CO 250 Stuff I Need to Memorize

Sami Yousef

April 9, 2023

Contents

Chapter 1

Linear Programs

1.1 Certificates

For any LP, we know it must have one of three outcomes:

1. **Optimal**: The LP is feasible and the optimal value is attained.
2. **Infeasible**: The LP is infeasible.
3. **Unbounded**: The LP is feasible but unbounded.

We can find a certficate for each of these cases.

Definition. Certficate of Optimality

Definition 1.1.1 (Certificate of Optimality I (Bound)). If the objective function is of the form $\underbrace{c^T}_{<0}\underbrace{x}_{\geq 0}+z$, then we can say z is an upperbound (or lowerbound if LP is min) on the objective function and if we have a solution with value z, then it must be optimal.

Definition 1.1.2 (Certificate of Optimality II (Weak Duality)). Let x and y be solutions to the primal (P) and dual (D) LPs, respectively. Then we have x and y are optimal for (P) and (D) $\iff c^Tx + z_1 = b^Ty + z_2$, where $c^Tx + z_1$ is the objective function of (P) and $b^Ty + z_2$ is the objective function of (D).

Definition 1.1.3 (Certificate of Optimality III (CS Conditions)). Let x and y be solutions to the primal (P) and dual (D) LPs, respectively. Then we have x and y are optimal for (P) and (D) $\iff$ x and y are feasible and satisfy all the CS conditions.

Definition 1.1.4 (Certificate of Infeasibility (Farkas' Lemma)). Let A be the constraint matrix of an LP in SEF. Then we have the LP is infeasible $\iff$ there exists a vector y such that $y^TA \geq \mathbb{0}^T$ and $y^Tb < 0$. This y vector can be found at the end of the Simplex Algorithim: If the ending basis is B, then $y^T = c_B^T A^{-1}$. If the dual has a feasible solution, then that solution can also be used as a certificate.

Definition 1.1.5 (Certificate of Unboundedness). A max LP is unbounded if and only if there exists a family of solutions $x(t) = \bar{x} + td$ for some feasible solution $\bar{x}$ and vector d such that:

1. $d \geq 0$
2. $Ad = \mathbb{0}$.
3. $c^Td > 0$.

More generally, the LP is unbounded if and only if there exists a family of solutions $x(t)$ such

that $x(t)$ is feasible for all $t \geq 0$ and $c^T x(t) \to \infty$ (or $-\infty$ if min) as $t \to \infty$.
$\bar{x}$ and d can be found at the end of the Simplex Algorithim.

1.2 The Simplex Algorithm

For this section, we will assume the LP is in SEF: $\{\max c^T x : Ax = b, x \geq 0\}$. If we are given a starting basis or BFS, we can skip Phase I and proceed with Phase II.

Definition 1.2.1 (Simplex: Phase I). In Phase I, we add auxiliary variables and construct an auxiliary LP with the following form:

$$\min(\mathbb{0}, -\mathbb{1}) \begin{pmatrix} x \\ a \end{pmatrix}$$
$$\text{subject to } \begin{pmatrix} A & I \end{pmatrix} \begin{pmatrix} x \\ a \end{pmatrix} = b$$
$$\begin{pmatrix} x \\ a \end{pmatrix} \geq 0$$

where I is the identity matrix and a are the auxiliary variables. Then we solve the auxiliary problem using the auxiliary variables as the starting basis. We can only proceed to Phase II if the optimal value at the end is zero. Otherwise, the original LP is infeasible.

Definition 1.2.2 (Simplex: Phase II). In Phase II, we solve the original LP using the auxiliary variables as the starting basis. In each iteration, we enter the index left-most positive value of c^T into the basis, and depart the index of min $\frac{b_i}{A_{ji}}$ where j is the index of the entering variable. After entering and departing a variable, we convert the LP into canonical form and repeat until we're done. The canonical form can be achieved by:

1. Finding $y^T = c_B^T A^{-1}$.
2. Replacing the objective function, $c^T x + z_1$, with $[c^T - y^T A]x + z_1 + y^T b$.
3. Multiplying the constraints by A_B^{-1}.

