The University of Texas at Austin
Department of Electrical and Computer Engineering

ECE 381V: Large-Scale Optimization II — Spring 2022

LECTURE 2

Caramanis & Mokhtari Monday, January 24, 2022

Goal: In the previous lecture. we talked about the general agenda of this course. In this lecture, we recap some basics of optimization theory that we will need throughout this course, including optimality conditions, and definitions of convex functions and convex sets.

1 Optimality Conditions

A differentiable function f at any point $\mathbf{x}$ can be linearly approximated by

$$\hat{f}_{1st}(\mathbf{x}; \hat{\mathbf{x}}) := f(\hat{\mathbf{x}}) + \nabla f(\hat{\mathbf{x}})^\top (\mathbf{x} - \hat{\mathbf{x}})$$

where by Taylor's expansion the error of approximation is $o(\|\hat{\mathbf{x}} - \mathbf{x}\|)$. In other words we have

$$\begin{aligned} f(\mathbf{x}) &= f(\hat{\mathbf{x}}) + \nabla f(\hat{\mathbf{x}})^\top (\mathbf{x} - \hat{\mathbf{x}}) + o(\|\mathbf{x} - \hat{\mathbf{x}}\|) \\ &= \hat{f}_{1st}(\mathbf{x}; \hat{\mathbf{x}}) + o(\|\mathbf{x} - \hat{\mathbf{x}}\|) \end{aligned} \tag{1}$$

Using this observation we can define the first-order optimality condition for unconstrained problems.

Theorem 1. *(First-Order Optimality Condition) Let $\mathbf{x}^*$ be a local minimum of a differentiable function f. Then,*

$$\nabla f(\mathbf{x}^*) = \mathbf{0}.$$

Proof. As $\mathbf{x}^*$ is a local minimum of f there exists a neighborhood around $\mathbf{x}^*$ defined as $\mathcal{N}_{\mathbf{x}^*} = \{\mathbf{x} \mid \|\mathbf{x} - \mathbf{x}^*\| \leq r\}$, where $r > 0$, such that

$$f(\mathbf{x}^*) \leq f(\mathbf{x}) \qquad \text{for all } \mathbf{x} \in \mathcal{N}_{\mathbf{x}^*}$$

Suppose $\nabla f(\mathbf{x}^*) \neq \mathbf{0}$. Now consider the point $\mathbf{y} = \mathbf{x}^* - \alpha \nabla f(\mathbf{x}^*)/\|\nabla f(\mathbf{x}^*)\|$. Hence, we have $\|\mathbf{y} - \mathbf{x}^*\| = \alpha$. Now if we select $\alpha < r$, then $\mathbf{y} \in \mathcal{N}_{\mathbf{x}^*}$. Moreover, we have:

$$\begin{aligned} f(\mathbf{y}) &= f(\mathbf{x}^*) + \nabla f(\mathbf{x}^*)^\top (\mathbf{x}^* - \mathbf{y}) + o(\|\mathbf{x}^* - \mathbf{y}\|) \\ &= f(\mathbf{x}^*) - \alpha \|\nabla f(\mathbf{x}^*)\| + o(\alpha) \end{aligned} \tag{2}$$

Hence, by making α sufficiently small we will have $f(\mathbf{y}) < f(\mathbf{x}^*)$ which is a contradiction. $\square$

Remark 1. $\nabla f(\mathbf{x}^*) = \mathbf{0}$ *is a necessary condition for a local minimum* $\mathbf{x}^*$.

A twice differentiable function f at any point $\mathbf{x}$ can be quadratically approximated by

$$\hat{f}_{2nd}(\mathbf{x}; \hat{\mathbf{x}}) := f(\hat{\mathbf{x}}) + \nabla f(\hat{\mathbf{x}})^\top (\mathbf{x} - \hat{\mathbf{x}}) + \frac{1}{2}(\mathbf{x} - \hat{\mathbf{x}})^\top \nabla^2 f(\hat{\mathbf{x}})(\mathbf{x} - \hat{\mathbf{x}})$$

where by Taylor's expansion the error of approximation is $o(\|\hat{\mathbf{x}} - \mathbf{x}\|^2)$. In other words, we have

$$\begin{aligned} f(\mathbf{x}) &= f(\hat{\mathbf{x}}) + \nabla f(\hat{\mathbf{x}})^\top(\mathbf{x} - \hat{\mathbf{x}}) + \frac{1}{2}(\mathbf{x} - \hat{\mathbf{x}})^\top \nabla^2 f(\hat{\mathbf{x}})(\mathbf{x} - \hat{\mathbf{x}}) + o(\|\mathbf{x} - \hat{\mathbf{x}}\|^2) \\ &= \hat{f}_{2nd}(\mathbf{x}; \hat{\mathbf{x}}) + o(\|\mathbf{x} - \hat{\mathbf{x}}\|^2) \end{aligned} \tag{3}$$

