Convex sets, convex functions, & some of their properties. (Part I)

John MacLaren Walsh, Ph.D.

April 16, 2014

1 References

- S. Boyd and L. Vandenberghe, *Convex Optimization*, Cambridge Univ. Press, 2004.
- R. T. Rockafellar, *Convex Analysis*, Princeton Univ. Press, 1970.

The material in this lecture follows the first three chapters of these two books closely, as well as several other chapters from Rockafellar.

2 Affine Sets, Affine Hulls, & Dimension

Consider two points $\mathbf{x}_1, \mathbf{x}_2 \in \mathbb{R}^N$, $\mathbf{x}_1 \neq \mathbf{x}_2$. The line through these two points may be represented as $\{\theta\mathbf{x}_1 + (1-\theta)\mathbf{x}_2 | \theta \in \mathbb{R}\}$. A set $\mathcal{A} \subset \mathbb{R}^N$ is *affine* if the line through any two points in $\mathcal{A}$ is itself in $\mathcal{A}$. In other words, $\mathcal{A}$ is affine if whenever $\mathbf{x}_1, \mathbf{x}_2 \in \mathcal{A}$, $\theta\mathbf{x}_1 + (1-\theta)\mathbf{x}_2 \in \mathcal{A}$ for all $\theta \in \mathbb{R}$.

More generally, an affine set contains all *affine combinations* of its points, in that if $\mathbf{x}_1, \dots, \mathbf{x}_n \in \mathcal{A}$ for an affine set $\mathcal{A}$, then

$$\sum_{i=1}^{n} \theta_i \mathbf{x}_i \in \mathcal{A} \quad \forall(\theta_1, \dots, \theta_n) \text{ such that} \sum_{i=1}^{n} \theta_i = 1 \tag{1}$$

Affine sets containing the origin $\mathbf{0}$ are vector subspaces (observe that they contain all linear combinations of points within them by including $\mathbf{0}$ in an associated affine combination).

Two affine sets $\mathcal{A}_1$ and $\mathcal{A}_2$ are said to be parallel if there exists a vector $\mathbf{x}$ such that $\mathcal{A}_2 = \mathcal{A}_1 - \mathbf{x}$.

Every affine set $\mathcal{A}$ is parallel to a unique vector subspace, which may be found as $\mathcal{A} - \mathbf{x}_0$, for any $\mathbf{x}_0 \in \mathcal{A}$. To see this, observe that the new set will again be affine and will now contain $\mathbf{0}$.

Put another way, every affine set $\mathcal{A}$ can be represented $\mathcal{A} = \mathcal{V} + \mathbf{v}$, i.e. as a vector subspace $\mathcal{V}$ plus some offset $\mathbf{v}$.

From these facts follow the definition of the *dimension* of an affine set $\mathcal{A}$, which is simply the dimension of the associated vector space $\mathcal{V}$.

Take an arbitrary set $\mathcal{C} \subset \mathbb{R}^N$. The *affine hull* of $\mathcal{C}$ denoted by $\text{aff}(\mathcal{C})$ is the set of affine combinations of points in $\mathcal{C}$. It is the smallest affine set containing $\mathcal{C}$.

The *affine dimension* of some set $\mathcal{C} \subset \mathbb{R}^N$ is then the dimension of its affine hull.

Affine hulls have some topological benefits as they allow us to define an appropriate extension of the idea of interior points of a set when dealing with sets that live in affine subsets of $\mathbb{R}^N$. The *relative interior* of a set $\mathcal{C}$ is its interior when regarded within the metric space on its affine hull, that is

$$\{\mathbf{x} \in \mathcal{C} \,|\, B(\mathbf{x}, r) \cap \text{aff}(\mathcal{C}) \subseteq \mathcal{C}, \text{ for some } r > 0\} \tag{2}$$

Naturally, the *relative boundary* of a set $\mathcal{C}$ is the set of those points in the closure of $\mathcal{C}$ not in the relative interior of $\mathcal{C}$.

3 Convex Sets

When dealing with affine sets, we were interested in constructions built from lines, whereas convex theory is concerned with constructions built from *line segments*. Formally, given two points $\mathbf{x}_1, \mathbf{x}_2 \in \mathbb{R}^N$, the line segment between these two points can be represented as $\{\lambda\mathbf{x}_1 + (1-\lambda)\mathbf{x}_2 \,|\lambda \in [0, 1]\}$.

A *convex set* is a set which contains all of the line segments whose endpoints are in it. Formally, a set $\mathcal{C} \subset \mathbb{R}^N$ is said to be convex if for any $\mathbf{x}_1$ and $\mathbf{x}_2$ in $\mathcal{C}$ the point $\lambda\mathbf{x}_1 + (1-\lambda)\mathbf{x}_2 \in \mathcal{C}$ for any $\lambda \in [0, 1]$.

