

## TOPIC 3

# Optimality, convexity, and gradient descent

## 3 Convexity

---

Course: [Math 535 \(http://www.math.wisc.edu/~roch/mmids/\)](http://www.math.wisc.edu/~roch/mmids/) - Mathematical Methods in Data Science (MMiDS)

Author: [Sebastien Roch \(http://www.math.wisc.edu/~roch/\)](http://www.math.wisc.edu/~roch/), Department of Mathematics, University of Wisconsin-Madison

Updated: Oct 10, 2020

Copyright: © 2020 Sebastien Roch

---

Our optimality conditions have only concerned local minimizers. Indeed, in the absence of global structure, local information such as gradients and Hessians can only inform us about the immediate neighborhood of points. Here we consider convexity, under which local minimizers are also global minimizers.

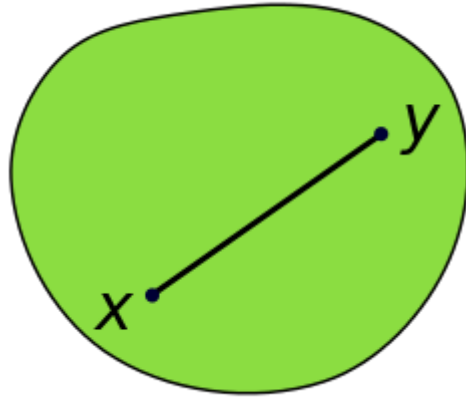
### 3.1 Convex sets

We start with convex sets.

**Definition (Convex Set):** A set  $D \subseteq \mathbb{R}^d$  is convex if for all  $\mathbf{x}, \mathbf{y} \in D$  and all  $\alpha \in [0, 1]$   
 $(1 - \alpha)\mathbf{x} + \alpha\mathbf{y} \in D$ .

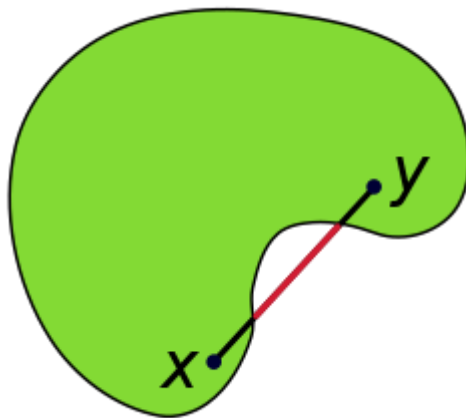
<

The following is a convex set:



(Source ([https://commons.wikimedia.org/wiki/File:Convex\\_polygon\\_illustration1.svg](https://commons.wikimedia.org/wiki/File:Convex_polygon_illustration1.svg)))

The following is not a convex set:



(Source ([https://commons.wikimedia.org/wiki/File:Convex\\_polygon\\_illustration2.svg](https://commons.wikimedia.org/wiki/File:Convex_polygon_illustration2.svg)))

**Example:** Any open ball in  $\mathbb{R}^d$  is convex. Indeed, let  $\delta > 0$  and  $\mathbf{x}_0$ . For any  $\mathbf{x}, \mathbf{y} \in B_\delta(\mathbf{x}_0)$  and any  $\alpha \in [0, 1]$ , we have

$$\begin{aligned} \|(1 - \alpha)\mathbf{x} + \alpha\mathbf{y} - \mathbf{x}_0\| &= \|(1 - \alpha)(\mathbf{x} - \mathbf{x}_0) + \alpha(\mathbf{y} - \mathbf{x}_0)\| \\ &\leq \|(1 - \alpha)(\mathbf{x} - \mathbf{x}_0)\| + \|\alpha(\mathbf{y} - \mathbf{x}_0)\| \\ &= (1 - \alpha)\|\mathbf{x} - \mathbf{x}_0\| + \alpha\|\mathbf{y} - \mathbf{x}_0\| \\ &\leq (1 - \alpha)\delta + \alpha\delta \\ &= \delta \end{aligned}$$

where we used the triangle inequality on the second line. Hence we have established that  $(1 - \alpha)\mathbf{x} + \alpha\mathbf{y} \in B_\delta(\mathbf{x}_0)$ .  $\triangleleft$

