EECS 127/227AT Optimization Models in Engineering UC Berkeley Fall 2023

Homework 6

Self grades are due at 11 PM one week after the homework.

1. Convex or Concave

Determine whether the following functions are convex, strictly convex, concave, strictly concave, both or neither.

(a) $f(x) = \mathrm{e}^x - 1$ on $\mathbb{R}$.

Solution: $f(x) = \mathrm{e}^x - 1$ on $\mathbb{R}$.
This is strictly convex since $\frac{\mathrm{d}^2 f}{\mathrm{d}x^2}(x) = \mathrm{e}^x > 0$ for all $x \in \mathbb{R}$.

(b) $f(x_1, x_2) = x_1 x_2$ on $\mathbb{R}^2_{++}$ (i.e. when $x_1 > 0$ and $x_2 > 0$).

Solution: $f(x_1, x_2) = x_1 x_2$ on $\mathbb{R}^2_{++}$.
This is neither convex nor concave. The Hessian of f is

$$\nabla^2 f(x) = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \tag{1}$$

which has eigenvalues ± 1 which implies the Hessian is neither positive semidefinite nor negative semidefinite.

(c) The log-likelihood of a set of points $\{x_1, \ldots, x_n\}$ that are normally distributed with mean μ and finite variance $\sigma > 0$ is given by:

$$f(\mu, \sigma) = n \log\left(\frac{1}{\sqrt{2\pi}\sigma}\right) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - \mu)^2 \tag{2}$$

i. Show that if we view the log likelihood for fixed σ as a function of the mean, i.e

$$g(\mu) = n \log\left(\frac{1}{\sqrt{2\pi}\sigma}\right) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - \mu)^2 \tag{3}$$

then g is strictly concave (equivalently, we say f is strictly concave in μ).

ii. **(OPTIONAL)** Show that if we view the log likelihood for fixed μ as a function of the inverse of the variance, i.e

$$h(z) = n \log\left(\frac{\sqrt{z}}{\sqrt{2\pi}}\right) - \frac{z}{2}\sum_{i=1}^{n}(x_i - \mu)^2 \tag{4}$$

then h is strictly concave (equivalently, we say f is strictly concave in $z = \frac{1}{\sigma^2}$). Note that we have used the dummy variable z to denote $\frac{1}{\sigma^2}$.

iii. **(OPTIONAL)** Show that f is not jointly concave in $\mu, \frac{1}{\sigma^2}$. *HINT: We say a function $w(x, y)$ with $x \in \mathcal{R}^m$ and $y \in \mathcal{R}^n$ is jointly convex if*

$$w\left(\lambda(x_1, y_1) + (1-\lambda)(x_2, y_2)\right) \le \lambda w((x_1, y_1)) + (1-\lambda) w((x_2, y_2)). \tag{5}$$

This is the same as letting $z = (x, y)$ and saying f is convex in z. We can define joint concavity in a similar fashion by reversing the inequalities.

Solution: For $g(\mu)$ we have,

$$\nabla g(\mu) = \sum_{i=1}^{n} \frac{x_i - \mu}{\sigma^2} \tag{6}$$
$$\nabla^2 g(\mu) = -\frac{n}{\sigma^2} < 0. \tag{7}$$

Since σ is finite, g is strictly concave (equivalently f is strictly concave in μ).
For $h(z)$ we have,

$$\nabla h(z) = \frac{n}{2z} - \sum_{i=1}^{n} \frac{(x_i - \mu)^2}{2} \tag{8}$$
$$\nabla^2 h(z) = -\frac{n}{2z^2} < 0. \tag{9}$$

Since z^2 is finite ($\sigma > 0$), h is strictly concave (equivalently f is strictly concave in σ^2). For $f(\mu, \frac{1}{\sigma^2})$, we find the second order partial derivatives and stack them in the Hessian. We have,

$$\nabla^2 f(\mu, \frac{1}{\sigma^2}) = \begin{bmatrix} -\frac{n}{\sigma^2} & \sum_{i=1}^{n}(x_i - \mu) \\ \sum_{i=1}^{n}(x_i - \mu) & -\frac{n\sigma^4}{2} \end{bmatrix}. \tag{10}$$

