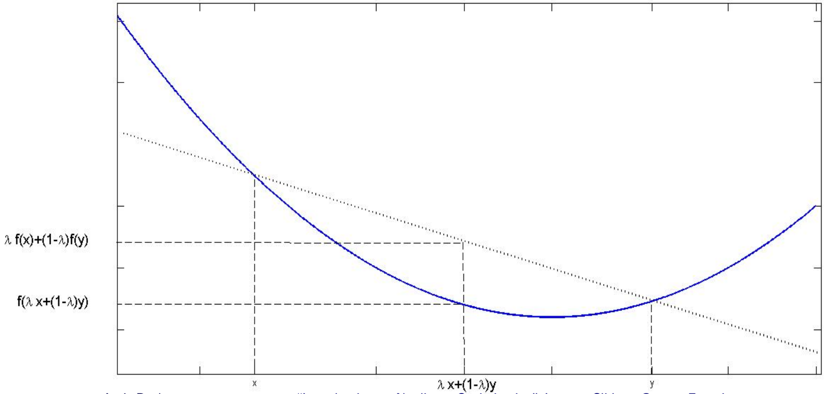
Lecture 7 - Convex Functions

Definition A function $f : C \to \mathbb{R}$ defined on a convex set $C \subseteq \mathbb{R}^n$ is called **convex** (or **convex over** C) if

$$f(\lambda \mathbf{x} + (1-\lambda)\mathbf{y}) \le \lambda f(\mathbf{x}) + (1-\lambda) f(\mathbf{y}) \text{ for any } \mathbf{x}, \mathbf{y} \in C, \lambda \in [0,1].$$



Convexity, Strict Convexity and Concavity

- In case where no domain is specified, we naturally assume that f is defined over the entire space $\mathbb{R}^n$.
- A function $f : C \to \mathbb{R}$ defined on a convex set $C \subseteq \mathbb{R}^n$ is called strictly convex if

$$f(\lambda \mathbf{x} + (1-\lambda)\mathbf{y}) < \lambda f(\mathbf{x}) + (1-\lambda) f(\mathbf{y}) \text{ for any } \mathbf{x} \neq \mathbf{y} \in C, \lambda \in (0,1).$$

- A function is called concave if $-f$ is convex. Similarly, f is called strictly concave if $-f$ is strictly convex.
- We can also define concavity directly: a function f is concave if and only if for any $\mathbf{x}, \mathbf{y} \in C$ and $\lambda \in [0,1]$,

$$f(\lambda \mathbf{x} + (1-\lambda)\mathbf{y}) \geq \lambda f(\mathbf{x}) + (1-\lambda) f(\mathbf{y}).$$

Examples of Convex Functions

- **Affine Functions.** $f(\mathbf{x}) = \mathbf{a}^T\mathbf{x} + b$, where $\mathbf{a} \in \mathbb{R}^n$ and $b \in \mathbb{R}$.
- **Norms.** $g(\mathbf{x}) = \|\mathbf{x}\|$.
- **Convexity of f:** Take $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ and $\lambda \in [0, 1]$. Then

$$\begin{aligned} f(\lambda\mathbf{x} + (1-\lambda)\mathbf{y}) &= \mathbf{a}^T(\lambda\mathbf{x} + (1-\lambda)\mathbf{y}) + b \\ &= \lambda(\mathbf{a}^T\mathbf{x}) + (1-\lambda)(\mathbf{a}^T\mathbf{y}) + \lambda b + (1-\lambda)b \\ &= \lambda(\mathbf{a}^T\mathbf{x} + b) + (1-\lambda)(\mathbf{a}^T\mathbf{y} + b) \\ &= \lambda f(\mathbf{x}) + (1-\lambda)f(\mathbf{y}), \end{aligned}$$

- **Convexity of g:** Take $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ and $\lambda \in [0, 1]$. Then

$$\begin{aligned} g(\lambda\mathbf{x} + (1-\lambda)\mathbf{y}) &= \|\lambda\mathbf{x} + (1-\lambda)\mathbf{y}\| \\ &\le \|\lambda\mathbf{x}\| + \|(1-\lambda)\mathbf{y}\| \\ &= \lambda\|\mathbf{x}\| + (1-\lambda)\|\mathbf{y}\| \\ &= \lambda g(\mathbf{x}) + (1-\lambda)g(\mathbf{y}), \end{aligned}$$

Jensen's Inequality

Theorem. Let $f : C \to \mathbb{R}$ be a convex function where $C \subseteq \mathbb{R}^n$ is a convex set. Then for any $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_k \in C$ and $\boldsymbol{\lambda} \in \Delta_k$, the following inequality holds:

$$f\left(\sum_{i=1}^{k} \lambda_i \mathbf{x}_i\right) \le \sum_{i=1}^{k} \lambda_i f(\mathbf{x}_i).$$

Proof very similar to the proof that any convex combination of pts. in a convex sets is in the set – see the proof of Theorem 7.5 on pages 118,119 of the book.

The Gradient Inequality

Theorem. Let $f : C \to \mathbb{R}$ be a continuously differentiable function defined on a convex set $C \subseteq \mathbb{R}^n$. Then f is convex over C if and only if

$$f(\mathbf{x}) + \nabla f(\mathbf{x})^T(\mathbf{y} - \mathbf{x}) \le f(\mathbf{y}) \text{ for any } \mathbf{x}, \mathbf{y} \in C. \tag{1}$$

Proof.

