Convex Optimization: Theory (CS5580)

Instructor: Saketh

Contents

Lecture 1

Definition of Mathematical Program

In order to formally study optimization problems we introduced a new mathematical symbol/structure, known as Mathematical Program (MP), which is of the following form:

$$\begin{aligned} \min_{x \in \mathcal{X}} \quad & f(x) \\ \text{s.t.} \quad & g_i(x) \leq 0 \ \forall\ i = 1, \ldots, m. \end{aligned} \tag{1.1}$$

Here, x is a (dummy) variable, $\mathcal{X}$ is any set, and f, g_i are all (extended) real-valued functions, with common domain $\mathcal{X}$.

Further, in the context of the MP (1.1), the following are defined:

1. We say x is the variable, $\mathcal{X}$ is the domain, $f : \mathcal{X} \mapsto \mathbb{R}_{ext}$ is[1] the objective (function) of the MP (1.1).

2. The inequalities $g_i(x) \leq 0$ as refered to as the constraints, while the functions $g_i : \mathcal{X} \mapsto \mathbb{R}_{ext}$ are called as the constraint functions.

3. The set $\mathcal{F} \equiv \{x \in \mathcal{X} \mid g_i(x) \leq 0 \ \forall\ i = 1, \ldots, m\}$ is known as the feasibility set, and each member of the feasibility set is called as a feasible solution/point, of the MP (1.1).

4. The value of the MP (1.1) is defined as $\inf\left(\{f(x) \mid x \in \mathcal{F}\}\right)$, with the understanding that the value is defined as $-\infty$ if the set of feasible function values,

[1] $\mathbb{R}_{ext}$ denotes the extended real numbers i.e., $\mathbb{R} \cup \{-\infty, \infty\}$. The reason for including $\pm\infty$ will be clear when discussing cascaded MPs etc.

$\{f(x) \mid x \in \mathcal{F}\}$, is not bounded below, and is defined as ∞ if the feasibility set is empty. This value is also sometimes called as the optimal value.

5. If an MP's value is $-\infty$, i.e., the set of values of the objective function over the feasibility set is not bounded below, then the MP is said to be unbounded. If an MP's value is ∞, i.e., the feasibility set is empty, then the MP is said to be infeasible.

6. We say x^* is an optimal solution iff x^* is a feasible solution, i.e., $x^* \in \mathcal{F}$, and the optimal value is attained at it, i.e., $f(x^*) =$ value of (1.1).

7. The set of all optimal solutions is called as the (optimal) solution set and is denoted by:

$$\begin{aligned} \arg\min_{x \in \mathcal{X}} \quad & f(x) \\ \text{s.t.} \quad & g_i(x) \le 0 \ \forall \ i = 1, \ldots, m. \end{aligned} \tag{1.2}$$

   This set could be empty or singleton or have multiple elements etc.

8. An MP is said to be solvable iff its solution set is non-empty. Further, it is uniquely solvable iff its solution set is a singleton.

Lecture 2

Comparing MPs

We looked at many examples of MPs and realized that there is no loss of generality due to its standard form.

The notion of optimal value naturally defines a total order on the set of all MPs:

1. We say MP1=MP2 iff the value of MP1 is equal to the value of MP2.
2. We say MP1>MP2 iff the value of MP1 is greater than the value of MP2.
3. By $MP1 = 2$, we mean the value of MP1 is 2. etc.

This also helped us to write down functions of MPs: $h(MP)$ is nothing but h evaluated at the value of the MP. It is an exercise to show that if h is a monotonically-non-decreasing continuous, then:

$$h\left(\begin{array}{cc} \min_{x\in\mathcal{X}} & f(x) \\ \text{s.t.} & g_i(x) \le 0\ \forall\ i \end{array}\right) = \begin{array}{cc} \min_{x\in\mathcal{X}} & h(f(x)) \\ \text{s.t.} & g_i(x) \le 0\ \forall\ i \end{array}$$

We also took an example, $h(x) = x^2$, which is non-monotonic and showed that the equality only if the objective is non-negative.

We also defined an alternative but equivalent form for an MP:

$$\begin{array}{cc} \max\limits_{x\in\mathcal{X}} & f(x) \\ \text{s.t.} & g_i(x) \le 0\ \forall\ i = 1,\ldots,m. \end{array} \tag{2.1}$$

1. All the definitions of variable, domain, objective, constraints, feasibility set remain the analogous.

2. The value of (2.1) is defined as the $\sup(\{f(x) \mid x \in \mathcal{F}\})$, with the understanding that the value is defined as ∞ if the set of feasible function values, $\{f(x) \mid x \in \mathcal{F}\}$, is not bounded above, and is defined as $-\infty$ if the feasibility set is empty.

3. Since for any set S, we have that $\inf(S) = -\sup(-S)$, both forms[1] are equivalent.

We clarified that, in complete contrast to CS Algorithms courses where domains are discrete, this course is about continuous optimization, where the domain is an uncountable set. Further, we are only interested in domains that are uncountable subsets of $\mathbb{R}^n$. Hence we review Euclidean spaces.

Read section 4.1 in Boyd and Vandenberghe [2004] for getting familiar with playing around with MPs.

[1] $-S$ is the set with members as the negatives of those in S.

Lecture 3

Review of Euclidean Space

1. We argued that understanding the special structure implied by the objective function, feasibility set, and more fundamentally, the underlying space in which the variable lives, is important to understand the nature of the associated MP. Hence we began with a review of vector spaces (since in continuous optimization, the variables are assumed to live in Euclidean spaces)[1]:

   (a) The most basic structure of the Euclidean space is that of a Vector space, which is formally defined in page 9 of Sheldon Axler [1997].

   (b) We highlighted linear combination as the important operation in the context of a Vector space. Linear combination in turn involves the more basic vector-addition and scalar-multiplication operations. Given vectors $v \in \mathbb{R}^n, w \in \mathbb{R}^n$ and scalars $\alpha \in \mathbb{R}, \beta \in \mathbb{R}$, the vector $\alpha v + \beta w$ is the corresponding linear combination.

   (c) The next is the structure of an inner-product space, which is defined formally in pg 98-101 of [Sheldon Axler, 1997]. Let V denote the set of all vectors in a vector space, then a function $\langle \cdot, \cdot \rangle : V \times V \mapsto \mathbb{R}$ is defined as an inner-product iff $\langle \cdot, \cdot \rangle$ satisfies the following three properties:

   Non-negativity: i. $\langle v, v \rangle \geq 0 \; \forall \; v \in V$.

   ii. $\langle v, v \rangle = 0 \iff v = 0$.

   Symmetry: $\langle v, w \rangle = \langle w, v \rangle \; \forall \; v, w \in V$.

   Linearity: $\langle \alpha v + \beta w, u \rangle = \alpha \langle v, u \rangle + \beta \langle w, u \rangle \; \forall \; u, v, w \in V, \;\; \alpha, \beta \in \mathbb{R}$.

   (d) We gave many examples of inner-products with Euclidean vectors, Matrices. In particular, we noted that $\langle v, w \rangle_M \equiv v^\top M w (M \succ 0)$ is the

[1] Go through pages 1–13 in [Sheldon Axler, 1997]. Also go through related exercises.

general form of inner-products in Euclidean spaces (and hence analogous in any finite dimensional space). Since M induces the entire geometry (as we see below), it is called the kernel.

(e) Innerproduct naturally induces various geometrical notions:

orthogonality: $v \perp w \iff \langle v, w \rangle = 0$.

Angle: $\angle v, w \equiv \arccos\left(\frac{\langle u,v\rangle}{\sqrt{\langle u,u\rangle}\sqrt{\langle v,v\rangle}}\right)$. Due to the Cauchy-Schwartz inequality[2], this angle formula is well-defined.

Norm (Length): $||v|| \equiv \sqrt{\langle v, v\rangle}$. One can verify that this is a valid definition as it satisfies the three conditions for a norm. A function $||\cdot|| : V \mapsto \mathbb{R}$ is a norm iff

i. $||x|| \geq 0 \ \forall \ x \in V$ and $||x|| = 0 \Leftrightarrow x = 0$ (Non-negativity and positive-definiteness).

ii. $||\lambda x|| = |\lambda| ||x|| \ \forall \ \lambda \in \mathbb{R}$ (Absolute homogeneity).

iii. $||x + y|| \leq ||x|| + ||y|| \ \forall \ x, y \in V$ (Triangle inequality).

In case all the above except the condition "$||x|| = 0 \Rightarrow x = 0$" are satisfied, then it is called a semi-norm.

Distance: Distance between $u, v \in V$ is defined as $||v - w||$.

Projection: Projection of $v \in V$ onto a set $S \subseteq V$, denoted by $P_S(v)$, is defined by $\arg\min_{s\in S} ||s - v||$.

Geometrical objects: Sphere $\mathcal{S} \equiv \{s \in V \mid ||v|| \leq 1\}$, Ellipse $\equiv \{Au + b \mid ||u|| \leq 1\}$ etc.

(f) Also, one can prove other basis geometric results like Pythagorean, Parallelogram theorem etc.

(g) Also, analysis definitions like Cauchy, convergent sequence, limits naturally follow. An inner-product space that is complete (all Cauchy sequences converge) is called a Hilbert space. In this course we will be concerned with variables living in finite-dimensional Hilbert spaces.

[2] See 6.6 in Sheldon Axler [1997] for a proof .

Lecture 4

Linear Sets

All through we assume $\mathcal{V} = (V, +, \cdot, \langle\cdot,\cdot\rangle)$ is the underlying (finite-dimensional) Hilbert space.

1. Linear Sets (L): sets that are closed under linear combinations[1]. We call this the primal definition/characterization of linear sets.

