

Chapter 1

Optimality Conditions: Unconstrained Optimization

1.1 Differentiable Problems

Consider the problem of minimizing the function $f : \mathbb{R}^n \rightarrow \mathbb{R}$ where f is twice continuously differentiable on \mathbb{R}^n :

$$\mathcal{P} \quad \underset{x \in \mathbb{R}^n}{\text{minimize}} \quad f(x)$$

We wish to obtain constructible first- and second-order necessary and sufficient conditions for optimality. Recall the following elementary results.

THEOREM 1.1.1 [*First- Order Necessary Conditions for Optimality*]

Let $f : \mathbb{R}^n \rightarrow \mathbb{R}$ be differentiable at a point $\bar{x} \in \mathbb{R}^n$. If \bar{x} is a local solution to the problem \mathcal{P} , then $\nabla f(\bar{x}) = 0$.

PROOF: From the definition of the derivative we have that

$$f(x) = f(\bar{x}) + \nabla f(\bar{x})^T(x - \bar{x}) + o(\|x - \bar{x}\|)$$

where $\lim_{x \rightarrow \bar{x}} \frac{o(\|x - \bar{x}\|)}{\|x - \bar{x}\|} = 0$. Let $x := \bar{x} - t\nabla f(\bar{x})$. Then

$$0 \leq \frac{f(\bar{x} - t\nabla f(\bar{x})) - f(\bar{x})}{t} = -\|\nabla f(\bar{x})\|^2 + \frac{o(t\|\nabla f(\bar{x})\|)}{t}.$$

Taking the limit as $t \downarrow 0$ we obtain

$$0 \leq -\|\nabla f(\bar{x})\|^2 \leq 0.$$

Hence $\nabla f(\bar{x}) = 0$. ■

THEOREM 1.1.2 [Second-Order Optimality Conditions]

Let $f : \mathbb{R}^n \rightarrow \mathbb{R}$ be twice differentiable at the point $\bar{x} \in \mathbb{R}^n$.

1. (necessity) If \bar{x} is a local solution to the problem \mathcal{P} , then $\nabla f(\bar{x}) = 0$ and $\nabla^2 f(\bar{x})$ is positive semi-definite.
2. (sufficiency) If $\nabla f(\bar{x}) = 0$ and $\nabla^2 f(\bar{x})$ is positive definite, then there is an $\alpha > 0$ such that $f(x) \geq f(\bar{x}) + \alpha\|x - \bar{x}\|^2$ for all x near \bar{x} .

PROOF:

1. We make use of the second-order Taylor series expansion

$$(1.1.1) f(x) = f(\bar{x}) + \nabla f(\bar{x})^T(x - \bar{x}) + \frac{1}{2}(x - \bar{x})^T \nabla^2 f(\bar{x})(x - \bar{x}) + o(\|x - \bar{x}\|^2).$$

Given $d \in \mathbb{R}^n$ and $t > 0$ set $x := \bar{x} + td$, plugging this into (1.1.1) we find that

$$0 \leq \frac{f(x + td) - f(x)}{t^2} = \frac{1}{2}d^T \nabla^2 f(\bar{x})d + \frac{o(t^2)}{t^2}$$

since $\nabla f(\bar{x}) = 0$ by Theorem 1.1.1. Taking the limit as $t \rightarrow 0$ we get that

$$0 \leq d^T \nabla^2 f(\bar{x})d.$$

Now since d was chosen arbitrarily we have that $\nabla^2 f(\bar{x})$ is positive semi-definite.

2. From (1.1.1) we have that

$$(1.1.2) \quad \frac{f(x) - f(\bar{x})}{\|x - \bar{x}\|^2} = \frac{1}{2} \frac{(x - \bar{x})^T}{\|x - \bar{x}\|} \nabla^2 f(\bar{x}) \frac{(x - \bar{x})}{\|x - \bar{x}\|} + \frac{o(\|x - \bar{x}\|^2)}{\|x - \bar{x}\|^2}.$$

If $\lambda > 0$ is the smallest eigenvalue of $\nabla^2 f(\bar{x})$, choose $\epsilon > 0$ so that

$$(1.1.3) \quad \left| \frac{o(\|x - \bar{x}\|^2)}{\|x - \bar{x}\|^2} \right| \leq \frac{\lambda}{4}$$

whenever $\|x - \bar{x}\| < \epsilon$. Then for all $\|x - \bar{x}\| < \epsilon$ we have from (1.1.2) and (1.1.3) that

$$\begin{aligned} \frac{f(x) - f(\bar{x})}{\|x - \bar{x}\|^2} &\geq \frac{1}{2}\lambda + \frac{o(\|x - \bar{x}\|^2)}{\|x - \bar{x}\|^2} \\ &\geq \frac{1}{4}\lambda. \end{aligned}$$

Consequently, if we set $\alpha = \frac{1}{4}\lambda$, then

$$f(x) \geq f(\bar{x}) + \alpha\|x - \bar{x}\|^2$$

whenever $\|x - \bar{x}\| < \epsilon$.