- Suppose first that f is convex. Let $\mathbf{x}, \mathbf{y} \in C$ and $\lambda \in (0, 1]$. If $\mathbf{x} = \mathbf{y}$, then (1) trivially holds. We will therefore assume that $\mathbf{x} \neq \mathbf{y}$.
- $\frac{f(\mathbf{x}+\lambda(\mathbf{y}-\mathbf{x}))-f(\mathbf{x})}{\lambda} \le f(\mathbf{y}) - f(\mathbf{x})$.
- Taking $\lambda \to 0^+$, we obtain

$$f'(\mathbf{x}; \mathbf{y} - \mathbf{x}) \le f(\mathbf{y}) - f(\mathbf{x}).$$

- Since f is continuously differentiable, $f'(\mathbf{x}; \mathbf{y} - \mathbf{x}) = \nabla f(\mathbf{x})^T(\mathbf{y} - \mathbf{x})$, and (1) follows.

Proof Contd.

- To prove the reverse direction, assume that that the gradient inequality holds.
- Let $\mathbf{z}, \mathbf{w} \in C$, and let $\lambda \in (0, 1)$. We will show that $f(\lambda \mathbf{z} + (1 - \lambda)\mathbf{w}) \leq \lambda f(\mathbf{z}) + (1 - \lambda) f(\mathbf{w})$.
- Let $\mathbf{u} = \lambda \mathbf{z} + (1 - \lambda)\mathbf{w} \in C$. Then

$$\mathbf{z} - \mathbf{u} = \frac{\mathbf{u} - (1 - \lambda)\mathbf{w}}{\lambda} - \mathbf{u} = -\frac{1 - \lambda}{\lambda}(\mathbf{w} - \mathbf{u}).$$

- We have

$$\begin{aligned} f(\mathbf{u}) + \nabla f(\mathbf{u})^T (\mathbf{z} - \mathbf{u}) & \leq f(\mathbf{z}), \\ f(\mathbf{u}) - \frac{\lambda}{1 - \lambda} \nabla f(\mathbf{u})^T (\mathbf{z} - \mathbf{u}) & \leq f(\mathbf{w}). \end{aligned}$$

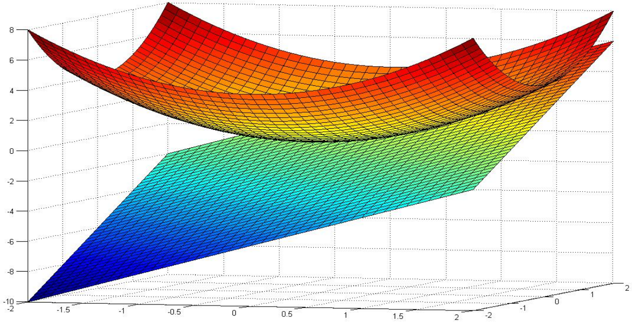
- Thus,

$$f(\mathbf{u}) \leq \lambda f(\mathbf{z}) + (1 - \lambda) f(\mathbf{w}).$$

The Gradient Inequality for Strictly Convex Functions

Proposition Let $f : C \to \mathbb{R}$ be a continuously differentiable function defined on a convex set $C \subseteq \mathbb{R}^n$. Then f is strictly convex over C if and only if

$$f(\mathbf{x}) + \nabla f(\mathbf{x})^T(\mathbf{y} - \mathbf{x}) < f(\mathbf{y}) \text{ for any } \mathbf{x}, \mathbf{y} \in C \text{ satisfying } \mathbf{x} \neq \mathbf{y}$$



Stationarity $\Rightarrow$ Global Optimality

A direct result of the gradient inequality is that the first order optimality condition $\nabla f(\mathbf{x}^*) = 0$ is sufficient for global optimality.

Proposition Let f be a continuously differentiable function which is convex over a convex set $C \subseteq \mathbb{R}^n$. Suppose that $\nabla f(\mathbf{x}^*) = \mathbf{0}$ for some $\mathbf{x}^* \in C$. Then $\mathbf{x}^*$ is the global minimizer of f over C.

Proof. In class

This is why convex optimization problems are relatively easy to solve on computers/ analyze mathematically.

Convexity of Quadratic Functions with Positive Semidefinite Matrices

Theorem. Let $f : \mathbb{R}^n \to \mathbb{R}$ be the quadratic function given by $f(\mathbf{x}) = \mathbf{x}^T\mathbf{A}\mathbf{x} + 2\mathbf{b}^T\mathbf{x} + c$ where $\mathbf{A} \in \mathbb{R}^{n\times n}$ is symmetric, $\mathbf{b} \in \mathbb{R}^n$ and $c \in \mathbb{R}$. Then f is (strictly) convex if and only if $\mathbf{A} \succeq \mathbf{0}$ ($\mathbf{A} \succ \mathbf{0}$).

Proof.

