Topics in Applied Optimization

Optimization Algorithms for ML and Data Sciences

Pawan Kumar

CSTAR, IIIT-H

1st order or 2nd order cvxty can't be used because "max" is not differentiable

Max Functions. $f(x) = \max\{x_1, x_2, \ldots, x_n\}$ is convex on $\mathbb{R}^n$

$$f(\theta x + (1-\theta)y) = \max_i \left(\theta x_i + (1-\theta) y_i \right)$$

$$\leq \max_i \theta x_i + \max_i (1-\theta) y_i$$

$$= \theta \max_i x_i + (1-\theta) \max_i y_i$$

$$= \theta f(x) + (1-\theta) f(y)$$

$\Rightarrow$ f is convex

2 variables $\quad f: \mathbb{R}^2 \to \mathbb{R} \quad \boxed{\nabla^2 f(x) \succeq 0}$

Quadratic over linear: $f(x, y) = x^2/y$, with

$$\textbf{dom } f = \mathbb{R} \times \mathbb{R}_{++} = \{(x, y) \in \mathbb{R}^2 \mid y > 0\}$$

is convex function.

$\begin{bmatrix} y \\ -x \end{bmatrix} \begin{bmatrix} y & -x \end{bmatrix}$ = outer product

$\begin{bmatrix} - \end{bmatrix} \cdot \begin{bmatrix} | \end{bmatrix}$ = inner prod

$\begin{bmatrix} | \end{bmatrix} \begin{bmatrix} - \end{bmatrix}$ = outer prod.

2nd order convexity. That is, check $\nabla^2 f(x)$

$H = \nabla^2 f(x) = \frac{2}{y^3} \begin{bmatrix} y \\ -x \end{bmatrix} \begin{bmatrix} y \\ -x \end{bmatrix}^T$

of the form $c\, u u^T$

claim $uu^T \succeq 0$

$x^T (u u^T) x = (u^T x)^T (u^T x) = \|u^T x\|^2 \geq 0$

$$= \begin{bmatrix} \frac{\partial^2 f}{\partial x^2} & \frac{\partial^2 f}{\partial x \partial y} \\ \frac{\partial^2 f}{\partial y \partial x} & \frac{\partial^2 f}{\partial y^2} \end{bmatrix} = \begin{bmatrix} 2x/y \\ -\frac{x^2}{y^2} \end{bmatrix}$$

$$\frac{\partial}{\partial y}\left(x^2 y^{-2}\right) = -x^2(-2) y^{-3} = \frac{2x^2}{y^3}$$

$$= \begin{bmatrix} 2/y & -\frac{2x}{y^2} \\ -\frac{2x}{y^2} & \frac{2x^2}{y^3} \end{bmatrix} = \frac{2}{y^3} \begin{bmatrix} y^2 & -xy \\ -xy & x^2 \end{bmatrix} = \frac{2}{y^3} \begin{bmatrix} y \\ -x \end{bmatrix} \begin{bmatrix} y \\ -x \end{bmatrix}^T$$

Quadratic over linear: $f(x, y) = x^2/y$, with

$$\mathbf{dom}\ f = \mathbb{R} \times \mathbb{R}_{++} = \{(x, y) \in \mathbb{R}^2 \mid y > 0\}$$

is convex function.

We have

$$\nabla^2 f(x, y) = 2/y^2 \begin{bmatrix} y^2 & -xy \\ -xy & x^2 \end{bmatrix} = 2/y^3 \begin{bmatrix} y \\ -x \end{bmatrix} \begin{bmatrix} y \\ -x \end{bmatrix}^T \geq 0$$

Log-sum-exp: $f(x) = \log(e^{x_1} + \cdots + e^{x_n})$ is convex on $\mathbb{R}^n$

$v^T (\mathrm{diag}\, z) v$

$n=2$ $\quad \nabla^2 f(x) = \frac{1}{(1^T z)^2}\left[(1^T z)\,\mathrm{diag}(z) - z z^T \right]$

