IIIT Hyderabad **Scribed By: Shilpak Deb (2024201072)**
- **Digvijay Rathore (2024201078)**

Optimization Methods(CS1.404) **Lecture #8**
Instructor: Dr. Naresh Manwani 13^{th} **Feb 2025**

Outline. *This scribe delves into fundamental properties of convex functions, including the convexity of level sets, the gradient inequality, the sufficiency of stationary points for global optimality, and the monotonicity of gradients.*

1 Convexity of Sublevel Sets Theorem

1.1 Level and Sublevel Sets

Let $S \subseteq \mathbb{R}^n$ and $f : S \to \mathbb{R}$.

Then the *level set* of f corresponding to a threshold $\alpha \in \mathbb{R}$ is defined as follows:

$$L_\alpha = \{x \in \mathbb{R}^n \mid f(x) = \alpha\}.$$

Similarly, the *sublevel set* at level α is given by

$$S_\alpha = \{x \in \mathbb{R}^n \mid f(x) \leq \alpha\}.$$

A key distinction between convex and non-convex functions lies in their sublevel sets. For any value of α, all convex functions have convex sublevel sets, whereas non-convex functions exhibit non-convex sublevel sets.

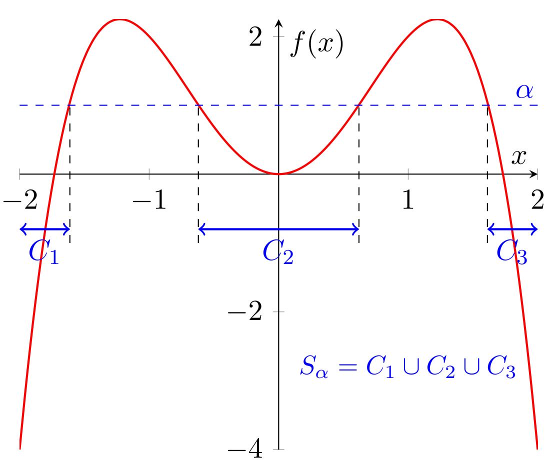

Figure 1: A non-convex function $f(x) = -x^4 + 3x^2$ with non-convex sublevel sets

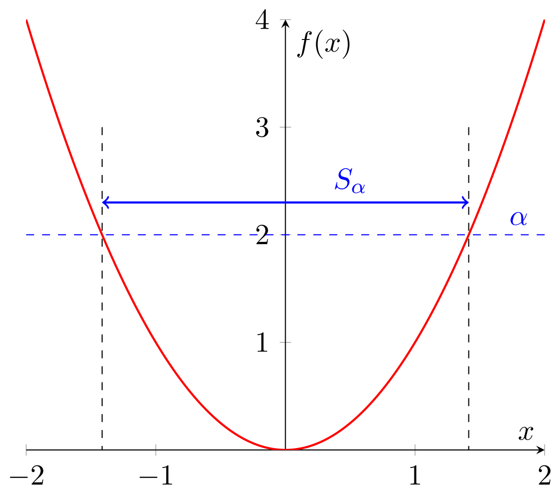

Figure 2: A convex function $f(x) = x^2$ with convex sublevel sets

1.2 Theorem Statement

Theorem: Let $S \subseteq \mathbb{R}^n$ be a convex set and $f : S \to \mathbb{R}$ be a convex function. Then, the *sublevel set* S_α defined by:

$$S_\alpha = \{x \in \mathbb{R}^n \mid f(x) \leq \alpha\}.$$

is also a convex set $\forall \alpha \in \mathbb{R}$.

1.3 Proof

Proof: Let $S \subseteq \mathbb{R}^n$ be a convex set and $f : S \to \mathbb{R}$ be a convex function. Let $\alpha \in \mathbb{R}$ and $S_\alpha = \{x \in \mathbb{R}^n \mid f(x) \leq \alpha\}$ be the *level set* of f corresponding to α.

