WINTER SCHOOL 2024 AT TEXAS STATE UNIVERSITY: CONVEX GEOMETRY AND ANALYSIS

RENE CABRERA

Abstract. In this lecture we will talk about some key points concerning the concept of convex geometry and analysis. We talk about convex sets and functions and their geometric interpretations and properties. That when a function is not necessarily differentiable at certain points, we can at least talk about their "subdifferentiability"—a notion that is more general thatn differentiation. Having discussed these concepts, we intoduce the Legendre-Fenchel transform, an important concept in the theory of convex optimization. Then towards the end we tie these concepts together to discuss optimal transport, an important convex optimization problem that is applicable in several branches of math such as analysis, Riemannian geometry, probability, and statistics.

1. Introduction

Convexity either of a set or a function, has the following two important notions: (1) and (2).
Let $\Omega \subset \mathbb{R}^d$ be a *convex* and open subset and $f : \Omega \to (-\infty, +\infty]$. Convexity of a function f is

$$f((1-\lambda)x + \lambda y) \leq (1-\lambda)f(x) + \lambda f(y) \quad \forall\, x, y \in \Omega \tag{1}$$

and

$$\nabla^2 f(x) \geq 0 \quad \forall x \in \Omega. \tag{2}$$

The latter one is called the Hessian and it is a matrix.

While *convexity* of a set K is: given any two distinct points $x, y \in K$, their line segment connecting x and y,

$$[x, y] := (1-t)x + ty \in K \quad \forall\, 0 \leq t \leq 1.$$

The latter can be interpreted as a geometrical notion, while the former analytical. (Here provide pictures to reference the latter—and later the tangent line being below the graph of f for (1)). The notion of convexity can help us determine extremal values of a "convex" function.

1.1. **Motivating example coming from Discrete optimal transport.** First let us put the example of the problem we want to solve; then give it an applicable interpretation. To that end, consider the discrete spaces $X = \{1, 2, \ldots, n\}$ and $Y := \{1, 2, \ldots, m\}$; where n, m are positive integers. We want to investigate what would be the "cost" of moving mass at location x to location y. Allowing the mass to split, the transportation occurs under a *plan* π defined on $X \times Y$ that tells you how much mass is transported from an initial location (the source) to a final one (the target).

As an example, suppose we have an Amazon warehouse at location $x = 0$ with 1 unit of resource and distribution centers at $y = 0, 1$ receiving $1/3$ and $2/3$ units of the resources, respectively. To wit, the plan is such that the resources that we have at $x = 0$, $1/3$ of it goes to $y = 0$, while $2/3$ goes to $y = 1$. This means $\pi(0,0) = 1/3$, $\pi(0,1) = 2/3$, and $\pi(0,\mathbb{R}) = 1$. In general, if $A \subset X$ and $B \subset Y$, $\pi(A, B)$ measures how much mass is transported from A to B. Let us state the problem of optimizing the transportation cost under the plan constraint more precisely.

Let μ and ν be vectors in the set

$$\mathbf{P}_d := \left\{ \omega \in \mathbb{R}^d : \omega \geq 0, \sum_{i=1} \omega_i = 1 \right\}. \tag{3}$$

Given matrix $c_{ij} \in \mathbb{R}^{n\times m}$, the problem is

$$\begin{aligned} &\min_{\pi\in\mathbb{R}^{n\times m}} \sum_{i=1}^{n}\sum_{j=1}^{m} \pi_{ij} c_{ij} \\ \text{subject to constraints } &\pi \geq 0 \\ &\sum_{j=1}^{m} \pi_{ij} = \mu_i \ \forall i \\ &\sum_{i=1}^{n} \pi_{ij} = \nu_j \ \forall j. \end{aligned} \tag{4}$$

Notice that we are minimizing over a convex set

$$\Pi_{ij} := \left\{ \mu \in \mathbf{P}_n, \nu \in \mathbf{P}_m \ : \ \sum_{i=1}^{n} \pi_{ij} = \nu_j, \sum_{j=1}^{m} \pi_{ij} = \mu_i \ \forall (i,j) \right\} \tag{5}$$

Hence, the problem is feasible. The dual problem to (4), is thus

$$\max \left\{ \sum_{i=1}^{n} \phi_i \mu_i + \sum_{j=1}^{m} \psi_j \nu_j : c_{ij} \geq \phi_i + \psi_j \ \forall (i,j) \right\}. \tag{6}$$

We will prove that, assuming that the primal and dual have exactly one extremal point, show that the optimal value of the primal (minimization problem) and dual objectives coincide. [The proof will be done after going over the necessary mathematical machinery]

Proof. Let $b = (\mu_1, \dots, \mu_n, \nu_1, \dots, \nu_m)$, let $\pi = (\pi_{11}, \pi_{12}, \dots, \pi_{1n}; \pi_{21}, \pi_{22}, \dots, \pi_{nm})$, and let $\lambda = (\phi_1, \dots, \phi_n; \psi_1, \dots, \psi_m)$ be the Lagrangian multipliers. The constraints that the matrix $(\pi_{ij})_{i.j}$ ought to satisfy are

$$\sum_{i=1}^{n} \pi_{ij} = \nu_j \ \forall j \quad \text{and} \quad \sum_{j=1}^{m} \pi_{ij} = \mu_i \ \forall i$$

Then the Lagrange function is given by

$$\Lambda(\pi_{ij},\phi_i,\psi_j) := \sum_{i=1}^{n}\sum_{j=1}^{m}\pi_{ij}c_{ij} - \sum_{i=1}^{n}\phi_i\left[\sum_{j=1}^{m}\pi_{ij}-\mu_i\right] - \sum_{j=1}^{m}\psi_j\left[\sum_{i=1}^{n}\pi_{ij}-\nu_j\right]$$
$$= \sum_{i=1}^{n}\phi_i\mu_i + \sum_{j=1}^{m}\psi_j\nu_j + \sum_{i,j=1}^{n,m}(c_{ij}-\phi_i-\psi_j)\pi_{ij}$$

The dual functional can be defined by the minimum of Λ over the $\pi_{i,j}$:

$$\mathfrak{D}(\phi_i,\psi_j) := \min_{\pi_{ij}}\Lambda(\pi_{ij},\phi_i,\psi_j) = \min_{\pi_{ij}\geq 0}\left\{\sum_{i=1}^{n}\phi_i\mu_i + \sum_{j=1}^{m}\psi_j\nu_j + \sum_{i,j=1}^{n,m}(c_{ij}-\phi_i-\psi_j)\pi_{ij}\right\}.$$

