Peeter Joot
peeterjoot@protonmail.com

ECE1505H Convex Optimization. Lecture 6: First and second order conditions. 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, covering ch. 1 [1] content.

Today

- First and second order conditions for convexity of differentiable functions.
- Consequences of convexity: local and global optimality.
- Properties.

Quasi-convex F_1 and F_2 convex implies $\max(F_1, F_2)$ convex.

Figure 1.1: Min and Max

Note that $\min(F_1, F_2)$ is NOT convex.

If $F : \mathbb{R}^n \rightarrow \mathbb{R}$ is convex, then $F(\mathbf{x}_0 + t\mathbf{v})$ is convex in $t \, \forall t \in \mathbb{R}, \mathbf{x}_0 \in \mathbb{R}^n, \mathbf{v} \in \mathbb{R}^n$, provided $\mathbf{x}_0 + t\mathbf{v} \in \operatorname{dom} F$.

Idea: Restrict to a line (line segment) in dom F. Take a cross section or slice through F alone the line. If the result is a 1D convex function for all slices, then F is convex.

This is nice since it allows for checking for convexity, and is also nice numerically. Attempting to test a given data set for non-convexity with some random lines can help disprove convexity. However, to show that F is convex it is required to test all possible slices (which isn't possible numerically, but is in some circumstances possible analytically).

Differentiable (convex) functions

Definition 1.1: First order condition

If

$$F : \mathbb{R}^n \to \mathbb{R}$$

is differentiable, then F is convex iff dom F is a convex set and $\forall \mathbf{x}, \mathbf{x}_0 \in \operatorname{dom} F$

$$F(\mathbf{x}) \geq F(\mathbf{x}_0) + (\boldsymbol{\nabla} F(\mathbf{x}_0))^{\mathrm{T}} (\mathbf{x} - \mathbf{x}_0).$$

This is the first order Taylor expansion. If $n = 1$, this is $F(x) \geq F(x_0) + F'(x_0)(x - x_0)$.

The first order condition says a convex function always lies above its first order approximation, as sketched in fig. 1.2.

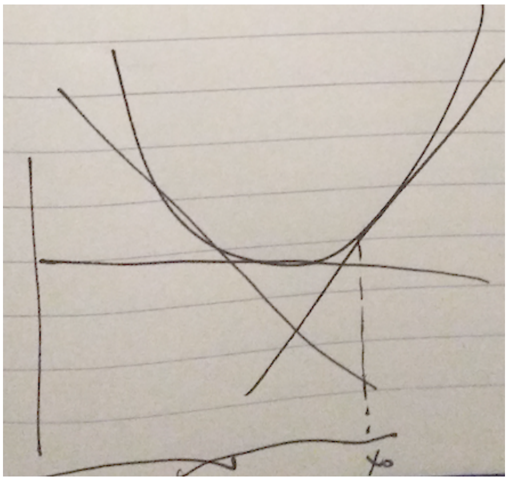

Figure 1.2: First order approximation lies below convex function

When differentiable, the supporting plane is the tangent plane.

Definition 1.2: Second order condition

If $F : \mathbb{R}^n \to \mathbb{R}$ is twice differentiable, then F is convex iff dom F is a convex set and $\boldsymbol{\nabla}^2 F(\mathbf{x}) \geq$

$0\,\forall \mathbf{x} \in \operatorname{dom} F$.

The Hessian is always symmetric, but is not necessarily positive. Recall that the Hessian is the matrix of the second order partials $(\boldsymbol{\nabla} F)_{ij} = \partial^2 F/(\partial x_i \partial x_j)$.

The scalar case is $F''(x) \geq 0\,\forall x \in \operatorname{dom} F$.

An implication is that if F is convex, then $F(x) \geq F(x_0) + F'(x_0)(x - x_0)\,\forall x, x_0 \in \operatorname{dom} F$

Since F is convex, $\operatorname{dom} F$ is convex.