Theorem 1.2.1 (Bland's Rule). If there are multiple entering variables, we choose the one with the smallest index. The simplex algorithm will always terminate if we use Bland's Rule. Note: this theorem was named Bland's rule because it's bland af.

Definition. Degeneracy

Definition 1.2.3 (Degenerate Iteration). A degenerate iteration is an iteration where the BFS does not change after the iteration.

Definition 1.2.4 (Degenerate Solution). A degenerate solution is one where x_i is zero for some $i \in B$.

1.3 Geometry

Definition 1.3.1 (Polyhedron). A polyhedron is a set of points in $\mathbb{R}^n$ that can be described by a finite number of linear inequalities (i.e. an intersection of halfspaces). So every polyhedron can take the form of $P = \{Ax \leq b\}$

Remark. The feasible region of an LP is a polyhedron.

Chapter 2

Duality

2.1 Obtaining the Dual

Definition 2.1.1 (Dual of an LP)**.** Let P be an LP with objective function $c^T x$ and a constraint matrix A with right-hand side b. Then we can use the table below to find the dual of P. The dual will have objective function $b^T y$, constraint matrix A^T, and right-hand side c.

Note.

max		min
$\leq$ constraint	$\Leftrightarrow$	$\geq$ variable
$\geq$ constraint	$\Leftrightarrow$	$\leq$ variable
$=$ constraint	$\Leftrightarrow$	free variable
free variable	$\Leftrightarrow$	$=$ constraint
$\geq$ variable	$\Leftrightarrow$	$\geq$ constraint
$\leq$ variable	$\Leftrightarrow$	$\leq$ constraint

Note that for a max LP, when we go from a constraint to a variable, the sign flips. When we go from a variable to a constraint, the sign does not flip. This can be a helpful way to memorize this.

Remark. The dual of the dual is the primal.

Example (Finding the Dual)**.** Find the dual of:

$$\begin{aligned} &\min \begin{pmatrix} 1 & 2 & 3 \end{pmatrix} x \\ &\text{subject to} \\ &\begin{pmatrix} 3 & 1 & 4 \\ 1 & 5 & 9 \\ 2 & 6 & 5 \end{pmatrix} x \leq \begin{pmatrix} 10 \\ 24 \\ 16 \end{pmatrix} \\ &x_1 \geq 0, x_2 \leq 0, x_3 \text{ free} \end{aligned}$$

Answer. Since we are starting with a min LP, we will look at the right side of the table. We can see that we have $\leq$ constraints, so we must have $\leq$ variables. Since we have $x_1 \geq 0$, $x_2 \leq 0$, and x_3 free, the constraints in the dual should be $\leq, \geq, =$ in order. So we have:

$$\begin{aligned} &\max \begin{pmatrix} 10 & 24 & 16 \end{pmatrix} y \\ &\text{subject to} \\ &\begin{pmatrix} 3 & 1 & 2 \\ 1 & 5 & 6 \\ 4 & 9 & 5 \end{pmatrix} y \begin{matrix} \leq \\ \geq \\ = \end{matrix} \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} \\ &y \leq \mathbb{0} \end{aligned}$$

⊛

2.2 Results of Duality

Theorem 2.2.1 (Weak Duality Theorem)**.** Let (P) be $\{\max c^T x : Ax = b, x \geq \mathbb{0}\}$ and (D) be its dual $\{\min b^T y : A^T y \geq c\}$. If x and y are feasible solutions to (P) and (D) respectively, then $c^T x \leq b^T y$. Moreover, if $c^T x = b^T y$, then x and y are optimal for (P) and (D) respectively.

Theorem 2.2.2 (Strong Duality Theorem)**.** Let (P) and (D) be as in the Weak Duality Theorem. If (P) has an optimal solution, then so does (D). Moreover, (P) and (D) will have the same optimal value

Corollary 2.2.1 (Simple Results of Duality)**.** By the duality theorems above, we have:

1. (P) is infeasible $\Rightarrow$ (D) is infeasible or unbounded.
2. (P) is unbounded $\Rightarrow$ (D) is infeasible.
3. (P) has an optimal solution $\Rightarrow$ (D) has an optimal solution.

Reminder that the dual of the dual is the primal. For example, if we have (D) is infeasible, then (P) is infeasible or unbounded.