In particular,

$$\mathfrak{D}(\phi_i,\psi_j) = \begin{cases} \sum_{i=1}^{n}\phi_i\mu_i + \sum_{j=1}^{m}\psi_j\nu_j & \text{if } c_{ij}-\phi_i-\psi_j \geq 0 \\ -\infty & \text{otherwise} \end{cases}$$

Adding the constraint of $\phi_i+\psi_j \leq c_{ij}$ for all (i,j) to the dual problem, we acquire

$$\mathcal{D}(\phi_i,\psi_j) = \max\left\{\sum_{i=1}^{n}\phi_i\mu_i + \sum_{j=1}^{m}\psi_j\nu_j : c_{ij}\geq\phi_i+\psi_j\ \forall(i,j)\right\}.$$

The Lagrange multiplier function $\Lambda(\pi;\phi,\psi)$ can be rewritten

$$\Lambda(\pi;\phi,\psi) = \langle c,\pi\rangle - \phi^\intercal(\pi\mathbb{I}-\mu) - \psi^\intercal(\pi^\intercal\mathbb{I}-\nu),$$

where $\langle\cdot,\cdot\rangle$ indicates the Euclidean inner product on $\mathbb{R}^n$ and $\mathbb{I}$ the vector consisting of all ones. Taking the gradient of the Lagrangian with respect to π, shifting to vector notation, yields

$$0 = \nabla_x\Lambda(\pi;\phi,\psi) = c - \phi\mathbb{I}^\intercal - \mathbb{I}\psi^\intercal.$$

The Karush-Kuhn-Tucker infers that x is a critical point for minimizing $\Lambda(\pi;\phi,\psi)$ whenever there exist multipliers ϕ and ψ such that $0 = \nabla_\pi\Lambda(\pi;\phi,\psi)$. To this end, say π^* is a critical point of $\Lambda(\pi;\phi,\psi)$ for which $c-\phi\mathbb{I}^\intercal-\mathbb{I}^\intercal\psi = 0$. Substituting this into the Lagrange multiplier function Λ, we get

$$\Lambda(\pi_{ij};\phi_i,\psi_j) = \sum_{i=1}^{n}\phi_i\mu_i + \sum_{j=1}^{m}\psi_j\nu_j.$$

The constraint $A^\intercal\lambda \leq c$ turns out to be $\phi_i+\psi_j \leq c_{ij}$ for which we have let $b = (\mu_1,\dots,\mu_n,\nu_1,\dots,\nu_m)$ be in $\mathbb{R}^n\times\mathbb{R}^m$ and $A^\intercal = \sum_{j,i=1}^{m,n}\pi_{ji}$ and $\lambda := (\phi_1,\dots,\phi_n;\psi_1,\dots,\psi_m) \in \mathbb{R}^n\times\mathbb{R}^m$ the Lagrange multipliers. $\square$

Remark 1.1. That (5) is convex follows easily from the following computation. Firtly, notice that $\Pi_{ij}\neq\emptyset$. Just simply take the sum $\pi_{ij} = \mu_i+\nu_j$ for all (i,j), which is in Π_{ij}. That Π_{ij} is convex, let T_{ij} and S_{ij} for all (i,j) be members of Π_{ij}. Then

$$\sum_j T_{ij} = \mu_i,\ \sum_i T_{ij} = \nu_j\ \forall(i,j) \quad\text{and}\quad \sum_j S_{ij} = \mu_i,\ \sum_i S_{ij} = \nu_j\ \forall(i,j).$$

We have to show that $R_{ij} := (1-t)S_{ij} + tT_{ij}$ for all $0 \le t \le 1$ belongs to Π_{ij}. To this end,

$$\begin{aligned}\sum_i R_{ij} &= (1-t)\sum_i S_{ij} + t\sum_i T_{ij}\\ &= (1-t)\nu_j + t\nu_j\\ &= \nu_j \in \Pi_{ij}\ \forall j.\end{aligned}$$

Similarly,

$$\begin{aligned}\sum_j R_{ij} &= (1-t)\sum_j S_{ij} + t\sum_j T_{ij}\\ &= (1-t)\mu_i + t\mu_i\\ &= \mu_i \in \Pi_{ij}\ \forall i.\end{aligned}$$

Therefore, $R_{ij} \in \Pi_{ij}$ for all (i,j).

Remark 1.2. The solution above shows that one can recover the relation

$$\max\left\{\sum_{i,j=1}^{n,m} b_{i,j}y_{i,j} :\ Ay \le c\right\} = \min\left\{\sum_{i,j=1}^{n,m} c_{i,j}\pi_{i,j} : A\pi = b;\ \pi \ge 0\right\}.$$

The KKT conditions of the minimization problem above were used to solve for the multipliers of the maximization one.

The computation above, in the discrete setting, belongs to the more general mathematical theory of Fenchel-Rockafellar duality or the Legendre-Fenchel transform. In short, this means that given a convex function $\varphi : \Omega \to (-\infty, +\infty]$, the Legendre-Fenchel transform of φ is the function φ^*, and it is defined by

$$\varphi^*(z^*) := \max_{z\in\mathbb{R}^n}\{\langle z^*, z\rangle - \varphi(z)\}. \tag{7}$$

2. Convex sets

Before we define what a convex set is, we need some preliminary set up to set the stage. Let $\mathbb{R}^n$ be the Euclidean space, and the usual vector space of real n-tuples $x = (x_1, x_2, \ldots, x_n)$. The (Euclidean) inner product of two vectors x and y is defined by

$$\langle x, y\rangle = x_1y_1 + \cdots + x_ny_n.$$

Furthermore, we obtain some properties.

