CS 435, Linear Optimization

Lecture Plan, 2015

Introduction

1. **Introduction to optimization**. Domains and functions. Classification of domains as discrete and continuous. Examples of optimization problems:

   - **The Time-tabling problem**. How to assign slots to actual times. The domain as bijections and the function to save on transit time. The travelling salesman problem.
   - **The Electricity Grid Problem**. Given villages, how to find the cheapest electricity grid. Modelling this as the min-cost spanning tree. The question of redundancy and the 2-connected min-cost subgraph problem.
   - **The Powai Lake Aerator Problem**. Locating a point to put the aerator. Largest inscribed circle.

   The separation of Optimization into Modelling, Solution and core Mathematics. An example of the mean-value theorem in 1-D and 2-D. The antipode pressure and temperature problem.

2. The construction of rationals $\mathbb{Q}$ through integers $\mathbb{Z}$. Downward closed and upward closed ideals. Principal ideals. The lemma of complements: Either of (i) both I and $\bar{I}$ are non-principal, or (ii) only one of them is principal. $\mathbb{R}$ as non-principal down-closed ideals and $\mathbb{Q}$ as those whose complement is principal. The centroid of Powai lake as an approximation of reals by rationals.

Topology

1. Introduction to $B_\epsilon(x)$ as basic open sets. Definition of open sets. Finite intersection and arbitrary union prroperties. Definition of closed sets. Standard examples. Rationals as neither. A limit point of a set. Definition of closed sets in terms of closed under taking limits.

2. Basic open sets in $\mathbb{R}^n$. Definition of the topology of $\mathbb{R}^n$. Examples of varieties. Powai lake. The euclidean metric and the euclidean basic open sets. equivalence of the two sets of open sets. More examples of closed and open sets in $\mathbb{R}^n$. Domain of triangles in Powai Lake.

3. **Definition**. Let X be a set and Γ be a collection of subsets of X which we call as **open**. We say that Γ is a topology iff (i) $X, \phi \in \Gamma$, (ii) finite intersections of open sets is open, and (iii) arbitrary union of open subsets is open.

4. Let BOS (also known as *basic open sets*) be a family of subsets of X such that (i)$X, \phi \in BOS$, (ii) for every $U, V \in BOS$ and every $x \in U \cap V$, there is a $W \in BOS$ so that $x \in W \subseteq U \cap W$. Let Γ be the collection of all aribitrary unions of elements of BOS. Show that Γ is a topology.

5. Limit points and closed sets. The definition of limits: $(x_k) \rightarrow x$ iff for every U containing x, there is an N (depending on x and U) for all $j > N$, we have $x_j \in U$.

6. Lemma: C is closed iff for all $(x_i) \rightarrow x$ then $x \in C$. Let us prove this here. In the forward direction, suppose C is closed and $(x_n) subset C$ and $x_n \rightarrow x$. If $x \notin C$ then $x \in \overline{C}$ which is open. Thus there is an open set $\overline{C}$ which separates (x_n) from x! In the other direction, we show that if C enjoys the limit property then C must be closed. We show that $\overline{C}$ is open. Suppose not. Then there is an $x \in \overline{C}$ such that for all U open, with $x \in U$, there is a $x_U \in U \cap C$. Let us consider the BOS $B_{1/n}(x)$ and let $x_n \in B_{1/n}(x) \cap C$. Then $(x_n) \subseteq C$, $(x_n) \rightarrow x$ and yet $x \notin C$!

Functions: Continuity and the mean-value theorem

1. The open-open definition of continuity. The limit definition of continuity at a point. Equivalence of the same. Projections, polynomials, closest distance. Basic properties of continuous functions.

2. Let us prove here the equivalence of the two definitions. Recall that $f : X \rightarrow Y$ is continuous iff for all U open in Y, $f^{-1}(U)$ is open in X. We claim that this is equivalent to the condition that for all sequence $(x_n) \rightarrow x$ in X, we must have that $f(x_n) \rightarrow f(x)$.

