C477: CW1 Revision

Ruth Misener
r.misener@imperial.ac.uk

Panos Parpas
p.parpas@imperial.ac.uk

Computational Optimisation Group
Department of Computing

December 7, 2018

Excellent Job on CW1!

Everyone seems fairly comfortable with the first half of the course; this is important since the course builds on itself. In particular:

- Well done mastering: convexity, optimality conditions, one-dimensional optimisation, and first-order optimisation methods;
- Well done justifying your work so that we can give partial credit;
- Well done on the Matlab portion of the coursework.

Outline

- **Proving Convexity / Concavity**
  - From the Definition
  - First Derivative Test (Gradient Inequality)
  - Second Derivative Test
  - Counter examples [refute convexity / concavity]
- **Reminders**
  - Vector & Matrix Manipulations
  - Feasible directions
  - Optimality conditions
- **Partial Credit on CW1**
- **Feedback for C477**

Detecting Convexity [Second Deriv Test]

Hessian: Sufficient Conditions for Strict Convexity

A function $f : S \subset \mathbb{R}^n \to \mathbb{R}$ in $\mathcal{C}^2$ is strictly convex on S if, at each $\boldsymbol{x} \in S$, the Hessian $\boldsymbol{H}(\boldsymbol{x}) \succ 0$ (positive definite)

Characterise the convexity of $f(x) = \|\boldsymbol{a}x - \boldsymbol{b}\|_2^2$, where $x \in \mathbb{R}$ is a variable, $\boldsymbol{a}$, $\boldsymbol{b} \in \mathbb{R}^n$ are given parameters, and $\boldsymbol{a} \neq \boldsymbol{0}$

$$
\begin{aligned}
f(x) &= \|\boldsymbol{a}x - \boldsymbol{b}\|_2^2 = \langle \boldsymbol{a}x - \boldsymbol{b},\, \boldsymbol{a}x - \boldsymbol{b} \rangle = \|\boldsymbol{a}\|_2^2 x^2 - 2x\langle \boldsymbol{a},\, \boldsymbol{b} \rangle + \|\boldsymbol{b}\|_2^2 \\
f'(x) &= 2\|\boldsymbol{a}\|_2^2 x - 2\langle \boldsymbol{a},\, \boldsymbol{b} \rangle \\
f''(x) &= 2\|\boldsymbol{a}\|_2^2
\end{aligned}
$$

Since $\boldsymbol{a} \neq \boldsymbol{0}$, $f''(x) > 0 \;\forall\, x \in \mathbb{R}$ and the function is strictly convex.

Sanity Check

Should it be that f is strictly convex if and only if $\boldsymbol{H}(\boldsymbol{x}) \succ 0$?

Detecting Convexity [From the Definition]

Convex Functions

A function $f : S \to \mathbb{R}$, defined on a convex set $S \subset \mathbb{R}^n$, is **convex on S** if the line segment connecting $f(\boldsymbol{x})$ and $f(\boldsymbol{y})$ at **any** two points $\boldsymbol{x}$, $\boldsymbol{y} \in S$ satisfies: $f(\alpha\, \boldsymbol{x} + (1-\alpha)\, \boldsymbol{y}) \le \alpha f(\boldsymbol{x}) + (1-\alpha) f(\boldsymbol{y}), \quad \forall \alpha \in (0,\, 1)$