Proposition 2.1. *Given x, y, and z in $\mathbb{R}^n$, we have*
(1)

$$\langle x, x\rangle \ge 0 \quad \textit{and} \quad \langle x, x\rangle = 0 \textit{iff} \quad x = 0.$$

(2) $\langle x, y\rangle = \langle y, x\rangle$
(3) $\langle \lambda x, y\rangle = \lambda\langle x, y\rangle$ *(*$\lambda \in \mathbb{R}$*)*
(4) $\langle x + z, y\rangle = \langle x, y\rangle + \langle z, y\rangle$
(5) Euclidean norm:

$$\begin{aligned}\|x\| &= \sqrt{\langle x, x\rangle} \\ &= \sqrt{x_1^2 + \cdots x_n^2}.\end{aligned}$$

Moreover, for any $x, y \in \mathbb{R}^d$*,*
(6) $\|x\| \geq 0$*,* $\|x\| = 0$ *iff* $x = 0$*.*
(7) $\|\lambda x\| = |\lambda|\|x\|$ *(*$\lambda \in \mathbb{R}$*).*
(8) $\|x + y\| \leq \|x\| + \|y\|$ *(Triangle inequality)*
(9) $|\langle x, y\rangle| \leq \|x\|\,\|y\|$ *(Cauchy-Schwartz).*

We are now prepared to define a convex set.

Definition 2.2. A subset K of $\mathbb{R}^n$ is *convex* if

$$(1-\lambda)x + \lambda y \in K$$

whenever $x, y \in K$ and $\lambda \in [0, 1]$.

From this definition we can se the following. For any $x, y \in \mathbb{R}^n$, the *line segment* connecting x to y is defined by

$$[x, y] := \{(1-\lambda)x\lambda y : 0 \leq \lambda \leq 1\}.$$

Notice that if $\lambda = 0$, then $(1-\lambda)x + \lambda y = x$; when $\lambda = 1$, $(1-\lambda)x + \lambda y = y$; when $\lambda = \frac{1}{2}$, $(1-\lambda)x + \lambda y = \frac{x+y}{2}$—the midpoint of the interval connecting x and y.

Remark 2.3. A subset K of $\mathbb{R}^n$ is *convex* iff the line segment connecting x and y, $[x, y] \subset K$, whenever $x, y \in K$.

Examples. 1. Let $K := \{x \in \mathbb{R}^n : \|x\| \leq 1\}$. To show that K is convex, fix any $x, y \in K$, $\lambda \in (0, 1)$. Then

$$\|x\| \leq 1, \|y\| \leq 1.$$

An application of Proposition 2.1 (8) and (7) show

$$\begin{aligned}\|(1-\lambda)x + \lambda y\| &\leq \|(1-\lambda)x\| + \|\lambda y\| \\ &\leq (1-\lambda)\|x\| + \lambda\|y\| \\ &\leq 1 - \lambda + \lambda = 1.\end{aligned}$$

Therefore, we can see that $(1-\lambda)x + \lambda y \in K$, and consequently, K is a convex set.

2. Given any $v \in \mathbb{R}^n$, define the set

$$K := \{x \in \mathbb{R}^n : \langle v, x\rangle \leq r,\ r \in \mathbb{R}\}. \tag{8}$$

Then K is a convex set. Indeed, fix any $x, y \in K$, $\lambda \in (0, 1)$. Then, by definition

$$\langle v, x\rangle \leq r, \quad \langle v, y\rangle \leq r.$$

Then according to Proposition 2.1 (2)-(4)

$$\begin{aligned}\langle v,(1-\lambda)x+\lambda y\rangle &= \langle (1-\lambda)x+\lambda y, v\rangle \\ &= \langle (1-\lambda)x, v\rangle + \langle \lambda y, v\rangle \\ &= (1-\lambda)\langle x, v\rangle + \lambda\langle y, v\rangle \\ &= (1-\lambda)\langle v, x\rangle + \lambda\langle v, y\rangle \\ &= (1-\lambda)r + \lambda r = r.\end{aligned}$$

But this says that the line segment $(1-\lambda)x+\lambda y$ belongs in K. So therefore, K is convex.

2.1. **Algebra of Convex sets.** Next, we want to talk about some convex sets operations. For starters, if C is a convex subset of $\mathbb{R}^n$, then convexity of such set is invariant under translation and scalar multiplication. Namely, every translate $C + a$ ($a \in \mathbb{R}$) and scalar multiple $\lambda C + \{\lambda z : z \in C\}$ are convex as well. (exercise).

Let K_1 and K_2 be convex subsets of $\mathbb{R}^n$. Define the product set of convex sets as

$$K_1 \times K_2 := \{(k_1, k_2) : k_1 \in K_1, k_2 \in K_2\} \subset \mathbb{R}^n \times \mathbb{R}^n. \tag{9}$$

Proposition 2.4. *If K_1 is a convex set in $\mathbb{R}^n$ and K_2 is a set in $\mathbb{R}^n$, then $K_1 \times K_2$ is a convex set.*

Proof. Let $K := K_1 \times K_2$. We wish to show K is convex. To this end, fix any $w = (w_1, w_2) \in K$ and $z = (z_1, z_2) \in K_2$. Then, we see that

$$w_1, z_1 \in K_1, \quad \text{and} \quad w_2, z_2 \in K_2.$$

For any $\lambda \in (0, 1)$, we have that

$$\begin{aligned}(1-\lambda)w + \lambda z &= (1-\lambda)(w_1, w_2) + \lambda(z_1, z_2) \\ &= ((1-\lambda)w_1, (1-\lambda)w_2) + (\lambda z_1, \lambda z_2) \\ &= ((1-\lambda)w_1 + \lambda z_1, (1-\lambda)w_2 + \lambda z_2) \in K\end{aligned}$$

by noticing that $(1-\lambda)w_1 + \lambda z_1 \in K_1$ and $(1-\lambda)w_2 + \lambda z_2 \in K_2$. Consequently, K is convex. □

Affine sets. Let A be an $n \times m$ matrix and $b \in \mathbb{R}^n$. Define

$$B(x) := Ax + b, \ x \in \mathbb{R}^m.$$

Note that Ax is a $n \times 1$ matrix, then $B : \mathbb{R}^m \to \mathbb{R}^n$. So $B(x)$ is called an *affine mapping*. I.e., a sum of a linear mapping with a constant. (Pictures: line $y = 2x + 1$). The goal here is to show that convexity of sets is preserved under the mapping B.

Proposition 2.5. *Let $B : \mathbb{R}^m \to \mathbb{R}^n$ be an affine mapping.*
(1) If $K \subset \mathbb{R}^m$ is convex then the image

$$B(K) := \{B(x) : x \in K\}$$

is a convex set in $\mathbb{R}^n$.
(2) If E is convex set in $\mathbb{R}^n$, then the inverse image

$$B^{-1}(E) := \{x \in \mathbb{R}^m : B(x) \in E\}$$

is convex.

