Peeter Joot
peeterjoot@protonmail.com

ECE1505H Convex Optimization. Lecture 7: Examples of convex and concave functions, local and global minimums. Taught by Prof. Stark Draper

Disclaimer Peeter's lecture notes from class. These may be incoherent and rough.

These are notes for the UofT course ECE1505H, Convex Optimization, taught by Prof. Stark Draper, from [1].

Today

- Local and global optimality
- Compositions of functions
- Examples

Example:

$$\begin{aligned} F(x) &= x^2 \\ F''(x) &= 2 > 0 \end{aligned} \tag{1.1}$$

strictly convex.

Example:

$$\begin{aligned} F(x) &= x^3 \\ F''(x) &= 6x. \end{aligned} \tag{1.2}$$

Not always non-negative, so not convex. However x^3 is convex on $\operatorname{dom} F = \mathbb{R}_+$.

Example:

$$\begin{aligned} F(x) &= x^\alpha \\ F'(x) &= \alpha x^{\alpha - 1} \\ F''(x) &= \alpha(\alpha - 1) x^{\alpha - 2}. \end{aligned} \tag{1.3}$$

This is convex on $\mathbb{R}_+$, if $\alpha \ge 1$, or $\alpha \le 0$.

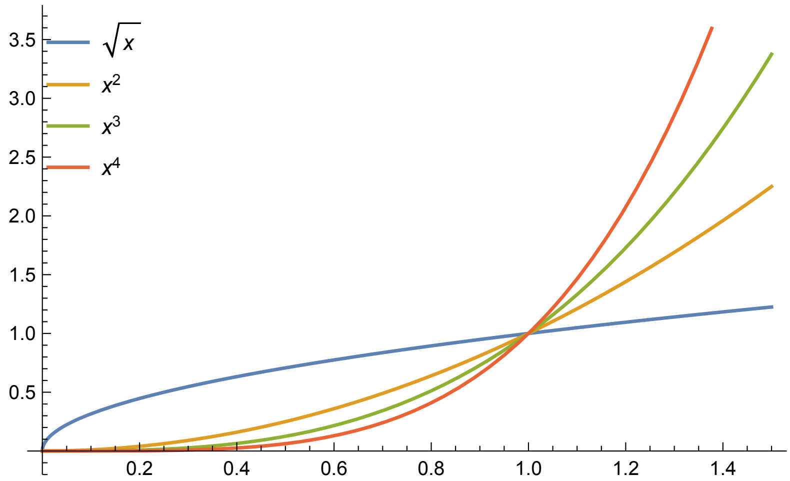

Figure 1.1: Powers of x.

Example:

$$\begin{aligned} F(x) &= \log x \\ F'(x) &= \frac{1}{x} \\ F''(x) &= -\frac{1}{x^2} \leq 0 \end{aligned} \tag{1.4}$$

This is concave.

Example:

$$\begin{aligned} F(x) &= x \log x \\ F'(x) &= \log x + x\frac{1}{x} = 1 + \log x \\ F''(x) &= \frac{1}{x} \end{aligned} \tag{1.5}$$

This is strictly convex on $\mathbb{R}_{++}$, where $F''(x) \geq 0$.

Example:

$$\begin{aligned} F(x) &= e^{\alpha x} \\ F'(x) &= \alpha e^{\alpha x} \\ F''(x) &= \alpha^2 e^{\alpha x} \geq 0 \end{aligned} \tag{1.6}$$

Such functions are plotted in fig. 1.2, and are convex function for all α.