- The convexity of f is equivalent to

$$f(\mathbf{y}) \geq f(\mathbf{x}) + \nabla f(\mathbf{x})^T(\mathbf{y}-\mathbf{x}) \text{ for any } \mathbf{x}, \mathbf{y} \in \mathbb{R}^n$$

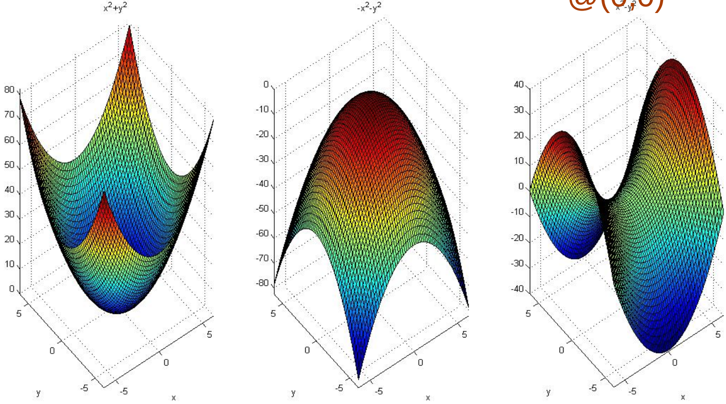
- Same as $\mathbf{y}^T\mathbf{A}\mathbf{y} + 2\mathbf{b}^T\mathbf{y} + c \geq \mathbf{x}^T\mathbf{A}\mathbf{x} + 2\mathbf{b}^T\mathbf{x} + c + 2(\mathbf{A}\mathbf{x}+\mathbf{b})^T(\mathbf{y}-\mathbf{x})$ for any $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$.
- $(\mathbf{y}-\mathbf{x})^T\mathbf{A}(\mathbf{y}-\mathbf{x}) \geq 0$ for any $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$.
- Equivalent to the inequality $\mathbf{d}^T\mathbf{A}\mathbf{d} \geq 0$ for any $\mathbf{d} \in \mathbb{R}^n$.
- Same as $\mathbf{A} \succeq \mathbf{0}$.
- Similar arguments show that strict convexity is equivalent to

$$\mathbf{d}^T\mathbf{A}\mathbf{d} > 0 \text{ for any } \mathbf{0} \neq \mathbf{d} \in \mathbb{R}^n,$$

namely to $\mathbf{A} \succ \mathbf{0}$.

Illustration

saddle point
@(0,0)



Much easier to picture these with the help of the spectral theorem (and rotational invariane of convexity)

Monotonicity of the Gradient

Theorem. Suppose that f is a continuously differentiable function over a convex set $C \subseteq \mathbb{R}^n$. Then f is convex over C if and only if

$$(\nabla f(\mathbf{x}) - \nabla f(\mathbf{y}))^T(\mathbf{x} - \mathbf{y}) \geq 0 \text{ for any } \mathbf{x}, \mathbf{y} \in C.$$

See the proof of Theorem 8.11 on pages 122,123 of the book.

Second-Order Characterization of Convexity

Theorem. Let f be a twice continuously differentiable function over an open convex set $C \subseteq \mathbb{R}^n$. Then f is convex over C if and only if $\nabla^2 f(\mathbf{x}) \succeq \mathbf{0}$ for any $\mathbf{x} \in C$.

Proof.

- Suppose that $\nabla^2 f(\mathbf{x}) \succeq \mathbf{0}\ \forall \mathbf{x} \in C$. Let $\mathbf{x}, \mathbf{y} \in C$, then $\exists \mathbf{z} \in [\mathbf{x}, \mathbf{y}] \in C$:

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

- $(\mathbf{y}-\mathbf{x})^T \nabla^2 f(\mathbf{z})(\mathbf{y}-\mathbf{x}) \geq 0 \Rightarrow f(\mathbf{y}) \geq f(\mathbf{x}) + \nabla f(\mathbf{x})^T(\mathbf{y}-\mathbf{x}) \Rightarrow f$ convex.
- Suppose that f is convex over C. Let $\mathbf{x} \in C$ and let $\mathbf{y} \in \mathbb{R}^n$.
- C is open $\Rightarrow \exists \varepsilon > 0$ such that $\mathbf{x} + \lambda \mathbf{y} \in C\ \forall \lambda \in (0, \varepsilon)$.

$$f(\mathbf{x} + \lambda \mathbf{y}) \geq f(\mathbf{x}) + \lambda \nabla f(\mathbf{x})^T \mathbf{y}.$$

- $f(\mathbf{x} + \lambda \mathbf{y}) = f(\mathbf{x}) + \lambda \nabla f(\mathbf{x})^T \mathbf{y} + \frac{\lambda^2}{2} \mathbf{y}^T \nabla^2 f(\mathbf{x}) \mathbf{y} + o(\lambda^2 \|\mathbf{y}\|^2)$.
- Thus, $\frac{\lambda^2}{2} \mathbf{y}^T \nabla^2 f(\mathbf{x}) \mathbf{y} + o(\lambda^2 \|\mathbf{y}\|^2) \geq 0$ for any $\lambda \in (0, \varepsilon)$.
- Dividing by λ^2, $\frac{1}{2} \mathbf{y}^T \nabla^2 f(\mathbf{x}) \mathbf{y} + \frac{o(\lambda^2 \|\mathbf{y}\|^2)}{\lambda^2} \geq 0$.
- Taking $\lambda \to 0^+$, we have $\mathbf{y}^T \nabla^2 f(\mathbf{x}) \mathbf{y} \geq 0 \forall \mathbf{y} \in \mathbb{R}^n$.
- Hence $\nabla^2 f(\mathbf{x}) \succeq \mathbf{0}$ for any $\mathbf{x} \in C$.