Consider any 2 points $x, y \in \operatorname{dom} F$, and $\theta \in [0, 1]$. Define

$$z = (1-\theta)x + \theta y \in \operatorname{dom} F, \tag{1.1}$$

then since $\operatorname{dom} F$ is convex

$$\begin{aligned} F(z) &= F((1-\theta)x + \theta y) \\ &\leq (1-\theta)F(x) + \theta F(y) \end{aligned} \tag{1.2}$$

Reordering

$$\theta F(x) \geq \theta F(x) + F(z) - F(x), \tag{1.3}$$

or

$$F(y) \geq F(x) + \frac{F(x + \theta(y-x)) - F(x)}{\theta}, \tag{1.4}$$

which is, in the limit,

$$F(y) \geq F(x) + F'(x)(y - x) \qquad \square \tag{1.5}$$

To prove the other direction, showing that

$$F(x) \geq F(x_0) + F'(x_0)(x - x_0), \tag{1.6}$$

implies that F is convex. Take any $x, y \in \operatorname{dom} F$ and any $\theta \in [0, 1]$. Define

$$z = \theta x + (1-\theta)y, \tag{1.7}$$

which is in $\operatorname{dom} F$ by assumption. We want to show that

$$F(z) \leq \theta F(x) + (1-\theta)F(y). \tag{1.8}$$

By assumption

(i) $F(x) \geq F(z) + F'(z)(x - z)$

(ii) $F(y) \geq F(z) + F'(z)(y - z)$

Compute

$$\begin{aligned} \theta F(x) + (1-\theta)F(y) &\geq \theta\left(F(z) + F'(z)(x-z)\right) + (1-\theta)\left(F(z) + F'(z)(y-z)\right) \\ &= F(z) + F'(z)\left(\theta(x-z) + (1-\theta)(y-z)\right) \\ &= F(z) + F'(z)\left(\theta x + (1-\theta)y - \theta z - (1-\theta)z\right) \\ &= F(z) + F'(z)\left(\theta x + (1-\theta)y - z\right) \\ &= F(z) + F'(z)\,(z - z) \\ &= F(z). \end{aligned} \tag{1.9}$$

Proof of the 2nd order case for $n = 1$ Want to prove that if

$$F : \mathbb{R} \to \mathbb{R} \tag{1.10}$$

is a convex function, then $F''(x) \geq 0 \, \forall x \in \operatorname{dom} F$.
By the first order conditions $\forall x \neq y \in \operatorname{dom} F$

$$F(y) \geq F(x) + F'(x)(y - x) F(x) \quad \geq F(y) + F'(y)(x - y) \tag{1.11}$$

Can combine and get

$$F'(x)(y - x) \leq F(y) - F(x) \leq F'(y)(y - x) \tag{1.12}$$

Subtract the two derivative terms for

$$\frac{(F'(y) - F'(x))(y - x)}{(y - x)^2} \geq 0, \tag{1.13}$$

or

$$\frac{F'(y) - F'(x)}{y - x} \geq 0. \tag{1.14}$$

In the limit as $y \to x$, this is

$$\boxed{F''(x) \geq 0 \, \forall x \in \operatorname{dom} F.} \tag{1.15}$$

Now prove the reverse condition:
If $F''(x) \geq 0 \, \forall x \in \operatorname{dom} F \subseteq \mathbb{R}$, implies that $F : \mathbb{R} \to \mathbb{R}$ is convex.
Note that if $F''(x) \geq 0$, then $F'(x)$ is non-decreasing in x.
i.e. If $x < y$, where $x, y \in \operatorname{dom} F$, then

$$F'(x) \leq F'(y). \tag{1.16}$$

Consider any $x, y \in \operatorname{dom} F$ such that $x < y$, where

$$\begin{aligned} F(y) - F(x) &= \int_x^y F'(t) dt \\ &\geq F'(x) \int_x^y 1 dt \\ &= F'(x)(y - x). \end{aligned} \tag{1.17}$$

This tells us that

$$F(y) \geq F(x) + F'(x)(y - x), \tag{1.18}$$

which is the first order condition. Similarly consider any $x, y \in \operatorname{dom} F$ such that $x < y$, where

$$\begin{aligned} F(y) - F(x) &= \int_x^y F'(t) dt \\ &\leq F'(y) \int_x^y 1 dt \\ &= F'(y)(y - x). \end{aligned} \tag{1.19}$$

This tells us that

$$F(x) \geq F(y) + F'(y)(x - y). \tag{1.20}$$

Vector proof: F is convex iff $F(\mathbf{x} + t\mathbf{v})$ is convex $\forall \mathbf{x}, \mathbf{v} \in \mathbb{R}^n, t \in \mathbb{R}$, keeping $\mathbf{x} + t\mathbf{v} \in \operatorname{dom} F$.

