

# Section 05: Solutions

---

## 1. Convexity

Convexity is defined for both sets and functions. For today we'll focus on discussing the convexity of functions.

**Definition 1** (convex functions). A function  $f : \mathbb{R}^d \rightarrow \mathbb{R}$  is **convex** on a set  $A$  if for all  $x, y \in A$  and  $\lambda \in [0, 1]$ :

$$f(\lambda x + (1 - \lambda)y) \leq \lambda f(x) + (1 - \lambda)f(y)$$

When the this definition holds with the inequality being reversed, then  $f$  is said to be concave. From the definition, it is clear that a function  $f$  is convex if and only if  $-f$  is concave.

(a) Why do we care whether a function is convex or not?

**Solution:**

Many numerical methods or algorithms were developed for finding local minima while in machine learning we are typically interested in finding the global minimum. Convex functions are useful because local minima are always the global minimum. Here is a short proof of this result: let  $x^*$  is a local minimizer for a convex function  $f$  and suppose  $f(x_0) < f(x^*)$ . Now note that there exists a  $\lambda \in (0, 1)$  such that  $y = \lambda x^* + (1 - \lambda)x_0$  and  $f(y) \geq f(x^*)$  (i.e.,  $y$  is close to  $x^*$  enough). We now have a contradiction:

$$f(y) \leq \lambda f(x^*) + (1 - \lambda)f(x_0) < f(x^*) \leq f(y).$$

In words, a line segment between any arbitrary point  $x_0$  and a local minimizer  $x^*$  should be entirely above the function by definition of convexity, ensuring that  $f(x_0) < f(x^*)$  cannot happen.

(b) Which of the following functions are convex? (Hint: draw a picture)

(i)  $|x|$  (ii)  $\cos(x)$  (iii)  $x^T x$

**Solution:**

$|x|$  and  $x^T x$  are both convex.  $\cos(x)$  is not convex since we can draw a line at two points (from say  $\frac{\pi}{2}$  to  $2\pi + \frac{\pi}{2}$ ) that is not entirely above the function.

Proof that  $|x|$  is convex:

$$\begin{aligned} f(\lambda x + (1 - \lambda)y) &= |\lambda x + (1 - \lambda)y| \\ &\leq \lambda|x| + (1 - \lambda)|y| \end{aligned}$$

Proof that  $x^T x$  is convex:

We begin by examining the definition: whenever  $\lambda \in [0, 1]$ , we have

$$\begin{aligned} (\lambda x + (1 - \lambda)y)^T (\lambda x + (1 - \lambda)y) &= \lambda^2 x^T x + (1 - \lambda)^2 y^T y + 2\lambda(1 - \lambda)x^T y \\ &= \lambda x^T x + (1 - \lambda)y^T y - \lambda(1 - \lambda)(x^T x - 2x^T y + y^T y) \\ &= \lambda x^T x + (1 - \lambda)y^T y - \lambda(1 - \lambda)(x - y)^T (x - y) \\ &\leq \lambda x^T x + (1 - \lambda)y^T y, \end{aligned}$$

where the inequality holds because  $(x - y)^T (x - y) = \|x - y\|_2^2 \geq 0$ . So our function is convex.

(c) Can a function be both convex and concave on the same set? If so, give an example. If not, describe why not.

**Solution:**

Linear functions (i.e. functions such that  $f(\lambda x + (1 - \lambda)y) = \lambda f(x) + (1 - \lambda)f(y)$ ) are both convex and concave.

## 2. Practical Methods for Checking Convexity

Using the definition to check whether a function is convex or not can be a tedious task in many situations. Some basic methods that can help us achieve the task in an efficient way are introduced below:

- for differentiable function, examine  $f(y) \geq f(x) + \nabla f(x)^T(y - x)$  for any  $x, y$  in the domain of  $f$ .
- for twice differentiable functions, examine  $\nabla^2 f(x) \succeq 0$  (i.e., the Hessian matrix is positive semidefinite).
- nonnegative weighted sum
- composition with affine function
- composition with convex function
- pointwise maximum and supremum

Note: there are even more such methods, which are covered in a convex optimization course or textbook.

- (a) If  $f$  is differentiable, then  $f$  is convex if and only if  $f(y) \geq f(x) + \nabla f(x)^T(y - x)$  for any  $x, y$  in the domain of  $f$ . A geometric interpretation of this characterization is that any tangent plane of a convex function  $f$  must lie entirely below  $f$ . One interesting application of this characterization is one of the most important inequalities in probability and statistics: the Jensen's inequality, which states that  $\mathbb{E}f(X) \geq f(\mathbb{E}(X))$  when  $f$  is convex.

**Solution:**

Let  $\mu = \mathbb{E}(X)$ , then since  $f$  is convex, we have

$$f(X) \geq f(\mu) + \nabla f(\mu)^T(X - \mu)$$

with probability 1. This means that taking expectation on both sides preserves the inequality:  $\mathbb{E}f(X) \geq f(\mu) = f(\mathbb{E}X)$ .

