MA 614: LINEAR AND NONLINEAR PROGRAMMING
COURSE NOTES

1. Introduction
We survey theory, modeling, computation, and applications associated with mathematical techniques of optimization. We begin with linear programming. The standard and canonical forms will be covered first, followed by the techniques used to convert more general problems into these forms. Duality and the simplex method will be covered next. We will cover the computational aspects of the simplex method to complete the first part of the course. In the last part of the course we will cover extensions of linear programming. The topics covered will be chosen from quadratic, stochastic, integer, non-linear, interior point programming and/or particular classes of applications depending on the interests of the class. Financial applications will be emphasized. Each student will be expected to carry out a project, which can be in theory, modeling, computation, or an application of mathematical programming. Students will be provided access to open source and other free optimization software that they can use for future projects and applications.

2. The Generic Problem
The generic mathematical optimization problem is:

\[ \text{Minimize } f(x) \text{ subject to } x \in D \]  \hspace{1cm} (1)

The real valued function \( f \) is called the objective function, and \( D \) the constraint set, or decision set. We choose to minimize “cost.” A simple modeling trick allows us to maximize “profit” instead because \( \text{-Minimum } -f(x) \text{ subject to } x \in D = \text{Maximum } f(x) \text{ subject to } x \in D \). Another simple trick is to observe that \( \text{Minimum } \{ f(x) + c \} = \text{Minimum } \{ f(x) \} + c \), for any constant \( c \). Thus, you will rarely see an objective function with a constant term. There are many such (but more sophisticated)
modeling tricks we will introduce that make our techniques widely applicable.

**Special Cases:**
- Unconstrained optimization
- Equality constrained optimization
- Linear programming
- Combinatorial (0-1) programming
- Integer programming
- Quadratic Programming
- Convex programming
- Non-linear programming
- Stochastic programming
- Functional optimization (dynamic programming, optimal control theory, calculus of variations) Notice the development of this last category predated much of the finite dimensional work.

Even with the trivial structure of (1), we can already say something about it.

---

### 2.1 The Relaxation Lemma

**Lemma:** Let \( f' : D' \to \mathbb{R} \) where \( D' \supseteq D \) and \( f'(x) \leq f(x) \) for \( x \in D \). This is called a relaxation of \( f, D \) (Figure 1). Suppose that \( x^* \in D' \) and that \( f(x^*) \leq f(x) \) for any \( x \in D' \) (that is, \( x^* \) solves (1) with \( f', D' \) replacing \( f, D \)). Then if \( x^* \in D \), and \( f(x^*) = f(x^*) \) then \( x^* \) solves the original problem (1).

**Proof:** By hypothesis, \( x^* \in D \), and since \( f(x^*) \leq f(x) \) for any \( x \in D' \), it must also hold on \( D \) since \( D' \supseteq D \).
Examples:
Integer Programming:

One form of linear programs is:

$$\text{Min } f(x) = cx$$

Subject to $$D' = \{x | Ax = b, x \geq 0\}$$

where \(x\) is an \(n\)-vector of real numbers, \(c\) is an \(n\)-vector of given parameters, \(b\) is an \(m\)-vector of given parameters, and \(A\) is a given \(m \times n\) matrix.

An integer linear program is given by:

$$\text{Min } f(x) = cx$$

Subject to $$D = \{x | Ax = b, x \geq 0, x \text{ integer}\}$$
where \( x \) is an \( n \)-vector of integer numbers, \( c \) is an \( n \)-vector of given parameters, \( b \) is an \( m \)-vector of given parameters, and \( A \) is a given \( m \times n \) matrix.

Then the linear program can be interpreted as a relaxation of the integer program, with \( f = f' \) and \( D \) and \( D' \) as given. Then, if we solve the relaxation (the linear program) and the result happens to be a vector of integers, then the solution of the linear program also solves the integer program.

This is valuable because, in the language of computer science, algorithms for linear programming exist with polynomial time worst case performance, while integer programming is NP hard, which strongly suggests that worst case time is exponential in the size of the data.

