Lecture Notes Supplement: Convex Functions

MATH XXX - Course Name
(Rewritten based on lecture of [Original Lecture Date])

April 22, 2025

Administrative Notes:

Welcome back, everyone! Before we dive into today's fascinating topic, a few quick notes:

- **Homework X:** Due [Date] at [Time] via [Submission Method]. Remember to start early!

- **Office Hours:** My office hours this week are [Time] on [Day] in [Location]. [TA Name]'s office hours are [Time] on [Day] in [Location].

- **Midterm Exam:** Scheduled for [Date] during class time in [Location]. More details on coverage next week.

- **Reading:** For this topic, please refer to Chapter Y, Section Z in the textbook. These notes supplement the reading.

(Note: Please replace the bracketed information above with the actual details from the original lecture if they were provided. If no administrative details were given in the original source, this section serves as a template placeholder.)

---

1 Convexity: Definition and Geometric Intuition

Today, we explore a fundamental property of functions known as *convexity*. You might have an intuitive sense of what "convex" means from geometry (like a convex lens or a convex polygon). In the context of functions, convexity captures a specific kind of "bending" behavior of the graph. Why is this important? Convex functions appear everywhere – in optimization problems (where they guarantee that a local minimum is also a global minimum), in probability theory (leading to powerful inequalities like Jensen's), and in many areas of analysis and geometry. Understanding convexity provides us with powerful tools for analyzing functions.

Let's formalize this idea.

Definition 1.1 (Convex Function (Def 1.12))**.** Let $I \subseteq \mathbb{R}$ be an interval. A function $g : I \to \mathbb{R}$ is called **convex** on I if for all $x_1, x_2 \in I$ and for all $\lambda \in [0, 1]$, the following inequality holds:

$$g(\lambda x_1 + (1 - \lambda)x_2) \leq \lambda g(x_1) + (1 - \lambda)g(x_2)$$

If the inequality holds strictly ($<$) for all $x_1 \neq x_2$ and $\lambda \in (0, 1)$, the function is called **strictly convex**.

A function g is called **concave** on I if the inequality is reversed ($\geq$), and **strictly concave** if the inequality is strictly reversed ($>$) for $x_1 \neq x_2$ and $\lambda \in (0, 1)$.

Remark 1.2. Notice that g is concave if and only if $-g$ is convex. This simple observation means that any property or theorem we establish for convex functions has a direct analogue for concave functions, usually just by flipping signs or inequalities appropriately. Therefore, we will often focus our attention on convex functions.

Geometric Interpretation of the Definition: What does the inequality in Definition 1.1 really say? The expression $\lambda x_1 + (1 - \lambda)x_2$ represents any point between x_1 and x_2 (inclusive) as λ varies from 0 to 1. The point on the graph corresponding to this x-value is $(\lambda x_1 + (1 - \lambda)x_2, g(\lambda x_1 + (1 - \lambda)x_2))$.

On the right side, the expression $\lambda g(x_1) + (1 - \lambda)g(x_2)$ represents the y-value on the line segment (the "chord") connecting the points $(x_1, g(x_1))$ and $(x_2, g(x_2))$ on the graph of g, evaluated at the same x-coordinate $\lambda x_1 + (1 - \lambda)x_2$.

So, the definition $g(\lambda x_1 + (1 - \lambda)x_2) \leq \lambda g(x_1) + (1 - \lambda)g(x_2)$ means that for any two points on the graph of a convex function, the graph of the function between these points lies *below or on* the chord connecting them. Picturing this helps build intuition!

2 An Alternative Characterization of Convexity

The definition is fundamental, but sometimes another characterization is more useful, especially when dealing with differentiable functions. The following lemma provides such an alternative viewpoint.

