Many of these slides are taken from Ryan Tibshirani’s convex optimization class

References:
Duality in linear programs

Suppose we want to find lower bound on the optimal value in our convex problem, $B \leq \min_{x \in C} f(x)$

E.g., consider the following simple LP

$$\min_{x,y} \quad x + y$$

subject to $x + y \geq 2$

$x, y \geq 0$

What’s a lower bound? Easy, take $B = 2$

But didn’t we get “lucky”?
Duality in linear programs

Try again:

\[
\begin{align*}
\min_{x,y} & \quad x + 3y \\
\text{subject to} & \quad x + y \geq 2 \\
& \quad x, y \geq 0
\end{align*}
\]

\[
\begin{align*}
x + y & \geq 2 \\
+ & \quad 2y \geq 0 \\
= & \quad x + 3y \geq 2
\end{align*}
\]

Lower bound \( B = 2 \)

More generally:

\[
\begin{align*}
\min_{x,y} & \quad px + qy \\
\text{subject to} & \quad x + y \geq 2 \\
& \quad x, y \geq 0
\end{align*}
\]

\[
\begin{align*}
a + b & = p \\
a + c & = q \\
a, b, c & \geq 0
\end{align*}
\]

Lower bound \( B = 2a \), for any \( a, b, c \) satisfying above
What’s the best we can do? Maximize our lower bound over all possible \( a, b, c \):

\[
\begin{align*}
\min_{x,y} & \quad px + qy \\
\text{subject to} & \quad x + y \geq 2 \\
& \quad x, y \geq 0
\end{align*}
\]

\[
\begin{align*}
\max_{a,b,c} & \quad 2a \\
\text{subject to} & \quad a + b = p \\
& \quad a + c = q \\
& \quad a, b, c \geq 0
\end{align*}
\]

Called **primal** LP

Called **dual** LP

Note: number of dual variables is number of primal constraints
Duality in linear programs

Try another one:

\[
\begin{align*}
\min_{x,y} & \quad px + qy \\
\text{subject to} & \quad x \geq 0 \\
& \quad y \leq 1 \\
& \quad 3x + y = 2
\end{align*}
\]

Primal LP

\[
\begin{align*}
\max_{a,b,c} & \quad 2c - b \\
\text{subject to} & \quad a + 3c = p \\
& \quad -b + c = q \\
& \quad a, b \geq 0
\end{align*}
\]

Dual LP

Note: in the dual problem, \( c \) is unconstrained.
Duality in linear programs

Given \( c \in \mathbb{R}^n, A \in \mathbb{R}^{m \times n}, b \in \mathbb{R}^m, G \in \mathbb{R}^{r \times n}, h \in \mathbb{R}^r \)

\[
\begin{align*}
\min_{x \in \mathbb{R}^n} & \quad c^T x \\
\text{subject to} & \quad Ax = b \\
& \quad Gx \leq h
\end{align*}
\]

Primal LP

\[
\begin{align*}
\max_{u \in \mathbb{R}^m, v \in \mathbb{R}^r} & \quad -b^T u - h^T v \\
\text{subject to} & \quad -A^T u - G^T v = c \\
& \quad v \geq 0
\end{align*}
\]

Dual LP

Explanation: for any \( u \) and \( v \geq 0 \), and \( x \) primal feasible,

\[
\begin{align*}
\sum_{i=1}^{m} u_i (Ax_i - b_i) + \sum_{j=1}^{r} v_j (Gx_j - h_j) & \leq 0, \quad \text{i.e.,} \\
-A^T u - G^T v)^T x & \geq -b^T u - h^T v
\end{align*}
\]

