726: Nonlinear Optimization I
Mathematical Background Material
Fall, 2008

Important: Review Appendix A of [3] and make sure you understand it. The text below hits a few important topics from that material and adds a few more relevant items. Some of it will be covered in class but it is up to you to fill in the rest. Many of the things discussed here will be referred to during the semester.

1. Linear Algebra.

DEFINITION 1.1. *The vectors $x_1, x_2, \ldots, x_m$ in $\mathbb{R}^n$ are* linearly dependent *if there exist scalars $\alpha_1, \alpha_2, \ldots, \alpha_m$, not all zero, such that $\sum_{i=1}^m \alpha_i x_i = 0$. Otherwise, they are linearly independent.*

LEMMA 1.2. *Suppose that $y_1, y_2, \ldots, y_{m+1}$ are vectors in $\mathbb{R}^n$ that are each a linear combination of the vectors $x_1, x_2, \ldots, x_m$. Then $y_1, y_2, \ldots, y_{m+1}$ are linearly dependent.*

Proof. See Mangasarian [2, p. 177]. □

COROLLARY 1.3.

- *Any $m > n$ vectors in $\mathbb{R}^n$ are linearly dependent.*
- *Consider the system $Ax = b$, where $A \in \mathbb{R}^{m \times n}$ with $m < n$. If this system has a solution, then it has infinitely many solutions.*

Proof. First statement follows by noting that each of the m vectors can be expressed as linear combinations of the unit vectors e_i, $i = 1, 2, \ldots, n$ where e_i is the vector with all 0s except for a 1 in the ith position. Second statement follows from the fact that $Az = 0$ has infinitely many solutions. □

DEFINITION 1.4. *A set $S \subset \mathbb{R}^n$ is a* subspace *if for any $x, y \in S$ we have $\alpha x + \beta y \in S$ for all scalars α and β.*

DEFINITION 1.5. *Given a set $S \subset \mathbb{R}^n$, a* basis *is the maximal set of linearly independent vectors that can be chosen from S.*

Bases are used mostly when S is a subspace.

LEMMA 1.6. *The linearly independent vectors $x_1, x_2, \ldots, x_r$ are a basis for S if and only if any vector $y \in S$ is a linear combination of the x_is.*

DEFINITION 1.7. *Given a matrix $A \in \mathbb{R}^{m \times n}$, the* rank *of A is the maximum number of linearly independent rows in A.*

This is the same as the number of linearly independent columns, so $\text{rank}(A) = \text{rank}(A^T)$.

LEMMA 1.8. *If $A \in \mathbb{R}^{r \times n}$ with $r \le n$ has (full) rank r, then the system $Ax = b$ has a solution for any b. If in addition $r < n$, it has infinitely many solutions.*

DEFINITION 1.9. *A matrix $A \in \mathbb{R}^{n \times n}$ is* nonsingular *if it has rank n and* singular *otherwise.*

A is nonsingular if and only if $Ax = b$ has a unique solution for any b.

DEFINITION 1.10. *A matrix $A \in \mathbb{R}^{n \times n}$ is* positive definite *if $x^T A x > 0$ for any $x \neq 0$. It is* positive semidefinite *if $x^T A x \ge 0$ for any x.*

Any positive definite matrix A is nonsingular.

LEMMA 1.11. *If A is symmetric positive semidefinite, the matrix*

$$\begin{bmatrix} A & B^T \\ -B & 0 \end{bmatrix}$$

is positive semidefinite for any B (but symmetric only if B is vacuous or zero). The *matrix*

$$\begin{bmatrix} A & B^T \\ B & 0 \end{bmatrix}$$

is symmetric but not positive semidefinite unless B is vacuous or zero.