2. Examples of linear sets are: $\{0\}, \{\lambda v \mid \lambda \in \mathbb{R}\}$ for some fixed $v \in V$ (called as line (parallel to v, passing through origin), $\{\lambda v + \mu w \mid \lambda, \mu \in \mathbb{R}\}$ for some fixed $v, w \in V$ (called as plane (parallel to v, w, passing through origin), ..., V.

3. We define linear span of a set, S, as the set containing all possible linear combinations with vectors in S. i.e., $LIN(S) = \{\sum_{i=1}^{n} \lambda_i v_i \mid v_i \in V, \lambda_i \in \mathbb{R}, n \in \mathbb{N}\}$. In other words, L is linear iff $L = LIN(L)$.

4. Infact, one can efficiently reconstruct a linear set L with linearly independent vectors using this notion of linear span. A set of vectors, $v_1, \ldots, v_m$, is linear independent iff $\sum_{i=1}^{m} \lambda_i v_i = 0 \Rightarrow \lambda_i = 0\ \forall\ i = 1, \ldots, m$. We say S is a spanning set of L iff $L = LIN(S)$. Further, we say a spanning set B is a basis iff B is a linearly independent set. One can show that all bases[2] of a linear set will be are equal size. This common size is known as the dimension of the linear set. Needless to say, Basis of L is the minimal way of representing L using the notion of linear combinations. We say $L = LIN(B)$ is the primal/inner representation/description of L.

[1] Refer sections A.1.2, A.1.3, A.1.4, A.3.4 in Ben-Tal and Nemirovski [2021] for material on Linear Sets.

[2] Please refer pages 21-36 of Sheldon Axler [1997]

5. We realized that every linear set can also be described using the notion of orthogonality. Let L be a linear set and B be a basis of subspace induced by it. Let us define the orthogonal complement[3] of a set S, as, $S^\perp \equiv \{v \in \mathcal{V} \mid \langle l, v\rangle = 0 \;\forall\; l \in S\}$. The following statements are true:

   (a) $L^\perp$ is a linear set (follows from linearity of the inner product). Infact, $L^\perp$ is a linear set even if L is not a linear (but an arbitrary) set. Let us denote the basis of $L^\perp$ as $B^\perp$. Needless to say, $B^\perp \cap B = \{0\}$.

   (b) By the rank-nullity theorem[4], it follows that $dim(L) + dim(L^\perp) = dim(V)$[5]. From this and orthogonality, it follows that $B^\perp \cup B$ is a basis for V.

   (c) A key result in duality is:

   Theorem 4.0.1. *$L = \left(L^\perp\right)^\perp$, whenever L is a linear set.*

   . This follows from the rank-nullity theorem. Note that the theorem need not be true if L is not a linear set, in which case $L \subset \left(L^\perp\right)^\perp$.

   (d) From the above it follows that $L = \left\{v \in \mathcal{V} \mid \langle l, v\rangle = 0 \;\forall\; l \in B^\perp\right\}$. We call this the dual/outer representation/description of L. $B^\perp$ is also known as the dual basis of L.

   (e) We call linear sets of the form $\mathcal{H}_u \equiv \{v \mid \langle v, u\rangle = 0\}$ as hyperplanes (passing through origin; $u \neq 0$).

   (f) From above discussion it is also clear that linear sets are nothing but intersections of hyperplanes through origin. We call this the dual definition/characterization of a linear set.

6. If $dim(L) \leq \left\lfloor \frac{n}{2} \right\rfloor$, then one would describe L as $LIN(B)$, else one would describe L as $\left\{v \in V \mid \langle v, b\rangle = 0 \;\forall\; b \in B^\perp\right\}$. Thus one would at the maximum require $\left\lfloor \frac{n}{2} \right\rfloor$ vectors to represent any Linear set!

[3] Refer A2.3 in Ben-Tal and Nemirovski [2021].

[4] Theorem 3.4 in Sheldon Axler [1997].

[5] this justifies the name complement!

Lecture 5

Linear Sets: Calculus & Topology; Affine Sets

1. We discussed[1] operations that preserve linearity of sets:

   (a) Given an arbitrary collection of sets $S_\lambda,\ \lambda \in \Lambda$, where Λ is the index set[2], we define (arbitrary) intersection, $\cap_{\lambda \in \Lambda} S_\lambda \equiv \{x \mid x \in S_\lambda\ \forall\ \lambda \in \Lambda\}$. It is easy to see that (arbitrary) intersection of linear sets is linear.

   (b) Given an arbitrary collection of sets $S_\lambda,\ \lambda \in \Lambda$, where Λ is the index set[3], we define (arbitrary) union, $\cup_{\lambda \in \Lambda} S_\lambda \equiv \{x \mid \exists \lambda \in \Lambda\ \ni x \in S_\lambda\}$. It is easy to give counter examples where union of two linear sets is not linear.

   (c) Given sets $S_1, \ldots, S_n$ and reals $\lambda_1, \ldots, \lambda_n$, we define their linear combination as $\sum_{i=1}^{n} \lambda_i S_i \equiv \{\sum_{i=1}^{n} \lambda_i v_i \mid v_i \in S_i,\ \forall\ i = 1, \ldots, n\}$. It is easy to show that linear combinations of linear sets are same as a simple summation of the same sets, and are linear sets. Infact, $LIN\left(S_1 \cup S_2\right) = S_1 + S_2$.

   (d) If L is linear, then it's complement, $L^c \equiv \{v \in V \mid v \notin L\}$, will not be linear (infact L^c will not even contain 0).

   (e) Given two Linear sets L_1, L_2, their Cartesian product, $L_1 \times L_2 \equiv \{(v_1, v_2) \mid v_1 \in L_1,\ v_2 \in$ is also a linear set[4].

[1] We encourage readers to think about two different proof strategies henceforth. One based on primal definition, and the other based on dual.

[2] Index set could be finite, countably infinite or uncountable.

[3] Index set could be finite, countably infinite or uncountable.

[4] This is a sub-result used in proving that Direct sum is a valid Hilbert space.

(f) Given two sets S_1, S_2, we define their set difference as $S_1 \backslash S_2 \equiv \{v_1 \in S_1 \mid v_1 \notin S_2\}$. Again, $L_1 \backslash L_2$ will not be linear for linear L_1, L_2 (infact $L_1 \backslash L_2$ will not even contain 0).

2. We introduced some topological notions:

Closure: Given a set, S, closure[5], $Cl(S)$, is defined as the set comprised of the limits of all convergent sequences formed with elements of S.

Closed set: S is closed iff $S = Cl(S)$.

Interior Point: Given a set, S, a point $x \in S$ is said to be an interior point of S iff $B_\epsilon(x) \subseteq S$ for some $\epsilon > 0$, where $B_\epsilon(x) \equiv \{v \in V \mid ||v - x|| \leq \epsilon\}$ is the ball of radius ϵ centered at x.

Interior: The set of all interior points of S is defined as the interior, $int(S)$. A set is said to have interior iff its interior is non-empty.

Boundary: Given a set S, boundary, $\delta(S) \equiv Cl(S) \backslash int(S)$.

Bounded Set: A set S is bounded iff $B_r(0) \subseteq S$ for some finite $r > 0$.

Compact: A set S is compact iff it is closed and bounded.

3. Here are some standard results in topology:

   (a) Complementarity of open and closed sets: S is closed if and only if S^c is open.

   (b) (arbitrary) Intersection of closed sets is closed; (arbitrary) union of open sets is open.

   (c) Finite Union of closed sets is closed and finite intersection of open sets is open.

   (d) (arbitrary) intersection of bounded sets is bounded. Finite union of bounded sets is bounded.

4. Linear sets are closed[6].

5. Linear sets, except the entire set of vectors, are not open. But as we will see later, they are relatively open.

6. Linear sets, except the one containing only 0, are not bounded.

We now study a slight generalization of linear sets, called affine sets.

[5] B.1.6.A. in Ben-Tal and Nemirovski [2021].

[6] As all finite dimensional spaces are equivalent to Euclidean space, which is complete.

1. We defined affine sets as shifted linear sets: A is affine[7] iff there exists a linear set L and $a \in V$, such that $A = \{a\} + L$.

2. We defined affine combination as linear combination with the restriction that the combining coefficients sum to unity.

3. We defined affine hull:

$$AFF(S) \equiv \left\{ \sum_{i=1}^{m} \lambda_i v_i \mid \lambda_i \in \mathbb{R},\ v_i \in S\ \forall\ i = 1, \dots, m, \sum_{i=1}^{m} \lambda_i = 1, m \in \mathbb{N} \right\},$$

   i.e., the set of all vectors which can be expressed as affine combinations of those in the set.

4. We proved that A is affine iff $A = AFF(A)$, which we took as the primal definition/characterization of Affine sets. It was easy to define notions of affinely spanning set, affine independence and affine basis (refer section A.3 in Ben-Tal and Nemirovski [2021] for all related discussions/proofs). We will call affine basis as the primal/inner representation/description.

5. We defined dimension, $dim(A) \equiv dim(L)$, which turned out to be one less that the number of elements in the affine basis.

6. We proved the dual characterization/definition: A is affine with associated linear set as L, with $B^{\perp} = \{b_1, \dots, b_m\}$ as the basis for $L^{\perp}$, iff there exist numbers $\alpha_i \in \mathbb{R},\ i = 1, \dots, m$, such that $A = \{v \mid \langle v, b_i \rangle = \alpha_i,\ \forall\ i = 1, \dots, m\}$. In particular, this shows that affine sets are nothing but solution sets of (non-homogeneous) linear equations. We call $(B^{\perp}, \alpha)$ as the dual/outer representation/description of A, where α is the vector with entries as α_i.