Lemma 2.1 (Characterization via Linear Support (Lemma 1.13)). *Let $I \subseteq \mathbb{R}$ be an interval and $g : I \to \mathbb{R}$. The function g is convex on I if and only if for every $x \in I$, there exists a real number $v(x)$ such that for all $y \in I$,*

$$g(y) \geq g(x) + v(x)(y - x)$$

Geometric Interpretation of the Lemma: The expression $L_x(y) = g(x) + v(x)(y - x)$ defines a line that passes through the point $(x, g(x))$ on the graph of g. The lemma states that g is convex if and only if, for any point x on the graph, we can find a line passing through $(x, g(x))$ (with slope $v(x)$) such that the entire graph of g lies *above or on* this line. Such a line is often called a supporting line. If g is differentiable at x, we will see that the natural choice for $v(x)$ is the derivative $g'(x)$, meaning a convex function lies above all of its tangent lines.

Proof of Lemma 2.1. We need to prove both directions of the "if and only if" statement.

($\implies$) Assume g is convex. We need to show that for any $x \in I$, such a $v(x)$ exists. Let $x \in I$ be fixed. Consider any $y \in I$ with $y \neq x$. The definition of convexity involves the inequality stating that secant lines lie above the graph. More fundamentally, for a convex function, the slopes of secant lines starting at x are non-decreasing as the other endpoint moves away from x. That is, for $y_1 < y_2 < x < y_3 < y_4$ in I, we have

$$\frac{g(y_2) - g(y_1)}{y_2 - y_1} \leq \frac{g(x) - g(y_1)}{x - y_1} \leq \frac{g(y_3) - g(x)}{y_3 - x} \leq \frac{g(y_4) - g(x)}{y_4 - x}$$

This implies that the limit of the slopes of secants from the left and the limit from the right both exist, although they might not be equal if g is not differentiable at x. Let $g'_-(x)$ and $g'_+(x)$ denote

the left-hand and right-hand derivatives, respectively. It's a property of convex functions that $g'_-(x) \leq g'_+(x)$ for x in the interior of I. Furthermore, for any $y < x < z$ in I,

$$\frac{g(x) - g(y)}{x - y} \leq g'_-(x) \leq g'_+(x) \leq \frac{g(z) - g(x)}{z - x}$$

We can choose *any* value $v(x)$ between the left and right derivatives (inclusive), i.e., $v(x) \in [g'_-(x), g'_+(x)]$. For instance, let's choose $v(x) = g'_+(x)$ (if x is not the right endpoint) or $v(x) = g'_-(x)$ (if x is not the left endpoint). If x is an interior point, both exist, and we can choose either, or any value in between. Let's take $v(x)$ such that $g'_-(x) \leq v(x) \leq g'_+(x)$. If $y > x$, then $\frac{g(y) - g(x)}{y - x} \geq g'_+(x) \geq v(x)$. Since $y - x > 0$, this implies $g(y) - g(x) \geq v(x)(y - x)$. If $y < x$, then $\frac{g(x) - g(y)}{x - y} \leq g'_-(x) \leq v(x)$. Since $x - y > 0$, this implies $g(x) - g(y) \leq v(x)(x - y)$. Rearranging gives $g(y) \geq g(x) - v(x)(x - y) = g(x) + v(x)(y - x)$. In both cases (and trivially if $y = x$), we have $g(y) \geq g(x) + v(x)(y - x)$. Thus, such a $v(x)$ exists. (Note: If g is differentiable at x, then $g'_-(x) = g'_+(x) = g'(x)$, and we must choose $v(x) = g'(x)$.)