Convexity of the log-sum-exp function

- $f(\mathbf{x}) = \log\left(e^{x_1} + e^{x_2} + \ldots + e^{x_n}\right), \quad \mathbf{x} \in \mathbb{R}^n$
- $\frac{\partial f}{\partial x_i}(\mathbf{x}) = \frac{e^{x_i}}{\sum_{j=1}^n e^{x_j}}, \quad i = 1, 2, \ldots, n,$
- $\frac{\partial^2 f}{\partial x_i \partial x_j}(\mathbf{x}) = \begin{cases} -\frac{e^{x_i} e^{x_j}}{\left(\sum_{j=1}^n e^{x_j}\right)^2}, & i \neq j, \\ -\frac{e^{x_i} e^{x_j}}{\left(\sum_{j=1}^n e^{x_j}\right)^2} + \frac{e^{x_i}}{\sum_{j=1}^n e^{x_j}}, & i = j \end{cases}$
- We can thus write the Hessian matrix as

$$\nabla^2 f(\mathbf{x}) = \operatorname{diag}(\mathbf{w}) - \mathbf{w}\mathbf{w}^T, \quad \mathbf{w} = \left(\frac{e^{x_i}}{\sum_{j=1}^n e^{x_j}}\right)_{i=1}^n \in \Delta_n.$$

- For any $\mathbf{v} \in \mathbb{R}^n$: $\mathbf{v}^T \nabla^2 f(\mathbf{x}) \mathbf{v} = \sum_{i=1}^n w_i v_i^2 - (\mathbf{v}^T \mathbf{w})^2 \geq 0$ since defining $s_i = \sqrt{w_i} v_i, t_i = \sqrt{w_i}$, we have

$$(\mathbf{v}^T \mathbf{w})^2 = (\mathbf{s}^T \mathbf{t})^2 \leq \|\mathbf{s}\|^2 \|\mathbf{t}\|^2 = \left(\sum_{i=1}^n w_i v_i^2\right)\left(\sum_{i=1}^n w_i\right) = \sum_{i=1}^n w_i v_i^2.$$

- Thus, $\nabla^2 f(\mathbf{x}) \succeq \mathbf{0}$ and hence f is convex over $\mathbb{R}^n$.

Convexity of quad-over-lin

$$f(x_1, x_2) = \frac{x_1^2}{x_2}$$

defined over $\mathbb{R} \times \mathbb{R}_+ = \{(x_1, x_2) : x_2 > 0\}$.

In class

Operations Preserving Convexity

- Convexity is preserved under several operations such as summation, multiplication by positive scalars and affine change of variables.

Theorem.

- Let f be a convex function defined over a convex set $C \subseteq \mathbb{R}^n$ and let $\alpha \geq 0$. Then αf is a convex function over C.
- Let $f_1, f_2, \ldots, f_p$ be convex functions over a convex set $C \subseteq \mathbb{R}^n$. Then the sum function $f_1 + f_2 + \ldots + f_p$ is convex over C.
- Let f be a convex function defined on a convex set $C \subseteq \mathbb{R}^n$. Let $\mathbf{A} \in \mathbb{R}^{n \times m}$ and $\mathbf{b} \in \mathbb{R}^n$. Then the function g defined by

$$g(\mathbf{y}) = f(\mathbf{A}\mathbf{y} + \mathbf{b}).$$

is convex over the convex set $D = \{\mathbf{y} \in \mathbb{R}^m : \mathbf{A}\mathbf{y} + \mathbf{b} \in C\}$.

See the proofs of Theorems 7.16 and 7.17 of the book.

Example: Generalized quadratic-over-linear

The generalized quad-over-lin function

$$g(\mathbf{x}) = \frac{\|\mathbf{A}\mathbf{x}+\mathbf{b}\|^2}{\mathbf{c}^T\mathbf{x}+d} \quad (\mathbf{A} \in \mathbb{R}^{m\times n}, \mathbf{b} \in \mathbb{R}^m, \mathbf{c} \in \mathbb{R}^n, d \in \mathbb{R})$$

is convex over $D = \{\mathbf{x} \in \mathbb{R}^n : \mathbf{c}^T\mathbf{x} + d > 0\}$.
In class

Examples of Convex Functions

- $$f(x_1, x_2) = x_1^2 + 2x_1x_2 + 3x_2^2 + 2x_1 - 3x_2 + e^{x_1}.$$
- $$f(x_1, x_2, x_3) = e^{x_1 - x_2 + x_3} + e^{2x_2} + x_1$$
- $$f(x_1, x_2) = -\log(x_1x_2)$$

  over $\mathbb{R}^2_{++}$

In class

Preservation of Convexity under Composition

Theorem. Let $f : C \to \mathbb{R}$ be a convex function defined over the convex set $C \subseteq \mathbb{R}^n$. Let $g : I \to \mathbb{R}$ be a one-dimensional nondecreasing convex function over the interval $I \subseteq \mathbb{R}$. Assume that the image of C under f is contained in I: $f(C) \subseteq I$. Then the composition of g with f defined by

$$h(\mathbf{x}) \equiv g(f(\mathbf{x}))$$

is convex over C.