7. We call affine sets of dimensionality one less than the highest, as Hyperplane. Needless to say, the dual characterization is the most efficient: $\mathbb{H}_w \equiv \{x \mid \langle w, x \rangle = b\}$, where $w \neq 0, b \in \mathbb{R}$. It follows that all affine sets, apart from V, are either hyperplanes or their intersections.

8. We gave examples of affine sets, hyperplanes, and identified their primal and dual representations.

9. The operations that preserve affinity and the topology remains analogous to linear sets.

[7] Please refer sections A.3 in Ben-Tal and Nemirovski [2021] and optionally, section 1 in Rockafellar [1996] for material on Affine sets.

Lecture 6

Cones

1. We defined conic combination as linear combination with the restriction that the combining coefficients must be non-negative.

2. We defined conic hull:

$$CONIC(S) \equiv \left\{ \sum_{i=1}^{m} \lambda_i v_i \mid \lambda_i \in \mathbb{R}^+,\ v_i \in S \ \forall\, i = 1, \ldots, m, m \in \mathbb{N} \right\},$$

   i.e., the set of all vectors which can be expressed as conic combinations of those in the set.

3. We say that K is a cone/conic-set iff $K = CONIC(K)$, which we took as the primal definition/characterization of Conic sets.

4. We say S is a conicly spanning set of K iff $K = CONIC(S)$. We also saw examples like the ice-cream cone (in 3d) and the psd cone (in space of Symmetry matrices), that are NOT polyhedral cones. In each case we identified a "minimal" conicly spanning set:

   (a) For the ice-cream cone (in 3d), a minimally conicly spanning set is the unit circle at unit height.

   (b) For the psd cone, a minimally conicly spanning set is the set of all symmetric-rank-one matries i.e., matrices of the form $xx^\top$, $x \in \mathbb{R}^n$.

Lecture 7

Cones: Duality

1. We realized examples of cones with finitely sized conicly spanning sets, which we henceforth call as Polyhedral Cones.

2. We then generalized the notion of an orthogonal complement, and defined the dual cone, S^*, of a set S: $S^* \equiv \{v \in V \mid \langle v, s\rangle \geq 0 \; s \in S\}$. It is an easy exercise to show that S^* is indeed a cone for any set S. We gave examples of dual cones, and noted that the ice-cream and psd cones are dual to themselves and hence are called as self-dual cones.

3. We proved that S^* is always a closed set.

Lecture 8

Cones: Duality & Algebra

1. We them attempted proving an important duality result:

   Theorem 8.0.1. *For a closed cone K, we have $K = (K^*)^*$.*

   While it was easy to see that $K \subseteq (K^*)^*$, we said it is not straightforward to show the converse. We noted that a separation theorem, which we will state and prove in coming lectures on convex sets, will help proving it. Infact we mentioned all duality concepts including that of notion of subgradients for convex functions follow from this basic, fundamental, separation theorem.

2. For now, we assumed that the above conjecture is true and hence dual description of a closed cone is immediate:

   Theorem 8.0.2. *K is a closed cone if and only if it is intersection of halfspaces through the origin.*

   Hence, we take this as the dual definition/characterization of closed Conic sets.

3. Another important result in duality is:

   Theorem 8.0.3. *K is a polyhedral cone if and only if it has a finite dual description.*

   This can be proved (refer theorem 4.5.1 in LAURITZEN [2009]) using the Fourier-Motzkin's algorithm (theorem 1.2.2 in LAURITZEN [2009]). Refer also theorem B.2.5 inBen-Tal and Nemirovski [2021].

4. The following results about algebra with cones K_1, K_2 are true (not detailed in lecture, in interest of time):

(a) (Arbitrary) intersection of cones is a cone.

(b) Union of cones need not be a cone. However, $CONIC(K_1 \cup K_2) = K_1 + K_2$.

(c) (Any) linear combination of cones is a cone.

(d) Cartesian product of cones is a cone, and $(K_1 \times K_2)^* = K_1^* \times K_2^*$.

(e) Complement of a cone is never a cone.

(f) $K_1 \subseteq K_2 \Rightarrow K_2^* \subseteq K_1^*$.

(g) Milutin-Dubovitski lemma: $(K_1 \cap K_2)^* = K_1^* + K_2^*$, for closed cones K_1, K_2 whose sum is also closed[1].

5. Following topological results hold for cones:

   (a) Cones can be closed, open, neither, both.

   (b) Cones are unbounded.

   (c) Refer to exercise B.16 in Ben-Tal and Nemirovski [2021].

6. Refer sections B.1.4, B.2.7.B in Ben-Tal and Nemirovski [2021], section 2.6.1 in Boyd and Vandenberghe [2004], and optionally relevant parts in sections 2, 14 in Rockafellar [1996], for discussion on cones.

[1] Refer section B.2.7.C in Ben-Tal and Nemirovski [2021].

Lecture 9

Convex sets: Definition & Examples

1. We say C is a convex set iff $x, y \in C, \lambda \in [0,1] \Rightarrow \lambda x + (1-\lambda)y \in C$ i.e., if two points are in the set, then the entire line segment induced by them is also in the set.

2. Motivated by above, we defined convex combination as linear combination with the restriction that the combining coefficients must be non-negative and must sum to unity.

3. We defined convex hull:

$$CONV(S) \equiv \left\{ \sum_{i=1}^{m} \lambda_i v_i \mid \lambda_i \in \mathbb{R}^+,\ v_i \in S\ \forall\ i = 1, \ldots, m, \sum_{i=1}^{m} \lambda_i = 1, m \in \mathbb{N} \right\},$$

   i.e., the set of all vectors which can be expressed as convex combinations of those in the set.

4. Using induction, it was simple to show that C is convex if and only if $C = CONV(C)$, which we took as the primal definition/characterization of Convex sets.

5. We looked at several examples of convex sets:

   Norm Ball: $\{x \mid \|x\| \leq r\}$. Here, $\|\cdot\|$ is some valid norm (need not be an inner-product induced one). r is the radius.

   Ellipse: $\{x \mid \langle x-a, x-a \rangle \leq r\}$. Here, $\langle \cdot, \cdot \rangle$ is some valid inner-product (need not be the default inner-product). a, r are given center and radius.

Displaced Cones: Sets of the form $\{a\} + K$, where a is some vector and K is some cone. Most important sub-example of this is the halfspace (that need not pass through origin).

Conic-section: Intersection of a cone and an affine set.

Simplex: Intersection of $\{x \mid x \geq 0\}$ cone and the hyperplane $\left\{x \mid \mathbf{1}^\top x = 1\right\}$.

Spectrahedron: Intersection of the cone of psd matrices and the hyperplane of unit-trace matrices.

Birkhoff polytope: `https://en.wikipedia.org/wiki/Birkhoff_polytope` in the matrix space.

The above examples motivated us to define a polytope: P is a polytope iff $\exists\ S \ni P = CONV(S),\ |S| \in \mathbb{N}$. We argued that the set of permutation matrices ($n!$ matrices) generates the Birkhoff polytope. The set of all matrices with every row having a one in exactly one column position (n^n matrices) generates the set of all Stochastic matrices.

6. Similarly, we defined Polyhedron as the intersection of finite number of halfspaces.

7. Refer sections B.1.1-B.1.3 in Ben-Tal and Nemirovski [2021], sections 2.1-2.3 in Boyd and Vandenberghe [2004]. Optionally, sections 2,3 in Rockafellar [1996].

Lecture 10

Convex Sets: Polyhedral and 1-d characterization

1. The relationship between polyhodron and polytope is nicely summarized by the so-called Minkowski-Weyl theorem (refer theorem B.2.9 in Ben-Tal and Nemirovski [2021]):

   Theorem 10.0.1. *A set $\mathcal{P}$ is polyhedral if and only if there exist finite sets K, C such that $\mathcal{P} = CONIC(K) + CONV(C)$.*

2. Some important conclusions from this theorem are:

   - Every polytope is a polyhedron.
   - Not all polyhedrons are polytopes.
   - Bounded polyhdron is a polytope.

3. We then noted a result that says convexity is essentially a 1-d concept:

   Theorem 10.0.2. *C is convex if and only if $C \cap l$ is either empty or convex for every line l.*

   This helps us in proving some sets are convex/non-convex.

4. We then generalized the notion of dual cones to convex sets: given a set $S \subseteq V$, we define its polar as $S^\circ \equiv \{v \in V \mid \langle v, s \rangle \leq 1 \; \forall \; s \in S\}$ (refer section B.2.7.A in Ben-Tal and Nemirovski [2021]).

5. For many sets we visualized who the polar would look like. In particular, it was easy to see that:

(a) Polar of a cone is same as (negative of) dual cone. Polar of a linear set is same as its orthogonal complement.

(b) Polar of a set is a convex set, even if the set is non-convex.

(c) Polar of a set is a closed set, even if the set is not closed.

(d) Polar of a set contains origin, even if the set does not.

(e) $S^\circ = (CONV(S))^\circ$.

Lecture 11

Convex Sets: Polar, Dual Characterization

6. We then began proving the most important duality result for convex sets:

   Theorem 11.0.1. *If C is a closed convex set containing origin, then $(C^{\circ})^{\circ} = C$. As a consequence, $(K^{*})^{*} = K$, whenever K is a closed cone. And, $\left(L^{\perp}\right)^{\perp} = L$, whenever L is a linear set.*

   Proving $C \subseteq (C^{\circ})^{\circ}$ was easy. The other way, proved in proposition B.2.5 in Ben-Tal and Nemirovski [2021], requires the so-called separation theorem.