Let

$$h(t; \mathbf{x}, \mathbf{v}) = F(\mathbf{x} + t\mathbf{v}) \tag{1.21}$$

then $h(t)$ satisfies scalar first and second order conditions for all $\mathbf{x}, \mathbf{v}$.

$$\begin{aligned} h(t) &= F(\mathbf{x} + t\mathbf{v}) \\ &= F(g(t)), \end{aligned} \tag{1.22}$$

where $g(t) = \mathbf{x} + t\mathbf{v}$, where

$$\begin{aligned} F &: \mathbb{R}^n \rightarrow \mathbb{R} \\ g &: \mathbb{R} \rightarrow \mathbb{R}^n. \end{aligned} \tag{1.23}$$

This is expressing $h(t)$ as a composition of two functions. By the first order condition for scalar functions we know that

$$h(t) \geq h(0) + h'(0)t. \tag{1.24}$$

Note that

$$h(0) = \left. F(\mathbf{x} + t\mathbf{v}) \right|_{t=0} = F(\mathbf{x}). \tag{1.25}$$

Let's figure out what $h'(0)$ is. Recall hat for any $\tilde{F} : \mathbb{R}^n \rightarrow \mathbb{R}^m$

$$D\tilde{F} \in \mathbb{R}^{m \times n}, \tag{1.26}$$

and

$$D\tilde{F}(\mathbf{x})_{ij} = \frac{\partial \tilde{F}_i(\mathbf{x})}{\partial x_j} \tag{1.27}$$

This is one function per row, for $i \in [1, m], j \in [1, n]$. This gives

$$\begin{aligned} \frac{d}{dt} F(\mathbf{x} + \mathbf{v}t) &= \frac{d}{dt} F(g(t)) \\ &= \frac{d}{dt} h(t) \\ &= Dh(t) \\ &= DF(g(t)) \cdot Dg(t) \end{aligned} \tag{1.28}$$

The first matrix is in $\mathbb{R}^{1 \times n}$ whereas the second is in $\mathbb{R}^{n \times 1}$, since $F : \mathbb{R}^n \rightarrow \mathbb{R}$ and $g : \mathbb{R} \rightarrow \mathbb{R}^n$. This gives

$$\frac{d}{dt} F(\mathbf{x} + \mathbf{v}t) = \left. DF(\tilde{\mathbf{x}}) \right|_{\tilde{\mathbf{x}} = g(t)} \cdot Dg(t). \tag{1.29}$$

That first matrix is

$$
\begin{aligned}
DF(\tilde{\mathbf{x}})\big|_{\tilde{\mathbf{x}}=g(t)} &= \left(\begin{bmatrix} \frac{\partial F(\tilde{\mathbf{x}})}{\partial \tilde{x}_1} & \frac{\partial F(\tilde{\mathbf{x}})}{\partial \tilde{x}_2} & \cdots & \frac{\partial F(\tilde{\mathbf{x}})}{\partial \tilde{x}_n} \end{bmatrix}\right)\Big|_{\tilde{\mathbf{x}}=g(t)=\mathbf{x}+t\mathbf{v}} \\
&= (\boldsymbol{\nabla} F(\tilde{\mathbf{x}}))^{\mathrm{T}}\Big|_{\tilde{\mathbf{x}}=g(t)} \\
&= \left(\boldsymbol{\nabla} F(g(t))\right)^{\mathrm{T}}.
\end{aligned} \tag{1.30}
$$