Let us prove the forward direction. Suppose that $(x_n) \to x$, and let us denote $f(x_n)$ by y_n. If $(y_n) \not\to y$ then there must be an open set U such that $y \in U$ and there are infinitely many y_{n_j} such that y_{n_j} is not in U. Let $V = f^{-1}(U)$. Since f is continuous, V is open and $\overline{V}$ is closed. Moreover, $x \notin \overline{V}$ while the sequence $(x_{n_j}) \subseteq \overline{V}$. However, it is easy to show that $(x_{n_j}) \to x$. This contradicts that $\overline{V}$ is closed.

The reverse direction is easier. Let us take $U \subset X$ open and suppose that $Y = f^{-1}(U)$ is not open. This means that there is a sequence $(x_n) \to x$ where $x_n \notin V$ while $x \in V$. Applying f to both sides, we see that $(f(x_n)) \to f(x)$ and yet U separates $f(x_n)$ from $f(x)$.

3. Composition of continuous functions. Cross-product of continuous functions $f_i : X \to Y_i$. Continuous functions $f : \mathbb{R}^m \to \mathbb{R}^n$ as a tuple of continuous functions $(f_1, \ldots, f_n) : \mathbb{R}^m \to \mathbb{R}$.

4. Lemma: Let (y_n) be a bounded increasing sequence then (y_n) is convergent. Proof: Let $I = \{q \in \mathbb{Q} | \exists y_k s.t. q \leq y_k\}$. Clearly I is an ideal. If I is principal then $I = y \downarrow$ where $y \in \mathbb{Q}$, or else I is an irrational real y. In either case $(y_n) \to y$.

5. **Mean value theorem**. Let $f : \mathbb{R} \to \mathbb{R}$ be continuous. Let $[x_0, y_0]$ be a closed interval such that $f(x_0) \geq 0$ and $f(y_0) \leq 0$. Then there is a $x \in [x_0, y_0]$ so that $f(x) = 0$. **Proof**. Put $n = 0$.Suppose that $f(x_n) < 0$ and $f(y_0) > n0$. Let $z_n = (x_n + y_n)/2$. If $f(z_n) = 0$ then we are done. If $f(z_n) < 0$ define $x_{n+1} = z_n$ and $y_{n+1} = y_n$. If $f(z_n) > 0$ define $x_{n+1} = x_n$ and $y_{n+1} = z_n$. We have now constructed an increasing sequence (x_n) and a decreasing (y_n). Argue that there is an x so that $(x_n) \to x$ and $(y_n) \to x$. Now apply f.

6. Examples. Stereographic projection. Let $D = \mathbb{R}^3 - \{(0,0,0)\}$ and $S^2 = \{(x,y,z) | x^2 + y^2 + z^2 = 1\}$ be the unit sphere. Let $s : D \to S^2$ be given by:

$$\begin{aligned} s(x,y,z) &= \left(\tfrac{x}{\sqrt{x^2+y^2+z^2}}, \tfrac{y}{\sqrt{x^2+y^2+z^2}}, \tfrac{z}{\sqrt{x^2+y^2+z^2}}\right) \\ &= (f_X(x,y,z), f_Y(x,y,z), f_Z(x,y,z)) \end{aligned}$$

where each $f_i : D \to \mathbb{R}$ is a continuous function.

Next let $S^1 = \{(x,y) | x^2 + y^2 = 1\}$ and let $L = S^1 \times [0, \pi/2] \subseteq \mathbb{R}^3$. This is a cylinder sitting on the unit circle of height $\pi/2$. Given any point

$(\theta, \alpha) \in L$, let us call θ the longitude and α the latitude. Define the map $p : L \to S^2$ given by:

$$p(\theta, \alpha) = (\cos(\alpha)\cos(\theta), \cos(\alpha)\sin(\theta), \sin\alpha)$$

The map p imprints the latitude and longitude on the globe S^2. Note that $p(\theta, \alpha) \in S^2$ and also that $p(\theta, \pi/2) = (0, 0, 1)$. Thus, the whole upper rim of the cylinder goes to the pole. On the remaining part of the cylinder, the map p is 1-1. Compute the inverse of the map.

