1. Convex Sets

Definition 1.1. *Convex Set*
A subset $C \subseteq \mathbb{R}^n$ is a convex set if $\forall x, y \in C, \forall \alpha \in [0,1]$

$$(1-\alpha)x + \alpha y \in C$$

Definition 1.2. *Unit Simplex*
The unit simplex $\Delta_{k-1} \subseteq \mathbb{R}^k$ is the set

$$\Delta_{k-1} = \{\alpha \in \mathbb{R}^n_+ : \sum^k \alpha_i = 1\}$$

Definition 1.3. *Convex Combination*
A convex combination of $\{x_1, x_2, ..., x_k\} \subseteq \mathbb{R}^n$ is any vector of the form

$$x = \sum^k \alpha_i x_i$$

where $\alpha = (\alpha_1, ..., \alpha_k) \in \Delta_{k-1}$

Proposition 1.1. *$C \in \mathbb{R}^n$ is convex $\iff \forall x_1, ..., x_k \subseteq C, \forall \alpha \in \Delta_{k-1}, \sum^k \alpha_i x_i \in C$*

Definition 1.4. *Set of all possible convex combinations*
For $A \in \mathbb{R}^n$, let $K(A) = \{$all possible convex combinations of points in $A\} = \{\sum^k \alpha_i x_i : k \geq 1, \{x_1, ..., x_k\} \subseteq A, \alpha \in \Delta_{k-1}\}$

Proposition 1.2. *$K(A) \supseteq A$ is convex.*

Proposition 1.3. *If $\{C_\alpha\}_{\alpha \in A}$ is a collection of convex sets in $\mathbb{R}^n$, then $\bigcap_{\alpha \in A} C_\alpha \subseteq \mathbb{R}^n$ is convex.*

Proposition 1.4. *If $C \subseteq \mathbb{R}^n, D \subseteq \mathbb{R}^m$ are convex, then $C \times D \subseteq \mathbb{R}^n \times \mathbb{R}^m$ is convex*

Definition 1.5. *Convex Hull*
For any $A \subseteq \mathbb{R}^n$, the convex hull of A is the smallest convex set that contains A. That is,

$$co(A) = \bigcap \{C \subseteq \mathbb{R}^n : C \text{ is convex}, A \subseteq C\}$$

In other words, the convex hull of A is the intersection of all convex sets that contain A. Trivially, if A is convex, then $co(A) = A$.

Proposition 1.5. *$co(A) = K(A)$*

- *$K(A)$ is like building the convex hull from within, while $co(A)$ is like shrinking the set until we get the smallest convex set.*

Proposition 1.6. *If $\{A_\alpha\}$ is a collection of convex sets, then $\bigcap_\alpha A_\alpha$ is also convex, where the intersection may be over a countable or uncountable collection.*

- *Implication: The convex hull $co(A)$ is well-defined and also convex, since the $co(A)$ is the intersection of all convex sets containing A.*

Theorem 1.1. *(Carathéodory)*
Let $A \subseteq \mathbb{R}^n, x \in co(A)$. Then, $\exists\{x_1, ..., x_{n+1}\} \subseteq A$ and $\alpha \in \Delta_n$ such that $x = \sum^{n+1} \alpha_i x_i$.

- *In essence, we are starting with a convex combination and altering it such that we get the same convex combination but with one fewer point. So, if x is some convex combination of points in A, then we can find some other convex combination using one fewer vector such that we get the same point.*

Definition 1.6. *Hyperplane*
A hyperplane in $\mathbb{R}^n$ is a set

$$H(\alpha, b) = \{x \in \mathbb{R}^n : \langle \alpha, x \rangle = b, \alpha \in \mathbb{R}^n, \alpha \neq \mathbf{0}, b \in \mathbb{R}\}$$

- *Recall that $\langle \alpha, x \rangle = \alpha^T x_i = \sum^n c_i x_i$ is the inner product.*
- *Equivalently, $\langle \alpha, x \rangle = \|\alpha\| \cdot \|x\| \cos\theta$, where θ is the angle between x and α.*
- $x \perp \alpha \iff \langle \alpha, x \rangle = 0$.
- *More intuitively, we can think of a hyperplane as an $(n-1)$ dimensional subspace shifted up/down in $\mathbb{R}^n$*

Proposition 1.7. *Take any $x_0 \in H$. Then, $\langle \alpha, x \rangle = b$, so I can write*

$$H = \{x \in \mathbb{R}^n : \langle \alpha, x - x_0 \rangle = 0, x_0 \in H\}$$

Remark 1.1.1. *In the definition of a hyperplane, we call α the "normal vector", implying that α is the vector perpendicular to the hyperplane. In particular, we see this with the definition*

$$H = \{x \in \mathbb{R}^n : \langle \alpha, x - x_0 \rangle = 0, x_0 \in H\}$$

We can then generate the entire hyperplane by varying the vector x and taking all of the vectors whose projection onto the space spanned by α is c.

Definition 1.7. *Closed Halfspace*
The space above the hyperplane and space below the hyperplane (above/below vector α), each containing the plane.

(1) $H^+ = \{x \in \mathbb{R}^n : \langle \alpha, x \rangle \geq b\}$

(2) $H^- = \{x \in \mathbb{R}^n : \langle \alpha, x \rangle \leq b\}$

Definition 1.8. *Interior of Halfspace*

(1) $int(H^+) = \{x \in \mathbb{R}^n : \langle \alpha, x \rangle > b\}$

(2) $int(H^-) = \{x \in \mathbb{R}^n : \langle \alpha, x \rangle < b\}$

Definition 1.9. *Linear Manifold*
A linear manifold L is an intersection of hyperplanes. That is,

$$L = \{x \in \mathbb{R}^n : Ax = b, A \in \mathbb{R}^{m+n}, b \in \mathbb{R}^m\}$$

We can also think of a linear manifold as a translation of a linear subspace. That is, if $x_0 \in L$, implying that $Ax_0 = b$, then

$$L = \{x \in \mathbb{R}^n : A(x - x_0) = 0\} = \{x \in \mathbb{R}^n : \langle a_i^T, x - x_0 \rangle = 0\} = x_0 + null(A)$$

If there does not exist such an x_0, then L is the empty set.

Definition 1.10. *Affine Set*
A set L is affine $\iff$ L contains all lines through any two points in the set (i.e., includes all linear combinations of elements of the set). That is,

$$L = \{\sum^k \alpha_i x_i \in L : \forall \{x_1, ..., x_k\} \subseteq L, \sum \alpha_i = 1, k \in \mathbb{N}\}$$

Note that this differs from the definition of a convex set by the restriction on the coefficients. For a convex set, $\sum \alpha_i = 1, |\alpha_i| > 0$, while for an affine set, $\sum \alpha_i = 1$, meaning that we may have negative α_i's.

