Convex Optimization Note

Garrick Lin

March 26, 2019

1 Affine Set

1.1 Definition

A set $\mathcal{A}$ is called affine set when it satisfied:

If $\mathbf{x_1} \in \mathcal{A}$ and $\mathbf{x_2} \in \mathcal{A}$, then, $\forall \theta \in \mathcal{R}$, $x = \theta \mathbf{x_1} + (1-\theta)\mathbf{x_2}$ also belong to $\mathcal{A}$.

The expression $\theta \mathbf{x_1} + (1-\theta)\mathbf{x_2}$ represent a line that cross through $\mathbf{x_1}$ and $\mathbf{x_2}$. Hence, affine set can be explained intuitively as:

Affine set is a set that contain the line which cross through any two point within this set.

1.2 Properties

Assume a affine set $\mathcal{A} \subseteq \mathcal{R}^n$. If $\mathcal{A}$ contain original point($\mathbf{0}$), then, $\mathcal{A}$ is a subspace. In order to prove this, $\mathcal{A}$ must satisfied following three rules:

1. $\mathcal{A}$ contain original point.
2. If $\mathbf{v} \in \mathcal{A}$, then, $\forall \theta \in \mathcal{R}$, $\theta \mathbf{v} \in \mathcal{A}$.
3. If $\mathbf{v_1}, \mathbf{v_2} \in \mathcal{A}$, then, $\mathbf{v_1} + \mathbf{v_2} \in \mathbf{A}$.

The first rule is already satisfied. For the second rule, $\forall \theta \in \mathcal{R}$, $\theta \mathbf{v}$ satisfied:

$$\theta \mathbf{v} = \theta \mathbf{v} + (1-\theta)\mathbf{0}$$

Since $\mathcal{A}$ is affine and $\mathbf{v}, \mathbf{0} \in \mathcal{A}$, $\theta \mathbf{v} \in \mathcal{A}$ always true. For the third rule, there is:

$$\mathbf{v_1} + \mathbf{v_2} = 2(\frac{1}{2}\mathbf{v_1} + \frac{1}{2}\mathbf{v_2})$$

$\frac{1}{2}\mathbf{v_1}+\frac{1}{2}\mathbf{v_2}$ is in $\mathcal{A}$ as $\mathcal{A}$ is affine and $\frac{1}{2}\mathbf{v_1}, \frac{1}{2}\mathbf{v_2} \in \mathcal{A}$(According to the second rule, $\mathbf{v_1}, \mathbf{v_2} \in \mathcal{A}$, then, $\frac{1}{2}\mathbf{v_1}, \frac{1}{2}\mathbf{v_2} \in \mathcal{A}$). According to the second rule, $2(\frac{1}{2}\mathbf{v_1} + \frac{1}{2}\mathbf{v_2}) \in \mathcal{A}$

Therefore, any affine set $\mathcal{A}'$ can be thought as a subspace $\mathcal{V}$ with a transportation $\mathbf{v}$, which means, $\mathcal{A}' = \mathcal{V} + \mathbf{v}$. On the other hand, an affine set subtract some constant vector ($\mathcal{A}' - \mathbf{v}$), which make the new set contain original point, can form a subspace. Formally, for a affine set $\mathcal{A} = \{\mathbf{a_1}, \mathbf{a_2}, \cdots, \mathbf{a_n}\}$ and $\mathbf{a_i} \in \mathcal{A}$, the set:

$$\mathcal{A} - \mathbf{a_i} = \{\mathbf{a_1} - \mathbf{a_i}, \mathbf{a_2} - \mathbf{a_i}, \cdots, \mathbf{a_{i-1}}, \mathbf{0}, \mathbf{a_{i+1}}, \cdots, \mathbf{a_n} - \mathbf{a_i}\} \tag{1}$$

form a subspace. Furthermore, any affine set $\mathcal{A}$ has the form:

$$\mathcal{A} = \{\mathbf{x}|\, \mathcal{B}\mathbf{x} = \mathbf{c}\} \tag{2}$$

Proof: According to equation 1, $\mathcal{A} - \mathbf{a_i}$ is a subspace. Denote this subspace as $\mathcal{L} = \mathcal{A} - \mathbf{a_i}$. Assume the space that perpendicular to $\mathcal{L}$, denote as $\mathcal{L}^{\perp}$, has basis $\{\mathbf{b_1}, \mathbf{b_2}, \cdots, \mathbf{b_n}\}$. Then, $\mathcal{L}$ is the set of vectors that perpendicular to the basis of $\mathcal{L}^{\perp}$. As a result, if we write the basis of $\mathcal{L}^{\perp}$ in matrix form, which is:

$$\mathcal{B} = \begin{bmatrix} \mathbf{b_1}^T \\ \mathbf{b_2}^T \\ \vdots \\ \mathbf{b_n}^T \end{bmatrix}$$

We have:

$$\mathcal{L} = \{\mathbf{y}|\, \mathcal{B}\mathbf{y} = 0\} \tag{3}$$

Equation 3 always hold as all vector in $\mathcal{L}^{\perp}$ is the linear combination of the basis, hence, if a vector $\mathbf{v}$ satisfied $\mathbf{b_1}^T\mathbf{v} = 0, \mathbf{b_2}^T\mathbf{v} = 0, \cdots, \mathbf{b_n}^T\mathbf{v} = 0$, then, $(\theta_1\mathbf{b_1} + \theta_2\mathbf{b_2} + \cdots \theta_n\mathbf{b_n})^T\mathbf{v} = 0$, which means,$\mathbf{v}$ is perpendicular to any vector in $\mathcal{L}^{\perp}$ if $\mathbf{v}$ is perpendicular to the basis of $\mathcal{L}^{\perp}$.

Since the vector in $\mathcal{L}$ come from $\mathcal{A}$ subtract $\mathbf{a_i}$, for all $\mathbf{y} \in \mathcal{L}$, there must exists $\mathbf{x} \in \mathcal{A}$ such that:

$$\mathbf{y} = \mathbf{x} - \mathbf{a_i} \tag{4}$$

By combining equation 4 and equation 3, we have:

$$\mathcal{A} = \{\mathbf{x}|\, \mathcal{B}(\mathbf{x} - \mathbf{a_i}) = 0\}$$

Which is equivalent to :

$$\mathcal{A} = \{\mathbf{x}|\, \mathcal{B}\mathbf{x} = \mathcal{B}\mathbf{a_i}\}$$

and has the same form as equation 2, where $\mathbf{c} = \mathcal{B}\mathbf{a_i}$.

2 Hyper-plane

2.1 Definition

In a $\mathcal{R}^n$ space, the hyperplane is defined as a $n-1$ dimension subspace plus a translate $\mathbf{x_0}$, which is perpendicular to a one dimension vector $\mathbf{w}$. The formal mathematical expression of hyper-plane $\mathcal{H}$ is:

$$\mathcal{H} = \{\mathbf{y} + \mathbf{x_0}|\, \mathbf{w}^T\mathbf{y} = 0\} \tag{5}$$

We can denote $\mathbf{x} = \mathbf{y} + \mathbf{x_0}$, hence, equation 5 become:

$$\mathcal{H} = \{\mathbf{x}|\, \mathbf{w}^T(\mathbf{x} - \mathbf{x_0}) = 0\} \tag{6}$$

It is clearly that when $\mathbf{x} = \mathbf{x_0}$, $\mathbf{w}^T(\mathbf{x_0} - \mathbf{x_0}) = 0$ always hold, hence, $\mathbf{x_0}$ is a point within hyper-plane. As a result, equation 6 can be explained intuitively as: Hyper-plane is a set of any point $\mathbf{x}$ which satisfied the vector $\mathbf{x} - \mathbf{x_0}$(start from a fixed point $\mathbf{x_0}$ which belong to this hyper-plane, end at $\mathbf{x}$) is perpendicular to a normal vector $\mathbf{w}$.

Equation 6 give us another way to represent hyper-plane:

$$\mathcal{H} = \{\mathbf{x}|\, \mathbf{w}^T\mathbf{x} = \mathbf{b}\} \tag{7}$$

Where $\mathbf{b} = \mathbf{w}^T\mathbf{x_0}$.