The Separation Theorem

1. **Definition**. A set $D \subseteq \mathbb{R}^n$ is called convex iff for $x, y \in$, the line segment connecting x and y also belongs to D, i.e., $\{\lambda x + (1-\lambda)y | \lambda \in [0, 1]\} \subseteq D$. Examples of convex and non-convex sets. Convex hull of set of points. Exercise: convex hulls of points in the convex hull. Polytopes and polyhedra-by inequalities and by convex hull of points.
2. Hyperplanes and their use as witnesses. Examples. **Theorem**. Let D be a closed convex domain in $\mathbb{R}^n$ and $p \in \mathbb{R}^n$. Then $p \notin C$ iff there is an $x \in \mathbb{R}^n$ and $a \in \mathbb{R}$ such that $x \cdot p > a$ and $x \cdot y \leq a$ for all $y \in D$.
3. **Proof**: Suppose $p \notin D$. Step I: There exists a unique closest point $q \in D$ and that $\|p - q\| > 0$. Step II. $x = p - q$ does the job.
4. Applications of the separation theorem. The dye-inventory problem.

The closest point algorithm

1. The closest point algorithm when D is given as a convex hull of finite number of points. Inputs and outputs. The initialization and the iterator. Proofs that the sequence is convergent and within the polytope.
2. Convergence problem of the stopping condition. Plots of $log(d(i))$ when point is inside and point is outside. Problem when D is given by $Ax^T \leq b$.

The linear optimization problem on domains of the type $D = \{x | Ax^T \leq b\}$

1. Definitions of $I(x)$, $Dir(x)$ for any $x \in D$.
2. Cones. Representation by rays and $Ax^T \leq 0$. Examples. Cones of infinitesimal movements in domains given by $Ax^T \leq b$. Proof that $Dir(x)$ is indeed this.
3. Separation theorem for cones that it allows $a = 0$. Polar cones and the double polar theorem. Definitions of $Normal(x)$ and proof that these are indeed the normals at a point x.
4. The generic b condition that $rank([A_I b_I]) \neq rank(A_I)$ and its consequence. Classification of points of a polytope D.
5. The linear KKT theorem for closest point in $Ax^T \leq b$. The dichotomy that either $c \in Normal(x)$ or there is a direction d such that $d \cdot c^T > 0$.
6. Implementation of the condition by $-(A_I)^{-1}$ in the full rank case. The Simplex algorithm. Implementation in the non-full rank. Example of closet point problem.
7. The structure of polyhedra of the type $Ax^T \leq b$. Removal of redundant inequalities for the proof. Structure of faces. Degrees of freedom. The conical decomposition of polytopes of the form $Ax^T \leq b$. The boundedness condition that $cone(A) = \mathbb{R}^n$.
8. The statement without proof that convex hull of finite set can be put in the form of $Ax^T \leq b$. Combinatorial outline of the proof.
9. Standard form of the LP. Complimentary slackness and the dual LP. The primal-dual form. Examples of LPs and their duals.

Derivative and the Gradient

1. **Optimization**. The three parts of a typical optimization program are as follows: (i) **Modelling**, i.e., setting up the domain D and the objective function f, (ii) **Iteration**, i.e., getting to a better $x_{n+1} \in D$ from a given $x_n \in D$, and finally (iii) the **stopping condition**, i.e., mathematical conditions to *prove* that you are at the maxima, or at least, a *local maxima*. The first (i) is to do about the clever use of functions to model and design D and f. Parts (ii) and (iii) depend

more on *differential* properties of f and of the boundary of D, i.e., the functions, say $g_1, \ldots, g_k$ used to define D.

2. **Beer and Water**. Consider the case when you go to the market with 1 unit of money. There is beer and water in the market on which you want to spend this money. Each of beer and water has a utility function $u_b(y)$ and $u_w(y)$, where $u_a(y)$ is the amount of happiness obtained by consuming y units of quantity a. Suppose, e.g., $u_b(y) = y$ while $u_w(y) = \sqrt{y}$. If you decided to devote x units to beer and the remaining to water. Thus $D = [0, 1]$ and $f = u_b(x) + u_w(1 - x)$.

   Let x_0 be the optimal purchase. We see that, either (i) the $x_0 \in (0, 1)$ in which case $\partial u/\partial x = 0$ or (ii) $x_0 = 0$ or $x_0 = 1$. This explains that the conditions which define the optimal come from either the domain or the objective functions (or both). In this case, we have $u(0) = u(1) = 1$. On the other hand:

$$\partial u/\partial x = 1 - \frac{1}{2\sqrt{1-x}} = 0$$

   This gives us $x = 3/4$ and $u(3/4) = 5/4$. This is the optimum.