$(b^T b) \cdot (a^T a)$ $\quad \Rightarrow (a^T b)^2$

$v^T z z^T v$ $\quad (z^T v)^T (z^T v) = \|z^T v\|^2$

$$v^T \nabla^2 f(x) v = \frac{1}{(1^T z)^2}\left(\sum_{i=1}^n z_i \right)\left(\sum_{i=1}^n v_i^2 z_i \right) - \left(\sum v_i z_i \right)^2 \geq 0$$

$z^T v = (z_1 v_1 + z_2 v_2 + \cdots + z_n v_n)$

Recall: C-S ineq. $\quad (a^T a)(b^T b) \geq (a^T b)^2 \quad \forall a, b$

$\nabla^2 f(x) \succeq 0$

$a_i = v_i \sqrt{z_i}$, $\quad b_i = \sqrt{z_i}$

$a^T a = \sum_{i=1}^n v_i^2 z_i$, $\quad b^T b = \sum_{i=1}^n z_i$

$a^T b = \sum_{i=1}^n v_i z_i$

$u u^T = \|u\|^2 \geq 0$

$$z z^T = \begin{bmatrix} e^{x_1} \\ e^{x_2} \end{bmatrix} \begin{bmatrix} e^{x_1} & e^{x_2} \end{bmatrix} = \begin{bmatrix} e^{2x_1} & e^{x_1+x_2} \\ e^{x_1+x_2} & e^{2x_2} \end{bmatrix}$$

Log-sum-exp: $f(x) = \log(e^{x_1} + \cdots + e^{x_n})$ is convex on $\mathbb{R}^n$

- The graph is

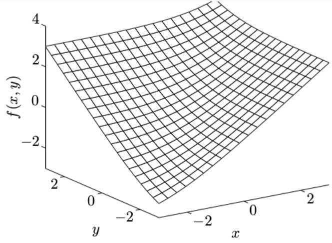

Figure 3.4 Graph of $f(x, y) = \log(e^x + e^y)$.

Log-sum-exp: $f(x) = \log(e^{x_1} + \cdots + e^{x_n})$ is convex on $\mathbb{R}^n$

- The graph is

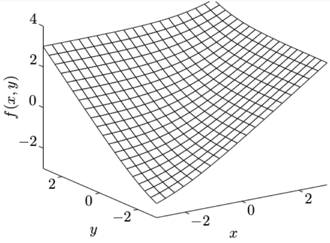

Figure 3.4 Graph of $f(x, y) = \log(e^x + e^y)$.

- Hessian of $f(x)$ is

Log-sum-exp: $f(x) = \log(e^{x_1} + \cdots + e^{x_n})$ is convex on $\mathbb{R}^n$

- The graph is

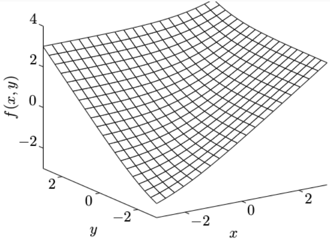

Figure 3.4 Graph of $f(x, y) = \log(e^x + e^y)$.

- Hessian of $f(x)$ is

$$\nabla^2 f(x) = \frac{1}{(1^T z)^2} \left((1^T z)\, \mathbf{diag}(z) - z z^T \right),$$

Log-sum-exp: $f(x) = \log(e^{x_1} + \cdots + e^{x_n})$ is convex on $\mathbb{R}^n$

- The graph is

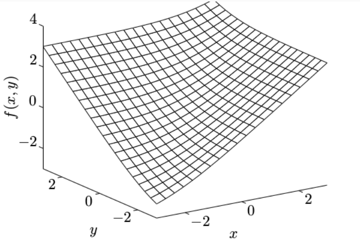

Figure 3.4 Graph of $f(x,y) = \log(e^x + e^y)$.

- Hessian of $f(x)$ is

$$\nabla^2 f(x) = \frac{1}{(\mathbf{1}^T z)^2}\left((\mathbf{1}^T z)\,\mathbf{diag}(z) - zz^T\right),$$

where $z = (e^{x_1}, \ldots, e^{x_n})$.