Proposition 1.8. *A set L is a linear manifold $\iff$ L is an affine set. That is,*

$$\{x \in \mathbb{R}^n : \langle a_i^T, x - x_0 \rangle = 0\} = \{\sum^k \alpha_i x_i \in L : \forall \{x_1, ..., x_k\} \subseteq L, \sum \alpha_i = 1, k \in \mathbb{N}\}$$

Definition 1.11. *Polyhedron*

$$S = \{x \in \mathbb{R}^n : Ax \leq b\}$$

Note the similarity of the set definition of a polyhedron with that of a linear manifold. The only difference is that $Ax = b$ *for a linear manifold, while* $Ax \leq b$ *for a polyhedron.*

Definition 1.12. *Polytope*

A polytope is a bounded polyhedron. Note that this implies that not all polyhedron are bounded.

2. Topological Properties of Convex Sets

Definition 2.1. *Affine Hull*
For any set $A \subseteq \mathbb{R}^n$, the affine hull of A, denoted aff(A) is the smallest linear manifold (affine set) that contains A. That is,

$$aff(A) = \bigcap \{L \supseteq A : L\, is\ a\ linear\ manifold\} = \bigcap \{L \supseteq A : L\, is\ an\ affine\ set\}$$

Proposition 2.1. *aff*$(A) = \{\sum^k \alpha_i x_i : \{x_1, ..., x_k\} \subseteq A, \sum^k \alpha_i = 1, k \in \mathbb{N}\}$

Definition 2.2. *Interior*
The interior of a set C is the set of all points of C around which you can draw an epsilon ball and remain in C. That is,

$$int(C) = \bigcup \{A : A \subseteq S, A\, is\ open\}$$

Definition 2.3. *Relative Interior*
Let $C \subseteq \mathbb{R}^n$ be convex. The relative interior of C is defined as

$$ri(C) = \{x \in C : \exists \varepsilon > 0\, s.t.\, N_\varepsilon \bigcap aff(C) \subseteq C\}$$

In other words, the relative interior is the interior of the hyperplane containing C.

Definition 2.4. *Closure*
A closure of a set A is the set A and all of its limit points. That is,

$$cl(A) = \{x : \forall \varepsilon > 0, N_\varepsilon \bigcap A \neq \emptyset\}$$

Equivalently, $cl(A)$ is the union of A and its boundary, as well as the intersection of all closed sets containing A.

Lemma 2.0.1. *Let $C \subseteq \mathbb{R}^n$ be convex. Let $x_1 \in ri(C)$ and $x_2 \in cl(C)$. Then $[x_1, x_2) \subseteq ri(C)$.*

- *In words, if I take the ball around a point $z \in [x_1, x_2)$, then the intersection of that ball and the affine hull of A will be contained in C.*
- *Think of $[x_1, x_2)$ as all convex combinations of x_1 and x_2, not including x_2.*
- *Two useful claims when proving this:*
  (1) *For $z \in [x_1, x_2)$, i.e. $z = \alpha x_1 + (1-\alpha)x_2$ for $\alpha \in (0, 1]$, $N_{\alpha \cdot \varepsilon} \bigcap aff(C) \subseteq C$.*
  (2) $\alpha y + (1-\alpha)x \in aff(C) \iff y \in aff(C)$

Corollary 2.0.1. *If $C \subseteq \mathbb{R}^n$ is convex, then $ri(C)$ is also convex.*

Corollary 2.0.2. *Assume that C is convex. Then,*

(1) $cl(C) = cl(ri(C))$
(2) $ri(C) = ri(cl(C))$

Note that this is not typically true for an arbitrary set C.

Lemma 2.0.2. *If C is convex, then $cl(C)$ is also convex.*

Lemma 2.0.3. *If $O \subseteq \mathbb{R}^n$ is open, then $co(O)$ is also open.*

Lemma 2.0.4. *If $K \subseteq \mathbb{R}^n$ is compact, then $co(K)$ is compact.*

Remark 2.0.1. *If K is closed only (i.e., not bounded), then $co(K)$ may not be closed.*

3. Projection onto Convex Sets

Definition 3.1. *Orthogonal Projection*
Let $C \subseteq \mathbb{R}^n$ be convex and $x \in \mathbb{R}^n$. If $\exists x^* \in C$ such that $\|x - x^*\| < \|x - z\| \forall z \in C,\ z \neq x^*$, then we say that x^* is the orthogonal projection of x onto C, denoted as $x^* = \mathbb{P}_C(x)$.

- *Trivially, if $x \in C$, then $x^* = x$.*
- *Note that $\|x - x^*\| \leq \|x - z\|,\ \forall z \in C, z \neq x^*$ asserts uniqueness.*

Theorem 3.1. *Let $C \subseteq \mathbb{R}^n$ be closed and convex. Then,*

(1) *Existence: $\mathbb{P}_C(x)$ exists $\forall x \in \mathbb{R}^n$.*
(2) *Uniqueness: $x^* = \mathbb{P}_C(x)$*
(3) *Characterization: $x^* \in C$ and $\langle x - x^*, z - x^*\rangle \leq 0,\ \forall z \in C$.*

In other words, the existence of $\mathbb{P}_C(x)$ implies that $\inf_{z\in C} \|x - z\| = x^*,\ x \in \mathbb{R}^n$ is attained and x^* is unique.

4. Projection onto a Linear Subspace

Remark 4.0.1. *A linear subspace is convex and closed.*

Corollary 4.0.1. *Let $Y \subseteq \mathbb{R}^n$ be a linear subspace. Then, for any $x \in \mathbb{R}^n$,*

$$x^* \in Y \iff \langle x - x^*, y\rangle = 0, \forall y \in Y$$

Definition 4.1. *Orthogonal Set*
Let $W \subseteq \mathbb{R}^n$. We define the orthogonal set to W as

$$W^\perp = \{x \in \mathbb{R}^n : \langle x, w\rangle = 0, \forall w \in W\}$$

Proposition 4.1. *Let Y be a linear subspace. Then,*

(1) *$Y^\perp$ is a linear subspace*
(2) *If $Y = span\{x_1, ..., x_n\}$, i.e. all linear combinations of vectors $x_1, ..., x_n$, then $Y^\perp = \{x \in \mathbb{R}^n : \langle x, x_i\rangle = 0, i = 1, ..., n\}$*
(3) *If Y is a linear subspace, then $Y \bigcap Y^\perp = \{\mathbf{0}\}$*
(4) *If Y is a linear subspace, then $[Y^\perp]^\perp = Y$*
(5) *If Y is a linear subspace, then for any $x \in \mathbb{R}^n$, there exists a unique decomposition $x = y + z$, where $y = \mathbb{P}_Y(x) \in Y$ and $z = \mathbb{P}_Y^\perp(x) \in Y^\perp$*

Corollary 4.0.2. *If $Y = span\{x_1, ..., x_n\}$, then*

$$x^* = \mathbb{P}_Y(x) \iff \exists \alpha_j \in \mathbb{R},\ j = 1, ..., n \text{ s.t. } x^* = \sum^l \alpha_j x_j \text{ and } \sum \alpha_j \langle x_j, x_i\rangle = \langle x, x_i\rangle\ \forall i = 1, ..., n$$

Note that $x^* \in Y \iff x^*$ can be written as a linear combination of the basis vectors $x_1, ..., x_n$. Also observe that the projection x^* is perpendicular to each one of the basis vectors.

