Lecture #7

1. Recap: We saw last time, a few suitable Lagrangian duals: weak-duality; strong-duality; minimality.

2. Today, let me quickly mention these optimality conditions again, before we discuss specific examples.

3. Recall our last result: \( \partial D_f(x) \) is nonempty (but not necessarily) for \( x \) to be a local (global) for \( f \).

4. In the realm of convexity, let's look at other optimality conditions, all of which are ultimately based around giving meaningful ways to characterize/verify \( \partial D_f(x) \).

5. Lemma: \( \partial D_f(x) \) iff \( f \) is finite at \( x \) and \( f'(x,y) \geq 0 \) for all \( y \).

   Proof: we know that \( \partial D_f(x) \) iff

   \[ f'(x,y) \geq \langle g, y \rangle \quad \forall y. \]

   Recall: \( f'(x,y) = \lim_{h \to 0} \frac{f(x+hy) - f(x)}{h} = \inf_{h > 0} \frac{f(x+hy) - f(x)}{h} \)

6. Why does this hold? Since \( f \) is concave, the function \( f(x+hy) - f(x) \) is linear.

   Consider the function:

   \[ h(y) = \frac{(x+hy) - f(x)}{h} \]

   \( h(y) \) is also concave (push homotopy).

   The function \( g(x) = \frac{f(x+hy) - f(x)}{h} \) is an upper bound for \( g \).

   Because \( f \) is concave, it has a unique slope.

   Alternative: \( h(y) = f(x+hy) - f(x) \) is concave;

   \( x \cdot h(y) \) is monotonic.

   Why does "\( \inf \) hold? Since \( f \) is concave, the

   "Alt. Slope" (Bruno) \( x \to \frac{f(x+hy) - f(x)}{h} \) is decreasing.
Thus we set \( f(\bar{x} + h) \) for \( h > 0 \), any \( y \)

\[
\begin{align*}
 f(\bar{x} + h) &> f(\bar{x}) + \langle g, h \rangle \\
 f(\bar{x} + h) - f(\bar{x}) &> \langle g, h \rangle \\
 \frac{f(\bar{x} + h) - f(\bar{x})}{h} &> \langle g, y \rangle \\
 \Rightarrow \lim_{h \to 0} \frac{f(\bar{x} + h) - f(\bar{x})}{h} &= \langle g, y \rangle.
\end{align*}
\]

(Other details as before)

**Theorem.** Let \( f \) be a closed, lower \((\neq +\infty)\) CER for \( f \).

Then \( \inf f = -f^*(0) \). Thus, \( f \) is bounded below.

Let \( 0 \in \text{dom } f^* \).

**Proof:** \( f^*(z) = \sup_x \langle z, x \rangle - f(x) \), hence \( z = 0 \).

(5) The minimum must \( z \) if \( f \) is \( \partial f^*(0) \). Then the \( \inf \) is attained \( \iff f^* \) is subdifferentiable at \( 0 \).

**Proof:** Recall that \( z \) is a \( \min \) \( f \) iff \( 0 \in \partial f^*(a) \)

But \( g \in \partial f^*(g) \iff \exists \bar{x} \in \partial f^*(0) \) \( \Rightarrow \exists \bar{x} \in \partial f^*(0) \) \( \Rightarrow \exists \bar{x} \in \partial f^*(0) \)

(\( \exists \bar{x} \in \partial f^*(0) \))

(5) For \( \inf \) of \( f \) to be finite but unattained

it is neccessary and sufficent that \( f^*(0) \) be finite but \( f^*(0, y) = \infty \)

for some direction \( y \).
Let \( f^{**}(y) = \emptyset \) be the dual problem, where

\[
\text{for } x \text{ fixed.}
\]

\[
\text{unbounded when } \exists y \text{ s.t. } f^{**}(0; y) = -\infty \text{ bounds.}
\]

**Theorem:**
\[
\exists f(x) = \{ g \in \mathbb{R}^n \mid f'(x; y) > (g, y), \forall y \in \mathbb{R}^n \}
\]

**Reason:**
The minimum set of \( f' \) is nonempty and bounded if and only if \( 0 \in \text{int}(\text{dom } f^{**}) \) and

The main set is a singleton if \( f^{**} \) is differentiable at 0 and

\[
x^* = \nabla f^{**}(0)
\]

(Note: \( f^{**}(0) \) is a singleton \( \Rightarrow f^{**} \) is diff at 0.)