By applying this property several times, we observe that a convex set $\mathcal{C}$ must contain any *convex combination*

$$\sum_{i=1}^{k} \lambda_i \mathbf{x}_i, \text{such that} \quad \lambda_i \geq 0 \; \forall i \in \{1, \ldots, k\} \quad \text{and} \; \sum_{i=1}^{k} \lambda_i = 1 \tag{3}$$

of points $\mathbf{x}_1, \ldots, \mathbf{x}_k$ of points in $\mathcal{C}$.

Given any set $\mathcal{C} \subset \mathbb{R}^N$, the *convex hull* of this set $\text{conv}(\mathcal{C})$ is the set of convex combinations of points in $\mathcal{C}$

$$\text{conv}(\mathcal{C}) = \left\{ \sum_{i=1}^{k} \lambda_i \mathbf{x}_i \;\middle|\; \mathbf{x}_i \in \mathcal{C}, \; \lambda_i \geq 0 \quad \forall i \in \{1, \ldots, k\}, \sum_{i=1}^{k} \lambda_i = 1, \; k \in \mathbb{N} \right\} \tag{4}$$

$\text{conv}(\mathcal{C})$ is the smallest convex set containing $\mathcal{C}$, and is hence also the intersection of all convex sets containing $\mathcal{C}$. *Carathéodory's theorem* tells us that in forming the convex hull, it suffices to consider convex combinations of order $k \leq N+1$, i.e. one only needs combine no more than a number of points equal to the dimension of the space plus 1, so that

$$\text{conv}(\mathcal{C}) = \left\{ \sum_{i=1}^{N+1} \lambda_i \mathbf{x}_i \;\middle|\; \mathbf{x}_i \in \mathcal{C}, \; \lambda_i \geq 0 \quad \forall i \in \{1, \ldots, N+1\}, \sum_{i=1}^{N+1} \lambda_i = 1 \right\} \tag{5}$$

Some familiar convex sets we discussed in class include polyhedra and norm balls.

3.1 Convex Cones

A set $\mathcal{C} \subset \mathbb{R}^N$ is called a *convex cone* if for any $\mathbf{x}_1, \mathbf{x}_2 \in \mathcal{C}$, the point

$$\mu_1 \mathbf{x}_1 + \mu_2 \mathbf{x}_2 \in \mathcal{C} \tag{6}$$

for any $\mu_1, \mu_2 \geq 0$. Apply this definition recursively, we observe that a convex cone $\mathcal{C}$ contains all *conic combinations*

$$\sum_{i=1}^{k} \mu_i \mathbf{x}_i, \quad \mu_i \geq 0, \; i \in \{1, \ldots, k\} \tag{7}$$

of points $\mathbf{x_1}, \ldots, \mathbf{x}_k$ in $\mathcal{C}$.

Given an arbitrary set $\mathcal{C} \subset \mathbb{R}^N$, the *conic hull* of $\mathcal{C}$, denoted by $\text{cone}(\mathcal{C})$, is the set of conic combinations of points in $\mathcal{C}$

$$\text{cone}(\mathcal{C}) = \left\{ \sum_{i=1}^{k} \mu_i \mathbf{x}_i \mid \mathbf{x}_i \in \mathcal{C}, \; \mu_i \geq 0, \; i \in \{1, \ldots, k\}, \; k \in \mathbb{N} \right\} \tag{8}$$

and is also the smallest convex cone containing $\mathcal{C}$.

Some familiar convex cones include the solution set to a system of homogeneous linear inequalities $\{\mathbf{x} \in \mathbb{R}^N \mid \mathbf{A}\mathbf{x} \leq \mathbf{0}\}$, and the set of positive semidefinite symmetric matrices (sometimes denoted by $\mathbb{S}_+^N$).

3.2 Operations on Convex Sets

The intersection of a collection of convex sets is itself also convex, that is, if $\mathcal{C}_i$ are convex sets for $i \in \{1, \ldots, I\}$, then $\bigcap_{i=1}^{I} \mathcal{C}_i$ is also convex.

Additionally, convexity is preserved through images under affine functions. Formally, a function $f : \mathbb{R}^N \to \mathbb{R}^M$ is *affine* if it can be represented as $f(\mathbf{x}) = \mathbf{A}\mathbf{x} + \mathbf{b}$. If $\mathcal{C}$ is a convex set, then $f(\mathcal{C}) = \{\mathbf{y} \mid \exists \mathbf{x} \in \mathbb{R}^N, \; \mathbf{y} = f(\mathbf{x})\}$.