Lemma 4.0.1. *Every linear subspace of $\mathbb{R}^n$ has an orthonormal basis, i.e. basis vectors are linearly independent with norm 1.*

Proposition 4.2. *For $x_1, ..., x_k \in \mathbb{R}^n$ and $x \in \mathbb{R}^n$, denoting $Y = span\{x_1, ..., x_k\}$*

$$x_0 - \mathbb{P}_Y(x) = \mathbf{0} \iff \langle x - x_0, x_i\rangle = 0, \forall i = 1, ..., k$$

Remark 4.0.2. *In summary,*

(1) *For $S \subseteq \mathbb{R}^n$, S convex, and $x \in \mathbb{R}^n$, $x_0 = \mathbb{P}_S(x) \iff \langle x - x_0, y - x_0 \rangle \leq 0 \; \forall y \in S$*

(2) *For $Y \subseteq \mathbb{R}^n$, Y linear subspace, and $x \in \mathbb{R}^n$, $x_0 = \mathbb{P}_Y(x) \iff \langle x - x_0, y \rangle = 0, \; \forall y \in Y$*

(3) *For $x_1, ..., x_k \in \mathbb{R}^n$, let $S = span(x_1, ..., x_k)$. Then, for $x \in \mathbb{R}^n$, $x_0 - \mathbb{P}_S(x) \iff \langle x - x_0, x_i \rangle = 0, \forall i = 1, ..., k$*

Definition 4.2. *Gram Matrix*
For any $x_1, ..., x_k \in \mathbb{R}^n$, we define the matrix of inner products to be the Gram Matrix as follows

$$G_{x_1,...,x_k} = \begin{bmatrix} \langle x_1, x_1 \rangle & \cdots & \langle x_1, x_k \rangle \\ \vdots & \ddots & \vdots \\ \langle x_k, x_1 \rangle & \cdots & \langle x_k, x_k \rangle \end{bmatrix}$$

Note that G is symmetric and has dimensions $(k \times k)$.

Definition 4.3. *Linear Independence*
A collection of vectors $x_1, ., , x_k$ are linearly independent if

$$\sum \lambda_i x_i = \mathbf{0} \iff \lambda_1 = \cdots = \lambda_k = 0$$

Lemma 4.0.2. *$G_{x_1,...,x_k}$ is invertible $\iff \{x_1, ..., x_k\}$ are linearly independent.*

Theorem 4.1. *Let $x_1, ..., x_k \in \mathbb{R}^n$ be linearly independent vectors and take any $x \in \mathbb{R}^n$. Define $Y =$span$(x_1, ..., x_k)$. Then,*

$$x_0 = \mathbb{P}_Y(x) \iff x_0 = \lambda_1 x_1 + \cdots \lambda_k x_k,$$

where $\lambda = [\lambda_1, ..., \lambda_k]^T$ constitute the unique solution to the system

$$G_{x_1,...,x_k} \lambda = [\langle x_i, x \rangle]$$

and $[\langle x_i, x \rangle] = [\langle x_1, x \rangle, ..., \langle x_k, x \rangle]^T$.

Definition 4.4. *Range of a Matrix*
Let $A \in \mathbb{R}^{k \times n}$. Then the range of A, denoted $rng(A)$, is

$$rng(A) = \{Ax : x \in \mathbb{R}^n\} = span(x_1, ..., x_n)$$

where $x_1, ..., x_n$ are the column vectors of A. Note that $rng(A)$ is a subspace of $\mathbb{R}^k$, since each column vector contains k elements. We can also think of $rng(A)$ as the range of a function defined by A, since A is a linear operator.

Definition 4.5. *(Column) Rank of a Matrix*
Let $A \in \mathbb{R}^{k \times n}$. Then, $rank(A)$ is the number of independent columns of A. If $rank(A) = \min(k, n)$, then we say that A is full rank.

Application 4.1. *Let $A \in \mathbb{R}^{k \times n}$ and $b \in \mathbb{R}^k$. Suppose we want to solve $Ax = b$, but A is $(k \times n)$ and b is $(k \times 1)$ (i.e. $Ax = b$ unsolvable). Then, we want to find the next best approximation of b in $rng(A)$.*

- *Problem: Minimize $\|Ax - b\|$ with respect to x*
- *Solution: Least Squares*

  (1) *x_0 solves this problem $\iff Ax_0 = \mathbb{P}_{rng(A)}(b)$*

  (2) *$\iff \langle b - Ax_0, y \rangle$, $y \in rng(A)$*

(3) $= \langle b - Ax_0, Ax\rangle$

(4) $= 0,\ \forall x \in \mathbb{R}^n$

Proposition 4.3. *For any* $A \in \mathbb{R}^{k\times n}$*,* $y \perp rng(A) \iff (Ay)^T = \mathbf{0}$

- *In essence,* $rng(A)^{\perp} = null(A^T)$

Corollary 4.1.1. *Following from the proposition above,* $\langle b - Ax_0, Ax\rangle = 0,\ \forall x \in \mathbb{R}^n \iff A^T(b - Ax_0) = 0 \iff A^T Ax_0 = A^T b$*.*

Lemma 4.1.1. $A^T A$ *is invertible* $\iff$ $rank(A) = n$*, so* A *is full rank*

- A *being full rank* $\implies$ *its columns are linearly independent*

Theorem 4.2. *If* $A \in \mathbb{R}^{k\times n}$ *has* $rank(A) = n$ *("identification condition"), then the solution to the problem* $Ax = b$ *is*

$$x_0 = (A^T A)^{-1} A^T b$$

5. Separation by Hyperplane and Convex Cones

Theorem 5.1. *Hyperplane Separation Thm. (of Closed, Convex set from Point)*
Let S be a nonempty, closed, and convex set in $\mathbb{R}^n$. Let $x \in \mathbb{R}^n \setminus S$. Then, $\exists \alpha \in \mathbb{R}^n \setminus \{\mathbf{0}\}$ such that $\langle \alpha, x \rangle > \sup_{y \in S} \langle \alpha, y \rangle$.

- *In words, a nonempty, closed, and convex set can be strictly separated by a hyperplane from any given point that lies outside that set.*
- *Proof idea is to project x onto our set S to get a supporting hyperplane of S and then to shift the hyperplane up so that it strictly separates x from S.*
- *And since S is closed and convex, there exists a unique nearest point in S to a point x outside of S, namely the projection of x onto S, $\mathbb{P}_S(x)$.*
- *$\langle \alpha, x \rangle > \langle \alpha, y \rangle$ just means that x and y are on different sides of the hyperplane.*
- *$\alpha \in \mathbb{R} \setminus \{0\}$ is the normal vector in the definition of the hyperplane.*

Application 5.1. *How to Find a Separating Hyperplane*

(1) *For existence, use hyperplane separation theorem*

(2)

Definition 5.1. *Cone*
A nonempty set $C \subseteq \mathbb{R}^n$ is a cone if $\lambda x \in C$ for all $x \in C$ and $\lambda \geq 0$.