Proof. Given any vectors x and y of appropriate dimensions, we have for the first claim that

$$\begin{bmatrix} x \\ y \end{bmatrix}^T \begin{bmatrix} A & B^T \\ -B & 0 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = x^T A x \geq 0.$$

For the second claim, let the indices i and j be such that $B_{ij} \neq 0$. Then if we do symmetric pivoting to bring rows/columns j and $n+i$ to the top left 2×2 submatrix we have that this submatrix has the form

$$\begin{bmatrix} A_{jj} & B_{ij} \\ B_{ij} & 0 \end{bmatrix}$$

Since it is symmetric, this matrix has real eigenvalues $(A_{jj} \pm \sqrt{A_{jj}^2 + 4B_{ij}^2})/2$, one of which is positive and the other of which is negative. Choosing z to be the eigenvector in $\mathbf{R}^2$ that corresponds to this negative eigenvalue, we have that

$$\begin{bmatrix} z \\ 0 \end{bmatrix}^T P^T \begin{bmatrix} A & B^T \\ B & 0 \end{bmatrix} P \begin{bmatrix} z \\ 0 \end{bmatrix} = z^T \begin{bmatrix} A_{jj} & B_{ij} \\ B_{ij} & 0 \end{bmatrix} z < 0,$$

demonstrating that the matrix in question is not positive semidefinite. □

QR Decomposition. Given any matrix $A \in \mathbf{R}^{m \times n}$ with $m \geq n$, there is an $n \times n$ permutation matrix P, an $m \times m$ orthogonal matrix Q and an $m \times m$ upper triangular matrix R with nonnegative diagonals such that

$$AP = Q \begin{bmatrix} R \\ 0 \end{bmatrix} = \begin{bmatrix} Q_1 & Q_2 \end{bmatrix} \begin{bmatrix} R \\ 0 \end{bmatrix} = Q_1 R.$$

(The zero matrix has dimension $(n-m) \times m$, and Q_1 is $n \times m$.) When A has full (column) rank, R has all positive diagonal elements. When A is rank deficient (say, rank r), R has the form $R = \begin{bmatrix} R_1 & R_2 \\ 0 & 0 \end{bmatrix}$, where R_1 is $r \times r$ upper triangular with positive diagonal elements and R_2 is $r \times (m-r)$. In this case we can write the factorization as

$$AP = Q \begin{bmatrix} R \\ 0 \end{bmatrix} = \begin{bmatrix} \bar{Q}_1 & \bar{Q}_2 \end{bmatrix} \begin{bmatrix} R_1 & R_2 \\ 0 & 0 \end{bmatrix} = \bar{Q}_1 \begin{bmatrix} R_1 & R_2 \end{bmatrix}.$$

DEFINITION 1.12. *Given a matrix $A \in \mathbf{R}^{m \times n}$, the* range space *of A is the set $R(A) \stackrel{\text{def}}{=} \{Av \,|\, v \in \mathbf{R}^n\}$. (It is also known as the* image of A *and is a subspace of $\mathbf{R}^m$.) The* null space $N(A)$ *is the set of all vectors z such that $Az = 0$. (It is also known as the* kernel of A *and is a subspace of $\mathbf{R}^n$.)*

THEOREM 1.13. *(Fundamental Theorem of Algebra.) If $A \in \mathbf{R}^{m \times n}$, $R(A) \oplus N(A^T) = \mathbf{R}^m$.*

Given a matrix $A \in R^{m \times n}$ with $m < n$ and full rank m, find a matrix $Z \in \mathbf{R}^{n \times (n-m)}$ that spans the null space $N(A)$, that is, $N(A) = \{Zy \,|\, y \in \mathbf{R}^{n-m}\}$.

Method 1. Perform a column permutation of A described by permutation matrix P, such that

$$AP = \begin{bmatrix} A_1 & A_2 \end{bmatrix},$$

where A_1 is $m \times m$ nonsingular and A_2 is $m \times (n-m)$. Then set $Z = P \begin{bmatrix} A_1^{-1} A_2 \\ -I \end{bmatrix}$.

Method 2. Perform a QR decomposition of A^T, to get

$$A^T P = \begin{bmatrix} Q_1 & Q_2 \end{bmatrix} \begin{bmatrix} R \\ 0 \end{bmatrix}.$$

Then $AQ_2 = PP^T AQ_2 = P[Q_2^T A^T P] = 0$. Since Q_2 has linearly independent columns and the maximal number of them $(n-m)$ its columns span $N(A^T)$.

Norms on $\mathbf{R}^n$: Euclidean norm (a.k.a. 2-norm): $\|x\|_2 = \sqrt{\sum_{i=1}^n x_i^2}$; 1-norm: $\|x\|_1 = \sum_{i=1}^n |x_i|$; ∞-norm: $\|x\|_\infty = \max_{i=1,2,\ldots,n} |x_i|$.