Let $x_1, x_2 \in S_\alpha$ and $\lambda \in [0, 1]$. Then, due to the convexity of the function f,

$$f(\lambda x_1 + (1 - \lambda)x_2) \;\; \leq \;\; \lambda f(x_1) + (1 - \lambda)f(x_2) \qquad \textbf{...(1)}$$

Since $x_1, x_2 \in S_\alpha$,

$$f(x_1) \leq \alpha \qquad \textit{and} \qquad f(x_2) \leq \alpha$$

Using this in eq. (1),

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

$$\Rightarrow f(\lambda x_1 + (1 - \lambda)x_2) \;\; \leq \;\; \alpha$$

$\therefore \lambda x_1 + (1 - \lambda)x_2 \in S_\alpha, \;\; \forall x_1, x_2 \in S_\alpha$ and $\lambda \in [0, 1]$

Thus, S_α is a convex set. □

Question: Is the converse of the above result true; i.e. for a function f, if all its sub-level sets are convex, then can we say f itself is convex?

Answer: *No.*

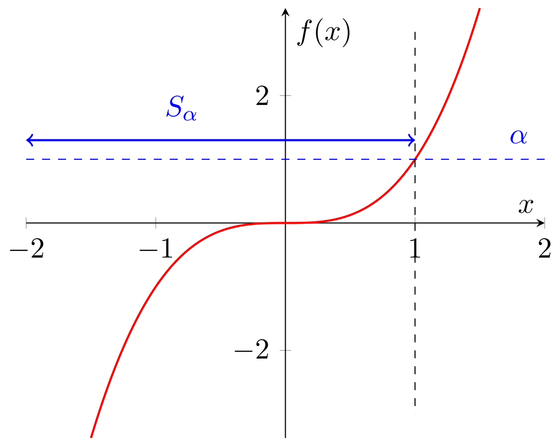

Figure 3: $f(x) = x^3$ has convex sublevel sets but is not convex.

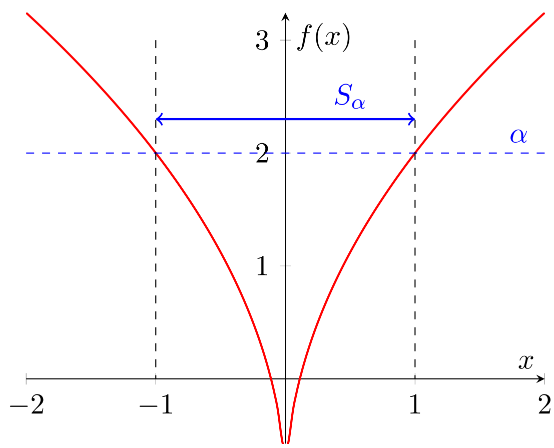

Figure 4: $f(x) = 3|x|^{1/2} - 1$ has convex sublevel sets but is not convex.

The examples above illustrate functions whose sublevel sets S_α are convex for all $\alpha \in \mathbb{R}$, yet the functions themselves are not convex.

2 Gradient Inequality for Convex Functions

2.1 Approximation of a Convex Function at a Point

Let $S \subseteq \mathbb{R}^n$ be a convex set and $f : S \to \mathbb{R}$ be a convex function. Then for a tangent drawn at a particular point $(x, f(x))$, the tangent which can be referred to as $\hat{f}(x; x_1)$, gives a lower bound of $f(x)\ \forall x \in S$, i.e.,

$$\hat{f}(x; x_1) \leq f(x) \qquad \forall x \in S$$

Here $\hat{f}(x; x_1)$ may also be called the approximation of $f(x)$ at x_1. Here x_1 was arbitrarily chosen, thus this property holds true for all points $x_1 \in S$.