Example: For symmetric $P \in S^n$

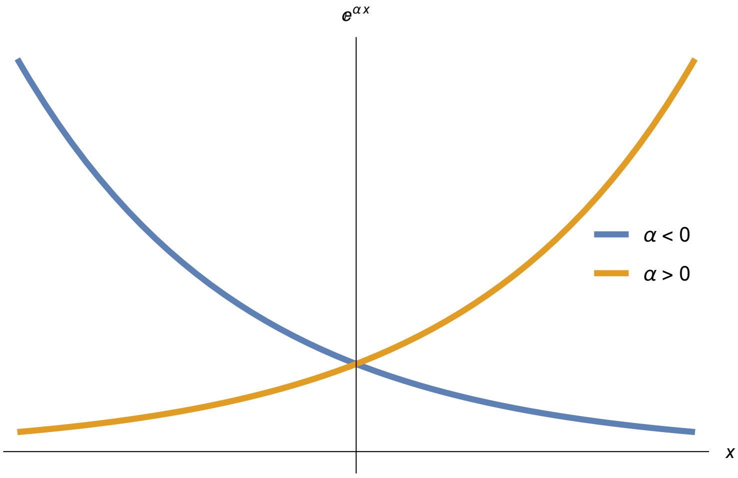

Figure 1.2: Exponential.

$$\begin{aligned} F(\mathbf{x}) &= \mathbf{x}^{\mathrm{T}} P \mathbf{x} + 2\mathbf{q}^{\mathrm{T}} \mathbf{x} + r \\ \boldsymbol{\nabla} F &= (P + P^{\mathrm{T}})\mathbf{x} + 2\mathbf{q} = 2P\mathbf{x} + 2\mathbf{q} \\ \boldsymbol{\nabla}^2 F &= 2P. \end{aligned} \tag{1.7}$$

This is convex(concave) if $P \geq 0$ ($P \leq 0$).

Example: A quadratic function

$$F(x, y) = x^2 + y^2 + 3xy, \tag{1.8}$$

that is neither convex nor concave is plotted in fig. 1.3

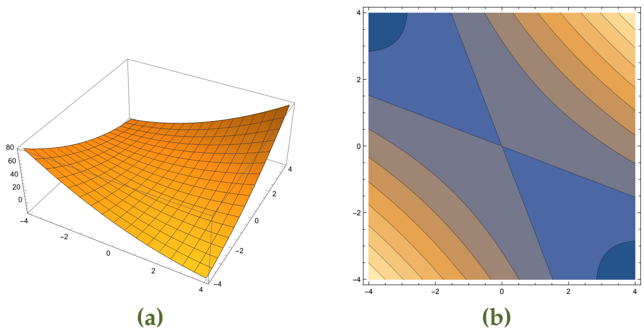

Figure 1.3: Function with saddle point (3d and contours).

This function can be put in matrix form

$$\begin{aligned} F(x, y) &= x^2 + y^2 + 3xy \\ &= \begin{bmatrix} x & y \end{bmatrix} \begin{bmatrix} 1 & 1.5 \\ 1.5 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}, \end{aligned} \tag{1.9}$$

and has the Hessian

$$\begin{aligned} \nabla^2 F &= \begin{bmatrix} \partial_{xx}F & \partial_{xy}F \\ \partial_{yx}F & \partial_{yy}F \end{bmatrix} \\ &= \begin{bmatrix} 2 & 3 \\ 3 & 2 \end{bmatrix} \\ &= 2P. \end{aligned} \tag{1.10}$$

From the plot we know that this is not PSD, but this can be confirmed by checking the eigenvalues

$$\begin{aligned} 0 &= \det(P - \lambda I) \\ &= (1-\lambda)^2 - 1.5^2, \end{aligned} \tag{1.11}$$

which has solutions

$$\begin{aligned} \lambda &= 1 \pm \frac{3}{2} \\ &= \frac{3}{2}, -\frac{1}{2}. \end{aligned} \tag{1.12}$$

This is not PSD nor negative semi-definite, because it has one positive and one negative eigenvalues. This is neither convex nor concave.

Along $y = -x$,

$$\begin{aligned} F(x, y) &= F(x, -x) \\ &= 2x^2 - 3x^2 \\ &= -x^2, \end{aligned} \tag{1.13}$$

so it is concave along this line. Along $y = x$

$$\begin{aligned} F(x, y) &= F(x, x) \\ &= 2x^2 + 3x^2 \\ &= 5x^2, \end{aligned} \tag{1.14}$$

so it is convex along this line.