2.2 Half-space

Intuitively, a hyper-plane split the whole space into two part, each of them is called half-space. The formal mathematical expression for half-space, base on equation 6, is:

$$\{\mathbf{x}|\, \mathbf{w}^T(\mathbf{x}-\mathbf{x_0}) < 0\}$$

or

$$\{\mathbf{x}|\, \mathbf{w}^T(\mathbf{x}-\mathbf{x_0}) > 0\}$$

Which corresponding to the two split space. In geometric, two vector $\mathbf{v_1}$ and $\mathbf{v_2}$ satisfied: $\mathbf{v_1}^T\mathbf{v_2} > 0$ if their angle smaller than 90 degree, otherwise, $\mathbf{v_1}^T\mathbf{v_2} < 0$. As a result, the half-space $\{\mathbf{x}|\, \mathbf{w}^T(\mathbf{x}-\mathbf{x_0}) < 0\}$ is the set of point $\mathbf{x}$ which satisfied the vector $\mathbf{x}-\mathbf{x_0}$ have a angle greater than 90 degree with a normal vector $\mathbf{w}$. The half-space $\{\mathbf{x}|\, \mathbf{w}^T(\mathbf{x}-\mathbf{x_0}) > 0\}$ can be explained in the same way except the angle is smaller than 90 degree.

Furthermore, according to 7, the expression of half-space can also be written as:

$$\{\mathbf{x}|\, \mathbf{w}^T\mathbf{x} < \mathbf{b}\}$$

Where $\mathbf{b} = \mathbf{w}^T\mathbf{x_0}$.

3 Cone

3.1 Definition

A cone $\mathcal{K}$ is the set of any point $\mathbf{x}$ that satisfied: If $\mathbf{x} \in \mathcal{K}$, then, for any $\theta \geq 0$, $\theta\mathbf{x} \in \mathcal{K}$. A **convex** cone, as its' name suggest, is a cone which is also a convex set.

3.2 Property of Convex Cone

Assume a cone $\mathcal{K}$ is a convex cone, then, it satisfied:

> If $\mathbf{x_1} \in \mathcal{K}$ and $\mathbf{x_2} \in \mathcal{K}$, then, for any $0 \leq \theta \leq 1$, $\theta\mathbf{x_1} + (1-\theta)\mathbf{x_2} \in \mathcal{K}$.

Furthermore, for any $p, q > 0$, the expression $p\mathbf{x_1} + q\mathbf{x_2}$ can be written as:

$$p\mathbf{x_1} + q\mathbf{x_2} = (p+q)(\frac{p}{p+q}\mathbf{x_1} + \frac{q}{p+q}\mathbf{x_2}) \tag{8}$$

Following the property of convex set, the point $(\frac{p}{p+q}\mathbf{x_1} + \frac{q}{p+q}\mathbf{x_2}) \in \mathcal{K}$. Additionally, since $(p+q) \in \mathcal{R}$ and $p+q > 0$, the definition of cone indicate $(p+q)(\frac{p}{p+q}\mathbf{x_1} + \frac{q}{p+q}\mathbf{x_2}) \in \mathcal{K}$. As a result, it can be seen within a convex cone $\mathcal{K}$:

> If $\mathbf{x_1}, \mathbf{x_2} \in \mathcal{K}$, then, for any $\theta_1, \theta_2 > 0$, $\theta_1\mathbf{x_1} + \theta_2\mathbf{x_2} \in \mathcal{K}$

3.3 Dual Cone

In geometric, a point $\mathbf{x}$ can be defined as the end point of a vector that start with the original point($\mathbf{0}$) and terminate at point $\mathbf{x}$, which means, any point $\mathbf{x}$ can also be thought as a vector $\mathbf{x}-\mathbf{0}$. Base on this, the dual cone of a cone $\mathcal{K}$, denote as $\mathcal{K}^*$, is defined as the set of any point λ which satisfied $\lambda^T\mathbf{x} > 0$ for all $\mathbf{x} \in \mathcal{K}$, which means, the **vector** λ have the angle smaller than 90 degree for any **vector** $\mathbf{x}-\mathbf{0}$ defined by the point $\mathbf{x}$ in $\mathcal{K}$.