7. We then covered s definition related to this theorem: We say two sets $S_1, S_2 \subseteq V$ are strictly separated iff there exists a $w \neq 0 \in V$, such that:

$$\min_{s_1 \in S_1} \langle w, s_1 \rangle > \max_{s_2 \in S_2} \langle w, s_2 \rangle.$$

   Also, in this case, we say "w strictly separates S_1, S_2".

8. From theorem 11.0.1, it follows that:

   Theorem 11.0.2. *C is closed convex if and only if it is an intersection of half spaces (that need not pass through origin).*

   We take this as the dual definition/characterization of (closed) convex sets. The proof follows by shifting origin such that the set contains origin and applying theorem 11.0.1 and then shifting back the origin.

9. Refer section B.2.6.A in Ben-Tal and Nemirovski [2021], and section 14 in Rockafellar [1996].

Lecture 12

Separation theorem, Supporting Hyperplane, Tangent and Normal Cones

1. We began by stating and proving[1] the separation theorem (refer theorem B.2.9 inBen-Tal and Nemirovski [2021]):

   Theorem 12.0.1. *Let C be a closed convex set and $x_0 \notin C$. Then*

   (a) $\Pi_C(x_0)$ *exists and is unique.*

   (b) $\langle x_0 - \Pi_C(x_0), x - \Pi_C(x_0) \rangle \leq 0 \; \forall \; x \in C$.

   As a consequence, $x_0 - \Pi_C(x_0)$ *strictly separates C and x_0.*

2. Motivated by separation theorem's proof, we defined the notion of a supporting hyperplane: Given a set $S \subseteq V$ and a point on the boundary, $x_0 \in \partial S$, we say that the hyperplane $\{x \in V \mid \langle w, x - x_0 \rangle = 0\}$ is a supporting hyperplane of S at x_0 iff $\langle w, x - x_0 \rangle \leq 0 \; \forall \; x \in S$.

3. We then desired to show that all closed convex sets have a supporting hyperplane at all boundary points. We argued that this will need defining two cones:the tangent and the normal.

4. We defined tangent cone[2] of a set S at a point $s_0 \in S$ as all those directions along which one can move from s_0 and stay inside S. Formally, $\mathcal{T}_S(s_0) \equiv \{h \in V \mid \exists t > 0 \ni x_0 + th \in S\}$.

[1] Some proofs, like this one, appear in previous offering's notes: `https://1drv.ms/b/s!Au6Zdrbq2x4phu1rCuc-ZBseLtnnuA`, `https://1drv.ms/b/s!Au6Zdrbq2x4pgc9YPLmTTUMOHwfemg`.

[2] Ben-Tal and Nemirovski [2021] calls this the radial cone.

5. After some examples, we easily showed that:

   Theorem 12.0.2. *For a convex set, tangent cone at any point is indeed a cone. Moreover,* $\mathcal{T}_S(s_0) = CONIC(\{s - s_0 \mid s \in S\})$*.*

6. We then defined its dual cone as the normal cone[3]: $\mathcal{N}_S(s_0) \equiv (\mathcal{T}_S(s_0))^*$.

7. Since by definition of a boundary point, x_0, of a closed convex set, C, there is atleast one direction moving along which one cannot stay inside the set (for any small movement), it is clear that the tangent cone is not V. Hence the Normal cone cannot be $\{0\}$, and there consequently there exists a $w \neq 0 \in \mathcal{N}_C(x_0)$. By definition of Normal cone, it follows that $\{x \in V \mid \langle w, x - x_0 \rangle = 0\}$ is a supporting hyperplane of C at x_0. We summarize this as the following important theorem:

   Theorem 12.0.3. *Let C be a closed convex set and $x_0 \in \partial C$. Then there exists a supporting hyperplane for C at x_0.*

8. Read sections B.2.6.C in Ben-Tal and Nemirovski [2021] and section 2.5 in Boyd and Vandenberghe [2004]. Optional reading: section 11 in Rockafellar [1996].

[3] Boyd and Vandenberghe [2004] defines normal cone as the negative of the dual cone of the tangent cone.

Lecture 13

Convex Sets: Calculus and Topology

We quickly wrapped up our discussion on convex sets by noting:

1. The following results about algebra with convex sets C_1, C_2 are true (refer section B.1.5 in Ben-Tal and Nemirovski [2021]):

    (a) (Arbitrary) intersection of convex sets is a convex.

    (b) Union of convex sets need not be convex. However, $CONV(C_1 \cup C_2) = C_1 + C_2$.

    (c) (Any) linear combination of convex sets is a convex set.

    (d) Cartesian product of convex sets is a convex set.

    (e) Consider an Affine mapping defined by $y = Ax + b \in \mathbb{R}^m, x \in \mathbb{R}^n$ where A is $m \times n$ and $b \in \mathbb{R}^m$.

        i. $C \subseteq \mathbb{R}^n$ is convex $\Rightarrow$ its image under the affine mapping, i.e., $\{y = Ax + b \mid x \in C\}$ is convex.

        ii. $C \subseteq \mathbb{R}^m$ is convex $\Rightarrow$ its pre-image under the affine mapping, i.e., $\{x \mid Ax + b \in C\}$ is convex.

    (f) Complement of a convex set is never a convex set.

2. Following topological results hold for convex sets:

    (a) Convex sets can be closed, open, neither, both.

    (b) Convex sets can be bounded, unbounded.

(c) We defined relatively interior point x_0 of S iff $B_\epsilon(x_0) \cap AFF(S) \subseteq S$. The set of all relatively interior points are relative interior $rint(S)$. We argued that all convex sets have non-empty relative interior (as they contain simplices).

(d) Refer to section B.1.6 in Ben-Tal and Nemirovski [2021] for further details.

Lecture 14

Real-Valued Functions over Hilbert Spaces

1. We then began study of the final ingredient of a MP, which is a real-valued function over a subset in a Hilbert space i.e., $f : V \mapsto \mathbb{R}_{ext}$[1]. We define domain of f as $dom(f) \equiv \{x \in V \mid -\infty < f(x) < \infty\})$.

2. We defined (and gave examples) of some special sets associated with functions:

   (a) $graph(f) \equiv \{(x, f(x)) \mid x \in dom(f)\}$. We assume that this set lies in the direct sum of the Hilbert space in which the domain lies, and the space of reals.

   (b) $epi(f) \equiv \{(x, y) \mid x \in dom(f),\ f(x) \leq y\}$. We assume that this set lies in the direct sum of the Hilbert space in which the domain lies, and the space of reals.

   (c) Level set of f at $t \in \mathbb{R}$: $\mathcal{L}_t(f) \equiv \{x \in dom(f) \mid f(x) \leq t\}$. By definition this set lies in the space same as the domain.

3. We then defined some topologically related concepts:

   (a) f is said to be closed iff its epigraph is a closed set.

   (b) f is said to be bounded above iff $\max_{x \in dom(f)} f(x) < \infty$. f is said to be bounded below iff $\min_{x \in dom(f)} f(x) > -\infty$.

   (c) f is said to be continuous at $x_0 \in dom(f)$ iff for every convergent sequence in the domain to it, $\{x_n \in dom(f)\} \to x_0$, we have that

[1] We consider the extended reals as the co-domain because we already know that the objective could itself be defined as the value of an MP (like in Cascaded MPs), which could be $\pm\infty$.

$\{f(x_n)\} \rightarrow f(x_0)$. f is said to be continuous (everywhere) iff it is continuous at every point in its domain.

(d) f is said to be L-Lipschitz continuous (or simply L-conts) iff $x, y \in dom(f) \Rightarrow |f(x) - f(y)| \leq L\|x - y\|$. We showed that every L-conts function is continuous. However functions like the simple 1-d quadratic is continuous but not L-conts.

(e) f is said to be differentiable at $x_0 \in int(dom(f))$ iff

$$\exists\ \nabla f(x_0) \in V \ \ni \ \lim_{x \to x_0} \frac{f(x) - f(x_0) - \langle \nabla f(x_0), x - x_0 \rangle}{\|x - x_0\|} = 0.$$

If such a $\nabla f(x_0)$ exists, then it will be unique and it is called as the gradient vector. It is a simple exercise to show that $\langle \nabla f(x_0), u \rangle = \lim_{h \to 0} \frac{f(x_0+hu) - f(x_0)}{h} \equiv D_f(x_0; u)$, the directional derivative of f at x_0 in the direction[2] u. More specifically:

Theorem 14.0.1. *The i^{th} entry of $\nabla f(x)$ is $\frac{\partial f(x)}{\partial x_i}$.*

(f) $f : \mathbb{R}^n \mapsto \mathbb{R}$ is said to be twice-differentiable at $x_0 \in int(dom(f))$ iff $\exists\ \nabla f(x_0) \in \mathbb{R}^n \ \ \nabla^2 f(x_0) \in \mathbb{R}^{n \times n} \ \ni$

$$\lim_{x \to x_0} \frac{f(x) - f(x_0) - \nabla f(x_0)^\top (x - x_0) - \frac{1}{2}(x - x_0)^\top \nabla^2 f(x_0)(x - x_0)}{\|x - x_0\|^2} = 0.$$

If such a $\nabla^2 f(x_0)$ exists, then it will be unique and it is called as the Hessian matrix. A basic result in calculus says that:

Theorem 14.0.2. *The $(i,j)^{th}$ entry in $\nabla^2 f(x)$ is $\frac{\partial^2 f(x)}{\partial x_i \partial x_j}$. Now define functions $g_{x_0 u}(t) \equiv f(x_0 + tu)$. Then, $\frac{\mathrm{d}^2 g_{x_0 u}(t)}{dt^2} = u^\top \nabla^2 f(x_0 + tu) u$.*

[2] `http://people.whitman.edu/~hundledr/courses/M225/Ch14/Example_DirectionalDeriv.pdf` provides an example where all directional derivatives exist but the function is NOT differentiable!