Example:

$$F(\mathbf{x}) = \sqrt{x_1 x_2}, \tag{1.15}$$

on $\operatorname{dom} F = \{x_1 \geq 0, x_2 \geq 0\}$

For the Hessian

$$\begin{aligned} \frac{\partial F}{\partial x_1} &= \frac{1}{2} x_1^{-1/2} x_2^{1/2} \\ \frac{\partial F}{\partial x_2} &= \frac{1}{2} x_2^{-1/2} x_1^{1/2} \end{aligned} \tag{1.16}$$

The Hessian components are

$$
\begin{aligned}
\frac{\partial}{\partial x_1}\frac{\partial F}{\partial x_1} &= -\frac{1}{4}x_1^{-3/2}x_2^{1/2} \\
\frac{\partial}{\partial x_1}\frac{\partial F}{\partial x_2} &= \frac{1}{4}x_2^{-1/2}x_1^{-1/2} \\
\frac{\partial}{\partial x_2}\frac{\partial F}{\partial x_1} &= \frac{1}{4}x_1^{-1/2}x_2^{-1/2} \\
\frac{\partial}{\partial x_2}\frac{\partial F}{\partial x_2} &= -\frac{1}{4}x_2^{-3/2}x_1^{1/2}
\end{aligned}
\tag{1.17}
$$

or

$$
\boldsymbol{\nabla}^2 F = -\frac{\sqrt{x_1 x_2}}{4}\begin{bmatrix} \frac{1}{x_1^2} & -\frac{1}{x_1 x_2} \\ -\frac{1}{x_1 x_2} & \frac{1}{x_2^2} \end{bmatrix}. \tag{1.18}
$$

Checking this for PSD against $\mathbf{v} = (v_1, v_2)$, we have

$$
\begin{aligned}
\begin{bmatrix} v_1 & v_2 \end{bmatrix}\begin{bmatrix} \frac{1}{x_1^2} & -\frac{1}{x_1 x_2} \\ -\frac{1}{x_1 x_2} & \frac{1}{x_2^2} \end{bmatrix}\begin{bmatrix} v_1 \\ v_2 \end{bmatrix} &= \begin{bmatrix} v_1 & v_2 \end{bmatrix}\begin{bmatrix} \frac{1}{x_1^2}v_1 - \frac{1}{x_1 x_2}v_2 \\ -\frac{1}{x_1 x_2}v_1 + \frac{1}{x_2^2}v_2 \end{bmatrix} \\
&= \left(\frac{1}{x_1^2}v_1 - \frac{1}{x_1 x_2}v_2\right)v_1 + \left(-\frac{1}{x_1 x_2}v_1 + \frac{1}{x_2^2}v_2\right)v_2 \\
&= \frac{1}{x_1^2}v_1^2 + \frac{1}{x_2^2}v_2^2 - 2\frac{1}{x_1 x_2}v_1 v_2 \\
&= \left(\frac{v_1}{x_1} - \frac{v_2}{x_2}\right)^2 \\
&\ge 0,
\end{aligned}
\tag{1.19}
$$

so $\boldsymbol{\nabla}^2 F \le 0$. This is a negative semi-definite function (concave). Observe that this check required checking PSD for all values of $\mathbf{x}$.

This is an example of a more general result

$$
F(x) = \left(\prod_{i=1}^{n} x_i\right)^{1/n}, \tag{1.20}
$$

which is concave (prove on homework).

Summary. If F is differentiable in $\mathbb{R}^n$, then check the curvature of the function along all lines. i.e. At all locations and in all directions.