Characterise the convexity of $f(x) = \|\boldsymbol{a}x - \boldsymbol{b}\|_2^2$, where $x \in \mathbb{R}$ is a variable, $\boldsymbol{a}$, $\boldsymbol{b} \in \mathbb{R}^n$ are given parameters, and $\boldsymbol{a} \neq \boldsymbol{0}$

$$
\begin{aligned}
&\alpha\|\boldsymbol{a}x - \boldsymbol{b}\|_2^2 + (1-\alpha)\|\boldsymbol{a}y - \boldsymbol{b}\|_2^2 - \|\boldsymbol{a}\left(\alpha x + (1-\alpha)y\right) - \boldsymbol{b}\|_2^2 \\
&= \alpha\|\boldsymbol{a}\|_2^2 x^2 + (1-\alpha)\|\boldsymbol{a}\|_2^2 y^2 - \|\boldsymbol{a}\|_2^2(\alpha x + (1-\alpha)y)^2 \\
&= \|\boldsymbol{a}\|_2^2 \left(\alpha(1-\alpha)x^2 - 2\alpha(1-\alpha)xy + \alpha(1-\alpha)y^2\right) \\
&= \|\boldsymbol{a}\|_2^2 \alpha(1-\alpha)(x-y)^2 \ge 0
\end{aligned}
$$

Function is defined on convex set $x \in \mathbb{R}$, the inequality is strict for all $x \neq y$ and $\alpha \in (0,\, 1)$, so the function is strictly convex.

Detecting Convexity [First Deriv Test]

Suppose that $C \subset \mathbb{R}^n$ is a convex set, and that $f : C \to \mathbb{R}$ is differentiable in $\mathbb{R}^n$. Then, f is convex on C if and only if for any $\widehat{\boldsymbol{x}} \in C$,

$$f(\boldsymbol{x}) \geq f(\widehat{\boldsymbol{x}}) + \nabla f(\widehat{\boldsymbol{x}})^\top (\boldsymbol{x} - \widehat{\boldsymbol{x}}), \quad \forall \boldsymbol{x} \in C.$$

Characterise the convexity of $f(x) = \|\boldsymbol{a}x - \boldsymbol{b}\|_2^2$, where $x \in \mathbb{R}$ is a variable, $\boldsymbol{a}$, $\boldsymbol{b} \in \mathbb{R}^n$ are given parameters, and $\boldsymbol{a} \neq \boldsymbol{0}$

$$\begin{aligned}
f(x) - f(\widehat{x}) - \nabla f(\widehat{x})^\top (x - \widehat{x}) & \\
=& \|\boldsymbol{a}x - \boldsymbol{b}\|_2^2 - \|\boldsymbol{a}\widehat{x} - \boldsymbol{b}\|_2^2 - \left(2\|\boldsymbol{a}\|_2^2 \widehat{x} - 2\langle \boldsymbol{a}, \boldsymbol{b} \rangle\right)(x - \widehat{x}) \\
=& \|\boldsymbol{a}\|_2^2 (x^2 - \widehat{x}^2) - 2(x - \widehat{x})\langle \boldsymbol{a}, \boldsymbol{b} \rangle - \left(2\|\boldsymbol{a}\|_2^2 \widehat{x} - 2\langle \boldsymbol{a}, \boldsymbol{b} \rangle\right)(x - \widehat{x}) \\
=& \|\boldsymbol{a}\|_2^2 (x^2 - \widehat{x}^2) - 2\|\boldsymbol{a}\|_2^2 \widehat{x}(x - \widehat{x}) = \|\boldsymbol{a}\|_2^2 \left(x^2 - 2x\widehat{x} + \widehat{x}^2\right) \\
=& \|\boldsymbol{a}\|_2^2 (x - \widehat{x})^2 \geq 0 \text{ (since } \boldsymbol{a} \neq \boldsymbol{0}\text{, the inequal is strict when } x \neq \widehat{x})
\end{aligned}$$

Detecting Convexity [Finding counter-examples]

Convex Functions

A function $f : S \to \mathbb{R}$, defined on a convex set $S \subset \mathbb{R}^n$, is **convex on S** if the line segment connecting $f(\boldsymbol{x})$ and $f(\boldsymbol{y})$ at **any** two points $\boldsymbol{x}$, $\boldsymbol{y} \in S$ satisfies: $f(\alpha\, \boldsymbol{x} + (1-\alpha)\, \boldsymbol{y}) \leq \alpha f(\boldsymbol{x}) + (1-\alpha) f(\boldsymbol{y}), \quad \forall \alpha \in (0,\, 1)$

How to find counter-examples

Think of values for $\boldsymbol{x}$, $\boldsymbol{y} \in S$ and $\alpha \in (0,\, 1)$ violating the inequality.