Proof Outline. Let $\mathbf{x}, \mathbf{y} \in C$ and let $\lambda \in [0,1]$. Then

$$\begin{aligned} h(\lambda\mathbf{x} + (1-\lambda)\mathbf{y}) &= g(f(\lambda\mathbf{x} + (1-\lambda)\mathbf{y})) \\ &\leq g(\lambda f(\mathbf{x}) + (1-\lambda)f(\mathbf{y})) \\ &\leq \lambda g(f(\mathbf{x})) + (1-\lambda)g(f(\mathbf{y})) \\ &= \lambda h(\mathbf{x}) + (1-\lambda)h(\mathbf{y}), \end{aligned}$$

thus establishing the convexity of h. $\square$

Examples

- $h(\mathbf{x}) = e^{\|\mathbf{x}\|^2}$
- $h(\mathbf{x}) = (\|\mathbf{x}\|^2 + 1)^2$

In class

Point-Wise Maximum of Convex Functions

Theorem. Let $f_1, f_2, \ldots, f_p : C \to \mathbb{R}$ be p convex functions over the convex set $C \subseteq \mathbb{R}^n$. Then the maximum function

$$f(\mathbf{x}) \equiv \max_{i=1,2,\ldots,p} \{f_i(\mathbf{x})\}$$

is convex over C.

Proof Outline Let $\mathbf{x}, \mathbf{y} \in C$ and $\lambda \in [0,1]$. Then

$$\begin{aligned} f(\lambda\mathbf{x} + (1-\lambda)\mathbf{y}) &= \max_{i=1,2,\ldots,p} f_i(\lambda\mathbf{x} + (1-\lambda)\mathbf{y}) \\ &\le \max_{i=1,2,\ldots,p} \{\lambda f_i(\mathbf{x}) + (1-\lambda) f_i(\mathbf{y})\} \\ &\le \lambda \max_{i=1,2,\ldots,p} f_i(\mathbf{x}) + (1-\lambda) \max_{i=1,2,\ldots,p} f_i(\mathbf{y}) \\ &= \lambda f(\mathbf{x}) + (1-\lambda) f(\mathbf{y}). \end{aligned}$$

Examples.

- $f(\mathbf{x}) = \max\{x_1, x_2, \ldots, x_n\}$ is convex.
- For a given vector $\mathbf{x} = (x_1, x_2, \ldots, x_n)^T \in \mathbb{R}^n$, let $x_{[i]}$ denote the i-th largest value in $\mathbf{x}$. For any $k \in \{1, 2, \ldots, n\}$ the function

$$h_k(\mathbf{x}) = x_{[1]} + x_{[2]} + \ldots + x_{[k]},$$

is convex. why?

Preservation of Convexity Under Partial Minimization

Theorem. Let $f : C \times D \to \mathbb{R}$ be a convex function defined over the set $C \times D$ where $C \subseteq \mathbb{R}^m$ and $D \subseteq \mathbb{R}^n$ are convex sets. Let

$$g(\mathbf{x}) = \min_{\mathbf{y} \in D} f(\mathbf{x}, \mathbf{y}), \quad \mathbf{x} \in C,$$

where we assume that the minimum is finite. Then g is convex over C.

Proof. Let $\mathbf{x}_1, \mathbf{x}_2 \in C$ and $\lambda \in [0,1]$. Take $\varepsilon > 0$. Then $\exists \mathbf{y}_1, \mathbf{y}_2 \in D$:

$$f(\mathbf{x}_1, \mathbf{y}_1) \leq g(\mathbf{x}_1) + \varepsilon, f(\mathbf{x}_2, \mathbf{y}_2) \leq g(\mathbf{x}_2) + \varepsilon.$$

By the convexity of f we have

$$\begin{aligned} f(\lambda \mathbf{x}_1 + (1-\lambda)\mathbf{x}_2, \lambda \mathbf{y}_1 + (1-\lambda)\mathbf{y}_2) &\leq \lambda f(\mathbf{x}_1, \mathbf{y}_1) + (1-\lambda) f(\mathbf{x}_2, \mathbf{y}_2) \\ &\leq \lambda(g(\mathbf{x}_1) + \varepsilon) + (1-\lambda)(g(\mathbf{x}_2) + \varepsilon) \\ &= \lambda g(\mathbf{x}_1) + (1-\lambda) g(\mathbf{x}_2) + \varepsilon. \end{aligned}$$

Since the above inequality holds for any $\varepsilon > 0$, it follows that $g(\lambda \mathbf{x}_1 + (1-\lambda)\mathbf{x}_2) \leq \lambda g(\mathbf{x}_1) + (1-\lambda) g(\mathbf{x}_2)$.

Example: The distance function from a convex set $d_C(\mathbf{x}) \equiv \inf_{\mathbf{y} \in C} \|\mathbf{x} - \mathbf{y}\|$ is convex.