■

1.2 Convex Problems

Observe that Theorem 1.1.1 establishes first-order necessary conditions while Theorem 1.1.2 establishes both second-order necessary and sufficient conditions. What about first-order sufficiency conditions? For this we introduce the following definitions.

DEFINITION 1.2.1 [*Convex Sets and Functions*]

1. A subset $C \subset \mathbb{R}^n$ is said to be convex if for every pair of points x and y taken from C , the entire line segment connecting x and y is also contained in C , i.e.,

$$[x, y] \subset C \quad \text{where} \quad [x, y] = \{(1 - \lambda)x + \lambda y : 0 \leq \lambda \leq 1\}.$$

2. A function $f : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{\pm\infty\}$ is said to be convex if the set

$$\text{epi}(f) = \{(\mu, x) : f(x) \leq \mu\}$$

is a convex subset of \mathbb{R}^{1+n} . In this context, we also define the set

$$\text{dom}(f) = \{x \in \mathbb{R}^n : f(x) < +\infty\}$$

to be the essential domain of f .

LEMMA 1.2.1 *The function $f : \mathbb{R}^n \rightarrow \mathbb{R}$ is convex if and only if for every two points $x_1, x_2 \in \text{dom}(f)$ and $\lambda \in [0, 1]$ we have*

$$f(\lambda x_1 + (1 - \lambda)x_2) \leq \lambda f(x_1) + (1 - \lambda)f(x_2).$$

That is, the secant line connecting $(x_1, f(x_1))$ and $(x_2, f(x_2))$ lies above the graph of f .

EXAMPLE: The following functions are examples of convex functions: $c^T x$, $\|x\|$, e^x , x^2

The significance of convexity in optimization theory is illustrated in the following result.

THEOREM 1.2.1 *Let $f : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{\pm\infty\}$ be convex. If $\bar{x} \in \text{dom}(f)$ is a local solution to the problem \mathcal{P} , then \bar{x} is a global solution to the problem \mathcal{P} .*

PROOF: If $f(\bar{x}) = -\infty$ we are done, so let us assume that $-\infty < f(\bar{x})$. Suppose there is a $\hat{x} \in \mathbb{R}^n$ with $f(\hat{x}) < f(\bar{x})$. Let $\epsilon > 0$ be such that $f(\bar{x}) \leq f(x)$ whenever $\|x - \bar{x}\| \leq \epsilon$. Set $\lambda := \epsilon(2\|\bar{x} - \hat{x}\|)^{-1}$ and $x_\lambda := \bar{x} + \lambda(\hat{x} - \bar{x})$. Then $\|x_\lambda - \bar{x}\| \leq \epsilon/2$ and $f(x_\lambda) \leq (1 - \lambda)f(\bar{x}) + \lambda f(\hat{x}) < f(\bar{x})$. This contradicts the choice of ϵ , hence no such \hat{x} exists. ■

If f is a differentiable convex function, then a better result can be established. In order to obtain this result we need the following lemma.

LEMMA 1.2.2 *Let $f : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$ be convex.*

1. *Given $x \in \text{dom}(f)$ and $d \in \mathbb{R}^n$ the difference quotient*

$$(1.2.4) \quad \frac{f(x + td) - f(x)}{t}$$

is a non-decreasing function of t on $(0, +\infty)$.