- (b) Let  $\alpha \geq 0$  and  $\beta \geq 0$ , and if  $f$  and  $g$  are convex, then  $\alpha f$ ,  $f + g$ ,  $\alpha f + \beta g$  are all convex. One application: When a (possibly complicated) objective function can be expressed as a sum (e.g., the negative log-likelihood function), then showing the convexity of each individual term is typically easier.

**Solution:**

From definition,  $\alpha f$  and  $f + g$  are easily proved convex. To show that  $\alpha f + \beta g$  is convex, first note that  $\alpha f$  and  $\beta g$  are both convex, hence their sum is convex as well.

- (c) Suppose  $f(\cdot)$  is convex, then  $g(x) := f(Ax + b)$  is convex. Use this method to show that  $\|Ax + b\|_1$  is convex (in  $x$ ), where  $\|z\|_1 = \sum_i |z_i|$ .

**Solution:**

With this method, we only need to show the convexity of  $\|x\|_1$ . This is true from definition by observing that

$$\|\lambda x + (1 - \lambda)y\|_1 = \sum_i |\lambda x_i + (1 - \lambda)y_i| \leq \sum_i \lambda |x_i| + (1 - \lambda)|y_i| = \lambda \|x\|_1 + (1 - \lambda)\|y\|_1,$$

where the inequality holds because of triangular inequality.

- (d) Suppose  $f(\cdot)$ ,  $g(\cdot)$  are both convex and  $f$  is non-decreasing. Show that  $(f \circ g)(\cdot)$  is also convex. **Solution:**

Using the definition of convexity, we have that

$$\begin{aligned}(f \circ g)(\lambda x + (1 - \lambda)y) &= f(g(\lambda x + (1 - \lambda)y)) \\ &\leq f(\lambda g(x) + (1 - \lambda)g(y)) \\ &\leq \lambda f(g(x)) + (1 - \lambda)f(g(y)) \\ &= \lambda(f \circ g)(x) + (1 - \lambda)(f \circ g)(y)\end{aligned}$$

- (e) Suppose you know that  $f_1$  and  $f_2$  are convex functions on a set  $A$ . The function  $g(x) := \max\{f_1(x), f_2(x)\}$  is also convex on  $A$ . In general, if  $f(x, y)$  is convex in  $x$  for each  $y$ , then  $g(x) := \sup_y f(x, y)$  is convex. Use this method to show that the largest eigenvalue of a matrix  $X$ ,  $\lambda_{\text{Max}}(X)$ , is convex in  $X$  (Using the definition of convexity would make this question quite difficult).

**Solution:**

Consider  $f(v, X) := v^T X v$ , then for each  $v$ , we have

$$f(v, \lambda X + (1 - \lambda)Y) = \lambda f(v, X) + (1 - \lambda)f(v, Y),$$

suggesting that  $f(v, X)$  is convex in  $X$  for each  $v$ . Then  $g(X) := \lambda_{\text{Max}}(X) = \sup_{\|v\|_2=1} f(v, X)$  is convex in  $X$  using this method.

- (f) Does the same result hold for  $h(x) := \min\{f_1(x), f_2(x)\}$ ? If so, give a proof. If not, provide convex functions  $f_1, f_2$  such that  $h$  is not convex.

**Solution:**

No, consider  $f_1(x) = x^2$ ,  $f_2(x) = (x - 1)^2$ . Then  $h(0) = h(1) = 0$ , but  $h(0.5) = 0.25$ , so  $h(0.5 * 0 + 0.5 * 1) = 0.25 > 0 = 0.5 * h(0) + 0.5 * h(1)$ . So the minimum of two convex functions is not convex in general.

- (g) Show that the objective function in linear regression  $\|Y - Xw\|_2^2$  is convex in  $w$ .

**Solution:**

We present two ways to solve this problem.

First, we use the property shown in (c), where it is sufficient if we show  $f(x) = \|x\|_2^2$  is a convex function since  $Y - Xw$  is an affine transformation of  $w$ . Furthermore, since  $g(y) = y^2$  is a convex function, by property shown in (d), we are only left to show that  $h(x) = \|x\|_2$  is convex. With triangle inequality, we finish the proof:

$$\begin{aligned}\|\lambda x + (1 - \lambda)y\|_2 &\leq \|\lambda x\|_2 + \|(1 - \lambda)y\|_2 \\ &= \lambda\|x\|_2 + (1 - \lambda)\|y\|_2\end{aligned}$$

Another way to show this is by show the Hessian is positive semi-definite. Let  $f(w) = (Y - Xw)^T(Y - Xw)$ , then

$$\nabla^2 f(w) = 2(X^T X),$$

which is clearly a positive semidefinite matrix.