Theorem 2. *(Second-Order Optimality Condition) Let $\mathbf{x}^*$ be a local minimum of a twice differentiable function f. Then,*

$$\nabla f(\mathbf{x}^*) = \mathbf{0}, \qquad \nabla^2 f(\mathbf{x}^*) \succeq \mathbf{0}.$$

Proof. As $\mathbf{x}^*$ is a local minimum of f there exists a neighborhood around $\mathbf{x}^*$ defined as $\mathcal{N}_{\mathbf{x}^*} = \{\mathbf{x} \mid \|\mathbf{x} - \mathbf{x}^*\| \leq r\}$, where $r > 0$, such that

$$f(\mathbf{x}^*) \leq f(\mathbf{x}) \qquad \text{for all } \mathbf{x} \in \mathcal{N}_{\mathbf{x}^*}$$

Further, we know that $\nabla f(\mathbf{x}^*) = \mathbf{0}$. Suppose $\nabla^2 f(\mathbf{x}^*) \succeq \mathbf{0}$ does not hold and therefore we have $\lambda_{min} \nabla^2(f(\mathbf{x}^*)) < 0$. Now consider the point $\mathbf{y} = \mathbf{x}^* - \alpha\mathbf{v}$, where $\mathbf{v}$ is the eigenvector corresponding to the smallest eigenvalue of $\nabla^2 f(\mathbf{x}^*)$. Then, we have $\|\mathbf{y} - \mathbf{x}^*\| = \alpha\|\mathbf{v}\|$. Hence, if we select $\alpha < r$, then $\mathbf{y} \in \mathcal{N}_{\mathbf{x}^*}$. Moreover, we have:

$$\begin{aligned} f(\mathbf{y}) &= f(\mathbf{x}^*) + \frac{1}{2}(\mathbf{y} - \mathbf{x}^*)^\top \nabla^2 f(\mathbf{x}^*)(\mathbf{y} - \mathbf{x}^*) + o(\|\mathbf{x}^* - \mathbf{y}\|^2) \\ &= f(\mathbf{x}^*) - \alpha^2 \lambda_{min}(\nabla^2 f(\mathbf{x}^*)) + o(\alpha^2) \end{aligned} \tag{4}$$

Hence, by making α sufficiently small we will have $f(\mathbf{y}) < f(\mathbf{x}^*)$ which is a contradiction. □

Remark 2. $\nabla f(\mathbf{x}^*) = \mathbf{0}$ *and* $\nabla^2 f(\mathbf{x}^*) \succeq \mathbf{0}$ *are necessary conditions for a local minimum* $\mathbf{x}^*$.

Theorem 3. *(Sufficient Conditions) Let a function f be twice differentiable on $\mathbb{R}^n$. If a point $\mathbf{x}^*$ satisfies the conditions*

$$\nabla f(\mathbf{x}^*) = \mathbf{0}, \qquad \nabla^2 f(\mathbf{x}^*) \succ \mathbf{0},$$

then $\mathbf{x}^*$ is a strict local minimum of f.

Proof. Consider the neighborhood around $\mathbf{x}^*$ with radius r where r is small enough such that

$$-\frac{1}{4} r^2 \lambda_{min} \leq o(r^2) \leq \frac{1}{4} r^2 \lambda_{min}$$

Consider an arbitrary vector $\mathbf{d} \in \mathbb{R}^n$ with norm 1, i.e., $\|\mathbf{d}\| = 1$. Further, consider a point in the neighborhood $\|\mathbf{x}^* - \mathbf{y}\| \leq r$ defined as $\mathbf{y} = \mathbf{x}^* + \alpha\mathbf{d}$ where $\alpha \in (0, r]$. As $\alpha \in (0, r]$, we have

$$-\frac{1}{4}\alpha^2 \lambda_{min} \leq o(\alpha^2) \leq \frac{1}{4}\alpha^2 \lambda_{min}$$