Furthermore, the expression $\lambda^T \mathbf{x} > 0$ define a half-space according to equation 6, where the fixed point $\mathbf{x_0}$ is the original point. As a result, the dual cone of $\mathcal{K}$ can also be illustrated as:

> A point λ is in $\mathcal{K}^*$ if the half-space with the **vector** λ as normal vector, which expressed as $\lambda \mathbf{x} > 0$, contain the original cone $\mathcal{K}$.

3.4 Generalized Inequality

We can define the generalized inequality based on the concept of cone:

> $\mathbf{x}$ is smaller than $\mathbf{y}$ respect to a specific cone $\mathcal{K}$ if $\mathbf{y} - \mathbf{x} \in \mathcal{K}$. Denote as: $\mathbf{x} \prec_{\mathcal{K}} \mathbf{y}$

Start from the definition of generalized inequality, we can define **minimum** and **minimal** value of a set $\mathcal{S}$ with respect to a specific cone $\mathcal{K}$:

> The point $\mathbf{x_0} \in \mathcal{S}$ is said to be the **minimum** of $\mathcal{S}$ if for any other $\mathbf{y} \in \mathcal{S}$, the expression $\mathbf{x} \preceq_{\mathcal{K}} \mathbf{y}$ always holds.

Before we move to the definition of **minimal** point, we have to realize an important fact with respect to generalized inequality: When $\mathbf{y} \npreceq_{\mathcal{K}} \mathbf{x}$ hold, it **does not** mean $\mathbf{x} \preceq_{\mathcal{K}} \mathbf{y}$, which means, even though $\mathbf{y}$ is not smaller than $\mathbf{x}$ (with respect to a specific cone $\mathcal{K}$), it **does not** implies that $\mathbf{x}$ is smaller or equal to $\mathbf{y}$. This fact implies that a point smaller than any other element of a set is not equal to there is no other element within this set that smaller than the point. We can now introduce the definition of **minimal** point with respect to cone $\mathcal{K}$:

> A point $\mathbf{x_0}$ is minimal point of $\mathcal{S}$ if for any $\mathbf{y} \in \mathcal{S}$, $\mathbf{y} \npreceq_{\mathcal{K}} \mathbf{x_0}$ always hold unless $\mathbf{y} = \mathbf{x_0}$

Which means, if $\mathbf{x_0}$ is a minimal point of $\mathcal{S}$, there exist no element $\mathbf{y}$ within $\mathcal{S}$ which satisfied $\mathbf{y} \preceq_{\mathcal{K}} \mathbf{x_0}$(There is no element in $\mathcal{S}$ that smaller than $\mathbf{x_0}$). Furthermore, according to the fact that mentioned above, for some $\mathbf{y} \in \mathcal{S}$, even though $\mathbf{y} \npreceq_{\mathcal{K}} \mathbf{x_0}$, $\mathbf{x_0}$ does **not** need to satisfy $\mathbf{x_0} \preceq_{\mathcal{K}} \mathbf{y}$. As a result, $\mathbf{x_0}$ does not need to be the minimum point.

3.5 Minimum and Minimal with Dual Cone Perspective

If a point $\mathbf{x_0}$ is a minimum point of $\mathcal{S}$ with respect to cone $\mathcal{K}$, then, for any $\mathbf{x} \in \mathcal{S}$, there is:

$$\mathbf{x} - \mathbf{x_0} \in \mathcal{K}$$

According to the definition of dual cone, for the element $\mathbf{v}$ within $\mathcal{K}$, it must satisfy:

$$\lambda^T \mathbf{b} > 0$$

Where $\lambda \in \mathcal{K}^*$. Since $\mathbf{x} - \mathbf{x_0} \in \mathcal{K}$, we have:

$$\lambda^T (\mathbf{x} - \mathbf{x_0}) > 0$$

Which define a hyper-plane and a half-space with $\mathbf{x_0}$ as the fixed point according to equation 6. As a result, the minimum point of a set $\mathcal{S}$ with respect to a cone $\mathcal{K}$ and its' dual cone $\mathcal{K}^*$ can be illustrated as:

> If $\mathbf{x_0}$ is minimum point, then, for any $\lambda \in \mathcal{K}^*$, the hyper-plane and half-space defined by λ and $\mathbf{x_0}$: $\lambda^T (\mathbf{x} - \mathbf{x_0})$ should contain entire $\mathcal{S}$.