Moreover, for important special cases, of integer programming, the usual methods of solving the linear programming relaxation automatically yields integer solutions. For example when the matrix \( A \) is \textit{totally unimodular} this happens. This class includes transportation, assignment, and maximum flow problems, all of which are of practical interest. For a complete discussion of this, see Chapters 19-21 of [Schrijver, 1986].

Even if one is not guaranteed an integer solution of the relaxation, most integer programming solution methods start with solving the linear programming relaxation.

\textbf{Convex Programming}

If \( f \) is a differentiable convex function in some convex domain \( D \), with non-empty interior in \( \mathbb{R}^n \), then for any \( x_0 \) we have \( f(x) \geq L(x, x^0) = f(x^0) + \nabla f(x^0)(x - x^0) \) where \( \nabla f(x^0) \) is the gradient of \( f \) at \( x^0 \). Thus \( L(x, x^0) \) is
a relaxation of \( f \). Even if \( f \) is not differentiable we can always use the subgradient in the same way. So if we are solving (1) where \( f \) is convex and \( D \) is defined by linear constraints, we can try to replace \( f \) by its linearization \( L(x, x^0) \). If we solve the relaxation obtaining \( x^* \), and \( L(x^*, x^0) = f(x^*) \), then we have solved the original problem. Note, in this case, \( x^* = x^0 \). In fact, under general assumptions we have that \( L(x^*, x^*) = f(x^*) \) at the optimum, \( x^* \). This leads to a special case of the Karush-Kuhn-Tucker conditions that we will discuss later in the course.

2.2 Convex/Concave Programming:

**Definition:** Let \( D \subseteq \mathbb{R}^n \) in (1). Then \( x^* \) in \( D \) is said to be a (global) minimum if \( x \in D \) implies \( f(x) \geq f(x^*) \); \( x^* \) is a local minimum if there exists \( \varepsilon > 0 \) so that if \( x \in D \) and \( ||x-x^*|| < \varepsilon \) then \( f(x) \geq f(x^*) \).

**Theorem:** Suppose \( f \), and \( D \) in (1) are convex, suppose further that \( D \) is closed, then any local optimum is a global optimum. On the other hand, if \( f \) is concave and \( D \) contains no lines, then, if the infimum \( \{f(x)|x \in K\} \) is achieved by a point in \( K \), then it is achieved at an extreme point (Figure 2).

![Figure 2: Local Optima](image-url)
**Proof:** Suppose $f$ is convex, and let $x^*$ be a local minimum point. Suppose there is an $x^0$ in $D$ so that $f(x^0) < f(x^*)$. The segment connecting $x^*$ and $x^0$ must be contained in $D$. Then $f(x) \leq (1-\lambda)f(x^*) + \lambda f(x^0) < f(x^*)$ for $0 < \lambda < 1$. Therefore, for any positive $\lambda$ no matter how small, $f(x)$ will be strictly less than $f(x^*)$, and thus $x^*$ does not provide a local minimum—contradiction.

Now suppose $f$ is concave. Let $x^*$ be the point at which an infimum is attained. If $K = \{x^*\}$, is a singleton $x^*$ is an extreme point and we're done. If $x^*$ is an interior point of $K$ then $f$ is constant on $K$ (Lemma A.9). Replace $x^*$ with a boundary point of $K$. There must exist one since $K$ contains no lines. Just pick any point $x$ in $K$, and consider the line through $x$ and $x^*$. In one direction or the other, you must leave $K$ at a boundary point. Therefore, without loss of generality we can assume that $x^*$ is a boundary point of $K$. Let $H$ be a supporting hyperplane for $K$ at the resulting $x^*$. Consider $K' = K \cap H$. $K'$ satisfies the assumptions of the theorem for $K$, that is, $K'$ is closed convex, and contains no lines. However, the dimension of $K'$ is one less than the dimension of $K$. By induction we can repeat the process until we reach a singleton since every extreme point of $K'$ is an extreme point of $K$ (Lemma A.8).

**Corollary:** If $f$ in (1) is linear, any local optimum is a global optimum and if there is a global (local optimum) there is one at an extreme point.