Proof. (1) Fix any $a, b \in B(x)$ and $0 < \lambda < 1$. Then we have

$$B(x) = a \quad \text{for some} \quad x \in K \quad B(z) = b \quad \text{for some} \quad z \in K.$$

Then

$$\begin{aligned}(1-\lambda)a + \lambda b &= (1-\lambda)B(x) + \lambda B(z)\\ &= (1-\lambda)(Ax+b) + \lambda(Az+b)\\ &= A((1-\lambda)x + \lambda z) + b\\ &= B((1-\lambda)x + \lambda z) \in B(K),\end{aligned}$$

since $(1-\lambda)x + \lambda z \in K$.
Part (2) exercise. □

2.2. **More properties of Convex sets.** Consider two convex subsets C_1 and C_2 of $\mathbb{R}^n$. The sum of such sets C_1 and C_2 is denoted and defined by

$$C_1 + C_2 := \{x_1 + x_2 : \ x_1 \in C_1, x_2 \in C_2\}. \tag{10}$$

The first result says that the sum of two convex sets is convex.

Theorem 2.6. *If C_1 and C_2 are convex subsets of $\mathbb{R}^n$, then so is their sum, $C_1 + C_2$. Moreover, if $\alpha \in \mathbb{R}$, $\alpha C := \{\alpha x : x \in C\}$ is convex.*

Proof. Let x and y be two points in $C_1 + C_2$. Then, there exist vectors $x_1, y_1 \in C_1$ and $x_2, y_2 \in C_2$ such that

$$x = x_1 + x_2 \quad \text{and} \quad y = y_1 + y_2.$$

For $\lambda \in (0,1)$, we have

$$(1-\lambda)x + \lambda y = [(1-\lambda)x_1 + \lambda y_1] + [(1-\lambda)x_2 + \lambda y_2];$$

and b the convexity of C_1 and C_2, we have

$$(1-\lambda)x_1 + \lambda y_1 \in C_1 \quad \text{and} \quad (1-\lambda)x_2 + \lambda y_2 \in C_2.$$

This says $(1-\lambda)x + \lambda y \in C_1 + C_2$.

The second conclusion is done in exercises. □

Next, we discuss a very important concept in *convex analysis.*

Let $w_1, \ldots w_n$ be elements on $\mathbb{R}^n$. An element x of $\mathbb{R}^n$ is called an *convex combination* of $w_1, \ldots w_n$ if

$$x = \sum_{i=1}^{n} \lambda_i w_i,$$

where $\lambda_i \geq 0$ and $\sum_{i=1}^{n} \lambda_i = 1$.

Example. Take $x = \frac{1}{3}w_1 + \frac{1}{3}w_2 + \frac{1}{3}w_3$ is a convex combination of w_1, w_2, and w_3. We can see this from the definition given. Each $\lambda_i = \frac{1}{3} \geq 0$ for $i = 1, 2, 3$. This example forms what is called an $n = 3$-*symplex.*

Proposition 2.7. *A subset C of $\mathbb{R}^n$ is convex iff it contains all convex combinations of its elements.*

Proof. Suppose firstly that C contains all convex combinations of its elements. Our goal is to show that C is convex. Fix any $x, y \in C$ and $\lambda \in (0,1)$. Then $(1-\lambda)x + \lambda y$. Denoting $\lambda_1 = \lambda$ and $1 - \lambda = \lambda_2$, then clearly $\lambda_1, \lambda_2 \geq 0$ and their sum $\lambda_1 + \lambda_2 = 1$. Therefore, $(1-\lambda)x + \lambda y \in C$, since it is a convex combination of $x, y \in C$. Therefore, C is convex.

For the converse, simply take any finite set $\{w_1, w_2, \ldots, w_k\} \subset C$ and parameters $\lambda_1, \ldots, \lambda_k \geq 0$ with $\lambda_1 + \lambda_k = 1$ then $\lambda_1 w_1 + \cdots + \lambda_k w_k$ must be contained in C: this can be seen by induction on k and using

$$\lambda_1 w_1 + \cdots + \lambda_k w_k = (1-\lambda_k)\left(\frac{\lambda_1}{1-\lambda_k} w_1 + \cdots + \frac{\lambda_{k-1}}{1-\lambda_k} w_{k-1}\right) + \lambda_k w_k \quad \text{for } \lambda_k < 1.$$

□

Last concept of convex set. Every intersection of convex sets is convex, and $\mathbb{R}^n$ is convex. Thus, we have

Definition 2.8. For any $C \subset \mathbb{R}^n$, the "smallest" convex set containing C is called the *convex hull* of C, and can be constructed as the intersection of all convex sets that contain C:

$$\text{conv(C)} := \bigcap \{K \subset \mathbb{R}^n : C \subset K,\ K \text{ is convex}\}$$

3. Convex functions

In this lecture we discuss *convex* functions.

Definition 3.1. Let $f : \mathbb{R}^n \to \mathbb{R} \cup \{\pm\infty\}$ be a function. f is called *convex* if it holds

$$f((1-\lambda)x + \lambda y) \leq (1-\lambda)f(x) + \lambda f(y) \quad \forall x, y \in \mathbb{R}^n\ \lambda \in (0,1).$$

Graph a convex function.

No we discuss some basic properties of convex functions.

Let $f : \mathbb{R}^n \to (-\infty, +\infty]$ be a convex function. We define the *domain* of f is defined by

$$\text{dom}(f) := \{x \in \mathbb{R}^n : f(x) < +\infty\}. \tag{11}$$

The *epigraph* of f is the set

$$\text{epi}(f) := \{(x, \mu) \in \mathbb{R}^n \times \mathbb{R} : \mu \geq f(x)\}. \tag{12}$$

Examples. 1. Take $f : \mathbb{R} \to (-\infty, +\infty]$ be given as

$$f(x) = \begin{cases} 0 & |x| \leq 1 \\ +\infty & o.w \end{cases} \tag{13}$$

graph: the domain of f is $(-1, 1)$, while epi$f$ is the swet–shaded region between $x = -1$ and $x = 1$, and $x \geq 0$. Hence,

$$\text{dom}(f) = [-1, 1], \quad \text{epi} := [-1, 1] \times [0, \infty).$$

Proposition 3.2. *Let* $f : \mathbb{R}^n \to (-\infty, \infty]$.
(1) If f *is convex, then the doman of* f *is convex as well.*
(2) f *is convex iff*

$$f((1-\lambda)x + \lambda y) \le (1-\lambda)f(x) + \lambda f(y) \quad \forall x, y \in dom(f)\ \forall \lambda \in (0,1).$$

Remark 3.3. The difference between the definition of convexity and (2) is we only require that the above inequality be true for x, y in $\mathrm{dom}(f)$, instead of the whole Euclidean space $\mathbb{R}^n$.

