SYS 6003: Optimization	**Fall 2016**
Lecture 8	
Instructor: Quanquan Gu	*Date: Sep* 19^{th}

In the following example, we apply the fact that "a function is convex if and only if its restriction to any line is convex" to prove that log determinant function is a concave function (i.e., negative log determinant function is convex).

Example 1 (Log Determinant Function) $f(\mathbf{X}) = \log\det(\mathbf{X})$ *is concave, where* $\mathbf{X}$ *is a positive definite matrix.*
To see that, first define: $g(t) = \log\ \det(\mathbf{X} + t\mathbf{V})$*, such that* $\mathbf{X} + t\mathbf{V}$ *is a positive definite matrix. Since* $\mathbf{X}$ *is positive definite, there exists* $\mathbf{X}^{1/2}$ *such that* $\mathbf{X} = \mathbf{X}^{1/2}\mathbf{X}^{1/2}$*. We then have*

$$\begin{aligned} g(t) &= \log\det(\mathbf{X}^{\frac{1}{2}}\mathbf{X}^{\frac{1}{2}} + t\mathbf{X}^{\frac{1}{2}}\mathbf{X}^{-\frac{1}{2}}\mathbf{V}\mathbf{X}^{-\frac{1}{2}}\mathbf{X}^{\frac{1}{2}}) \\ &= \log\det(\mathbf{X}^{\frac{1}{2}}(\mathbf{I} + t\mathbf{X}^{-\frac{1}{2}}\mathbf{V}\mathbf{X}^{-\frac{1}{2}})\mathbf{X}^{\frac{1}{2}}). \end{aligned}$$

Recall that $\det(\mathbf{AB}) = \det(\mathbf{A})\det(\mathbf{B})$*, it then follows that*

$$\begin{aligned} g(t) &= \log\big(\det(\mathbf{X})\det(\mathbf{I} + t\mathbf{X}^{-\frac{1}{2}}\mathbf{V}\mathbf{X}^{-\frac{1}{2}})\big) \\ &= \log\det(\mathbf{X}) + \log\det(\mathbf{I} + t\mathbf{X}^{-\frac{1}{2}}\mathbf{V}\mathbf{X}^{-\frac{1}{2}}). \end{aligned} \tag{1}$$

Note that $\mathbf{X}$ *and* $\mathbf{X} + t\mathbf{V}$ *are positive semidefinite, so are* $\mathbf{X}^{-1/2}$ *and* $\mathbf{I} + t\mathbf{X}^{-1/2}\mathbf{V}\mathbf{X}^{-1/2}$*. Assume the eigenvalues of* $\mathbf{X}^{-1/2}\mathbf{V}\mathbf{X}^{-1/2}$ *are* $\lambda_1, \lambda_2, \ldots, \lambda_d$*, then*

$$\log\det(\mathbf{I} + t\mathbf{X}^{-\frac{1}{2}}\mathbf{V}\mathbf{X}^{-\frac{1}{2}}) = \log\prod_{i=1}^{d}(1 + t\lambda_i) = \sum_{i=1}^{d}\log(1 + t\lambda_i).$$

Combining this with (1) *gives*

$$g(t) = \log\det(\mathbf{X}) + \sum_{i=1}^{d}\log(1 + t\lambda_i).$$

Notice that the second order derivative of $-g(t)$ *is*

$$-g''(t) = \sum_{i=1}^{d}\frac{\lambda_i^2}{(1 + t\lambda_i)^2} \geq 0.$$

Thus, $-g(t)$ *is convex, so is* $-f(\mathbf{X})$*. We then know that* $f(\mathbf{X})$ *is concave.*

Remark 1 *In the above proof, we do not require* $\mathbf{V}$ *to be positive definite. We only require* $\mathbf{X} + t\mathbf{V}$ *to be positive definite. Therefore,* $\lambda_1, \ldots, \lambda_d$ *are not necessarily positive. But* $1 + t\lambda_1, \ldots, 1 + t\lambda_d$ *are positive.*

Remark 2 *The negative log determinant function is often used as a regularization term in optimization problem, to make* $\mathbf{X}$ *positive semidefinite implicitly.*

Definition 1 (Sublevel set)

$$L_C(f) = \big\{\mathbf{x} : f(\mathbf{x}) \leq C, \mathbf{x} \in \mathbf{dom}(f)\big\},$$

where C *is a constant.*

The following theorem states that the sublevel set of a convex function is a convex set.