Level Sets

Definition. Let $f : S \to \mathbb{R}$ be a function defined over a set $S \subseteq \mathbb{R}^n$. Then the level set of f with level α is given by

$$\mathrm{Lev}(f, \alpha) = \{\mathbf{x} \in S : f(\mathbf{x}) \leq \alpha\}.$$

Theorem. Let $f : C \to \mathbb{R}$ be a convex function over the convex set $C \subseteq \mathbb{R}^n$. Then for any $\alpha \in \mathbb{R}$ the level set $\mathrm{Lev}(f, \alpha)$ is convex.

Proof.

But the converse is not true. And we have something called 'quasi-convex functions' because of it (next slide.)

- Let $\mathbf{x}, \mathbf{y} \in \mathrm{Lev}(f, \alpha)$ and $\lambda \in [0, 1]$.
- Then $f(\mathbf{x}), f(\mathbf{y}) \leq \alpha$. Hence,

$$f(\lambda\mathbf{x} + (1-\lambda)\mathbf{y}) \leq \lambda f(\mathbf{x}) + (1-\lambda)f(\mathbf{y}) \leq \lambda\alpha + (1-\lambda)\alpha = \alpha,$$

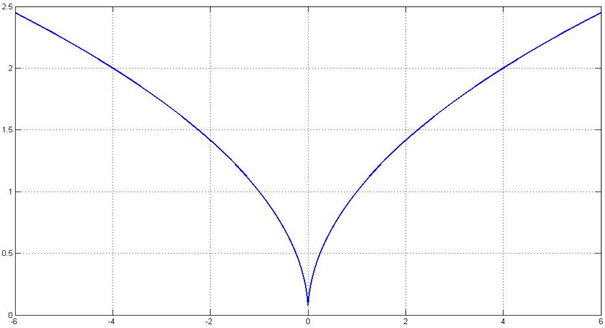
- $\lambda\mathbf{x} + (1-\lambda)\mathbf{y} \in \mathrm{Lev}(f, \alpha)$, and we have established the convexity of $\mathrm{Lev}(f, \alpha)$.

Quasi-Convex Functions

- Definition. A function $f : C \to \mathbb{R}$ defined over the convex set $C \subseteq \mathbb{R}^n$ is called quasi-convex if for any $\alpha \in \mathbb{R}$ the set $\mathrm{Lev}(f, \alpha)$ is convex.

Examples:

- $f(x) = \sqrt{|x|}$.
- $f(\mathbf{x}) = \frac{\mathbf{a}^T\mathbf{x}+b}{\mathbf{c}^T\mathbf{x}+d}$, over $C = \{\mathbf{x} \in \mathbb{R}^n : \mathbf{c}^T\mathbf{x} + d > 0\}$. where $\mathbf{a}, \mathbf{c} \in \mathbb{R}^n$ and $b, d \in \mathbb{R}$.



Continuity of Convex Functions

Theorem. Let $f : C \to \mathbb{R}$ be a convex function defined over a convex set $C \subseteq \mathbb{R}^n$. Let $\mathbf{x}_0 \in \text{int}(C)$. Then there exist $\varepsilon > 0$ and $L > 0$ such that $B[\mathbf{x}_0, \varepsilon] \subseteq C$ and

$$|f(\mathbf{x}) - f(\mathbf{x}_0)| \leq L\|\mathbf{x} - \mathbf{x}_0\| \text{ for any } \mathbf{x} \in B[\mathbf{x}_0, \varepsilon]$$

Proof.

- Take $\varepsilon > 0$ such that $B_\infty[\mathbf{x}_0, \varepsilon] \equiv \{\mathbf{x} \in \mathbb{R}^n : \|\mathbf{x} - \mathbf{x}_0\|_\infty \leq \varepsilon\} \subseteq C$.
- Let $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_{2^n}$ be the 2^n extreme points of $B_\infty[\mathbf{x}_0, \varepsilon]$.
- For any $\mathbf{x} \in B_\infty[\mathbf{x}_0, \varepsilon]$ there exists $\boldsymbol{\lambda} \in \Delta_{2^n}$ such that $\mathbf{x} = \sum_{i=1}^{2^n} \lambda_i \mathbf{v}_i$.By Jensen's inequality,

$$f(\mathbf{x}) = f\left(\sum_{i=1}^{2^n} \lambda_i \mathbf{v}_i\right) \leq \sum_{i=1}^{2^n} \lambda_i f(\mathbf{v}_i) \leq M,$$

  where $M = \max_{i=1,2,\ldots,2^n} f(\mathbf{v}_i)$.
- $B_2[\mathbf{x}_0, \varepsilon] = B[\mathbf{x}_0, \varepsilon] = \{\mathbf{x} \in \mathbb{R}^n : \|\mathbf{x} - \mathbf{x}_0\|_2 \leq \varepsilon\} \subseteq B_\infty[\mathbf{x}_0, \varepsilon]$.
- We therefore conclude that $f(\mathbf{x}) \leq M$ for any $\mathbf{x} \in B[\mathbf{x}_0, \varepsilon]$.

Continuity of Convex Functions Contd.

- Let $\mathbf{x} \in B[\mathbf{x}_0, \varepsilon]$ be such that $\mathbf{x} \neq \mathbf{x}_0$. Define

$$\mathbf{z} = \mathbf{x}_0 + \frac{1}{\alpha}(\mathbf{x} - \mathbf{x}_0), \quad \alpha = \frac{1}{\varepsilon}\|\mathbf{x} - \mathbf{x}_0\|$$

.

