CO250 Final Exam Guide

Richard Wu

Contents

1 Modelling

Word problems can be modelled or formulated into mathematical programming problem.

In general, to formulate an programming problem, we identify the **variables**, **max or min objective function**, and the **constraints**.

1.1 Linear Programming

A **linear program** has linear (affine) constraints and objective function. The constraints must be:

1. Inequalities/equalities (but not strict inequalities $<, >$)
2. Degree of at most 1 terms

For **multiperiod models** (e.g. modelling oil supply/demand per month, we can introduce t_i variables to denote the leftover oil that is carried over per month, where each month's supply and demand is an equality constraint with supply and demand).

1.2 Integer Programming

Similar to linear programming but there are integral constraints on certain variables. Note one can specify a integer constraint as

$$x_1 \text{ integer}$$

To bound the value e.g. $x_1 \in \{0, 2\}$, one can write

$$\begin{aligned} &z_1 \in \{0, 1\} \text{ or } 0 \leq z_1 \leq 1, z_1 \text{ integer} \\ &x_1 = 2z_1 \end{aligned}$$

1.3 Graph Problems

One can also model graph problems like the minimum cost perfect matching as an LP. For a given graph $G = (V, E)$ with vertices V and edges E, the minimum cost matching problem is

$$min \qquad \sum (c_e x_e : e \in E)$$

subject to

$$\begin{aligned} &\sum (x_e : e \in \delta(v)) = 1 && v \in V \\ &x_e \geq 0 && e \in E \\ &x_e \text{ integer} && e \in E \end{aligned}$$

where c_e is the cost of edge e and $\delta(v)$ is the cut of vertex v.

The **cut** $\delta(v)$ of a vertex v or a set of vertices V are all the edges that have exactly one endpoint in v.

1.4 Nonlinear Programming

Optimization problem of the form

$$min \qquad z = f(x)$$

s.t.

$$\begin{aligned} g_1(x) &\leq 0 \\ g_2(x) &\leq 0 \\ &\vdots \\ g_m(x) &\leq 0 \end{aligned}$$

where $f : \mathbb{R}^n \to \mathbb{R}$ and $g_i : \mathbb{R}^n \to \mathbb{R}$. That is the functions may be nonlinear.

2 Solving Linear Programs

Linear programs can have three outcomes:

Infeasible There are no feasible solutions.

Optimal There are optimal solution(s) that give the optimal value.

Unbounded There is no "best" solution and there is at least one feasible solution.

A solution x is **feasible** if all the constraints hold.

Note that the following propositions and insights are derived by looking at the signage of the constraints and the signage of x.

Since they are specific to the signage and max/min type of the LP, use them with discretion (try to derive/prove the propositions to gain an intuition of why they work).

We will later show how to derive all of these with the Simplex method.

2.1 Infeasibility

For a given system (constraints) $Ax = b, x \geq 0$, the LP is infeasible if there exists a *certificate of infeasibility* or vector y such that:

1. $y^T A \geq 0^T$, and
2. $y^T b < 0$

This follows from the signage of x and the equality.

2.2 Unboundedness

For a given LP $\max\{c^T x : Ax = b, x \geq 0\}$, it is unbounded if there exists *certificate of unboundedness* $(\bar{x}, d)$ where $\bar{x}$ is a feasible solution such that:

1. $Ad = 0$
2. $d \geq 0$
3. $c^T d > 0$

which gives us the unbounded solution $x = \bar{x} + td$, $t \geq 0$. Note that $c^T d < 0$ if the LP is a min problem (to ensure that the solution x tends towards $-\infty$ as $t \to \infty$) and $d \leq 0$ if $x \leq 0$ (so x remains < 0).

2.3 Optimality

For a given LP $\max\{c^T x : Ax = b, x \geq 0\}$ it is optimal if there exists a *certificate of optimality* y such that:

1. $y^T A \geq c^T$, and
2. $c^T x = y^T b$

2.4 Standard Equality Form (SEF)

An LP $\max\{c^T x : Ax = b, x \geq 0\}$ is in Standard Equality Form (SEF) since it is:

1. A **maximization** problem
2. Other than nonnegativity constraints, constraints are equalities
3. Every variable has nonnegativity constraint

The SEF are equivalent (that is SEF of an LP is unbounded *if and only if* the LP is unbounded, similarly for the other states).

