Journal of the Nigerian Association of Mathematical Physics
Volume 26, (March, 2014), pp 22 – 27

Conditions for the Convexity of a Class of Twice Continuously Differentiable Real-Valued Functions of Several Variables

Peter E. Ezimadu

Department of Mathematics and Computer Science,
Delta State University, Abraka.

Abstract

The importance of convex functions in optimization theory calls for the availability of results that ensure their easy recognition. In response to this demand this work presents a characterization which gives an interplay associated with the first order condition, the second order condition, the epigraph, the integral and monotonicity for convex functions. Specifically, it presents a characterization of convex functions by extending an existing result to include the Hessian and epigraph of convex functions; thus providing an equivalence for the definitions of convex functions, and consequently a better horizon for understanding convex functions.

Keywords: Convex Function, Epigraph, Monotone Mapping, Gradient Vector, Hessian.

1.0 Introduction

The theory of convex functions is very important in many real-world problems. For instance, constrained control and estimation problems are convex. In this work we will promote these concepts and thus provide a better horizon for recognizing convex functions.

Now, let us consider the following definitions which are fundamental to the comprehension of this work.

Definition 1.1 *Let the set $D \subseteq \mathbb{R}^n$. If for any $x, y \in D$ we have that*

$$z = \theta x + (1-\theta)y \in D, \qquad \forall\, \theta \in [0,1] \tag{1}$$

then D is said to be convex [1,2,3].

It follows that D can have no re-entrant corners. This means that for any two points $x, y \in D$, the line segment joining x and y is entirely contained in D. It also states that D is path-connected. That is two arbitrary points in D can be linked by a continuous path. A more general definition of convex set which readily follows is that $\forall\, x_i \in D,\ i = 1, \ldots, k,$

$$z = \sum_{i=1}^{k} \theta_i x_i \in D \tag{2}$$

where $\sum_{i=1}^{k} \theta_i = 1,\ \theta_i \geq 0$. The vector z in (1) or (2) is referred to as a convex combination of the points $x_1, x_2, \ldots, x_k$.

Definition 1.2 *Let $D \subset \mathbb{R}^n$ be a nonempty convex set. A function $f: D \to \mathbb{R}$ is said to be convex on D if for any $x, y \in D$ and all $0 \leq \theta \leq 1$, we have*

$$f(\theta x + (1-\theta)y) \leq \alpha f(x) + (1-\theta) f(y) \tag{3}$$

If strict inequality holds in (3) *for all $x \neq y$, then f is called a strict convex function* [1,2,4].

The geometric interpretation of convexity is simple. For a convex function the function values are below the corresponding chord, that is, the values of a convex function at points on the line segment $\theta x + (1-\theta)y$ are less or equal to the height of the chord joining the points $(x, f(x))$ and $(x, f(y))$.

A function is convex if and only if it is convex when restricted to any line that intersects its domain. Rephrased, *f is convex if and only if for all $x \in D$ and for all v, the function $h(t) = f(x + tw)$ is convex on* $\{t: x + tw \in D\}$. This property is very useful in testing whether a function is convex by restricting it to a line.

Note: If f is convex (strictly convex) function then $-f$ is a concave (strictly concave) function.

Corresponding author: ***Peter E. Ezimadu,*** E-mail, Tel.: +2348028831349

Definition 1.3 *Let $D \subset \mathbb{R}^n$ be a nonempty set, and $f: D \longrightarrow \mathbb{R}^n$. The epigraph of f is a subset of $\mathbb{R}^{n+1}$ defined by*

$$epif = \{(x,t): f(x) \leq t, \ \ x \in D, \ \ t \in \mathbb{R}\} \tag{4}$$

Definition 1.4 *A function is concave if and only if its hypograph, defined as*

$$hypf = \{(x,t): f(x) \geq t, \ \ x \in D, \ \ t \in \mathbb{R}\} \tag{5}$$

is a convex set.

The link between convex sets and convex functions is via the epigraph: *A function is convex if and only if its epigraph is a convex set* [5,6].