Log-sum-exp: $f(x) = \log(e^{x_1} + \cdots + e^{x_n})$ is convex on $\mathbb{R}^n$

- The graph is

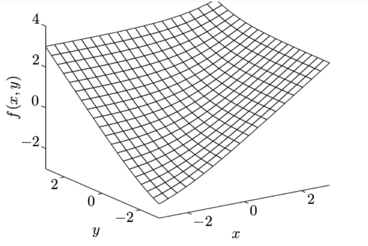

Figure 3.4 Graph of $f(x, y) = \log(e^x + e^y)$.

- Hessian of $f(x)$ is

$$\nabla^2 f(x) = \frac{1}{(1^T z)^2}\left((1^T z)\,\mathbf{diag}(z) - zz^T\right),$$

where $z = (e^{x_1}, \ldots, e^{x_n})$. Show that

$$v^T \nabla^2 f(x) v \geq 0$$

Proof that Hessian of log-sum-exp is Convex...

Scratch Space

Scratch Space

Example: Geometric mean is a concave function

Geometric mean: $f(x) = (\prod_{i=1}^{n} x_i)^{1/n}$ is concave on **dom** $f = \mathbb{R}^n_{++}$

$$\frac{\partial f}{\partial x_k^2} = -(n-1)\frac{\left(\prod_{i=1}^{n} x_i\right)^{1/n}}{n^2 x_k^2}$$

$$\frac{\partial^2 f}{\partial x_k \partial x_l} = \frac{\left(\prod_{i=1}^{n} x_i\right)^{1/n}}{n^2 x_k x_l}$$

for $k \neq l$

$\nabla^2 f(x)$

Example: Geometric mean is a concave function

Geometric mean: $f(x) = (\prod_{i=1}^n x_i)^{1/n}$ is concave on **dom** $f = \mathbb{R}^n_{++}$

Proof: The Hessian H is given by (Why?)

$$H(k,k) = \frac{\partial^2 f(x)}{\partial x_k^2} = -(n-1)\frac{(\prod_{i=1}^n x_i)^{1/n}}{n^2 x_k^2}$$

$$H(k,l) = \frac{\partial^2 f}{\partial x_k \partial x_l} = \frac{(\prod_{i=1}^n x_i)^{1/n}}{n^2 x_k x_l} \quad \text{for } k \neq l$$

Example: Geometric mean is a concave function

Geometric mean: $f(x) = (\prod_{i=1}^n x_i)^{1/n}$ is concave on **dom** $f = \mathbb{R}^n_{++}$

Proof: The Hessian H is given by (Why?)

$$H(k,k) = \frac{\partial^2 f(x)}{\partial x_k^2} = -(n-1)\frac{(\prod_{i=1}^n x_i)^{1/n}}{n^2 x_k^2}$$

$$H(k,l) = \frac{\partial^2 f}{\partial x_k \partial x_l} = \frac{(\prod_{i=1}^n x_i)^{1/n}}{n^2 x_k x_l} \quad \text{for } k \neq l$$

$$n\,1/x_i^2 - \frac{1}{x_i^2} = (n-1)/x_i^2$$

We have

$$\nabla^2 f(x) = -\frac{\prod_{i=1}^n x_i^{1/n}}{n^2}\left(n\,\mathbf{diag}(1/x_1^2, 1/x_2{}^2, \ldots, 1/x_n{}^2) - qq^T\right), \quad q_i = 1/x_i$$

$q_i = 1/x_i$

$$qq^T = \begin{bmatrix} 1/x_1 \\ 1/x_2 \\ \vdots \\ 1/x_n \end{bmatrix} \begin{bmatrix} 1/x_1 & 1/x_2 & \cdots & 1/x_n \end{bmatrix} = \begin{bmatrix} 1/x_1^2 & 1/x_1x_1 & \cdots & 1/x_1x_n \\ & 1/x_2^2 & & \\ & & \ddots & \\ \frac{1}{x_nx_1} & \cdots & & 1/x_n^2 \end{bmatrix}$$

Example: Geometric mean is a concave function

Geometric mean: $f(x) = (\prod_{i=1}^n x_i)^{1/n}$ is concave on **dom** $f = \mathbb{R}^n_{++}$

Proof: The Hessian H is given by (Why?)