The corollary has strong practical implications for linear programming. It says we need only look at extreme points. Moreover, at an extreme point all we have to do is check local optimality. It will turn out that the constraint set for linear programming has a finite number of extreme points, so this implies that solving linear programs might be relatively easy. So let’s look!

**APPENDIX A: ANALYSIS AND CONVEXITY**

**A Couple of Results from Analysis:**
Theorem A.1: Let $D$ be a closed, bounded (compact) set in $\mathbb{R}^n$. Let
\[ \{x^j \mid j = 0, 1, \ldots \} \]
be an infinite sequence of points in $D$. Then there exists
a subsequence of the sequence, $\{x^i \mid i = 0, 1, \ldots \}$ which converges to
a point in $D$.

Theorem A.2: Let $D$ be a closed, bounded (compact) set in $\mathbb{R}^n$, and let $f$
be a continuous function from $D$ to $\mathbb{R}$. Then $f$ achieves its infimum
(supremum) at a point in $D$.
For proofs see your favorite real analysis book.

Convexity:
First some definitions:

The line through two distinct given points, $x^0$ and $x^1$ in $\mathbb{R}^n$ is:
\[ \{x \mid x = (1-\lambda)x^0 + \lambda x^1, \lambda \in \mathbb{R} \} \]

The half line from $x^0$ through another point $x^1$ in $\mathbb{R}^n$ is:
\[ \{x \mid x = (1-\lambda)x^0 + \lambda x^1, \lambda \geq 0 \} \]

The line segment connecting two distinct given points, $x^0$ and $x^1$ in $\mathbb{R}^n$ is:
\{x \mid x = (1-\lambda)x^0 + \lambda x^1, \quad 0 \leq \lambda \leq 1\}

A set \(A\) in \(\mathbb{R}^n\) is an affine subspace (a translate of a linear subspace) if the line through every two distinct points of \(D\) is in \(D\).

A set \(C\) in \(\mathbb{R}^n\) is a cone at \(x^0\) if \(x^0\) is in \(D\), and the half-line from \(x^0\) to \(x^1\) is in \(D\) for every point \(x^1\) in \(D\). A cone \(C\) is pointed if it contains no non-trivial affine subspace.

A set \(K\) in \(\mathbb{R}^n\) is convex if the segment from \(x^0\) to \(x^1\) is in \(D\) for every \(x^0\) and \(x^1\) in \(D\). \(K\) is bounded if it contains no non-trivial half lines.

**Definition:** \(x\) is an extreme point of a convex set \(K\), if \(x = (1-\lambda)x^0 + \lambda x^1\) with \(x^0 \in K\), \(x^1 \in K\), and \(0 < \lambda < 1\) is only possible when \(x = x^0 = x^1\).

Note that if \(K = \{x\}\), then \(x\) is trivially an extreme point.

A function \(f:D \rightarrow \mathbb{R} (D \subseteq \mathbb{R}^n)\) is convex if its epigraph = \{\(y,x\)\mid \(y \geq f(x)\), \(x \in D\)\} is convex in \(\mathbb{R}^{n+1}\).

![Figure A.2: Epigraph](image_url)
A function \( f: D \rightarrow \mathbb{R} \) (\( D \subseteq \mathbb{R}^n \)) is **concave** if \(-f\) is convex.

**Proposition A.3:** A function \( f: D \rightarrow \mathbb{R} \) is convex iff \( D \) is convex and for all \( x^0 \in D, \ x^1 \in D, \) and \( 0 \leq \lambda \leq 1, \) we have \( f(x) \leq (1-\lambda)f(x^1) + \lambda f(x^0). \)