4 Convex Function

4.1 Definition

A function $f(\mathbf{x})$ is called convex function if it satisfied:

$$f(\theta\mathbf{x_2} + (1-\theta)\mathbf{x_2}]) \leq \theta\mathbf{x_1} + (1-\theta)\mathbf{x_2} \tag{9}$$

for any $\mathbf{x_1}, \mathbf{x_2} \in \mathbf{dom}f$ and $0 \leq \theta \leq 1$.

4.2 Conditions

A function $f(\mathbf{x})$ is a convex function if it satisfied one of the following conditions:

4.2.1 First-Order Condition

If $f(\mathbf{x})$ is differentiable, then, for any $\mathbf{x} \in \mathbf{dom}f$ and an arbitrary $\mathbf{x_0}$, the first-order derivative if f should satisfied:

$$f(\mathbf{x}) \geq f(\mathbf{x_0}) + \nabla f(\mathbf{x_0})(\mathbf{x} - \mathbf{x_0}) \tag{10}$$

The right hand side of this inequality is the first-order Taylor expression, which can be written as:

$$g(\mathbf{x}) = f(\mathbf{x_0}) + \nabla f(\mathbf{x_0})(\mathbf{x} - \mathbf{x_0})$$

In $\mathcal{R}^2$, $g(\mathbf{x})$ is a line that go through $(\mathbf{x_0}, f(\mathbf{x_0}))$ and 'support' f at this point. As a result, the first-order condition can be explained as: For any point $\mathbf{x}$, the value of function f at this point is greater or equal to the value of g at the same point $\mathbf{x}$. Moreover, in $\mathcal{R}^n$, $g(\mathbf{x})$ indeed defien a hyper-plane, the prove can be shown as: The function $g(\mathbf{x})$ can be written as:

$$\nabla f(\mathbf{x_0})(\mathbf{x} - \mathbf{x_0}) - (g(\mathbf{x}) - f(\mathbf{x_0})) = 0$$

Which has the form:

$$\begin{bmatrix} \nabla f(\mathbf{x_0}) \\ -1 \end{bmatrix}^T \left(\begin{bmatrix} \mathbf{x} \\ g(\mathbf{x}) \end{bmatrix} - \begin{bmatrix} \mathbf{x_0} \\ f(\mathbf{x_0}) \end{bmatrix}\right) = 0 \tag{11}$$

It is obvious that the above equation define a hyper-plane. Hence, the first-order condition can also be expression as: For any $\mathbf{x} \in \mathbf{dom}f$, the point $(\mathbf{x}, f(\mathbf{x}))$ is above the hyper-plane defined by equation 11 through some fixed point $(\mathbf{x_0}, \mathbf{f}(\mathbf{x_0}))$, which means, it is in the half-space given by this hyper-plane. This can be proved through:

$$\begin{bmatrix} \nabla f(\mathbf{x_0}) \\ -1 \end{bmatrix}^T \left(\begin{bmatrix} \mathbf{x} \\ f(\mathbf{x}) \end{bmatrix} - \begin{bmatrix} \mathbf{x_0} \\ f(\mathbf{x_0}) \end{bmatrix}\right) \leq 0$$

Furthermore, when $\nabla f(\mathbf{x_0}) = 0$ in some point $(\mathbf{x_0}, f(\mathbf{x_0}))$, the equation 10 become:

$$f(\mathbf{x}) \geq f(\mathbf{x_0})$$

Hence, $f(\mathbf{x_0})$ is the minimum value of function f, which is why for convex function, the point that first-order derivative equal to zero is the minimizer of this function.