The determinant of the Hessian is given by,

$$\det\left(\nabla^2 f\right) = \frac{n^2\sigma^2}{2} - (\sum_{i=1}^{n}(x_i - \mu))^2. \tag{11}$$

and the trace of the Hessian is given by,

$$\operatorname{tr}\left(\nabla^2 f\right) = -\frac{n}{\sigma^2} - \frac{n\sigma^4}{2} < 0 \tag{12}$$

Note that the trace is the sum of the eigenvalues, and the determinant is the product of the eigenvalues. Since the trace is always negative, if the determinant is negative it must imply that one eigenvalue is positive and another is negative; that is, we have f is neither convex nor concave. It is easy to see that $\det\left(\nabla^2 f\right)$ can sometimes be negative – for example, if we choose σ^2 to be close to zero and μ away from x_i, the second negative term dominates and makes $\det\left(\nabla^2 f\right) \leq 0$. **Aside:** Note however, in the maximum likelihood estimates, the Hessian is negative semi-definite implying that locally the function is concave. More concretely, at

$$\widehat{\mu} = \frac{1}{n}\sum_{i=1}^{n} x_i, \quad \widehat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \widehat{\mu})^2 \tag{13}$$

we have $\nabla^2 f(\widehat{\mu}, 1/\widehat{\sigma}^2) \preceq 0$

(d) Prove that $f(x) = \log(1 + \mathrm{e}^x)$ is convex. Note that this implies that $g(x) = -f(x) = \log\left(\frac{1}{1+\mathrm{e}^x}\right)$ is concave. Compare this to $h(x) = \frac{1}{1+\mathrm{e}^x}$, is $h(x)$ convex or concave?

Solution: We will do this by verifying the second order sufficient conditions for convexity. We have the derivatives of f can be computed using the chain rule as follows:

$$f'(x) = \frac{\partial f}{\partial x}(x) = \frac{\mathrm{e}^x}{1 + \mathrm{e}^x}$$
$$f''(x) = \frac{\partial^2 f}{\partial x^2}(x) = \frac{\mathrm{e}^x}{1 + \mathrm{e}^x} + \frac{-\mathrm{e}^x}{(1 + \mathrm{e}^x)^2}\mathrm{e}^x = \frac{\mathrm{e}^x}{(1 + \mathrm{e}^x)^2} > 0.$$

Since we have $f''(x) > 0$ for all x, we conclude that the function f is strictly convex.

Now consider $h(x) = \frac{1}{1+e^x}$. We use the second order condition for convexity, and calculate

$$\nabla h(x) = \frac{-e^x}{(1+e^x)^2}; \quad \nabla^2 h(x) = \frac{(e^x - 1)e^x}{(e^x + 1)^3}.$$

The second derivative is positive for $x > 0$, and negative for $x < 0$, hence the function is neither convex nor concave.

2. Further characterizations of convexity

Show that $\sigma_1 : \mathbb{R}^{m\times n} \to \mathbb{R}_+$, the function that maps a matrix to its largest singular value, is a convex function, with domain $\mathbb{R}^{m\times n}$.

HINT: You may express $\sigma_1(A)$ using the ℓ^2 operator norm of A:

$$\sigma_1(A) = \max_{\vec{x}\in\mathbb{R}^n : \|\vec{x}\|_2 = 1} \|A\vec{x}\|_2 ,$$

This question proves that this norm is convex, so you may not use the fact that norms are convex.

Solution: We have

$$\sigma_1(A) = \max_{\vec{x}\in\mathbb{R}^n : \|\vec{x}\|_2 = 1} \|A\vec{x}\|_2 ,$$

which is the characterization of the largest singular value of a matrix as its induced ℓ^2 norm. Note that this expresses the function

$$\sigma_1 : \mathbb{R}^{m\times n} \to \mathbb{R}_+,$$

as the supremum of a family of function, one for each $x \in \mathbb{R}^n$ with $\|\vec{x}\|_2 = 1$, which we may temporarily call ψ_x for convenience, given by

$$\psi_x(A) \coloneqq \|A\vec{x}\|_2 \, \forall A \in \mathbb{R}^{m\times n}.$$

If we could prove that each $\psi_x : \mathbb{R}^{m\times n} \to \mathbb{R}_+$ is convex then we could be done, because the supremum of a family of convex functions is convex.