**Proof:** Suppose \( D \) is not convex, then the epigraph of \( f \) cannot be convex. If \( f(x) > (1-\lambda)f(x^1) + \lambda f(x^0) \) for some \( x^0 \in D, \ x^1 \in D, \) then the segment between \((f(x^1), x^1), \) and \((f(x^0), x^0)\) is not contained entirely in the epigraph. Suppose for all \( x^0 \in D, \ x^1 \in D, \) and \( 0 \leq \lambda \leq 1, \) we have \( f(x) \leq (1-\lambda)f(x^1) + \lambda f(x^0), \) and let \((y^0, x^0), \) and \((y^1, x^1)\) be two points in the epigraph of \( f, D. \) Finally, let \( y_\lambda = (1-\lambda)y^1 + \lambda y^0, \) and \( x_\lambda = (1-\lambda)x^1 + \lambda x^0. \) Then \( y_\lambda = (1-\lambda)y^1 + \lambda y^0 \geq (1-\lambda)f(x^1) + \lambda f(x^0) \geq f(x_\lambda). \)

**Corollary A.4:** A function \( f: D \rightarrow \mathbb{R} \) is convex iff \( D \) is convex and for all \( x^0 \in D, \ x^1 \in D, \) and \( 0 \leq \lambda \leq 1, \) we have \( f(x) \geq (1-\lambda)f(x^1) + \lambda f(x^0). \)

**Quadratic Forms:**
Note for all quadratic forms $x'Qx$ we will always assume that $Q$ is real symmetric. We can do this because our data, and decision variables will always be real, and $x'Qx = \frac{1}{2} x'(Q + Q')x$.

From matrix theory, we know that any real symmetric matrix, $Q$, can be written in the form $Q = E'DE$ where $D$ is a real diagonal matrix (that is, $d_{ij} = 0$ if $i \neq j$, and $d_{ii}$ is real) and $E$ is a non-singular matrix with $E^\top = E'$.

A square matrix $Q$ is positive semi-definite if $x'Qx \geq 0$ for all $x$. It is positive definite if $x'Qx > 0$ for all non-zero $x$.

**Proposition A.5:** A positive semi-definite matrix $Q$ is positive definite if and only if it is non-singular.

**Proof:** If $Q$ is singular, there exists a non-zero $x$ so that $Qx = 0$.

Therefore, for this $x$ $x'Qx = 0$, and $Q$ is not positive definite. On the other hand suppose $x'Qx = 0$. We can write this as $x'E'DEx = 0$. Let $y = Ex \neq 0$, then $y' = x'E'$, and $y'Dy = \sum d_{ii}y_i^2 = 0$. This implies $d_{ii} = 0$ for some $i$. Now let $y = e_i$, the $i$th unit vector. Then $y'Dy = 0$. Finally, let $x = E'y$.

Then $x'Qx = y'EQE'y = y'EE'DEE'y = y'Dy = 0$.

**Definition:** $H = \{x | \pi x = p\}$ is a supporting hyperplane for $x^0$ in a convex set $K$ if $\pi x^0 = p$, and $\pi x \geq p$ for all $x \in K$. If $K \not\subset H$ then the supporting hyperplane is proper.

**Lemma A.6:** Let $K$ be a closed convex set and $y \not\in K$. Then there exist $\pi$ such that $\inf\{\pi x | x \in K\} > \pi y$.

**Proof:** The euclidean norm $\|y - x\|$ is a continuous function of $x$ hence it attains its minimum in $K$, at say $x^0$. Clearly $x^0$ is a boundary point of $K$. Let $\pi = x^0 - y$. Then $\pi y = -(x^0 - y)(x^0 - y) + \pi x^0 < \pi x^0$ since $y \neq x^0$.

On the other hand if $x \in K$, then $x^0 + \lambda(x - x^0) \in K$ for $0 \leq \lambda \leq 1$. Squaring
we obtain after cancellations and dropping the common factor $\lambda \leq -2(y - x^0)(x- x^0) + \lambda(x - x^0)^2$. Letting $\lambda \to 0$ we have $(x^0 - y)(x - x^0) \leq 0$ hence $\pi x \geq \pi x^0 > \pi y$.

**Theorem A.7:** Let $K$ be a convex set and $y$ a boundary point then there exists $\pi$ such that $\pi y = \inf \{\pi x | x \in K\}$, i.e., through every boundary point of $K$ there passes a supporting hyperplane ($p = \pi y$).

