6FMAI19 Nonlinear Optimization Spring, 2022

Lecture 2 — 26/1, 2022

Lecturer: Yura Malitsky *Scribe: Max Johansson*

1 Convex Sets

If a set $X \subset \mathbb{R}^n$ is convex, then for any two points in the set, all points on the straight line between x and y will also be in X. Examples of convex sets are intervals on $\mathbb{R}$, and the interior of a circle in $\mathbb{R}^2$. Pictorially, sets can be represented as shapes as shown in Fig. 1, where it is clear that X is convex, while X' is not.

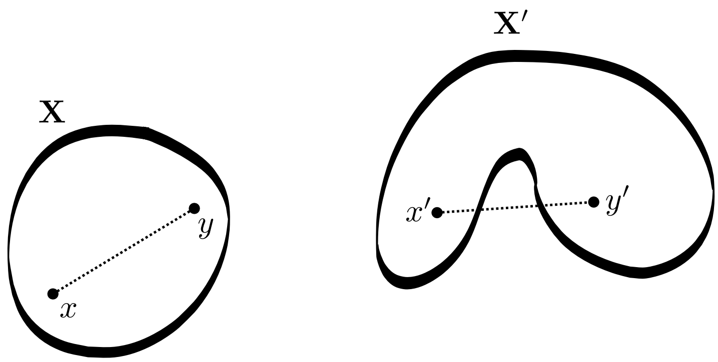

Figure 1: Convex and non-convex sets.

The formal definition of a convex set follows.

Definition 1 (Convex Set). *A set $X \subset \mathbb{R}^n$ is convex if for all $x, y \in X$, $\alpha \in [0,1]$, it follows that*

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

Fig. 2 shows the intersection of three convex sets X_1, X_2 and X_3. As one might expect, the intersection of X_1, X_2 and X_3 is also a convex set.

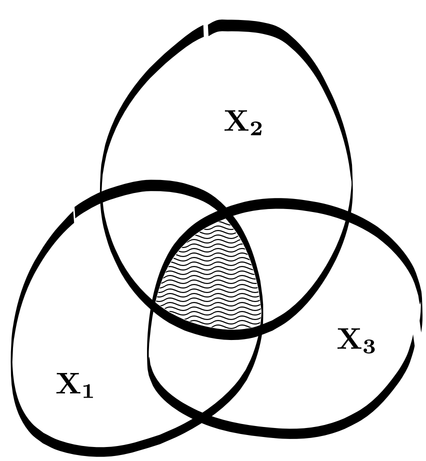

Figure 2: Intersection of three convex sets.

Proposition 2 (Intersection of convex sets)**.** *Let $X_1, X_2, \ldots$ be convex sets, and*

$$X = \bigcap_{i=1}^{\infty} X_i.$$

Then X is convex.

Proof. Assume X is non-convex. Then there exists $x, y \in X$ such that for some $\alpha \in [0, 1]$, $\alpha x + (1-\alpha)y \notin X$. It follows that for at least one X_i, $\alpha x + (1-\alpha)y \notin X_i$. However, $x, y \in X_i$, and therefore X_i is non-convex, a contradiction. □

2 Convex Functions

The convexity of functions is a property that resembles convexity of sets. If a function $f : \mathbb{R} \to \mathbb{R}$ is convex, then we can pick any two points $x, y \in \mathbb{R}$, and the straight line between $f(x)$ and $f(y)$ will lie above the curve of the function. Two functions are shown in Fig. 3. Clearly, f is convex while g is not.