To put an LP in SEF:

1. If minimization problem, multiply objective function by -1 and change to maximization
2. Add slack variables for inequalities. If $\alpha x \leq \beta$, introduce $x_{n+1} \geq 0$ where the modified constraint is $\alpha x = \beta$ where $\alpha_{n+1} = 1$. Similarly for $\alpha x \geq \beta$, introduce $x_{n+1} \geq 0$ where $\alpha_{n+1} = -1$.
3. If variables are negative (e.g. $x_i \leq 0$), flip the signs of all the correspond coefficients α_i and make $x_i \geq 0$
4. If variables are free (no constraint on x_i or x_i free), replace x_i with $x_i^+, x_i^- \geq 0$ where $x_i = x_i^+ - x_i^-$. The coefficients corresponding to x_i^+ is $+\alpha_i$ and coefficients corresponding to x_i^- is $-\alpha_i$ (thus the one column A_i becomes two columns).

2.5 Bases and Canonical Form

Given that $A \in \mathbb{R}^{m \times n}$, a basis B for A are n linear independent columns of A (1 for each dimension of $x \in \mathbb{R}^n$).

It can be represented as the set of the column indices e.g. $B = \{1, 2, 4\}$.

The other column indices $\{1, \ldots, n\} \setminus B$ is denoted as N.

Note all bases have a unique basic solution. That is $A_B x = \beta$ (A restricted to the basis B or the columns of A that correspond to the B) has one and only one solution x.

Whether x is feasible depends on the nonnegativity constraint.

The **canonical form** of an LP in SEF $\max\{c^T x + \bar{z} : Ax = b, x \geq 0\}$ satisfies:

1. $A_B = I$ (identity matrix)
2. $c_B = 0$ (objective function coefficients corresponding to B is 0)

One can use row operations or inverse matrices (A_B^{-1}) to derive $A_B = I$. One can subtract a linear combination (vector y) of the LHS of the canonical constraints ($A_B^{-1} Ax - A_B^{-1} b = 0$) to get $c_B = 0$.

More formally, for a given LP in SEF $\max\{z(x) = c^T x + \bar{z} : Ax = b, x \geq 0\}$ and a given basis B of A, we find A_B^{-1} and vector y such that we get the canonical form:

$$max \qquad \bar{z} + y^T b + (c^T - y^T A)x$$

subject to

$$\begin{aligned} &A_B^{-1} Ax = A_B^{-1} b \\ &x \geq 0 \end{aligned}$$

where we want the parentheses expression in the expression to result in $c_B = 0$, or

$$\begin{aligned} &c_B^T - y^T A = 0 \\ &y = A_B^{-T} c_B \end{aligned}$$

Note: This y is the certificate of optimality and infeasibility for the final basis we get when doing **Simplex Iteration** (see below).

Note that the basic solution $\bar{x}$ for a basis B is

$$\begin{aligned} \bar{x_B} &= b \\ \bar{x_N} &= 0 \end{aligned}$$

2.6 Simplex Iteration

Given a feasible basis B (we can derive this later using an auxiliary LP in **Two-Phase Simplex**), the Simplex Algorithm is as follows:

1. Convert the LP to canonical form corresponding to the current basis B

2. If $c_N \leq 0$ (all coefficients in objective function are non-positive), then we found the **optimal solution** (the basic feasible solution).

   Otherwise by **Bland's Rule**, find the first non-neagtive coefficient in objective function $c^T x + \bar{z}$. We choose the correspond x_k variable to pivot on. Set $x_k = t \geq 0$ (we use this t to figure out which x_B we need to act on).

3. If $A_k \leq 0$ (that is all coefficients corresponding to x_k is non-positive), then the LP is **unbounded** (the $d \geq 0$ vector consists of the coefficient corresponding to A_k).

   For each row i of the constraint, there is one $x_{iB} \in x_B$ where $\alpha_{iB} = 1$. For that row, we have $x_{iB} + A_{ik}x_k = b_i$ or $x_{iB} = b_i - tA_{ik}$. We let

   $$t = min\Big\{\frac{b_i}{A_{ik}} : A_{ik} > 0\Big\}$$

   Thus the corresponding x_{iB} leaves the basis and the corresponding x_k enters the basis. We return to step 1 for the new basis.

