Peeter Joot
peeterjoot@protonmail.com

ECE1505H Convex Optimization. Lecture 8: Local vs. Global, and composition of functions. Taught by Prof. Stark Draper

1.1 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].

1.2 Today

- Finish local vs global.
- Compositions of functions.
- Introduction to convex optimization problems.

1.3 Continuing proof:

We want to prove that if

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

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

Proof:

Again, using Taylor approximation

$$F(\mathbf{x}^* + \mathbf{v}) = F(\mathbf{x}^*) + (\nabla F(\mathbf{x}^*))^\mathrm{T} \mathbf{v} + \frac{1}{2} \mathbf{v}^\mathrm{T} \nabla^2 F(\mathbf{x}^*) \mathbf{v} + o(\|\mathbf{v}\|^2) \tag{1.1}$$

The linear term is zero by assumption, whereas the Hessian term is given as > 0. Any direction that you move in, if your move is small enough, this is going uphill at a local optimum.

1.4 Summarize:

For twice continuously differentiable functions, at a local optimum $\mathbf{x}^*$, then

$$\begin{aligned} \nabla F(\mathbf{x}^*) &= 0 \\ \nabla^2 F(\mathbf{x}^*) &\geq 0 \end{aligned} \tag{1.2}$$

If, in addition, F is convex, then $\nabla F(\mathbf{x}^*) = 0$ implies that $\mathbf{x}^*$ is a global optimum. i.e. for (unconstrained) convex functions, local and global optimums are equivalent.

- It is possible that a convex function does not have a global optimum. Examples are $F(x) = e^x$ (fig. 1.1) , which has an inf, but no lowest point.

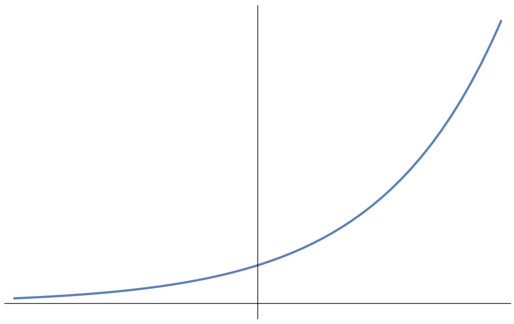

Figure 1.1: Exponential has no global optimum.

- Our discussion has been for unconstrained functions. For constrained problems (next topic) is not not necessarily true that $\nabla F(\mathbf{x}) = 0$ implies that $\mathbf{x}$ is a global optimum, even for F convex.

  As an example of a constrained problem consider

$$\begin{aligned} &\min 2x^2 + y^2 \\ &x \geq 3 \\ &y \geq 5. \end{aligned} \tag{1.3}$$

  The level sets of this objective function are plotted in fig. 1.2. The optimal point is at $\mathbf{x}^* = (3, 5)$, where $\nabla F \neq 0$.

1.5 Projection

Given $\mathbf{x} \in \mathbb{R}^n, \mathbf{y} \in \mathbb{R}^p$, if $h(\mathbf{x}, \mathbf{y})$ is convex in $\mathbf{x}, \mathbf{y}$, then

$$F(\mathbf{x}_0) = \inf_{\mathbf{y}} h(\mathbf{x}_0, \mathbf{y}) \tag{1.4}$$

is convex in $\mathbf{x}$, as sketched in fig. 1.3.

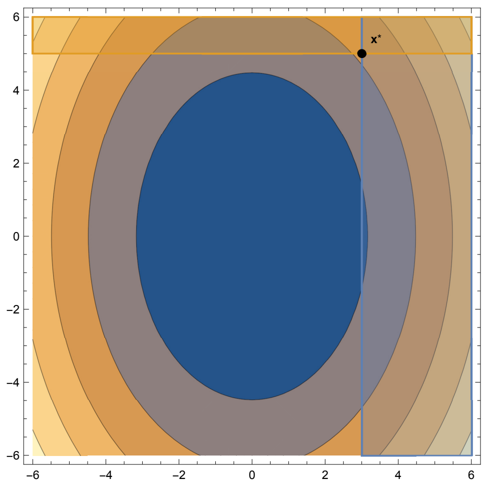

Figure 1.2: Constrained problem with optimum not at the zero gradient point.

