Last Update: 03/18/2021, 7:38pm

ECE 236B Convex Optimization (Notes)

Winter 2021, Instructor: Lieven Vandenberghe, TA: Xin Jiang, UCLA
Notetaker: Zhiying Li

for review purpose only

Contents

1 Schur Complement

2 Convex Sets

There are not too many examples

3 Convex Functions

Commonly, the most important concept or a topic in this part is: **how to prove a function is a convex function?** And there are a lots of methods to prove the convexity of a function.

3.1 Verifying the definition

Definition. (convex function) $f : \mathbf{R}^n \to \mathbf{R}$ is convex if

1. $\mathrm{dom}\, f$ is a convex set
2. (Jenson's inequality) $f(\theta x + (1-\theta)y) \le \theta f(x) + (1-\theta)f(y)$ for all $x, y \in \mathrm{dom}\, f,\ 0 \le \theta \le 1$

remark.

1. Students tend to ignore the first requirement ($\mathrm{dom}\, f$ is a convex set).
2. (Alternative definition) Many machine learning textbooks (are lazy and) tend to use definition that only involves Jenson's inequality:

$$f(\theta x + (1-\theta)y) \le \theta f(x) + (1-\theta)f(y) \text{ for all } x, y \in \mathrm{dom}\, f,\ 0 \le \theta \le 1$$

   Their arguments tends to be: by writing $f(\theta x + (1-\theta)y)$, it implicitly implies $\theta x + (1-\theta)y \in \mathrm{dom}\, f$ for all $x, y \in \mathrm{dom}\, f$, and hence $\mathrm{dom}\, f$ is a convex set.

3. **Useful Property**: Extended-value extension $\tilde{f}$ of f:

$$\tilde{f}(x) = \begin{cases} f(x) & x \in \mathrm{dom}\, f \\ \infty & x \notin \mathrm{dom}\, f \end{cases}$$

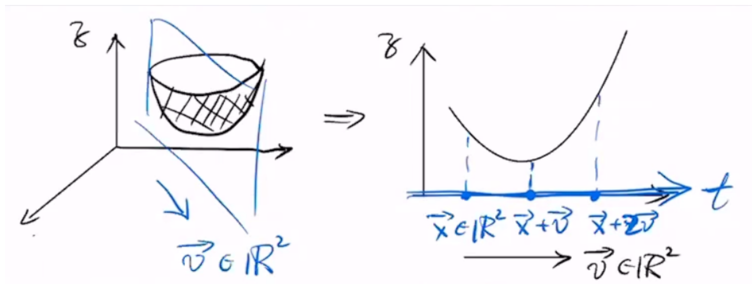
   often simplifies notation:

$$\tilde{f}(\theta x + (1-\theta)y) \le \theta \tilde{f}(x) + (1-\theta)\tilde{f}(y) \text{ for all } \cancel{x, y \in \mathrm{dom}\, f},\ 0 \le \theta \le 1$$

Practically, we rarely use this definition to prove complexity on non-trivial problem, because directly using defintion either makes most non-trivial cases too complicated, and we normally alternatively have some more advanced tools.

However, this definition would imply a trick of restricting to a arbitrary line:

3.2 Restricting to an arbitrary line



As is illustrated by the graph above, we can check convexity of f by checking convexity of functions of one variable, making use of the following property:

> $f : \mathbf{R}^n \to \mathbf{R}$ is convex if and only if the function $g : \mathbf{R} \to \mathbf{R}$
>
> $$g(t) = f(x + tv), \qquad \operatorname{dom} g = \{t \mid x + tv \in \operatorname{dom} f\}$$
>
> is convex (in t) for any $x \in \operatorname{dom} f, \quad v \in \mathbf{R}^n$.

remark.

1. t does not have to be in $\operatorname{dom} f$.
2. It reduces checking Jenson's inequality in high dimensional to 1-dimensinoal problem. So it's espeically useful for matrix functions (functions of eigenvalue, or eigen-function/ spectral function). Such as examples below:

Example. $f(x) = \operatorname{tr} X^{-1} = \sum_{i=1}^{n} \frac{1}{\lambda_i(X)}$, with $\operatorname{dom} f = \mathbf{S}^n_{++}$ is convex.