2.7 Two-Phase Simplex Algorithm/Method

We first need to find the feasible basis of B of A which we can iterate on. To do this, we create an **auxiliary LP** by introducing m additional variables. That is for a given LP in SEF $\max\{c^T x + \bar{z} : Ax = b, x = 0\}$, the auxilliary LP is

$$max \qquad - x_{n+1} - \ldots - x_{n+m}$$

subject to

$$\begin{aligned} &\big(A\big|I_m\big)x = b \\ &x \geq 0 \end{aligned}$$

We perform Simplex iteration on this auxiliary LP to find an optimal solution.

Intuitively if the original basis is feasible, there is a solution where $x_{n+1} = \ldots = x_{n+m} = 0$ giving an optimal value of 0.

That is: if the optimal value is 0, then the original LP is feasible for the final optimal basis B. Otherwise the objective value is < 0 and the original LP is infeasible.

3 Geometry of LPs

Note in LPs, we had linear constraints of the form $\alpha x = \beta$ and WLOG $\alpha x \leq \beta$. They correspond to **hyperplanes** (plane in n-space) and **halfspaces** (the space under the plane), respectively.

3.1 Feasible Regions

Constraints that are $\alpha x \leq \beta$ or halfspaces are interesting since they form **feasible regions** in n-space. That is: the feasible region of an LP is a *polyhedron* or equivalently the intersection of a finite number of halfspaces.

To convert $\alpha x = \beta$ to halfspaces, we can let introduce $\alpha x \leq \beta$ and $-\alpha x \geq -\beta$.

To convert non-negativity constraint $x \geq 0$ to a halfspace, we simply reverse the sign ($-x \leq 0$).

3.2 Convexity

There are certain nice properties to convex feasible regions (which we will see later).

We can define a line through points $x^{(1)}, x^{(2)} \in \mathbb{R}^n$ as a set of points

$$\{x = \lambda x^{(1)} + (1 - \lambda)x^{(2)} : \lambda \in \mathbb{R}\}$$

To constrain this into a line segment between the two points, we limit λ (the factor that creates the linear combination)

$$\{x = \lambda x^{(1)} + (1 - \lambda)x^{(2)} : 0 \leq \lambda \leq 1\}$$

A region or subset C of $\mathbb{R}^n$ is **convex** if the line segment between every point $x^{(1)}, x^{(2)} \in C$ is inside C (we can demonstrate this by algebraic manipulation).

Halfspaces are convex. The intersection of a finite or infinite set of convex sets is also convex. Thus feasible regions for LPs are convex.

3.3 Extreme Points

What is the geometric interpretation of basic feasible solutions? Think of the very extreme points of our feasible region.

$x \in C$ is **not** an **extreme point** of C *if and only if*

$$x = \lambda x^{(1)} + (1 - \lambda)x^{(2)}$$

for some *distinct points* $x^{(1)}, x^{(2)} \in C$ and $0 < \lambda < 1$.

Thus in order for x to be an extreme point, it cannot be *properly contained* in any line segments of C (it must be one of the end points).

Note the points along a boundary curve are extreme points. The point along a straight boundary line is obviously not extreme.

Note for a given polyhedron $P = \{x \in \mathbb{R}^n : Ax \leq b\}$, and $A^= x = b^=$ be the set of tight constraints for $\bar{x} \in P$. Then $\bar{x}$ is an extreme point *if and only if* $rank(A^=) = n$.

Relating this back to LPs, $\bar{x}$ is an extreme point *if and only if* $\bar{x}$ is a basic feasible solution to $Ax = b$.

4 Duality

The purpose of duality is to create a sister (dual) LP to a given LP that helps bound the optimal value. For example, given an LP $\max\{c^T x : Ax = b, x \geq 0\}$, note we previously derived a nice certificate of optimality where $y^T A \geq c^T$ (or $A^T y \geq c$) and $c^T x = y^T b$ (or $x^T c = b^T y$).