2. *For every $x \in \text{dom}(f)$ and $d \in \mathbb{R}^n$ the directional derivative $f'(x; d)$ always exists and is given by*

$$(1.2.5) \quad f'(x; d) := \inf_{t>0} \frac{f(x + td) - f(x)}{t}.$$

3. *For every $x \in \text{dom}(f)$, the function $f'(x; \cdot)$ is sublinear, i.e. $f'(x; \cdot)$ is positively homogeneous,*

$$f'(x; \alpha d) = \alpha f'(x; d) \quad \forall d \in \mathbb{R}^n, 0 \leq \alpha,$$

and subadditive,

$$f'(x; u + v) \leq f'(x; u) + f'(x; v).$$

PROOF: We assume (1.2.4) is true and show (1.2.5). If $x + td \notin \text{dom}(f)$ for all $t > 0$, then the result obviously true. Therefore, we may as well assume that there is a $\bar{t} > 0$ such that $x + td \in \text{dom}(f)$ for all $t \in (0, \bar{t}]$. Recall that

$$(1.2.6) \quad f'(x; d) := \lim_{t \downarrow 0} \frac{f(x + td) - f(x)}{t}.$$

Now if the difference quotient (1.2.4) is non-decreasing in t on $(0, +\infty)$, then the limit in (1.2.6) is necessarily given by the infimum in (1.2.5). This infimum always exists and so $f'(x; d)$ always exists and is given by (1.2.5).

We now prove (1.2.4). Let $x \in \text{dom}(f)$ and $d \in \mathbb{R}^n$. If $x + td \notin \text{dom}(f)$ for all $t > 0$, then the result is obviously true. Thus, we may assume that

$$0 < \bar{t} = \sup\{t : x + td \in \text{dom}(f)\}.$$

Let $0 < t_1 < t_2 < \bar{t}$ (we allow the possibility that $t_2 = \bar{t}$ if $\bar{t} < +\infty$). Then

$$\begin{aligned} f(x + t_1 d) &= f\left(x + \left(\frac{t_1}{t_2}\right) t_2 d\right) \\ &= f\left[\left(1 - \left(\frac{t_1}{t_2}\right)\right) x + \left(\frac{t_1}{t_2}\right) (x + t_2 d)\right] \\ &\leq \left(1 - \frac{t_1}{t_2}\right) f(x) + \left(\frac{t_1}{t_2}\right) f(x + t_2 d). \end{aligned}$$

Hence

$$\frac{f(x + t_1 d) - f(x)}{t_1} \leq \frac{f(x + t_2 d) - f(x)}{t_2}.$$

We now show Part 3 of this result. To see that $f'(x; \cdot)$ is positively homogeneous let $d \in \mathbb{R}^n$ and $\alpha > 0$ and note that

$$f'(x; \alpha d) = \alpha \lim_{t \downarrow 0} \frac{f(x + (t\alpha)d) - f(x)}{(t\alpha)} = \alpha f'(x; d).$$

To see that $f'(x; \cdot)$ is subadditive let $u, v \in \mathbb{R}^n$, then

$$\begin{aligned} f'(x; u + v) &= \lim_{t \downarrow 0} \frac{f(x + t(u + v)) - f(x)}{t} \\ &= \lim_{t \downarrow 0} \frac{f(x + \frac{t}{2}(u + v)) - f(x)}{t/2} \\ &= \lim_{t \downarrow 0} 2 \frac{f(\frac{1}{2}(x + tu) + \frac{1}{2}(x + tv)) - f(x)}{t} \\ &\leq \lim_{t \downarrow 0} 2 \frac{\frac{1}{2}f(x + tu) + \frac{1}{2}f(x + tv) - f(x)}{t} \\ &= \lim_{t \downarrow 0} \frac{f(x + tu) - f(x)}{t} + \frac{f(x + tv) - f(x)}{t} \\ &= f'(x; u) + f'(x; v). \end{aligned}$$

■

From Lemma 1.2.2 we immediately obtain the following result.

THEOREM 1.2.2 *Let $f : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$ be convex and suppose that $\bar{x} \in \mathbb{R}^n$ is a point at which f is differentiable. Then \bar{x} is a global solution to the problem \mathcal{P} if and only if $\nabla f(\bar{x}) = 0$.*

PROOF: If \bar{x} is a global solution to the problem \mathcal{P} , then, in particular, \bar{x} is a local solution to the problem \mathcal{P} and so $\nabla f(\bar{x}) = 0$ by Theorem 1.1.1. Conversely, if $\nabla f(\bar{x}) = 0$, then, by setting $t := 1$, $x := \bar{x}$, and $d := y - \bar{x}$ in (1.2.5), we get that

$$0 \leq f(y) - f(\bar{x}),$$

or $f(\bar{x}) \leq f(y)$. Since y was chosen arbitrarily, the result follows. ■

As Theorems 1.2.1 and 1.2.2 demonstrate, convex functions are very nice functions indeed. This is especially so with regard to optimization theory. Thus, it is important that we be able to recognize when a function is convex. For this reason we give the following result.