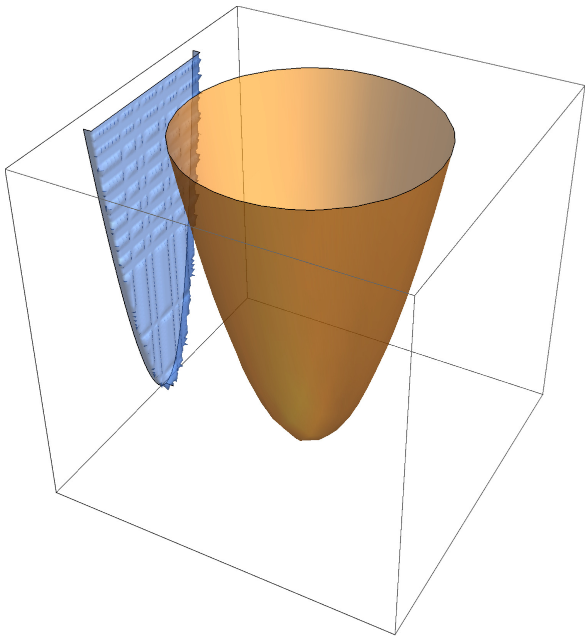

Figure 1.3: Epigraph of h is a filled bowl.

The intuition here is that shining light on the (filled) "bowl". That is, the image of epi h on the $\mathbf{y} = 0$ screen which we will show is a convex set.

Proof:

Since h is convex in $\begin{bmatrix}\mathbf{x}\\ \mathbf{y}\end{bmatrix} \in \operatorname{dom} h$, then

$$\operatorname{epi} h = \left\{(\mathbf{x}, \mathbf{y}, t) | t \geq h(\mathbf{x}, \mathbf{y}), \begin{bmatrix}\mathbf{x}\\ \mathbf{y}\end{bmatrix} \in \operatorname{dom} h\right\}, \tag{1.5}$$

is a convex set.

We also have to show that the domain of F is a convex set. To show this note that

$$\begin{aligned}\operatorname{dom} F &= \left\{\mathbf{x} | \exists \mathbf{y} s.t. \begin{bmatrix}\mathbf{x}\\ \mathbf{y}\end{bmatrix} \in \operatorname{dom} h\right\}\\ &= \left\{\begin{bmatrix}I_{n\times n} & 0_{n\times p}\end{bmatrix}\begin{bmatrix}\mathbf{x}\\ \mathbf{y}\end{bmatrix} | \begin{bmatrix}\mathbf{x}\\ \mathbf{y}\end{bmatrix} \in \operatorname{dom} h\right\}.\end{aligned} \tag{1.6}$$

This is an affine map of a convex set. Therefore dom F is a convex set.

$$\begin{aligned}\operatorname{epi} F &= \left\{\begin{bmatrix}\mathbf{x}\\ \mathbf{y}\end{bmatrix} | t \geq \inf h(\mathbf{x}, \mathbf{y}), \mathbf{x} \in \operatorname{dom} F, \mathbf{y} : \begin{bmatrix}\mathbf{x}\\ \mathbf{y}\end{bmatrix} \in \operatorname{dom} h\right\}\\ &= \left\{\begin{bmatrix}I & 0 & 0\\ 0 & 0 & 1\end{bmatrix}\begin{bmatrix}\mathbf{x}\\ \mathbf{y}\\ t\end{bmatrix} | t \geq h(\mathbf{x}, \mathbf{y}), \begin{bmatrix}\mathbf{x}\\ \mathbf{y}\end{bmatrix} \in \operatorname{dom} h\right\}.\end{aligned} \tag{1.7}$$

