Team 2: Jackson Lyons, [redacted]
MTH 320: Analysis I
April 26, 2023

# Project 5
# Exploring continuity and differentiability of convex functions

## Definitions:

### Convex Sets:

A set $C \subset \mathbb{R}$ is a convex set if it includes the line segment joining any two of its points. That is, $C$ is convex if for every real $a$ with $0 \leq a \leq 1$ and every $x, y \in C$,

$$(1-a)x + ay \in C.$$

If $a = 0$ then $(1-a)x + ay = x$ and if $a = 1$ then $(1-a)x + ay = y$, so $C \subset k(C \times C \times [0,1])$. Thus $C$ is convex $\iff$ $k(C \times C \times [0,1]) = C$ $\iff$ $k(C \times C \times (0,1)) \subset C$. Note that the empty set is convex. [3]

### Convex Functions:

A function $f : I \to \mathbb{R}$ is convex if whenever $a \leq x \leq b$ for $a, x, b \in I$ we have $f(x) \leq (f(a)\frac{b-x}{b-a} + f(b)\frac{x-a}{b-a})$ In other words, if the line drawn between $(a, f(a))$ and $(b, f(b))$ is above the graph of $f$. [1]

Alternative Definition: A function $f : \mathbb{R}^n \to \mathbb{R}$ is convex if its domain is a convex set and for all $x, y$ in its domain, and all $\lambda \in [0,1]$, we have

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

The function $f$ is strictly convex if

$$f(\lambda x + (1-\lambda)y) < \lambda f(x) + (1-\lambda)f(y)$$

And is strongly convex if $\exists a > 0$ such that $f(x) - a||x||^2$ is convex. A function $f$ is convex if it is strictly convex, and a function is strictly convex if it is strongly convex. [2]

Examples on $\mathbb{R}$:
- $e^{ax} \quad \forall a \in \mathbb{R}$
- $-log(x)$
- $|x|^p \quad \forall p \geq 1$

Examples on $\mathbb{R}^n$:
- $f(x) = a^T x + b \quad \forall a \in \mathbb{R}^n, b \in \mathbb{R}$
- $||x||_p = (\sum_{i=1}^n |x_i|^p)^{1/p} \quad \forall p \geq 1$

## Continuity:

**Theorem.** Let $C \subseteq \mathbb{R}$ be an open interval. Then $f : C \to \mathbb{R}$ is convex if and only if for any $a, b, c \in C$, with $a < b < c$,

$$\frac{f(b)-f(a)}{b-a} \leq \frac{f(c)-f(b)}{c-b},$$

and

$$\frac{f(b)-f(a)}{b-a} \leq \frac{f(c)-f(a)}{c-a}.$$

Proof. Take some $a, b, c \in C$ such that $a < b < c$. Since $b-a$ and $c-b > 0$, then our first inequality holds if and only if $[f(b)-f(a)](c-b) \leq [f(c)-f(b)](b-a)$, which holds if and only if $f(b)(c-a) \leq f(a)(c-b)+f(c)(b-a)$, which holds if and only if $f(b) \leq \frac{c-b}{c-a}f(a)+\frac{b-a}{c-a}f(c)$. Note that this is our first definition of convexity, so if the first inequality holds, then $f$ is convex, and vice versa. Similarly, the second inequality holds if and only if $f(b) \leq \frac{b-a}{c-a}f(c)+\frac{c-b}{c-a}f(a)$. Again, if $f$ is convex, this holds, and vice versa. $\square$ [4]

**Theorem.** Let $C \subseteq \mathbb{R}^n$ be non-empty, open, and convex and let $f : C \to \mathbb{R}$ be convex on $C$. Then $f$ is continuous on $C$. [4]

Proof. Let $f$ be convex, and $a, b, c \in C$ with $a < b < c$. We can take any element in a closed sub-interval $x \in [a, b]$, and by our previous theorem,

$$f(b) - \frac{f(b)-f(a)}{b-a}(b-x) \geq f(x) \geq f(b) - \frac{f(c)-f(b)}{c-b}(b-x),$$

and for any $y \in [b, c]$,

$$f(b) + \frac{f(b)-f(a)}{b-a}(x-b) \leq f(y) \leq f(b) + \frac{f(c)-f(b)}{c-b}(x-b).$$

These imply that $f$ is continuous at $b$, and since $a < b < c$, $f$ is continuous on the interval $(a, b)$. $\square$ [4]

**Theorem.** Let $f : \mathbb{R}^n \to (-\infty, \infty]$ be convex with $x \in \mathbb{R}^n$. If $f$ is bounded above on $\mathbb{B}(x; \delta)$ for some $\delta > 0$, then $f$ is Lipschitz continuous on $\mathbb{B}(x; \frac{\delta}{2})$. [5]

We will skip the proof of this theorem. This result is interesting, however, as it allows us to state and prove an important theorem for defining Lipschitz continuity of convex functions.