Lecture 15

Linear and Affine Functions

1. A function $f : L \subseteq V \mapsto \mathbb{R}$ is linear[1] iff L is a linear set, and $f(\sum_{i=1}^{n} \lambda_i x_i) = \sum_{i=1}^{n} \lambda_i f(x_i)\ \forall\ x_i \in L, \lambda_i \in \mathbb{R}, n \in \mathbb{N}$ i.e., Image of a linear combination of some points under the function is the same linear combination of images of those points. Basically, functions where linear intra-extrapolation is exact. We take this as the primal definition.

2. After giving some examples we noted the following important result that was very easy to prove:

   Theorem 15.0.1. *f is linear if and only if $graph(f)$ is a hyperplane*[2] *through origin.*

   (a) We first showed f is linear if and only if graph(f) is a linear set. This was straight-forward to prove. The proof also showed that if graph of a function is linear, then the function must be of the form $f(x) = \langle w, x \rangle$ for some $w \in L$, which is itself a linear function.

   (b) We then noted that $dim(dom(f)) \leq dim(graph(f)) \leq dim(dom(f)) + 1$. Also, since $(x, y) \notin graph(f)$ whenever $y \neq f(x)$, the dimensionality of the linear set is not $dim(dom(f)) + 1$. Hence $dim(graph(f))$ must be $dim(dom(f))$.

3. From the above, the dual definition follows:

[1] For the extended real number counterpart, the definition reads like: A function $f : V \mapsto \mathbb{R}_{ext}$ is linear iff $f(\sum_{i=1}^{n} \lambda_i x_i) = \sum_{i=1}^{n} \lambda_i f(x_i)\ \forall\ x_i \in dom(f), \lambda_i \in \mathbb{R}, n \in \mathbb{N}$, and $dom(f)$ is a linear set. For linear functions, we follow the convention that $f(x) = \infty\ \forall\ x \notin dom(f)$.

[2] This hyperplane lies in the space of the direct sum of that containing the domain and the reals space. This is the default space in which the graph is defined.

Theorem 15.0.2. ***Riesz representation theorem:*** *A function $f : L \mapsto \mathbb{R}$, where L is linear, is linear iff there exists*[3] *a $w \in L$ such that $f(x) = \langle w, x \rangle \ \forall \ x \in L$.*[4] *Moreover*[5]*, the space of linear functions on L, called the dual space, is equivalent to the space induced by L itself.*

4. Once linear functions are studied, affine[6] functions (and analogous results) are immediate: A function $f : A \mapsto \mathbb{R}$ is affine iff A is affine and $f(\sum_{i=1}^{n} \lambda_i x_i) = \sum_{i=1}^{n} \lambda_i f(x_i) \ \forall \ x_i \in A, \lambda_i \in \mathbb{R} \ \ni \ \sum_{i=1}^{n} \lambda_i = 1, n \in \mathbb{N}$ i.e., Image of an affine combination of some points under the function is the same affine combination of images of those points. We take this as the primal definition. Needless to say, all linear functions are affine.

5. Again, we can show:

   Theorem 15.0.3. *f is affine if and only if $graph(f)$ is a hyperplane. If L_A is the linear set associated with A, f is affine*[7] *if and only if there exists a $u \in L_A, b \in \mathbb{R}$ such that $f(x) = \langle u, x \rangle + b$. This is the dual definition.*

[3]This statemen can also be alternatively proved using orthonormal basis for L.

[4]For the extended real number counterpart, the dual definition reads like: A function $f : V \mapsto \mathbb{R}_{ext}$ is linear iff (a) $L \equiv dom(f)$ is a linear set, (b) there exists a $w \in L$ such that $f(x) = \langle w, x \rangle \ \forall \ x \in dom(f)$, and $dom(f)$ is a linear set. For linear functions, we follow the convention that $f(x) = \infty \ \forall \ x \notin dom(f)$.

[5]This additional qualification is left as an exercise to be proven.

[6]For the extended real number counterpart, the definition reads like: A function $f : V \mapsto \mathbb{R}_{ext}$ is affine iff $f(\sum_{i=1}^{n} \lambda_i x_i) = \sum_{i=1}^{n} \lambda_i f(x_i) \ \forall \ x_i \in dom(f), \lambda_i \in \mathbb{R} \ \ni \ \sum_{i=1}^{n} \lambda_i = 1, n \in \mathbb{N}$, and $dom(f)$ is an affine set. For affine functions, we follow the convention that $f(x) = \infty \ \forall \ x \notin dom(f)$.

[7]For the extended real number counterpart, everything is the same with the additional convention that $f(x) = \infty \ \forall \ x \notin dom(f)$.

Lecture 16

Conic Functions: Norms, Semi-Norms, Support Functions

1. A function $f : K \mapsto \mathbb{R}$ is conic[1] iff K is a cone and $f(\sum_{i=1}^{n} \lambda_i x_i) \leq \sum_{i=1}^{n} \lambda_i f(x_i) \; \forall \; x_i \in K, \lambda_i \geq 0, n \in \mathbb{N}$ i.e., Image of a conic combination of some points under the function under-estimates the same conic combination of images of those points. We take this as the primal definition. Needless to say, all linear functions are conic.

2. We noted that norms are conic functions. i.e., let $f(x) = ||x||$, where $|| \cdot ||$ is an arbitrary (yet valid) norm. Then, f is a conic function.

3. We noted that semi-norms are conic functions. Semi-norms necessarily satisfy all properties of norms, except that they may be zero for a non-zero (non-additive-identity) vector. For e.g., $f(x) = \sqrt{x^\top M x}$ is a semi-norm for any $M \succeq 0$.

4. We then proved the following useful characterization of conic functions:

   Theorem 16.0.1. *f is conic (function) if and only if epi(f) is conic (set).*

5. We then defined support functions and then from the above theorem it was clear that support functions are closed conic functions. We say $\mathcal{S}_A$ is support function[2] of A iff $\mathcal{S}_A(x) \equiv \max_{y \in A} \langle x, y \rangle$.

[1] For the extended real number counterpart, everything is the same with the additional convention that $f(x) = \infty \; \forall \; x \notin dom(f)$.

[2] While with any inner-product the function will be closed conic, the inner-product in the definition of support function is the default inner-product of the domain's space.

For e.g., $f(X) \equiv maxeig(X), X \in S^n$ is a support function, as $f(X) = \max_{y \in \mathbb{R}^n} \langle yy^\top, X \rangle_F \ s.t. \ ||y|| = 1$. Hence it is also a closed conic function.

Lecture 17

Dual Definition Conic Functions and Dual Norms

1. We defined a huge family of functions: Support function of a set $C \subseteq V$, evaluated at $x \in V$, is defined as $\mathcal{S}_C(x) \equiv \max_{y \in C} \langle x, y \rangle$. It was easy to show that support function is always a conic function. Moreover, it is also easy to show that its a closed function (as its epigraph is defined by an intersection of halfspaces).

2. From the dual definition of closed cones, it was clear that:

   Theorem 17.0.1. *A function is closed conic if and only if it is a support function (for some set). In other words, a function is closed conic if and only if it is pointwise maximum of a set of linear minorants of it.*

   g is said to be a minorant of f iff $g(x) \leq f(x)\ \forall\ x \in V$. This theorem provides the dual definition for (closed) conic functions.

3. After providing many examples of support functions, we defined the support function of a unit-norm ball (centered at origin) as the dual norm:

$$\begin{aligned} \|x\|_* \equiv \max_{y \in V} \quad & \langle x, y \rangle, \\ \text{s.t.} \quad & \|y\| \leq 1. \end{aligned}$$

   It was easy to show that dual norm is indeed as norm.

4. We then defined the dual function, f^*: a function whose epigraph is the dual cone of the epigraph of a given function[1], f. We noted examples of functions,

[1] Note that dual function can be defined for non-conic functions too!

whose dual function does not exist, by citing functions whose dual cone can never be a (valid) epigraph. Then we showed that:

Theorem 17.0.2. *Let f be a closed conic function whose dual function, f^*, exists. Then:*

$$f^*(x) = \max_{y \in V} \quad \langle x, -y \rangle, \\ \text{s.t.} \quad f(y) \leq 1.$$

Moreover, $(f^*)^* = f$. For such functions, Theorem 17.0.1 is hence a corollary of this theorem i.e., Every closed conic function, f, is the support function of the set, $\{x \mid f^*(-x) \leq 1\}$, provided f^* exists.

5. The proof follows from that written for theorem 17.0.1 and the fact that $f(\lambda x) = \lambda f(x)$ if $\lambda \geq 0$.

6. Refer section 13 in Rockafellar [1996] for conic functions.

Lecture 18

Convex Functions

1. A function $f : C \mapsto \mathbb{R}$ is convex[1] iff C is convex and $f(\lambda x + (1-\lambda) y) \leq \lambda f(x) + (1-\lambda) f(y) \; \forall \; \lambda \in [0,1]$. Using mathematical induction we showed that: If $dom(f)$ is convex, then f is convex if and only if $f(\sum_{i=1}^{n} \lambda_i x_i) \leq \sum_{i=1}^{n} \lambda_i f(x_i) \; \forall \; x_i \in A, \lambda_i \geq 0 \;\; \ni \;\; \sum_{i=1}^{n} \lambda_i = 1, n \in \mathbb{N}$. We take this as the primal definition. Needless to say, all linear, affine, conic functions are convex.

2. We gave our first non-conic example of a convex function as $f(x) = ||x||^2$, where $||\cdot||$ is any valid norm (in some abstract space). It was easy to show this from the primal definition. Nevertheless, we soon realized we will need more definitions if we need to give more examples.