($\Longleftarrow$) Assume that for every $x \in I$, there exists $v(x)$ such that $g(y) \geq g(x) + v(x)(y - x)$ for all $y \in I$. We want to show g is convex. Let $x_1, x_2 \in I$ and $\lambda \in [0, 1]$. Let $x = \lambda x_1 + (1 - \lambda)x_2$. By our assumption, there exists $v(x)$ such that the inequality holds for $y = x_1$ and $y = x_2$:

$$g(x_1) \geq g(x) + v(x)(x_1 - x)$$
$$g(x_2) \geq g(x) + v(x)(x_2 - x)$$

Multiply the first inequality by $\lambda \geq 0$ and the second by $(1 - \lambda) \geq 0$:

$$\lambda g(x_1) \geq \lambda g(x) + \lambda v(x)(x_1 - x)$$
$$(1 - \lambda)g(x_2) \geq (1 - \lambda)g(x) + (1 - \lambda)v(x)(x_2 - x)$$

Now, add these two inequalities:

$$\lambda g(x_1) + (1 - \lambda)g(x_2) \geq (\lambda + (1 - \lambda))g(x) + v(x)[\lambda(x_1 - x) + (1 - \lambda)(x_2 - x)]$$

Let's examine the term in the square brackets:

$$\lambda(x_1 - x) + (1 - \lambda)(x_2 - x) = \lambda x_1 - \lambda x + (1 - \lambda)x_2 - (1 - \lambda)x$$
$$= (\lambda x_1 + (1 - \lambda)x_2) - (\lambda + 1 - \lambda)x$$
$$= x - 1 \cdot x$$
$$= 0$$

So, the $v(x)$ term vanishes, and we are left with:

$$\lambda g(x_1) + (1 - \lambda)g(x_2) \geq g(x)$$

Substituting $x = \lambda x_1 + (1 - \lambda)x_2$ back gives:

$$\lambda g(x_1) + (1 - \lambda)g(x_2) \geq g(\lambda x_1 + (1 - \lambda)x_2)$$

This is precisely the definition of convexity (Definition 1.1). $\qquad\square$

3 Convexity and Derivatives

Lemma 2.1 becomes particularly powerful when the function g is differentiable. As hinted in the proof, if g is differentiable at x, the supporting line is simply the tangent line, $v(x) = g'(x)$.

Corollary 3.1 (Convexity and the First Derivative). *Let $g : I \to \mathbb{R}$ be differentiable on the interval I. Then g is convex on I if and only if its derivative g' is non-decreasing on I.*

Proof. ($\implies$) Assume g is convex and differentiable. By Lemma 2.1 with $v(x) = g'(x)$, for any $x, y \in I$, we have $g(y) \geq g(x) + g'(x)(y - x)$. Also, swapping roles of x and y, $g(x) \geq g(y) + g'(y)(x - y)$. Adding these inequalities gives:

$$g(x) + g(y) \geq g(x) + g(y) + g'(x)(y - x) + g'(y)(x - y)$$

This simplifies to $0 \geq g'(x)(y - x) - g'(y)(y - x)$, or

$$0 \geq (g'(x) - g'(y))(y - x)$$

If $y > x$, then $y - x > 0$, so we must have $g'(x) - g'(y) \leq 0$, which means $g'(x) \leq g'(y)$. If $y < x$, then $y - x < 0$, so we must have $g'(x) - g'(y) \geq 0$, which also means $g'(x) \geq g'(y)$. In either case, if $x_1 < x_2$ are points in I, then $g'(x_1) \leq g'(x_2)$. Thus, g' is non-decreasing on I.

($\impliedby$) Assume g' is non-decreasing. We want to show $g(y) \geq g(x) + g'(x)(y - x)$ for all $x, y \in I$. By the Mean Value Theorem, for any $x \neq y$ in I, there exists some c strictly between x and y such that $g(y) - g(x) = g'(c)(y - x)$.

- If $y > x$, then $x < c < y$. Since g' is non-decreasing, $g'(c) \geq g'(x)$. Multiplying by the positive quantity $y - x$, we get $g'(c)(y - x) \geq g'(x)(y - x)$. Therefore, $g(y) - g(x) \geq g'(x)(y - x)$.