THEOREM 1.14. *Let H and A be matrices with the following properties:*

(i) H is symmetric;

(ii) A has full row rank;

(iii) If Z is any matrix whose columns span $N(A)$, then $Z^T HZ$ is positive definite.

Then the following matrix is nonsingular:

$$\begin{bmatrix} H & A^T \\ A & 0 \end{bmatrix}.$$

Proof. We prove the result by showing that $(x, y) = (0, 0)$ is the only possible solution of the following linear system:

$$\begin{bmatrix} H & A^T \\ A & 0 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}.$$

Since $Ax = 0$, we have $x \in N(A)$ and so $x = Zu$ for some vector u. Hence, using the fact that $AZ = 0$ and that $Z^T HZ$ is nonsingular, we have

$$Hx + A^T y = 0 \;\Rightarrow\; HZu + A^T y = 0 \;\Rightarrow\; Z^T HZu = 0 \;\Rightarrow\; u = 0 \;\Rightarrow\; x = 0.$$

Hence,

$$Hx + A^T y = 0 \;\Rightarrow\; A^T y = 0.$$

Since A^T has full column rank, we have $y = 0$, completing the proof. □

Matrix norms: For $p = 1, 2, \infty$, induced norm is

$$\|A\|_p = \max_{x \neq 0} \|Ax\|_p / \|x\|_p.$$

Frobenius norm is

$$\|A\|_F \sqrt{\sum_{i=1}^n \sum_{j=1}^n A_{ij}^2}.$$

2. Topology.

Basic background on topology of $\mathbb{R}^n$.

DEFINITION 2.1. *Given $x \in \mathbb{R}^n$, and the open ball of radius ϵ around x is defined by $B_\epsilon(x) = \{z \mid \|z - x\|_2 < \epsilon\}$.*

DEFINITION 2.2. *A set $x \in \mathbb{R}^n$ is in the interior of $\Gamma \subset \mathbb{R}^n$ (denoted $x \in \mathrm{int}(\Gamma)$ if there is $\epsilon > 0$ such that $B_\epsilon(x) \subset \Gamma$.*

DEFINITION 2.3. *The* closure *of the set $\Gamma \subset \mathbb{R}^n$ (denoted $\mathrm{cl}(\Gamma)$) is the set of points x with the property that $B_\epsilon(x) \cap \Gamma \neq \emptyset$ for all $\epsilon > 0$.*

DEFINITION 2.4. *A set $\Gamma \subset \mathbb{R}^n$ is* open *if $x \in \mathrm{int}(\Gamma)$ for all $x \in \Gamma$. Γ is* closed *if $\Gamma = \mathrm{cl}(\Gamma)$.*

DEFINITION 2.5. *Given two sets $\Gamma \subset \Lambda \subset \mathbb{R}^n$, we say that Γ is open (closed) relative to Λ if there is some open (closed) set Σ such that $\Gamma = \Lambda \cap \Sigma$.*

Example: The set $\{(0, t_2) \mid t_2 > 0\}$ is open relative to $\{0\} \times \mathbb{R}$, though it is not itself open. (Take $\Sigma = \{(t_1, t_2) \mid t_1 \in \mathbb{R}, t_2 > 0\}$.) The set $(0, 1) \subset \mathbb{R}$ is closed relative to itself; take $\Sigma = [0, 1]$.

THEOREM 2.6.

- *Every union of open sets is open.*
- *Every finite interesection of open sets is open.*
- *0 and $\mathbb{R}^n$ are open.*
- *Every intersection of closed sets is closed.*
- *Every finite union of closed sets is closed.*
- *0 and $\mathbb{R}^n$ are closed.*

The last three results follow immediately from the first three if we use the fact that the complement of any open set is closed.

3. Sequences.

Now discuss sequences $\{x^k\}_{k=1}^\infty$ of real vectors. A subsequence is defined by an infinite subset of integers $\mathcal{K} \stackrel{\text{def}}{=} \{k_1, k_2, k_3, \ldots\}$, such that $k_1 < k_2 < k_3 < \cdots$. Write the subsequence as $\{x^{k_j}\}_{j=1}^\infty$ or as $\{x^k\}_{k \in \mathcal{K}}$.