For such point we have

$$\begin{aligned} f(\mathbf{y}) &= f(\mathbf{x}^*) + \frac{1}{2}(\mathbf{y} - \mathbf{x}^*)^\top \nabla^2 f(\mathbf{x}^*)(\mathbf{y} - \mathbf{x}^*) + o(\|\mathbf{x}^* - \mathbf{y}\|^2) \\ &= f(\mathbf{x}^*) + \frac{1}{2}\alpha^2 \mathbf{d}^\top \nabla^2 f(\mathbf{x}^*)\mathbf{d} + o(\alpha^2) \\ &\geq f(\mathbf{x}^*) + \frac{1}{2}\alpha^2 \lambda_{min}(\nabla^2 f(\mathbf{x}^*)) + o(\alpha^2) \\ &\geq f(\mathbf{x}^*) + \frac{1}{4}\alpha^2 \lambda_{min}(\nabla^2 f(\mathbf{x}^*)) \\ &> f(\mathbf{x}^*). \end{aligned} \tag{5}$$

Hence, $\mathbf{x}^*$ is a strict local minimum of f. □

2 Convex Functions

We give three definitions of convex functions: A basic definition that requires no differentiability of the function; a first-order definition of convexity that uses the gradient; and a second-order condition of convexity. We show that (for smooth functions) these three are equivalent. Later in the course, it will be important to deal with non-differentiable functions.
First, we define the set of points where a function is finite.

Definition 1 (Domain of a function)**.** *The domain of a function* $f : \mathbb{R}^n \to \mathbb{R}$ *is denoted by* $\mathrm{dom}(f)$*, and is defined as the set of points where a function* f *is finite:*

$$\mathrm{dom}(f) = \{\mathbf{x} \in \mathbb{R}^n \,:\, f(\mathbf{x}) < \infty\}.$$

2.1 Zeroth-order definition

Definition 2 (Convexity I)**.** *A function* $f : \mathbb{R}^n \to \mathbb{R}$ *is convex if* $dom(f)$ *is convex and* for all $\mathbf{x}_1, \mathbf{x}_2 \in \mathrm{dom}(f) \subseteq \mathbb{R}^n$*,* $\lambda \in [0, 1]$*, we have:*

$$f\left(\lambda \mathbf{x}_1 + (1-\lambda)\,\mathbf{x}_2\right) \;\le\; \lambda f(\mathbf{x}_1) + (1-\lambda)\, f(\mathbf{x}_2). \tag{6}$$

This inequality is illustrated in Figure 1.

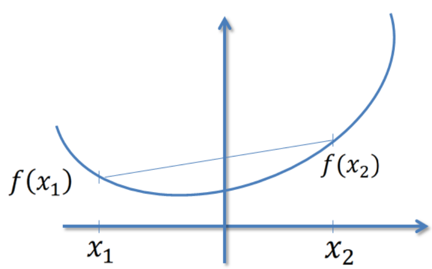

Figure 1: Convex functions

Remark 3. *We say that* f *is concave if* $-f$ *is convex.*

Definition 3 (Strict Convexity)**.** *A function* $f : \mathbb{R}^n \to \mathbb{R}$ *is strictly convex if* $dom(f)$ *is convex and* for all $\mathbf{x}_1, \mathbf{x}_2 \in \mathrm{dom}(f) \subseteq \mathbb{R}^n$ *where* $\mathbf{x}_1 \neq \mathbf{x}_2$ *and* $\lambda \in (0, 1)$*, we have:*

$$f\left(\lambda \mathbf{x}_1 + (1-\lambda)\,\mathbf{x}_2\right) \;<\; \lambda f(\mathbf{x}_1) + (1-\lambda)\, f(\mathbf{x}_2). \tag{7}$$

Show that the following functions are convex:

$$f(\mathbf{x}) = \mathbf{a}^\top \mathbf{x} + b, \qquad f(\mathbf{x}) = \|\mathbf{x}\|_p, \qquad f(\mathbf{X}) = \mathbf{tr}(\mathbf{A}^\top \mathbf{X}) + b = \sum_{i=1}^{m}\sum_{j=1}^{n} A_{ij}X_{ij} + b$$

2.2 First-order definition

We now give the first-order condition for convexity.