**Proof:** Let $y^j$ approach $y$ such that $y^j \not\in K$. Then there exists $\pi^j$ satisfying the lemma. We can assume without loss of generality that $\|\pi^j\| = 1$. Hence there exists $\pi$ which is an accumulation point of the $\pi^j$. Since $\pi^j(x - y^j) > 0$ for all $j$ and $x \in K$ we have $\pi(x - y) \geq 0$.

**Lemma A.8:** Let $H = \{x | \pi x = p\}$ be a supporting hyperplane for $D$ and let $D' = H \cap D$. Then any extreme point of $D'$ is an extreme point of $D$.

**Proof:** Let $x^*$ be an extreme point of $D$. Suppose $x^*$ is not an extreme point of $D$, then there is $x^0, x^1$, with $x^0 \in D, x^1 \in D, x^0 \neq x^1, 0 < \lambda < 1$ so that $x^* = (1 - \lambda)x^0 + \lambda x^1$. But then we have $p = \pi x^* = (1 - \pi)\pi x^0 + \lambda \pi x^1$.

Since $x^0$ is not in $H$, but is in $D$, $\pi x^0 > p$, then we must have $\pi x^1 < p$ which implies $x^1 \not\in D$, contradiction.

**Lemma A.9:** Let $f$ be a concave function $f: K \to R$, where is convex. If $f$ achieves its infimum relative to $K$ at some point of the interior of $K$, then $f$ is actually constant throughout $K$.

**Proof:** Let $x^* \in K$ be a point at which $f$ achieves its infimum. Let $x^i$ be any other point in $K$. Because $x^*$ is an interior point, there exists $x^0$ and $x^i$ so that $x^0 \neq x^i, 0 < \lambda < 1$, and $x^* = (1 - \lambda)x^0 + \lambda x^i$. By concavity: $f(x^*) \geq (1 - \lambda)f(x^0) + \lambda f(x^i)$. If $f(x^0) = f(x^i) = f(x^*)$ does not hold then $f(x^0)$ or $f(x^i)$ is strictly less than $f(x^*)$ contradicting the hypothesis that $x^*$ achieves the infimum.

**Convex Functions:**
Let $f$ be convex, and $f: D \to \mathbb{R}$, with $D$ closed and convex. Let $E$ be the epigraph of $f$, $D$. Let $x^0 \in D$, then $<f(x^0), x^0>$ is a boundary point of $E$. Then $E$ has a supporting hyperplane defined by, say, $(\sigma, \pi)$ at $<f(x^0), x^0>$ meaning (using the definition of an epigraph) that $\sigma y + \pi^0 x \geq \sigma f(x) + \pi^0 x^0$ for all $x \in K$. Note that $\sigma > 0$, otherwise $y$ could be arbitrarily small. So we divide through by $\sigma$ and change the sign of the remaining multipliers ($\pi = -\pi^0/\sigma$). This yields a linear function $h(x, x^0) = f(x^0) + \pi (x-x^0)$ which is said to be a subgradient of $f$ at $x^0$ if $h(x^0, x^0) = f(x^0)$, and $h(x, x^0) \leq f(x)$ for all $x \in K$. We have just showed that a subgradient exists for each point $x$ in $K$.

![Figure A.4: Subgradient](image)

We state without proof [Rockafellar, 1970, p. 242].

**Theorem A.10:** If $f$ is convex on a convex set $D$ with non-empty interior, it is differentiable at $x$ if and only if it has a unique subgradient at $x$.

Another useful theorem we state without proof is (see for example, [Bertsekas, 1995, p. 561-562]).
**Theorem A.11:** If $f$ is continuously differentiable then the conditions (a) and (b) below are equivalent. If $f$ is twice differentiable, then the conditions (a), (b) and (c) below are equivalent:

(a) $f$ is convex

(b) For all $x$, and $g$ in $D$: $f(y) \geq f(x) + \nabla f(x)(y-x)$

(c) For all $x$ the Hessian $\nabla^2 f(x)$ is a positive semi-definite matrix.

**REFERENCES:**