2.3 Proving Optimality with Duality

Recall from Optimality Certficate by Weak Duality, that a feasible solution to the dual, y can be used as a certficate of optimality for the primal. We can show this as follows:

$$
\begin{aligned}
&Ax = b \\
\Rightarrow &\underbrace{y^T A}_{\geq^? c^T} x = y^T b
\end{aligned}
$$

Then we have $c^T x \leq^? y^T b$. Therefore, $y^T b$ is an upper bound (or lower-bound for min, hence the ?) on the optimal value of the primal. If we have a feasible solution $\bar{x}$ with value $y^T b$, then we know it is optimal.

2.4 Complementary Slackness (CS) Conditions

Definition 2.4.1 (Complementary Slackness Conditions)**.** Let (P) and (D) be a primal and dual pair with feasible solutions x and y. Then the CS Conditions are:

1. $x_i = 0$ or the corresponding constraint in (D) is tight.
2. $y_i = 0$ or the corresponding constraint in (P) is tight.

As we saw in Optimality Certficate by CS Conditions, if x and y are feasible solutions to (P) and (D) respectively, then x and y are optimal if and only if they satisfy all the CS conditions.

2.5 Trust Me Bro Theorem

This theorem did not have a name in the course notes, so I made a name for it. It's pretty important so it should really have a name.

Theorem 2.5.1 (Trust Me Bro Theorem). Let $\bar{x}$ be a feasible solution to $\{\max c^T x : Ax \leq b\}$. Then $\bar{x}$ is optimal if and only if c is in the cone of tight constraints for $\bar{x}$. The generealization of this is the KKT Theorem

Proof. Trust me bro. ■

Corollary 2.5.1 (Trust Me Bro Theorem Extended). If the LP takes the form of $\{\max c^T x : Ax\ ?\ b\}$, then $\bar{x}$ is an optimal solution if and only if $c = \lambda_1 g_1 + \cdots + \lambda_k g_k$, where g_i are all the tight constraints for $\bar{x}$ and

1. $\lambda_i \geq 0$ for all $i = 1 \ldots k$ if $Ax \leq b$.
2. $\lambda_i \leq 0$ for all $i = 1 \ldots k$ if $Ax \geq b$.
3. λ_i free for all $i = 1 \ldots k$ if $Ax = b$.

Proof. 1. follows from the Trust Me Bro Theorem.

2. Suppose the LP is $\{\max c^T x : Ax \geq b\}$ with a feasible solution $\bar{x}$. Then the dual is $\{\min b^T y : A^T y = c, y \leq \mathbb{0}\}$.

($\Rightarrow$) Suppose $\bar{x}$ is optimal. Without loss of generality, we can assume the first n constraints are tight. Since $\bar{x}$ is optimal, then by CS Conditions, we know $y_{n+1} = \cdots = y_k = 0$ since their corresponding constraint is not tight. Then from the dual, we have $c = A_1^T y_1 + \cdots + A_n^T y_n$, and $y \leq \mathbb{0}$. Here A_i^T are the tight constraints and y_i is λ_i. That concludes the proof.

($\Leftarrow$) Proof is trivial and thus omitted. But, trust me bro, it's true.

3. Follows a similar proof to 2. ■

Chapter 3

Integer Programming

3.1 Shortest Path algorithm

3.1.1 The Problem

Definition 3.1.1 (Shortest Path Problem)**.** Let $G = (V, E)$ be a directed graph with n vertices and m edges. Let $c : E \to \mathbb{R}$ be a cost function. The shortest path problem is to find a path from $s \in V$ to $t \in V$ with the minimum cost. Without loss of generality, we can assume the objective function $c^T x$ (i.e. it is linear).

> **Corollary 3.1.1** (Result of Menger's Theorem)**.** If P is an s,t path and $\delta(S)$ is any s,t cut, then P must have at least one edge in $\delta(S)$. This is a result of a variation of Menger's Theorem not shown here.