So if \( c = -A^T u - G^T v \), we get a bound on primal optimal value
Another perspective on LP duality

\[
\begin{align*}
\text{Primal LP} & \quad \text{Dual LP} \\
\min_{x \in \mathbb{R}^n} & \quad \max_{u \in \mathbb{R}^m, v \in \mathbb{R}^r} \\
\quad c^T x & \quad -b^T u - h^T v \\
\text{subject to} & \quad \text{subject to} \\
Ax = b & \quad -A^T u - G^T v = c \\
Gx \leq h & \quad v \geq 0
\end{align*}
\]

Explanation #2: for any \( u \) and \( v \geq 0 \), and \( x \) primal feasible

\[
c^T x \geq c^T x + u^T (Ax - b) + v^T (Gx - h) := L(x, u, v)
\]

So if \( C \) denotes primal feasible set, \( f^* \) primal optimal value, then for any \( u \) and \( v \geq 0 \),

\[
f^* \geq \min_{x \in C} L(x, u, v) \geq \min_{x \in \mathbb{R}^n} L(x, u, v) \quad \text{:=} \quad g(u, v)
\]
Another perspective on LP duality

In other words, \( g(u, v) \) is a lower bound on \( f^* \) for any \( u \) and \( v \geq 0 \)

Note that

\[
g(u, v) = \begin{cases} 
-b^T u - h^T v & \text{if } c = -A^T u - G^T v \\
-\infty & \text{otherwise}
\end{cases}
\]

Now we can maximize \( g(u, v) \) over \( u \) and \( v \geq 0 \) to get the tightest bound, and this gives exactly the dual LP as before

This last perspective is actually completely general and applies to arbitrary optimization problems (even nonconvex ones)
Consider general minimization problem

\[
\min_{x \in \mathbb{R}^n} f(x)
\]
subject to \( h_i(x) \leq 0, \ i = 1, \ldots m \)
\( \ell_j(x) = 0, \ j = 1, \ldots r \)

Need not be convex, but of course we will pay special attention to convex case

We define the **Lagrangian** as

\[
L(x, u, v) = f(x) + \sum_{i=1}^{m} u_i h_i(x) + \sum_{j=1}^{r} v_j \ell_j(x)
\]

New variables \( u \in \mathbb{R}^m, v \in \mathbb{R}^r \), with \( u \geq 0 \) (implicitly, we define \( L(x, u, v) = -\infty \) for \( u < 0 \))
Important property: for any \( u \geq 0 \) and \( v \),

\[
f(x) \geq L(x, u, v) \quad \text{at each feasible } x
\]

Why? For feasible \( x \),

\[
L(x, u, v) = f(x) + \sum_{i=1}^{m} u_i h_i(x) + \sum_{j=1}^{r} v_j \ell_j(x) \leq f(x)
\]

- Solid line is \( f \)
- Dashed line is \( h \), hence feasible set \( \approx [-0.46, 0.46] \)
- Each dotted line shows \( L(x, u, v) \) for different choices of \( u \geq 0 \) and \( v \)

(From B & V page 217)
Let $C$ denote primal feasible set, $f^*$ denote primal optimal value. Minimizing $L(x, u, v)$ over all $x \in \mathbb{R}^n$ gives a lower bound:

$$f^* \geq \min_{x \in C} L(x, u, v) \geq \min_{x \in \mathbb{R}^n} L(x, u, v) := g(u, v)$$

We call $g(u, v)$ the Lagrange dual function, and it gives a lower bound on $f^*$ for any $u \geq 0$ and $v$, called dual feasible $u, v$.
Consider quadratic program (QP, step up from LP!)

\[ \min_{x \in \mathbb{R}^n} \frac{1}{2} x^T Q x + c^T x \]

subject to \( Ax = b, \ x \geq 0 \)

where \( Q > 0 \). Lagrangian:

\[ L(x, u, v) = \frac{1}{2} x^T Q x + c^T x - u^T x + v^T (Ax - b) \]

Lagrange dual function:

\[ g(u, v) = \min_{x \in \mathbb{R}^n} L(x, u, v) = -\frac{1}{2} (c-u+A^Tv)^T Q^{-1} (c-u+A^Tv) - b^T v \]

For any \( u \geq 0 \) and any \( v \), this is lower a bound on primal optimal value \( f^* \)
We choose $f(x)$ to be quadratic in 2 variables, subject to $x \geq 0$. Dual function $g(u)$ is also quadratic in 2 variables, also subject to $u \geq 0$.