Remark 5.1.1. *Cones can never be bounded.*

- *We can conceptualize them as sets made up of rays that go between the origin and a point contained in the set C.*
- *A cone always contains the origin.*
- *A cone need not be convex.*

Definition 5.2. *Convex Cone*
If C is a cone and a convex set, then we call C a convex cone.

Lemma 5.1.1. *A set $C \subseteq \mathbb{R}^n, C \neq$ is a convex cone $\iff \forall k \in \mathbb{N}$, $\forall x_1, ..., x_k \in C$, and $\forall \alpha_1, ..., \alpha_k \geq 0$, $\sum^k \alpha_i x_i \in C$, i.e. linear combination of points from C is contained in C*

Proposition 5.1. *Separation of a Closed, Convex Cone from a Point via Hyperplane*
Let C be a closed, convex cone in $\mathbb{R}^n$ and let $x \in \mathbb{R}^n \setminus C$. Then, $\exists \alpha \in \mathbb{R}^n \setminus \{\mathbf{0}\}$ such that

$$\langle \alpha, x \rangle > 0 = \max_{y \in C} \langle \alpha, y \rangle$$

- *In words, a nonempty, closed, convex cone cane be separated from any given point that lies outside that cone by a hyperplane that goes through the origin.*
- *This is just a particular instance of the hyperplane separation theorem.*

Definition 5.3. *Conical Hull*
For any $S \subseteq \mathbb{R}^n$, we define the following set as the conical hull, the smallest convex cone containing S

$$cone(S) = \bigcap \{C \subseteq \mathbb{R}^n : C \text{ is a convex cone}, S \subseteq C\}$$

- *We can think of the conical hull as refining the convex cones containing S until we reach the smallest one – building from the outside inward.*

Lemma 5.1.2.

$$cone(S) = \{\sum^{k} \alpha_i x_i : k \in \mathbb{N}; x_1, ..., x_k \in S; \alpha_1, ..., \alpha_k \geq 0\}$$

Remark 5.1.2. *A convex set S may have a conical hull $cone(S)$ that is not closed. For example, tangencies to axes can yield $cone(S)$ not closed.*

Definition 5.4. *Cone Generated by Set*
We say that a convex cone $C \subseteq \mathbb{R}^n$ is generated by a set $S \subseteq \mathbb{R}^n$ if $C = cone(S)$.

- *In essence, convex cone C is given to you and we want to find the S that generates it.*
- *In other words, the cone generated by a set is the conical hull of the set.*

Definition 5.5. *Finitely-Generated Cone*
We say that cone $C = cone(S)$ is finitely-generated if $\mid S \mid < \infty$.

Definition 5.6. *Basic Cone*
We say that $C = cone(S)$ is basic if S consists of linearly independent vectors (and thus finitely-generated, since a set of linearly independent vectors in $\mathbb{R}^n$ can consist of at most n vectors).

Lemma 5.1.3. *$cone(\{x\})$ is closed $\forall x \in \mathbb{R}^n$.*

- *Proof Sketch:*

  (1) *We proceed by constructing a sequence (y_m) in $cone(\{x\})$, and we WTS that if $y_m \to y$, then it converges within the set, i.e. $y \in \mathbb{R}^n$.*

  (2) *We are constructing a sequence in $cone(\{x\})$ whose elements are all scalar multiples of some fixed vector x, since the cone is generated by scaling up/down some vector x. This implies that we have a sequence of scalars that accompany each y_m.*

  (3) *More formally, for every $m \in \mathbb{N}$, there exists $\lambda_m \geq 0$ such that $y_m = \lambda_m x$. Since $\|\cdot\|$ is continuous*[1]*, $\|y_m\| \to \|y\|$*
  $\implies \|\lambda_m x\| \to \|y\|$.
  $\implies \lambda_m \|x\| \to \|y\|$
  $\implies \lim_{m\to\infty} \lambda_m \|x\| \to \|y\|$
  $\implies \lambda = \frac{\|y\|}{\|x\|}$

  (4) *Then, by finding that $\lambda_m \to \lambda = \frac{\|y\|}{\|x\|}$, we have $y_m = \lambda_m x \to \lambda x$, so long as $x \neq \mathbf{0}$.*
  $\implies y = \lambda x \in cone(\{x\})$

  (5) *Note that if we did in fact have $x = \mathbf{0}$, then $cone(\{x\}) = \{\mathbf{0}\}$ is a singleton $\implies$ any multiple of x is also $\mathbf{0}$ and we thus have a constant sequence which trivially converges to the constant, $\mathbf{0}$ in this case.*

Lemma 5.1.4. *Every basic cone in $\mathbb{R}^n$ is closed.*

Lemma 5.1.5. *Let $x, x_1, ..., x_k \in \mathbb{R}^n$ such that $x = \sum^k \theta_i x_i$, where $\theta_i > 0$, $\forall i = 1, ..., k$. Then, x can be expressed as a nonnegative linear combination of $k-1$ many vectors from $\{x_1, ..., x_k\} \iff \exists \lambda \in \mathbb{R}^k \setminus \{\mathbf{0}\}$ such that $\sum^k \lambda_i x_i = 0$.*

Lemma 5.1.6. *For any set $S \subseteq \mathbb{R}^n$,*

$$cone(S) = \bigcap \{cone(T) : T \subseteq S,\ T \text{ linearly independent}\}$$

Theorem 5.2. *Every finitely generated convex cone in $\mathbb{R}^n$ is closed.*

Corollary 5.2.1. *Every linear subspace of $\mathbb{R}^n$ is closed.*

---

[1] Recall that continuous functions preserve limits.

6. The Farkas Lemma

Theorem 6.1. *Farkas Lemma*
Let $A \in \mathbb{R}^{k \times n}$ and $b \in \mathbb{R}^k$. Then, either

$$\exists x \in \mathbb{R}^n, x \geq \mathbf{0} \text{ such that } Ax = b$$

or

$$\exists w \in \mathbb{R}^k \text{ such that } \langle w, b \rangle > 0 \text{ and } A^T w \leq \mathbf{0}$$

– *This implies that if b is in the convex cone generated by the column vectors of A, cone($\{a_1, ..., a_n\}$), then there exists a solution x to $Ax = b$. And if b sits outside the convex cone generated by the column vectors of A, then we can find a hyperplane separating b from cone($\{a_1, ..., a_n\}$).*

– *Proving Farkas Lemma requires two parts:*

(1) *Exclusivity: Suppose both parts are true and derive a contradiction*

(2) *If not p, then q: Suppose one part fails. Show that the other must hold.*

(3) *If p, then not q: Suppose one part holds. Show the other must fail.*

Corollary 6.1.1. *We can also state Farkas Lemma as follows:*
For any $A \in \mathbb{R}^{k \times n}$ and $b \in \mathbb{R}^k$

$$\exists x \geq \mathbf{0} \text{ such that } Ax = b \iff \forall w \in \mathbb{R}^k, \text{ either } \langle w, b \rangle \leq 0, \text{ or } A^T w > \mathbf{0} \text{ or both } \langle w, b \rangle \leq 0 \;\&\; A^T w > 0$$

