h(x) = 0$ $(\in \mathbb{R})$ as $h(x) \le 0$ and $-h(x) \le 0$

Each gets own lagrange multiplier, λ_+ and λ_-

So, add in

$$\lambda_+ h(x) + \lambda_- (-h(x)) = (\lambda_+ - \lambda_-) h(x)$$

For $\lambda_+, \lambda_- \ge 0$, can write any real as $\lambda_+ - \lambda_-$

So, replace by ν.

For $\min_x f(x)$ s.t. $g_i(x) \le 0 \quad 1 \le i \le k$

$h_i(x) = 0 \quad 1 \le i \le j$

$$L(x, \lambda, \nu) = f(x) + \sum_{i=1}^{k} \lambda(i) g_i(x) + \sum_{i=1}^{j} \nu(i) h_i(x)$$

$$q(\lambda, \nu) = \inf_x L(x, \lambda, \nu)$$

$$d^* = \max q(\lambda, \nu) \quad \lambda \in \mathbb{R}_+^k \quad \nu \in \mathbb{R}^j$$

Examples

LP $\max c^T x$ s.t. $a_i^T x \le b_i$

rewrite (by changing c) as

$\min c^T x$ s.t. $g_i(x) \le 0$, $g_i(x) = a_i^T x - b_i$

$$L(x, \lambda) = c^T x + \sum_i \lambda_i (a_i^T x - b_i)$$

$$= c^T x + \lambda^T A x - \lambda^T b$$
$$= (c^T + \lambda^T A) x - \lambda^T b$$

$$q(\lambda) = \inf_x (c^T + \lambda^T A) x - \lambda^T b$$

if $c^T + \lambda^T A \neq 0$, $\inf_x (c^T + \lambda^T A) x = -\infty$

so, $q(\lambda) = \begin{cases} -\infty & \text{if } c^T + \lambda^T A \neq 0 \\ -\lambda^T b & \text{o.w.} \end{cases}$

$$\max_{\lambda \ge 0} q(\lambda) = \max_{\lambda \ge 0} -\lambda^T b \quad \text{s.t. } A^T \lambda = -c$$

Is same dual as derived before
Not a new proof, as KKT proof relies on LP duality

Other forms of LP: $\min c^T x$ s.t. $Ax = b$, $x \geq 0$

$g_i(x) = -x_i$

$$L(x, \lambda, v) = c^T x + \sum_{i=1}^{n} \lambda(i)\left(-x(i)\right) + \sum_{i=1}^{k} v(i)\left(a_i^T x - b_i\right)$$

$$= c^T x - \lambda^T x + v^T(Ax - b)$$

$$= \left(c - \lambda + A^T v\right)^T x - v^T b$$

$q(\lambda, v) = -\infty$ unless $c - \lambda + A^T v = 0$

can eliminate $\lambda \geq 0$: $\exists \lambda \geq 0$ s.t. $c + A^T v = \lambda$

iff $c + A^T v \geq 0$

so, $\max\limits_{\lambda, v} q(\lambda, v) = \max\limits_{v} -b^T v$ s.t. $A^T v + c \geq 0$

lowest norm point on hyperplane

$$\min \|x\|_2^2 \quad \text{s.t.} \quad Ax = b$$

$$L(x, \nu) = \|x\|_2^2 + \nu^T(Ax - b) = \|x\|_2^2 + \nu^T A x - \nu^T b$$

$g(\nu) = \inf_x L(x, \nu)$ find this by setting $\nabla_x = 0$

$\nabla_x L(x, \nu) = 2x + A^T \nu$ so, $x = -\frac{1}{2} A^T \nu$

which gives $\frac{1}{4} \nu^T A A^T \nu - \frac{1}{2} \nu^T A A^T \nu - \nu^T b$

$$= \frac{1}{4} \nu^T A A^T \nu - \nu^T b$$

every ν gives a lower bound

For arbitrary norms $\min \|x\|$ s.t. $Ax=b$

Need notion of a dual norm:

$$\|y\|_* = \max_x x^T y \text{ s.t. } \|x\| \le 1$$

(will achieve with $\|x\| = 1$)

Examples dual of $\|\cdot\|_2$ is $\|\cdot\|_2$

because Cauchy-Schwartz $x^T y \le \|x\|_2 \|y\|_2 \le \|y\|_2$

with equality only when $y = \lambda x$, $\lambda > 0$

so, set $x = y/\|y\|_2$ to get

$x^T y = y^T y/\|y\|_2 = \|y\|_2$

dual of $\|\cdot\|_\infty$ is $\|\cdot\|_1$

given y, set $x = \begin{cases} 1 & y(i) \ge 0 \\ -1 & y(i) \le 0 \end{cases}$ so $\|x\|_\infty = 1$

and $x^T y = \sum_i x(i) y(i) = \sum_i |y(i)| = \|y\|_1$

In finite dimensions, $\|\cdot\|_{**} = \|\cdot\|$

so dual of $\|\cdot\|_1 = \|\cdot\|_\infty$

For $\frac{1}{p}+\frac{1}{q}=1$, $\|\cdot\|_p$ and $\|\cdot\|_q$ are dual,
where $\|x\|_p = \left(\sum |x(i)|^p\right)^{1/p}$