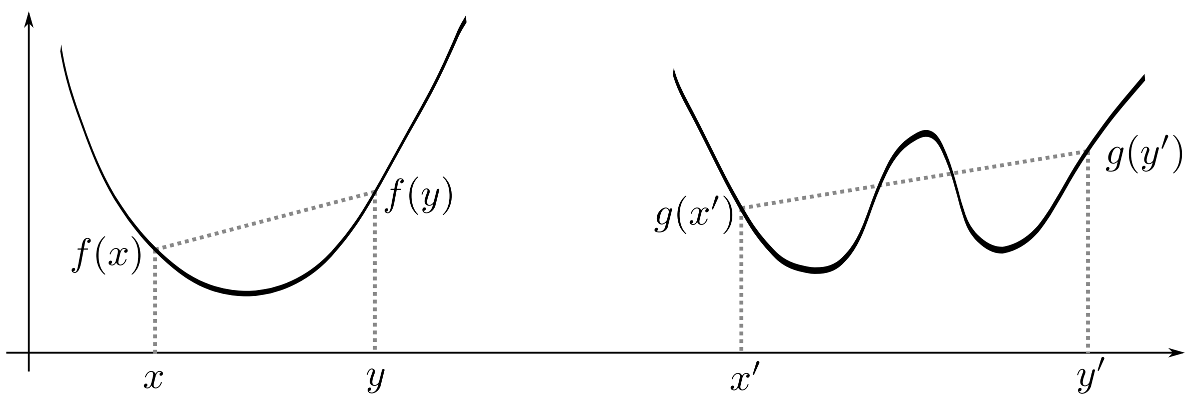

Figure 3: A convex and non-convex function.

A convex function only makes sense if its domain is a convex set. This property is included in its definition.

Definition 3 (Convex function)**.** *Let $X \subset \mathbb{R}^n$ be a convex set. The function $f : X \to \mathbb{R}$ is convex if for all $x, y \in X$, $\alpha \in [0, 1]$, it follows that*

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

Note that, in Fig. 3, the left hand side of the inequality in Def. 3 corresponds to the straight line between $f(x)$ and $f(y)$, while the right hand side is the function applied to the interval $[x, y]$.

The epigraph of a function f, denoted by epi f, is the set of ordered pairs (x, y) such that $y \geq f(x)$, $x \in X$, $y \in \mathbb{R}$. If X is the domain of f, then epi $f \subset X \times \mathbb{R}$. The epigraph of an example function $f : [x_1, x_2] \to \mathbb{R}$ is shown in Fig. 4. It is clear that neither the function f nor the set epi f is convex. It turns out that these properties are equivalent.

Proposition 4 (Equivalence of convex functions and epigraphs)**.** *Let $X \subset \mathbb{R}^n$ be a convex set. $f : X \to \mathbb{R}$ is a convex function if and only if the epigraph of f is a convex set.*

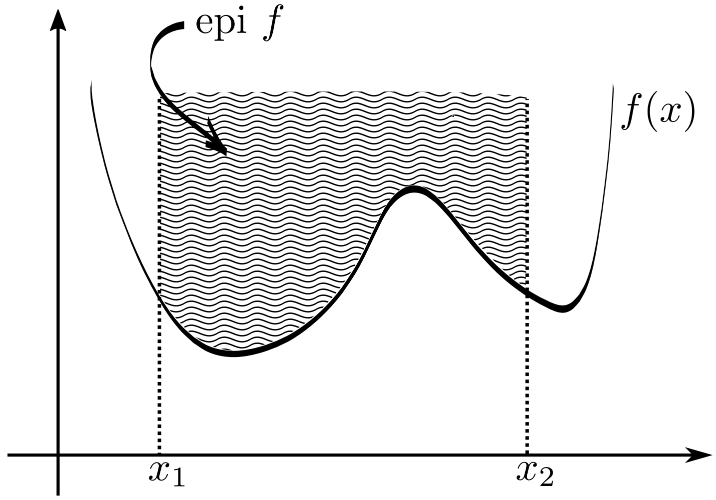

Figure 4: The epigraph of a function.