3.3 Faces, Exposed Faces, Extreme Points

A *face* $\mathcal{F}$ of a convex set $\mathcal{C}$ is a convex set $\mathcal{F} \subseteq \mathcal{C}$ such that every line segment in $\mathcal{C}$ with a relative interior point in $\mathcal{F}$ must have both endpoints in $\mathcal{F}$. Put another way, a face $\mathcal{F}$ of a convex set $\mathcal{C}$ is a convex subset $\mathcal{F} \subset \mathcal{C}$ such that whenever $\lambda \mathbf{x}_1 + (1-\lambda)\mathbf{x}_2 \in \mathcal{F}$ for some $\lambda \in (0,1), \; \mathbf{x}_1, \mathbf{x}_2 \in \mathcal{C}$, we also have $\mathbf{x}_1, \mathbf{x}_2 \in \mathcal{F}$. The empty set and $\mathcal{C}$ itself are faces of $\mathcal{C}$.

The zero dimensional faces of a convex set $\mathcal{C}$ are called *extreme points* of $\mathcal{C}$.

If $\mathcal{F}$ is a half-line $\mathcal{F} = \{\mathbf{x} + \gamma \mathbf{y} \mid \gamma \geq 0\}$that is a face of $\mathcal{C}$, then $\mathbf{y}$ is called an *extreme direction* of $\mathcal{C}$.

If $\mathcal{E}$ is the set of points where a certain linear function $h(\mathbf{x}) = \langle \mathbf{a}, \mathbf{x} \rangle$ attains its maximum over a convex set $\mathcal{C}$, then $\mathcal{E}$ is a face of $\mathcal{C}$, and is called an *exposed face.*

Exposed faces that are points are called are called *exposed points*, and the directions of exposed faces that are half-lines are called *exposed directions.*

Not all faces need to be exposed. As an example, consider $\mathcal{C}$ to be the convex hull of the torus. The relative boundary of the top disc and the bottom disc are extreme points of $\mathcal{C}$, but are not exposed points (although they are contained within the closed discs which are exposed faces).

Straszewicz's Theorem tells us that the set of exposed points of a closed convex $\mathcal{C}$ is a dense subset of the extreme points of $\mathcal{C}$. Thus, any extreme point of $\mathcal{C}$ has a sequence of exposed points of $\mathcal{C}$ which converges to it.

Just as polyhedra could possible contain lines and halflines, their generalization convex sets can as well.

In particular, we can define the *lineality space* $\mathcal{L}$ of a convex set $\mathcal{C} \subset \mathbb{R}^N$ to be the set of $\mathbf{y} \in \mathbb{R}^N$ such that for all $\mathbf{x} \in \mathcal{C}$, the line $\{\mathbf{x} + \alpha \mathbf{y} |\ \alpha \in \mathbb{R}\} \subset \mathcal{C}$.

The *recession cone* $\mathcal{C}^\infty$ of a convex set $\mathcal{C} \subset \mathbb{R}^N$ is defined as the set of all $\mathbf{y} \in \mathbb{R}^N$ such that for every $\mathbf{x} \in \mathcal{C}$ the halfline $\{\mathbf{x} + \gamma \mathbf{y} | \gamma \geq 0\} \subset \mathcal{C}$. The recession cone of a convex set is a convex cone.

Under this definition, it is clear that the lineality space of a convex set can be thought of the intersection of the recession cone with its negative, $\mathcal{C}^\infty \cap -\mathcal{C}^\infty$.

3.4 Representations of Convex Sets

The insights from the dual structure of polyhedra are extensible to closed convex sets.

Let $\mathcal{C}$ be a closed convex set containing no lines (i.e. intersect with the orthogonal complement of the lineality space if it did), then $\mathcal{C}$ is the convex hull of the set of its extreme points and extreme directions. Formally, any closed convex set $\mathcal{C}$ can be represented as

$$\mathcal{C} = \mathcal{L} + \mathrm{conv}\,(\mathcal{S}) + \mathrm{cone}\,(\mathcal{T}) \tag{9}$$

where $\mathcal{L}$ is the lineality space of $\mathcal{C}$, $\mathcal{S}$ is the set of extreme points of $\mathcal{C} \cap \mathcal{L}^\perp$, and the $\mathcal{T}$ forms the extreme directions of $\mathcal{C} \cap \mathcal{L}^\perp$.

Further, building on Straszewicz, a closed convex set $\mathcal{C}$ with no lines is the closure of the convex hull of its exposed points and exposed directions.

4 Convex Functions

A function $f : \mathcal{C} \to \mathbb{R}$, with domain $\mathcal{C} \subset \mathbb{R}^N$ is said to be a *convex function* if for any $\mathbf{x}, \mathbf{y} \in \mathcal{C}$,

$$f(\lambda \mathbf{x} + (1-\lambda)\mathbf{y}) \leq \lambda f(\mathbf{x}) + (1-\lambda) f(\mathbf{y}) \tag{10}$$

Furthermore, f is said to be strictly convex if $\mathcal{C}$ is convex and

$$f(\lambda \mathbf{x} + (1-\lambda)\mathbf{y}) < \lambda f(\mathbf{x}) + (1-\lambda) f(\mathbf{y}) \tag{11}$$

for any $\mathbf{x}, \mathbf{y} \in \mathcal{C}$ with $\mathbf{x} \neq \mathbf{y}$.