Proof. Consider $X \in \mathbf{S}^n_{++}$, and $t > 0$, and $V \in \mathbf{R}^n$ such that $X + tV \in \mathbf{S}^n_{++}$, we want to show that $g(t) = \operatorname{tr}(X + tV)^{-1}$ is convex.
We use the fact that $X \in \mathbf{S}^n_{++} \implies X = X^{1/2}X^{1/2}$, then

$$\begin{aligned}
g(t) = \operatorname{tr}(X + tV)^{-1} &= \operatorname{tr}(X^{1/2}X^{1/2} + tV)^{-1} \\
&= \operatorname{tr}\left[X^{1/2}\left(I + tX^{-1/2}VX^{-1/2}\right)X^{1/2}\right]^{-1} \\
&= \operatorname{tr}\left[X^{-1/2}\left(I + tX^{-1/2}VX^{-1/2}\right)^{-1}X^{-1/2}\right] \\
&= \operatorname{tr}\left[X^{-1}\left(I + tX^{-1/2}VX^{-1/2}\right)^{-1}\right]
\end{aligned}$$

Since $X + tV \in \mathbf{S}^n_{++}$ and $X \in \mathbf{S}^n_{++}$, these implies $V \in \mathbf{S}^n$. Then, $X^{-1/2}VX^{-1/2} \in \mathbf{S}^n$. Then we can apply eigen-decomposition: $X^{-1/2}VX^{-1/2} = Q\Lambda Q^T$, then

$$\begin{aligned}
g(t) = \operatorname{tr}(X + tV)^{-1} &= \operatorname{tr}\left[X^{-1}\left(I + tQ\Lambda Q^T\right)^{-1}\right] \\
&= \operatorname{tr}\left[X^{-1}Q\left(I + t\Lambda\right)^{-1}Q^T\right] \\
&= \operatorname{tr}\left[Q^TX^{-1}Q\left(I + t\Lambda\right)^{-1}\right]
\end{aligned}$$

If we let $Y = Q^T X^{-1} Q$, then notice $Y \in \mathbf{S}_{++}^n$. Also notice Y and $(I + t\Lambda)$ are diagnol matrix.

$$\begin{aligned} g(t) = \operatorname{tr}(X + tV)^{-1} &= \operatorname{tr}\left[Y\left(I + t\Lambda\right)^{-1}\right] \\ &= \sum_{i=1}^{n} \frac{Y_{ii}}{1 + t\lambda_i} \end{aligned}$$

Since $Y \in \mathbf{S}_{++}^n$ and diagnal, then $Y_{ii} > 0$ for all i.
Since $X + tV \in \mathbf{S}_{++}^n$ and $X + tV = X^{1/2}\left(I + tX^{-1/2}VX^{-1/2}\right)X^{1/2}$ and $X^{1/2} \in \mathbf{S}_{++}^n$, then we have $\left(I + tX^{-1/2}VX^{-1/2}\right) \succ 0$, hence $1 + t\lambda_i > 0$ for all i. Therefore, $g(t)$ is a convex function. And this proves $\operatorname{tr} X^{-1}$ is convex. □

Example. $f(x) = \log\det X = \sum_{i=1}^{n} \log \lambda_i(X)$, with $\operatorname{dom} f = \mathbf{S}_{++}^n$ is concave.

Proof.

$$\begin{aligned} g(t) = \log\det(X + tV) &= \log\det\left(X^{1/2}\left(I + tX^{-1/2}VX^{-1/2}\right)X^{1/2}\right) \\ &= \log\left[\det\left(X^{1/2}\right)\det\left(I + tX^{-1/2}VX^{-1/2}\right)\det\left(X^{1/2}\right)\right] \\ &= \log\left[\det X \det\left(I + tX^{-1/2}VX^{-1/2}\right)\right] \\ &= \log\det X + \log\det\left(I + tX^{-1/2}VX^{-1/2}\right) \end{aligned}$$

Again, let $X^{-1/2}VX^{-1/2} \in \mathbf{S}^n$, and by eigen-decomposition, $X^{-1/2}VX^{-1/2} = Q\Lambda Q^T$, then

$$\begin{aligned} g(t) &= \log\det X + \log\det\left[Q(I + t\Lambda)Q^T\right] \\ &= \log\det X + \log\det\left(I + t\Lambda\right) \\ &= \log\det X + \log\prod_{i=1}^{n}(1 + t\lambda_i) \\ &= \log\det X + \sum_{i=1}^{n}\log(1 + t\lambda_i) \end{aligned}$$