2.0 Jensen's Inequality and Extensions

The basic inequality (3), that is

$$f(\theta x + (1-\theta)y) \leq \theta f(x) + (1-\theta)f(y),$$

is sometimes called Jensen's Inequality. It is easily extended to convex combinations of more than two points: If f is convex, $x_1, \dots, x_n \in D$, and $\theta_1, \dots, \theta_n \geq 0$ with $\theta_1 + \cdots + \theta_n = 1$, then

$$f(\theta_1 x_1 + \cdots + \theta_k x_k) \leq \theta_1 f(x_1) + \cdots + \theta_k f(x_k) \tag{6}$$

As in the case of convex sets, the inequality extends to infinite sums, integrals, and expected values. For example, if $p(z) \geq 0$ on D, $\int_D p(z)dz = 1$, then

$$f\left(\int_D p(z)z dz\right) \leq \int_D f(z)p(z)dz \tag{7}$$

provided the integrals exist. In the most general case we can take any probability measure with support in D. If z is a random variable such that $z \in D$ with probability one, and f is convex, then we have

$$f(Ez) \leq Ef(z), \tag{8}$$

provided the expectations exist. We can recover the basic inequality (3) from this general form, by taking the random variable x to have support $\{x, y\}$, with $\text{prob}(z = x) = \theta$, $\text{prob}(z = y) = 1 - \theta$.

Thus the inequality (9) characterizes convexity: If f is not convex, there is a random variable $z \in D$ with probability one, such that $f(Ez) > Ef(z)$.

All of these inequalities are now called Jensen's Inequality, even though the inequality studied by Jensen was the very simple one

$$f\left(\frac{x+y}{2}\right) \leq \frac{f(x)+f(y)}{2} \text{ [4].} \tag{9}$$

3.0 Gradient and Hessian Matrices of Several Variables Functions

Convex functions need not be necessarily differentiable however differentiable convex functions can be characterized using their gradient vectors and Hessian matrices.

Now a continuous function $f: \mathbb{R}^n \longrightarrow \mathbb{R}$ is said to be continuously differentiable at $x \in \mathbb{R}^n$, if $\left(\frac{\partial f}{\partial x_i}\right)(x)$ exists and is continuous, $i = 1, \dots, n$.

Definition 3.1 *The gradient of f at x is defined as*

$$\nabla f(x) = \left(\frac{\partial f}{\partial x_1}(x), \dots, \frac{\partial f}{\partial x_n}(x)\right)^T \text{ [7].} \tag{10}$$

If f is continuously differentiable at every point of an open set $D \subset \mathbb{R}^n$, then f is said to be continuously differentiable on D and denoted by $f \in C^1(D)$.

A continuously differentiable function $f: \mathbb{R}^n \longrightarrow \mathbb{R}$ is called twice continuously differentiable at $x \in \mathbb{R}^n$ if $\frac{\partial^2 f}{\partial x_i \partial x_j}(x)$ exists and is continuous, $i = 1, \dots, n$.

Definition 3.2 *The $n \times n$-symmetric matrix*

$$Hf(x) = \left(\frac{\partial^2 f(x)}{\partial x_i x_j} \quad i,j = 1,2,\dots,n\right) \tag{11}$$

of all second-order partial derivatives evaluated at x is called the Hessian of f at x [7].

For example, consider the function

$$f(x) = e^{x_1 - x_2} + e^{x_2 - x_1} + e^{x_1^2} + x_3^2. \tag{12}$$

where $x = (x_1, x_2, x_3)$

$$\nabla f(x) = \begin{pmatrix} e^{x_1 - x_2} - e^{x_2 - x_1} + 2x_1 e^{x_1^2} \\ -e^{x_1 - x_2} + e^{x_2 - x_1} \\ 2x_3 \end{pmatrix}$$

and

$$Hf(x) = \begin{pmatrix} e^{x_1-x_2} + e^{x_2-x_1} + 4x_1^2 e^{x_1^2} + 2e^{x_1^2} & -e^{x_1-x_2} - e^{x_2-x_1} & 0 \\ -e^{x_1-x_2} - e^{x_2-x_1} & e^{x_1-x_2} + e^{x_2-x_1} & 0 \\ 0 & 0 & 2 \end{pmatrix}$$

If f is twice continuously differentiable at every point in an open set $D \subset \mathbb{R}^n$, then f is said to be twice continuously differentiable on D and is denoted by $f \in C^{(2)}(D)$.