Definition 4 (Convexity II)**.** *Suppose a function $f : \mathbb{R}^n \to \mathbb{R}$ is differentiable. Then it is convex if and only if dom(f) is convex and*

$$f(\mathbf{y}) \geq f(\mathbf{x}) + \nabla f(\mathbf{x})^T (\mathbf{y} - \mathbf{x}) \,. \tag{8}$$

Intuitively speaking, equation (8) states that the first-order Taylor approximation is in fact a global underestimator of the function, as illustrated in figure 2.

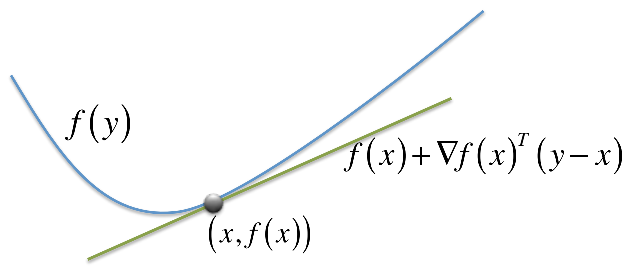

Figure 2: If function f is convex and differentiable, then eq. (8) is a global underestimator of f.

Proposition 1. *For differentiable functions, definition 2 and definition 4 are equivalent.*

Proof. If f is convex, by definition I, we know that for any $\lambda \in [0, 1]$ we have

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

Rearranging terms and dividing by λ, we obtain:

$$f(\mathbf{y}) - f(\mathbf{x}) \geq \frac{f(\mathbf{x} + \lambda(\mathbf{y} - \mathbf{x})) - f(\mathbf{x})}{\lambda}.$$

By sending $\lambda \to 0$ we obtain that

$$f(\mathbf{y}) - f(\mathbf{x}) \geq \nabla f(\mathbf{x})^\top(\mathbf{y} - \mathbf{x}).$$

Conversely, suppose now that function f satisfies the first-order condition. Let $\bar{\mathbf{x}} = \lambda\mathbf{x}_1 + (1-\lambda)\mathbf{x}_2$ be some point in the convex hull ($\lambda \in (0, 1)$). Hence, we have

$$\begin{aligned} f(\mathbf{x}_1) &\geq f(\bar{\mathbf{x}}) + \nabla f(\bar{\mathbf{x}})^\top(\mathbf{x}_1 - \bar{\mathbf{x}}), \\ f(\mathbf{x}_2) &\geq f(\bar{\mathbf{x}}) + \nabla f(\bar{\mathbf{x}})^\top(\mathbf{x}_2 - \bar{\mathbf{x}}). \end{aligned}$$

Multiplying the first inequality by λ, the second by $(1 - \lambda)$ and adding, we obtain

$$\begin{aligned} \lambda f(\mathbf{x}_1) + (1-\lambda)f(\mathbf{x}_2) &\geq f(\bar{\mathbf{x}}) + \lambda\nabla f(\bar{\mathbf{x}})^\top(\mathbf{x}_1 - \bar{\mathbf{x}}) + (1-\lambda)\nabla f(\bar{\mathbf{x}})^\top(\mathbf{x}_2 - \bar{\mathbf{x}}) \\ &= f(\bar{\mathbf{x}}) + \lambda\nabla f(\bar{\mathbf{x}})^\top \left(\lambda\mathbf{x}_1 + (1-\lambda)\mathbf{x}_2 - \bar{\mathbf{x}}\right) \\ &= f(\bar{\mathbf{x}}) \\ &= f(\lambda\mathbf{x}_1 + (1-\lambda)\mathbf{x}_2) \end{aligned}$$

□

2.3 Second-order definition

We now give the final definition of convexity.