Proof. The first thing that we wish to prove is (1): to show that $\mathrm{dom}(f)$ is a convex in the sense of the previous section. To this end, suppose f is convex (as definition suggests). Fix any $x, y \in \mathrm{dom}(f)$, $\lambda \in (0,1)$. Then by definition of the domain of f, we have that $f(x) < \infty$ and $f(y) < \infty$. Then

$$f((1-\lambda)x + \lambda y) \le (1-\lambda)f(x) + \lambda f(y) < \infty.$$

Moreover, by the definition, so $(1-\lambda)x + \lambda y \in \mathrm{dom}(f)$. Therefore, $\mathrm{dom}(f)$ is convex.

(2) Exercise. Hint, if f is convex, then the inequality is true for $x, y \in \mathrm{dom}(f)$ $\lambda \in (0,1)$. As $\mathrm{dom}(f) \subset \mathbb{R}^n$. For the other direction, suppose the inequality holds true for the conditions. Taking any $x, y \in \mathbb{R}^n$. If x, y are in the domain of f, then the result holds. On the other hand if x, y fail to be in $\mathrm{dom}(f)$; in this case, say $x \notin \mathrm{dom}(f)$, then $f(x) = +\infty$. Then the right hand side of the inequality is $+\infty$ which is always $\ge$ the left hand side. □

3.1. **Geometric Characterization for Convex function.** We look at the epi graph, namely if the epi graph of f is convex then automatically f is convex.

Proposition 3.4. *Let* $f : \mathbb{R}^n \to (-\infty, \infty]$ *be a convex function. Then* f *is convex iff* $epi(f) \subset \mathbb{R}^n \times \mathbb{R} = \mathbb{R}^{n+1}$ *is convex.*

Proof. First suppose that f is convex. We have to show that $epi(f) \subset \mathbb{R}^{n+1}$ is a convex set. Fix $(x, \mu), (y, \nu) \in \mathrm{epi}(f)$, and $\lambda \in (0,1)$. Then

$$f(x) \le \mu,\ f(y) \le \nu.$$

We have, then,

$$\begin{aligned} f((1-\lambda)x + \lambda y) &\le (1-\lambda)f(x) + \lambda f(y) \\ &\le (1-\lambda)\mu + \lambda\nu. \end{aligned}$$

This means that, by definition of the epigraph of f,

$$((1-\lambda)x + \lambda y, (1-\lambda)\mu + \lambda\nu) \in \mathrm{epi}(f).$$

We see this from the following simplification:
$((1-\lambda)x + \lambda y, (1-\lambda)\mu + \lambda\nu) = (1-\lambda)(x, \mu) + \lambda(y, \nu)$ which is in $\mathrm{epi}(f)$. Hence, $\mathrm{epi}f$ is convex.

Conversely, suppose that $\mathrm{epi}f$ is convex. Fix any $x, y \in \mathrm{dom}(f)$ and $\lambda \in (0,1)$. We shall show that

$$f((1-\lambda)x + \lambda y) \le (1-\lambda)f(x) + \lambda f(y).$$

We have $f(x) < \infty$ and $(x, f(x)) \in \text{epi}(f)$ (denoting $f(x) = \mu$). Similarly, $(y, f(y)) \in \text{epi}(f)$. Then

$$(1-\lambda)(x, f(x)) + \lambda(y, f(y)) \in \text{epi}(f).$$

However,

$$(1-\lambda)(x, f(x)) + \lambda(y, f(y)) = ((1-\lambda)x + \lambda y, (1-\lambda)f(x) + \lambda f(y)) \in \text{epi}(f);$$

Denoting $z = (1-\lambda)x + \lambda y$ and $\tau = (1-\lambda)f(x) + \lambda f(y)$, $f(z) \leq \tau$. Therefore,

$$f((1-\lambda)x + \lambda y) \leq (1-\lambda)f(x) + \lambda f(y).$$

But this says that f is convex! □

The following is great example of a convexity property.

Proposition 3.5. *Let $f : \mathbb{R}^n \to (-\infty, \infty]$ be a convex function. Then f is convex iff for any number $\lambda_i > 0$ for $i = 1, 2, \ldots, m$ and any element $x_i \in \mathbb{R}^n$ for $i = 1, 2, \ldots, m$ with $\sum_{i=1}^m \lambda_i = 1$, then*

$$f\left(\sum_{i=1}^m \lambda_i x_i\right) \leq \sum_{i=1}^m \lambda_i f(x_i). \tag{14}$$

Remark 3.6. Notice that from the definition convexity property of convex functions, (14) is true for m a positive integer. when $m = 2$: (14) holds true if f is convex. Now we can prove a more general property: if f is a convex function. Namely, if f is convex, then (14) holds true for all $m \in \mathbb{N}$. Conversely, if the inequality holds for all $m \in \mathbb{N}$, then f is convex.

Proof. One direction is easy. That is, if (14) holds for all m, $\lambda_i > 0$, and $x_i \in \mathbb{R}^n$, then for the special case $m = 2$, we have

$$f(\lambda_1 x_1 + \lambda_2 x_2) \leq \lambda_1 f(x_1) + \lambda_2 f(x_2).$$

But this is exactly the definition of convexity of f given in (3.1). Therefore, we need only show that if f is convex, then (14) holds.