Thus to find such optimal solution, we can formulate another LP where $A^T y \geq c$ is one of the constraints and we want to provide an upper bound on $x^T c = b^T y$. Thus the dual for such an LP is

$$min \qquad b^T y$$

subject to

$$\begin{aligned} &A^T y \geq c \\ &y \text{ free} \end{aligned}$$

4.1 Finding the Dual

For a given LP (P) (our **primal**), how exactly do we find the dual (D)? Obviously, the dual must have a complementary objective function(min if original is max and vice versa).

If the objective function of (P) is $c^T x$ (with constraints $Ax = b$), we've shown previously that we'd like $b^T y$ to bound (P), thus $b^T y$ is always our objective function in (D).

The signage of the constraints of (D) require some intuition. Suppose for a (P) that is max $c^T x$, we rewrite it in SEF such that we have something like

$$max \qquad \bar{z} + y^T b + (c^T - y^T A)x$$

Ideally we want $\bar{z} + y^T b$ to be the optimal value, hence

$$(c^T - y^T A)x \leq 0$$

The expression in the parentheses depends on the bound on x:

$x \geq 0$ Then $c^T - y^T A \leq 0$ thus our dual (D) would have constraint $y^T A \geq c^T$ (or $A^T y \geq c$).

$x \leq 0$ $c^T - y^T A \geq 0$ thus our dual (D) would have constraint $A^T y \leq c$.

x **free** $c^T - y^T A = 0$ thus our dual (D) would have constraint $A^T y = c$.

A similar procedure can be used for a primal (P) that is a minimization problem (but with $(c^T - y^T A)x \geq 0$).

To figure out the signage of the dual variables y, we need to perform the same process using the partially constructed dual, where instead of $(c^T - y^T A)x$ we have $(b^T - x^T A)y$.

The results can be summarized in the following table:

$(\mathrm{P}_{\max})$					
max	$c^T x$	$\leq$ constraint	≥ 0 variable	min	$b^T y$
subject to		$=$ constraint	free variable	subject to	
	$Ax \, ? \, b$	$\geq$ constraint	≤ 0 variable		$A^T y \, ? \, c$
	$x \, ? \, \mathbb{0}$	≥ 0 variable	$\geq$ constraint		$y \, ? \, \mathbb{0}$
		free variable	$=$ constraint		
		≤ 0 variable	$\leq$ constraint		

4.2 Weak Duality

Continuing from the maximization primal (P) and minimization dual (D), weak duality states that for feasible solutions $\bar{x}$ and $\bar{y}$ for (P) and (D), respectively

1. $c^T \bar{x} \leq b^T \bar{y}$
2. If $c^T \bar{x} = b^T \bar{y}$, then $\bar{x}$ and $\bar{y}$ is an optimal solution to (P) and (D), respectively

A corollary to this is that if (P) is unbounded, then (D) must be infeasible (since (D) was an upper bound to (P)). Similarly if (D) is unbounded, then (P) is infeasible. Thirdly, if (P) and (D) are both feasible, then they must both have optimal solutions (by Fundamental Theorem of Linear Programming).

4.3 Strong Duality

Very similar to Weak Duality, except it provides a stronger argument that if (P) has an optimal solution $\bar{x}$ then there exists an optimal solution $\bar{y}$ of (D). Moreover, the value of $\bar{x}$ in (P) equals the value of $\bar{y}$ in (D).

4.4 Possible States of Primal-Dual Pair

From Weak and Strong Duality, we have the following possible states:

	Dual		
Primal	Infeasible	Unbounded	Optimal
Infeasible	Yes	Yes	No
Unbounded	Yes	No	No
Optimal	No	No	Yes

4.5 Complementary Slackness

Recall from Weak Duality, we showed that if $c^T \bar{x} = b^T \bar{y}$ for feasible solutions $\bar{x}, \bar{y}$ for (P) and (D) respectively, then they are optimal solutions.