Definition 3.3 *Let $f: \mathbb{R}^n \longrightarrow \mathbb{R}$ be continuously differentiable on an open set $D \in \mathbb{R}^n$. Then for $x \in D$ and $d = y - x \in \mathbb{R}^n$, the directional derivative of f at x in the direction d is defined as*

$$f'(x;d) \overset{\text{def}}{=} \lim_{\theta \to 0} \frac{f(x+\theta d) - f(x)}{\theta} = \nabla f(x)^T d,$$

where $\nabla f(x)$ is the gradient of f at x, an $n \times 1$ vector.
For any $x, y \in D$, if $f \in C^1(D)$, then

$$f(y) = f(x) + \int_x^y \nabla f(z) dz. \qquad (13)$$

Thus

$$f(y) = f(x) + \nabla f(z)^T (y - x), \qquad z \in (x, y). \qquad (14)$$

Similarly, for $x, y \in D$, with $\theta \in (0,1)$ we have

$$f(y) = f(x) + \nabla f(x + \theta(y - x))^T (y - x), \qquad (15)$$

or

$$f(y) = f(x) + \nabla f(x)(y - x) + o(\| y - x \|) \qquad (16)$$

Definition 3.4 *Let $F: D \subset \mathbb{R}^n \to \mathbb{R}^n$ and $D_0 \subset D$, then F is monotone on D if for any $x, y \in D_0$*

$$\left(F(x) - F(y)\right)^T (x - y) \geq 0, \qquad (17)$$

and *F is strictly monotone on D if for any $x, y \in D_0$ $x \neq y$*

$$\left(F(x) - F(y)\right)^T (x - y) > 0. \qquad (18)$$

Definition 3.5 *A symmetric matrix A is called:*

(*i*) *Positive definite if for all $x \neq 0$ $x^T Ax > 0$. (If $-A$ is positive definite, we say A is negative definite).*

(*ii*) *Positive semidefinite (or nonnegative definite) if for all $x \neq 0$ $x^T Ax \geq 0$. (If $-A$ is nonnegative definite, that is $x^T Ax \leq 0$ for all x, we say that A is negative semidefinite or nonpositive definite)*[7,8].

4.0 Characterizations of Convex Functions

Now consider the following characterizations which help in defining convexity.

Theorem 4.1 First Order Characterization of Convex Functions

Let $D \subset \mathbb{R}^n$ be a nonempty open convex set and $f: D \to \mathbb{R}$ be a differentiable function, then:

(*i*) *f is convex if, and only if, for any $x, y \in D$*

$$f(y) \geq f(x) + \nabla f(x)^T (y - x) \qquad (19)$$

(*ii*) *f is strictly convex on D if, and only if, for any $x, y \in D$ with $x \neq y$*

$$f(y) > f(x) + \nabla f(x)^T (y - x), \quad [4,7,11]. \qquad (20)$$

Theorem 4.2 Monotonicity Characterization of Convex Function

Assume that $f: C \subset \mathbb{R}^n \longrightarrow \mathbb{R}$ is differentiable on the convex set C. Then f is convex on C if and only if its gradient ∇f is monotone, that is

$$\left(\nabla f(x_2) - \nabla f(x_1)\right)^T (x_2 - x_1) \geq 0 \qquad (21)$$

for any $x_1, x_2 \in C$.

f is strictly convex on C if and only if its gradient ∇f is strictly monotone, that is

$$\left(\nabla f(x_2) - \nabla f(x_1)\right)^T (x_2 - x_1) > 0 \qquad (22)$$

for any $x_1, x_2 \in C, x_1 \neq x_2$ [10].

Although these characterizations define convexity, characterizations which combine them will place us at a better view-point. This will enhance easy recognition of convex functions. This is achieved in the following result [11].