To that end, suppose that f is convex. Fix $\lambda_i > 0$, $x_i \in \mathbb{R}^n$ with $i = 1, 2, \ldots, m$ and $\sum_{i=1}^m \lambda_i = 1$. According to Proposition 3.4 as f is convex, then so is epi(f). One property taken for granted is the fact that *affine subspaces* can be described as the set of all affine combinations of a finite set of points:

$$F := \left\{\mathbf{v} \in \mathbb{R}^n : \mathbf{v} = \lambda_1 v_1 + \lambda_2 v_2 + \cdots + \lambda_n v_n, \text{ for } \lambda_i > 0, \sum_{i=1}^n \lambda_i = 1\right\}.$$

That is, if we take any finite number of points in the convex set—epi(f)—then any convex combination of those points will remain inside the convex set. Now we can assume without loss of generality that $x_i \in \text{dom}(f)$ for all $i = 1, \ldots, m$. Since, if not $x_i \notin \text{dom}(f)$, then $f(x_i) = \infty$, and thus the right hand side of (14) is ∞ which is greater than or equal to the left hand side and Jensen inequality is satisfied. Then as $x \in \text{dom}(f)$ we see that

$$(x_i, f(x_i)) \in \text{epi}(f)\ \forall i.$$

Since we know that epi(f) is convex, then the convex combination

$$\sum_{i=1}^{m} \lambda_i (x_i, f(x_i)) \in \operatorname{epi}(f)$$

Furthermore, we can rewrtie this as follows

$$\sum_{i=1}^{m} \lambda_i (x_i, f(x_i)) = \left(\sum_{i=1}^{m} \lambda_i x_i, \sum_{i=1}^{m} \lambda_i f(x_i) \right) := (z, \tau) \in \operatorname{epi}(f).$$

The definition of the epigraph of f tells you that $f(z) \leq \tau$. But this is

$$f\left(\sum_{i=1}^{m} \lambda_i x_i \right) \leq \sum_{i=1}^{m} \lambda_i f(x_i) \ \forall m$$

□

4. Subdifferential

We start the section by discussing the *subgradient* of a convex function. In calculus we study the derivatives of differentiable functions, but what happens when we encounter functions that are not differentiable? In optimization problems, finding extremal points is important. Nondifferentiable functions happen to arise in many areas of applied mathematics and is germane to optimization. For example, the absolute value function, $f(x) = |x|$ is not differentiable at $x = 0$—we have a corner there. However, we can talk about its (sub)tangent line at $x = 0$ that falls under the graph of $|x|$. The subtlety is that there are infinitely many options for such a "subtangent" line. To wit, unlike calculus, a convex function with a "corner point" can have variable slopes.

graph a general function with corners and give geometrical intuition regarding the *subgradient* and *subdifferential*.

Definition 4.1. Let $f : \mathbb{R}^n \to (-\infty, \infty]$ be a convex function and let $z \in \operatorname{dom}(f)$. A vector $v \in \mathbb{R}^n$ is a *subgradient* of f at z if

$$\langle v, x - z \rangle \leq f(x) - f(z) \ \forall x \in \mathbb{R}^n. \tag{15}$$

The collection of all subgradients of f at z is called the *subdifferential* of f at z. It is denoted by $\partial f(z)$.

Notice that the subdifferential consists of the slopes of " all tangents line" to f. It is a set-valued function. For the regular tangent line to a "smooth" curve there is one point, whereas for the absolute value at 0, there are infinitely many "subtangent" lines to it.

Examples. (1) Take the absolute value function $f(x) = |x|$. First thing to notice is that f is not differentiable at $x = 0$. Notice that we have options of putting a subtangent line at $x = 0$. In fact, there are many ways to accomplish this, and this is what will give us the subgradient of this function at zero.

Claim. Given $f(x) = |x|, x \in \mathbb{R}$. Then $\partial f(0) = [-1, 1]$.

We use the definition. Fix any $v \in \partial f(0)$. Then

$$\langle v, z \rangle \leq f(x) - f(z) \Longleftarrow v(x - 0) \leq f(x) - f(0) \Longleftrightarrow vx \leq |x| \quad \forall x \in \mathbb{R}.$$

Since this is true for all x, then it is true for all $x = 1$. In this case we have,

$$v \cdot 1 \leq |1| = 1.$$

The above inequality also holds for $x = -1$. In this case we ave,

$$v(-1) \leq |-1| = 1.$$

Or equivalently,

$$-v \leq 1 \Longleftarrow -1 \leq v.$$

Combining this inequality with the previous one says $-1 \leq v \leq 1$. This then tells you that $\partial f(0) \subset [-1, 1]$. To get equality, we must show the reverse inequality.

To that end, fix any $v \in [-1, 1]$. Then

$$|v| \leq 1.$$

For any $x \in \mathbb{R}$, we have the following

$$v(x - 0) = vx \leq |vx| = |v|\ |x| \leq 1\ x = x \leq |x|.$$

But this says

$$v(x - 0) \leq |x| - |0|.$$

Hence, $v \in \partial f(0)$, and hence $[-1, 1] \subset \partial f(0)$, for which equality now follows at once.
(2) Consider the Euclidean norm: $f(x) = \|x\|$, where $\|x\| = \sqrt{\langle x, x\rangle} = \sqrt{x_1^2 + \cdots + x_n^2}$.

Claim. $\partial f(0) = \mathbb{B} := \{v \in \mathbb{R}^n : \|v\| \leq 1\}$, called the closed unit ball.

Fix any $v \in \partial f(0)$. Then

$$\langle v, x - 0\rangle \leq f(x) - f(0) = \|x\|\ \forall x \in \mathbb{R}^n. \tag{16}$$

This is equivalent to

$$\langle v, x\rangle \leq \|x\| \quad \forall x \in \mathbb{R}^n. \tag{17}$$

So for $v = x$, we have the following

$$\langle v, v\rangle \leq \|v\| \implies \|v\|^2 \leq \|v\|.$$

This implies

$$\|v\| \leq 1.$$

If $v = 0$, then certainly $0 \leq 1$ or equivalently $0 \leq 0$. But this says $v \in \mathbb{B}$. Consequently, $\partial f(0) \subset \mathbb{B}$.

For the reverse inclusion, fix $v \in \mathbb{B}$. Then $\|v\| \leq 1$. Then for any $x \in \mathbb{R}^n$,

$$\langle v, x - 0\rangle = \langle v, x\rangle \leq \|v\|\|x\| \leq \|x\| = \|x\| - \|0\| = f(x) - f(0) \quad \forall x \in \mathbb{R}^n.$$

Hence, $v \in \partial f(0)$, and the opposite inclusion holds, and thus we get $\partial f(0) = \mathbb{B}$.

Note that both these examples involved real-values. Note that both these examples involved real-values. In the next exercise we take into account the extended plane.