Example: The function

$$F(\mathbf{x}) = \inf_{\mathbf{y}\in C} \|\mathbf{x} - \mathbf{y}\|, \tag{1.8}$$

over $\mathbf{x} \in \mathbb{R}^n, \mathbf{y} \in C$, ,is convex if C is a convex set. Reason:

- $\mathbf{x} - \mathbf{y}$ is linear in $(\mathbf{x}, \mathbf{y})$.
- $\|\mathbf{x} - \mathbf{y}\|$ is a convex function if the domain is a convex set
- The domain is $\mathbb{R}^n \times C$. This will be a convex set if C is.
- $h(\mathbf{x}, \mathbf{y}) = \|\mathbf{x} - \mathbf{y}\|$ is a convex function if dom h is a convex set. By setting dom $h = \mathbb{R}^n \times C$, if C is convex, dom h is a convex set.
- $F()$

1.6 Composition of functions

Consider

$$
\begin{aligned}
F(\mathbf{x}) &= h(g(\mathbf{x})) \\
\operatorname{dom} F &= \{\mathbf{x} \in \operatorname{dom} g | g(\mathbf{x}) \in \operatorname{dom} h\} \\
F &: \mathbb{R}^n \to \mathbb{R} \\
g &: \mathbb{R}^n \to \mathbb{R} \\
h &: \mathbb{R} \to \mathbb{R}.
\end{aligned} \tag{1.9}
$$

Cases:

(a) g is convex, h is convex and non-decreasing.

(b) g is convex, h is convex and non-increasing.

Show for 1D case ($n = 1$). Get to $n > 1$ by applying to all lines.

(a)

$$
\begin{aligned}
F'(x) &= h'(g(x))g'(x) \\
F''(x) &= h''(g(x))g'(x)g'(x) + h'(g(x))g''(x) \\
&= h''(g(x))(g'(x))^2 + h'(g(x))g''(x) \\
&= (\geq 0) \cdot (\geq 0)^2 + (\geq 0) \cdot (\geq 0)\,,
\end{aligned} \tag{1.10}
$$

since h is respectively convex, and non-decreasing.

(b)

$$
F'(x) = (\geq 0) \cdot (\geq 0)^2 + (\leq 0) \cdot (\leq 0)\,, \tag{1.11}
$$

since h is respectively convex, and non-increasing, and g is concave.

1.7 Extending to multiple dimensions

$$
\begin{aligned}
F(\mathbf{x}) &= h(g(\mathbf{x})) = h(g_1(\mathbf{x}), g_2(\mathbf{x}), \cdots g_k(\mathbf{x})) \\
g &: \mathbb{R}^n \to \mathbb{R} \\
h &: \mathbb{R}^k \to \mathbb{R}.
\end{aligned} \tag{1.12}
$$

is convex if g_i is convex for each $i \in [1, k]$ and h is convex and non-decreasing in each argument.
Proof:
again assume $n = 1$, without loss of generality,

$$
\begin{aligned}
g &: \mathbb{R} \to \mathbb{R}^k \\
h &: \mathbb{R}^k \to \mathbb{R}
\end{aligned} \tag{1.13}
$$

$$
F''(\mathbf{x}) = \begin{bmatrix} g_1(\mathbf{x}) & g_2(\mathbf{x}) & \cdots & g_k(\mathbf{x}) \end{bmatrix} \boldsymbol{\nabla}^2 h(g(\mathbf{x})) \begin{bmatrix} g_1'(\mathbf{x}) \\ g_2'(\mathbf{x}) \\ \vdots \\ g_k'(\mathbf{x}) \end{bmatrix} + \left(\boldsymbol{\nabla} h(g(x))\right)^{\mathrm{T}} \begin{bmatrix} g_1''(\mathbf{x}) \\ g_2''(\mathbf{x}) \\ \vdots \\ g_k''(\mathbf{x}) \end{bmatrix} \tag{1.14}
$$

The Hessian is PSD.

Example:

$$F(x) = \exp(g(x)) \\ = h(g(x)), \tag{1.15}$$

where g is convex is convex, and $h(y) = e^y$. This implies that F is a convex function.