Theorem 6.2. *Fredholm Alternative Thm.*
Let $A \in \mathbb{R}^{k \times n}$ and $b \in \mathbb{R}^k$. Then, either

$$\exists x \in \mathbb{R}^n \text{ such that } Ax = b$$

or

$$\exists w \in \mathbb{R}^k \text{ such that } \langle w, b \rangle \neq 0 \text{ and } A^T w = \mathbf{0}$$

Corollary 6.2.1. *Note that for any $(k \times n)$ matrix A,*

$$\exists x \in \mathbb{R}^n \text{ s.t. } Ax = b \implies b \in range(A)$$

On the other hand,

$$\exists w \in \mathbb{R}^k \text{ such that } \langle w, b \rangle \neq 0 \text{ and } A^T w = \mathbf{0} \implies b \notin null(A^T)^{\perp}$$

Therefore, the Fredholm Alternative Theorem tells us that either

$$b \in rng(A)$$

or

$$b \notin null(A^T)^{\perp}$$

And this holds if and only if $rng(A) = null(A^T)^{\perp}$. The latter equality holds as a consequence of the projection theorem.

Remark 6.2.1. *Every vector x can be expressed as the difference between two nonnegative vectors $u, v \geq 0$. Thus, if $Ax = b$, then there exist $u, v \in \mathbb{R}^n$, where $u, v \geq 0$, such that $x = u - v$.*

– *We can manipulate Farkas Lemma by using this fact and introducing slack/surplus variables.*

Corollary 6.2.2. *Let $A \in \mathbb{R}^{k \times n}$ and $b \in \mathbb{R}^k$. Then, either*

$$\exists x \in \mathbb{R}^n \text{ such that } Ax \geq b$$

or

$$\exists w \in \mathbb{R}^k, w \geq \mathbf{0}, \text{ such that } \langle w, b \rangle > 0 \text{ and } A^T w = \mathbf{0}$$

– *Proof Sketch:*

(1) *We first need to figure out a way to construct an analogous system of equations that (i) holds with equality (i.e. get rid of $\geq$) and (ii) guarantees a nonnegative solution vector so that we can apply Farkas Lemma.*

(2) *Let $y_1, y_2, y_3 \in \mathbb{R}^n$. Let $x = y_1 - y_2$. Then, $Ax = b \implies A(y_1 - y_2) \geq b$. This holds with equality if we choose some nonnegative vector y_3 and subtract it from the LHS. In particular, we choose slack variable such that $y_3 = Ax - b$.*
$\implies A(y_1 - y_2) - y_3 = b$

(3) *By constructing an augmented matrix, we can reinterpret this as a system of equations*
$\implies [A - A - I]z = b$, *where z is a $3n$ vector, $z \geq \mathbf{0}$, and $[A - A - I]$ is an augmented matrix.*

(4) *Note that each of our k equations takes the form: $a_i y_1 - a_i y_2 - y_3 = b$ for a_i row vectors of A for $i = 1, ..., k$.*

(5) *Now we have our new system of equations that agrees with the structure of Farkas Lemma.*

(6) *Furthermore, suppose $Ax \ngeq b$, i.e. part (i) is false. We WTS that part (ii) must hold.*

(7) $Ax \neq b \implies [A - A - I]z \neq b$. *Therefore, we may apply Farkas Lemma to conclude that for some $w \in \mathbb{R}^k$,* $\langle b, w \rangle > 0 \ \& \ [A - A - I]^T w \leq 0$
$\implies \langle b, w \rangle > 0 \ \& \ A^T w = 0 \ \& \ w \geq \mathbf{0}$ *for some $w \in \mathbb{R}^k$*

   – *Side-note: we know that $w \geq \mathbf{0}$ since if $\langle b, w \rangle > 0$, then neither b nor w can be $\mathbf{0}$.*

Lemma 6.2.1. *Stiemke Lemma*
Let $A \in \mathbb{R}^{k \times n}$ and $b \in \mathbb{R}^k$. Then, either

$$Ax = \mathbf{0}, \text{ for some } x \in \mathbb{R}^n, x >> \mathbf{0}$$

or

$$A^T w > \mathbf{0}, \text{ for some } w \in \mathbb{R}^k$$

– *Proof Sketch:*

(1) *To prove this lemma, we need to introduce slack variables and construct an augmented matrix as in the proof of the corollary above.*

(2) *Note that $Ax = \mathbf{0}$ for some $x >> \mathbf{0}$ is the same as saying that $Ax = \mathbf{0}$ for some $x \geq \mathbf{1}$.*

   – *If all components of x are strictly positive, we could always scale the vector by some constant λ so that all values of the vector are strictly greater than 1. In other words, we don't have to worry about the case in which $x_i = 0 \implies \lambda x_i 0$ for $\lambda > 0$.*

(3) $x \geq \mathbf{1} \implies -x \leq (-1, -1, ..., -1) \implies -x + y = -1$ *for some $y \in \mathbb{R}^n$*

(4) *Therefore,* $Ax = \mathbf{0} \iff Ax = \mathbf{0} \ \& \ -x + y = -1$ *for some $x, y \in \mathbb{R}^n$, $x, y \geq 0$*
$\implies a_i x = 0$ *for row vectors $a_1, ..., a_k$* $\& \ -x + y = -1$

(5) *Thus, our $Ax = b$ system becomes:*

$$\begin{bmatrix} a_{11} & \cdots & a_{1n} & 0 & \cdots & 0 \\ \vdots & & \vdots & \vdots & & \vdots \\ a_{k1} & \cdots & a_{kn} & 0 & \cdots & 0 \\ -1 & \cdots & 0 & 1 & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & -1 & 0 & \cdots & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_n \\ y_1 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ -1 \\ \vdots \\ -1 \end{bmatrix}$$

And now our new system has the structure that Farkas Lemma requires, where A refers to the augmented version above and $b = [0, ..., 0, -1, ..., -1]^T \in \mathbb{R}^{k+n}$.