THEOREM 1.2.3 *Let $f : \mathbb{R}^n \rightarrow \mathbb{R}$.*

1. *If f is differentiable on \mathbb{R}^n , then the following statements are equivalent:*

- (a) *f is convex,*
- (b) *$f(y) \geq f(x) + \nabla f(x)^T(y - x)$ for all $x, y \in \mathbb{R}^n$*
- (c) *$(\nabla f(x) - \nabla f(y))^T(x - y) \geq 0$ for all $x, y \in \mathbb{R}^n$.*

2. *If f is twice differentiable then f is convex if and only if f is positive semi-definite for all $x \in \mathbb{R}^n$.*

PROOF: (a) \Rightarrow (b) If f is convex, then 1.2.3 holds. By setting $t := 1$ and $d := y - x$ we obtain (b).

(b) \Rightarrow (c) Let $x, y \in \mathbb{R}^n$. From (b) we have

$$f(y) \geq f(x) + \nabla f(x)^T(y - x)$$

and

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

By adding these two inequalities we obtain (c).

(c) \Rightarrow (b) Let $x, y \in \mathbb{R}^n$. By the mean value theorem there exists $0 < \lambda < 1$ such that

$$f(y) - f(x) = \nabla f(x_\lambda)^T(y - x)$$

where $x_\lambda := \lambda y + (1 - \lambda)x$. By hypothesis,

$$\begin{aligned} 0 &\leq [\nabla f(x_\lambda) - \nabla f(x)]^T(x_\lambda - x) \\ &= \lambda[\nabla f(x_\lambda) - \nabla f(x)]^T(y - x) \\ &= \lambda[f(y) - f(x) - \nabla f(x)^T(y - x)]. \end{aligned}$$

Hence $f(y) \geq f(x) + \nabla f(x)^T(y - x)$.

(b) \Rightarrow (a) Let $x, y \in \mathbb{R}^n$ and set

$$\alpha := \max_{\lambda \in [0,1]} \varphi(\lambda) := [f(\lambda y + (1 - \lambda)x) - (\lambda f(y) + (1 - \lambda)f(x))].$$

We need to show that $\alpha \leq 0$. Since $[0, 1]$ is compact and φ is continuous, there is a $\lambda \in [0, 1]$ such that $\varphi(\lambda) = \alpha$. If λ equals zero or one, we are done. Hence we may as well assume that $0 < \lambda < 1$ in which case

$$0 = \varphi'(\lambda) = \nabla f(x_\lambda)^T(y - x) + f(x) - f(y)$$

where $x_\lambda = x + \lambda(y - x)$, or equivalently

$$\lambda f(y) = \lambda f(x) - \nabla f(x_\lambda)^T(x - x_\lambda).$$

But then

$$\begin{aligned} \alpha &= f(x_\lambda) - (f(x) + \lambda(f(y) - f(x))) \\ &= f(x_\lambda) + \nabla f(x_\lambda)^T(x - x_\lambda) - f(x) \\ &\leq 0 \end{aligned}$$

by (b).

2) Suppose f is convex and let $x, d \in \mathbb{R}^n$, then by (b) of Part (1),

$$f(x + td) \geq f(x) + t\nabla f(x)^T d$$

for all $t \in \mathbb{R}$. Replacing the left hand side of this inequality with its second-order Taylor expansion yields the inequality

$$f(x) + t\nabla f(x)^T d + \frac{t^2}{2}d^T \nabla^2 f(x)d + o(t^2) \geq f(x) + t\nabla f(x)^T d$$

or equivalently

$$\frac{1}{2}d^T \nabla^2 f(x)d + \frac{o(t^2)}{t^2} \geq 0.$$

Letting $t \rightarrow 0$ yields the inequality

$$d^T \nabla^2 f(x)d \geq 0.$$

Since d was arbitrary, $\nabla^2 f(x)$ is positive semi-definite.

Conversely, if $x, y \in \mathbb{R}^n$, then by the mean value theorem there is a $\lambda \in (0, 1)$ such that

$$f(y) = f(x) + \nabla f(x)^T(y - x) + \frac{1}{2}(y - x)^T \nabla^2 f(x_\lambda)(y - x)$$

where $x_\lambda = \lambda y + (1 - \lambda)x$. Hence

$$f(y) \geq f(x) + \nabla f(x)^T(y - x)$$

since $\nabla^2 f(x_\lambda)$ is positive semi-definite. Therefore, f is convex by (b) of Part (1). ■

We have established that $f'(x; d)$ exists for all $x \in \text{dom}(f)$ and $d \in \mathbb{R}^n$, but we have not yet discussed to continuity properties of f . We give a partial result in this direction in the next lemma.