To show that $\psi_x(\cdot)$ is convex, consider two matrices $A, B \in \mathbb{R}^{m\times n}$ and their linear combination $\theta A + (1-\theta)B$, where $\theta \in [0, 1]$.

$$\begin{aligned}\psi_x(\theta A + (1-\theta)B) &= \|(\theta A + (1-\theta)B)\vec{x}\|_2 = \|\theta A\vec{x} + (1-\theta)B\vec{x}\|_2 \\ &\le \theta \|A\vec{x}\|_2 + (1-\theta)\|B\vec{x}\|_2 = \theta\psi_x(A) + (1-\theta)\psi_x(B),\end{aligned}$$

where the inequality comes from the triangle inequality on the ℓ^2 norm. Hence, $\psi_x(\cdot)$ is convex.

3. Convex and strictly convex functions

(a) Recall that a function $f : \mathbb{R}^n \to \mathbb{R}$ is said to be strictly convex if it satisfies Jensen's inequality with strict inequality, i.e., $\forall \vec{x} \neq \vec{y} \in \mathbb{R}^n$ and $\forall t \in (0,1)$, we have

$$f(t\vec{x} + (1-t)\vec{y}) < tf(\vec{x}) + (1-t)f(\vec{y})$$

Show that for a strictly convex function $f : \mathbb{R}^n \to \mathbb{R}$, the problem

$$\min_{\vec{x} \in \mathbb{R}^n} f(\vec{x}) \tag{14}$$

has at most one solution.

HINT: Try to argue by contradiction assuming that there are two solutions $\vec{x}_1, \vec{x}_2$ which achieve the minimum value. Argue that using these two points you can find another point in $\mathbb{R}^n$ with strictly smaller function value.

Solution: Assume that $f^* = \min_{\vec{x} \in \mathbb{R}^n} f(\vec{x})$ has at least two different optimal solutions $\vec{x}_1, \vec{x}_2 \in \mathbb{R}^n$. Hence $f^* = f(\vec{x}_1) = f(\vec{x}_2)$. Consider $\vec{z} = (\vec{x}_1 + \vec{x}_2)/2$.

$$\begin{aligned} f(\vec{z}) &= f((\vec{x}_1 + \vec{x}_2)/2) \\ &< (f(\vec{x}_1) + f(\vec{x}_2))/2 \\ &= f^* \end{aligned}$$

where the inequality follows from strict convexity. Hence we've shown that $\vec{z}$ has functional value strictly smaller than f^* and hence f^* was not the optimal value giving us a contradiction.

(b) Prove that for all convex optimization problems $\min_{\vec{x} \in \mathcal{X}} f(\vec{x})$, where f is a convex function and $\mathcal{X}$ is a convex subset of its domain, all local minima are global minima. You may not assume that f is differentiable.

HINT: Start with assuming $\vec{x}^\star$ is a local minimum that is not global. Then there must exist some $\vec{\tilde{x}}$ satisfying $f(\vec{\tilde{x}}) < f(\vec{x}^\star)$. Use the definition of the convexity of a function to prove by contradiction.

Solution: To arrive at a contradiction, suppose $\vec{x}^\star$ is a local minimum that is not global. Let $\vec{\tilde{x}} \in \mathcal{X}$ be given such that $f(\vec{\tilde{x}}) < f(\vec{x}^\star)$. Then by convexity $\lambda\vec{x}^\star + (1-\lambda)\vec{\tilde{x}} \in \mathcal{X}$ and hence

$$\begin{aligned} f(\lambda\vec{x}^\star + (1-\lambda)\vec{\tilde{x}}) &\leq \lambda f(\vec{x}^\star) + (1-\lambda)f(\vec{\tilde{x}}) \\ &< \lambda f(\vec{x}^\star) + (1-\lambda)f(\vec{x}^\star) \\ &= f(\vec{x}^\star) \end{aligned}$$

Thus, for all $\lambda \in [0,1)$, we have $f(\lambda\vec{x}^\star + (1-\lambda)\vec{\tilde{x}}) < f(\vec{x}^\star)$. Then we can make $\lambda\vec{x}^\star + (1-\lambda)\vec{\tilde{x}}$ arbitrarily close to $\vec{x}^\star$ (by choosing λ to be arbitrarily close to 1), contradicting that $\vec{x}^\star$ is a local minimum.