3. We noted the famous Jensen's inequality, from which many other fundamental inequalities can be derived[2]:

   Theorem 18.0.1. *If f is convex and X is a random variable such that $\mathbb{E}[f(X)] < \infty$, then: $f(\mathbb{E}[X]) \leq \mathbb{E}[f(X)]$.*

   Note that the condition in Jensen's inequality with a discrete random variable taking finite values is same as the primal definition (Hence this inequality can be taken as a "Stochastic" definition for convex functions). We mentioned that many fundamental inequalities like the (generalized) AM-GM, Holders etc., are a consequence of Jensens inequality (with the convex function[3] as $-\log(x)$.

[1] For the extended real number counterpart, everything is the same with the additional convention that $f(x) = \infty \; \forall \; x \notin dom(f)$.

[2] Refer section 3.1.9 in Boyd and Vandenberghe [2004]. See proof2 in `https://en.wikipedia.org/wiki/Jensen%27s_inequality#Proofs`.

[3] We will prove $-\log(x)$ is a convex function later.

4. Again, it was easy to show that:

   Theorem 18.0.2. *f is a convex function if and only if epi(f) is a convex set. Also, all (non-empty) level sets of a convex function are convex*[4]*.*

5. We named a special convex function: Indicator function of a set S evaluated at $x \in V$ is defined as $I_S(x) \equiv \begin{cases} 0 & \text{if } x \in S, \\ \infty & \text{else.} \end{cases}$. Needless to say, I_C, is convex if and only if C is convex.

6. Sections 3.1.1, 3.1.7,3.1.8,3.1.9 in Boyd and Vandenberghe [2004]; C.1 in Ben-Tal and Nemirovski [2021]; relevant parts in section 4 in Rockafellar [1996].

[4] However, there are many non-convex functions (e.g., the inverted bell), whose level-sets are all convex.

Lecture 19

Convex functions: Sub-differentiability

A closer look at the supporting hyperplanes of epigraphs of convex functions motivated us to define:

1. A vector v is called as a sub-gradient of f at $x_0 \in dom(f)$ iff the so-called sub-gradient inequality:

$$f(x) \geq f(x_0) + \langle v, x - x_0 \rangle$$

   is satisfied for all $x \in dom(f)$.

2. The set of all sub-gradients of f at $x_0 \in dom(f)$ is called the sub-differential set of f at x_0 and is denoted by $\partial f(x_0)$.

3. f is said to be sub-differentiable at x_0 iff $\partial f(x_0)$ is non-empty.

We then proved the following important results:

Theorem 19.0.1. *If f is convex and $x_0 \in relint(dom(f))$, then $\partial f(x_0)$ is:*

1. *non-empty*
2. *convex*
3. *closed*
4. *bounded*

Lecture 20

Convex Functions: First Order Characterization

1. Gradient and Sub-gradient have a very close connection that is exposed by the following theorem:

   Theorem 20.0.1. *Let f be a convex function and $x_0 \in int(dom(f))$. f is differentiable at x_0 if and only if the gradient is the only sub-gradient at x_0 i.e., $\partial f(x_0) = \{\nabla f(x_0)\}$.*

   The proof for "only if" part is easy[1]: for any subgradient v (at x_0), we must have $\langle v, u\rangle \leq \frac{f(x_0+hu)-f(x_0)}{h}$. Because it is differentiable at x_0, talking limits on both sides of the inequality gives: $\langle v, u\rangle \leq D_f(x_0; u) = \langle \nabla f(x_0), u\rangle$. Since this is true for all u, we have $v = \nabla f(x_0)$. The "if" part is purely a result in calculus that is independent of convexity: please refer theorem 25.1 in Rockafellar [1996].

2. The following theorem is handy in computing sub-differential sets:

   Theorem 20.0.2. *Let $f = \max(f_1, \ldots, f_n)$. Let all f_i be convex, then f is also convex. Moreover, if $I_0 \subseteq \{1, \ldots, n\}$ is an index set such that $f(x_0) = f_i(x_0)\ \forall\ i \in I_0,\ f(x_0) > f_j(x_0)\ \forall\ j \in I_0^c$, then*

   $$\partial f(x_0) = CONV\left(\cup_{i\in I_0} \partial f_i\left(x_0\right)\right).$$

   Again, we easily proved RHS $\subseteq$ LHS and left the converse as bonus a exercise. In case there is an arbitrary collection of convex functions, then the generalization of the above theorem is called the Danskin's theorem: refer section 3 in `https://www.cs.cmu.edu/~yaoliang/mynotes/dv.pdf`.

[1] Prop. C.6.5 in Ben-Tal and Nemirovski [2021] provides an alternate proof.

3. The following gives the first-order characterization of convex functions:

   Theorem 20.0.3. *Let f be a continuous function defined on a convex domain, then f is convex if and only if it is sub-differentiable in the domain's (relative) interior.*

   As a corollary of this and theorem 20.0.1, we have that for a continuous and differentiable function, the notion of convexity is same as gradient satisfying the sub-gradient inequality everywhere.

4. Read sections C.6.2, C.3 in Ben-Tal and Nemirovski [2021]; section 3.1.3 in Boyd and Vandenberghe [2004], section 23 in Rockafellar [1996].

Lecture 21

Convex Functions: Second Order Characterizations

1. We then moved to second-order characterization and proved the following theorem:

   Theorem 21.0.1. *$f : (a, b) \mapsto \mathbb{R}$ is convex*[1] *if and only if $\frac{\mathrm{d}^2 f(t)}{dt^2} \geq 0 \ \forall \ t \in (a, b)$. Moreover, a continuous function $g : [a, b] \mapsto \mathbb{R}$ is convex if and only if $\frac{\mathrm{d}^2 g(t)}{dt^2} \geq 0 \ \forall \ t \in (a, b)$.*

2. From the definition of convex functions and the above, the following theorem is immediate:

   Theorem 21.0.2. *Let f be a function defined over a convex domain that is twice-differentiable in the interior of its domain and is continuous everywhere. For every $x_0, u \in C$, define the 1-d restriction $g_{x_0 u}$ given by: $g_{x_0 u}(t) \equiv f(x_0 + tu) \ \forall \ t \ni x_0 + tu \in C$. f is convex if and only if $\frac{\mathrm{d}^2 g_{x_0 u}(t)}{dt^2} \geq 0 \ \forall \ t \in int(dom(g_{x_0 u})), \ \forall \ x_0, u$.*

   In the lecture, we mentioned that the above turns out to be an "easy" definition for many example functions, especially the ones in complicated Hilbert spaces.

3. From the above and theorem 14.0.2, it follows that:

   Theorem 21.0.3. *A continuous function defined on a convex domain and that is twice-differentiable in the domain's interior is convex if and only if the Hessian is psd at any point in the domain's interior.*

[1] For the first statement in the theorem, "a" may be $-\infty$ and/or "b" may be ∞.

4. Refer sections C.2.2 in Ben-Tal and Nemirovski [2021]; 3.1.4 in Boyd and Vandenberghe [2004], relevant parts in sections 24-25 in Rockafellar [1996].

Lecture 22

Conjugacy and Convexity preserving Operations

1. We then generalized the notion of support function, which is nothing but a pointwise maximum of a set of linear functions, to the notion of Fenchel dual/Conjugate/Legendre Transformation, $f^{'}$, which is a pointwise maximum of a set of affine functions:

$$f^{'}(x) \;\equiv\; \max_{y \in V} \langle x, y \rangle - f(y), \tag{22.1}$$

   Note that indeed conjugate generalizes support function: $I_C^{'} = S_C$. In other words, for (the restricted class of) indicator functions, the notion of conjugate is exactly same as that of Support function. After giving examples, we note the following properties:

   (a) Conjugate of *any* function is closed, convex.

   (b) For closed conic functions whose dual exists, we have that: $f^{'}(x) = I_{-\mathcal{L}_1(f^*)}(x)$, where $\mathcal{L}_t(g)$ is the level set of g at level t.

   (c) Proposition C6.6 in Ben-Tal and Nemirovski [2021] summarized below:

   Theorem 22.0.1. *If f is closed convex, then $\left(f^{'}\right)^{'} = f$.*

2. From the above theorem, the dual definition of (closed) convex function is immediate:

   Theorem 22.0.2. *f is closed convex if and only if it is conjugate of some functions. In other words a function is closed convex if and only if it is pointwise maximum of a set of affine minorants of it.*

3. We mentioned that global properties of a function turn out to be local properties of the conjugate and vice-versa. This is the key advantage of this duality relationship. For example, $f'(0)$, which is a local property of conjugate is equal to $-\min_{y \in V} f(y)$, which is a global property of the original function.

4. The notion of conjugate also gives the following inequality: $f(x) + f'(y) \geq \langle x, y \rangle \; \forall \; x, y \in V$. This is called as the Fenchel's inequality. Again, many fundamental inequalities can be derived from this.

5. Read sections: C6.3 in Ben-Tal and Nemirovski [2021], 3.3 in Boyd and Vandenberghe [2004], 12,26 in Rockafellar [1996].

6. We then discussed section 3.2 inBoyd and Vandenberghe [2004], which presents operations that preserve convexity. These will be very handy in proving convexity of functions.

7. We then noted that convex functions have nice topological properties: refer section C.4 in Ben-Tal and Nemirovski [2021]. In particular, we have:

   Theorem 22.0.3. *Any Convex function is continuous in the relint of it's domain. Moreover, it is Lipschtz continuous in any compact subset of relint of domain. Hence it is also bounded in such a compact subset.*

Lecture 23

Convex Programs

1. An MP (1.1) is said to be a Convex Program (CP) iff its objective, f, and all constraint functions, g_i, are convex. As a result, domain (as well as feasibility set) will also be convex.