For the primal (P) such as $\max\{c^T x : Ax \leq b\}$ and dual (D) $\min\{b^T y : A^T y = b, y \geq 0\}$, we introduce a slack variable s to (P)

$$\{c^T x : Ax + s = b, s \geq 0\}$$

Assume our feasible solutions $\bar{x}, \bar{y}$ and $\bar{s} = b - A\bar{x}$. Then we have

$$b^T \bar{y} = \bar{y}^T b = \bar{y}^T (A\bar{x} + s) = (\bar{y}^T A)\bar{x} + \bar{y}^T s = c^T \bar{x} + \bar{y}^T s$$

In order to maintain the invariant of Strong Duality, we need

$$\bar{y}^T s = \sum_{i=1}^{m} \bar{y}_i s_i$$

where $\bar{y}_i, s_i \geq 0$ for all i. Thus either (or both) must be 0 in order for our condition to hold.

That is, **Complementary Slackness** states that feasible solutions $\bar{x}, \bar{y}$ are optimal *if and only if* for every row i of A, constraint i is tight for $\bar{x}$ and/or corresponding dual variable $\bar{y}_i = 0$.

The reverse for the dual must hold too.

5 Geometry of Optimal Solutions

With a better understanding of optimal solutions via Weak and Strong Duality, we can investigate the geometric interpretation.

5.1 Cones

A cone generated by $a^{(1)}, \ldots, a^{(k)} \in \mathbb{R}^n$ as the set

$$C = \Big\{ \sum_{i=1}^{k} \lambda_i a^{(i)} : \lambda_i \geq 0 \text{ for all } i = 1, \ldots, k \Big\}$$

Note a feasible solution $\bar{x}$ to the primal (P) is optimal *if and only if* c is in the cone of the tight constraints of $\bar{x}$.

This follows from complementary slackness and how we have either tight constraint for $\bar{x}_i$ or $\bar{y}_i = 0$.

5.2 Farka's Lemma

Essentially, **Farka's Lemma** states a solution either has a feasible solution or there exists a certificate of optimality y. That is one of the following statements hold:

1. The system $Ax = b, x \geq 0$ has a solution
2. There exists a vector y such that $A^T y \geq 0$ and $b^T y < 0$

The proof is done by contradiction and the use of a Primal-Dual pair.

6 Minimum Cost Perfect Matching

6.1 Matching

A matching is a set of edges in a graph where every vertex is incident to at most one edge in the matching.

A perfect matching is a matching where every vertex in the graph is matched.

6.2 Intuition for Lower Bound

In any graph, we want to find the perfect matching (assuming one exists) with the minimum cost given costs c_e per edge $e \in E$.

The proof for why the algorithm works is derived from a primal-dual pair (where primal (P) is a minimization problem for the cost of the edges, and the dual (D) maximizes the cost reduction or the sum of the potentials y).

Intuitively, we want to iteratively reduce the costs of the edges (maintaining that the edges have cost ≥ 0) in the graph by assigning potentials to each vertex and reducing the cost of each edge e by the potentials of its incidental vertices u, v. That is, the *reduced cost* of an edge $e = uv$ is

$$\bar{c}_{uv} = c_{uv} - y_u - y_v$$

Eventually we end up with *equality edges* (edges with a reduced cost of 0).

If the potentials y are feasible ($y_u + y_v \leq c_{uv}$) and all edges of a perfect matching M are equality edges with respect to y, then M is a minimum cost perfect matching.

The respective Primal-Dual pair for min-cost perfect matching is

$$min\{c^T x : Ax = 1, x \geq 0\}$$

where c is the vector of edge costs, and matrix A has vertices as rows and edges as columns and $A_{v,e} = 1$ if the vertex v is an endpoint of e, otherwise 0.

x corresponds to which edges we choose in our perfect matching.

6.3 Bipartite Graphs

A bipartite graph is a graph that be bipartitioned in sets of vertices U, W where every edge has an end in U and an end in W.

Thus a perfect matching must have $|U| = |W|$.

6.4 Hall's Theorem

Let $S \subseteq U$. Denote $N(S)$ as the adjacent vertices of S in W. A **deficient set** S occurs when $|S| > |N(S)|$ in a bipartite graph. It is called deficient because clearly one cannot form a perfect matching since there are not enough edges for the vertices in S to be matched with.

Hall's Theorem states that for a bipartite graph G with bipartition U, W and $|U| = |W|$: there exists a perfect matching M in G *if and only if* there are no deficient sets $S \subseteq U$.