$\bar{x} \in \mathbb{R}^n$ is the *limit* of the sequence $\{x^k\}$ if for any $\epsilon > 0$ there is an integer K such that $\|x^k - \bar{x}\| \le \epsilon$ for all $k \ge K$. Write

$$\lim_{k \to \infty} x^k = \bar{x}.$$

We say that $\{x^k\}$ *converges* to $\bar{x}$. Example: $x^k = 1/k$ has limit 0.

$\bar{x} \in \mathbb{R}^n$ is an *accumulation point* of the sequence $\{x^k\}$ if for any $\epsilon > 0$ and any positive integer K, there is some $k > K$ such that $\|x^k - \bar{x}\| \le \epsilon$. Example: $x^k = (-1)^k + 1/2^k$ has accumulation points -1 and 1, but no limit. Example: $x^k = k\pi - \lfloor k\pi \rfloor$ has every point in $[0, 1]$ as an accumulation point (but has no limit).

THEOREM 3.1.

(i) If $\bar{x}$ is an accumulation point, there is a subsequence for which $\bar{x}$ is the limit.

(ii) If $\bar{x}$ is the limit, the sequence can have no other accumulation points.

If $\bar{x} \in \mathrm{cl}(C)$ for some set C, there is a sequence $\{x^k\}$ of points in C for which $\bar{x}$ is the limit. Any accumulation point of a sequence $\{x^k\}$ in C must be in $\mathrm{cl}(C)$.

Now consider sequences of real numbers $\{\alpha_k\}$.

We say that α_L is the lim inf of $\{\alpha_k\}$ (written $\liminf_k \alpha_k = \alpha_L$) if

$$\alpha_L = \lim_{n \to \infty} (\inf_{k \ge n} \alpha_k).$$

Similarly, we have

$$\limsup_k \alpha_k = \lim_{n\to\infty} (\sup_{k\geq n} \alpha_k).$$

AXIOM 1. *Any nonempty set of real numbers Λ which has a lower (upper) bound has a greatest lower (least upper) bound.*

THEOREM 3.2. *Every bounded nondecreasing sequence of real numbers has a limit.*

Proof. Let β be an upper bound on $\{\alpha_k\}$. By the axiom, we can choose β to be the least possible upper bound. For any $\epsilon > 0$ there is an integer K such that $x^K > \beta - \epsilon$ (otherwise, we could decrease β to $\beta - \epsilon$, which contradicts the choice of β as the least upper bound). Because $\{\alpha_k\}$ is nondecreasing, we must have that

$$\beta \geq x^k \geq x^K \geq \beta - \epsilon, \qquad \text{for all } k \geq K.$$

Hence, β is in fact the limit of $\{\alpha_k\}$, as claimed. □

A sequence converges to a limit if and only if it is a *Cauchy sequence*, that is, one for which for all $\epsilon > 0$ there is an integer K such that $\|x_n - x_m\| \leq \epsilon$ for all $m, n \geq K$.

A set $\Lambda \subset \mathbb{R}^n$ is *bounded* if there is $\beta \geq 0$ such that $\|x\| \leq \beta$ for all $x \in \Lambda$.

A *compact set* Λ is defined by any one of the following equivalent properties:

(i) Λ is closed and bounded.

(ii) Every sequence of points in Λ has an accumulation point in Λ.

(iii) For every family of open sets Λ_i, $i = 1, 2, 3, \ldots$ such that $\Lambda \subset \Lambda_1 \cup \Lambda_2 \cup \Lambda_3 \cup \cdots$, there is a finite subfamily $i_1, i_2, i_3, \ldots, i_m$ such that $\Lambda \subset \Lambda_{i_1} \cup \Lambda_{i_2} \cup \Lambda_{i_3} \cup \cdots \cup \Lambda_{i_m}$.

Rates of convergence. Consider a sequence $\{x^k\}$ that converges to a limit x^*. We say that the convergence is *Q-linear* if $\lim sup \|x^{k+1} - x^*\|/\|x^k - x^*\| \leq r$ for some $r < 1$. We say that it is Q-superlinear if $\lim \|x^{k+1} - x^*\|/\|x^k - x^*\| = 0$. We say that it is *Q-quadratic* if $\|x^{k+1} - x^*\|/\|x^k - x^*\|^2$ is bounded. More generally, we say that it has Q-order τ for any $\tau > 1$ if

$$\liminf(\log \|x^{k+1} - x^*\| / \log \|x^k - x^*\|) \geq \tau.$$

We say the sequence $\{x^k\}$ converges *R-linearly* (resp. *R-superlinearly*, *R-quadratically*) to x^* if there is a sequence $\{\alpha_k\}$ of real numbers such that $\|x^k - x^*\| \leq \alpha_k$ and α_k converges Q-linearly (resp. *Q-superlinearly*, *Q-quadratically*) to zero.