Since $X + tV \in \mathbf{S}_{++}^n$ and $X + tV = X^{1/2}\left(I + tX^{-1/2}VX^{-1/2}\right)X^{1/2}$ and $X^{1/2} \in \mathbf{S}_{++}^n$, then we have $\left(I + tX^{-1/2}VX^{-1/2}\right) \succ 0$, hence $1 + t\lambda_i > 0$ for all i. Therefore, $g(t)$ is a concave function. And this proves $\log\det X$ is concave. □

3.3 Hessian (second-order conditions)

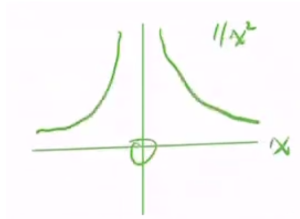
(second-order conditions) f is convex if and only if

1. f is twice differentiable
2. $\operatorname{dom} f$ is a convex set
3. $\nabla^2 f(x) \succ 0 \quad$ for all $x \in \operatorname{dom} f$

remark.

1. "dom f being convex set" is a condition that *cannot* be ignore.

   Counterexample: $f(x) = \dfrac{1}{x^2}$, $\quad \operatorname{dom} f = \{x \in \mathbf{R} \mid x \neq 0\}$. The Hessian $f''(x) = \dfrac{6}{x^4}$ for all $x \in \operatorname{dom} f$. However, dom f is not convex.



2. Need to be differentiable, counterexample: $f(x) = |x|$

3. This method is useful for:

   - Trivial cases: scalar functions, $\nabla^2 f \in \mathbf{S}^2$, etc
     - *Quadratic function*: $f(x) = (1/2)x^T Px + q^T x + r$ (with $P \in \mathbf{S}^n$)

       $$\nabla f(x) = Px + q, \quad \nabla^2 f(x) = P$$

       convex if $P \geq 0$
     - *Least squares objective*: $f(x) = \|Ax - b\|_2^2$

       $$\nabla f(x) = 2A^T(Ax - b), \quad \nabla^2 f(x) = 2A^T A$$

       convex (for any A)
     - *Quadratic-over-linear function*: $f(x, y) = x^2/y$

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

       convex for $y > 0$
   - "cyclicly symmetric" cases:
     - *Log-sum-exp function*: $f(x) = \log \sum_{k=1}^{n} \exp x_k$
     - *Geometric mean*: $f(x) = \left(\prod_{k=1}^{n} x_k \right)^{1/n}$

4. Although we can prove convexity by the first-order condition, in practice, we mostly use First-Order Condition as a property.

   (first-order conditions) f is convex if and only if

   (a) f is differentiable

   (b) dom f is a convex set

   (c) $f(y) \geqslant f(x) + \nabla f(x)^T(y - x) \quad$ for all $x \in \operatorname{dom} f$

3.4 Operations preserving convxity

3.4.1 Nonnegative weighted sum

If f_i for $i = 1, \ldots, m$ is convex, and $\alpha_i \geqslant 0$ for $i = 1, \ldots, m$, then $f = \sum_{i=1}^{m} \alpha_i f_i$ is convex.

3.4.2 Composition with affine mapping

$f(Ax + b)$ is convex if f is convex.

remark.

1. If matrix $A = [a_1, a_2, \ldots, a_n]$, then this is a affine mapping of $x = (x_1, x_2, \ldots, x_n)$. i.e.

$$Ax + b = x_1 a_1 + x_2 a_2 + \cdots + x_n a_n + b$$

2. This property can be more general, in the way that the affine mapping can also be with respect to matrix $A_i, B \in \mathbf{S}^m$ (*or* $\mathbf{R}^{p \times q}$ *or signals in general*):

$$\mathcal{A}(x) + B = x_1 A_1 + x_2 A_2 + \cdots + x_n A_n + B$$

   Notice, in this case $g(x) = f(\mathcal{A}(x) + B)$, the functions $f : \mathbf{S}^m \to \mathbf{R}$, and $g : \mathbf{R}^n \to \mathbf{R}$.