(3) Let K be a nonempty subset of $\mathbb{R}^n$. The indicator function associated with K is defined by

$$\delta_K(x) = \begin{cases} 0 & x \in K \\ +\infty & x \notin K \end{cases} \tag{18}$$

A few remarks are in order. The first, is that $\text{dom}(\delta_K) = K$. The $\text{epi}(\delta_K) = K \times [0, \infty)$. For any $z \in K$, $\partial\delta_K(z) = N(z, K)$. Here, the *normal cone* $N(z, K)$ is defined by

$$N(z, K) := \{v \in \mathbb{R}^n : \langle v, x - z\rangle \leq 0 : \forall x \in K\} \tag{19}$$

Indeed, fix $v \in \partial\delta_K(z)$, then

$$\begin{aligned} \langle v, x - z\rangle &\leq \delta_K(x) - \delta_K(z) \ \forall x \in \mathbb{R}^n \\ &= \delta_K(x) \ \forall x \in \mathbb{R}^n. \end{aligned}$$

But for any $x \in K$, we have

$$\langle v, x - z\rangle \leq \delta_K(x) = 0.$$

Thus, $v \in N(z, K)$. So $\partial\delta_K(z) \subset N(z, K)$.

4.1. **Subdifferentiable maximum rule.** Given three differentiable functions f_i, for $i = 1, 2, 3$; that intersect each other at some points, the maximum (max for short) defined by

$$f(x) := \max\{f_1(x), f_2(x), f_3(x)\} \tag{20}$$

may not be differentiable function. In fact, consider the following example. Let $f_1(x) = -x$, $f_2(x) = 1$, and $f_3(x) = x$. These functions are clearly differentable, but their max $f(x) := \max\{-x, 1, x\}$ is not. Look at the graph. Since any convex function, not necessarily differentiable, can be described in terms of "subgradient" vectors, which correspond to supporting hyperplanes to the epigraph of f, a natural question that arises is as follows.

How can we represent the subdifferential of the maximum function (20) at a point z in terms of the subdifferential? Namely, what is

$$\partial f(z) = \partial f_i(z) \quad \text{for } i = 1, 2, 3?$$

Let $f_i : \mathbb{R}^n \to (-\infty\, \infty]$ for $i = 1, \ldots, m$, be convex functions. Define the following function

$$f(x) := \max\{f_i(x) : \quad i = 1, \ldots, m\}.$$

Given $z \in \mathbb{R}^n$, we define the *active index set* at a point z as follows

$$I(z) := \{i = 1, \ldots, m : \ f_i(z) = f(z)\}.$$

In order to understand the active index set, we shall look at the previous example where the functions were $f_1(x) = -x$, $f_2(x) = 1$, and $f_3(x) = x$, and for $z = 1$. Indeed substituting $z = 1$ in the functions $f_i(z)$, for $i = 1, 2, 3$, we get $f_1(1) = -1$, $f_2(1) = 1$, and $f_3(1) = 1$. Thus,

$$f(1) = \max\{-1, 1, 1\} = 1.$$

Hence, using the index set, the set that tells you where—at what index—the max was attained, we see the following: $I(1) = \{2, 3\}$.

Let's do another example (*n class—here good idea to ask for the wrong answer*), where this time $z = -1$. Here substituting $z = -1$ in $f_i(z)$ for $i = 1, 2, 3$, we see

$$f(-1) = \max\{1, 1, -1\} \implies I(-1) = \{1, 2\}.$$

So similarly, we can obtain the formula for computing the active index set I for any z in the domain of the function. We will next get the main formula for this discussion. We first have to assume that each function f_i is continuous at z to obtain a formula in order to represent the subdifferetinal of the function f in terms of the function f_i for $i = 1, \ldots, m$.

Suppose each f_i for $i = 1, \ldots, m$ is *continuous* at z, then we have the following

$$\partial f(z) = \operatorname{conv}\left\{ \bigcup_{i \in I(x)} \partial f_i(z) \right\}.$$

We will look at examples to understand some computations and what the formula says in terms of subdifferentials. In order to understand the formula, let's return to our previous example.

Example. Consider $f_1(x) = -x$, $f_2(x) = 1$, and $f_3(x) = x$ and $z = 1$. Since the function f_i for $i = 1, 2, 3$ are continuous, we can apply the formula above. Since we already computed the active index set for $z = 1$, we have

$$I(z) = \{2, 3\}.$$

Now looking at the function

$$f_2(x) = 1 \implies \partial f_2(z) = \{0\}.$$

Similarly, for $f_3(x) = x$, then

$$\partial f_3(z) = \{1\}.$$

Because of the formula above, we have

$$\partial f(z) = \operatorname{conv}\{\partial f_2(z) \cup \partial f_3(z)\} = \{0, 1\} = [0, 1].$$

The above construction was merely meant to see the importance of the convexity of a function with the convex hull of a set.

5. Legendre-Fenchel Transform

We start with a graphical example to see where the *Legendre-Transform* "comes" from. Consider the graph of a convex function, and look for a particular slope. In fact, we will think, in some sense, of the slopes y as our independent variable. Let's continue by drawing a picture.

From the picture we get a nice formula for the y-intercept of this tangent line. It is $xy - f(x)$. Let's call this g. Namely, given x, let

$$g(y) = xy - f(x) \quad \forall y \in \mathbb{R}^n.$$

There is a relationship between y and x, namely that $y = f'(x)$—the slope of f at x. Let figure out if we can come up with some conditions for g. Since the relationship between x and y is the derivative, one can take a derivative of $xy - f(x)$ and set it equal to zero. Thus, this extra condition can also be obtained from differentiating g with respect to x and setting equal to zero. Recall that when we take derivative and set equal to zero is done when we wish to *optimize* a function, i.e., when we maximize or minimize a function. The conclusion is that this strongly resembles max or min g. In what comes next, we will need to make use of supremum and infimum. The supremum of a set is its least upper bound and the infimum is its greatest lower bound. But for all intents and purposes, we can think of supremum as a maximum and infimum as a minimum provided that our domain is closed and bounded (i.e., compact) such as $[0, 1]$.

Since f is convex, and by definition a *concave* function f on $\mathbb{R}^n$ is a function whose negative is convex, then we want to maximize g with respect to x (Otherwise, we would want to minimize as it is natural to minimize a convex function over a convex set). So we are interested in

$$g(y) := \max_x \{xy - f(x)\}.$$

This is the **Legendre-Fenchel Transform**.