Dual function $g(u)$ provides a bound on $f^*$ for every $u \geq 0$.

Largest bound this gives us: turns out to be exactly $f^*$ ... coincidence?

More on this later.
Weak duality

Given primal problem

$$\min_{x \in \mathbb{R}^n} f(x)$$

subject to $h_i(x) \leq 0, \ i = 1, \ldots, m$

$$\ell_j(x) = 0, \ j = 1, \ldots, r$$

Our constructed dual function $g(u, v)$ satisfies $f^* \geq g(u, v)$ for all $u \geq 0$ and $v$. Hence best lower bound is given by maximizing $g(u, v)$ over all dual feasible $u, v$, yielding Lagrange dual problem:

$$\max_{u \in \mathbb{R}^m, v \in \mathbb{R}^r} g(u, v)$$

subject to $u \geq 0$

Key property, called weak duality: if dual optimal value $g^*$, then

$$f^* \geq g^*$$

Note that this always holds (even if primal problem is nonconvex)
Another key property: the dual problem is a **convex optimization** problem (as written, it is a concave maximization problem)

Again, this is always true (even when primal problem is not convex)

By definition:

\[
g(u, v) = \min_{x \in \mathbb{R}^n} \left\{ f(x) + \sum_{i=1}^{m} u_i h_i(x) + \sum_{j=1}^{r} v_j \ell_j(x) \right\}
\]

\[
= -\max_{x \in \mathbb{R}^n} \left\{ -f(x) - \sum_{i=1}^{m} u_i h_i(x) - \sum_{j=1}^{r} v_j \ell_j(x) \right\}
\]

pointwise maximum of convex functions in \((u, v)\)

I.e., \(g\) is concave in \((u, v)\), and \(u \geq 0\) is a convex constraint, hence dual problem is a concave maximization problem
Strong duality

Recall that we always have \( f^* \geq g^* \) (weak duality). On the other hand, in some problems we have observed that actually

\[
f^* = g^*
\]

which is called strong duality

Slater’s condition: if the primal is a convex problem (i.e., \( f \) and \( h_1, \ldots h_m \) are convex, \( \ell_1, \ldots \ell_r \) are affine), and there exists at least one strictly feasible \( x \in \mathbb{R}^n \), meaning

\[
h_1(x) < 0, \ldots h_m(x) < 0 \quad \text{and} \quad \ell_1(x) = 0, \ldots \ell_r(x) = 0
\]

then strong duality holds

This is a pretty weak condition. (And it can be further refined: need strict inequalities only over functions \( h_i \) that are not affine)
Strong duality for LPs

For linear programs:

- Easy to check that the dual of the dual LP is the primal LP
- Refined version of Slater’s condition: strong duality holds for an LP if it is feasible
- Apply same logic to its dual LP: strong duality holds if it is feasible
- Hence strong duality holds for LPs, except when both primal and dual are infeasible

In other words, we pretty much always have strong duality for LPs
KKT Conditions
What we have seen so far

Given a minimization problem

$$\min_{x \in \mathbb{R}^n} f(x)$$

subject to $h_i(x) \leq 0, \ i = 1, \ldots m$
$$\ell_j(x) = 0, \ j = 1, \ldots r$$

we defined the Lagrangian:

$$L(x, u, v) = f(x) + \sum_{i=1}^{m} u_i h_i(x) + \sum_{j=1}^{r} v_j \ell_j(x)$$

and Lagrange dual function:

$$g(u, v) = \min_{x \in \mathbb{R}^n} L(x, u, v)$$
The subsequent dual problem is:

$$\max_{u \in \mathbb{R}^m, v \in \mathbb{R}^r} g(u, v)$$

subject to $u \geq 0$

Important properties:

- Dual problem is always convex, i.e., $g$ is always concave (even if primal problem is not convex)
- The primal and dual optimal values, $f^*$ and $g^*$, always satisfy weak duality: $f^* \geq g^*$
- Slater’s condition: for convex primal, if there is an $x$ such that

$$h_1(x) < 0, \ldots, h_m(x) < 0 \text{ and } \ell_1(x) = 0, \ldots, \ell_r(x) = 0$$

then strong duality holds: $f^* = g^*$. (Can be further refined to strict inequalities over the nonaffine $h_i, i = 1, \ldots, m$)
Given primal feasible $x$ and dual feasible $u, v$, the quantity
\[ f(x) - g(u, v) \]
is called the duality gap between $x$ and $u, v$. Note that
\[ f(x) - f^* \leq f(x) - g(u, v) \]
so if the duality gap is zero, then $x$ is primal optimal (and similarly, $u, v$ are dual optimal).

From an algorithmic viewpoint, provides a stopping criterion: if $f(x) - g(u, v) \leq \epsilon$, then we are guaranteed that $f(x) - f^* \leq \epsilon$.

Very useful, especially in conjunction with iterative methods ... more dual uses in coming lectures.
Subgradients

Remember that for convex $f : \mathbb{R}^n \to \mathbb{R}$,

$$f(y) \geq f(x) + \nabla f(x)^T (y - x) \quad \text{all } x, y$$

i.e., linear approximation always underestimates $f$

A subgradient of convex $f : \mathbb{R}^n \to \mathbb{R}$ at $x$ is any $g \in \mathbb{R}^n$ such that

$$f(y) \geq f(x) + g^T (y - x), \quad \text{all } y$$

- Always exists
- If $f$ differentiable at $x$, then $g = \nabla f(x)$ uniquely
- Actually, same definition works for nonconvex $f$ (however, subgradients need not exist)
Consider $f : \mathbb{R} \rightarrow \mathbb{R}, \ f(x) = |x|$

- For $x \neq 0$, unique subgradient $g = \text{sign}(x)$
- For $x = 0$, subgradient $g$ is any element of $[-1, 1]$
Set of all subgradients of convex $f$ is called the subdifferential:

$$\partial f(x) = \{g \in \mathbb{R}^n : g \text{ is a subgradient of } f \text{ at } x\}$$

- $\partial f(x)$ is closed and convex (even for nonconvex $f$)
- Nonempty (can be empty for nonconvex $f$)
- If $f$ is differentiable at $x$, then $\partial f(x) = \{\nabla f(x)\}$
- If $\partial f(x) = \{g\}$, then $f$ is differentiable at $x$ and $\nabla f(x) = g$
For any $f$ (convex or not),

$$f(x^*) = \min_{x \in \mathbb{R}^n} f(x) \iff 0 \in \partial f(x^*)$$

I.e., $x^*$ is a minimizer if and only if 0 is a subgradient of $f$ at $x^*$

Why? Easy: $g = 0$ being a subgradient means that for all $y$

$$f(y) \geq f(x^*) + 0^T(y - x^*) = f(x^*)$$

Note implication for differentiable case, where $\partial f(x) = \{\nabla f(x)\}$
Given general problem

$$\min_{x \in \mathbb{R}^n} f(x)$$

subject to

$$h_i(x) \leq 0, \quad i = 1, \ldots, m$$
$$\ell_j(x) = 0, \quad j = 1, \ldots, r$$

The Karush-Kuhn-Tucker conditions or KKT conditions are:

- $0 \in \partial f(x) + \sum_{i=1}^{m} u_i \partial h_i(x) + \sum_{j=1}^{r} v_j \partial \ell_j(x)$ (stationarity)
- $u_i \cdot h_i(x) = 0$ for all $i$ (complementary slackness)
- $h_i(x) \leq 0$, $\ell_j(x) = 0$ for all $i, j$ (primal feasibility)
- $u_i \geq 0$ for all $i$ (dual feasibility)
Let $x^*$ and $u^*, v^*$ be primal and dual solutions with zero duality gap (strong duality holds, e.g., under Slater’s condition). Then

$$f(x^*) = g(u^*, v^*)$$

$$= \min_{x \in \mathbb{R}^n} f(x) + \sum_{i=1}^{m} u_i^* h_i(x) + \sum_{j=1}^{r} v_j^* \ell_j(x)$$

$$\leq f(x^*) + \sum_{i=1}^{m} u_i^* h_i(x^*) + \sum_{j=1}^{r} v_j^* \ell_j(x^*)$$

$$\leq f(x^*)$$

In other words, all these inequalities are actually equalities.
Two things to learn from this:

- The point $x^*$ minimizes $L(x, u^*, v^*)$ over $x \in \mathbb{R}^n$. Hence the subdifferential of $L(x, u^*, v^*)$ must contain 0 at $x = x^*$—this is exactly the stationarity condition.

- We must have $\sum_{i=1}^{m} u_i^* h_i(x^*) = 0$, and since each term here is $\leq 0$, this implies $u_i^* h_i(x^*) = 0$ for every $i$—this is exactly complementary slackness.

Primal and dual feasibility obviously hold. Hence, we’ve shown:

If $x^*$ and $u^*, v^*$ are primal and dual solutions, with zero duality gap, then $x^*, u^*, v^*$ satisfy the KKT conditions.

(Note that this statement assumes nothing a priori about convexity of our problem, i.e., of $f, h_i, \ell_j$)
Sufficiency

If there exists $x^*, u^*, v^*$ that satisfy the KKT conditions, then

$$
g(u^*, v^*) = f(x^*) + \sum_{i=1}^{m} u_i^* h_i(x^*) + \sum_{j=1}^{r} v_j^* \ell_j(x^*)
$$

$$
= f(x^*)
$$

where the first equality holds from stationarity, and the second holds from complementary slackness.

Therefore duality gap is zero (and $x^*$ and $u^*, v^*$ are primal and dual feasible) so $x^*$ and $u^*, v^*$ are primal and dual optimal. I.e., we’ve shown:

If $x^*$ and $u^*, v^*$ satisfy the KKT conditions, then $x^*$ and $u^*, v^*$ are primal and dual solutions.
In summary, KKT conditions:

- always sufficient
- necessary under strong duality

Putting it together:

For a problem with strong duality (e.g., assume Slater’s condition: convex problem and there exists \( x \) strictly satisfying non-affine inequality constraints),

\[
x^* \text{ and } u^*, v^* \text{ are primal and dual solutions}
\Rightarrow \quad x^* \text{ and } u^*, v^* \text{ satisfy the KKT conditions}
\]

(Warning, concerning the stationarity condition: for a differentiable function \( f \), we cannot use \( \partial f(x) = \{\nabla f(x)\} \) unless \( f \) is convex)
Consider for $Q \succeq 0$,

$$\min_{x \in \mathbb{R}^n} \frac{1}{2} x^T Q x + c^T x$$

subject to $Ax = 0$

E.g., as in Newton step for $\min_{x \in \mathbb{R}^n} f(x)$ subject to $Ax = b$

Convex problem, no inequality constraints, so by KKT conditions: $x$ is a solution if and only if

$$\begin{bmatrix} Q & A^T \\ A & 0 \end{bmatrix} \begin{bmatrix} x \\ u \end{bmatrix} = \begin{bmatrix} -c \\ 0 \end{bmatrix}$$

for some $u$. Linear system combines stationarity, primal feasibility (complementary slackness and dual feasibility are vacuous)