Example. We know $f(Z) = -\log\det Z$ is convex on $\mathbf{S}^m_{++}$. Then,

$$\begin{aligned} g(x) &= f(\mathcal{A}(x) + A_0) \\ &= -\log\det\left(x_1 A_1 + x_2 A_2 + \cdots + x_n A_n + A_0\right) \end{aligned}$$

is a convex function with $\operatorname{dom} g = \{x \in \mathbf{R}^n \mid x_1 A_1 + x_2 A_2 + \cdots + x_n A_n + A_0 \succ 0\}$.

3.4.3 Pointwise maximum

if $f_1, \ldots, f_m$ are convex, then $f(x) = \max\{f_1(x), \ldots, f_m(x)\}$ is convex

Example 1. (*piecewise-linear function*): $f(x) = \max_{i=1,\ldots,m}\left(a_i^T x + b_i\right)$ is convex
Example 2. (*sum of r largest components of $x \in \mathbf{R}^n$*):

$$f(x) = x_{[1]} + x_{[2]} + \cdots + x_{[r]}$$

is convex $\left(x_{[i]}\right.$ is i th largest component of $\left.x\right)$

Proof. $f(x) = \max\{x_{i_1} + x_{i_2} + \cdots + x_{i_r} \mid 1 \leq i_1 < i_2 < \cdots < i_r \leq n\}$ □

Example 3. (*general form of Example 2.*): If $\alpha_1 \geqslant \alpha_2 \geqslant \cdots \geqslant \alpha_r$, then

$$f(x) = \alpha_1 x_{[1]} + \alpha_2 x_{[2]} + \cdots + \alpha_r x_{[r]}$$

is convex $\left(x_{[i]}\right.$ is i th largest component of $\left.x\right)$

3.4.4 Pointwise supremum

If $f(x, y)$ is convex in x for each $y \in \mathcal{A}$, then $g(x) = \sup_{y \in \mathcal{A}} f(x, y)$ is convex.

remark.

1. $\mathcal{A}$ does not have to be convex.
2. $f(x, y)$ does not have to be jointly convex in (x, y). $f(x, y)$ only has to be convex for x when given y.

3.4.5 Composition rule

(**Composition with scalar functions**) composition of $g : \mathbf{R}^n \to \mathbf{R}$ and $h : \mathbf{R} \to \mathbf{R}$:

$$f(x) = h(g(x))$$

f is convex if

- g convex, h convex, $\tilde{h}$ nondecreasing
- g concave, h convex, $\tilde{h}$ nonincreasing

remark.

(**Vector composition**) composition of $g : \mathbf{R}^n \to \mathbf{R}^k$ and $h : \mathbf{R}^k \to \mathbf{R}$:

$$f(x) = h(g(x)) = h\left(g_1(x), g_2(x), \ldots, g_k(x)\right)$$

f is convex if

- g_i convex, h convex, $\tilde{h}$ nondecreasing in each argument
- g_i concave, h convex, $\tilde{h}$ nonincreasing in each argument

remark.

3.4.6 Minimization

If $f(x, y)$ is convex in (x, y) and C is a convex set, then $g(x) = \inf_{y \in C} f(x, y)$ is convex.

remark.

3.4.7 Persepective

The perspective of a function $f : \mathbf{R}^n \to \mathbf{R}$ is the function $g : \mathbf{R}^n \times \mathbf{R} \to \mathbf{R}$

$$g(x, t) = t f(x/t), \quad \operatorname{dom} g = \{(x, t) \mid x/t \in \operatorname{dom} f, t > 0\}$$

g is convex if f is convex

remark.

3.4.8 Epigraph

Epigraph of $f : \mathbf{R}^n \to \mathbf{R}$

$$\mathrm{epi}\, f = \left\{(x,t) \in \mathbf{R}^{n+1} \mid x \in \mathrm{dom}\, f, f(x) \le t\right\}$$

f is convex if and only if epi f is a convex set

remark.

4 Convex Optimization Problems

4.1 Equivalent problem

4.2 Robust QP

4.3 SOCP

4.4 Geometric Programing

4.5 SDP

5 Duality

5.1 Lagrange Dual

5.2 Two-way partition example

5.3 Strong Duality and Slater's Condition

5.3.1 Geometric Interpretation

5.4 Optimality Conditions

5.5 Examples of Duality

5.5.1 Example 1. Summation of r largest elements

5.5.2 Example 2. Duality and SDPs

5.5.3 Example 3. Exact Penalty

6 Algorithms

6.1 Unconstrained Minimization

6.2 Minimization with Equality Constraints

6.3 Interior Point Method