(h) Show that the objectives for LASSO  $\|Y - X\omega\|_2^2 + \lambda\|\omega\|_1$  is convex in  $\omega$ . **Solution:**

By part (g), we know  $\|Y - X\omega\|_2^2$  is convex. By part (b) and (c), we know  $\lambda\|\omega\|_1$  is convex. By part (b) again, we know  $\|Y - X\omega\|_2^2 + \lambda\|\omega\|_1$  is convex.

(i) If  $f$  is twice differentiable with convex domain, then  $f$  is convex if and only if

$$\nabla^2 f(x) \succeq 0,$$

for any  $x$  in the domain of  $f$ .

Recall that the objective for logistic regression is

$$\arg \min_w \sum_{i=1}^n \log(1 + \exp(-y_i x_i^T w)) + \lambda \|w\|_2^2$$

Use this method to show that the objective function above is convex in  $w$ .

**Solution:**

So far we've shown that  $\|w\|_2^2$  is convex, and since sum of convex functions are still convex, we only left to show  $g(w) = \sum_{i=1}^n \log(1 + \exp(-y_i x_i^T w))$  being convex in  $w$ . Also, note that by property shown in (c), let  $z = y_i x_i^T w$  for each  $i$ , we can further reduce the problem to showing  $f(z) = \log(1 + \exp(z))$  being convex. Indeed, we have

$$\begin{aligned} f(z) &= \log(1 + \exp(z)) \\ f'(z) &= \frac{\exp(z)}{1 + \exp(z)} = 1 - \frac{1}{1 + \exp(z)} \\ f''(z) &= \frac{\exp(z)}{(1 + \exp(z))^2} \geq 0 \end{aligned}$$

### 3. Subgradients

We start with the definition of subgradients before discussing the motivation and its usefulness.

**Definition 2** (subgradients). A vector  $g \in \mathbb{R}^d$  is a subgradient of a convex function  $f : D \rightarrow \mathbb{R}$  at  $x \in D \subseteq \mathbb{R}^d$  if

$$f(y) \geq f(x) + g^T(y - x) \quad \text{for all } y \in D.$$

One interpretation of subgradient  $g$  is that the affine function (of  $y$ )  $f(x) + g^T(y - x)$  is a global underestimator of  $f$ . Note that if a convex function  $f$  is differentiable at  $x$  (i.e.,  $\nabla f(x)$  exists), then  $f(y) \geq f(x) + \nabla f(x)^T(y - x)$  is true for all  $y \in D$ , meaning that  $\nabla f(x)$  is a subgradient of  $f$  at  $x$ . But a subgradient can exist even when  $f$  is not differentiable at  $x$ .

(a) Why are subgradients useful in optimization? If  $g = 0$  is a subgradient of a function  $f$  at  $x^*$ , what does it imply?

**Solution:**

Consider the problem of minimizing a function  $f$ . If  $f$  is differentiable, we know that  $\nabla f(x) = 0$  is a necessary condition for  $x$  to be a local extremum. Together with the convexity of  $f$ ,  $\nabla f(x) = 0$  becomes a sufficient condition for  $x$  to be a local minimizer, and hence global minimizer. If solving for  $\nabla f(x) = 0$  analytically is difficult or infeasible, we have numerical methods such as gradient descent to obtain the

solution(s). These results are very useful in minimizing a convex differentiable function and subgradients can be treated as a generalization to situations when the underlying function (to be minimized) is nondifferentiable. In the analytical aspect, if  $g = 0$  is a subgradient of  $f$  at  $x^*$ , then from the definition above, we have

$$f(y) \geq f(x^*) + g^T(y - x^*) = f(x^*) \quad \text{for all } y \in D,$$

indicating that  $f(x^*)$  is a global minimum. To obtain the solution(s) numerically, we can consider subgradient descent.

- (b) What are the subgradients of  $f(x) = |x|$  at 0? (Hint: draw a picture and note that subgradients at a point might not be unique)

**Solution:**

From definition, a scalar  $g$  is a subgradient of  $f$  at  $x = 0$  if  $|y| = f(y) \geq f(0) + g(y - 0) = gy$  for any  $y \in \mathbb{R}$ . Solving  $|y| \geq gy$  yields that  $g \in [-1, 1]$ . That is, for any  $g \in [-1, 1]$ ,  $g$  is a subgradient of  $f$  at 0.

- (c) Some important results about subgradients are

- If  $f$  is convex, then a subgradient of  $f$  at  $x \in \text{int}(D)$  (interior of the domain of  $f$ ) always exists.
- If  $f$  is convex, then  $f$  is differentiable at  $x$  if and only if  $\nabla f(x)$  is the only subgradient of  $f$  at  $x$ .
- A point  $x^*$  is a global minimizer of a function  $f$  (not necessarily convex) if and only if  $g = 0$  is a subgradient of  $f$  at  $x^*$ .