Proof. Let $x, y \in X,\ \alpha \in [0,1]$.

$$\begin{aligned}
&\text{epi } f \text{ is a convex set} \Leftrightarrow\\
&\alpha(x, f(x)) + (1-\alpha)(y, f(y)) = (\alpha x + (1-\alpha)y,\ \alpha f(x) + (1-\alpha)f(y)) \subset \text{epi } f \Leftrightarrow\\
&\alpha f(x) + (1-\alpha)f(y) \geq f(\alpha x + (1-\alpha)y) \Leftrightarrow\\
&f \text{ is a convex function}
\end{aligned}$$

□

There are certain combinations of convex functions that are also convex. Addition of convex functions has a simple proof.

Proposition 5 (Addition of convex functions)**.** *If $f_1, f_2, \ldots f_n$ are convex functions on some common convex domain X. Then $f : X \to \mathbb{R}$ given by $f(x) = f_1(x) + f_2(x) + \cdots + f_n(x)$ is convex.*

Proof. If $f_1, f_2 \ldots f_n$ are convex functions, then for all $x, y \in X,\ \alpha \in [0,1]$ we have

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

It follows that

$$\begin{gathered}
\sum_{i=1}^{n}[\alpha f_i(x) + (1-\alpha)f_i(y)] \geq \sum_{i=1}^{n} f_i(\alpha x + (1-\alpha)y) \implies\\
\alpha \sum_{i=1}^{n} f_i(x) + (1-\alpha)\sum_{i=1}^{n} f_i(y) \geq \sum_{i=1}^{n} f_i(\alpha x + (1-\alpha)y) \implies\\
\alpha f(x) + (1-\alpha)f(y) \geq f(\alpha x + (1-\alpha)y)
\end{gathered}$$

□

Another case is the supremum of several convex functions.

Proposition 6 (Supremum of convex functions)**.** *If $f_1, f_2 \ldots f_n$ are convex functions, then $f = \sup\{f_1, f_2 \ldots f_n\}$ is convex.*

Proof. If $f_1, f_2 \ldots f_n$ are convex functions, then their epigraphs are convex sets. The epigraph of f is

$$\text{epi } f = \bigcap_{i=1}^{n} \text{epi } f_i,$$

and by Prop. 2, epi f is convex. Therefore f is a convex function. □

There is a geometric intution to the proof given in Fig. 5.

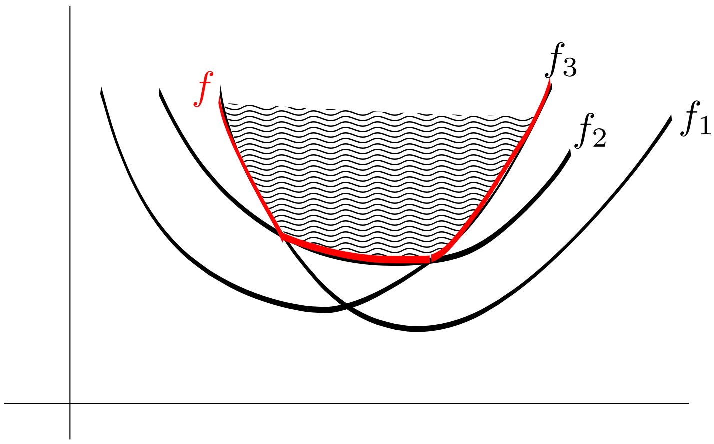

Figure 5: Intersection of epigraphs and supremum of functions.

Developing the notion of convexity required some work, but the effort is justified by the next theorem. Two definitions precede it.