*Exercise:* A convex combination of  $\mathbf{z}_1, \dots, \mathbf{z}_m \in \mathbb{R}^d$  is a linear combination of the form

$$\mathbf{w} = \sum_{i=1}^m \alpha_i \mathbf{z}_i$$

where  $\alpha_i \geq 0$  for all  $i$  and  $\sum_{i=1}^m \alpha_i = 1$ . Show that a set is convex if and only if it contains all convex combinations of its elements. [Hint: Use induction on  $m$ .]  $\triangleleft$

*Exercise:* A set  $C \subseteq \mathbb{R}^d$  is a cone if, for all  $\mathbf{x} \in C$  and all  $\alpha \geq 0$ ,  $\alpha\mathbf{x} \in C$ . A conic combination of  $\mathbf{z}_1, \dots, \mathbf{z}_m \in \mathbb{R}^d$  is a linear combination of the form

$$\mathbf{w} = \sum_{i=1}^m \alpha_i \mathbf{z}_i$$

where  $\alpha_i \geq 0$  for all  $i$ . Show that a set  $C$  is a convex cone if and only if it contains all conic combinations of its elements.  $\triangleleft$

*Exercise:* Show that any linear subspace of  $\mathbb{R}^d$  is convex.  $\triangleleft$

*Exercise:* Show that the set

$$\mathbb{R}_+^d = \{\mathbf{x} \in \mathbb{R}^d : \mathbf{x} \geq \mathbf{0}\}$$

is convex.  $\triangleleft$

**Example:** Here is an important generalization of the previous exercise. Think of the space of  $n \times n$  symmetric matrices as a linear subspace of  $\mathbb{R}^{n^2}$

$$\mathbf{S}^n = \{X \in \mathbb{R}^{n \times n} : X = X^T\}.$$

The dimension of  $\mathbf{S}^n$  is  $n(n-1)/2$ , the number of free parameters under the symmetry assumption. Consider now the set of all positive semidefinite matrices in  $\mathbf{S}^n$

$$\mathbf{S}_+^n = \{X \in \mathbf{S}^n : X \succeq 0\}.$$

Note that  $\mathbf{S}_+^n$  is not the same as the set of symmetric matrices with nonnegative elements. Here is a counter-example:

```
In [1]: # Julia version: 1.5.1
        using LinearAlgebra
```

```
In [2]: A = [0. 1. 0. 1.; 1. 0. 1. 0.; 0. 1. 0. 1.; 1. 0. 1. 0.]
```

```
Out[2]: 4×4 Array{Float64,2}:
 0.0  1.0  0.0  1.0
 1.0  0.0  1.0  0.0
 0.0  1.0  0.0  1.0
 1.0  0.0  1.0  0.0
```

```
In [3]: F = eigen(A)
        F.values
```

```
Out[3]: 4-element Array{Float64,1}:
 -1.9999999999999997
  0.0
  1.7763568394002505e-15
  2.0
```

We claim that the set  $\mathbf{S}_+^n$  is convex. Indeed let  $X, Y \in \mathbf{S}_+^n$  and let  $\alpha \in [0, 1]$ . Then by positive semidefiniteness of  $X$  and  $Y$ , for any  $\mathbf{v} \in \mathbb{R}^n$

$$\langle \mathbf{v}, [(1-\alpha)X + \alpha Y]\mathbf{v} \rangle = (1-\alpha)\langle \mathbf{v}, X\mathbf{v} \rangle + \alpha\langle \mathbf{v}, Y\mathbf{v} \rangle \geq 0.$$

This shows that  $(1-\alpha)X + \alpha Y \succeq 0$  and hence that  $\mathbf{S}_+^n$  is convex.  $\triangleleft$

The following exercise explains the choice of  $A$  in the previous example.