- Then obviously $\alpha \leq 1$ and $\mathbf{z} \in B[\mathbf{x}_0, \varepsilon]$, and in particular $f(\mathbf{z}) \leq M$.
- $\mathbf{x} = \alpha\mathbf{z} + (1 - \alpha)\mathbf{x}_0$.
- Consequently,

$$f(\mathbf{x}) \leq \alpha f(\mathbf{z}) + (1-\alpha)f(\mathbf{x}_0) \leq f(\mathbf{x}_0) + \alpha(M - f(\mathbf{x}_0)) = f(\mathbf{x}_0) + \frac{M - f(\mathbf{x}_0)}{\varepsilon}\|\mathbf{x} - \mathbf{x}_0\|.$$

- Thus, $f(\mathbf{x}) - f(\mathbf{x}_0) \leq L\|\mathbf{x} - \mathbf{x}_0\|$, where $L = \frac{M - f(\mathbf{x}_0)}{\varepsilon}$.
- We need to show that $f(\mathbf{x}) - f(\mathbf{x}_0) \geq -L\|\mathbf{x} - \mathbf{x}_0\|$.
- Define $\mathbf{u} = \mathbf{x}_0 + \frac{1}{\alpha}(\mathbf{x}_0 - \mathbf{x})$. Since $\mathbf{u} \in B[\mathbf{x}_0, \varepsilon]$, then $f(\mathbf{u}) \leq M$.
- $\mathbf{x} = \mathbf{x}_0 + \alpha(\mathbf{x}_0 - \mathbf{u})$. Therefore,

$$\begin{aligned} f(\mathbf{x}) &= f(\mathbf{x}_0 + \alpha(\mathbf{x}_0 - \mathbf{u})) \geq f(\mathbf{x}_0) + \alpha(f(\mathbf{x}_0) - f(\mathbf{u})) \\ &= f(\mathbf{x}_0) - \frac{M - f(\mathbf{x}_0)}{\varepsilon}\|\mathbf{x} - \mathbf{x}_0\| \\ &= f(\mathbf{x}_0) - L\|\mathbf{x} - \mathbf{x}_0\| \end{aligned}$$

Existence of Directional Derivatives of Convex Functions

Theorem. Let $f : C \to \mathbb{R}$ be a convex function over the convex set $C \subseteq \mathbb{R}^n$. Let $\mathbf{x} \in \operatorname{int}(C)$. Then for any $\mathbf{d} \neq \mathbf{0}$, the directional derivative $f'(\mathbf{x};\mathbf{d})$ exists.

Proof.

- Let $\mathbf{x} \in \operatorname{int}(C)$ and let $\mathbf{d} \neq \mathbf{0}$. Then the directional derivative (if exists) is the limit

$$\lim_{t \to 0^+} \frac{g(t) - g(0)}{t} \quad (g(t) = f(\mathbf{x} + t\mathbf{d})) \tag{2}$$

- Defining $h(t) = \frac{g(t)-g(0)}{t}$, (2) is the same as $\lim_{t \to 0^+} h(t)$.
- We will take an $\varepsilon > 0$ for which $\mathbf{x} + t\mathbf{d}, \mathbf{x} - t\mathbf{d} \in C$ for all $t \in [0, \varepsilon]$.
- Let $0 < t_1 < t_2 \leq \varepsilon$. Then $f(\mathbf{x} + t_1\mathbf{d}) \leq \left(1 - \frac{t_1}{t_2}\right) f(\mathbf{x}) + \frac{t_1}{t_2} f(\mathbf{x} + t_2\mathbf{d})$.
- Consequently, $\frac{f(\mathbf{x}+t_1\mathbf{d})-f(\mathbf{x})}{t_1} \leq \frac{f(\mathbf{x}+t_2\mathbf{d})-f(\mathbf{x})}{t_2}$.
- Thus, $h(t_1) \leq h(t_2) \Rightarrow h$ is monotone nondecreasing over $\mathbb{R}_{++}$. All that is left is to show that it is bounded below over $(0, \varepsilon]$.

Proof Contd.

- Take $0 < t \leq \varepsilon$. Note that

$$\mathbf{x} = \frac{\varepsilon}{\varepsilon + t}(\mathbf{x} + t\mathbf{d}) + \frac{t}{\varepsilon + t}(\mathbf{x} - \varepsilon\mathbf{d}).$$

- Hence,

$$f(\mathbf{x}) \leq \frac{\varepsilon}{\varepsilon + t} f(\mathbf{x} + t\mathbf{d}) + \frac{t}{\varepsilon + t} f(\mathbf{x} - \varepsilon\mathbf{d}).$$

- After some rearrangement of terms,

$$h(t) = \frac{f(\mathbf{x} + t\mathbf{d}) - f(\mathbf{x})}{t} \geq \frac{f(\mathbf{x}) - f(\mathbf{x} - \varepsilon\mathbf{d})}{\varepsilon}.$$

- h is bounded below over $(0, \varepsilon]$.
- Since h is nondecreasing and bounded below over $(0, \varepsilon]$, the limit $\lim_{t \to 0^+} h(t)$ exists $\Rightarrow$ the directional derivative $f'(\mathbf{x}; \mathbf{d})$ exists.

