Lecture ~~[illegible]~~ Tuesday, March 17.

Today

- Dual feasible solutions as optimality "certificates"
- Complementary Slackness.
- KKT conditions (Karush-Kuhn-Tucker)
- Water filling example.

Primal

$$p^* = \min f_0(x) \quad = \min_{x} \max_{\lambda, \nu} L(x, \lambda, \nu)$$
$$\text{s.t. } f_i(x) \le 0 \quad i = 1, \ldots, m$$
$$h_i(x) = 0$$

Dual:

$$d^* = \max g(\lambda, \nu) \quad = \max_{\lambda, \nu} \min_{x} L(x, \lambda, \nu)$$
$$\text{s.t. } \lambda \succeq 0$$

Example: Dual of an LP.

$$p^* = \text{minimize } \vec{c}^T \vec{x}$$
$$\text{s.t. } A\vec{x} \preceq \vec{b}$$
$$(A\vec{x} - \vec{b}) \preceq 0$$

$$d^* = \text{maximize } -\vec{b}^T \vec{\lambda}$$
$$\text{s.t. } A^T\vec{\lambda} + \vec{c} = 0$$
$$\vec{\lambda} \succeq 0$$

$$L(\vec{x}, \vec{\lambda}) = c^T \vec{x} + \lambda^T (A\vec{x} - \vec{b})$$
$$= -b^T\lambda + (A^T\lambda + c)^T \vec{x}$$

$$g(\lambda) = \min_x L(x, \lambda) = \begin{cases} -b^T\lambda & \text{if } (A^T\lambda + c) = 0 \\ -\infty & \text{otherwise} \end{cases}$$

So why do we care about the dual?

"Certificate" function. (B&V. 5.5.1)

If I can find (λ_1, ν_1) some λ_1 and ν_1 that are dual feasible, i.e. satisfy the constraints of the dual, then I can say $p^* \geq g(\lambda_1, \nu_1)$.

If x_1 is a primal feasible point, then I can say:

$$f_0(x_1) - p^* \leq \underset{\substack{\uparrow \\ \text{any primal feasible point}}}{f_0(x_1)} - g(\underset{\substack{\uparrow \\ \text{any dual}}}{\lambda_1}, \underset{\substack{\uparrow \\ \text{feasible points.}}}{\nu_1})$$

If $f_0(x_1)$ is ε close to any dual feasible point, then it must be ε close to the primal optimal p^* itself.

→ This is very useful when thinking of stopping criteria for different algorithms. So far we have seen Gradient Descent as one example of such a search algorithm - if you also have dual information you can know how close to the optimal you are.

Complementary Slackness (B & V 5.5.2)

Now instead of what happen at *any* primal feasible and dual feasible point, let us consider the primal optimal $(\vec{x}^*)$ and the dual optimal $(\vec{\lambda}^*, \vec{\nu}^*)$ points.

Assume strong duality holds, so $p^* = d^*$

$p^* = f_0(\vec{x}^*)$ $\qquad$ $d^* = g(\vec{\lambda}^*, \vec{\nu}^*)$

$p^* = d^*$ (strong duality) $\Rightarrow$

$$f_0(\vec{x}^*) \overset{(1)}{=} g(\vec{\lambda}^*, \vec{\nu}^*)$$

(Definition of dual)
$$\overset{(2)}{=} \min_{\vec{x}} \underbrace{\left(f_0(\vec{x}) + \sum_{i=1}^{m} \lambda_i^* f_i(\vec{x}) + \sum_{i=1}^{p} \nu_i^* h_i(\vec{x})\right)}_{L(\vec{x}, \vec{\lambda}^*, \vec{\nu}^*)}$$

$L(\vec{x}^*, \vec{\lambda}^*, \vec{\nu}^*) \leq L(\vec{x}, \vec{\lambda}^*, \vec{\nu}^*)$
$$\overset{(3)}{\leq} \underbrace{f_0(x^*) + \sum_{i=1}^{m} \lambda_i^* f_i(x^*) + \sum_{i=1}^{p} \nu_i^* h_i(\vec{x})}_{L(\vec{x}^*, \vec{\lambda}^*, \vec{\nu}^*)}$$

$$\overset{(4)}{\leq} f_0(x^*) + \underbrace{0}_{\substack{f_i(\vec{x}^*) \leq 0 \\ \lambda_i^* \geq 0}} + \underbrace{0}_{h_i(\vec{x}^*) = 0}$$

$$= f_0(\vec{x}^*)$$

But this means inequalities (3) and (4) must in fact be equalities!