4. Linear Programming.

Consider the linear program in standard form:

$$\min_x c^T x \quad \text{subject to} \quad Ax = b, \;\; x \geq 0, \tag{4.1}$$

for which the dual form is:

$$\max_u b^T u \quad \text{subject to} \quad A^T u \leq c. \tag{4.2}$$

There is a rich mathematical theory known as *duality theory* that relates these two problems. This theory is also useful in the construction of some algorithms, e.g. primal-dual interior-point methods.

THEOREM 4.1 (Weak Duality). *If x is feasible for (4.1) and u is feasible for (4.2), then $c^T x \geq b^T u$.*

Proof.

$$b^T u = (Ax)^T u = x^T A^T u \leq x^T c,$$

where the last inequality follows from $A^T u - c \leq 0$ and $x \geq 0$. □

THEOREM 4.2 (Strong Duality). *Exactly one of the following three statements is true:*

(i) Both primal and dual problems are feasible and both have optimal solutions with equal extrema.

(ii) Exactly one of the problems is infeasible and the other has unbounded objective on its feasible region.

(iii) Both problems are infeasible.

This result is proved in [1] by invoking the simplex method with a pivot rule that prevents cycling.

We say that a function $f(x)$ "attains its minimum" over a feasible region C if there exists $x^* \in C$ such that

$$f(x^*) = \inf_{x \in C} f(x).$$

A function can be bounded below and yet not attain its minimum. When C is not closed this is obvious; for example if $f(x)$ is any monotonically increasing function and $C = (0, 1]$. It can also happen for closed C; for example $f(x) = e^{-x}$ and $C = \{x \,|\, x \geq 0\}$.

COROLLARY 4.3. *Suppose the linear program (4.1) is feasible and bounded below. Then it attains its minimum.*

Proof. Case (iii) of the strong duality result does not hold since the primal problem (4.1) is feasible. Case (ii) does not hold either, because the primal is bounded below. Therefore case (i) holds, so we conclude that (4.1) has a solution x^* that attains the minimum. □

The Karush-Kuhn-Tucker (KKT) conditions give another important way of relating and recognizing primal and dual solutions.

THEOREM 4.4 (KKT conditions). *x solves (4.1) and u solves (4.2) if and only if x and u satisfy the following relationships:*

$$Ax = b, \quad x \geq 0, \quad A^T u \leq c,$$

(feasibility) and

$$x^T (A^T u - c) = 0$$

(complementarity).

5. Convex Sets and Projections. (From Robinson [4].)

A set $C \subset \mathbb{R}^n$ is *convex* if for each x and y in C and each $\lambda \in [0, 1]$, we have $(1 - \lambda)x + \lambda y \in C$.

A *half space* is a set of the form $\{x \,|\, \langle x, y \rangle \geq \eta\}$ where $y \in \mathbb{R}^n$ and $\eta \in \mathbb{R}$ are constants. A *hyperplane* is a set of the form $\{x \,|\, \langle x, y \rangle = \eta\}$.

A *polyhedral convex set* in $\mathbb{R}^n$ is the intersection of finitely many half-spaces.

A *cone* is a set with the property that $x \in K \Rightarrow \alpha x \in K$ for all $\alpha \geq 0$. A *convex cone* is a cone that is a convex set.

If C is a convex set, a vector y is said to be *normal* to C at x if for all $c \in C$ we have $\langle y, c - x \rangle \leq 0$.

The *normal cone* to C at a point x, denoted $N_C(x)$, is the set of all y that are normal to C at x. (If $x \notin C$, then $N_C(x) = \emptyset$.)

If K is a nonempty cone, the *polar* of K is defined by

$$K^\circ = \{x \mid \langle x, k\rangle \le 0 \ \text{ for all } k \in K\}.$$

The polar is always closed (by continuity of the inner product).

If K and L are two cones in $\mathbf{R}^n$ with $\emptyset \ne K \subset L$, then $L^\circ \subset K^\circ$.