The second Jacobian is

$$
Dg(t) = D \begin{bmatrix} g_1(t) \\ g_2(t) \\ \vdots \\ g_n(t) \end{bmatrix} = D \begin{bmatrix} x_1 + tv_1 \\ x_2 + tv_2 \\ \vdots \\ x_n + tv_n \end{bmatrix} = \begin{bmatrix} v_1 \\ v_1 \\ \vdots \\ v_n \end{bmatrix} = \mathbf{v}. \tag{1.31}
$$

so

$$
\begin{aligned}
h'(t) &= Dh(t) \\
&= \left(\boldsymbol{\nabla} F(g(t))\right)^{\mathrm{T}} \mathbf{v},
\end{aligned} \tag{1.32}
$$

and

$$
\begin{aligned}
h'(0) &= \left(\boldsymbol{\nabla} F(g(0))\right)^{\mathrm{T}} \mathbf{v} \\
&= (\boldsymbol{\nabla} F(\mathbf{x}))^{\mathrm{T}} \mathbf{v}.
\end{aligned} \tag{1.33}
$$

Finally

$$
\begin{aligned}
F(\mathbf{x} + t\mathbf{v}) &\ge h(0) + h'(0)t \\
&= F(\mathbf{x}) + (\boldsymbol{\nabla} F(\mathbf{x}))^{\mathrm{T}} (t\mathbf{v}) \\
&= F(\mathbf{x}) + \langle \boldsymbol{\nabla} F(\mathbf{x}), t\mathbf{v} \rangle.
\end{aligned} \tag{1.34}
$$

Which is true for all $\mathbf{x}, \mathbf{x} + t\mathbf{v} \in \operatorname{dom} F$. Note that the quantity $t\mathbf{v}$ is a shift.

Epigraph Recall that if $(\mathbf{x}, t) \in \operatorname{epi} F$ then $t \ge F(\mathbf{x})$.

$$
\begin{aligned}
t &\ge F(\mathbf{x}) \\
&\ge F(\mathbf{x}_0) + (\boldsymbol{\nabla} F(\mathbf{x}_0))^{\mathrm{T}} (\mathbf{x} - \mathbf{x}_0),
\end{aligned} \tag{1.35}
$$

or

$$
0 \ge -(t - F(\mathbf{x}_0)) + (\boldsymbol{\nabla} F(\mathbf{x}_0))^{\mathrm{T}} (\mathbf{x} - \mathbf{x}_0), \tag{1.36}
$$

In block matrix form

$$
0 \ge \begin{bmatrix} (\boldsymbol{\nabla} F(\mathbf{x}_0))^{\mathrm{T}} & -1 \end{bmatrix} \begin{bmatrix} \mathbf{x} - \mathbf{x}_0 \\ t - F(\mathbf{x}_0) \end{bmatrix} \tag{1.37}
$$

With $\mathbf{w} = \begin{bmatrix} (\boldsymbol{\nabla} F(\mathbf{x}_0))^{\mathrm{T}} & -1 \end{bmatrix}$, the geometry of the epigraph relation to the half plane is sketched in fig. 1.3.

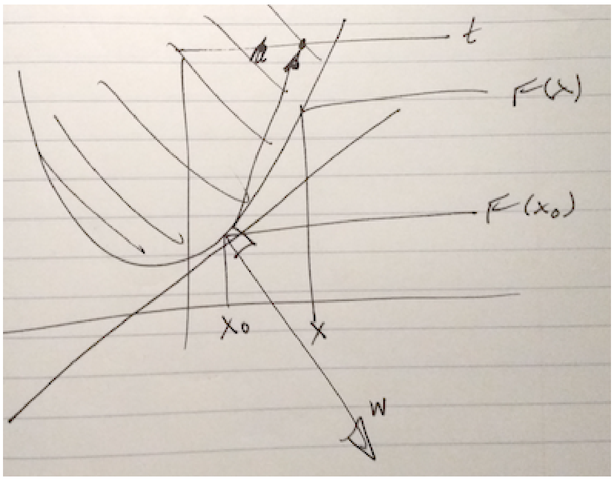

Figure 1.3: Half planes and epigraph.

Bibliography

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