11/8 : Duality & the dual LP

Consider minimization problem of the form:

$$\text{minimize } f(x)$$
$$\text{subject to } g(x) \leq 0$$

Standard way to solve is the method of Lagrange multipliers. Two cases to check:

1) $g(x) < 0$
2) $g(x) = 0$

If $g(x) < 0$, then x is in the interior of $\mathcal{X} = \{x \in \mathbb{R}^n : g(x) \leq 0\}$.

Hence, makes sense to check the first-order necessary conditions:

$$\boxed{\nabla f(x^*) = 0}$$

to find stationary points $x^* \in \mathcal{X}^\circ = \{x \in \mathbb{R}^n : g(x) < 0\}$.

On the other hand, if $x \in \partial\mathcal{X} = \{x \in \mathbb{R}^n : g(x) = 0\}$, then this no longer makes sense. Instead, we need to minimize f over $\partial\mathcal{X}$:

$$\text{minimize } f(x)$$
$$\text{subject to } g(x) = 0.$$

It will help to pause and consider the geometry of the set $\partial\mathcal{X}$.

To simplify things, we will assume $\partial\mathcal{X}$ is smooth.

Since ∂X is smooth (or at least C^1), we can reason as follows. Let $x_0 \in \partial X$ be a point on the boundary, and let $x(t)$ be a smooth curve such that $x(t) \in \partial X$ for all $t \in \mathbb{R}$ and $x(0) = x_0$. Then, we know:

$$g(x(t)) = 0 \quad \forall t.$$

Differentiating with respect to t gives:

$$\nabla g(x(t))^T x'(t) = 0.$$

Evaluating at $t=0$:

$$\nabla g(x_0)^T x'(0) = 0.$$

Note that $x'(0)$ is tangent to ∂X at x_0. Since $x(t)$ was arbitrary, we can see that $\nabla g(x_0)$ is orthogonal to <u>all</u> ~~tangent~~ vectors tangent to ∂X at x_0. So, it is <u>normal</u> to ∂X at x_0.

E.g., if ∂X is a 2D surface in $\mathbb{R}^3$, then $\nabla g(x_0)$ is a surface normal. In fact, since ∇g points in the direction of steepest ascent, we can see that it is ~~an~~ the outward-facing normal.

At the same time, we can consider minimizing f over another arbitrary curve $x(t) \in \partial X$. The first-order conditions of optimality are:

$$0 = \frac{d}{dt} f(x(t))\Big|_{t=0} = \nabla f(x(0))^T x'(0) = \nabla f(x_0)^T x'(0).$$

So we can also see that $\nabla f(x_0)$ is normal to ∂X at x_0!

However, since we're trying to solve a minimization problem, we can tell that $\nabla f(x_0)$ must be an <u>inward</u>-facing normal!

So, $\nabla f(x_0)$ and $\nabla g(x_0)$ are parallel and point in opposite directions. There must then exist a scalar $\lambda > 0$ such that:

$$\nabla f(x_0) = -\lambda \nabla g(x_0) \qquad (\text{or } \nabla f(x_0) + \lambda \nabla g(x_0) = 0).$$

Note that we can "generalize" this to the "$g(x) = 0$" case by allowing $\lambda = 0$ (in which case $\nabla f(x_0) = 0$). In this case we have two mutually exclusive cases:

1) $\lambda > 0$ and $g(x) = 0$
2) $\lambda = 0$ and $g(x) < 0$

Equivalently, this can be written:

$$\boxed{\lambda g(x) = 0 \quad \& \quad \lambda \geq 0 \quad \& \quad g(x) \leq 0}$$

Taking a slight leap of faith, if we write:

$$L(x, \lambda) = f(x) + \lambda g(x),$$

then a set of first-order necessary conditions for this optimization problem are:

$\dfrac{\partial L}{\partial x} = 0$	(stationarity)
$g(x) \leq 0$	(primal feasibility)
$\lambda \geq 0$	(dual feasibility)
$\lambda g(x) = 0$	(complementary slackness)

If we generalize this slightly, we get the $\overbrace{\text{KKT}}^{\text{Karush-Kuhn-Tucker}}$ conditions, which give first-order necessary conditions for constrained optimizations

Theorem: (KKT) Let x^* be a constrained local minimum of:

$$\begin{aligned} &\text{minimize} && f(x) \\ &\text{subject to} && g_i(x) \le 0, && i \in \mathcal{I} \\ & && g_i(x) = 0, && i \in \mathcal{E}, \end{aligned}$$

~~Then~~ and assume that f and each g_i are C^1 and that the set of gradients

We can generalize this a bit to get the KKT conditions, which are the first-order necessary conditions for a point to be a constrained optimum. Consider the minimization problem: (4)

$$\begin{aligned} &\text{minimize} && f(x) && \text{[cost function]} \\ &\text{subject to} && g_i(x) = 0 \quad i \in \mathcal{E} && \text{[equality constraints]} \\ & && g_i(x) \le 0 \quad i \in \mathcal{I} && \text{[inequality constraints]} \end{aligned} \qquad (*)$$

Each constraint has an associated index i.

Def: the active set at a point x is the set $\mathcal{A}(x) = \mathcal{E} \cup \{ i \in \mathcal{I} : g_i(x) = 0 \}$.

Def: a point $x \in \mathcal{X} = \{ x \in \mathbb{R}^n : g_i(x) = 0 \ (i \in \mathcal{E}) \ \& \ g_i(x) \le 0 \ (i \in \mathcal{I}) \}$ is a constrained local minimum if there exists a neighborhood of x, N, such that for all $y \in N \cap \mathcal{X}$, $f(x) \le f(y)$.

Theorem (KKT): Assume x^* is a constrained local minimum of $(*)$, and assume f and each g_i are C^1. Furthermore, assume that the set

$$\{ \nabla g_i(x^*) : i \in \mathcal{A}(x^*) \}$$

is linearly independent. Then, there exists $\lambda^* \ge 0$ such that for

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

we have that (x^*, λ^*) together satisfy:

$$\boxed{\left.\frac{\partial L}{\partial x}\right|_{(x^*, \lambda^*)} = 0, \quad g_i(x^*) = 0 \ \forall i \in \mathcal{E}, \quad g_i(x^*) \le 0 \ \forall i \in \mathcal{I}, \quad \lambda_i^* \ge 0 \ \forall i \in \mathcal{I}, \quad \lambda_i^* g_i(x^*) = 0 \ \forall i \in \mathcal{E} \cup \mathcal{I}.}$$

Note that this is a constrained system of ~~the~~ nonlinear equations in x^* and λ^*.

Back to LPs... Recall an LP in standard form:

$$\begin{aligned} \text{minimize} \quad & c^T x \\ \text{subject to} \quad & Ax = b \qquad (**) \\ & x \geq 0 \end{aligned}$$

Let's apply the KKT conditions to this LP: Write:

$$f(x) = c^T x$$
$$g(x) = Ax - b$$
$$h(x) = -x$$

so that $(**)$ becomes:

$$\begin{aligned} \text{minimize} \quad & f(x) \\ \text{subject to} \quad & g(x) = 0 \\ & h(x) \leq 0 \end{aligned}$$

Let λ be the Lagrange multiplier associated with g and s be the Lagrange multiplier associated with h. Then $L(x, \lambda, s) = f(x) + \lambda^T g(x) + s^T h(x)$.

The KKT conditions are:

$$\frac{\partial L}{\partial x} = c + A^T \lambda - s = 0$$

$$Ax = b, \quad x \leq 0, \quad s \geq 0, \quad x_i s_i = 0 \;\; \forall i = 1, \ldots, n.$$

Next time, we will see what these equations tell us about the LP.