Theorem 4.3 Characterization of Convex Functions through the Derivative and the Integral
Suppose $f: D \subset \mathbb{R}^n \to \mathbb{R}$, $f \in C^1(D)$ *and* $D \neq \emptyset$ *is open and convex, then the following statements are equivalent.*
(i) f *is convex.*
(ii) ∇f *is monotone.*
(iii) $f(y) - f(x) = \int_x^y \nabla f(z)dz$, $z \in D$.
(iv) $f(y) \geq f(x) + \nabla f(z)^T(y - x)$.

An important inference from the last result is that a presentation with any of (*i*) to (*iv*) without a pre-information on the nature of the function does not only imply that the function is convex but also a presentation with (*i*) to (*iv*).

The next two results involve the second order (Hessian) and epigraph characterizations of convex functions. They will be used to extend Theorem 4.4.

Theorem 4.4 Second Order Characterization of Twice Continuously Differentiable Convex Functions
Let $D \subset \mathbb{R}^n$ *be a nonempty open convex set and let* $f: D \to \mathbb{R}$ *be a twice continuously differentiable function, then* f *is convex if, and only if, its Hessian matrix is positive semi-definite at each point in D* [10].

Theorem 4.5 Epigraph Characterization of Convex Functions
Let $D \subset \mathbb{R}^n$ *be a nonempty convex set, and* $f: D \to \mathbb{R}$. *Then* f *is convex if, and only if the epigraph of* f *is a convex set* [5, 10].

Remark 4.6
Many results for convex functions can be proved (or interpreted) geometrically using epigraphs, and applying results for convex sets. As an example, consider the First Order Condition for convexity:

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

where f is convex and $x, y \in D$. We can interpret this basic inequality geometrically in terms of epi$f$: If $(y, t) \in$ epi$f$, then

$$t \geq f(y) \geq f(x) + \nabla f(x)^T(y - x). \quad (23)$$

We can express this as

$$(y, t) \in \text{epi}f \Longrightarrow \begin{bmatrix} \nabla f(x) \\ -1 \end{bmatrix}^T \left(\begin{bmatrix} y \\ t \end{bmatrix} - \begin{bmatrix} x \\ f(x) \end{bmatrix} \right) \leq 0 \quad (24)$$

This means that the hyperplane defined by $(\nabla f(y), -1)$ supports epi$f$ at the boundary point $(x, f(y))$.

5.0 Characterization of Convex Functions Through the Epigraph, Gradient, Integral, Monotonicity and Hessian.

We now present a characterization of convex functions which extends the main result in [9] to the Hessian and epigraph.

Theorem 5.1 *Suppose* $f: D \subset \mathbb{R}^n \to \mathbb{R}$, $f \in C^2(D)$ *and* $D \neq \emptyset$ *is open and convex, then the following statements are equivalent:*

(*i*) $f(\theta x + (1-\theta)y) \leq \theta f(x) + (1-\theta)f(y), \quad \theta \in (0,1)$ (25)
(*ii*) $(\theta x + (1-\theta)y,\ \theta\alpha_1 + (1-\theta)\alpha_2) \in epif$ (26)
for $x,\ y \in D$ *and* $(x, \alpha_1), (y, \alpha_2) \in epif$.
(*iii*) $f(y) \geq f(x) + \nabla f(x)^T(y - x)$, (27)
(*iv*) $\left(\nabla f(x) - \nabla f(y)\right)^T(x - y) \geq 0, \quad x, y \in D$ (28)
(*v*) $f(y) - f(x) = \int_x^y \nabla f(z)dz$, $z \in D$. (29)
(*vi*) $v^T Hf(z)v \geq 0$, $z \in D$, $\forall\ v \in \mathbb{R}^n$. (30)

Proof: We shall show that (*i*) ⟹ (*ii*) ⟹ (*iii*) ⟹ (*iv*) ⟹ (*v*) ⟹ (*vi*) ⟹ (*i*)
(*i*) ⟹ (*ii*).Suppose f is convex and let $x, y \in D$ and $(x, \alpha_1), (y, \alpha_2) \in$ epi$f$, then for any $\theta \in (0,1)$ we have

$$f(\theta x + (1-\theta)y) \leq \theta f(x) + (1-\theta)f(y) \leq \theta\alpha_1 + (1-\theta)\alpha_2 \quad (31)$$