Example:

$$F(x) = \frac{1}{g(x)}, \tag{1.16}$$

is convex if $g(x)$ is concave and positive. The most simple such example of such a function is $h(x) = 1/x$, $\operatorname{dom} h = \mathbb{R}_{++}$, which is plotted in fig. 1.4.

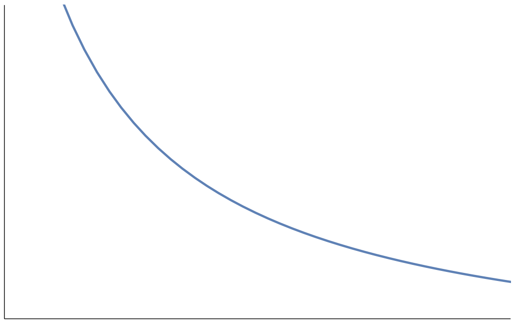

Figure 1.4: Inverse function is convex over positive domain.

Example:

$$F(x) = -\sum_{i=1}^{n} \log(-F_i(x)) \tag{1.17}$$

is convex on $\{x | F_i(x) < 0 \forall i\}$ if all F_i are convex.

- Due to $\operatorname{dom} F$, $-F_i(x) > 0 \, \forall x \in \operatorname{dom} F$
- $\log(x)$ concave on $\mathbb{R}_{++}$ so $-\log$ convex also non-increasing (fig. 1.5).

$$F(x) = \sum h_i(x) \tag{1.18}$$

but

$$h_i(x) = -\log(-F_i(x)), \tag{1.19}$$

which is a convex and non-increasing function ($-\log$), of a convex function $-F_i(x)$. Each h_i is convex, so this is a sum of convex functions, and is therefore convex.

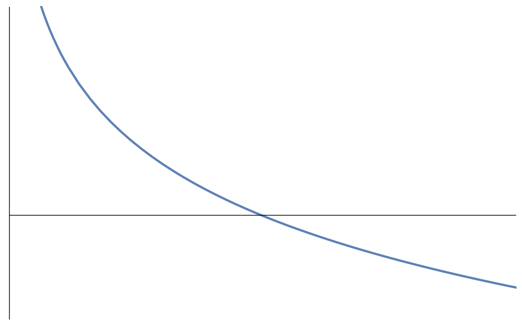

Figure 1.5: Negative logarithm convex over positive domain.

Example: Over $\operatorname{dom} F = S^n_{++}$

$$F(X) = \log\det X^{-1} \tag{1.20}$$

To show that this is convex, check all lines in domain. A line in S^n_{++} is a 1D family of matrices

$$\tilde{F}(t) = \log\det((X_0 + tH)^{-1}), \tag{1.21}$$

where $X_0 \in S^n_{++}, t \in \mathbb{R}, H \in S^n$.
F9
For t small enough,

$$X_0 + tH \in S^n_{++} \tag{1.22}$$

$$\begin{aligned}
\tilde{F}(t) &= \log\det((X_0 + tH)^{-1}) \\
&= \log\det\left(X_0^{-1/2}\left(I + tX_0^{-1/2}HX_0^{-1/2}\right)^{-1}X_0^{-1/2}\right) \\
&= \log\det\left(X_0^{-1}\left(I + tX_0^{-1/2}HX_0^{-1/2}\right)^{-1}\right) \\
&= \log\det X_0^{-1} + \log\det\left(I + tX_0^{-1/2}HX_0^{-1/2}\right)^{-1} \\
&= \log\det X_0^{-1} - \log\det\left(I + tX_0^{-1/2}HX_0^{-1/2}\right) \\
&= \log\det X_0^{-1} - \log\det\left(I + tM\right).
\end{aligned} \tag{1.23}$$