LEMMA 1.2.3 *Let $f : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$ be convex. Then f is bounded in a neighborhood of a point \bar{x} if and only if f is Lipschitz in a neighborhood of \bar{x} .*

PROOF: If f is Lipschitz in a neighborhood of \bar{x} , then f is clearly bounded above in a neighborhood of \bar{x} . Therefore, we assume local boundedness and establish Lipschitz continuity.

Let $\epsilon > 0$ and $M > 0$ be such that $|f(x)| \leq M$ for all $x \in \bar{x} + 2\epsilon\mathbb{B}$. Set $g(x) = f(x + \bar{x}) - f(\bar{x})$. It is sufficient to show that g is Lipschitz on $\epsilon\mathbb{B}$. First note that for all $x \in 2\epsilon\mathbb{B}$

$$0 = g(0) = g\left(\frac{1}{2}x + \frac{1}{2}(-x)\right) \leq \frac{1}{2}g(x) + \frac{1}{2}g(-x),$$

and so $-g(x) \leq g(-x)$ for all $x \in 2\epsilon\mathbb{B}$. Next, let $x, y \in \epsilon\mathbb{B}$ with $x \neq y$ and set $\alpha = \|x - y\|$. Then $w = y + \epsilon\alpha^{-1}(y - x) \in 2\epsilon\mathbb{B}$, and so

$$g(y) = g\left(\frac{1}{1 + \epsilon^{-1}\alpha}x + \frac{\epsilon^{-1}\alpha}{1 + \epsilon^{-1}\alpha}w\right) \leq \frac{1}{1 + \epsilon^{-1}\alpha}g(x) + \frac{\epsilon^{-1}\alpha}{1 + \epsilon^{-1}\alpha}g(w).$$

Consequently,

$$g(y) - g(x) \leq \frac{\epsilon^{-1}\alpha}{1 + \epsilon^{-1}\alpha}(g(w) - g(x)) \leq 2M\epsilon^{-1}\alpha = 2M\epsilon^{-1}\|x - y\|.$$

Since this inequality is symmetric in x and y , we obtain the result. ■

1.3 Convex Composite Problems

Convex composite optimization is concerned with the minimization of functions of the form $f(x) := h(F(x))$ where $h : \mathbb{R}^m \rightarrow \mathbb{R} \cup \{+\infty\}$ is a proper convex function and $F : \mathbb{R}^n \rightarrow \mathbb{R}^m$ is continuously differentiable. Most problems from nonlinear programming can be cast in this framework.

EXAMPLES:

- (1) Let $F : \mathbb{R}^n \rightarrow \mathbb{R}^m$ where $m > n$, and consider the equation $F(x) = 0$. Since $m > n$ it is highly unlikely that a solution to this equation exists. However, one might try to obtain a *best* approximate solution by solving the problem $\min\{\|F(x)\| : x \in \mathbb{R}^n\}$. This is a convex composite optimization problem since the norm is a convex function.
- (2) Again let $F : \mathbb{R}^n \rightarrow \mathbb{R}^m$ where $m > n$, and consider the inclusion $F(x) \in C$, where $C \subset \mathbb{R}^m$ is a non-empty closed convex set. One can pose this inclusion as the optimization problem $\min\{\text{dist}(F(x)|C) : x \in \mathbb{R}^n\}$. This is a convex composite optimization problem since the distance function

$$\text{dist}(y | C) := \inf_{z \in C} \|y - z\|$$

is a convex function.

- (3) Let $F : \mathbb{R}^n \rightarrow \mathbb{R}^m$, $C \subset \mathbb{R}^n$ a non-empty closed convex set, and $f_0 : \mathbb{R}^n \rightarrow \mathbb{R}$, and consider the constrained optimization problem $\min\{f_0(x) : F(x) \in C\}$. One can approximate this problem by the unconstrained optimization problem

$$\min\{f_0(x) + \alpha \text{dist}(f(x)|C) : x \in \mathbb{R}^n\}.$$

This is a convex composite optimization problem where $h(\eta, y) = \eta + \alpha \text{dist}(y|C)$ is a convex function. The function $f_0(x) + \alpha \text{dist}(f(x)|C)$ is called an *exact penalty function* for the problem $\min\{f_0(x) : F(x) \in C\}$. We will review the theory of such functions in a later section.