Follows from Hölder's Inequality
$x^T y \le \|x\|_p \|y\|_q$, with equality for positive x, y
when $y(i)^q = x(i)^p$

Back to $\min \|x\|$ s.t. $Ax=b$
$q(v) = \inf_x \|x\| - v^T A x - v^T b$

if $\|v^T A\|_* \le 1$, then $\|x\| \ge v^T A x$
and so $\inf_x \|x\| - v^T A x = 0$

if $\|v^T A\|_* > 1$, $\exists\, u$ s.t. $\|u\| = \underline{1}$,
$v^T A u = \|v^T A\|_* > 1$
considering $x = cu$ $c \to \infty$ shows
$\inf_x \|x\| - v^T A x = -\infty$

So, $q(v) = \begin{cases} -b^T v & \text{if } \|v^T A\|_* \le 1 \\ -\infty & \text{o.w.} \end{cases}$

Dual is $\max_v -b^T v$ s.t. $\|A^T v\|_* \le 1$

Generalized inequalities and Cones.

Issue: not all convex sets have simple description as $g_i(x) \le 0$ for differentiable convex g_i

Consider positive semidefinite matrices: M s.t. $x^T M x \ge 0$, $\forall x$.

S^n = n-by-n symmetric, S^n_+ = n-by-n psd.

Use ~~fact~~ is a proper cone.

K is a <u>cone</u> if $x \in K \Rightarrow tx \in K$ $\forall t > 0$

is proper if

a. is convex

b. is closed

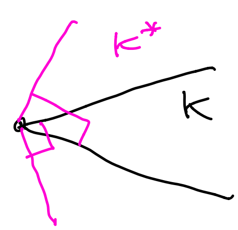
c. solid - has an interior

d. pointed: $x \in K$, $x \ne 0 \Rightarrow -x \notin K$

The dual cone is $K^* = \{x : x^T y \ge 0, \ \forall y \in K\}$

dual of $\mathbb{R}^n_+$ is $\mathbb{R}^n_+$

visualize



In finite dimensions, $K^{**} = K$

The dual cone of S_+^n is S_+^n

inner product of matrices X, Y obtain by writing as vectors. Get $\mathrm{Trace}(X^T Y)$

Generalized inequalities:

$$X \preceq_K Y \quad \text{if} \quad Y - X \in K$$

$$0 \preceq_K X \quad \text{iff} \quad X \in K \quad \text{iff} \quad -X \preceq_K 0$$

So, can write convex programs like

$$\min_x f(x) \quad \text{s.t.} \quad g_i(x) \le 0 \quad 1 \le i \le d$$
$$\text{and} \quad -X \preceq_{K_i} 0 \quad 1 \le i \le c$$

The Lagrange dual is

$$L(x, \lambda_0, \lambda_1, \dots, \lambda_j) = f(x) + \sum_{i=1}^{d} \lambda_0(i) g_i(x) + \sum_{i=1}^{c} \lambda_i^T X$$

All the same stuff holds.

Let's us handle semidefinite programming problems, like

$$\min_M \mathrm{Tr}(F^T M) \quad \text{s.t.} \quad M \in S_+^n$$
$$g_i(M) \le 0 \quad \text{for} \quad 1 \le i \le d.$$

proof S_+^n is self-dual

That is $\mathrm{Tr}(X^T Y) \geq 0$ for all $X \in S_+^n$ iff $Y \in S_+^n$

1. if $Y \notin S_+^n$, $\exists x$ s.t. $x^T Y x < 0$
   let $X = xx^T$
   $\mathrm{Tr}(X^T Y) = \mathrm{Tr}(xx^T Y) = \mathrm{Tr}(x^T Y x) = x^T Y x < 0$

2. If $Y \in S_n^+$ and $X \in S_n^+$, write $X = \sum_i x_i x_i^T$
   by $X = \sum_i \lambda_i v_i v_i^T$, $x_i = \sqrt{\lambda_i} v_i$
   good because $\lambda_i \geq 0$

So, $\mathrm{Tr}(X^T Y) = \mathrm{Tr}(\sum_i x_i x_i^T Y) = \sum_i \mathrm{Tr}(x_i x_i^T Y) \geq 0$