*Exercise:* A graph  $G = (V, E)$  is bipartite if there is a bipartition of the vertices  $V_1, V_2$  (i.e.  $V_1 \cap V_2 = \emptyset$  and  $V_1 \cup V_2 = V$ ) such that all edges are between  $V_1$  and  $V_2$ , that is, for any edge  $e = \{u, v\} \in E$  we have that  $e \cap V_1 \neq \emptyset$  and  $e \cap V_2 \neq \emptyset$ . A graph is  $\delta$ -regular if all its vertices have degree  $\delta$ . Show that if  $G$  is a  $\delta$ -regular, bipartite graph, then its adjacency matrix has an eigenvalue  $-\delta$ . [Hint: Try a vector that takes different values on  $V_1$  and  $V_2$ .]  $\triangleleft$

*Exercise:* Show that  $\mathbf{S}_+^n$  is a convex cone.  $\triangleleft$

A number of operations preserve convexity.

---

**Lemma (Operations that Preserve Convexity):** Let  $S_1$  and  $S_2$  be convex sets in  $\mathbb{R}^d$ . Let  $\beta \in \mathbb{R}$  and  $\mathbf{b} \in \mathbb{R}^d$ . The following sets are also convex:

(a)  $\beta S_1 = \{\beta \mathbf{x} : \mathbf{x} \in S_1\}$

(b)  $S_1 + \mathbf{b} = \{\mathbf{x} + \mathbf{b} : \mathbf{x} \in S_1\}$

(c)  $S_1 + S_2 = \{\mathbf{x}_1 + \mathbf{x}_2 : \mathbf{x}_1 \in S_1 \text{ and } \mathbf{x}_2 \in S_2\}$

(d)  $S_1 \times S_2 = \{(\mathbf{x}_1, \mathbf{x}_2) : \mathbf{x}_1 \in S_1 \text{ and } \mathbf{x}_2 \in S_2\}$

(e)  $T = \{\mathbf{x}_1 : (\mathbf{x}_1, \mathbf{x}_2) \in S_1 \times \mathbb{R}^d\}$

(f)  $S_1 \cap S_2 = \{\mathbf{x} : \mathbf{x} \in S_1 \text{ and } \mathbf{x} \in S_2\}$

---

*Proof:* We only prove (f). The other statements are left as an exercise. Suppose  $\mathbf{x}, \mathbf{y} \in S_1 \cap S_2$  and  $\alpha \in [0, 1]$ . Then, by the convexity of  $S_1$ ,  $(1 - \alpha)\mathbf{x} + \alpha\mathbf{y} \in S_1$  and, by the convexity of  $S_2$ ,  $(1 - \alpha)\mathbf{x} + \alpha\mathbf{y} \in S_2$ . Hence  $(1 - \alpha)\mathbf{x} + \alpha\mathbf{y} \in S_1 \cap S_2$ .