Since D is a convex set $(\theta x + (1-\theta)y) \in D$. Therefore

$$(\theta x + (1-\theta)y,\ \theta\alpha_1 + (1-\theta)\alpha_2) \in \text{epi}f \quad (32)$$

which means epi$f$ is convex.
(*ii*) ⟹ (*iii*). Since epi$f$ is convex

$$(\theta y + (1-\theta)x,\ \theta f(y) + (1-\theta)f(x)) \in \text{epi}f$$
$$\Longrightarrow f(\theta y + (1-\theta)x) \leq \theta f(y) + (1-\theta)f(x) = \theta f(y) + f(x) - \theta f(x)$$
$$\frac{f(\theta y + (1-\theta)x) - f(x)}{\theta} \leq f(y) - f(x).$$

As $\theta \to 0$, we have

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

$(iii) \Longrightarrow (iv)$. Now

$$f(y) \geq f(x) + \nabla f(x)^T (y - x) \qquad (33)$$

so that

$$f(x) \geq f(y) + \nabla f(y)^T (x - y). \qquad (34)$$

Summing (33) and (34) we have

$$\left(\nabla f(x) - \nabla f(y)\right)^T (x - y) \geq 0. \qquad (35)$$

$(iv) \Longrightarrow (v)$. Since ∇f is monotone, for any $x, y, z \in D$ with

$$z = x + \theta(y - x), \qquad \theta \in (0,1),$$

we have

$$\left(\nabla f(y) - \nabla f(z)\right)^T (y - z) \geq \left(\nabla f(z) - \nabla f(x)\right)^T (z - x) \geq 0$$
$$\Longrightarrow \nabla f(y)^T (y - z) + \nabla f(x)^T (z - x) \geq \nabla f(z)^T (y - z) + \nabla f(z)^T (z - x) \geq 0$$
$$\Longrightarrow f(y) - f(z) + o(\|y - z\|) + f(z) - f(x) - o(\|z - x\|) \geq \nabla f(z)^T (y - z)$$

By the mean value theorem

$$f(y) - f(x) \geq \int_x^y \nabla f(z) dz = f(y) - f(x)$$
$$f(y) - f(x) = \int_x^y \nabla f(z) dz. \qquad (36)$$

$(v) \Longrightarrow (vi)$. Since $f(z)$ is twice continuously differentiable at z, it follows that

$$f(y) = f(x) + \nabla f(x)^T (y - x) + \frac{1}{2}(y - x)^T Hf(z)(y - x),$$
$$f(y) - f(x) = \nabla f(x)^T (y - x) + \frac{1}{2}(y - x)^T Hf(z)(y - x) \quad \text{[By (v)]}$$
$$\nabla f(z)^T (y - x) = \nabla f(x)^T (y - x) + \frac{1}{2}(y - x)^T Hf(z)(y - x)$$
$$\nabla f(z)^T (y - z) - \nabla f(z)^T (x - z) = \nabla f(x)^T (y - x) + \frac{1}{2}(y - x)^T Hf(z)(y - x)$$
$$f(y) - f(z) - o(\|y - z\|) - \{f(x) - f(z) - o(\|x - z\|)\}$$
$$= f(y) - f(z) - o(\|y - x\|) + \frac{1}{2}(y - x)^T Hf(z)(y - x)$$
$$\Longrightarrow \qquad \frac{1}{2}(y - x)^T Hf(z)(y - x) = o(2(1 - \theta)\|y - x\|) \geq 0$$

since $(1 - \theta) > 0$,

$$\Longrightarrow \frac{1}{2}(y - x)^T Hf(z)(y - x) \geq 0. \qquad (37)$$

$(vi) \Longrightarrow (i)$. Now

$$f(y) = f(x) + \nabla f(x)^T (y - x) + \frac{1}{2}(y - x)^T Hf(z)(y - x)$$
$$\theta f(y) = \theta f(z) + \theta \nabla f(z)^T (y - z) + \frac{1}{2}\theta (y - z)^T Hf(z)(y - z)$$
$$(1 - \theta) f(x) = (1 - \theta) f(z) + (1 - \theta) \nabla f(z)^T (x - z)$$
$$+ \frac{1}{2}(1 - \theta)(x - z)^T Hf(z)(x - z)$$