Theorem 1 *If* f *is convex, then* $L_C(f)$ *is convex.*

Proof: For any $\mathbf{x}_1, \mathbf{x}_2 \in L_C(f)$ and any $\alpha \in [0,1]$, we have

$$f(\alpha\mathbf{x}_1 + (1-\alpha)\mathbf{x}_2) \leq \alpha f(\mathbf{x}_1) + (1-\alpha) f(\mathbf{x}_2) \leq \alpha C + (1-\alpha)C = C,$$

which implies that $\alpha\mathbf{x}_1 + (1-\alpha)\mathbf{x}_2 \in L_C(f)$. So, $L_C(f)$ is a convex set. ■

Note that Theorem 1 gives a necessary condition for convex functions. In other words, the converse of Theorem 1 does not necessarily hold. Quasi convex set is an example to show that why the the converse is not true.

Definition 2 (Quasi-convex) $f : \mathbb{R}^d \to \mathbb{R}$ *is a Quasi-convex function if,*
(i) its domain $\mathbf{dom} f$ *is a convex set;*
(ii) its sublevel set is a convex set for any $C \in \mathbb{R}$.

Note that convex function is quasi convex, but the converse is not necessarily true.
An example of quasi-convex function is the linear fractional function.

Example 2 (Linear Fractional Function)

$$f(\mathbf{x}) = \frac{\mathbf{a}^\top\mathbf{x} + b}{\mathbf{c}^\top\mathbf{x} + u}$$

is quasi-convex, where $\mathbf{x}, \mathbf{a}, \mathbf{c} \in \mathbb{R}^d, b, u \in \mathbb{R}$, $\mathbf{dom}(f) = \{\mathbf{x} : \mathbf{c}^\top\mathbf{x} + u \geq 0\}$. *Note that* $\mathbf{dom}(f)$ *is a half space and so is convex. For any real number* α, *consider the sublevel set:*

$$\begin{aligned} L_\alpha(f) &= \left\{\mathbf{x} : \frac{\mathbf{a}^\top\mathbf{x} + b}{\mathbf{c}^\top\mathbf{x} + u} \leq \alpha\right\} \\ &= \{\mathbf{x} : \mathbf{a}^\top\mathbf{x} + b \leq \alpha\mathbf{c}^\top\mathbf{x} + \alpha u\} \\ &= \{\mathbf{x} : (\mathbf{a} - \alpha\mathbf{c})^\top\mathbf{x} + b - \alpha u \leq 0\}. \end{aligned}$$

It is clear that $L_\alpha(f)$ *is a half space. Hence it is a convex set.*

A concept which is very likely to be confused with sublevel set is epigraph.

Definition 3 (Epigraph) *Epigraph of function* f *is* $\mathbf{epi}(f) = \big\{(\mathbf{x}, t) : f(\mathbf{x}) \leq t\big\}$.

The following theorems states that the convexity of a function and its epigraph are equivalent. They can be derived from one another.