(6) *Now, suppose that part (i) fails, i.e. $Ax \neq \mathbf{0}$. Then, $\nexists x, y \in \mathbb{R}^n$, $x, y \geq \mathbf{0}$ s.t. the augmented system above holds. Farkas Lemma tells us that there instead exists some vector $w \in \mathbb{R}^{k+n}$ such that $\langle b, w\rangle > 0$ & $A^T w \leq \mathbf{0}$.*

(7) *Since $b \in \mathbb{R}^{k+n}$, we know that w is the concatenation of some $(k \times 1)$ vector u and some $(n \times 1)$ vector v. Thus, $b = [u \;\; v]^T$, so we can rewrite the Farkas Lemma condition (ii): $\langle b, w\rangle = \langle [u \;\; v]^T, [\mathbf{0} \; -\mathbf{1}]^T\rangle > 0$.*
$\implies b_1 w_1 + ... + b_k w_k + b_{k+1} w_{k+1} + ... + b_{k+n} w_{k+n} = -v_1 - ... - v_n > 0$
$\implies -v_1 - ... - v_n > 0$
$\implies v_1 + ... + v_n < 0$

(8) *$A^T w \leq \mathbf{0}$ in the context of the Farkas Lemma $\implies A^T u - v \leq 0$ & $v \leq \mathbf{0}$, since our $A^T w$ in the context of this problem is:*

$$\begin{bmatrix} A^T & -\mathbf{1} \\ \mathbf{0} & \mathbf{1} \end{bmatrix} \begin{bmatrix} u \\ v \end{bmatrix} \leq 0$$

$\implies A^T u - v \leq 0$ & $v \leq \mathbf{0} \implies A^T u \leq v \leq \mathbf{0}$

(9) $v_1 + ... + v_n < 0$ *and* $v \leq \mathbf{0} \implies v < \mathbf{0}$

(10) *Therefore,* $A^T u \leq v < \mathbf{0} \implies A^T u < \mathbf{0}$

Remark 6.2.2. *$x >> \mathbf{0}$ means that x has all strictly positive components, so $x_i \neq 0$, $\forall i$. On the other hand, $x \geq \mathbf{0}$ means that x has all nonnegative components.*

- *And note that the converse of $x \geq \mathbf{0}$, i.e. $x \ngeq \mathbf{0}$, is that <u>at least one</u> x_i is not nonnegative.*

7. Concavity

Definition 7.1. *Concave*
Let S be a nonempty, convex subset of $\mathbb{R}^n$. We say that a function $f : S \to \mathbb{R}$ is concave if $\forall \alpha \in (0, 1), \forall x, y \in S, x \neq y$

$$f(\alpha x + (1 - \alpha)y) \geq \alpha f(x) + (1 - \alpha)f(y)$$

Definition 7.2. *Convex*
Let S be a nonempty, convex subset of $\mathbb{R}^n$. We say that $f : S \to \mathbb{R}$ is convex if $\forall \alpha \in (0, 1), \forall x, y \in S, x \neq y$

$$f(\alpha x + (1 - \alpha)y) \leq \alpha f(x) + (1 - \alpha)f(y)$$

Put more concisely, we say that a f is convex if $-f$ is concave.

Definition 7.3. *Affine*
A function f is affine if it is convex and concave, i.e. $\forall \alpha \in (0, 1), \forall x, y \in S, x \neq y$

$$f(\alpha x + (1 - \alpha)y) = \alpha f(x) + (1 - \alpha)f(y)$$

Proposition 7.1. *Let S and T be two nonempty convex subsets of $\mathbb{R}^n$ with $T \subseteq S$. If $f : S \to \mathbb{R}$ is a (strictly) concave function, then the restriction of f to T, $f\mid_T : T \to \mathbb{R}$, is (strictly) concave on T*

Proposition 7.2. *For any nonempty convex subset $S \subseteq \mathbb{R}^n$, a map $f : S \to \mathbb{R}$ is concave $\iff \{(x, t) \in \mathbb{R}^{n+1} : t \leq f(x)\}$ is a convex set.*

- *Geometrically, concavity (convexity) of a real map f on an interval I means that the line segment between any two 2-vectors of the form $(x, f(x))$ and $(y, f(y))$ lies everywhere below (above) the graph of f. Put differently: f is concave if and only if the area below (above) the graph of f constitutes a convex subset of $\mathbb{R}^2$.*

Example 7.1. $f : S \to \mathbb{R} : x \mapsto \min_{i=1,...,n} x_i$ *is a concave function.*

– Proof:

$$\begin{aligned} f(\alpha x + (1-\alpha)y) &= \min_i \alpha x_i + (1-\alpha)y_i \\ &\geq \min_i \alpha x_i + \min_i (1-\alpha)y_i \\ &= \alpha \min_i x_i + (1-\alpha)\min_i y_i \\ &= \alpha f(x) + (1-\alpha)f(y) \end{aligned}$$

Example 7.2. *$f : S \to \mathbb{R} : x \mapsto \max_{i=1,...,n} x_i$ is a convex function.*

Proposition 7.3. *Inverses of Convex Functions*
Let I be an interval and $f : I \to \mathbb{R}$ a strictly increasing and (strictly) convex function. Then, f^{-1} is a (strictly) concave function on $f(I)$.

– Proof Sketch:

(1) *Take any $x, y \in f(I)$ and $\alpha \in (0,1)$. Then, $\exists z, w \in I$ such that $f(z) = x$ and $f(w) = y$. Then, by convexity of f*

$$f(\alpha z + (1-\alpha)w) \leq \alpha f(z) + (1-\alpha)w = \alpha x + (1-\alpha)y$$

Since f and hence f^{-1} are strictly increasing, we can apply f^{-1} to both sides and preserve the inequality to get

$$\alpha z + (1-\alpha)w \leq f^{-1}(\alpha x + (1-\alpha)y)$$

And because $\alpha z + (1-\alpha)w = \alpha f(x) + (1-\alpha)f(y)$

$$\alpha f^{-1}(x) + (1-\alpha)f^{-1}(y) \leq f^{-1}(\alpha x + (1-\alpha)y)$$

thereby proving that f^{-1} is concave.

Example 7.3. *Let T be a closed and convex set in $\mathbb{R}^n$. Then, $dist(x,T) = \inf_{t\in T} \|x - t\|$ is a convex function.*

– Proof: *Let $x, y \in \mathbb{R}^n$ & $0 < \alpha < 1$. Set $x' = \mathbb{P}_T(x)$ and $y' = \mathbb{P}_T(y)$. Then,*

$$\begin{aligned} dist(\alpha x + (1-\alpha)y, T) &= \inf_{z\in T} \|\alpha x + (1-\alpha)y - z\| \\ &\leq \|\alpha x + (1-\alpha)y - \alpha x' + (1-\alpha)y'\| \\ &\leq \alpha\|x - x'\| + (1-\alpha)\|y - y'\| \\ &= \alpha\, dist(x,T) + (1-\alpha)\, dist(y,T) \end{aligned}$$

– *Note that $dist(x,T) = \inf_{t\in T} \|x - t\| = \|x - \mathbb{P}_T(x)\|$, since by definition, the projection of x onto a set T is the distance between x and the point in T closest to x.*

Example 7.4. *Convexity of Norms*
Let $\|\cdot\|$ be a norm on $\mathbb{R}^n$. Then, $\|\cdot\|$ is a convex function, since we know that $\|\alpha x + (1-\alpha)y\| \leq \alpha\|x\| + (1-\alpha)\|y\|$.

– *In fact, whatever norm you use on $\mathbb{R}^n$ is a convex function (all strictly convex aside from $\|\cdot\|_\infty$).*

Example 7.5. *Linearity of Inner Product*
The inner product with one of its arguments fixed $\langle\cdot, c\rangle$ is a linear function, and thus both convex and concave, since the inner product is linear in each of its arguments.

Example 7.6. *Let S be a nonempty convex subset of $\mathbb{R}^n$ and let k be a positive integer. If $f_1, ..., f_k$ are (strictly) concave functions on S, then for any $\lambda_1, ..., \lambda_k > 0$,*

$$\lambda_1 f_1 + ... + \lambda_k f_k$$

is a (strictly) concave function on S.