$a^Ta = n$

$b^Tb = \sum v_i^2/x_i^2$

$a^Tb = \left(\sum v_i/x_i\right)^2$

$$H(k,k) = \frac{\partial^2 f(x)}{\partial x_k^2} = -(n-1)\frac{(\prod_{i=1}^n x_i)^{1/n}}{n^2 x_k^2}$$

$$H(k,l) = \frac{\partial^2 f}{\partial x_k \partial x_l} = \frac{(\prod_{i=1}^n x_i)^{1/n}}{n^2 x_k x_l} \quad \text{for } k \neq l$$

$$\begin{bmatrix} (n-1)/x_1^2 & & 1/x_k x_l \\ & \frac{(n-1)}{x_2^2} & \\ & & \ddots \end{bmatrix}$$

$$\begin{bmatrix} 1/x_1^2 & -1/x_1x_2 & \cdots \\ & 1/x_2^2 & \\ & & 1/x_n^2 \end{bmatrix}$$

We have

$$\nabla^2 f(x) = -\frac{\prod_{i=1}^n x_i^{1/n}}{n^2}\left(n\,\mathbf{diag}(1/x_1^2, 1/x_2{}^2, \ldots, 1/x_n{}^2) - qq^T\right), \quad q_i = 1/x_i$$

We have after using **Cauchy-Schwarz inequality** (How?)

v^Tqq^Tv

$$v^T\nabla^2 f(x) v = -\frac{\prod_{i=1}^n x_i^{1/n}}{n^2}\left(n\sum_{i=1}^n v_i^2/x_i^2 - \left(\sum_{i=1}^n v_i/x_i\right)^2\right) \leq 0$$

$(a^Ta)\quad(b^Tb)$

$(a^Ta)(b^Tb) \geq (a^Tb)^2$

≥ 0

choose $a = 1$

$b_i = v_i/x_i$

Scratch Space

Scratch Space

Example: Geometric mean is a concave function

Geometric mean: $f(x) = (\prod_{i=1}^n x_i)^{1/n}$ is concave on **dom** $f = \mathbb{R}_{++}^n$

Proof: The Hessian H is given by (Why?)

$$H(k,k) = \frac{\partial^2 f(x)}{\partial x_k^2} = -(n-1)\frac{(\prod_{i=1}^n x_i)^{1/n}}{n^2 x_k^2}$$

$$H(k,l) = \frac{\partial^2 f}{\partial x_k \partial x_l} = \frac{(\prod_{i=1}^n x_i)^{1/n}}{n^2 x_k x_l} \quad \text{for } k \neq l$$

We have

$$\nabla^2 f(x) = -\frac{\prod_{i=1}^n x_i^{1/n}}{n^2}(n\,\mathbf{diag}(1/x_1^2, 1/{x_2}^2, \ldots, 1/{x_n}^2) - qq^T), \quad q_i = 1/x_i$$

We have after using **Cauchy-Schwarz inequality** (How?)

$$v^T \nabla^2 f(x) v = -\frac{\prod_{i=1}^n x_i^{1/n}}{n^2}\left(n\sum_{i=1}^n v_i^2/x_i^2 - \left(\sum_{i=1}^n v_i/x_i\right)^2\right) \leq 0$$

For CS, use $a = \mathbf{1}$ and $b_i = v_i/x_i$. Hence geometric mean is concave!

Example: Log-Determinant is a concave function

$f : S^n_{++} \to 1$

Log-determinant: $f(X) = \log\det X, \quad X > 0$ is a concave function

+ve defn:

$f : \mathbb{R}^{m\times n} \to \mathbb{R}$

$X \to \log\det(X)$

Proof:

- Consider arbitrary line $X = Z + tV, \quad Z, V \in \mathbb{S}^n$ and positive definite

symm. n×n

$g(t)$