A function g is said to be concave (or convex-down) if $-g$ is convex.

4.1 Differentiability and Convexity

If a function f is differentiable on an open domain $\mathcal{C}$, then it is convex if and only if

$$f(\mathbf{y}) \geq f(\mathbf{x}) + \langle \nabla f(\mathbf{x}), \mathbf{y} - \mathbf{x} \rangle \tag{12}$$

That is, f must lie above or on its first order Taylor series at each $\mathbf{x}$. Additionally, a differentiable f is strictly convex if and only if the inequality is strict for any $\mathbf{x} \neq \mathbf{y}$ with $\mathbf{x}, \mathbf{y} \in \mathcal{C}$.

As we will discuss more in a later lecture, using a strictly convex differentiable function, one can build a *divergence* (a function between two point reflecting some of the properties of a distance), called a *Bregman divergence* using this property through

$$D_f(\mathbf{y}, \mathbf{x}) = f(\mathbf{y}) - f(\mathbf{x}) - \langle \nabla f(\mathbf{x}), \mathbf{y} - \mathbf{x} \rangle \tag{13}$$

Observe that $D_f(\mathbf{x}, \mathbf{y}) \geq 0$ with equality if and only if $\mathbf{x} = \mathbf{y}$.

Further, if f is twice continuously differentiable at every point in a convex open domain $\mathcal{C}$, then it is convex if and only if the Hessian matrix $\nabla^2 f(\mathbf{x})$ is positive semi-definite for all $\mathbf{x} \in \mathcal{C}$.

Additionally, if $\nabla^2 f(\mathbf{x})$ is positive definite for all $\mathbf{x} \in \mathcal{C}$ for an open convex domain $\mathcal{C}$, then it is strictly convex. On the other hand, positive definiteness of the Hessian is not necessary for strict convexity.

4.2 Some Examples

Some examples of convex functions (straight from Boyd's list):

- Exponential $f(x) = e^{ax},\ a \in \mathbb{R}$ on domain $\mathbb{R}$.
- $f(x) = x^a$ on $[0, \infty)$ for $a \geq 1$
- $f(x) = -\log(x)$ on $(0, \infty)$.
- $f(x) = x \log(x)$ on $[0, \infty)$.
- Norms $f(\mathbf{x}) = \|\mathbf{x}\|$.
- Max $f(\mathbf{x}) = \max\{x_1, x_2, \ldots, x_N\}$.
- log-sum-exp (Soft-max) $f(\mathbf{x}) = \log\left(\exp(x_1) + \exp(x_2) + \cdots + \exp(x_N)\right)$.
- Log-det $f(\mathbf{X}) = \log\det \mathbf{X}$ on the domain $\mathbb{S}_{++}$.

4.3 Operations Preserving Convexity

Again straight from Boyd:

- **conic combinations** Any conic combination of convex function is again convex, i.e if $\gamma_1, \ldots, \gamma_N \geq 0$ and $f_1, \ldots, f_N$ are convex function on a common convex domain then

  $$\sum_{i=1}^{N} \gamma_i f(\mathbf{x}) \tag{14}$$

  is also a convex function.

- **composition with affine functions** if $g(\mathbf{x}) = f(\mathbf{A}\mathbf{y} + \mathbf{b})$ for some convex f, then g will be convex.

- **maximums:** if $f_1, \ldots, f_m$ are convex, then so is

  $$\max\{f_1(\mathbf{x}), \ldots, f_m(\mathbf{x})\} \tag{15}$$

  Similarly, if for every $\mathbf{y} \in \mathcal{A}$, $f(\mathbf{x}, \mathbf{y})$ is convex in $\mathbf{x}$, then

  $$\sup_{\mathbf{y} \in \mathcal{A}} f(\mathbf{x}, \mathbf{y}) \tag{16}$$

  is convex on $\{\mathbf{x} | (x, y) \in \mathcal{C}\ \forall \mathbf{y} \in \mathcal{A}, \sup_{\mathbf{y} \in \mathcal{A}} f(\mathbf{x}, \mathbf{y}) < \infty\}$.

- **minimization in an argument**: if $f(\mathbf{x}, \mathbf{y})$ is convex in (x, y) and $\mathcal{Y}$ is a convex nonempty set, then $g(\mathbf{x}) = \inf_{\mathbf{y} \in \mathcal{Y}} f(\mathbf{x}, \mathbf{y})$ is a convex function provided $\exists \mathbf{x}$ such that $g(\mathbf{x}) > -\infty$.

Other important ones, including rules for composition, can be found in Boyd.