Definition 5.1. Let $f : \mathbb{R}^n \to (-\infty, \infty]$ be a function (not necessarily convex), then the *Legendre-Fenchel* transform (or Fenchel *conjugate*) of f denoted by f^* is defined by

$$\begin{aligned} f^*(y) &= \sup\{\langle x, y\rangle - f(x) : x \in \mathbb{R}^n\} \\ &= -\inf\{f(x) - \langle x, y\rangle :\ x \in \mathbb{R}^n\} \quad y \in \mathbb{R}^n. \end{aligned} \tag{21}$$

Then, $f^* : \mathbb{R}^n \to [-\infty, +\infty]$. From the definition (21), f^* can be $\pm\infty$. Let us explore more properties of the Fenchel congujate of f. One important property that we shall soon see is that if we assume f is not necessarily convex, with nonempty domain, then its Fenchel congujate is always convex.

Theorem 5.2. *Let $f : \mathbb{R}^n \to (-\infty, \infty]$ be a function. Suppose that $dom(f) \neq \emptyset$. This means that there is $f(z) < +\infty$ for $z \in \mathbb{R}^n$. Then*

$$f^* : \mathbb{R}^n \to (-\infty, \infty]$$

is a convex function.

Proof. For any $y \in \mathbb{R}^n$, by definition we have

$$\begin{aligned} f^*(y) &= \sup\{\langle y, x\rangle - f(x)\ : x \in \mathbb{R}^n\} \\ &\geq \langle y, z\rangle - f(z) > -\infty. \end{aligned}$$

Next, we aim to show that f^* is actually a convex function. First notice that if $x \notin \mathrm{dom}(f)$, then $f(x) = \infty$, and thus $\langle y, x\rangle - f(x) = -\infty$, and so we have

$$\begin{aligned} f^*(y) &= \sup\{\langle y, x\rangle - f(x) :\ x \in \mathbb{R}^n\} = \sup\{\langle y, x\rangle - f(x)\ : x \in \mathrm{dom}(f)\} \\ &= \sup\{\varphi_x(y) : x \in \mathrm{dom}(f) \quad \text{where } \varphi_x(y) := \langle y, x\rangle - f(x)\}. \end{aligned}$$

Note that for $x \in \mathrm{dom}(f)$, the function $\varphi_x(y)$ is an affine function. An *affine* function on $\mathbb{R}^n$ is a function which is finite, convex, and concave. Moreover, this $f(x)$ can be seen as a constant with respect to y. Thus $\varphi_x(y)$ resembles something like $By + b$, from Section 2.1. Therefore, $f^*(y)$ is the supremum of a family of affine functions, which is always convex. Therefore, $f^*(y)$ is a convex function. $\square$

Examples. (1) Let $f(x) = 0$ for all $x \in \mathbb{R}$. Then

$$f^*(y) = \sup_{x\in\mathbb{R}}\{xy\}.$$

Note that $f^*(y) = 0$ if $y = 0$. In this case, $\mathrm{dom}(f^*) = \{0\}$. If $y \neq 0$, then $f^*(y) = \{xy\}$ is unbounded.

(2) The same function but defined over the domain $[-1, 1]$. Hence

$$f^*(y) = \sup_{x\in[-1,1]}\{xy\} = |y|.$$

This is because if $y > 0$, then I pick $x = 1$, as that is the maximum I can get. If $y < 0$, then pick $x = -1$, that's the biggest we can get. Now $\mathrm{dom}(f^*) = \mathbb{R}$.

An elementary proof that f^* is convex follows from the definition of Definition 7. Indeed, let $x_1^*, x_2^* \in \mathrm{dom}(f^*)$ and let $\lambda \in (0, 1)$. Then be definition

$$\begin{aligned}
f^*((1-\lambda)x_1^* + \lambda x_2^*) &= \sup_{x\in\mathbb{R}^n}\{\langle(1-\lambda)x_1^* + \lambda x_2^*, x\rangle - f(x)\} \\
&= \sup_{x\in\mathbb{R}^n}\{\langle(1-\lambda)x_1^* + \lambda x_2^*, x\rangle - (1-\lambda)f(x) - \lambda f(x)\} \\
&\leq \sup_{x\in\mathbb{R}^n}\{\langle(1-\lambda)x_1^*, x\rangle - (1-\lambda)f(x)\} \\
&+ \sup_{x\in\mathbb{R}^n}\{\langle\lambda x_2^*, x\rangle - \lambda f(x)\} \\
&= (1-\lambda)\sup_{x\in\mathbb{R}^n}\{\langle x_1^*, x\rangle - f(x)\} + \lambda \sup_{x\in\mathbb{R}^n}\{\langle x_2^*, x\rangle - f(x)\} \\
&= (1-\lambda)f^*(x_1^*) + \lambda f^*(x_2^*).
\end{aligned}$$

And this is convexity. Here we return to the convexity duality of the beginning of the lecture.

Let us tie together the concept of the Legendre-Fenchel transform with that of subdifferenitiability.

From the definition, a vector $x^* \in \mathbb{R}^n$ of a convex function f at $x \in \mathbb{R}^n$ satisfies the following inequality

$$f(z) \geq f(x) + \langle x^*, z - x\rangle \quad \forall z \in \mathbb{R}^n.$$

The collection of all x^* is the set

$$\partial f(x) := \{x^* : f(z) \geq f(x) + \langle x^*, z - x\rangle \quad \forall z \in \mathbb{R}^n\}.$$

This set is a multivalued set and it is convex. From this definition, we have that the following are equivalent.

1. $x^* \in \partial f(x)$
2. $\langle x^*, z\rangle - f(z)$ attains its maximum value at $z = x$.
3. $f^*(x^*) + f(x) \leq \langle x^*, x\rangle$.
4. $f^*(x^*) + f(x) = \langle x^*, x\rangle$.

The proof is not so bad. The subgradient inequality that defines 1. can be rewritten as

$$\langle x^*, x\rangle - f(x) \geq \langle x^*, z\rangle - f^*(z) \quad \forall z.$$

A closer look at this inequality gives you 2. Since, by definition of the Legendre-Fenchel transform, the maximum in part 2 is $f^*(x^*)$, 2. equivalent to 3. or 4.