Example 7.7. *Let S be a nonempty convex subset of $\mathbb{R}^n$ and define $h : S \to \mathbb{R}$ by*

$$h(x) = \|x\|^2$$

Then, h is a convex function.

- *Note that $h = g \circ f$, where $f : S \to [0, \infty) : x \mapsto \|x\|$ and $g : [0, \infty) \to \mathbb{R}] : x \mapsto x^2$. Since f is convex, and g is increasing and convex, we can conclude that h is convex.*

Example 7.8. *Concavity of Quadratic Forms*
Let Q be a symmetric $(n \times n)$ matrix, and let $f : \mathbb{R}^n \to \mathbb{R} : x \mapsto \langle Qx, x\rangle$. Let x, y be two arbitrarily chosen n-vectors, and let $\alpha \in (0, 1)$. Then,

$$\begin{aligned} & f(\alpha x + (1-\alpha)y) \geq \alpha f(x) + (1-\alpha)f(y) \\ \iff & \alpha(\alpha-1) < \langle Q(x-y), x-y\rangle \geq 0 \\ \implies & \langle Qz, z\rangle \leq \forall z \in \mathbb{R}^n \end{aligned}$$

(1) *f is concave $\iff$ Q is negative semidefinite, i.e. $\langle Qz, z\rangle \leq 0$, $\forall z \in \mathbb{R}^n$ (all eigenvalues are nonpositive; visually, diagonal has all nonpositive values)*

(2) *f is strictly concave $\iff$ Q is negative definite, i.e. $\langle Qz, z\rangle < 0$, $\forall z \in \mathbb{R}^n \setminus \{0\}$*

Proposition 7.4. *Let $f : S \to \mathbb{R}$ and I is an interval such that $f(S) \subseteq I$, and $g : I \to \mathbb{R}$. In addition, let f be concave and let g be concave and strictly increasing. Then, $g \circ f$ is concave.*

- *Strictly increasing and concave transformation of a concave function yields a concave function*

Proposition 7.5. *Let $\mathfrak{F}$ be the set of affine real functions on S. Assume that $\inf_{f\in\mathfrak{F}} f(x) > \infty$, $\forall x \in S$. Then, $x \mapsto \inf_{f\in\mathfrak{F}} f(x)$ is a concave function.*

- *Pointwise infimum of functions is concave.*
- *The converse is also true: If $\inf_{f\in\mathfrak{F}} f(x) < \infty$ $\forall x \in S$, then we can get a set of affine real functions that approximates our function.*
- *At the kinks of $\inf_{f\in\mathfrak{F}} f(x)$, use hyperplane separation theorem.*

Theorem 7.1. *Jensen's Inequality*
$f : S \to \mathbb{R}$ is concave $\iff$ $\forall k \geq 2$, $\forall x_1, ..., x_k \in S$, and $\forall \alpha \in \Delta^{k-1}$, $f(\sum^k \alpha_i x_i) \geq \sum^k \alpha_i f(x_i)$

Definition 7.4. *Quasiconcavity*
Let S be a nonempty convex subset of $\mathbb{R}^n$. We say that a function $f : S \to \mathbb{R}$ is quasiconcave if

$$f(\alpha x + (1-\alpha)y) \geq \min\{f(x), f(y)\}$$

for any distinct $x, y \in S$ and any $\alpha \in (0, 1)$

Remark 7.1.1. *Strictly increasing and concave transformations preserve concavity of a function. Strictly increasing transforms preserve quasiconcavity. And concave implies quasiconcavity, but the converse not generally true.*

Remark 7.1.2. *For a concave function, if you look at the area <u>below</u> the graph, the area (set) will be convex. Similarly, for a convex function, look at the area <u>above</u> the graph, and the area (set) will be convex.*

8. Continuity of Concave Functions

Remark 8.0.1. *A concave function on a convex set need not be continuous. However, discontinuities can occur only at the boundary points of the domain of the function.*

Lemma 8.0.1. *Locally Bounded*
Let O be a nonempty, open and convex subset of $\mathbb{R}^n$ and let $f : O \to \mathbb{R}$ be a concave function. Then, for every $x \in O$, $\exists \varepsilon, K > 0$ such that for every $y \in B(x, \varepsilon)$,

$$\mid f(y) \mid \leq K$$

Theorem 8.1. *Locally Lipschitz*
Let $f : O \to \mathbb{R}$ where $O \subseteq \mathbb{R}^n$ is open and convex. Let f be a concave function. Then, $\forall x \in O$, $\exists \varepsilon > 0$ & $K_x > 0$ such that

$$\mid f(a) - f(b) \mid \leq K_x \|a - b\|, \ \forall a, b \in B(x, \varepsilon)$$

- *In words, concave functions are locally Lipschitz in $\mathbb{R}^n$ (or more generally, on an open and convex set), and since the locally Lipschitz condition subsumes continuity, concave functions are implicitly continuous.*

Corollary 8.1.1. *Let S be a nonempty convex subset of $\mathbb{R}^n$ and $f : S \to \mathbb{R}$ a concave function. Then, f is continuous on int(S).*

- *Follows from the implication that a concave f is locally Lipschitz on int(S), since the interior of a convex set in $\mathbb{R}^n$ is also convex and the interior of a set is open by definition.*

Corollary 8.1.2. *If $f : O \to \mathbb{R}$ for $O \subseteq \mathbb{R}^n$ and f concave $\implies f$ is continuous.*

Proposition 8.1. *Let I be an open interval (and thus convex, since intervals are convex by def'n). Let $f : I \to \mathbb{R}$ be a concave function. Then,*

1. *f is left-differentiable (i.e. left derivative exists everywhere)*
2. *f is right-differentiable (i.e. right derivative exists everywhere)*
3. *For any $x, y \in I, x > y$,*
$$f'_+(x) \leq f'_-(x) \leq \frac{f(y) - f(x)}{y - x} \leq f'_+(y) \leq f'_-(y)$$
where $f'_-(x)$ denotes the left derivative at x, etc.
4. *f'_- and f'_+ are decreasing functions; and if f is twice differentiable, $f'' < 0$*

Corollary 8.1.3. *For an open interval $I \subseteq \mathbb{R}$, let $f : I \to \mathbb{R}$, differentiable (or twice-differentiable). Then, f is concave $\iff f'$ is decreasing (or $f'' \leq 0$ given that f'' exists).*

Proposition 8.2. *Let f, g be concave functions and g increasing. Then, $g \circ f$ is concave.*