Definition 7 (Local minimum and minimum point)**.** *Let $x \in X$, and $f : X \to \mathbb{R}$. x is a local minimum point if there exists a neighborhood N of x such that $y \in N$ implies that $f(x) \leq f(y)$. If $x \in X$ is a local minimum point, then $f(x)$ is a local minimum.*

Definition 8 (Global minimum and minimum point)**.** *If $x \in X$ is a local minimum point of $f : X \to \mathbb{R}$, and its associated neighborhood $N = X$, then x is a global minimum point. If $x \in X$ is a global minimum point, then $f(x)$ is a global minimum.*

Theorem 9 (Minima of convex functions)**.** *If $x \in X$ is a local minimum point of a convex function $f : X \to \mathbb{R}$, then x is a global minimum point and $f(x)$ is a global minimum.*

Proof. Assume that there exists a point $y \in X$ such that $f(y) < f(x)$. Then, since f is convex, we find that

$$\forall \alpha \in (0,1),\ f(\alpha x + (1-\alpha)y) \leq \alpha f(x) + (1-\alpha)f(y) < \alpha f(x) + (1-\alpha)f(x) = f(x) \quad (1)$$

Let $\alpha_n \in (0,1)$ define a sequence that converges to 1. Then there exists a number $M \in \mathbb{N}$ such that, whenever $n \geq M$,

$$\alpha_n x + (1-\alpha_n)y \in N,$$

where N is the neighborhood of x where $f(x)$ is a minima. However, since $f(\alpha_n x + (1-\alpha_n)y < f(x)$, we arrive at a contradiction. □

Another proposition concerns the set in X that minimizes a convex function.

Proposition 10 (Minimizing set of a convex function). *If $f : X \to \mathbb{R}$ is a convex function, then $\arg\min_x f(x)$ is a convex set.*

Proof. Let $x, y \in \arg\min_x f(x)$. Then $f(x) = f(y) = f^*$. Since f is convex we also find that for $\alpha \in [0,1]$,

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

So regardless of the value of α, the function f is minimized. This must mean that if $\alpha \in [0,1]$,

$$\alpha x + (1-\alpha)y \in \arg\min_x f(x)$$

which is the definition of a convex set. □

3 Differentiable Convex Functions

If we resttrict the investigation to functions that are differentiable, we find that there are connections between the convexity and the gradient. First, we need some facts from differential calculus.

Definition 11 (Differentiability). *If $D \subset \mathbb{R}^n$, the function $f : D \to \mathbb{R}$ is said to be differentiable at $x_0 \in D$ if*

$$\lim_{x \to x_0} \frac{\|f(x) - (f(x_0) + f'(x_0)(x - x_0))\|}{\|x - x_0\|} = 0,$$

where the derivative $f'(x_0) : \mathbb{R}^n \to \mathbb{R}$ is a linear map.

It follows that if a function f is differentiable at some point x_0, we can approximate the function at another point x as

$$f(x) \approx f(x_0) + f'(x_0)(x - x_0)$$

provided that x_0 and x are sufficiently close.

Definition 12 (Directional derivative). *If $D \subset \mathbb{R}^n$, then the directional derivative of $f : D \to \mathbb{R}$ at $x_0 \in D$, with respect to some vector $v \in \mathbb{R}^n$, is*

$$D_v f(x_0) = \lim_{t \to 0} \frac{f(x_0 + tv) - f(x_0)}{t}$$

provided that the limit exists.

The (total) derivative f' and the directional derivative $D_v f$ are connected by the following relation

$$D_v f(x_0) = f'(x_0)v$$

Definition 13 (Partial derivative). *If $D \subset \mathbb{R}^n$, and the function $f : D \to \mathbb{R}$ is differentiable, the partial derivative of f, at $x_0 \in D$, is*

$$\frac{\partial f}{\partial x_j}(x_0) = \lim_{t \to 0} \frac{f(x_0 + te_j) - f(x_0)}{t}$$

where e_j is the j'th standard basis vector of $\mathbb{R}^n$.

Definition 14 (Gradient)**.** *If $D \subset \mathbb{R}^n$, and the function $f : D \to \mathbb{R}$ is differentiable, the gradient of f is the column vector of its partial derivatives:*

$$\nabla f(x_0) = \Big[\frac{\partial f}{\partial x_1}(x_0),\ \frac{\partial f}{\partial x_2}(x_0),\ \dots\ ,\ \frac{\partial f}{\partial x_n}(x_0)\Big]^\mathsf{T}$$

The directional derivative and gradient are connected by the following relation

$$D_v f(x_0) = \langle \nabla f(x_0), v \rangle$$

where $\langle \cdot, \cdot \rangle$ is the familiar dot-product.