6.5 Min-Cost Perfect Matching Algorithm for Bipartite Graphs

The algorithm proceeds as such:

1. Fix bipartitions U, W. Initialize all potentials $\bar{y}_v$ for $v \in V(G)$ as
$$\bar{y}_v = \frac{1}{2} min\{c_e : e \in E\}$$
2. Construct graph H with vertices V and only equality edges ($\{uv \in E : c_{uv} = \bar{y}_v + \bar{y}_v)$
3. If H has a perfect matching, we are done.
4. Find a deficient set $S \subseteq U$ in H (note S can be just a vertex with no edges, still a deficient set!)
5. If all edges of G incident to S have endpoints only in $N_H(S)$, then S is deficient in G, no perfect matching.
6. Find the minimum reduced cost of an edge incident to S but not in H (that is, $\epsilon = \{c_{uv} - \bar{y}_u - \bar{y}_v : u \in S, v \notin N_H(S)\}$
7. We adjust the potentials of all vertices
$$\bar{y}_v = \begin{cases} \bar{y}_v + \epsilon & \text{for } v \in S \\ \bar{y}_v - \epsilon & \text{for } v \in N_H(S) \\ \bar{y}_v & \text{otherwise} \end{cases}$$
8. Return to step (2)

7 Solving Integer Programs

Note that we now know how to solve for feasible solutions (or infeasible or unbounded or optimal) for LPs. For IPs, we would like to bound this to integral answers only.

The method we will be using is called **cutting planes**. The intuition comes from geometric interpretation of LPs and their feasible regions.

7.1 Convex Hulls

Note for LPs, we have a convex polyhedron (all polyhedra are convex) as the feasible region.

For a set of points$S$ in $\mathbb{R}^n$, the **convex hull** of S (conv(S)) is the smallest convex set that contains S. One can imagine this as the halfspaces that link the outer points in S.

We can form a convex hull inside the feasible region P for the set of integral points S. Note for a polyhedron $P = \{x \in \mathbb{R}^n : Ax \leq b\}$ where all entries A, b are rational, the convex hull of S set of integral points is a polyedron Q also with rational entries.

7.2 LP Relaxation

Note the corresponding convex hull of integral points is called the **LP relaxation**.

Suppose there is an IP $\max\{c^T x : Ax \leq b, x \text{ integer }\}$.

For the convex set for S set of integral points satisfying the constraint $Ax \leq b$ in the IP

$$conv(S) = \{x : A'x \leq b'\}$$

the LP relaxation is an LP $\max\{c^T x : A'x \leq b'\}$. We can thus solve for the integral solutions as an LP problem.

The LP relaxation is equivalent to the original IP and satisfies:

1. IP is infeasible *if and only if* LP relaxation is infeasible
2. IP is unbounded *if and only if* LP relaxation is unbounded
3. every optimal solution to IP is optimal for LP relaxation
4. every optimal solution to LP relaxation that is an extreme point is optimal for IP

The last one is key (not every optimal solution to LP relaxation is optimal for IP).

7.3 Cutting Planes

We want to construct halfspaces that form a convex set around all the integral solutions in the feasible region of our IP.

The algorithm to do so is as follows:

1. Consider LP relaxation of IP problem
2. Solve LP relaxation with Simplex to obtain a final canonical form and basic solution $\bar{x}$.
3. If $\bar{x}$ is infaesible, LP relaxation and thus IP is infeasible. Otherwise $\bar{x}$ is optimal.
4. If $\bar{x}$ is integral, we have found our optimal solution for the IP.
5. Otherwise we need to derive cutting planes to bound our LP. One of the rows (just need one) in the final canonical form has a fractional RHS. Use this row to derive a cutting plane.
6. Add cutting plane constraint to LP and go to step (2).

To find a cutting plane for a given row i where $A_{i1}x_1 + \ldots A_{in}x_n = b_i$ and $x \geq 0$ (this is important as you will see!) the cutting plane is

$$\lfloor A_{i1} \rfloor x_1 + \ldots + \lfloor A_{in} \rfloor x_n = \lfloor b_i \rfloor$$

This works since $x_i \geq 0$ so decreasing the coefficient decreases the LHS (it is $\leq b_i$.

The LHS is now fully integral, thus it has to be $\leq \lfloor b_i \rfloor$.