Convexity of Log-Sum-Exp [Second Deriv Test]

$$\mathsf{LSE}(\boldsymbol{x}) = \left(\log \sum_{k=1}^{10} \exp\left(x_k\right) \right)$$

$$\nabla \mathsf{LSE}(\boldsymbol{x}) = \frac{1}{\sum_{k=1}^{10} \exp\left(x_k\right)} \left[\exp(x_1),\, \ldots,\, \exp(x_{10})\right]^\top$$

$$\nabla^2 \mathsf{LSE}(\boldsymbol{x}) = \frac{1}{\left(\sum_{k=1}^{10} \exp\left(x_k\right)\right)} \mathsf{diag}\left(\left[\exp(x_1),\, \ldots,\, \exp(x_{10})\right]\right) - \frac{1}{\left(\sum_{k=1}^{10} \exp\left(x_k\right)\right)^2} \begin{bmatrix} \exp(x_1) \\ \vdots \\ \exp(x_{10}) \end{bmatrix} \left[\exp(x_1),\, \ldots,\, \exp(x_{10})\right]$$

Cauchy-Schwartz Inequality [2^{nd} Deriv Test 1]

∇^2LSE($\boldsymbol{x}$) is positive semidefinite? [Answer available online]

Check that for every $\boldsymbol{z} \in \mathbb{R}^n$, we have $\boldsymbol{z}^\top \nabla^2 \mathsf{LSE}(\boldsymbol{x})\boldsymbol{z} \geq 0$. Set $e_i = \exp(x_i)$ and premultiply by $\left(\sum_{k=1}^{10} \exp(x_k)\right)^2$:

$$\left(\sum_{k=1}^{10} \exp(x_k)\right)^2 \boldsymbol{z}^\top \nabla^2 \mathsf{LSE}(\boldsymbol{x})\boldsymbol{z}$$

$$= \left(\sum_{k=1}^{10} \exp(x_k)\right)\left(\sum_{k=1}^{10} \exp(x_k)\, z_k^2\right) - \left(\sum_{k=1}^{10} \exp(x_k)\, z_k\right)^2 \geq 0$$

By the Cauchy-Schwartz inequality.

I don't know another proof route using only C477 material (any ideas?). But here are some alternative, clever answers from C477 students ...

Definition of Variance [2^{nd} Deriv Test 2]

∇^2LSE($\boldsymbol{x}$) is positive semidefinite? [Clever Answer by C477 Student]

Check that for every $\boldsymbol{z} \in \mathbb{R}^n$, we have $\boldsymbol{z}^\top \nabla^2 \mathsf{LSE}(\boldsymbol{x})\boldsymbol{z} \geq 0$. Set $e_i = \exp(x_i)/\sum_{k=1}^{10} \exp(x_k)$ and note that $\sum_{k=1}^{10} e_k = 1$.

$$\begin{aligned}
\boldsymbol{z}^\top \nabla^2 \mathsf{LSE}(\boldsymbol{x})\boldsymbol{z} &= \boldsymbol{z}^\top \left(\mathsf{diag}\left(\boldsymbol{e}\right) - \boldsymbol{e}\boldsymbol{e}^\top\right)\boldsymbol{z} \\
&= \sum_{k=1}^{10} \exp\left(x_k z_k^2\right) - \left(\sum_{k=1}^{10} \exp\left(x_k\right) z_k\right)^2 \\
&= \mathbb{E}_{\boldsymbol{z}}[\boldsymbol{e}^2] - \mathbb{E}_{\boldsymbol{z}}[\boldsymbol{e}]^2 \geq 0
\end{aligned}$$

Another Alternative

Can you use the Gershgorin circle theorem (`https://en.wikipedia.org/wiki/Gershgorin_circle_theorem`) to get a third 2^{nd} deriv test?