Now we have the tools to prove a powerful theorem of functions that are both convex and differentiable.

Theorem 15 (Convexity of differentiable functions)**.** *Let $f : X \to \mathbb{R}$ be a convex and differentiable function. Then the following statements are equivalent:*
i) f is convex
ii) $\forall x, y \in X,\ f(y) \geq f(x) + \langle \nabla f(x), y - x \rangle$
iii) $\forall x, y \in X,\ \langle \nabla f(x) - \nabla f(y), x - y \rangle \geq 0$

Proof. i) $\to$ ii) First note the following relations:

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

Since f is convex, we write the usual inequality using the relation above:

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

Now we write the inner product of the gradient of f, and the vector $x - y$, as

$$\langle \nabla f(y), x - y \rangle = \lim_{\alpha \to 0} \frac{f(y + \alpha(x - y)) - f(y)}{\alpha}$$

Using the convexity inequality, we get

$$\lim_{\alpha \to 0} \frac{f(y + \alpha(x - y)) - f(y)}{\alpha} \leq \lim_{\alpha \to 0} \frac{f(y) + \alpha(f(x) - f(y)) - f(y)}{\alpha} = \dots$$
$$\dots = \lim_{\alpha \to 0} \frac{\alpha(f(x) - f(y))}{\alpha} = f(x) - f(y)$$

and so

$$\langle \nabla f(y), x - y \rangle \leq f(x) - f(y)$$

Since this holds for all $x, y \in X$, we can swap the variables in the above inequality, which finishes the proof.

ii) $\to$ iii) Since the relation holds for all $x, y \in X$, we can swap x and y to produce two inequalities:

$$f(y) - f(x) \geq \langle \nabla f(x), y - x \rangle$$
$$f(x) - f(y) \geq \langle \nabla f(y), x - y \rangle$$

Adding them gives

$$0 \geq \langle \nabla f(x), y - x \rangle + \langle \nabla f(y), x - y \rangle$$

We can extract -1 from the second slot in the inner product, and using the distributativity in first slot property, we get

$$0 \geq \langle \nabla f(x), y - x \rangle - \langle \nabla f(y), y - x \rangle = \langle \nabla f(x) - \nabla f(y), y - x \rangle$$

□

If we further restrict our investigation to convex functions that are twice differentiable, we get conditions on the Hessian of f. To prove these results however, we need some additional machinery. First, the mean value theorem from single variable calculus.

Theorem 16 (Mean value theorem)**.** *If $g : [a, b] \to \mathbb{R}$ is continuous on $[a, b]$ and differentiable on (a, b), then there exists a point $c \in (a, b)$ so that*

$$g'(c) = \frac{g(b) - g(a)}{b - a}$$

Since we restrict our focus to twice differentiable functions, we can be sure that the requirements in the theorem holds. However, the theorem concerns functions with a 1-dimensional domain, and since we are working with functions that map subsets of $\mathbb{R}^n$ to $\mathbb{R}$, the theorem is not directly applicable.

Consider, as an example, a function $f : \mathbb{R}^2 \to \mathbb{R}$, such that the graph of f resembles some paraboloid. Fig. 6 is an example of such a function.