Now recall Lagrangian form last time:

\[
\min f(x) + \lambda^T h(x) = \text{Lagrangian}
\]

\[
L(x, \lambda) = f(x) + \lambda^T h(x)
\]

Let \( (P) \) be the primal problem:

\[
\min f(x) \quad \text{s.t.} \quad h_i(x) \leq 0, \quad i = 1, \ldots, m
\]

The Lagrangian of \( (P) \) is the function \( IR^n \times IR^m \) defined by:

\[
L(x, \lambda) = \begin{cases} f(x) + \lambda^T h(x), & \text{if } L(x, \lambda) \text{ is feasible} \\ -\infty, & \text{if } \exists \lambda \neq 0 \text{ s.t. } L(x, \lambda) = -\infty \\
+\infty, & \text{if } x \notin C. \end{cases}
\]
We have that
\[
\inf \sup_{z \in C} L(z, x) \leq \inf \sup_{z \in C} L(z, x) = \sup_{z \in C} \inf_{x \in E} L(z, x) = \sup_{z \in C} \inf_{x \in E} L(z, x).
\]

And we note that \( \sup_{z \in C} \inf_{x \in E} L(z, x) \) is a minimal setup

- \( L \) as defined above contains complete info about \((P)\).
- \( L \) is finite \( \mathcal{E} \times \mathcal{C} \) i.e. \( L \) finite
- \( f(x) = L(z, x) \)
- \( h_i(z) = L(z, x_i) - L(z, x) \quad \forall \ i = 1, \ldots, n, z \in C \)

Now recall a pair \((\overline{x}, \overline{z})\) called a saddle-point \( L \)

- if \( L(\overline{x}, \overline{z}) \leq L(\overline{x}, z) \leq L(x, z) \quad \forall x \in X, z \in C \)

(and for this \( \inf \sup L = \text{epi min} L = \min L \) \( \neq L(\overline{x}, \overline{z}) \))

---

**Main Theorem:** Let \((P)\) be a convex problem as stated above.
Let \( \overline{x}, \overline{z} \) be vectors in \( \mathbb{R}^n \times \mathbb{R}^m \). In order that \( \overline{z} \) be a VKT vector for \((P)\) and \( \overline{x} \) an optimal solution to \((P)\) it is necessary and sufficient that \((\overline{x}, \overline{z})\) is a saddle-point of the Lagrangian.

Moreover, this condition holds iff \( \overline{x} \) and \( \overline{z} \) satisfy

- \( \lambda_i > 0 \) for \( 1 \leq i \leq r \)
- \( h_i(\overline{x}) = 0 \) for \( 1 \leq i \leq r \)
- \( \lambda_i h_i(\overline{x}) = 0 \) for \( 1 \leq i \leq m \)
- \( \lambda_i h_i(\overline{x}) = 0 \) for \( i = m+1, \ldots, n \)
- \( c \in \left( \sum_{i=1}^{m} h_i(\overline{x}) + \lambda \omega(\overline{x}) \right) \)
Alternative derivation of the algorithm

Assume for simplicity that we know \( h_i(x) = 0 \) for \( 1 \leq i \leq m \)

We also assume that \( \exists \mathbf{x} \in \mathbb{R}^n \) such that

\[
\begin{align*}
&h_i(x) < 0, \quad \ldots \quad h_m(x) < 0 \\
&\text{Write: } C_i = \{ \mathbf{x} | h_i(x) \leq 0 \text{ for } 1 \leq i \leq m \}
\end{align*}
\]

\( (P) = \min_{\mathbf{x}} \phi(x) := f(x) + \sum_{i=1}^{m} \phi_i(x) \)

The optimal solution to \( (P) \) are vectors \( \mathbf{x} \) s.t.

\[
0 \in \partial \phi(x)
\]

The assumption \( \exists \mathbf{x} \text{ s.t. } h_i(x) < 0 \)

\[
\text{int } C_1 \cap \ldots \cap \text{int } C_m \neq \emptyset
\]

(From the facts that \( C_i \) are closed and \( h_i(x) < 0 \) for \( 1 \leq i \leq m \), it is bounded. Hence, by Rockafellar's theorem on subdifferential, we follow that)

\[
\partial \phi(x) = \partial f(x) + \sum_{i=1}^{m} \partial h_i(x)
\]

\( \partial \phi_i = \partial h_i \) (normal case) we derived it.