Hölder's Inequality [From the Definition]

∇^2LSE($\boldsymbol{x}$) is positive semidefinite? [Clever Answer by C477 Student]

$$
\begin{aligned}
\alpha f(\boldsymbol{x}) + (1-\alpha) f(\boldsymbol{y}) &= \alpha \log\left(\sum_{k=1}^{10} \exp\left(x_k\right)\right) + (1-\alpha)\log\left(\sum_{k=1}^{10} \exp\left(y_k\right)\right) \\
&= \log\left(\sum_{k=1}^{10} \exp\left(x_k\right)\right)^{\alpha} + \log\left(\sum_{k=1}^{10} \exp\left(y_k\right)\right)^{1-\alpha} \\
&= \log\left(\left(\sum_{k=1}^{10} \exp\left(x_k\right)\right)^{\alpha} \left(\sum_{k=1}^{10} \exp\left(y_k\right)\right)^{1-\alpha}\right) \\
&\geq \log\left(\sum_{k=1}^{10} \exp\left(\alpha x_k + (1-\alpha) y_k\right)\right)
\end{aligned}
$$

The last inequality is due to Hölder's Inequality

`https://en.wikipedia.org/wiki/Hölder%27s_inequality`

Reminder: Manipulating Vectors & Matrices

Find points satisfying the FONC

$$f'(x^*) = 0 \implies 2\|\boldsymbol{a}\|_2^2 x^* = 2\langle \boldsymbol{a},\, \boldsymbol{b}\rangle \implies x^* = \frac{\langle \boldsymbol{a}, \boldsymbol{b}\rangle}{\|\boldsymbol{a}\|_2^2}$$

Please revise so that you don't do this . . .

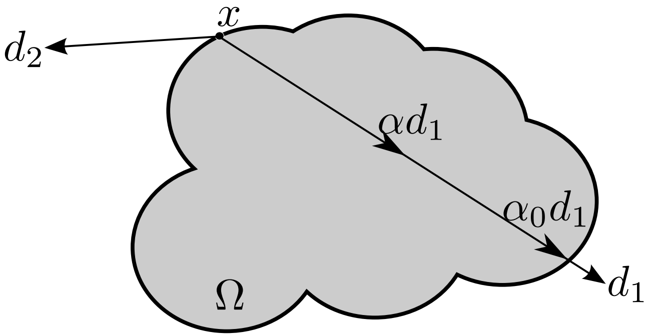
- $\dfrac{\langle \boldsymbol{a}, \boldsymbol{b}\rangle}{\|\boldsymbol{a}\|_2^2} = \dfrac{\langle \boldsymbol{a}, \boldsymbol{b}\rangle}{\langle \boldsymbol{a}, \boldsymbol{a}\rangle} = \dfrac{\boldsymbol{a}^\top \boldsymbol{b}}{\boldsymbol{a}^\top \boldsymbol{a}} = \dfrac{\sum_{i=1}^{n} a_i b_i}{\sum_{i=1}^{n} a_i a_i}$ (all valid alternatives);
- $\dfrac{\langle \boldsymbol{a}, \boldsymbol{b}\rangle}{\|\boldsymbol{a}\|_2^2} \neq \dfrac{\sum_{i=1}^{n} b_i}{\sum_{i=1}^{n} a_i}$ (can't cancel the a_i's);
- $\dfrac{\langle \boldsymbol{a}, \boldsymbol{b}\rangle}{\|\boldsymbol{a}\|_2^2} \neq \dfrac{\boldsymbol{b}}{\boldsymbol{a}}$ (can't divide a vector)

Revision materials: Background material; C145 notes (Piazza)

Reminder: Feasible Directions

Definition (Feasible Direction)

A vector $\boldsymbol{d} \in \mathbb{R}^n$, $\boldsymbol{d} \neq \mathbf{0}$ is a feasible direction at $\boldsymbol{x} \in \Omega$ if there exists an $\alpha_0 > 0$ such that $\boldsymbol{x} + \alpha\, \boldsymbol{d} \in \Omega$ for all $\alpha \in [0,\, \alpha_0]$.