Definition 5 (Convexity III)**.** *Suppose that a function $f : \mathbb{R}^n \to \mathbb{R}$ is twice differentiable. Then f is convex iff its Hessian is positive semidefinite:*

$$\bigtriangledown^2 f(\mathbf{x}) \succeq 0, \text{ for all } \mathbf{x} \in dom(f) \tag{9}$$

As an example, consider the function:

$$f(\mathbf{x}) = \|\mathbf{A}\mathbf{x} - \mathbf{b}\|_2^2.$$

Expanding, we have $f(\mathbf{x}) = \mathbf{x}^\top \mathbf{A}^\top \mathbf{A}\mathbf{x} - \mathbf{b}^\top \mathbf{A}\mathbf{x} - \mathbf{x}^\top \mathbf{A}^\top \mathbf{b} + \|\mathbf{b}\|_2^2$. The Hessian of this is $2\mathbf{A}^\top \mathbf{A}$, which is positive semidefinite, since for any vector $\mathbf{x}$, $\mathbf{x}^\top(\mathbf{A}^\top \mathbf{A})\mathbf{x} = \|\mathbf{A}\mathbf{x}\|_2^2 \geq 0$.

Proposition 2. *The definition give above is equivalent to Definitions 2 and 4. That is, a twice differentiable function f is convex if*

$$\nabla^2 f(\mathbf{x}) \in \mathcal{S}_+^n. \tag{10}$$

Proof. One way to prove this is via Taylor's theorem, using the Lagrange form of the remainder. To prove that a function with positive semidefinite Hessian is convex, using a second order Taylor expansion we have:

$$f(\mathbf{y}) = f(\mathbf{x}) + \nabla f(\mathbf{x})^T (\mathbf{y} - \mathbf{x}) + (\mathbf{y} - \mathbf{x})^T \nabla^2 f(\mathbf{x} + \alpha(\mathbf{y} - \mathbf{x}))(\mathbf{y} - \mathbf{x}) \tag{11}$$

for some value of $\alpha \in [0, 1]$. Now, since the Hessian is positive semidefinite,

$$(\mathbf{y} - \mathbf{x})^T \nabla^2 f(\mathbf{x} + \alpha(\mathbf{y} - \mathbf{x}))(\mathbf{y} - \mathbf{x}) \geq 0, \tag{12}$$

which leads to

$$f(\mathbf{y}) \geq f(\mathbf{x}) + \nabla f(\mathbf{x})^T (\mathbf{y} - \mathbf{x}), \tag{13}$$

which is the first-order condition of convexity. Hence, this proves that $f(\mathbf{x})$ is convex. For the reverse direction, showing that convexity implies positive semi definiteness of the Hessian, again we can use Taylor's theorem. However, there is a slightly delicate issue because of the Hessian. For this, we need to use some properties of symmetric matrices, to claim that if for some orthonormal basis $\{\mathbf{x}_i\}$, a matrix $\mathbf{A}$ satisfies $\mathbf{x}^\top \mathbf{A}\mathbf{x} \geq 0$, then $\mathbf{A}$ is positive semidefinite. □

2.4 Examples of Convex Functions

- ***Exponential*** $f(x) = e^{ax}$, for all $a \in \mathbb{R}$ is convex on $\mathbb{R}$. To show e^{ax} is convex for all $a \in \mathbb{R}$, we could simply see the second derivative of function f, which is $a^2 e^{ax} \geq 0$, for all $a \in \mathbb{R}$
- ***Powers*** $f(x) = x^a$ is convex on $\mathbb{R}_{++}$ when $a \geq 1$ *or* $a \leq 0$, concave otherwise. Since $f''(x) = a(a-1)x^{a-2}, x \in \mathbb{R}_{++}$ is non-negative when $a \geq 1$ *or* $a \leq 0$.
- ***Negative Logarithm*** $f(x) = -logx$ is convex on its domain $\mathbb{R}_{++}$, because $f''(x) = \frac{1}{x^2} > 0$, for all $x \in \mathbb{R}_{++}$
- ***Norms*** Every norms on $\mathbb{R}^n$ is convex. Using triangle inequality and positive homogenity, $f(\lambda x + (1-\lambda)y) \leq f(\lambda x) + f((1-\lambda)y) = \lambda f(x) + (1-\lambda)f(y)$, for all $x, y \in \mathbb{R}^n$, $0 \leq \lambda \leq 1$.