Theorem 2 *Function f is convex if and only if its epigraph is a convex set.*

Proof: (i) "$\Rightarrow$" : f is convex $\Rightarrow$ its epigraph is convex.
For any

$$\begin{pmatrix} \mathbf{x}_1 \\ t_1 \end{pmatrix}, \begin{pmatrix} \mathbf{x}_2 \\ t_2 \end{pmatrix} \in \mathbf{epi}(f),$$

and any $\alpha \in [0, 1]$, we want to show

$$\alpha \begin{pmatrix} \mathbf{x}_1 \\ t_1 \end{pmatrix} + (1-\alpha) \begin{pmatrix} \mathbf{x}_2 \\ t_2 \end{pmatrix} = \begin{pmatrix} \alpha\mathbf{x}_1 + (1-\alpha)\mathbf{x}_2 \\ \alpha t_1 + (1-\alpha)t_2 \end{pmatrix} \in \mathbf{epi}(f).$$

This is easy to prove since we have

$$f(\alpha\mathbf{x}_1 + (1-\alpha)\mathbf{x}_2) \leq \alpha f(\mathbf{x}_1) + (1-\alpha)f(\mathbf{x}_2) \leq \alpha t_1 + (1-\alpha)t_2.$$

(ii) "$\Leftarrow$" The epigraph of f is convex $\Rightarrow$ f is convex.
We want to show that $f(\alpha\mathbf{x}_1 + (1-\alpha)\mathbf{x}_2) \leq \alpha f(\mathbf{x}_1) + (1-\alpha)f(\mathbf{x}_2)$. Let us choose $t_1 = f(\mathbf{x}_1), t_2 = f(\mathbf{x}_2)$, then $(\mathbf{x}_1, t_1), (\mathbf{x}_2, t_2) \in \mathbf{epi}(f)$. By the convexity of $\mathbf{epi}(f)$, we have

$$\alpha \begin{pmatrix} \mathbf{x}_1 \\ t_1 \end{pmatrix} + (1-\alpha) \begin{pmatrix} \mathbf{x}_2 \\ t_2 \end{pmatrix} \in \mathbf{epi}(f), \quad i.e., \quad \begin{pmatrix} \alpha\mathbf{x}_1 + (1-\alpha)\mathbf{x}_2 \\ \alpha t_1 + (1-\alpha)t_2 \end{pmatrix} \in \mathbf{epi}(f),$$

which implies that $f(\alpha\mathbf{x}_1 + (1-\alpha)\mathbf{x}_2) \leq \alpha t_1 + (1-\alpha)t_2 = \alpha f(\mathbf{x}_1) + (1-\alpha)f(\mathbf{x}_2)$. Due to the arbitrariness of $\mathbf{x}_1, \mathbf{x}_2$ and α, we know $f(\mathbf{x})$ is indeed convex. ■

So far, we have learned different proof techniques for convex functions, which are summarized as follows

1) By definition
2) By first-order condition: $f(\mathbf{y}) \geq f(\mathbf{x}) + \nabla f(\mathbf{x})^\top (\mathbf{y} - \mathbf{x})$
3) By second-order condition: $\nabla^2 f(\mathbf{x}) \succeq 0$
4) By restriction of a function to any line)
5) By epigraphy(Theorem 2)

Before I am going to introduce one more proof technique for convex functions, I first introduce the extended value extension of a convex function.

Recall the definition of convex function: A function $f : \mathbb{R}^d \to \mathbb{R}$ is convex if $\mathbf{dom}f$ is convex, and for any $\mathbf{x}, \mathbf{y} \in \mathbf{dom}f$ and any $\alpha \in [0, 1]$ we have

$$f(\alpha\mathbf{x} + (1-\alpha)\mathbf{y}) \leq \alpha f(\mathbf{x}) + (1-\alpha)f(\mathbf{y}).$$

As can be seen, we require $\mathbf{dom}f$ is a convex set, and the inequality holds for any $\mathbf{x}, \mathbf{y} \in \mathbf{dom}f$. In order to simplify the notation, we introduce the extended value extension of convex functions, which extends the domain of a convex function $f : \mathbb{R}^d \to \mathbb{R}$ from $\mathbf{dom}f$ to $\mathbb{R}^d$.