- If $y < x$, then $y < c < x$. Since g' is non-decreasing, $g'(c) \leq g'(x)$. Multiplying by the negative quantity $y - x$, we must reverse the inequality: $g'(c)(y - x) \geq g'(x)(y - x)$. Therefore, $g(y) - g(x) \geq g'(x)(y - x)$ holds in this case too.

If $y = x$, the inequality $g(x) \geq g(x) + g'(x)(x - x)$ is trivially true ($g(x) \geq g(x)$). Thus, $g(y) \geq g(x) + g'(x)(y - x)$ holds for all $x, y \in I$. By Lemma 2.1, g is convex. $\square$

If the function is twice differentiable, we get an even simpler condition.

Theorem 3.2 (Convexity and the Second Derivative). *Let $g : I \to \mathbb{R}$ be twice differentiable on the interval I. Then g is convex on I if and only if $g''(x) \geq 0$ for all $x \in I$.*

Proof. This follows almost directly from Corollary 3.1. A differentiable function g' is non-decreasing on an interval I if and only if its derivative, $(g')' = g''$, is non-negative on I. This is a standard result from calculus.

Alternatively, we can directly use the Taylor expansion with integral remainder, as suggested in the original notes. Assume g is twice differentiable. By the Fundamental Theorem of Calculus applied twice:

$$g(y) = g(x) + \int_x^y g'(u)\, du$$
$$= g(x) + \int_x^y \left(g'(x) + \int_x^u g''(v)\, dv \right) du$$
$$= g(x) + g'(x) \int_x^y 1\, du + \int_x^y \left(\int_x^u g''(v)\, dv \right) du$$
$$= g(x) + g'(x)(y - x) + \int_x^y \int_x^u g''(v)\, dv\, du$$

Rearranging gives the expression for the difference between $g(y)$ and its tangent approximation at x:

$$g(y) - \big(g(x) + g'(x)(y - x)\big) = \int_x^y \int_x^u g''(v)\, dv\, du$$

Now, we analyze the sign of the double integral term. If $g''(v) \geq 0$ for all $v \in I$:

- If $y > x$, then u ranges from x to y. For each u in $(x, y]$, v ranges from x to u. Since $g''(v) \geq 0$, the inner integral $\int_x^u g''(v)\, dv \geq 0$. Then, integrating this non-negative function from x to y results in a non-negative value for the double integral.

- If $y < x$, let's rewrite the integration limits to be increasing. Let $F(u) = \int_x^u g''(v)dv = -\int_u^x g''(v)dv$. If $u < x$, then v ranges from u to x, so $\int_u^x g''(v)dv \geq 0$, which means $F(u) \leq 0$ for $u < x$. The double integral is $\int_x^y F(u)du = -\int_y^x F(u)du$. Since $F(u) \leq 0$ on (y, x), the integral $\int_y^x F(u)du \leq 0$. Multiplying by -1 gives a result ≥ 0.

In both cases ($y > x$ and $y < x$), if $g''(v) \geq 0$ for all v between x and y, then the double integral term is non-negative. Therefore, $g(y) - (g(x) + g'(x)(y - x)) \geq 0$, which means $g(y) \geq g(x) + g'(x)(y - x)$. By Lemma 2.1, this implies g is convex.

Conversely, assume g is convex. Since g is differentiable, by Corollary 3.1, g' is non-decreasing on I. Since g is twice differentiable, the derivative of g', which is g'', must be non-negative on I. That is, $g''(x) \geq 0$ for all $x \in I$. $\qquad \square$

This theorem provides a very practical and often easily applicable test for convexity for sufficiently smooth functions.

Example 3.3 (Examples of Convex Functions (Original Examples))**.** Let's use the second derivative test (Theorem 3.2) to verify the convexity of the functions mentioned in the original notes.

1. **The Exponential Function:** Consider $g(x) = e^x$ defined on the interval $I = \mathbb{R}$. We compute the derivatives: $g'(x) = e^x$ $g''(x) = e^x$ Since $e^x > 0$ for all real x, we have $g''(x) \geq 0$ for all $x \in \mathbb{R}$. Therefore, by Theorem 3.2, $g(x) = e^x$ is convex on $\mathbb{R}$. (In fact, since $g''(x) > 0$, it is strictly convex.)