**Theorem.** Let $f : \mathbb{R}^n \to (-\infty, \infty]$ be a convex function such that $\text{int}(\text{dom}f) \neq \varnothing$. Then $f$ is locally Lipschitz continuous on the interior of its domain $\text{int}(\text{dom}f)$. [5]

Proof. First, define $A := \{\bar{x} \pm \varepsilon e_i\}$, where $\{e_i \mid i = 1, ..., n\}$ is the standard orthonormal basis of $\mathbb{R}^n$. We can pick some $\bar{x} \in \text{int}(\text{dom}f)$ and choose some $\varepsilon > 0$ such that $\bar{x} \pm \varepsilon e_i \in \text{dom}f$ for each $i$. Let $M := \max\{f(a) \mid a \in A\}$ be finite. We express

$$x = \sum_{i=1}^{m} \lambda_i a_i \text{ with } \lambda_i \geq 0, \sum_{i=1}^{m} \lambda_i = 1, a_i \in A$$

for any $x \in \mathbb{B}(\bar{x}, \frac{\varepsilon}{n})$, which shows

$$f(x) \leq \sum_{i=1}^{m} \lambda_i f(a_i) \leq \sum_{i=1}^{m} \lambda_i M = M,$$

so $f$ is bounded above on $\mathbb{B}(\bar{x}, \frac{\varepsilon}{n})$. Hence, by our previous theorem, $f$ is Lipschitz continuous on $\mathbb{B}(\bar{x}, \frac{\varepsilon}{2n})$, and locally Lipschitz continuous on $\text{int}(\text{dom}f)$. $\square$ [5]

**Corollary.** Let $f : \mathbb{R}^n \to (-\infty, \infty]$ be a convex function and let $x \in \text{intdom}f$. Then the following are equivalent:

$(i)$ $f$ is continuous at $x$

$(ii)$ $x \in \text{int}(\text{dom}f)$

$(iii)$ $f$ is locally Lipschitz continuous around $x$ [5]

This is a direct consequence of the previous theorem, and as such, the proof is trivial. With this corollary we can say that, for any convex function $f : C \to \mathbb{R}$ where $C$ is an open interval, $f$ is Lipschitz continuous on $C$, and is consequently uniformly continuous and continous on $C$.

**Differentiability:**

**Theorem** (Mean Value Theorem) Let $f : [a, b] \to \mathbb{R}$ be a continuous function differentiable on $(a, b)$. Then there exists a point $c \in (a, b)$ such that

$$f(b) - f(a) = f'(c)(b - a)$$

Proof. Define the function $g : [a, b] \to \mathbb{R}$ by

$$g(x) := f(x) - f(b) - \tfrac{f(b)-f(a)}{b-a}(x - b)$$

The function $g$ is differentiable on $(a, b)$, continuous on $[a, b]$, such that $g(a) = 0$ and $g(b) = 0$. So $\exists c \in (a, b)$ such that $g'(c) = 0$, that is,

$$0 = g'(c) = f'(c) - \tfrac{f(b)-f(a)}{b-a} \implies f'(c) = \tfrac{f(b)-f(a)}{b-a}$$
$$\implies f'(c)(b - a) = f(b) - f(a)$$

□ [1]

**Theorem** Let $f$ be differentiable in I. It is convex if and only if $f'$ is increasing [8]

Proof. "$\implies$" Let $f$ be a convex function in $I$. Let $x_1, x_2, x_3, x_4 \in I$ such that $x_1 < x_2 < x_3 < x_4$. Next, by definition of convex functions,

$$\tfrac{f(x_2)-f(x_1)}{x_2-x_1} \leq \tfrac{f(x_3)-f(x_2)}{x_3-x_2} \leq \tfrac{f(x_4)-f(x_3)}{x_4-x_3}$$

By ignoring the middle term and letting $x_2 \to x_1^+$ and $x_3 \to x_4^-$, we get that $f'(x_1) \leq f'(x_4)$ Thus $f'$ is increasing on I.