The *tangent cone* to a convex set C at a point $x \in C$ is $T_C(x) = N_C(x)^\circ$.

In the text below, we assume $\|\cdot\| = \|\cdot\|_2$.

If S is any subset of $\mathbf{R}^n$, the *point-to-set distance to S* is defined by

$$d_S(x) = \inf\{\|x - s\| \mid s \in S\}. \tag{5.1}$$

If $S = \emptyset$ then $d_S(x) = \infty$ for all x.

If S is nonempty and closed, then the infimum in the formula above is attained. To see this, let s_0 be any element of S. Then

$$d_S = \inf\{\|x - s\| \mid s \in S,\ \|x - s\| \le \|x - s_0\|\}.$$

The infimum is taken over a compact set, so since $\|x - \cdot\|$ is continuous, it attains its minimum. (The minimum may be attained at more than one point, in general)

When S is convex, the minimum in (5.1) is attained at one point only, as we show below. We show first some properties of the point that attains the minimum.

Lemma 5.1. *Let C be a convex set in $\mathbf{R}^n$, and let $x_0 \in \mathbf{R}^n$. The function $\|x_0 - (\cdot)\|$ attains its minimum on C at a point $c_0 \in C$ if and only if for each $c \in C$, we have $\langle x_0 - c_0, c - c_0\rangle \le 0$.*

Proof. Note first that

$$\begin{aligned} &\|x_0 - c\|^2 \\ &\quad = \|(x_0 - c_0) - (c - c_0)\|^2 = \|x_0 - c_0\|^2 - 2\langle x_0 - c_0, c - c_0\rangle + \|c - c_0\|^2. \end{aligned} \tag{5.2}$$

($\Leftarrow$) Assume first that $\langle x_0 - c_0, c - c_0\rangle \le 0$. Rearranging (5.2), we have

$$\begin{aligned} 0 &\ge \langle x_0 - c_0, c - c_0\rangle \\ &= \frac{1}{2}\left(\|x_0 - c_0\|^2 - \|x_0 - c\|^2\right) + \frac{1}{2}\|c - c_0\|^2 \ge \frac{1}{2}\left(\|x_0 - c_0\|^2 - \|x_0 - c\|^2\right). \end{aligned}$$

Therefore, c_0 minimizes $\|x_0 - (\cdot)\|$ over C.

($\Rightarrow$) Assume that $\|x_0 - (\cdot)\|$ attains its minimum on C at a point $c_0 \in C$. Let c_1 be any point in C. Since C is convex, we have that for all $\lambda \in (0, 1)$ that $c(\lambda) \stackrel{\text{def}}{=} (1 - \lambda)c_0 + \lambda c_1 \in C$. Using (5.2) and the fact that c_0 is the minimizer, we have

$$\begin{aligned} 0 &\ge \|x_0 - c_0\|^2 - \|x_0 - c(\lambda)\|^2 \\ &= 2\langle x_0 - c_0, c(\lambda) - c_0\rangle - \|c(\lambda) - c_0\|^2 = 2\lambda\langle x_0 - c_0, c_1 - c_0\rangle - \lambda^2\|c_1 - c_0\|^2. \end{aligned}$$

Dividing both sides by λ and letting $\lambda \downarrow 0$, we obtain $\langle x_0 - c_0, c_1 - c_0\rangle \le 0$, as required. $\square$

6. Nonconvex sets. When Ω is nonconvex, we define the tangent cone first, then define the normal as the polar of the tangent.

A vector $w \in \mathbb{R}^n$ is tangent to Ω at $x \in \Omega$ if for all sequences $x_i \in \Omega$ with $x_i \to x$ and all positive scalar sequences $t_i \downarrow 0$, there is a sequence $w_i \to w$ such that $x_i + t_i w_i \in \Omega$ for all i.

REFERENCES

[1] M. C. Ferris, O. L. Mangasarian, and S. J. Wright. *Linear Programming with* MATLAB. SIAM, November 2007.

[2] O. L. Mangasarian. *Nonlinear Programming.* SIAM Publications, Philadelphia, PA, 1994. (Reprint of 1969 McGraw-Hill edition.).

[3] J. Nocedal and S. J. Wright. *Numerical Optimization.* Springer, New York, second edition, 2006.

[4] S. M. Robinson. Class notes for cs726. unpublished, 1987.