Heads Up!

The definition requires $\boldsymbol{d} \neq \mathbf{0}$.

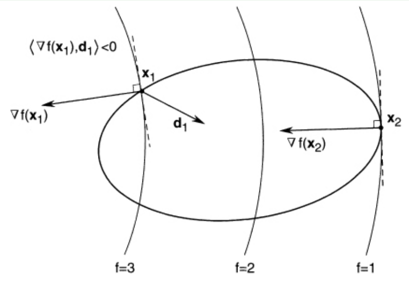
Reminder: First Order Necessary Condition (FONC)

Theorem (First Order Necessary Condition)

Let Ω be a subset of $\mathbb{R}^n$ and $f \in \mathcal{C}^1$ a real valued function on Ω. If $\boldsymbol{x}^*$ is a local minimiser of f over Ω, then for any feasible direction $\boldsymbol{d}$ at $\boldsymbol{x}^*$,

$$\boldsymbol{d}^\top \nabla f(\boldsymbol{x}^*) \geq 0.$$

Example: $\boldsymbol{x_1}$ does not satisfy the FONC, $\boldsymbol{x_2}$ does



Reminder: Second Order Necessary Condition (SONC)

Theorem

Let $\Omega \subset \mathbb{R}^n$, and $f \in \mathcal{C}^2$, $\boldsymbol{x}^*$ be a local minimiser of f over Ω and $\boldsymbol{d}$ be a feasible direction at $\boldsymbol{x}^*$. If $\boldsymbol{d}^\top \nabla f(\boldsymbol{x}^*) = 0$, then:

$$\boldsymbol{d}^\top \nabla^2 f(\boldsymbol{x}^*) \boldsymbol{d} \geq 0,$$

where $\nabla^2 f$ is the Hessian matrix of f.

Given a multivariable function $f : \mathbb{R}^n \to \mathbb{R}$ and a point $\boldsymbol{x} \in \mathbb{R}^n$, recall the Hessian, $\boldsymbol{H}(\boldsymbol{x})$, the matrix of second partial derivatives

$$\nabla^2 f(\boldsymbol{x}) = \boldsymbol{H}(\boldsymbol{x}) \triangleq \begin{pmatrix} \frac{\partial^2 f}{\partial x_1^2}(\boldsymbol{x}) & \frac{\partial^2 f}{\partial x_1 \partial x_2}(\boldsymbol{x}) & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n}(\boldsymbol{x}) \\ \frac{\partial^2 f}{\partial x_2 \partial x_1}(\boldsymbol{x}) & \frac{\partial^2 f}{\partial x_2^2}(\boldsymbol{x}) & \cdots & \frac{\partial^2 f}{\partial x_2 \partial x_n}(\boldsymbol{x}) \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1}(\boldsymbol{x}) & \frac{\partial^2 f}{\partial x_n \partial x_2}(\boldsymbol{x}) & \cdots & \frac{\partial^2 f}{\partial x_n^2}(\boldsymbol{x}) \end{pmatrix}$$

Reminder: Second Order Sufficient Condition (SOSC)

Theorem

Suppose that $f \in \mathcal{C}^2$ *in a region where* $\boldsymbol{x}^*$ *is an interior point. Suppose that,*

1. $\nabla f(\boldsymbol{x}^*) = \mathbf{0}$.
2. $\nabla^2 f(\boldsymbol{x}^*) \succ \mathbf{0}$, *i.e., the Hessian is positive definite at the point* $\boldsymbol{x}^*$.

Then $\boldsymbol{x}^*$ *is a strict local minimiser of* f.