2. After giving examples of CPs, we noted the following interesting results about CPs:

   (a) Below is a sufficient condition on the feasibility set for boundedness of objective:

   Theorem 23.0.1. *A CP is bounded whenever its feasibility set is bounded.*

   The converse need not be true.

   (b) Below is a sufficient condition for solvability of a CP:

   Theorem 23.0.2. *A CP is solvable whenever its feasibility set is compact and its objective is continuous.*

   Again, the converse need not be true.

   (c) Below is a sufficient condition for unique solvability:

   Theorem 23.0.3. *A CP is uniquely solvable whenever its feasibility set is compact and its objective is continuous, strictly convex.*

   Converse need not be true.

   (d) We defined strictly convex functions: A function $f : C \mapsto \mathbb{R}$ is strictly convex iff C is convex and $f(\lambda x + (1-\lambda)y) < \lambda f(x) + (1-\lambda) f(y)\ \forall\ \lambda \in (0,1)$. Needless to say, all strictly convex functions are convex.

Theorem 23.0.4. *Let $f : C \mapsto \mathbb{R}$ be continuous and C be convex. Then, f is strictly convex in $relint(C)$ if and only if $f(x) > f(x_0) + \langle \nabla f(x_0), x - x_0 \rangle \ \forall \ x \in C \neq x_0, \forall \ \nabla f(x_0) \in \partial f(x_0), \forall \ x_0 \in relint(C)$. In other words, strict convexity is same as strict sub-differentiability.*

Theorem 23.0.5. *Let $f : C \mapsto \mathbb{R}$, where C is convex and f is twice-differentiable in $relint(C)$. Then, $\nabla^2 f(x_0) \succ 0 \ \forall \ x_0 \in relint(C) \Rightarrow f$ is strictly convex. The converse need not be true.*

(e) We then defined Unconstrained Convex Programs: CPs whose domain is the entire set of vectors (that form a finite dimensional Hilbert space) and whose feasibility set is same as its domain. Equivalently, a CP whose domain is entire set of vectors and there are no constraints is an unconstrained CP i.e., CPs of the form $\min_{x \in V} f(x)$. The following theorem was easy to prove:

Theorem 23.0.6. *Let f be a convex function, such that $dom(f) = V$. Then,*

$$x^* \in \arg\min_{x \in V} f(x) \iff 0 \in \partial f(x^*).$$

Lecture 24

First order Optimality conditions

1. We gave intuition and proved the following general condition for optimality:

   Theorem 24.0.1. *Let (1.1) be a CP with a differential objective. Then,*

   $$x^* \text{ is an optimal solution to } (1.1) \iff \nabla f(x^*) \in \mathcal{N}_{\mathcal{F}}(x^*).$$

2. In the subsequent lecture we will write down simplified expressions for the normal cone of feasibility set for special classes of CPs and re-write the above theorem 24.0.1[1] appropriately.

3. Refer section 4.2.3 in Boyd and Vandenberghe [2004] for details.

[1] Direct 'A' grade to students who generalize this theorem to the case where the objective is NOT differentiable.

Lecture 25

Optimality Condition: Special cases

We wrote down various special cases (corollaries) of theorem 24.0.1:

1. Unconstrained CPs: if $\mathcal{F} = V$, then $\mathcal{N}_{\mathcal{F}}(x) = \mathcal{N}_V(x) = \{0\}\ \forall\ x \in V$. This gives back theorem 23.0.6 in differentiable objective case.

2. If $x \in int(\mathcal{F})$, then $\mathcal{N}_{\mathcal{F}}(x) = \{0\}$. This gives back theorem 23.0.6 in differentiable objective case. In particular, if $\mathcal{F}$ is an open set, then any feasible solution is in the interior.

3. If $\mathcal{F}$ is a hyperplane, then one can eliminate one variable and re-write the CP as an unconstrained one. In general, if some of the constraints in the CP are linear equalities, then one can perform gaussian elimination to eliminate few variables and re-write the CP involving lesser number of variables.

4. If $\mathcal{F}$ is the halfspace $\{x \mid \langle a, x\rangle \leq b\}$, then $\mathcal{N}_{\mathcal{F}}(x) = CONIC(\{-a\})$ if $\langle a, x\rangle = b$ (active constraint), and $\mathcal{N}_{\mathcal{F}}(x) = \{0\}$ if $\langle a, x\rangle < b$ (inactive constraint).

5. Next special case is: Polyhedrally Constrained Convex Programs (PCCPs) as CPs with an open domain, and the constrained functions are all restricted to be affine i.e., CPs of the form:

$$\begin{aligned} \min_{x \in \mathcal{X}} \quad & f(x) \\ \text{s.t.} \quad & \langle a_i, x\rangle \leq b_i\ \forall\ i = 1, \ldots, m, \end{aligned} \tag{25.1}$$

where $\mathcal{X}$ is open and f is convex. In this case we showed that $\mathcal{N}_{\mathcal{F}}(x) = CONIC\left(\cup_{i \in A(x)}\{-a_i\}\right)$, where $A(x)$ is the set of all active constraints at

$x \in \mathcal{F}$ i.e., $A(x) \equiv \{i \mid \langle a_i, x\rangle = b_i\}$. With this theorem 24.0.1 can be re-written as:

Theorem 25.0.1. *Let (25.1) be a CP with differentiable objective. Then, x^* is a solution to (25.1) if and only if there exists $\lambda^* \in \mathbb{R}^m$ such that:*

(a) $x^* \in \mathcal{X}$, $\langle a_i, x^*\rangle \leq b_i, \lambda_i^* \geq 0 \ \forall \ i = 1, \ldots, m$ *(feasibility conditions).*

(b) $\lambda_i^* \left(\langle a_i, x^*\rangle - b_i\right) = 0 \ \forall \ i = 1, \ldots, m$ *(complementary slackness conditions).*

(c) $\nabla f(x^*) + \sum_{i=1}^m \lambda_i^* a_i = 0$ *(gradient conditions).*

We then gave the CP corresponding to the maximum entropy models as a use case for the above conditions.

Lecture 26

KKT conditions

We then defined a regular CP as a CP (1.1) with the domain restricted to be (convex) open, all functions (objective, constraint) restricted to be (convex) differentiable and the Slater's condition is satisfied. Slater's condition says that there must exist an $x_0 \in \mathcal{X}$ such that $g_i(x_0) < 0$, whenever g_i is non-affine. We then defined a KKT point (x^*, λ^*) as any pair x^*, λ^* that satisfy the following three (sets of) conditions, known as the KKT conditions:

1. $x^* \in \mathcal{X}$, $g_i(x^*) \leq 0, \lambda_i^* \geq 0 \; \forall \; i = 1, \ldots, m$ (feasibility conditions).
2. $\lambda_i^* g_i(x^*) = 0 \; \forall \; i = 1, \ldots, m$ (complementary slackness conditions).
3. $\nabla f(x^*) + \sum_{i=1}^{m} \lambda_i^* \nabla g_i(x^*) = 0$ (gradient conditions).

By writing an expression for $\mathcal{N}_{\mathcal{F}}$ in this case, theorem 24.0.1 simplifies as:

Theorem 26.0.1. *x^* is a solution to a regular CP if and only if there exists $\lambda^* \in \mathbb{R}^m$ such that (x^*, λ^*) is a KKT point.*

Refer section D.2.3.B in Ben-Tal and Nemirovski [2021] and sections 5.9.2,5.9.3 in Boyd and Vandenberghe [2004] for more details.

Lecture 27

KKT Conditions: Example

1. We discussed an example where the KKT conditions can be used to derive analytical form for the solution: the optimization problem that arises in defining dual norm of (entrywise) p-norm. We understood that the KKT theorem can't be directly applied and made a series of 6-7 re-writings of the CP to arrive at a one where the KKT theorem can be applied.

Lecture 28

Duality in MPs

From our experience with notions of orthogonal complement, dual cone, polar, dual function, conjugate, we then noted desirable properties to define a "dual" MP:

Convexity: We insist that dual of any MP must be a convex program. For e.g., polar of any set is a convex set, conjugate of any function is convex etc.

Outer Description: We insist that dual of a min MP is a max MP such that value of the "primal" minimization MP at any feasible solution is less than that of the dual at any of its feasible solution. Then, the function values in primal will not "overlap" with those in dual. For e.g., vectors in orthogonal complement do not overlap with the set and provide an "outer description"! This is more formally called as principle of Weak Duality. This will also be useful for computationally intractable MPs: for example, the primal objective with any primal feasible solution gives an upper bound and the dual objective with any dual feasible solution will give a lower bound.

(A)symmetry: We insist that the optimal value of primal, if its a convex program, is equal to that of its dual. This is modeled from facts like: polar of polar of a convex set (that is closed and has origin) is the original set, conjugate of conjugate is original function. If $P \geq D \geq D(D)$ and $D(D) = P$, then $P = D$. So, we insist that the primal and dual have same optimal value for CPs. This is formally called as principle of Strong Duality.

Inheritance: We insist that when we define a dual, we reuse some older notions of duality like conjugate, dual cone etc. Using these existing notions, one should be able write down a dual for a given MP.

Related: Both P and D must use the same parameters (data). This makes them "physically related".

We attempted to dualize unconstrained MPs, leading to the degenerate MP: $f^{'}(0)$.