3. The derivative at a point in 1 dimension. Abstract properties.

   Definition. The derivative of a function $f : \mathbb{R} \to \mathbb{R}$ at the point a is given by the limit $\lim_{h\to 0} \frac{f(a+h)-f(a)}{h}$. $\lim_{h\to 0}$ is short form for *every sequence* (h_n) *such that* $(h_n) \to 0$. We similarly define left-hand and right-hand derivatives by restricting the sign of h. We say that f is differentiable on an interval $[a, b]$ iff it is differentiable at every point of (a, b), is RH-differentiable at a and LH-differentiable at b. The derivative $f' : [a, b] \to \mathbb{R}$ is the value of this derivative at each point $x \in [a, b]$.

4. **Rolle's Theorem**. Let $f : \mathbb{R} \to R$ be differentiable and so that its derivative f' is continuous. Let $f(a) = f(b)$. Then there is a point $y \in [a, b]$ such that $f'(y) = 0$. **Proof**. Let f achieve its maxima in $[a, b]$ (why?) at y. Show by a limiting argument that $f'(y) = 0$.

   Conversely, if $f : \mathbb{R} \to \mathbb{R}$ is differentiable so that its derivative f' is differentiaqble and f'' is continuous, and if $f'(x_0) = 0$ and $f''(x_0) < 0$ then there is an $\epsilon > 0$ so that for all y such that $|y - x_0| < \epsilon$ we have $f(y) \leq f(x_0)$, i.e., x_0 is a **local maxima**.

5. **Lemma**. Let f be differentiable so that f' is continuous. Suppose that $f'(a) = C > 0$ and $0 < c < C$, then there is an $\epsilon > 0$ such that for all $a < x < a + \epsilon$, we have $f(x) > f(a) + c(x - a)$ for all $a - \epsilon < y < a$ and $f(y) < f(a) + c(y - a)$.

6. Derivative $D_{a,v}$ in a direction v at a position a. Let $f : \mathbb{R}^n \to \mathbb{R}$ be a function and let $a \in \mathbb{R}^n$ be a **position** and $v \in \mathbb{R}^n$ be a **direction**. Define $g_{a,v}(x) = f(a + xv)$. The directional directive of f at a in the direction v is given by $g'_{a,v}(0)$, i.e., $D_{a,v}(f) = g'_{a,v}(0)$.

7. Properties of $D_{a,v}$. $D_{a,v+w} = D_{a,v} + D_{a,w}$. Proof:

$$\begin{aligned}
\frac{g_{a,v+w}(h) - g_{a,v+w}(0)}{h} &= \frac{1}{h}[f(a + vh + wh) - f(a)] \\
&= \frac{1}{h}[f(a + vh + wh) - f(a + vh)] \\
&= \frac{1}{h}[(f(a + vh + wh) - f(a + wh)) + (f(a + wh) - f(a + vh))] \\
&= \frac{1}{h}[(f(a + vh + wh) - f(a + wh) - f(a + vh) + f(a)) + (f(a + vh) \\
&= \frac{(f(a+vh+wh) - f(a+wh) - f(a+vh) + f(a)}{h} + \frac{f(a+vh) - f(a)}{h} + \frac{f(a+wh) - f(a+vh)}{h}
\end{aligned}$$

   Taking limits, we get:

$$D_{a,v+w}(f) = \frac{(f(a + vh + wh) - f(a + wh) - f(a + vh) + f(a)}{h} + D_{a,v}(f) + D_{a,w}(f)$$

   It is easily see that for many classes of functions, e.g., trigonometric, polynomials and exponentials, the first term vanishes.

8. Let $e_i = (0, 0, \ldots 0, 1, 0, \ldots, 0)$, where the 1 is in the i-th position. If $v = (x_1, \ldots, x_n)$ then $D_{a,v} = \sum_i x_i D_{a,e_i}(f)$, i.e.,

$$D_{a,v} = \sum_i x_i \frac{\partial f}{\partial x_i}(a)$$

   The vector

$$\nabla f = [\frac{\partial f}{\partial x_i}]$$

   is called the gradient. Thus $D_{a,v}(f) = v \cdot \nabla f(a)$.