Since $Hf(z)$ is positive semidefinite, at each point in D

$$\theta f(y) + (1 - \theta) f(x) \geq f(z) + \nabla f(z)^T [\theta y - \theta z + (1 - \theta)(x - z)]$$
$$= f(\theta y + (1 - \theta) x). \qquad (38)$$

Thus we are done.

Remark 5.2

This result shows the inter-connectivity among convexity and some useful concepts in optimization theory. We observe that any convex function can be described by an epigraph which consists of a set of points above the graph as well as fit exactly into the graph of the function which in this case (of convex function) is convex. A geometric interpretation of the third statement is that at any point the linear approximation based on a local derivative is a lower estimate of the function.

That is, a function whose epigraph is convex always lie above its tangent at any point. Such a tangent is called a supporting hyperplane of the convex function. It is a known fact that the gradient of a real-valued convex function of single variable is non-decreasing (monotone). This is also the case with several-variable function as indicated (in the fourth statement) that if a function always lie above or along its tangent at any point then the function is monotone. In fact, for the directional derivatives, this fact can also be proved to hold for non-differentiable convex functions. The integral of this gradient between any two points along any line is the same as the difference in the value of the function between the two points along this line. Finally the last statement can be interpreted geometrically as a requirement that the graph of the function have positive (upward) curvature at each point of the domain of the function. Thus this characterization gives equivalence for the definitions of convexity for real-valued convex functions of several variables.

6.0 Conclusion

As indicated above any function which satisfies any of the defining properties above equally satisfies all the other properties. Thus we are spared the problem of much guessing about the convexity of a function since we need not depend only on the conventional definitions. This result shows that if a given definition cannot be incorporated into a given scheme we can resort to another. Thus it serves as a safe haven for many computational schemes. It further suggests that a given scheme can be refined to incorporate a desired definition or concept. This results from the interplay among these concepts.

As indicated earlier these defining properties of convex functions already exist in optimization materials, however this characterization which combines these properties, thereby giving a wider definition of convexity has not been achieved. Thus this work places us at a better horizon for recognizing convex functions.

References

[1] Avriel, M., Diewert, W., Schaibe, S., and Zang, I. (1988), *Generalized Concavity*, Plenum Press.
[2] Berkovitz, L. D. (2002), *Convexity and optimization in* $\mathbb{R}^n$, John Wiley and Sons, Inc. New York.
[3] Cambini, A. and Martein, L. (2009), *Generalized Convexity and optimization: Theory and Applications*, Springer-Verlag, Berlin Heidelberg.
[4] Simon, B. (2011), *Convexity: An Analytic Viewpoint*, Cambridge University Press, New York.
[5] Boyd, S. and Vandenberg, L. (2004), *Convex Optimization*, Cambridge University Press.
[6] Jahn, J. (2007), *Introduction to the Theory of Nonlinear Optimization*, Springer-Verlag Berlin Heidelberg.
[7] Hiriart-Urruty, J.-B. and Lemarechal, C.(2001), *Fundamentals of Convex analysis*, Springer.
[8] Barvinok, A. (2002), *A Course in Convexity*, Volume 54 of Graduate Studies in Mathematics, American Mathematical Society.
[9] Hormander, L. (1994), *Notion of Convexity*, Birkhauser, Boston, (1994).
[10] Sun, W and Yuan, Y.(2006), *Optimization Theory and Method*, Springer Science+Business Media, LLC.
[11] Ezimadu, P. E. and Igabari, J. N. (2008), *On a Differential and Integral Characterization of Real-valued Convex Functions of Several Variables*, Journal of the Nigerian Association of Mathematical Physics, 13:83-86.