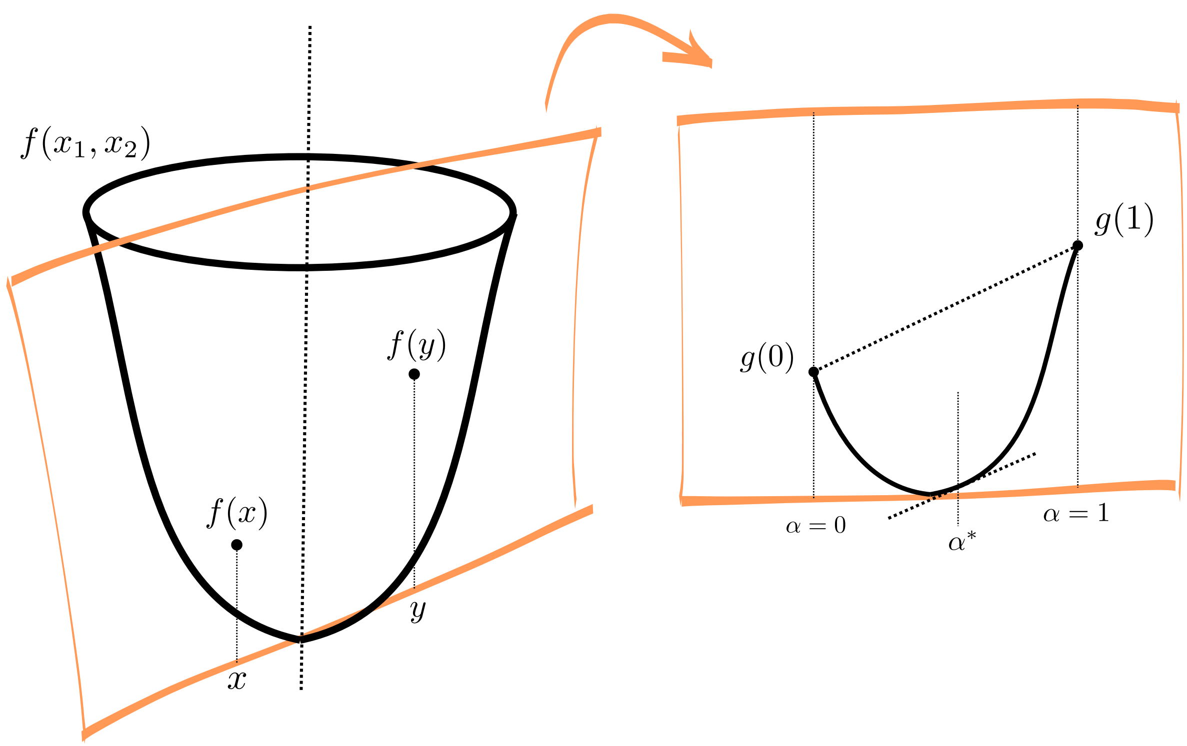

Figure 6: Example of the mean value theorem.

Pick two points $x, y \in \mathbb{R}^2$ and let $x, y, f(x)$ define a plane that slices through the paraboloid. Now if we let $z = \alpha x + (1-\alpha)y$, for $\alpha \in (0,1)$ we can picture the composition $g(\alpha) = f(z(\alpha))$ as depending on a single variable in the plane. By the mean value theorem, there exists a point α^*, with $z^* = \alpha^* x + (1-\alpha^*)y$, such that

$$g'(\alpha^*) = \frac{g(1) - g(0)}{1 - 0} = f(y) - f(x)$$

The derivative $g'(\alpha^*)$ corresponds to the directional derivative of $D_{y-x}f(z^*)$, or in terms of the gradient:

$$\langle \nabla f(z^*), y - x \rangle = f(y) - f(x)$$

Therefore, generalizing to higher dimensions, if we have $x, y \in X \subset \mathbb{R}^n$, and $f : X \to \mathbb{R}$ is differentiable, there exists a point $z \in X$ such that

$$f(y) = f(x) + \langle f(z), y - x \rangle$$

Definition 17 (Hessian matrix)**.** *Let $X \subset \mathbb{R}^n$, and $f : X \to \mathbb{R}$ be a twice differentiable function. The Hessian matrix of f is defined as the square matrix of second partial derivatives:*

$$H_f = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f}{\partial x_2 \partial x_1} & \frac{\partial^2 f}{\partial x_2^2} & \cdots & \frac{\partial^2 f}{\partial x_2 \partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1} & \frac{\partial^2 f}{\partial x_n \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_n^2} \end{bmatrix}$$