If the Hessian is PSD at all $\mathbf{x} \in \operatorname{dom} F$, that is

$$
\boldsymbol{\nabla}^2 F \ge 0 \, \forall \mathbf{x} \in \operatorname{dom} F, \tag{1.21}
$$

then the function is convex.

more examples of convex, but not necessarily differentiable functions

Example: Over dom $F = \mathbb{R}^n$

$$F(\mathbf{x}) = \max_{i=1}^{n} x_i \tag{1.22}$$

i.e.

$$\begin{aligned} F((1,2) &= 2 \\ F((3,-1) &= 3 \end{aligned} \tag{1.23}$$

Example:

$$F(\mathbf{x}) = \max_{i=1}^{n} F_i(\mathbf{x}), \tag{1.24}$$

where

$$F_i(\mathbf{x}) = ...? \tag{1.25}$$

max of a set of convex functions is a convex function.

Example:

$$F(x) = x_{[1]} + x_{[2]} + x_{[3]} \tag{1.26}$$

where
$x_{[k]}$ is the k-th largest number in the list
Write

$$F(x) = \max x_i + x_j + x_k \tag{1.27}$$

$$(i, j, k) \in \binom{n}{3} \tag{1.28}$$

Example: For $\mathbf{a} \in \mathbb{R}^n$ and $b_i \in \mathbb{R}$

$$\begin{aligned} F(\mathbf{x}) &= \sum_{i=1}^{n} \log(b_i - \mathbf{a}^{\mathrm{T}}\mathbf{x})^{-1} \\ &= -\sum_{i=1}^{n} \log(b_i - \mathbf{a}^{\mathrm{T}}\mathbf{x}) \end{aligned} \tag{1.29}$$

This $b_i - \mathbf{a}^{\mathrm{T}}\mathbf{x}$ is an affine function of $\mathbf{x}$ so it doesn't affect convexity.
Since log is concave, $-\log$ is convex. Convex functions of affine function of $\mathbf{x}$ is convex function of $\mathbf{x}$.

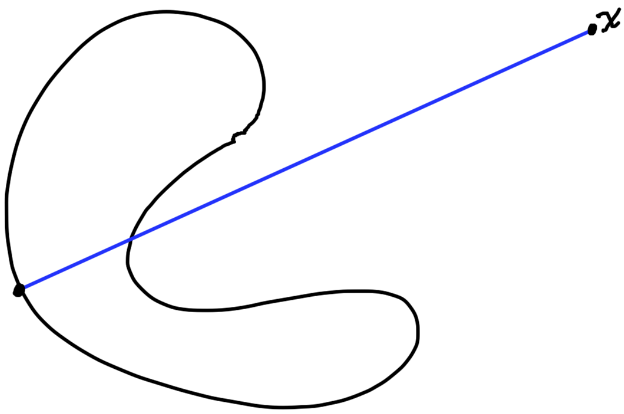

Figure 1.4: Max length function

Example:

$$F(\mathbf{x}) = \sup_{\mathbf{y} \in C} \|\mathbf{x} - \mathbf{y}\| \tag{1.30}$$

Here $C \subseteq \mathbb{R}^n$ is not necessarily convex. We are using sup here because the set C may be open. This function is the length of the line from $\mathbf{x}$ to the point in C that is furthest from $\mathbf{x}$.

- $\mathbf{x} - \mathbf{y}$ is linear in $\mathbf{x}$
- $g_{\mathbf{y}}(\mathbf{x}) = \|\mathbf{x} - \mathbf{y}\|$ is convex in $\mathbf{x}$ since norms are convex functions.
- $F(\mathbf{x}) = \sup_{\mathbf{y} \in C} \|\mathbf{x} - \mathbf{y}\|$. Each $\mathbf{y}$ index is a convex function. Taking max of those.

Example:

$$F(\mathbf{x}) = \inf_{\mathbf{y} \in C} \|\mathbf{x} - \mathbf{y}\| . \tag{1.31}$$

Min and max of two convex functions are plotted in fig. 1.5.
The max is observed to be convex, whereas the min is not necessarily so.

$$\begin{aligned} F(\mathbf{z}) &= F(\theta \mathbf{x} + (1 - \theta)\mathbf{y}) \\ &\geq \theta F(\mathbf{x}) + (1 - \theta)F(\mathbf{y}). \end{aligned} \tag{1.32}$$

This is not necessarily convex for all sets $C \subseteq \mathbb{R}^n$, because the inf of a bunch of convex function is not necessarily convex. However, if C is convex, then $F(\mathbf{x})$ is convex.