□

*Exercise:* Prove (a)-(e) from the *Operations that Preserve Convexity Lemma*. ◀

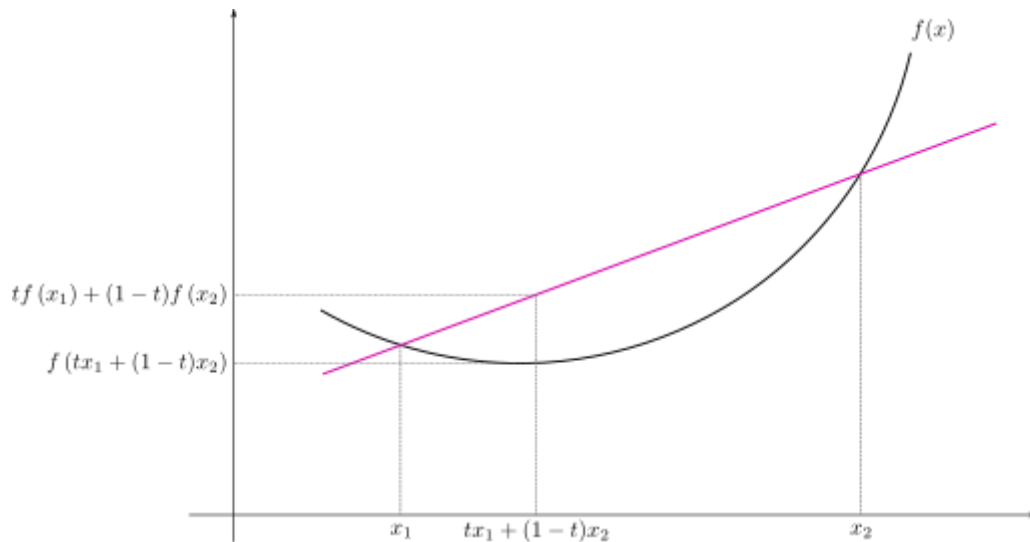
## 3.2 Convex functions

Our main interest is in convex functions.

**Definition (Convex Function):** A function  $f : \mathbb{R}^d \rightarrow \mathbb{R}$  is convex if, for all  $\mathbf{x}, \mathbf{y} \in \mathbb{R}^d$  and all  $\alpha \in [0, 1]$

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

More generally, a function  $f : D \rightarrow \mathbb{R}$  over a convex domains  $D \subseteq \mathbb{R}^d$  is convex if the definition above holds over all  $\mathbf{x}, \mathbf{y} \in D$ . ◀



(Source (<https://commons.wikimedia.org/wiki/File:ConvexFunction.svg>))

*Exercise:* Let  $f : \mathbb{R}^d \rightarrow \mathbb{R}$  be a function. The epigraph of  $f$  is the set  

$$\mathbf{epi} f = \{(\mathbf{x}, y) : \mathbf{x} \in \mathbb{R}^d, y \geq f(\mathbf{x})\}.$$

Show that  $f$  is convex if and only if  $\mathbf{epi} f$  is a convex set. ◁

**Lemma (Affine Functions are Convex):** Let  $\mathbf{w} \in \mathbb{R}^d$  and  $b \in \mathbb{R}$ . The function  $f(\mathbf{x}) = \mathbf{w}^T \mathbf{x} + b$  is convex.

*Proof:* For any  $\mathbf{x}, \mathbf{y} \in \mathbb{R}^d$  and  $\alpha \in [0, 1]$ ,

$$f((1 - \alpha)\mathbf{x} + \alpha\mathbf{y}) = \mathbf{w}^T [(1 - \alpha)\mathbf{x} + \alpha\mathbf{y}] + b = (1 - \alpha)[\mathbf{w}^T \mathbf{x} + b] + \alpha[\mathbf{w}^T \mathbf{y} + b]$$

which proves the claim. ◻

Exercise: Let  $f_1, \dots, f_m : \mathbb{R}^d \rightarrow \mathbb{R}$  be convex functions and let  $\beta_1, \dots, \beta_m \geq 0$ . Show that

$$f(\mathbf{x}) = \sum_{i=1}^m \beta_i f_i(\mathbf{x})$$

is convex.  $\triangleleft$

Exercise: Let  $f_1, f_2 : \mathbb{R}^d \rightarrow \mathbb{R}$  be convex functions. Show that the pointwise maximum function

$$f(\mathbf{x}) = \max\{f_1(\mathbf{x}), f_2(\mathbf{x})\}$$

is convex. [Hint: First show that  $\max\{\alpha + \beta, \eta + \phi\} \leq \max\{\alpha, \eta\} + \max\{\beta, \phi\}$ .]  $\triangleleft$

Exercise: Prove the following composition theorem: if  $f : \mathbb{R}^d \rightarrow \mathbb{R}$  is convex and  $g : \mathbb{R} \rightarrow \mathbb{R}$  is convex and nondecreasing, then the composition  $h = g \circ f$  is convex.  $\triangleleft$

A common way to prove that a function is convex is to look at its Hessian. We start with a first-order condition that will prove useful.