Definition 18 (Positive semi-definite)**.** *A symmetric real matrix $M \in \mathcal{L}(\mathbb{R}^n)$ is said to be positive-semi definite (PSD) if $x^T M x$ is non-negative for all vectors $x \in \mathbb{R}^n$.*

Important properties of M being PSD is that all eigenvalues of M are real, and

$$\forall x \in \mathbb{R}^n, \ \langle Mx, x \rangle \geq 0$$

which is also a necessary condition.

Now we can state and prove (in one direction) a theorem about twice differentiable convex functions.

Theorem 19 (Convexity of twice-differentiable functions)**.** *Let $X \subset \mathbb{R}^n$ be a convex set, and $f : X \to \mathbb{R}$ a twice differentiable function. Then f is convex if and only if the Hessian of f, H_f, is a positive semi-definite matrix on X.*

Proof. If f is a function on X, then the partial derivative $\frac{\partial f}{\partial x_j}$ is also a function on X, so the mean value theorem holds just the same. Therefore, for all $x, y \in X$ there exists a point $z = \alpha x + (1-\alpha)y \in X, \ \alpha \in (0,1)$ such that

$$\frac{\partial}{\partial x_j} f(y) = \frac{\partial}{\partial x_j} f(x) + \langle \nabla \frac{\partial}{\partial x_j} f(z), y - x \rangle$$

or equivalently, with the dot-product notation $\langle a, b \rangle = a \cdot b = a^\intercal b$

$$\frac{\partial}{\partial x_j} f(y) = \frac{\partial}{\partial x_j} f(x) + \left[\nabla \frac{\partial}{\partial x_j} f(z) \right]^\intercal (y - x)$$

By writing the gradient as a column-vector of partial derivatives we get

$$\frac{\partial}{\partial x_j} f(y) = \frac{\partial}{\partial x_j} f(x) + \Big[\frac{\partial^2}{\partial x_1 x_j} f(z), \frac{\partial^2}{\partial x_2 x_j} f(z), \ \dots\Big](y - x)$$

By letting $j = 1, 2 \dots n$, we get a system of equations involving the Hessian matrix

$$\nabla f(y) = \nabla f(x) + H_f(z)(y - x)$$

Moving $\nabla f(x)$ to the left hand side, and taking the inner product with respect to $x - y$ (on both sides) gives

$$\langle H_f(z)(y - x), y - x\rangle = \langle \nabla f(y) - \nabla f(x), y - x\rangle \geq 0$$

where we used property $iii)$ of Theorem 15 for the inequality.

If we restrict z to be in the interior of X, denoted by $\text{int}(X)$, we can always find points $x, y \in X$ such that the above inequality holds. Therefore, H_f is PSD on $\text{int}(X)$:

$$\forall z \in \text{int}(X), \ x \in \mathbb{R}^n, \ \langle H_f(z)x, x\rangle \geq 0$$

If z lies on the boundary of X, we can instead appeal to the continuity of H_f. Let z_n be a sequence of points in $\text{int}(X)$ such that $(z_n) \to z$, where z lies on the boundary of X. It follows that

$$\forall x \in \mathbb{R}^n, \ \langle H_f(z_n)x, x\rangle \geq 0$$

The sequence defined by the left hand side is in $\mathbb{R}$, so we can apply the order limit theorem to get

$$\forall x \in \mathbb{R}^n, \ \lim_{n\to\infty} \langle H_f(z_n)x, x\rangle \geq 0$$

and finally, by continuity,

$$\forall x \in \mathbb{R}^n, \ \langle H_f(z)x, x\rangle \geq 0$$

which completes the proof.

□