2. **Power Functions:** Consider $g(x) = x^p$. The original notes specify the case $p \geq 2$. We compute the derivatives: $g'(x) = px^{p-1}$ $g''(x) = p(p-1)x^{p-2}$ We need to consider the domain I and the condition $g''(x) \geq 0$.

   - Case 1: $p = 2$. Then $g(x) = x^2$. $g''(x) = 2(1)x^0 = 2$. Since $g''(x) = 2 \geq 0$ for all $x \in \mathbb{R}$, the function $g(x) = x^2$ is convex on $I = \mathbb{R}$.
   - Case 2: $p > 2$.
     - If we consider the domain $I = (0, \infty)$: Here $x > 0$, so x^{p-2} is positive and well-defined. Since $p > 2$, both p and $p - 1$ are positive. Thus, their product $p(p - 1)$ is positive. Therefore, $g''(x) = p(p - 1)x^{p-2} > 0$ for all $x \in (0, \infty)$. So, $g(x) = x^p$ is strictly convex on $(0, \infty)$ for $p > 2$.
     - If p is an integer ≥ 2: $g(x) = x^p$ is defined on $I = \mathbb{R}$. If p is even, then $p - 2$ is an even non-negative integer. Thus $x^{p-2} \geq 0$ for all $x \in \mathbb{R}$. Since $p(p-1) \geq 0$ (positive if $p > 2$), $g''(x) = p(p - 1)x^{p-2} \geq 0$ for all $x \in \mathbb{R}$. So $g(x) = x^p$ is convex on $\mathbb{R}$ for even integers $p \geq 2$. If p is odd and $p \geq 3$, then $p - 2$ is an odd positive integer. Thus x^{p-2} is positive for $x > 0$ and negative for $x < 0$. Since $p(p - 1) > 0$, $g''(x)$ changes sign at $x = 0$. It is positive for $x > 0$ (convex there) and negative for $x < 0$

(concave there). So $g(x) = x^p$ for odd integers $p \geq 3$ is not convex on $\mathbb{R}$ (only on $[0, \infty)$).

Summary for $g(x) = x^p, p \geq 2$: It is convex on $(0, \infty)$. It is convex on $\mathbb{R}$ if p is an even integer. It is convex on $[0, \infty)$ if p is any real number ≥ 2 (or even $p \in [1, 2)$), provided we define derivatives appropriately at $x = 0$ if needed, or just check the definition. The second derivative test $g''(x) = p(p-1)x^{p-2} \geq 0$ is most easily applied on $(0, \infty)$ where x^{p-2} is always positive.

These examples confirm the utility of the second derivative test for establishing convexity.

Remark 3.4 (Further Examples/Non-Examples).
- $g(x) = |x|$ is convex on $\mathbb{R}$ (check Definition 1.1), but not differentiable at $x = 0$. It satisfies Lemma 2.1 (e.g., $v(0)$ can be any value in $[-1, 1]$).

- $g(x) = -\ln(x)$ is convex on $(0, \infty)$. $g'(x) = -1/x$, $g''(x) = 1/x^2 > 0$.

- $g(x) = x^3$ is neither convex nor concave on $\mathbb{R}$ ($g''(x) = 6x$, which changes sign), but it is convex on $[0, \infty)$ and concave on $(-\infty, 0]$.

Remark 3.5. Convexity is a cornerstone concept with far-reaching implications. These equivalent characterizations—the fundamental secant line definition, the supporting line property, the non-decreasing first derivative, and the non-negative second derivative—provide a versatile toolkit for identifying and working with convex functions. We will leverage this understanding, particularly when exploring Jensen's inequality in probability theory and its role in optimization.