We can forumlate the problem as an integer program as follows:

$$\begin{aligned} & \min c^T x \\ \text{s.t.} \quad & \sum (x_e : e \in \delta) \geq 1 \text{ for all s,t cuts } \delta \\ & x_e \in \mathbb{N}_0, \ \text{ for all } e \in E \end{aligned}$$

3.1.2 The Algorithm

Algorithm 1 Shortest Path Algorithm

Input: $G = (V, E)$ **and** $c \in \mathbb{R}^m$ *input*
Output: **an s,t path with minimum cost** *output*

```
Set y := 0, U := {s}, T := {}
while t ∉ U do
    Calculate slack(uv) for all uv ∈ δ(U)
    Find the edge uv with minimum slack
                    ▷ If there are multiple edges with the same slack, choose one arbitrarily
    Set y_U := slack(uv)
    Set U := U ∪ {v}, and T := T ∪ {uv}
                    ▷ At the end of this loop, we have a shortest s,t path in T
output an s,t path in T.
```

3.2 Solving Integer Programs

3.2.1 Properties of Integer Programs

Definition 3.2.1 (Convex Hull)**.** Let $S \subseteq \mathbb{R}^n$ be a set of points. The convex hull of S is the smallest convex set containing S.

Corollary 3.2.1. Each set S has a unique convex hull.

Theorem 3.2.1 (Meyer's Theorem)**.** Let $P = \{Ax \leq b\}$, where A and b are rational. Then the convex hull of integer points in P is polyhedron.

Theorem 3.2.2 (IP and LP Results)**.** Let (IP) be $\{\max c^T x : Ax \leq b, x \in \mathbb{Z}\}$, where A, b are rational. Let the convex hull of feasible solutions be the polyhedron $P = \{x : A'x \leq b'\}$. Then (LP) be $\{\max c^T x : x \in P\}$. Then we have the following results:

1. (IP) is infeasible $\Leftrightarrow$ (LP) is infeasible.
2. (IP) is unbounded $\Leftrightarrow$ (LP) is unbounded.
3. an optimal solution to (IP) is an optimal solution to (LP).
4. an extreme optimal solution of (LP) is an optimal solution of (IP).

Proof. No. ■

3.2.2 Cutting Planes

Definition 3.2.2 (Cutting Planes)**.** We denote $\lfloor ax = b \rfloor$ as $\lfloor a \rfloor x = \lfloor b \rfloor$. Let $\bar{x}$ be the optimal solution for the relaxed IP.
For any constraint $ax = b$, if $\bar{x}$ does not satisfy $\lfloor ax \leq b \rfloor$, then $\lfloor ax \leq b \rfloor$ is a cutting plane. It can be written as $\lfloor a \rfloor x + s = \lfloor b \rfloor$, where s is a slack variable, and added to the IP.

3.2.3 Solving IPs with Relaxations

We can use relaxations and cutting planes to solve integer programs. The idea is to relax the integer program by removing the intergrality constraint. Then solving the new linear program. If the optimal solution is an integer, then we stop. Otherwise, we add a cutting plane and solve again. We repeat until we find an integer solution. This is a really bad way of solving IPs, but oh well.

Algorithm 2 Cutting Planes Algorithm

Input: An IP: $\{\max c^T x : Ax = b, x \in \mathbb{Z}\}$
Output: An optimal solution to the IP: $\bar{x}$

Solve the LP relaxation of the IP. Let $\bar{x}$ be the optimal solution.
while $\bar{x}$ is not integer **do**
  Find a cutting plane s for the IP.
  Add s to the IP.
  Solve the LP relaxation of the IP. Let $\bar{x}$ be the optimal solution.
return $\bar{x}$

Chapter 4

Nonlinear Programming

4.1 Gradients and Subgradients

Definition 4.1.1 (Subgradient). Let $f : \mathbb{R}^n \to \mathbb{R}$ be a convex function. We say s is a subgradient of f at $\bar{x}$ if $h(x) \leq f(x)$ for all $x \in \mathbb{R}^n$, where $h(x) := f(\bar{x}) + s^T(x - \bar{x})$

Definition 4.1.2 (Gradient). Let $f : \mathbb{R}^n \to \mathbb{R}$ be a convex function differentiable with respect to $\bar{x}_i$ for all $i = 1, \ldots, n$. We say $\nabla f(\bar{x}) = \left[\partial_{x_1} f, \ldots, \partial_{x_n} f\right]$ is a gradient of f at $\bar{x}$.

Proposition 4.1.1 (Properties of Gradients and Subgradients). .