If λ_i are eigenvalues of M, then $1 + t\lambda_i$ are eigenvalues of $I + tM$. i.e.:

$$\begin{aligned}
(I + tM)\mathbf{v} &= I\mathbf{v} + t\lambda_i\mathbf{v} \\
&= (1 + t\lambda_i)\mathbf{v}.
\end{aligned} \tag{1.24}$$

This gives

$$\begin{aligned}\tilde{F}(t) &= \log\det X_0^{-1} - \log\prod_{i=1}^{n}(1+t\lambda_i) \\ &= \log\det X_0^{-1} - \sum_{i=1}^{n}\log(1+t\lambda_i)\end{aligned} \tag{1.25}$$

- $1 + t\lambda_i$ is linear in t.
- $-\log$ is convex in its argument.
- sum of convex function is convex.

Example:

$$F(X) = \lambda_{\max}(X), \tag{1.26}$$

is convex on dom $F \in S^n$
(a)

$$\lambda_{\max}(X) = \sup_{\|\mathbf{v}\|_2 \leq 1} \mathbf{v}^{\mathrm{T}} X \mathbf{v}, \tag{1.27}$$

$$\begin{bmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n \end{bmatrix} \tag{1.28}$$

Recall that a decomposition

$$\begin{aligned} X &= Q\Lambda Q^{\mathrm{T}} \\ Q^{\mathrm{T}}Q = QQ^{\mathrm{T}} &= I \end{aligned} \tag{1.29}$$

can be used for any $X \in S^n$.
(b)
Note that $\mathbf{v}^{\mathrm{T}} X \mathbf{v}$ is linear in X. This is a max of a number of linear (and convex) functions, so it is convex.
Last example:
(non-symmetric matrices)

$$F(X) = \sigma_{\max}(X), \tag{1.30}$$

is convex on dom $F = \mathbb{R}^{m\times n}$. Here

$$\sigma_{\max}(X) = \sup_{\|\mathbf{v}\|_2 = 1} \|X\mathbf{v}\|_2 \tag{1.31}$$

This is called an operator norm of X. Using the SVD

$$
\begin{aligned}
X &= U\Sigma V^{\mathrm{T}} \\
U &= \mathbb{R}^{m\times r} \\
\Sigma &\in \text{diag} \in \mathbb{R}r \times r \\
V^T &\in \mathbb{R}^{r\times n}.
\end{aligned}
\tag{1.32}
$$

Have

$$
\begin{aligned}
\|X\mathbf{v}\|_2^2 &= \left\|U\Sigma V^{\mathrm{T}}\mathbf{v}\right\|_2^2 \\
&= \mathbf{v}^{\mathrm{T}} V\Sigma U^{\mathrm{T}} U\Sigma V^{\mathrm{T}}\mathbf{v} \\
&= \mathbf{v}^{\mathrm{T}} V\Sigma\Sigma V^{\mathrm{T}}\mathbf{v} \\
&= \mathbf{v}^{\mathrm{T}} V\Sigma^2 V^{\mathrm{T}}\mathbf{v} \\
&= \tilde{\mathbf{v}}^{\mathrm{T}}\Sigma^2\tilde{\mathbf{v}},
\end{aligned}
\tag{1.33}
$$

where $\tilde{\mathbf{v}} = \mathbf{v}^{\mathrm{T}} V$, so

$$
\begin{aligned}
\|X\mathbf{v}\|_2^2 &= \sum_{i=1}^{r} \sigma_i^2 \|\tilde{\mathbf{v}}\| \\
&\leq \sigma_{\max}^2 \|\tilde{\mathbf{v}}\|^2,
\end{aligned}
\tag{1.34}
$$

or

$$
\begin{aligned}
\|X\mathbf{v}\|_2 &\leq \sqrt{\sigma_{\max}^2} \|\tilde{\mathbf{v}}\| \\
&\leq \sigma_{\max}.
\end{aligned}
\tag{1.35}
$$

Set $\mathbf{v}$ to the right singular value of X to get equality.

Bibliography

[1] Stephen Boyd and Lieven Vandenberghe. *Convex optimization*. Cambridge university press, 2004. 1.1