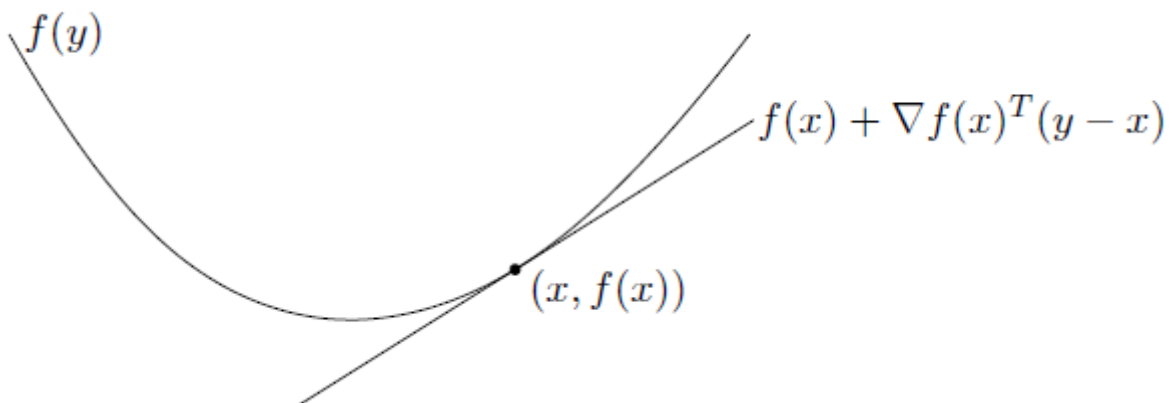
---

**Lemma (First-Order Convexity Condition):** Let  $f : \mathbb{R}^d \rightarrow \mathbb{R}$  be continuously differentiable. Then  $f$  is convex if and only if for all  $\mathbf{x}, \mathbf{y} \in \mathbb{R}^d$

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

---

The condition is illustrated below.



(Source (<https://math.stackexchange.com/questions/1518550/geometric-interpretation-of-first-order-condition>))

*Proof (First-Order Convexity Condition):* Suppose first that  $f(\mathbf{z}_2) \geq f(\mathbf{z}_1) + \nabla f(\mathbf{z}_1)^T(\mathbf{z}_2 - \mathbf{z}_1)$  for all  $\mathbf{z}_1, \mathbf{z}_2$ . For any  $\mathbf{x}, \mathbf{y}$  and  $\alpha \in [0, 1]$ , let  $\mathbf{w} = (1 - \alpha)\mathbf{x} + \alpha\mathbf{y}$ . Then taking  $\mathbf{z}_1 = \mathbf{w}$  and  $\mathbf{z}_2 = \mathbf{x}$  gives

$$f(\mathbf{x}) \geq f(\mathbf{w}) + \nabla f(\mathbf{w})^T(\mathbf{x} - \mathbf{w})$$

and taking  $\mathbf{z}_1 = \mathbf{w}$  and  $\mathbf{z}_2 = \mathbf{y}$  gives

$$f(\mathbf{y}) \geq f(\mathbf{w}) + \nabla f(\mathbf{w})^T(\mathbf{y} - \mathbf{w}).$$

Multiplying the first inequality by  $(1 - \alpha)$  and the second one by  $\alpha$ , and adding them up gives

$$(1 - \alpha)f(\mathbf{x}) + \alpha f(\mathbf{y}) \geq f(\mathbf{w}) + \nabla f(\mathbf{w})^T[(1 - \alpha)\mathbf{x} + \alpha\mathbf{y}] - \mathbf{w}) = f(\mathbf{w})$$

proving convexity.