Extended Real-Valued Functions

- Until now we have discussed functions that are real-valued, meaning that they take their values in $\mathbb{R} = (-\infty, \infty)$.
- We will now consider functions that take their values in $\mathbb{R} \cup \{\infty\} = (-\infty, \infty]$. Such functions are called extended real-valued functions.
- **Example:** the indicator function: given a set $S \subseteq \mathbb{R}^n$, the indicator function $\delta_S : \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$ is given by

$$\delta_S(\mathbf{x}) = \begin{cases} 0 & \text{if } \mathbf{x} \in S, \\ \infty & \text{if } \mathbf{x} \notin S. \end{cases}$$

- The effective domain of an extended real-valued function is the set of vectors for which the function takes a real value:

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

- An extended real-valued function $f : \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$ is called proper if is not always equal to infinity, meaning that there exists $\mathbf{x}_0 \in \mathbb{R}^n$ such that $f(\mathbf{x}_0) < \infty$.

Extended Real-Valued Functions Contd.

- An extended real-valued function is convex if for any $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ and $\lambda \in [0, 1]$ the following inequality holds:

$$f(\lambda \mathbf{x} + (1-\lambda)\mathbf{y}) \leq \lambda f(\mathbf{x}) + (1-\lambda) f(\mathbf{y}),$$

  where we use the usual arithmetic rules with ∞ such as

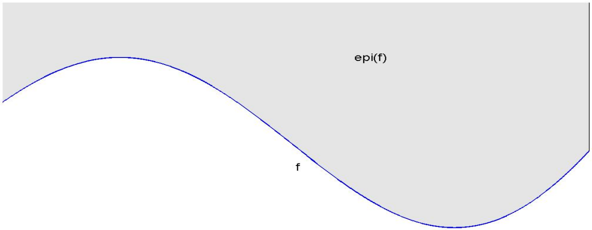
$$\begin{aligned} a + \infty &= \infty \text{ for any } a \in \mathbb{R}, \\ a \cdot \infty &= \infty \text{ for any } a \in \mathbb{R}_{++}. \end{aligned}$$

  In addition, we have the much less obvious rule that $0 \cdot \infty = 0$.
- It is easy to show that an extended real-valued function is convex iff $\text{dom}(f)$ is a convex set and the restriction of f to its effective domain is a convex real-valued function over $\text{dom}(f)$.
- As an example, the indicator function $\delta_C(\cdot)$ of a set $C \subseteq \mathbb{R}^n$ is convex if and only if C is a convex set.

The Epigraph

- Definition. Let $f : \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$. Then its epigraph $\text{epi}(f) \in \mathbb{R}^{n+1}$ is defined to be the set

$$\text{epi}(f) = \{(\mathbf{x}; t) : f(\mathbf{x}) \leq t\}.$$



It is not difficult to show that an extended real-valued function f is convex if and only if its epigraph set $\text{epi}(f)$ is convex.

Preservation of Convexity Under Supremum

Theorem. Let $f_i : \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$ be an extended real-valued convex functions for any $i \in I$ (I being an arbitrary index set). Then the function $f(\mathbf{x}) = \sup_{i \in I} f_i(\mathbf{x})$ is an extended real-valued convex function.

Proof. f_i convex for all $i \Rightarrow \mathrm{epi}(f_i)$ convex $\Rightarrow \mathrm{epi}(f) = \cap_{i \in I} \mathrm{epi}(f_i)$ convex $\Rightarrow$ $f(\mathbf{x}) = \sup_{i \in I} f_i(\mathbf{x})$ is convex.

- **Support Functions**. Let $S \subseteq \mathbb{R}^n$. The support function of S is the function

$$\sigma_S(\mathbf{x}) = \sup_{\mathbf{y} \in S} \mathbf{x}^T \mathbf{y}.$$

The support function is a convex function (regardless of whether S is convex or not).

Maximum of a Convex Fun. over a Compact Convex Set

Theorem. Let $f : C \to \mathbb{R}$ be convex over the nonempty convex and compact set $C \subseteq \mathbb{R}^n$. Then there exists at least one maximizer of f over C that is an extreme point of C.

Proof.

- Let $\mathbf{x}^*$ be a maximizer of f over C. If $\mathbf{x}^*$ is an extreme point of C, then the result is established. Otherwise,
- By Krein-Milman, $C = \text{conv}(\mathrm{ext}(C)) \Rightarrow \exists \mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_k \in \mathrm{ext}(C)$ and $\boldsymbol{\lambda} \in \Delta_k$ s.t.

$$\mathbf{x}^* = \sum_{i=1}^{k} \lambda_i \mathbf{x}_i.$$

relevent to linear programming (see Chapter 8)

- By convexity of f,

$$f(\mathbf{x}^*) \leq \sum_{i=1}^{k} \lambda_i f(\mathbf{x}_i),$$

- $\sum_{i=1}^{k} \lambda_i (f(\mathbf{x}_i) - f(\mathbf{x}^*)) \geq 0 \Rightarrow f(\mathbf{x}_i) = f(\mathbf{x}^*)$ (why?)