And an exercise shows that (by subdifferential calculus rules from Lect. 3)

\[
\partial \phi_i = \left\{ \begin{array}{ll}
\emptyset & \text{if } h_i(x) > 0 \\
\{ \mathbf{y} \in \mathbb{R}^n | h_i(x) + \mathbf{y} \leq 0 \} & \text{if } h_i(x) < 0
\end{array} \right.
\]

Using this expression, we can verify that the subdifferentials satisfy the subgradient conditions.

Therefore, the optimal solution is found where the subdifferentials vanish, i.e.,

\[
\partial \phi(x) = 0
\]

Hence, the algorithm terminates.
Thus, $\mathbf{\partial} \phi(x) \neq \emptyset$ only if

$x$ satisfies $h_i(x) \leq 0 \quad \forall i \in \mathcal{E}_m$

(since, contrary to what I said in Part 3, $\mathcal{A} + \phi = \emptyset$)

Thus, $\mathbf{\partial} \phi(x) = \bigcup \{ \mathbf{\partial} \phi(x) + \lambda_i \mathbf{h}_i(x) + \lambda_i \mathbf{h}_i(x) \}$

over any chain of $\lambda_i > 0$

such that

$\lambda_i h_i(x) = 0$

Since if $h_i(x) < 0$, $\mathbf{\partial} \phi(x) = \emptyset$;

while for $h_i(x) = 0$, $\mathbf{\partial} \phi(x) = \lambda_i \mathbf{h}_i(x) \mathbf{h}_i(x)$

So either $\lambda_i > 0$ or $h_i(x) > 0$ (we cannot have $h_i(x) = 0$).

Thus, $0 \in \mathbf{\partial} \phi(x)$ iff $x$ satisfies the given conditions.

Example: $\min \left\{ \frac{1}{2} \| \mathbf{A} x - \mathbf{b} \|_2 \right\}$

subject to $\| \mathbf{A} x - \mathbf{b} \|_2 \leq 1$

Let $\phi(x) = \frac{1}{2} \| \mathbf{A} x - \mathbf{b} \|_2^2 + \delta_c(x)$

$\mathbf{\partial} \phi(x) = \mathbf{A}^T (\mathbf{A} x - \mathbf{b}) + \mathbf{\partial} \delta_c(x)$

$\exists \lambda \in [0,1]$ such that $\mathbf{A}^T \mathbf{A} x + \lambda \mathbf{1} \mathbf{1}^T x = \mathbf{A} b$.

Clearly, $0 \in \mathbf{\partial} \phi(x)$.

Thus, $\mathbf{A}^T (\mathbf{A} x - \mathbf{b}) + \lambda \mathbf{1} \mathbf{1}^T x = \mathbf{A} b$

$\lambda \geq 0$; $\lambda \mathbf{1} \mathbf{1}^T x = \mathbf{0}$
Recall that \( DS_c(x) = N_c(x) \)

\[
= \{ g \mid 0 \geq \langle g, y-x \rangle \ \forall y \in C \}
\]

Since this is a cone if \( g \in N_c(x) \)

\[
= \{ g \in N_c(x) \mid A x \geq 0 \}
\]
Sometimes, we can directly write the KKT conditions (solving non-linear equations).

Usually, that is too hard, and we proceed to:

- Iterating below them (Newton method)
- Some generate a seq $(x^k)$ that satisfies
  KKT in the limit
- A few other variants

Let's try to see briefly extend above duality optimality
theory to the conic world.

**Facts about LP duality:**

- **Conjugate:** \[ \min \{ c^T x : A x \leq b \} \]
- **Dual:** \[ \max \{ b^T y : A^T y + c = 0 \} \]

A. If either $\mathbf{p}^* \text{ or } \mathbf{d}^*$ finite, then $\mathbf{p}^* = \mathbf{d}^*$ and

- both primal / dual solutions attained.