- ***Max Function*** $f(\mathbf{x}) = max\{x_1, x_2, \cdots, x_n\}$ is convex on $\mathbb{R}^n$. To show this, $f(\lambda\mathbf{x} + (1-\lambda\mathbf{y})) = max\{\lambda x_1 + (1-\lambda)y_1, \cdots, \lambda x_n + (1-\lambda)y_n\} = \lambda x_i + (1-\lambda)y_i \leq \lambda x_{max} + (1-\lambda)y_{max} = \lambda f(\mathbf{x}) + (1-\lambda)f(\mathbf{y})$, for all $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$.
- ***Quadratic over linear:*** $f(x, y) = \frac{x^2}{y}$ is convex for $y > 0$ since $\nabla^2 f(x, y) = [\frac{2}{y}, -\frac{2x}{y^2}; -\frac{2x}{y^2}, \frac{2x^2}{y^3}] = \frac{2}{y^3}[y; -x][y; -x]^\top \succeq \mathbf{0}$ for $y > 0$.

2.5 Side results for convex functions

Theorem 4. *Let f be a differentiable convex function. Then, $\mathbf{x}^*$ is a global minimizer of f if and only if $\nabla f(\mathbf{x}^*) = \mathbf{0}$.*

Proof. If $\mathbf{x}^*$ is a global minimizer, then it is a local minimizer. Hence, it must satisfy $\nabla f(\mathbf{x}^*) = \mathbf{0}$ as shown previously.
If $\nabla f(\mathbf{x}^*) = \mathbf{0}$, then by the definition of convexity, for any $\mathbf{y} \in \mathbb{R}^n$ we have

$$f(\mathbf{y}) \geq f(\mathbf{x}^*) + \nabla f(\mathbf{x}^*)^\top (\mathbf{y} - \mathbf{x}^*) = f(\mathbf{x}^*)$$

and hence $f(\mathbf{x}^*)$ is a global minimizer of f. □

Theorem 5. *Let f be a differentiable convex function. Then, for any $\mathbf{x}$ and $\mathbf{y}$ we have*

$$(\nabla f(\mathbf{x}) - \nabla f(\mathbf{y}))^\top (\mathbf{x} - \mathbf{y}) \geq 0.$$

Proof. By the definition of convexity, we have

$$f(\mathbf{y}) \geq f(\mathbf{x}) + \nabla f(\mathbf{x})^\top (\mathbf{y} - \mathbf{x})$$

and

$$f(\mathbf{x}) \geq f(\mathbf{y}) + \nabla f(\mathbf{y})^\top (\mathbf{x} - \mathbf{y})$$

If we add up both sides of the above expressions we obtain

$$0 \geq \nabla f(\mathbf{x})^\top (\mathbf{y} - \mathbf{x}) + \nabla f(\mathbf{y})^\top (\mathbf{x} - \mathbf{y})$$

and be regrouping the terms our claim follows. □

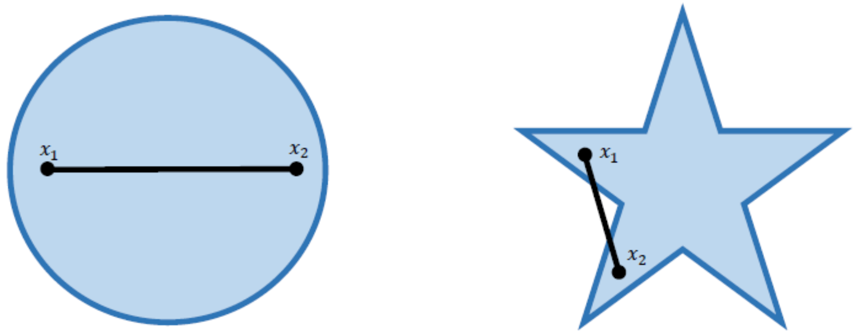

Figure 3: A convex set can be easily determined by examining whether the line segment between any two points in the set are in the set. Thus the figure on the left (circle) is in the set whereas the figure on the right (star) is not.

3 Convex Sets

Before we are able to formulate a convex optimization problem, we must understand what constitutes a convex set. We will begin by providing the general definition of a convex set and look at many different sets that are convex.

3.1 Definition of a Convex Set

Definition 6. *A* ***convex combination*** *of points* $\mathbf{x}_1, \ldots, \mathbf{x}_k$ *is described by*

$$\sum_{i=0}^{k} \theta_i \mathbf{x}_i, \tag{14}$$

where $\theta_1 + \cdots + \theta_k = 1$ *and* $\theta_i \geq 0$.