Consequences of convexity for differentiable functions

- Think about unconstrained functions dom $F = \mathbb{R}^n$.

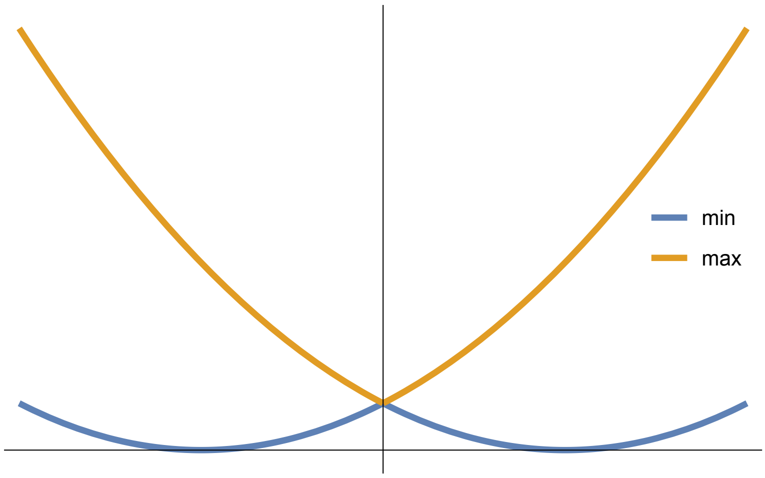

Figure 1.5: Min and max

- By first order condition F is convex iff the domain is convex and

$$F(\mathbf{x}) \geq (\nabla F(\mathbf{x}))^{\mathrm{T}} (\mathbf{y} - \mathbf{x}) \, \forall \mathbf{x}, \mathbf{y} \in \operatorname{dom} F. \tag{1.33}$$

If F is convex and one can find an $\mathbf{x}^* \in \operatorname{dom} F$ such that

$$\nabla F(\mathbf{x}^*) = 0, \tag{1.34}$$

then

$$F(\mathbf{y}) \geq F(\mathbf{x}^*) \, \forall \mathbf{y} \in \operatorname{dom} F. \tag{1.35}$$

If you can find the point where the gradient is zero (which can't always be found), then $\mathbf{x}^*$ is a global minimum of F.

Conversely, if $\mathbf{x}^*$ is a global minimizer of F, then $\nabla F(\mathbf{x}^*) = 0$ must hold. If that were not the case, then you would be able to find a direction to move downhill, contracting the optimality of $\mathbf{x}^*$.

Local vs Global optimum

Definition 1.1: Local optimum.

$\mathbf{x}^*$ is a local optimum of F if $\exists \epsilon > 0$ such that $\forall \mathbf{x}$, $\|\mathbf{x} - \mathbf{x}^*\| < \epsilon$, we have

$$F(\mathbf{x}^*) \leq F(\mathbf{x})$$

Theorem 1.1

Suppose F is twice continuously differentiable (not necessarily convex)