- A subgradient at $\bar{x}$ is unique if and only if f is differentiable at $\bar{x}$.
- A subgradient will always exist for a convex function with only bounded operators.
- Any gradient is also a subgradient, but the converse is not true.

4.2 NLP Relaxations

Definition 4.2.1 (Supporting Halfspace). Let $C = \{x : f(x) \leq 0\}$ and $\bar{x} = 0$. Then the supporting halfspace of C is $F = \{x : h(x) \leq 0\}$, where $h(x)$ is the same from Subgradient.

Proposition 4.2.1 (Supporting Halfspace of a Convex Set). Let C, F be as defined above. $C \subseteq F$.

Proof. Let $x \in C$. Then $f(x) \leq 0$. Then $h(x) \leq 0$ since $h(x) \leq f(x)$ for all $x \in \mathbb{R}^n$. Therefore, $x \in F$. ■

Corollary 4.2.1 (NLP Relaxation). Let (NLP) be an NLP in the form $\{\min c^T x : g_i(x) \leq 0, i = 1, \ldots, m\}$. If $g_i(x)$ has a subgradient s at $\bar{x}$, then the NLP obtained by replacing $g_i(x) \leq 0$ with $g_i(\bar{x}) + s^T(x - \bar{x}) \leq 0$ is a relaxation of (NLP).

4.3 KKT Conditions

The KKT approach is a generlization of Lagrange Multipliers, which only allow for equality constraints. KKT allows for both equality and inequality constraints.

Note. We assume that the objective function and all constraints are differentiable at the optimal point.

Definition. **KKT Conditions**

Note that the definitions that follow are all equivalent, although some only apply to NLPs without equality constraints.

Definition 4.3.1 (KKT Conditions for a General NLP)**.** Given a general NLP in the form of: $\{\min f(x) : g_i(x) \leq 0, h_j(x) = 0, i = 1, \ldots, m, j = 1, \ldots, n\}$, The KKT conditions are split into 4 parts:

1. Stationarity: $0 \in \partial f(x) + \sum_{i=1}^{m} \lambda_i \partial g_i(x) + \sum_{i=1}^{n} \mu_i \partial h_i(x)$
2. Primal Feasibility: $g_i(x) \leq 0$ and $h_j(x) = 0$ for all i, j
3. Dual Feasibility: $\lambda_i \geq 0, i = 1, \ldots, m$.
4. Complementary Slackness: $\lambda_i g_i(x) = 0, i = 1, \ldots, m$

Note that λ_i and μ_i are the Lagrange Multipliers for the ith constraint.

Definition 4.3.2 (KKT Conditions for a General NLP II)**.** We can rewrite the KKT conditions to make them a bit more palatable. Let $\bar{x}$ be a solution to the NLP. Also assume that the NLP has a Slater Point. Then the "only" KKT condition is:

$$-\nabla f(x) \in \text{cone}\{\nabla g_i(x), \nabla h_j(x) : i \in T, j = 1, \ldots, n\},$$

where T is the index set of tight constraints at $\bar{x}$.

Definition 4.3.3 (KKT Conditions for an NLP without Equality Constraints)**.** Given an NLP in the form of:
$\{\min f(x) : g_i(x) \leq 0, i = 1, \ldots, m\}$, The KKT conditions are split into 4 parts:

1. Stationarity: $0 \in \partial f(x) + \sum_{i=1}^{m} \lambda_i \partial g_i(x)$
2. Primal Feasibility: $g_i(x) \leq 0$ for all i
3. Dual Feasibility: $\lambda_i \geq 0, i = 1, \ldots, m$.
4. Complementary Slackness: $\lambda_i g_i(x) = 0, i = 1, \ldots, m$

Definition 4.3.4 (KKT Conditions for an NLP without Equality Constraints II)**.** Let $\bar{x}$ be a solution to the NLP. Also assume that the NLP has a Slater Point. Then the "only" KKT condition is:

$$-\nabla f(x) \in \text{cone}\{\nabla g_i(x) : i \in T\},$$

where T is the index set of tight constraints at $\bar{x}$.

Theorem 4.3.1 (KKT Theorem)**.** Let x^* be a solution to an NLP. Then x^* is optimal if and only if the KKT conditions hold.