Definition 7. *A set,* $\mathcal{X}$*, is called a* ***convex set*** *if and only if the convex combination of any two points in the set belongs to the set, i.e.*

$$\mathcal{X} \subseteq \mathbb{R}^n \textit{ is convex } \Leftrightarrow \textit{ if } \text{for all } \mathbf{x}_1, \mathbf{x}_2 \in \mathcal{X} \textit{ and } \text{for all } \lambda \in [0,1],\ \lambda \mathbf{x}_1 + (1-\lambda)\mathbf{x}_2 \in \mathcal{X}\ . \tag{15}$$

3.2 Examples of Convex Sets

Definition 8. *An* ***affine combination*** *of points* $\mathbf{x}_1, \ldots, \mathbf{x}_k$ *is described by*

$$\sum_{i=0}^{k} \theta_i \mathbf{x}_i, \tag{16}$$

where $\theta_1 + \cdots + \theta_k = 1$*. (Note: Affine combination lacks the nonnegative constraint on* θ_i*).*

Definition 9. *A set,* $\mathcal{A}$*, is called an* ***affine set*** *if and only if the affine combination of any two points in the set belongs to the set. In symbols:*

$$\mathcal{A} \subseteq \mathbb{R}^n \textit{ is affine } \Leftrightarrow \textit{ if } \text{for all } \mathbf{x}_1, \mathbf{x}_2 \in \mathcal{A},\ \lambda \mathbf{x}_1 + (1-\lambda)\mathbf{x}_2 \in \mathcal{A}\ . \tag{17a}$$

An equivalent definition, using the solution set of a system of linear equations is

$$\mathcal{A} \subseteq \mathbb{R}^n \textit{ is affine } \Leftrightarrow \mathcal{A} = \left\{x \mid \mathbf{A}\mathbf{x} = \mathbf{b}, \mathbf{A} \in \mathbb{R}^{m\times n}, \mathbf{b} \in \mathbb{R}^n\right\}. \tag{17b}$$

Remark 4. *Affine sets are convex.*

Proof. Let $\mathbf{x}_1, \mathbf{x}_2 \in \mathcal{A}$ and $\lambda \in [0, 1]$. Taking the convex combination of $\mathbf{x}_1$ and $\mathbf{x}_2$,

$$\begin{aligned} \mathbf{A}(\lambda\mathbf{x}_1 + (1-\lambda)\mathbf{x}_2) &= \lambda\mathbf{A}\mathbf{x}_1 + (1-\lambda)\mathbf{A}\mathbf{x}_2 \\ &= \lambda\mathbf{b} + (1-\lambda)\mathbf{b} \\ &= \mathbf{b} \end{aligned}$$

Since the convex combination of points in $\mathcal{A}$ is also in the set, $\mathcal{A}$ is a convex set. □

Definition 10. *A* ***hyperplane****, $\mathcal{H}$, is a set defined by*

$$\mathcal{H} = \{\mathbf{x} \in \mathbb{R}^n \mid \mathbf{s}^T\mathbf{x} = \mathbf{b}\}(\mathbf{s} \neq \emptyset), \tag{18}$$

where b is the offset and s is the normal vector.

Definition 11. *A* ***halfspace****, $\mathcal{H}_+$, is a set defined as*

$$\mathcal{H}_+ = \{\mathbf{x} \in \mathbb{R}^n \mid \mathbf{s}^T\mathbf{x} \leq \mathbf{b}\}(\mathbf{s} \neq \emptyset), \tag{19}$$

where b is the offset and s is the normal vector.

Definition 12. *A* ***polyhedron*** *is an intersection of finite number of halfspaces and hyperplanes. A* ***polytope*** *is a polyhedron that is also bounded.*

Definition 13. *$\mathbf{S}^n$ is the set of symmetric $n \times n$ matrices. $\mathbf{S}^n_+ = \{\mathbf{X} \in \mathbf{S}^n \mid \mathbf{X} \succeq \mathbf{0}\}$ is the set of positive semi-definite $n \times n$ matrices. $\mathbf{S}^n_{++} = \{\mathbf{X} \in \mathbf{S}^n \mid \mathbf{X} \succ \mathbf{0}\}$ is the set of positive definite $n \times n$ matrices.*

A very interesting set that turns out to be useful for many applications, and in general comes up frequently in many applications of convex optimization which we consider, is the set $\mathbf{S}^n_+$, of symmetric $n \times n$ matrices with non-negative eigenvalues. As a simple exercise, we can show that this set is in fact convex.