- *Proof:*
  1. $g \circ f(\lambda x + (1 - \lambda)y) \geq \lambda g \circ f(x) + (1 - \lambda) g \circ f(y)$
  2. *Since f is concave,*
  $f(\lambda x + (1 - \lambda)y) \geq \lambda f(x) + (1 - \lambda) f(y)$
  3. *Since g is increasing, we can apply $g(\cdot)$ to both sides and preserve the inequality*
  $g(f(\lambda x + (1 - \lambda)y)) \geq g(\lambda f(x) + (1 - \lambda) f(y))$
  4. *By the concavity of g,*
  $g(\lambda f(x) + (1 - \lambda) f(y)) \geq \lambda g \circ f(x) + (1 - \lambda) g \circ f(y)$
  $\implies g \circ f(\lambda x + (1 - \lambda)y) \geq \lambda g \circ f(x) + (1 - \lambda) g \circ f(y)$

Proposition 8.3. *Let $F : C \to \mathbb{R} \setminus \{-\infty\}$. If $F(x) = \inf\{f(x) : f \in A \subseteq$ affine functions$\}$, then F is concave. Furthermore, if F is concave and continuous, then $\exists A \subseteq$ affine functions such that $F(x) = \inf\{f : f \in A, A \subseteq$ affine functions$\}$.*

– *Proof uses supporting hyperplanes to construct our set of affine functions (since hyperplanes are affine) which approximate our concave function.*

9. Optimization

Definition 9.1. *Differentiable*
Let $O \subseteq \mathbb{R}^n$ be an open and convex set. $f : O \to \mathbb{R}$ is differentiable at $x \in O$ if for $f(y) = f(x) + \langle \nabla f(x), y - x\rangle + E(y - x)$,

$$\frac{E(y-x)}{\|y-x\|} \to 0 \text{ as } \|y-x\| \to 0,$$

where $E(y - x) = f(x + z) - f(x) - \langle \nabla f(x), z\rangle$ is the error function and $\nabla f(x)$ is the vector of partials evaluated at x. Therefore, the error $E(y - x)$ has to go to 0 faster than $\|y - x\|$ does.

Remark 9.0.1. *If f is differentiable, then $\nabla f(x)$ is a vector of partials, but even if f is not differentiable, the vector of partials may still be well-defined.*

Theorem 9.1. *Let $f : O \to \mathbb{R}$ be a (continuously) differentiable function. Then, f is concave $\iff$ $f(y) \leq f(x) + \langle \nabla f(x), y - x\rangle$, $\forall x, y \in O$, $x \neq y$.*

- *If $O \subseteq \mathbb{R}$, then this is just like saying the first derivative is decreasing.*
- *We have a strict inequality if f is strictly concave.*

Definition 9.2. *Local Maximizer*
Let $f : O \to \mathbb{R}$ be any function. We say that $x^* \in O$ is a local maximizer of f if $\exists \varepsilon > 0$ such that $f(x^*) \geq f(x)$ $\forall x \in B(x, \varepsilon)$

Definition 9.3. *Global Maximizer*
Similarly, we say that x^* is a global maximizer if $f(x^*) \geq f(x)$ $\forall x \in O$.

Theorem 9.2. *Let $f : O \to \mathbb{R}$ be a function differentiable at x^*. If x^* is a local maximizer of f, then $\nabla f(x^*) = \mathbf{0}$.*

Lemma 9.2.1. *Suppose $f : O \to \mathbb{R}$ is concave. Then, every local maximizer of f is a global maximizer (if convex, then global minimizer).*

Theorem 9.3. *Let $f : O \to \mathbb{R}$ be concave. Suppose f is differentiable at $x^* \in O$. Then, x^* is a global maximizer of $f \iff \nabla f(x^*) = \mathbf{0}$.*

Theorem 9.4. *KKT for Concave Programming*
Suppose we are interested in the following maximization problem:

$$\max f(x) \text{ s.t. } h_i(x) \geq 0 \ \forall i = 1, ..., k$$

Let O be a nonempty, open, and convex set in $\mathbb{R}^n$. Let $f, h_1, ..., h_k : O \to \mathbb{R}$ be differentiable, concave real-valued functions on O. Assume $\exists x \in O$ such that $h_i(x) > 0$ $\forall i = 1, ..., k$; i.e., there is at least one point in the interior of the constraint set (closed since concave implies continuous) for which all constraints hold strictly—<u>Slater's Condition</u>. Then, for any $x^* \in O$, we have

$$x^* \in \arg\max\{f(x) : x \in O \ \& \ h_i(x) \geq 0 \ \forall i\} \iff \exists \lambda^* \in [0, \infty)^k \text{ such that}$$

$$\text{1. } \nabla f(x^*) + \sum^k \lambda_i^* \nabla h_i(x^*) = \mathbf{0} \text{ [FOC]}$$
$$\text{2. } \lambda_i^* h_i(x) = 0 \forall i = 1, ..., k$$

- *Note that Slater's condition simply requires that we have some point x in the interior of the constraint set in which all constraints hold with strict inequality, i.e. $h_i(x) > 0$ $\forall i$.*
- *The condition of concave constraint functions implies that we have a convex constraint set. More generally, we just care that we have a convex constraint set $\bigcap h_i(X)$ (i.e. concavity of constraint functions not necessary).*

- *In words, the theorem tells us that if we have a solution x^* for a maximization problem with concave constraints $h_1, ..., h_k$ and a concave objective function f in which Slater's condition holds, then the KKT conditions (1) and (2) must also hold.*
- *And conversely, if we have a maximization problem with a convex constraint set and a concave objective function in which Slater's condition holds, then if the KKT conditions hold, then we have a solution x^*.*
- *Note that a solution x^* to a general maximization need not satisfy the KKT conditions (or Slater's condition for that matter).*
  - *e.g. Consider* $\max x$ *s.t.* $x^2 \leq 0$. *Slater's condition does not hold. By looking at the objective function, we can tell that $x = 0$ is an optimum, but we can't use the FOC since $1 + \lambda^* 0 = 1 \neq 0$.*
  - *e.g. Consider* $\max -(x_1^2 + x_2^2)$ *such that* $(x_1 - 1)^3 - x_2 \geq 0$.

Lemma 9.4.1. *Directional Derivative*
If $f : O \to \mathbb{R}$ is differentiable at $x \in O$, then $\partial_z f(x) = \langle \nabla f(z), z \rangle \ \forall z \in \mathbb{R}^n$

Lemma 9.4.2. *Let constraint set $A \subseteq \mathbb{R}^n$ be a nonempty, convex set. Let $O \subseteq \mathbb{R}^n$ be an open convex set such that $A \subseteq O$. Let $f : O \to \mathbb{R}$ be a concave function. Then, $x^* \in \arg\max\{f(x) : x \in A\} \iff \partial_y f(x^*) \leq 0 \ \forall y \in \mathbb{R}^n$ such that $\exists T > 0$ with $x^* + ty \in A \ \forall t \in (0, T)$*

- *The "then" part (ii) of the lemma just says that in any direction $y \in \mathbb{R}^n$ from proposed optimum x^* (so y is a vector of directions), we're moving downward/decreasing. And in particular, we are testing the behavior of the objective function f near x^* by seeing what f does when we perturb the input x^* with a perturbation of "magnitude" t and direction y, i.e. $x^* + ty$. The condition that $x^* + ty \in A$ just ensures that our perturbation is meaningful; i.e., the perturbed value still lives in the constraint set and is thus feasible.*