Most importantly:

$$\sum_{i=1}^{m} \lambda_i^* f_i(\vec{x}^*) = 0 \qquad \text{because } (3) = (4).$$

But recall $\lambda_i^* \geq 0$ and $f_i(\vec{x}^*) \leq 0$

$$\Rightarrow \quad \lambda_i f_i(\vec{x}^*) \leq 0.$$

The only way a sum of non-negative terms is 0 is if each term is zero. Therefore:

$$\lambda_i f_i(\vec{x}^*) = 0 \qquad \forall \; i = 1, 2, \ldots m.$$

If $\lambda_i > 0 \Rightarrow f_i(\vec{x}^*) = 0$

If $f_i(\vec{x}^*) < 0 \Rightarrow \lambda_i^* = 0$

This is called "<u>complimentary slackness</u>".

KKT Conditions Karush-Kuhn-Tucker.

Non-convex problem. Strong duality holds.
(ie. convexity is not needed). Differentiable

KKT Necessary conditions ie.

Let x^* be primal optimal, (λ^*, ν^*) be dual opt.

Then, their optimality implies:

$$f_i(x^*) \leq 0 \quad , \quad i=1, \dots m \quad \text{Primal feasible}$$

$$h_i(x^*) = 0 \quad , \quad i=1 \dots p.$$

$$\lambda_i^* \geq 0 \quad , \quad i=1 \dots m. \quad \text{Dual feasible.}$$

$$\lambda_i^* f_i(x^*) = 0 \quad , \quad i=1 \dots m. \quad \text{Complementary Slackness}$$

$$\nabla f_0(x^*) + \sum_{i=1}^{m} \lambda_i^* \nabla f_i(x^*) + \sum_{i=1}^{p} \nu_i^* \nabla h_i(x^*) = 0$$

$\hookrightarrow$ x^* minimizes $L(x, \lambda^*, \nu^*)$
fixed.

Hence, grdient w.r.t. x must be zero at optimum point x^*.

It is important to note that for non-convex problems, these are necessary conditions, ie. IF x^*, λ^*, ν^* are optimal they must satisfy the conditions, but there can be points that satisfy the conditions, but are not optimal.

KKT conditions for convex problems. + diffrentiable

"Sufficient Condition"

Now if $f_i(x)$ $(f_0(x), f_1(x) \dots f_m(x))$ are all convex, and $h_i(x)$ are affine, and $\tilde{x}, \tilde{\lambda}, \tilde{\nu}$ are points that satisfy the KKT conditions. i.e. if

$$f_i(\tilde{x}) \leq 0 \quad , \quad i=1,2, \dots m$$

$$h_i(\tilde{x}) = 0 \quad , \quad i=1,2 \dots p.$$

$$\tilde{\lambda}_i \geq 0 \quad , \quad i=1, \dots m$$

$$\tilde{\lambda}_i \cdot f_i(\tilde{x}) = 0 \quad , \quad i=1, \dots m$$

$$\nabla f_0(\tilde{x}) + \sum_{i=1}^{m} \tilde{\lambda}_i \nabla f_i(\tilde{x}) + \sum_{i=1}^{p} \tilde{\nu}_i \nabla h_i(\tilde{x}) = 0 .$$

then : $\tilde{x}$ is primal optimal and $(\tilde{\lambda}, \tilde{\nu})$ are dual opt.

$\tilde{x}, \tilde{\lambda}, \tilde{\nu}$ satisfy KKT $\iff$ $\tilde{x}, \tilde{\lambda}, \tilde{\nu}$ are primal and dual optimal.

+ problem is convex
+ h_i are affine
+ Slater's condition.

KKT conditions are necessary AND sufficient.

Proof:

Consider: $L(x, \tilde{\lambda}, \tilde{\nu})$ function of x.

Since $\tilde{\lambda} \succeq 0$, this is convex in x.

$$L(x, \tilde{\lambda}, \tilde{\nu}) = \underbrace{f_0(x)}_{\text{convex}} + \sum \underbrace{\tilde{\lambda}_i}_{\geq 0} \underbrace{f_i(x)}_{\text{convex}} + \sum \nu_i \underbrace{h_i(x)}_{\text{affine}}$$

$\therefore$ If $\nabla L(x, \tilde{\lambda}, \tilde{\nu})\Big|_{x=\tilde{x}} = 0$, then $\tilde{x}$ is minimum point.

$$g(\tilde{\lambda}, \tilde{\nu}) = \min_x L(x, \lambda, \tilde{\nu}) = L(\tilde{x}, \tilde{\lambda}, \tilde{\nu})$$

$$= f_0(\tilde{x}) + \sum \tilde{\lambda}_i f_i(\tilde{x}) + \sum \tilde{\nu}_i h_i(\tilde{x})$$

$$= f_0(\tilde{x}) + \underbrace{0}_{\substack{\text{Complementary} \\ \text{slackness} \\ \text{KKT}}} + \underbrace{0}_{h_i(\tilde{x})=0}$$

$\therefore (\tilde{x})$ and $(\tilde{\lambda}, \tilde{\nu})$ have zero duality gap.

- $g(\lambda, \nu)$ is a lower bound for every primal value $f_0(x)$. So if $g(\tilde{\lambda}, \tilde{\nu}) = f_0(\tilde{x})$ for some $\tilde{x}$, this must be the optimum.