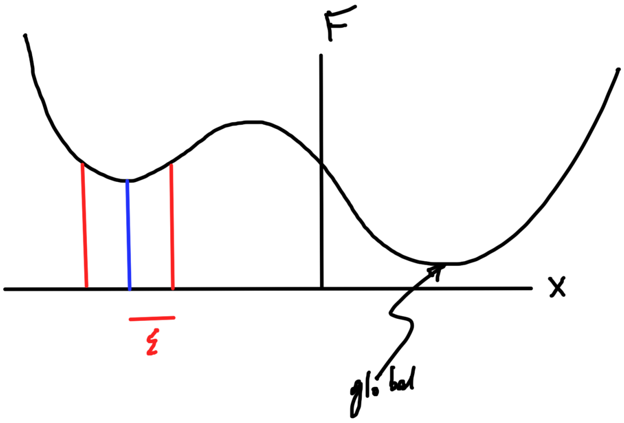

Figure 1.6: Global and local minimums

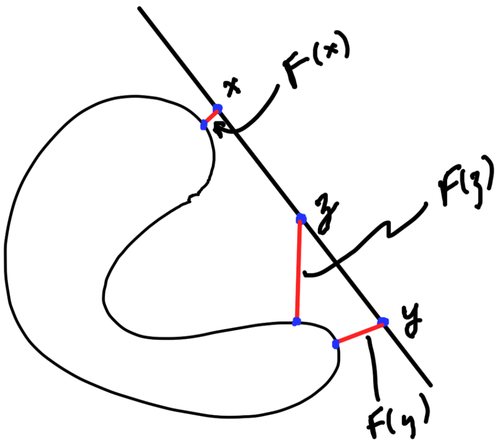

Figure 1.7: CAPTION: l7MinFuncFig5

- If $\mathbf{x}^*$ is a local optimum then

$$\begin{aligned}\boldsymbol{\nabla} F(\mathbf{x}^*) = 0 \\ \boldsymbol{\nabla}^2 F(\mathbf{x}^*) \geq 0\end{aligned}$$

- If

$$\begin{aligned}\boldsymbol{\nabla} F(\mathbf{x}^*) = 0 \\ \boldsymbol{\nabla}^2 F(\mathbf{x}^*) \geq 0\end{aligned},$$

  then $\mathbf{x}^*$ is a local optimum.

Proof:

- Let $\mathbf{x}^*$ be a local optimum. Pick any $\mathbf{v} \in \mathbb{R}^n$.

$$\lim_{t \to 0} \frac{F(\mathbf{x}^* + t\mathbf{v}) - F(\mathbf{x}^*)}{t} = \left(\boldsymbol{\nabla} F(\mathbf{x}^*)\right)^{\mathrm{T}} \mathbf{v} \geq 0. \tag{1.36}$$

Here the fraction is ≥ 0 since $\mathbf{x}^*$ is a local optimum.
Since the choice of $\mathbf{v}$ is arbitrary, the only case that you can ensure that $\geq 0, \forall \mathbf{v}$ is

$$\boldsymbol{\nabla} F = 0, \tag{1.37}$$

(or else could pick $\mathbf{v} = -\boldsymbol{\nabla} F(\mathbf{x}^*)$.
This means that $\boldsymbol{\nabla} F(\mathbf{x}^*) = 0$ if $\mathbf{x}^*$ is a local optimum.
Consider the 2nd order derivative

$$\begin{aligned}\lim_{t \to 0} \frac{F(\mathbf{x}^* + t\mathbf{v}) - F(\mathbf{x}^*)}{t^2} &= \lim_{t \to 0} \frac{1}{t^2} \left(F(\mathbf{x}^*) + t \left(\boldsymbol{\nabla} F(\mathbf{x}^*)\right)^{\mathrm{T}} \mathbf{v} + \frac{1}{2} t^2 \mathbf{v}^{\mathrm{T}} \boldsymbol{\nabla}^2 F(\mathbf{x}^*) \mathbf{v} + O(t^3) - F(\mathbf{x}^*) \right) \\ &= \frac{1}{2} \mathbf{v}^{\mathrm{T}} \boldsymbol{\nabla}^2 F(\mathbf{x}^*) \mathbf{v} \\ &\geq 0.\end{aligned} \tag{1.38}$$

Here the $\geq$ condition also comes from the fraction, based on the optimiality of $\mathbf{x}^*$. This is true for all choice of $\mathbf{v}$, thus $\boldsymbol{\nabla}^2 F(\mathbf{x}^*)$.

Bibliography

[1] Stephen Boyd and Lieven Vandenberghe. *Convex optimization*. Cambridge university press, 2004. 1