For the other direction, assume that  $f$  is convex. For any  $\mathbf{x}, \mathbf{y}$  and  $\alpha \in (0, 1)$ , by the *Multivariate Mean Value Theorem*, for some  $\xi \in (0, 1)$  it holds that

$$f(\mathbf{w}) = f(\mathbf{x} + \alpha(\mathbf{y} - \mathbf{x})) = f(\mathbf{x}) + \alpha(\mathbf{y} - \mathbf{x})^T \nabla f(\mathbf{x} + \xi\alpha(\mathbf{y} - \mathbf{x}))$$

while convexity implies

$$f(\mathbf{w}) \leq (1 - \alpha)f(\mathbf{x}) + \alpha f(\mathbf{y}).$$

Combining, rearranging and dividing by  $\alpha$  gives

$$(\mathbf{y} - \mathbf{x})^T \nabla f(\mathbf{x} + \xi\alpha(\mathbf{y} - \mathbf{x})) \leq f(\mathbf{y}) - f(\mathbf{x}).$$

Taking  $\alpha \rightarrow 0$  gives the claim.  $\square$

**Lemma (Second-Order Convexity Condition):** Let  $f : \mathbb{R}^d \rightarrow \mathbb{R}$  be twice continuously differentiable. Then  $f$  is convex if and only if, for all  $\mathbf{x} \in \mathbb{R}^d$ ,  $\mathbf{H}_f(\mathbf{x})$  is positive semidefinite.

*Proof:* Suppose first that  $\mathbf{H}_f(\mathbf{z}_1) \succeq 0$  for all  $\mathbf{z}_1$ . For any  $\mathbf{x}, \mathbf{y}$ , by the *Multivariate Taylor's Theorem*, there is  $\xi \in (0, 1)$  such that

$$\begin{aligned} f(\mathbf{y}) &= f(\mathbf{x}) + \nabla f(\mathbf{x})^T(\mathbf{y} - \mathbf{x}) + (\mathbf{y} - \mathbf{x})^T \mathbf{H}_f(\mathbf{x} + \xi(\mathbf{y} - \mathbf{x}))(\mathbf{y} - \mathbf{x}) \\ &\geq f(\mathbf{x}) + \nabla f(\mathbf{x})^T(\mathbf{y} - \mathbf{x}) \end{aligned}$$

where we used the positive semidefiniteness of the Hessian. By the *First-Order Convexity Condition*, it implies that  $f$  is convex.



For the other direction, assume that  $f$  is convex. For any  $\mathbf{x}$ ,  $\mathbf{w}$  and  $\alpha \in (0, 1)$ , by the *Multivariate Taylor's Theorem* again, for some  $\xi_\alpha \in (0, 1)$  it holds that

$$f(\mathbf{x} + \alpha\mathbf{w}) = f(\mathbf{x}) + \alpha\mathbf{w}^T \nabla f(\mathbf{x}) + \alpha^2 \mathbf{w}^T \mathbf{H}_f(\mathbf{x} + \xi_\alpha \alpha\mathbf{w}) \mathbf{w}$$

while the *First-Order Convexity Condition* implies

$$f(\mathbf{x} + \alpha\mathbf{w}) \geq f(\mathbf{x}) + \alpha\mathbf{w}^T \nabla f(\mathbf{x}).$$

Combining, rearranging and dividing by  $\alpha^2$  gives

$$\mathbf{w}^T \mathbf{H}_f(\mathbf{x} + \xi_\alpha \alpha\mathbf{w}) \mathbf{w} \geq 0.$$

Taking  $\alpha \rightarrow 0$  shows that  $\mathbf{w}^T \mathbf{H}_f(\mathbf{x}) \mathbf{w} \geq 0$ . Since  $\mathbf{w}$  is arbitrary, this implies that the Hessian is positive semidefinite at  $\mathbf{x}$ . This holds for any  $\mathbf{x}$ , which proves the claim.  $\square$