Below is an example for a convex scalar function f. *(A scalar function was chosen for ease of visualization and this can be easily extended to higher dimensions)*

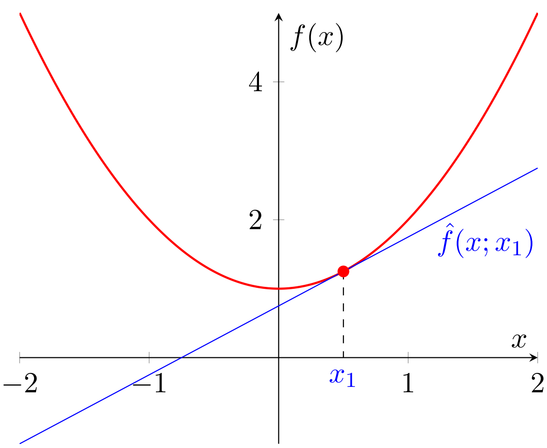

Figure 5: Approximation of a convex function at a particular point.

2.2 Theorem Statement

Theorem: Let $S \subseteq \mathbb{R}^n$ be a convex set and $f : S \to \mathbb{R}$ and $f \in \mathbb{C}^1(S)$. Then f is a convex function if and only if:

$$f(x_2) \geq f(x_1) + \nabla f(x_1)^T (x_2 - x_1) \qquad \forall x_1, x_2 \in S$$

2.3 Proof

Proof: (*Forward*) $\Rightarrow$ Let $S \subseteq \mathbb{R}^n$ be a convex set and $f : S \to \mathbb{R}$ be a convex function such that $f \in \mathbb{C}^1(S)$. Then, for $\forall x_1, x_2 \in S$ and $\lambda \in (0, 1]$,

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

(**Note:** open interval is taken in the left so as to keep λ non-zero)

$$\Rightarrow f(x_1 + \lambda(x_2 - x_1) \ \leq \ f(x_1) + \lambda(f(x_1) - f(x_2))$$

$$\Rightarrow \frac{f(x_1 + \lambda(x_2 - x_1)) - f(x_1)}{\lambda} \ \leq \ f(x_2) - f(x_1)$$

[We can divide by λ since we have taken $\lambda \in (0,1]$]

$\therefore$ Taking $\lim_{\lambda\to 0+}$ on both sides,

$$\Rightarrow \lim_{\lambda\to 0+} \frac{f(x_1+\lambda(x_2-x_1))-f(x_1)}{\lambda} \;\leq\; \lim_{\lambda\to 0+} f(x_2)-f(x_1)$$

Here $lim_{\lambda\to 0+}\frac{f(x_1+\lambda(x_2-x_1))-f(x_1)}{\lambda}$ is essentially the directional derivative of f at x_1 in the direction (x_2-x_1). Thus, the above inequality can be rewritten as:

$$\lim_{\lambda\to 0+} \nabla f(x_1)^T(x_2-x_1) \;\leq\; f(x_2)-f(x_1) \qquad \forall x_1,x_2\in S$$

(*Reverse*) $\Leftarrow$ For a convex set $S\subseteq \mathbb{R}^n$ and a function $f:S\to\mathbb{R}$ and $f\in\mathbb{C}^1(S)$, let the following property hold $\forall x_1,x_2\in S$:

$$f(x_2)-f(x_1) \;\geq\; \nabla f(x_1)^T(x_2-x_1) \qquad ...\mathbf{(1)}$$

Now we have to show that f is a convex function.

Let there be a new point $x_3\in S$ such that, for some arbitrarily chosen $\lambda\in[0,1]$:

$$x_3 = \lambda x_1 + (1-\lambda)x_2$$

$\therefore$ By the definition of f,

$$f(x_1) \geq f(x_3) + \nabla f(x_3)^T(x_1-x_3) \qquad ...\mathbf{(2)}$$

$$f(x_2) \geq f(x_3) + \nabla f(x_3)^T(x_2-x_3) \qquad ...\mathbf{(3)}$$

Now, using eq. (2) and (3), we can write:

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

$$\geq \;\lambda[f(x_3)+\nabla f(x_3)^T(x_1-x_3)] \;+\; (1-\lambda)[f(x_3)+\nabla f(x_3)^T(x_2-x_3)]$$

$$= \; f(x_3) \;+\; \nabla f(x_3)^T[\lambda(x_1-x_3)+(1-\lambda)(x_2-x_3)]$$

$$= \; f(x_3) \;+\; \nabla f(x_3)^T[(\lambda x_1+(1-\lambda)x_2)-x_3]$$

$$= \; f(x_3) \;+\; \nabla f(x_3)^T[x_3-x_3]$$

[We had defined x_3 as $\lambda x_1 + (1-\lambda)x_2$]

$$= \; f(x_3)$$

$$\therefore \; \lambda f(x_1)+(1-\lambda)f(x_2) \;\geq\; f(x_3)$$

$$\Rightarrow \; \lambda f(x_1)+(1-\lambda)f(x_2) \;\geq\; f(\lambda x_1+(1-\lambda)x_2)$$

Thus for any $x_1, x_2 \in S$ and any arbitrarily chosen $\lambda \in [0, 1]$,

$$\lambda f(x_1) + (1-\lambda) f(x_2) \;\; \geq \;\; f(\lambda x_1 + (1-\lambda) x_2)$$

...which implies that f is a convex function.

$\because$ Both directions hold true, we can conclude that for a convex set $S \subseteq \mathbb{R}^n$ and a function $f : S \to \mathbb{R}$ with $f \in \mathbb{C}^1(S)$, the function f is a convex function if and only if the following condition is satisfied:

$$f(x_2) \geq f(x_1) + \nabla f(x_1)^T (x_2 - x_1)$$

$\square$

2.4 Gradient Inequality for Strict Convexity

In case of strict convexity of a function f, the theorem statement changes slightly:

Theorem: Let $S \subseteq \mathbb{R}^n$ be a convex set and $f : S \to \mathbb{R}$ and $f \in \mathbb{C}^1(S)$. Then f is a strictly convex function if and only if:

$$f(x_2) > f(x_1) + \nabla f(x_1)^T (x_2 - x_1) \qquad\qquad \forall x_1, x_2 \in S, \;\; x_1 \neq x_2$$

3 Sufficiency of Stationarity under Convexity

3.1 Stationarity and Stationary Points

A *stationary point* of a function is a point where its gradient vanishes. Such points may correspond to a local minimum, a local maximum, or a saddle point, depending on the function's behavior in their neighborhood.

Mathematically, for a function f, a point x^* is said to be *stationary* if:

$$\nabla f(x^*) = 0$$

3.2 Proposition Statement

Proposition: Let $S \subseteq \mathbb{R}^n$ be a convex set and $f : S \to \mathbb{R}$ be a convex function such that $f \in \mathbb{C}^1(S)$. Suppose $\nabla f(x^*) = \bar{0}$ for some $x^* \in S$. Then x^* is a global minima.

3.3 Proof

Proof: Let $S \subseteq \mathbb{R}^n$ be a convex set and $f : S \to \mathbb{R}$ be a convex function such that $f \in \mathbb{C}^1(S)$.

Suppose $\nabla f(x^*) = \bar{0}$ for some $x^* \in S$.

$\therefore$ Using the gradient inequality for the function f, we get:

$$f(x) > f(x^*) + \nabla f(x^*)^T(x - x^*) \qquad \forall x \in S$$

$\therefore$ Using $\nabla f(x^*) = \bar{0}$, we get:

$$f(x) > f(x^*) \qquad \forall x \in S$$

...which implies that x^* is a global minima. $\square$

4 Monotonicity of the Gradients of Convex Functions

4.1 Monotonically Changing Gradient

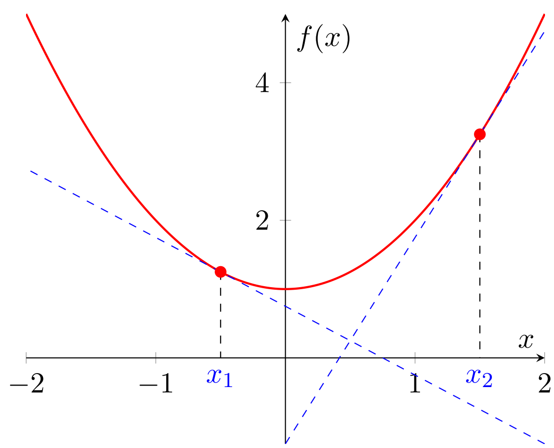

Figure 6: Monotonically increasing gradient $f'(x)$ with increasing x.

For a convex scalar function $f : \mathbb{R} \to \mathbb{R}$, it can be observed that the gradient of f, i.e. $f'(x)$ is monotonically non-decreasing with increasing x. Mathematically, for any 2 points $x_1, x_2 \in \mathbb{R}$, we can conclude that:

$$f'(x_2) - f'(x_1) \geq 0 \qquad \forall x_2 \geq x_1$$

and

$$f'(x_2) - f'(x_1) \leq 0 \qquad \forall x_2 \leq x_1$$

Thus, the terms $f'(x_2) - f'(x_1)$ and $x_2 - x_1$ are either both simultaneously non-negative (≥ 0) or non-positive (≤ 0). Together we can write them as:

$$(f'(x_2) - f'(x_1))(x_2 - x_1) \;\geq\; 0 \qquad \forall x_1, x_2 \in \mathbb{R}$$

This can be generalized to higher dimensions. For any convex set $S \subseteq \mathbb{R}^n$ set and a convex function $f : S \to \mathbb{R}$ and $f \in \mathbb{C}^1(S)$,

$$[\nabla f(x_2) - \nabla f(x_1)]^T(x_2 - x_1) \;\; \geq \;\; 0 \qquad\qquad \forall x_1, x_2 \in S$$

4.2 Theorem Statement

Theorem: Let $S \subseteq \mathbb{R}^n$ be a convex set and $f : S \to \mathbb{R}$ and $f \in \mathbb{C}^1(S)$. Then f is a convex function if and only if:

$$[\nabla f(x_2) - \nabla f(x_1)]^T(x_2 - x_1) \;\; \geq \;\; 0 \qquad x_1, x_2 \in S$$

4.3 Proof

Proof: (*Forward*) $\Rightarrow$ Let $S \subseteq \mathbb{R}^n$ be a convex set and $f : S \to \mathbb{R}$ be a convex function such that $f \in \mathbb{C}^1(S)$. Then, from gradient inequlaity for convex function, for $\forall x_1, x_2 \in S$ we have:

$$f(x_1) \geq f(x_2) + \nabla f(x_2)^T(x_1 - x_2) \qquad\qquad ...\mathbf{(1)}$$

$$f(x_2) \geq f(x_1) + \nabla f(x_1)^T(x_2 - x_1) \qquad\qquad ...\mathbf{(2)}$$

Adding (1) and (2),

$$f(x_1) + f(x_2) \;\; \geq \;\; f(x_2) + f(x_1) + \nabla f(x_2)^T(x_1 - x_2) + \nabla f(x_1)^T(x_2 - x_1)$$

Rearranging the terms, we get:

$$[\nabla f(x_2) - \nabla f(x_1)]^T(x_2 - x_1) \;\; \geq \;\; 0 \qquad\qquad \forall x_1, x_2 \in S$$

(*Reverse*) $\Leftarrow$ For a convex set $S \subseteq \mathbb{R}^n$ and a function $f : S \to \mathbb{R}$ and $f \in \mathbb{C}^1(S)$, let the following property hold $\forall x_1, x_2 \in S$:

$$[\nabla f(x_2) - \nabla f(x_1)]^T(x_2 - x_1) \;\; \geq \;\; 0 \qquad x_1, x_2 \in S$$