Proof. We use the fact that a symmetric matrix $\mathbf{M}$ is in $\mathbf{S}^n_+$ iff $\mathbf{x}^\top\mathbf{M}\mathbf{x} \geq 0$ for all $\mathbf{x} \in \mathbb{R}^n$. Using this, checking convexity becomes straightforward. For a convex combination of any two matrices $\mathbf{M}_1, \mathbf{M}_2 \in \mathcal{S}^n_+$ we have

$$\begin{aligned} \mathbf{x}^\top(\lambda\mathbf{M}_1 + (1-\lambda)\mathbf{M}_2)\mathbf{x} &= \lambda\mathbf{x}^\top\mathbf{M}_1\mathbf{x} + (1-\lambda)\mathbf{x}^\top\mathbf{M}_2\mathbf{x} \\ &\geq 0 \end{aligned}$$

Hence, $\lambda\mathbf{M}_1 + (1-\lambda)\mathbf{M}_2 \in \mathcal{S}^n_+$ for all $\lambda \in [0, 1]$. □

Definition 14. ***(Euclidean) ball*** *with center $\mathbf{x}_c$ and radius r is*

$$B(\mathbf{x}_c, r) = \{\mathbf{x} \mid \|\mathbf{x} - \mathbf{x}_c\|_2 \leq r\}$$

Indeed, a Euclidean ball is a convex set.

Definition 15. ***Ellipsoid*** *with center* $\mathbf{x}_c$ *is defined as*

$$\{\mathbf{x} \mid (\mathbf{x}-\mathbf{x}_c)^\top \mathbf{P}(\mathbf{x}-\mathbf{x}_c) \leq 1\}$$

with $\mathbf{P} \in \mathbf{S}^n_{++}$, *(*$\mathbf{P}$ *is symmetric positive definite).*

It is easy to verify that an ellipsoid is a convex set.

Definition 16. ***Norm:*** *a function* $\|.\|$*that satisfies the following properties:*

- $\|\mathbf{x}\| \geq 0$*;* $\|\mathbf{x}\| = 0$ *if and only if* $\mathbf{x} = \mathbf{0}$*.*
- $t\|\mathbf{x}\| = |t|\|\mathbf{x}\|$ *for* $t \in \mathbb{R}$
- $\|\mathbf{x}+\mathbf{y}\| \leq \|\mathbf{x}\| + \|\mathbf{y}\|$

Definition 17. ***Norm ball*** *with center* $\mathbf{x}_c$ *and radius* r*:* $\{\mathbf{x} \mid \|\mathbf{x}-\mathbf{x}_c\| \leq r\}$

Norm ball is a convex set!

Definition 18. *The* ***Norm Cone*** *associated with the norm* $\|.\|$ *is the set* $\{(\mathbf{x}, t) \mid \|\mathbf{x}\| \leq t\} \in \mathbb{R}^{n+1}$.

Norm cone is a convex cone. (it is a cone and also it is a convex set.)

Example: *The second-order cone* is the norm cone for the Euclidean norm, i.e.,

$$\begin{aligned} \mathcal{C} &= \{(\mathbf{x}, t) \mid \|\mathbf{x}\|_2 \leq t\} \in \mathbb{R}^{n+1} \\ &= \left\{ \begin{bmatrix} \mathbf{x} \\ t \end{bmatrix} \;\middle|\; \begin{bmatrix} \mathbf{x} \\ t \end{bmatrix}^\top \begin{bmatrix} \mathbf{I} & 0 \\ 0 & -1 \end{bmatrix} \begin{bmatrix} \mathbf{x} \\ t \end{bmatrix} \leq 0,\ t \geq 0 \right\} \end{aligned}$$

Definition 19. *A convex hull of a set* $\mathcal{C}$ *is the set of all convex combinations of points in* $\mathcal{C}$*. As the name implies, convex hulls are convex. See Fig. 4 for an illustration.*

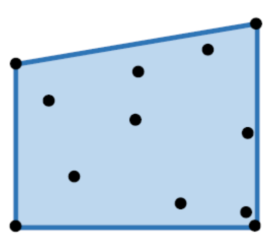

Figure 4: Convex hull