  (recall \( \vartheta(u) = \inf \{ c^T x : A x \leq b, x \geq u \} \))

  \( \vartheta(b) = p^* \Rightarrow \mathbf{p}^* \text{ is feasible} \)

- \( p^* = -\infty \) then \( d^* = -\infty \) (weak dual)
- \( p^* = +\infty \) then \( d^* = +\infty \) (weak dual)
Known results about LP duality:

1. \( p^* = -\infty \Rightarrow d^* = -\infty \)
2. \( d^* = +\infty \Rightarrow p^* = +\infty \)
3. If \( p^* \) or \( d^* \) finite, then \( p^* = d^* \) and both \((1) + (2)\) have optimal solutions.

Thus, if LP in strongly duality

if \( p^* = +\infty \) (inferimal, ran so fast)

SDP - duality:

(Original form) \( p^* = \inf \{ \langle C, x \rangle \mid \langle A_i, x \rangle = b_i, \; x \succeq 0 \} \)

How to handle the \( x \succeq 0 \) constraint?

- Introduce the "Canic Lagrangian"

\[ \ell(x, u, y) := \langle C, x \rangle + \sum_i y_i (\langle A_i, x \rangle - b_i) - \langle y, x \rangle \]

where \( y \in \mathbb{R}^m \), and \( x \geq 0 \): matrix dual variable.

- Implicit fact: If \( x, y \succeq 0 \) then \( \ell(x, y, y) > 0 \)

\[ \Rightarrow \quad p^* \geq \ell(x, y, y) \quad \text{finite} \quad x, y, y \]

\[ = \sup_{y \succeq 0} \ell(x, y, y) \]

\[ \inf \text{ as } y \text{ with} \quad p^* > d^* = \sup_{y \succeq 0} \inf_{x} \ell(x, y, y) \]
Simplex form \( L(x, y, z) \) yields

\[ g(y, z) = \begin{cases} 2 + b^Tz + 1 & \text{if } C - \sum x_i a_i - y = 0 \\ \infty & \text{otherwise} \end{cases} \]

So dual problem:

\[
\begin{align*}
\max & \quad b^Tz \\
\text{subject to} & \quad y + y_0 = C - \sum x_i a_i = 0 \\
& \quad y_0 \geq 0
\end{align*}
\]

This is called the canonical form of SDP.

Some points:
- Strong duality holds if either primal or dual is strictly feasible.
- (In LPs, more feasible solutions)
- (Example: examples with slides related to.)

Lemma: Let \( A, a_n, c \) be symmetric matrices.

The system:

\[ x_i A_i + x_n A_n = c \]

If \( x \) is a symmetric matrix \( y \) if:

\[ y(A_i + y) = 0 \]
\[ y(C + y) > 0 \]
\[ y > 0 \]

Ex. 5.44

Proof:

If \( \max x_i A_i + x_n A_n = c \) has no solution, the submatrix \( L \) has rank

in deficient from the interior of \( \mathbb{R}^n \) \( \implies \exists \) separating hyperplane

Let:

\[ \exists x \mid y^T(yx) = 0 \quad \text{we may assume that} \quad y^T(yx) > 0 \]

We have:

\[ y^T(x_i + A_i y) = 0 \]

\[ y^T(x_n + A_n y) = 0 \]

Ex. 5.44

BV
That this works
whenever
\[ H = 2X^t F_0(Yx) = 0 \]
we can define
\[ C = h(Yx) \]
\[ h((Y-c)x) = 0 \]
\[ \forall l < Yx \Rightarrow \langle c|x \rangle = ? \]
\[ \forall l < 0 \text{ or } \geq 0 \]

Most likely CRF in a LP^m : 
\[ \log \det(x) \]

The set
\[ A^0(x) := 2A_0 + x A_1 + \cdots + x_m A_m \mathbb{R}^{0,5} \]

Called a "Sherrahadon"

Coincide with solution set of an SDP

Rank of SDP

Barvinok: If LP SDP is solvable, then \( E \) a double when rank \( B \) satisfies \( \Delta(E) \leq m \)

\( m \) is \# constraints)

Reading assignment: Barvinok & Monteiro. A nonlinear long algorithm via low-rank factorization.