Definition 4 (Extended Value Extension) *Suppose $f : \mathbb{R}^d \to \mathbb{R}$ is convex, the extended value extension of f, denoted by $\tilde{f}$ is:*

$$\tilde{f}(\mathbf{x}) = \begin{cases} f(\mathbf{x}), & \text{if } \mathbf{x} \in \mathbf{dom} f \\ \infty, & \text{if } \mathbf{x} \notin \mathbf{dom} f \end{cases}. \tag{2}$$

Now if a function $f : \mathbb{R}^d \to \mathbb{R}$ is convex, we have for any $\mathbf{x}, \mathbf{y} \in \mathbb{R}^d$ and any $\alpha \in [0, 1]$,

$$\tilde{f}(\alpha\mathbf{x} + (1-\alpha)\mathbf{y}) \leq \alpha\tilde{f}(\mathbf{x}) + (1-\alpha)\tilde{f}(\mathbf{y}).$$

In addition, extended value extension will simplify the argument in some proofs. We will see it in some lectures.

Now, I am going to introduce another proof technique for convex functions, which is based on convexity preserving operation.

Convexity Preserving Operations for Convex Functions

1. Nonnegative Scaling: αf is convex, if f is convex and $\alpha \geq 0$.

2. Sum: $f_1 + f_2$ is convex in $\mathbf{dom}\ f_1 \cap \mathbf{dom}\ f_2$, if f_1 is convex and f_2 is convex.

3. Composition with affine function: If f is convex, then $f(\mathbf{A}\mathbf{x} + \mathbf{b})$ is convex in $\{\mathbf{x} | \mathbf{A}\mathbf{x} + \mathbf{b} \in \mathbf{dom}\ f\}$.

   Example 3 $f(\mathbf{x}) = -\sum_{i=1}^m \log(b_i - \mathbf{a}_i^\top \mathbf{x})$, *where* $\mathbf{a}_i \in \mathbb{R}^d, \mathbf{x} \in \mathbb{R}^d$, *and* $b_i \in \mathbb{R}$ *is convex.*

   Example 4 $f(\mathbf{x}) = \|\mathbf{A}\mathbf{x} + \mathbf{b}\|_2$ *is convex.*

4. Pointwise Maximum: If $f_1, f_2, ..., f_m$ are convex, then $f(\mathbf{x}) = \max\{f_1(\mathbf{x}), ..., f_m(\mathbf{x})\}$ is convex on $\mathbf{dom}\ f = \cap_{i=1}^m \mathbf{dom}\ f_i$
   Proof: For any $\mathbf{x}, \mathbf{y} \in \mathbf{dom}\ f$, and any $\alpha \in [0, 1]$, we have

$$\begin{aligned} f(\alpha\mathbf{x} + (1-\alpha)\mathbf{y}) &= \max\{f_1(\alpha\mathbf{x} + (1-\alpha)\mathbf{y}), ..., f_m(\alpha\mathbf{x} + (1-\alpha)\mathbf{y})\} \\ &\leq \max\{\alpha f_1(\mathbf{x}) + (1-\alpha)f_1(\mathbf{y}), ..., \alpha f_m(\mathbf{x}) + (1-\alpha)f_m(\mathbf{y})\} \\ &= \max_{1\leq i\leq m}[\alpha f_i(\mathbf{x}) + (1-\alpha)f_i(\mathbf{y})]. \end{aligned}$$

   Note that

$$\begin{aligned} \max_{1\leq i\leq m}[\alpha f_i(\mathbf{x}) + (1-\alpha)f_i(\mathbf{y})] &\leq \alpha \max_{1\leq i\leq m}\{f_i(\mathbf{x})\} + (1-\alpha) \max_{1\leq i\leq m}\{f_i(\mathbf{y})\} \\ &= \alpha f(\mathbf{x}) + (1-\alpha)f(\mathbf{y}). \end{aligned}$$

   Thus, by definition, $f(\mathbf{x})$ is convex. ■