$\therefore$ Let us construct a function $g(t)$ as follows:

$$g(t) \;\; = \;\; f(x_1 + t(x_2 - x_1)) \qquad\qquad \forall t \in [0, 1]$$

Using the fundamental theorem of calculus, we have:

$$g(1) = g(0) + \int_0^1 g'(t)dt$$

$$\Rightarrow \;\; f(x_2) \;\; = \;\; f(x_1) + \int_0^1 \nabla f(x_1 + t(x_2 - x_1))^T(x_2 - x_1)dt \qquad\qquad ...\mathbf{(1)}$$

We see that:

$$\int_0^1 \nabla f(x_1)^T(x_2 - x_1)dt$$

$$= \nabla f(x_1)^T(x_2 - x_1)\int_0^1 dt$$

$$= \nabla f(x_1)^T(x_2 - x_1)$$

Substituting this result in (1), we get:

$$f(x_2) = f(x_1) + \nabla f(x_1))^T(x_2 - x_1) + \int_0^1 [\nabla f(x_1 + t(x_2 - x_1))^T(x_2 - x_1) - \nabla f(x_1)^T(x_2 - x_1)]dt \qquad ...\textbf{(2)}$$

Here we also see that:

$$\nabla f(x_1 + t(x_2 - x_1))^T(x_2 - x_1) - \nabla f(x_1)^T(x_2 - x_1)$$

$$= [\nabla f(x_1 + t(x_2 - x_1)) - \nabla f(x_1)]^T(x_2 - x_1)$$

$$= \frac{1}{t}[\nabla f(x_1 + t(x_2 - x_1)) - \nabla f(x_1)]^T(x_1 + t(x_2 - x_1) - x_1)$$

$$\geq 0$$

[Both $x_1 + t(x_2 - x_1)$ and x_1 follow the monotonicity property]

Substituting this result in (2), we get:

$$f(x_2) \geq f(x_1) + \nabla f(x_1))^T(x_2 - x_1) + \int_0^1 0.dt$$

$$\Rightarrow f(x_2) \geq f(x_1) + \nabla f(x_1))^T(x_2 - x_1)$$

...since this holds for any $x_1, x_2 \in S$, thus by the gradient inequality, f is a convex function.

$\because$ Both directions holds, and we can conclude that for a convex set $S \subseteq \mathbb{R}^n$ and a function $f : S \to \mathbb{R}$ with $f \in \mathbb{C}^1(S)$, then f is convex if and only if the following is satisfied:

$$[\nabla f(x_2) - \nabla f(x_1)]^T(x_2 - x_1) \geq 0 \qquad \forall x_1, x_2 \in S$$

$\square$

4.4 Monotonicity for Strict Convexity

In case of strict convexity of a function f, the theorem statement changes slightly.

Theorem: Let $S \subseteq \mathbb{R}^n$ be a convex set and $f : S \to \mathbb{R}$ and $f \in \mathbb{C}^1(S)$. Then f is a strictly convex function if and only if:

$$[\nabla f(x_2) - \nabla f(x_1)]^T(x_2 - x_1) > 0 \qquad \forall x_1, x_2 \in S,\ x_1 \neq x_2$$

In this case, the gradient is either strictly increasing or decreasing (instead of being non-decreasing and non-increasing).

References

1. Boyd, S., & Vandenberghe, L. (2004). *Convex Optimization*. Cambridge University Press.

2. Bertsekas, D. P. (2010). *Convex Optimization Theory*. Universities Press.

3. Beck, A. (2014). *Introduction to Nonlinear Optimization*. MOS-SIAM Series on Optimization.

4. Wikipedia contributors. (n.d.). *Sublevel set*. In *Wikipedia, The Free Encyclopedia*. `https://en.wikipedia.org/wiki/Level_set`.

5. Wikipedia contributors. (n.d.). *Quasiconvex function*. In *Wikipedia, The Free Encyclopedia*. `https://en.wikipedia.org/wiki/Quasiconvex_function`.