8 Non-Linear Programming

NLP is hard. We only care for NLPs with convex feasible regions, or from the definition of convexity before and given an NLP

$$min\{c^T x : g_i(x) \leq 0, \forall i \in \{1, \ldots, m\}\}$$

Every $g_i(x)$ is convex (thus the intersection of convex sets is convex).

8.1 Convex Functions

A function $f(x)$ is **convex** if for every pair of points $x^{(1)}, x^{(2)}$

$$f(\lambda x^{(1)} + (1-\lambda)x^{(2)}) \leq \lambda f(x^{(1)}) + (1-\lambda) f(x^{(2)})$$

where $0 \leq \lambda \leq 1$.

Graphically, the line between two points on the function always lies above or on the function between the two points.

We define the **epigraph** of a function f as the shaded region above the function. That is

$$epi(f) = \left\{ \begin{pmatrix} \mu \\ x \end{pmatrix} \in \mathbb{R} \times \mathbb{R}^n : f(x) \leq \mu \right\}$$

Note a function is convex *if and only if* its epigraph is a convex set.

We also define the *level set* as domain (set of x) for a given $\beta \in \mathbb{R}$ where the function is below β. That is

$$\{x \in \mathbb{R}^n : g(x) \leq \beta\}$$

Note that the level set of a convex function is a convex set.

However, if every level set is convex, this does NOT imply the function is convex (e.g. x^3 level sets of negative and positive x values).

8.2 Subgradients

How do we bound the convex set of feasible solutions to an NLP as an intersection of halfspaces so we can solve it as an LP?

We introduce the idea of subgradients (think of them as halfspaces with a slope at a given point x). From calculus (or the equation of a line), we define **subgradient** $s \in \mathbb{R}^n$ of f at $\bar{x}$ where:

$$f(\bar{x}) + s^T(x - \bar{x}) \leq f(x)$$

The affine function on the LHS that we can use as a constraint in our LP relaxation.

We denote the gradient of a function f as ∇f (it will be the subgradient) and it is calculated as

$$\nabla f(x) = [\frac{\partial f(x)}{\partial x_1}, \ldots, \frac{\partial f(x)}{\partial x_n}]^T$$

8.3 Karush-Kuhn-Tucker (KKT) Theorem

A **Slater point** x' for an NLP satisfies $g_i(x') < 0$ (not tight) for every $i \in \{1, \ldots, m\}$.

Combining what we have learned, the KKT Theorem states that for a convex NLP of the form

$$min\{f(x) : g_i(x) \leq 0, \forall i \in \{1, \ldots, m\}\}$$

that has a Slater point, $\bar{x}$ a feasible solution is optimal *if and only if*

$$-\nabla f(\bar{x}) \in cone\{\nabla g_i(\bar{x}) : g_i(\bar{x}) = 0 \text{ (tight) }\}$$

That is the negative of the coefficients in the objective function is in the cone of the constraints tight for $\bar{x}$.

We will show the reverse direction proof: that is $-\nabla f(\bar{x})$ is in the cone of the subgradients of g_i tight for $\bar{x}$, then $\bar{x}$ is optimal.

Note that by the definition of cones for $y_i \geq 0$ (where $J(\bar{x})$ are the row indices tight for $\bar{x}$):

$$-\nabla f(\bar{x}) = \sum_{i \in J(\bar{x})} y_i \nabla g_i(\bar{x})$$

Furthermore from the constraints we have

$$0 \geq g_i(x) = g_i(\bar{x}) + \nabla g_i(\bar{x})(x - \bar{x})$$

Since $g_i(\bar{x}) = 0$ (tight for $\bar{x}$)), we have $\nabla g_i(\bar{x})(x - \bar{x}) \leq 0$.

Combining the two equations, we get

$$f(x) = f(\bar{x}) + \nabla f(\bar{x})(x - \bar{x}) = f(\bar{x}) - \sum_{i \in J(\bar{x})} y_i \nabla g_i(\bar{x})^T (x - \bar{x}) \geq f(\bar{x})$$

where the last inequality holds because $y_i \geq 0$ and $\nabla g_i(\bar{x})^T(x - \bar{x}) \leq 0$.

The proof for the converse ($\bar{x}$ optimal $\rightarrow$ cone equality) **requires a Slater point**.