*Exercise:* Show that  $f(x) = e^{\beta x}$ , where  $x, \beta \in \mathbb{R}$ , is convex.  $\triangleleft$

**Example:** Consider the quadratic function

$$f(\mathbf{x}) = \frac{1}{2} \mathbf{x}^T P \mathbf{x} + \mathbf{q}^T \mathbf{x} + r.$$

We showed previously that the Hessian is

$$\mathbf{H}_f(\mathbf{x}) = \frac{1}{2} [P + P^T].$$

So  $f$  is convex if and only if the matrix  $\frac{1}{2} [P + P^T]$  has only nonnegative eigenvalues.  $\triangleleft$

### 3.3 Convexity and unconstrained optimization

A key property of convex functions is the following.

---

**Theorem (Global Minimizers of Convex Functions):** Let  $f : \mathbb{R}^d \rightarrow \mathbb{R}$  be a convex function. Then any local minimizer of  $f$  is also a global minimizer.

---

*Proof:* By contradiction, suppose  $\mathbf{x}_0$  is a local minimizer, but not a global minimizer. Then there is  $\mathbf{y}$  such that

$$f(\mathbf{y}) < f(\mathbf{x}_0).$$

By convexity, for any  $\alpha \in (0, 1)$

$$f(\mathbf{x}_0 + \alpha(\mathbf{y} - \mathbf{x}_0)) \leq (1 - \alpha)f(\mathbf{x}_0) + \alpha f(\mathbf{y}) < f(\mathbf{x}_0).$$

But that implies that every open ball around  $\mathbf{x}_0$  contains a point taking a smaller value than  $f(\mathbf{x}_0)$ , a contradiction.  $\square$

In the continuously differentiable case, we get:

---

**Theorem (First-Order Sufficient Condition for Convex Functions):** Let  $f : \mathbb{R}^d \rightarrow \mathbb{R}$  be a continuously differentiable, convex function. If  $\nabla f(\mathbf{x}_0) = \mathbf{0}$  then  $\mathbf{x}_0$  is a global minimizer.

---

*Proof:* Assume  $\nabla f(\mathbf{x}_0) = \mathbf{0}$ . By the *First-Order Convexity Condition*, for any  $\mathbf{y}$

$$f(\mathbf{y}) - f(\mathbf{x}_0) \geq \nabla f(\mathbf{x}_0)^T (\mathbf{y} - \mathbf{x}_0) = 0.$$

That proves the claim.  $\square$

**Example:** Consider the quadratic function

$$f(\mathbf{x}) = \frac{1}{2} \mathbf{x}^T P \mathbf{x} + \mathbf{q}^T \mathbf{x} + r$$

where  $P$  is symmetric and positive definite. The Hessian is then

$$\mathbf{H}_f(\mathbf{x}) = \frac{1}{2} [P + P^T] = P$$

for any  $\mathbf{x}$ . So  $f$  is convex. Further the gradient is

$$\nabla f(\mathbf{x}) = P \mathbf{x} + \mathbf{q}$$

for all  $\mathbf{x}$ .

Any  $\mathbf{x}$  satisfying

$$P \mathbf{x} + \mathbf{q} = \mathbf{0}$$

is a global minimizer. If  $P = Q \Lambda Q^T$  is a spectral decomposition of  $P$ , where all diagonal entries of  $\Lambda$  are strictly positive, then  $P^{-1} = Q \Lambda^{-1} Q^T$  where the diagonal entries of  $\Lambda^{-1}$  are the inverses of those of  $\Lambda$ . That can be seen by checking that

$$Q \Lambda Q^T Q \Lambda^{-1} Q^T = Q I_{d \times d} Q^T = I_{d \times d}.$$

So the following is a global minimizer

$$\mathbf{x}^* = -Q \Lambda^{-1} Q^T \mathbf{q}.$$

$\triangleleft$