Lecture 29

Fenchel Duality

- We considered MPs of the following form[1]:

$$P \equiv \min_{x \in V} f(x) + g(x). \tag{29.1}$$

- Using the relation $f \geq (f')'$ (bi-conjugate), and interchanging min-max, we have $P \geq D$ (weak duality), where D, the Fenchel Dual, is given by:

$$D \equiv \max_{z \in V} -f'(z) - g'(-z). \tag{29.2}$$

- We noted that D is a convex program, even if P is not.
- We proved the following strong duality theorem:

  Theorem 29.0.1. *If f, g are closed convex, we were able to show $P = D$.*

- We noted the Fenchel dual of the popular (soft-margin) SVM.

[1] Popularly encountered in Machine Learning

Lecture 30

Lagrange Duality

1. We began with PCMPs (Polyhedrally Constrained Mathematical Programs):

$$\begin{aligned} \min_{x\in\mathbb{R}^n} &\quad f(x) \\ \text{s.t.} &\quad A^\top x \le b. \end{aligned}$$

   If further the objective is linear/quadratic, it is known as a Linear/Quadratic Program (LP/QP).

2. We noted that the Fenchel dual of a PCMP involves the conjugate of the objective and that of the Indicator function of the constraint set, I_C, where $C \equiv \left\{x \mid A^\top x \le b\right\}$. The later, in general, is a Linear program, which may not allow any anlytical form. We hence wished to explore dual forms that may avoid the LP. This motivated the use of support function function form (rather than the conjugate form): $I_C(x) = \max_{\lambda\ge 0} \lambda^\top(A^\top x - b)$.

3. After min-max interchange, we have weak duality with $\max_{\lambda\ge 0} -b^\top\lambda - f'(-A\lambda)$. Note that this is a convex program even if the given PCMP is not. Also, from KKT conditions it follows that strong duality holds whenever the objective f is convex and the PCMP is solvable. We call this the Lagrange dual.

4. The primal-dual pair in case $b = 0$ and f is linear (say, $f(x) = c^\top x$) reads as:

$$\begin{aligned} P = \quad & \min_{x\in\mathbb{R}^n} c^\top x \quad = & \max_{\lambda\ge 0} 0 = D \\ & \text{s.t.} A^\top x \le 0. & \text{s.t. } A\lambda + c = 0. \end{aligned}$$

   This says, if D is feasible, i.e., the system $A\lambda + c = 0, \lambda \ge 0$ is solvable, then P is bounded by 0, i.e., the system $A^\top x \le 0, c^\top x < 0$ is infeasible (and vice-versa). This infact is the statement of the popular Farka's Lemma, which we now know is a special case of the Lagrange duality. Such equivalences are interesting because verifying feasibility may be easy; whereas verifying infeasibility may not be[1].

5. We then generalized this to MPs (1.1), henceforth referred to as the primal, we defined Lagrangian: $\mathcal{L}(x,\lambda) \equiv f(x)+\sum_{i=1}^m \lambda_i g_i(x)$. It's domain is $\mathcal{X}\times\mathbb{R}^m_+$. We call x as primal variables, and λ_i as Lagrange multipliers or Lagrange Dual variables or simply, dual variables. We then define the Lagrange dual function: $\underline{\mathcal{L}}(\lambda) \equiv \min_{x\in\mathcal{X}} \mathcal{L}(x,\lambda)$. Finally, we define the Lagrange Dual Problem as:

$$\max_{\lambda\geq 0} \quad \underline{\mathcal{L}}(\lambda). \tag{30.1}$$

6. It was an easy exercise to show:

   Theorem 30.0.1. *Let P be the value of (an arbitrary, perhaps non-convex) MP given by (1.1), and D be that of it's Lagrange dual (30.1). Then, $P \geq D$ (Weak Duality). Moreover, (30.1) is (always) a Convex Program. If P is differentiable regular convex and solvable, then $P = D$ (strong duality).*

7. It is an easy exercise to show that the Lagrange dual of an LP is another LP (self-dual) and that of a strictly-convex QP is another strictly-convex QP (self-dual):

   (a) We defined a Linear Program (LP) as a special CP of the form:

   $$\begin{aligned} \min_{x\in V} \quad & \langle c, x\rangle, \\ \text{s.t.} \quad & \langle a_i, x\rangle \leq b_i \ \forall\ i=1,\dots,m. \end{aligned} \tag{30.2}$$

   Theorem 30.0.2. *The Lagrange dual of (30.2) is:*

   $$\begin{aligned} \max_{\lambda\in\mathbb{R}^m} \quad & -b^\top\lambda, \\ \text{s.t.} \quad & \lambda\geq 0,\ \ c+\textstyle\sum_{i=1}^m \lambda_i a_i = 0. \end{aligned} \tag{30.3}$$

   This again can be written as an LP. Hence self-duality holds for LPs.

   (b) We defined a Quadratic Program (QP) as an MP of the form:

   $$\begin{aligned} \min_{x\in\mathbb{R}^n} \quad & \tfrac{1}{2}x^\top P x + q^\top x, \\ \text{s.t.} \quad & a_i^\top x \leq b_i \ \forall\ i=1,\dots,m. \end{aligned} \tag{30.4}$$

[1] Imagine the difficult situation of a judge who needs to verify the claim of a witness that she does not know a particular language, say, English; compared to the ease in verifying the alternative claim that she knows English well. Also, see example 5.10 in Boyd and Vandenberghe [2004].

Theorem 30.0.3. *If $P \succ 0$, the Lagrange dual of (30.4) is:*

$$\begin{array}{lll} (30.5) & \max\limits_{\lambda \in \mathbb{R}^m} & -\frac{1}{2}\lambda^\top A^\top P^{-1} A\lambda - \lambda^\top \left(A^\top P^{-1} q + b\right) - \frac{1}{2} q^\top P^{-1} q, \\ & \text{s.t.} & \lambda \geq 0. \end{array}$$

This again can be written as a convex QP, hence self-duality holds.

8. Refer sections 5.1-5.2 in Boyd and Vandenberghe [2004] and sections 1.2,D.2,D.3 in Ben-Tal and Nemirovski [2021].

Lecture 31

Conic Programs and Duality

1. We derived the Lagrange dual problem for a general convex QP:

   Theorem 31.0.1. *If $P \succeq 0$, then the Lagrange dual of QP (30.4) is:*

   $$\begin{aligned} \max_{\lambda \in \mathbb{R}^m, t \in \mathbb{R}} \quad & t, \\ \text{s.t.} \quad & \begin{bmatrix} P & q + \sum_{i=1}^m \lambda_i a_i \\ q^\top + \sum_{i=1}^m \lambda_i a_i^\top & -2\left(\sum_{i=1}^m \lambda_i b_i + t\right) \end{bmatrix} \succeq 0, \lambda \geq 0. \end{aligned} \tag{31.1}$$

2. Motivated by the above (and generalizing LPs) we defined a Semi-Definite Program (SDP):

   $$\begin{aligned} \min_{x \in \mathbb{R}^n} \quad & c^\top x, \\ \text{s.t.} \quad & B - \textstyle\sum_{i=1}^n x_i A_i \succeq 0. \end{aligned} \tag{31.2}$$

   Here the matrices B, A_i are symmetric matrices of size m. The constraints of the form in SDP are known as Linear Matrix Inequalities (LMI).

3. We note that if all matrices B, A_i are diagonal matrices, then SDP is same as LP. Secondly, (30.5) can be written as an SDP.

4. SDP happen to be an enormous class of CPs, with huge number of applications. Section 3.2 in Ben-Tal and Nemirovski [2021] presents a host of sets that can be represented by LMIs.

5. Further generalizing SDPs, we defined Conic Programs:

   $$\begin{aligned} \min_{x \in V} \quad & \langle c, x \rangle_V, \\ \text{s.t.} \quad & b -_W l(x) \in K \subseteq W. \end{aligned} \tag{31.3}$$

Here V, W are vector sets from different vector spaces. For e.g., in SDPs, $V = \mathbb{R}^n$ and $W = \mathcal{S}_m$ (symmetric matrices). $a -_W b = a +_W (-1.b)$, where $+_W$ is the addition operator in W space. $\langle \cdot, \cdot \rangle_V$ is the inner-product in V space. $l : V \mapsto W$ is a linear function (definition same as with scalar valued functions). K is a closed cone.

6. It is easy to see that LPs, QPs, SDPs, can all be written in (31.3) form.

7. We defined conic dual of (31.3) as[1]:

$$
\begin{aligned}
\max_{y \in W} \quad & -\langle b, y \rangle_W, \\
\text{s.t.} \quad & l^\top(y) +_V c = 0, \;\; y \in K^* \subseteq W.
\end{aligned} \tag{31.4}
$$

8. Interestingly, in special cases of LP, QP etc., the Lagrange dual (when primal is written in (1.1) form) will match Conic Dual (when primal i written in (31.3) form).

   Theorem 31.0.2. *Conic dual is (always) convex and, the Value of (31.3) $\geq$ value of (31.4), even if K is arbitrary. If K is closed convex, their values are the same.*

   The proof follows from infimal convolution theorem and also highlights the optimality conditions for this case.

[1] Holds even in case K is an arbitrary set.

Bibliography

A. Ben-Tal and A. Nemirovski. Lectures On Modern Convex Optimization. Available freely at `https://www2.isye.gatech.edu/~nemirovs/LMCOLN2021WithSol.pdf`, 2021.

S. Boyd and L. Vandenberghe. *Convex Optimization*. Cambridge University Press, 2004.

NIELS LAURITZEN. LECTURES ON CONVEX SETS. Available at `https://www.fmf.uni-lj.si/~lavric/lauritzen.pdf`, 2009.

R. T. Rockafellar. *Convex Analysis*. Princeton University Press, 1996.

Sheldon Axler. *Linear Algebra Done Right*. Springer-Verlag, 1997.