Most of the first-order theory for convex composite functions is easily derived from the observation that

$$(1.3.7) \quad f(y) = h(F(y)) = h(F(x) + F'(x)(y - x)) + o(\|y - x\|).$$

This local representation for f is a direct consequence of h being locally Lipschitz:

$$\begin{aligned} |h(F(y)) - h(F(x) + F'(x)(y - x))| \\ \leq K\|y - x\| \int_0^1 \|F'(x + t(y - x)) - F'(x)\| dt \end{aligned}$$

for some $K \geq 0$. Equation (1) can be written equivalently as

$$(1.3.8) \quad h(F(x + d)) = h(F(x)) + \Delta f(x; d) + o(\|d\|)$$

where

$$\Delta f(x; d) := h(F(x) + F'(x)d) - h(F(x)).$$

From 1.3.8, one immediately obtains the following result.

LEMMA 1.3.1 *Let $h : \mathbb{R}^m \rightarrow \mathbb{R}$ be convex and let $F : \mathbb{R}^n \rightarrow \mathbb{R}^m$ be continuously differentiable. Then the function $f = h \circ F$ is everywhere directional differentiable and one has*

$$(1.3.9) \quad \begin{aligned} f'(x; d) &= h'(F(x); F'(x)d) \\ &= \inf_{\lambda > 0} \frac{\Delta f(x; \lambda d)}{\lambda}. \end{aligned}$$

This result yields the following optimality condition for convex composite optimization problems.

THEOREM 1.3.1 *Let $h : \mathbb{R}^m \rightarrow \mathbb{R}$ be convex and $F : \mathbb{R}^n \rightarrow \mathbb{R}^m$ be continuously differentiable. If \bar{x} is a local solution to the problem $\min\{h(F(x))\}$, then $d = 0$ is a global solution to the problem*

$$(1.3.10) \quad \min_{d \in \mathbb{R}^n} h(F(\bar{x}) + F'(\bar{x})d).$$

There are various ways to test condition 1.3.8. A few of these are given below.

LEMMA 1.3.2 *Let h and F be as in Theorem 1.3.1. The following conditions are equivalent*

- (a) $d = 0$ is a global solution to 1.3.10.
- (b) $0 \leq h'(F(x); F'(x)d)$ for all $d \in \mathbb{R}^n$.
- (c) $0 \leq \Delta f(x; d)$ for all $d \in \mathbb{R}^n$.

PROOF: The equivalence of (a) and (b) follows immediately from convexity. Indeed, this equivalence is the heart of the proof of Theorem 1.3.1. The equivalence of (b) and (c) is an immediate consequence of 1.3.2. ■

In the sequel, we say that $x \in \mathbb{R}^n$ satisfies the first-order condition for optimality for the convex composite optimization problem if it satisfies any of the three conditions (a)–(c) of Lemma 1.3.2.

1.3.1 A Note on Directional Derivatives

Recall that if $f : \mathbb{R}^n \rightarrow \mathbb{R}$ is differentiable, then the function $f'(x; d)$ is linear in d :

$$f'(x; \alpha d_1 + \beta d_2) = \alpha f'(x; d_1) + \beta f'(x; d_2) .$$

If f is only assumed to be convex and not necessarily differentiable, then $f'(x; \cdot)$ is sublinear and hence convex. Finally, if $f = h \circ F$ is convex composite with $h : \mathbb{R}^m \rightarrow \mathbb{R}$ convex and $F : \mathbb{R}^n \rightarrow \mathbb{R}^m$ continuously differentiable, then, by Lemma (1.3.1), $f'(x; \cdot)$ is also sublinear and hence convex. Moreover, the approximate directional derivative $\Delta f(x; d)$ satisfies

$$\lambda_1^{-1} \Delta f(x; \lambda_1 d) \leq \lambda_2^{-1} \Delta f(x; \lambda_2 d) \quad \text{for } 0 < \lambda_1 \leq \lambda_2,$$

by the non-decreasing nature of the difference quotients. Thus, in particular,

$$\Delta f(x; \lambda d) \leq \lambda \Delta f(x; d) \quad \text{for all } \lambda \in [0, 1].$$