"$\impliedby$" Let $f'$ be increasing on $I$. Let $x_1, x_2, x_3 \in I$ such that $x_1 < x_2 < x_3$. Applying the mean value theorem, we get that:

$$\exists a : \tfrac{f(x_2)-f(x_1)}{x_2-x_1} = f'(a)$$

and

$$\exists b : \tfrac{f(x_3)-f(x_2)}{x_3-x_2} = f'(b)$$

where $x_1 < a < x_2 < b < x_3$. $f'$ increasing $\implies f'(a) \leq f'(b)$. Therefore,

$$\tfrac{f(x_2)-f(x_1)}{x_2-x_1} \leq \tfrac{f(x_3)-f(x_2)}{x_3-x_2}$$

So by definition, $f$ is convex. □ [9]

**Theorem** A twice differentiable function $f : C \to \mathbb{R}$ is convex if and only if $f''(x) \geq 0, \forall x \in C$. [2]

Proof. This follows directly from the previous theorem. Let $f : C \to \mathbb{R}$ be a twice differentiable function. If $f$ is convex, then by the previous theorem, $f'$ is increasing $\implies f''(x) \geq 0 \quad \forall x \in C$. Suppose $f''(x) \geq 0 \quad \forall x \in C$ this implies that $f'$ is increasing. So by the previous theorem, f is convex. □ [8]

**Application:**

**Gradient descent** is an iterative optimization algorithm used to find a local minimum of a function. This algorithm is commonly used in Machine Learning to minimize a cost function (cost functions are explained further on page 4).

The general idea of gradient descent is to iteratively approach the point at which $f$ achieves its minimum value by moving in the direction of the negative gradient $(-\nabla f(x))$. The gradient $(\nabla f(x))$ is the direction in which the function $f(x)$ grows fastest. So the negative gradient is the direction in which $f(x)$ decreases fastest. This algorithm can be written as the following:

$$x_{t+1} = x_t - t_k \nabla f(x_t)$$

Where $t_k$ is the step size. [6]

**Theorem** For differentiable convex functions, the following three properties are equivalent:
$(i)$ $x$ is a local minimum of $f(x)$
$(ii$ $x$ is a global minimum of $f(x)$
$(iii)$ $\nabla f(x) = 0$

So when $f$ is convex, the stopping point for gradient descent would be when $\nabla f(x) = 0$. But for most functions, the gradient won't get to be exactly 0 in a reasonable time. So generally the point to stop further minimizing would be when the gradient $\nabla f(x)$ is sufficiently close to zero. More specifically when $||\nabla f(x)|| < \epsilon$. Common values for $\epsilon$ are around .01. [7]

As mentioned previously, gradient descent is commonly used in Machine Learning to minimize a cost function. Cost functions (also called Loss functions) evaluate the accuracy of a machine learning model given a set of data. In other words, the cost function explains the difference between the predicted value and the actual value. High accuracy is crucial for Machine Learning models so it is important to minimize the cost function.

One of the most popular cost functions used in Machine Learning is **Mean Squared Error (MSE)**. This function measures the average squared difference between actual and predicted values. One important factor about the MSE cost function is that it is **convex**. This means that the local minimum is the same as the global minimum. So when using gradient descent to minimize MSE, it will converge to the global minimum. [10]

**References:**

[1] Lebl Jiří. Basic Analysis: Introduction to Real Analysis. Jiří Lebl, 2016.

[2] Ahmadi, A. A., and G. Hall. " Theory of Convex Functions." Princeton University ORF 523. Lecture, March 1, 2016.
https://www.princeton.edu/~aaa/Public/Teaching/ORF523/S16/ORF523_S16_Lec7_gh.pdf

[3] Border, KC. "Convex Sets and Functions." California Institute of Technology Ec 181: Convex Analysis and Economic Theory. 2020. https://healy.econ.ohio-state.edu/kcb/Ec181/Lecture01.pdf

[4] Nachbar, John. "Concave and Convex Functions." Washington University. 2019.
https://math.ucr.edu/ res/math153-2019/concavity.pdf

[5] Nam Nguyen. "11 Continuity of Convex Functions." Convex Analysis, May 31, 2016.
https://maunamn.wordpress.com/11-continuity-of-convex-functions/.

[6] Alex. "Gradient Descent for Convex Optimization: The Basic Idea." Boostedml, October 30, 2021.
https://boostedml.com/2020/06/gradient-descent-for-convex-optimization-the-basic-idea.html

[7] Paul, Michael. "Optimization and Gradient Descent." INFO-4604, Applied Machine Learning University of Colorado Boulder. Lecture, September 11, 2018.
https://cmci.colorado.edu/classes/INFO-4604/files/slides-4_optimization.pdf

[8] "Convex Functions." Math 2060b. Lecture, 2015.
https://www.math.cuhk.edu.hk/course_builder/1415/math2060b/Notes%201%20Convex%20Functions%202015.pdf.

[9] Binmore, K. (1977) "Monotone Functions: Theorem 12.18," in Mathematical analysis: A straightforward approach. Cambridge: Cambridge Univ. Pr.

[10] Yashwanth, N.V.S. (2020) Why convexity is the key to optimization, Medium. Towards Data Science.
https://towardsdatascience.com/understand-convexity-in-optimization-db87653bf920#:~:text=MSE%20is%20a%20convex%20function,converge%20to%20the%20global%20minimum.