4. First Order Criteria for Convexity, Strict Convexity, and Strong Convexity

Let $f : \mathbb{R}^n \to \mathbb{R}$ be a differentiable function with domain $\mathrm{dom}(f)$. Note that requiring that f is differentiable automatically implies that we are assuming that $\mathrm{dom}(f)$ is an open set.

(a) Show that f is convex iff it holds that $\mathrm{dom}(f)$ is a convex set and for all $\vec{x}, \vec{y} \in \mathrm{dom}(f)$ we have

$$\left(\nabla f(\vec{y}) - \nabla f(\vec{x})\right)^T (\vec{y} - \vec{x}) \geq 0. \tag{15}$$

Remark: When a function $g : \mathbb{R}^n \to \mathbb{R}^n$ satisfies the condition $\left(g(\vec{y}) - g(\vec{x})\right)^T (\vec{y} - \vec{x}) \geq 0$ for all $\vec{x}, \vec{y} \in \mathrm{dom}(g)$, we say that that g is *monotone*. Note that this is consistent with the use of the term "monotone" to refer to a function $g : \mathbb{R} \to \mathbb{R}$ that is monotonically increasing (although one often uses the term in this case to also apply to a function $g : \mathbb{R} \to \mathbb{R}$ that is monotonically decreasing). Thus the condition in (15) is saying that ∇f is monotone.

Solution: Suppose f is convex. Then $\mathrm{dom}(f)$ is convex. Also, given any $\vec{x}, \vec{y} \in \mathrm{dom}(f)$, we have

$$f(\vec{y}) - f(\vec{x}) \geq \nabla f(\vec{x})^T (y - x),$$

and

$$f(\vec{x}) - f(\vec{y}) \geq \nabla f(\vec{y})^T (\vec{x} - \vec{y}).$$

Adding these two inequalities gives

$$0 \geq \left(\nabla f(\vec{x}) - \nabla f(\vec{y})\right)^T (\vec{y} - \vec{x}),$$

which is the same as (15).

Conversely, suppose the differentiable function f is such that $\mathrm{dom}(f)$ is a convex set and the condition in (15) holds for all $\vec{x}, \vec{y} \in \mathrm{dom}(f)$. Given $\vec{x}, \vec{y} \in \mathrm{dom}(f)$, we have $x + t(\vec{y} - \vec{x}) \in \mathrm{dom}(f)$ for all t in an open interval containing $[0, 1]$, because $\mathrm{dom}(f)$ is open and convex. Define

$$g(t) := f(\vec{x} + t(\vec{y} - \vec{x})),$$

for t in such an open interval containing $[0, 1]$. Then

$$g'(t) = \nabla f(\vec{x} + t(\vec{y} - \vec{x}))^T (\vec{y} - \vec{x}), \;\; t \in [0, 1].$$

We have

$$\begin{aligned}
& f(\vec{y}) - f(\vec{x}) \\
&= g(1) - g(0) \\
&= \int_0^1 g'(t) dt \\
&= \int_0^1 \nabla f(\vec{x} + t(\vec{y} - \vec{x}))^T (\vec{y} - \vec{x}) dt \\
&= \int_0^1 \left(\nabla f(\vec{x} + t(\vec{y} - \vec{x})) - \nabla f(\vec{x})\right)^T (\vec{y} - \vec{x}) dt + \nabla f(\vec{x})^T (\vec{y} - \vec{x}) \\
&\geq \nabla f(\vec{x})^T (\vec{y} - \vec{x}),
\end{aligned}$$

where we have used (15) in the last step (for the pair $x, x + t(\vec{y} - \vec{x}) \in \mathrm{dom}(f)$). Since this holds for all $\vec{x}, \vec{y} \in \mathrm{dom}(f)$ and since $\mathrm{dom}(f)$ is a convex set, we conclude that f is convex.

(b) Recall that a function $f : \mathbb{R}^n \to \mathbb{R}$ with domain $\text{dom}(f)$ is said to be strictly convex if $\text{dom}(f)$ is a convex set and for all $\vec{x} \neq \vec{y} \in \text{dom}(f)$ and $\lambda \in (0,1)$ we have

$$f(\lambda\vec{x} + (1-\lambda)\vec{y}) < \lambda f(\vec{x}) + (1-\lambda)f(\vec{y}).$$

Show that f is strictly convex iff it holds that $\text{dom}(f)$ is a convex set and for all $\vec{x} \neq \vec{y} \in \text{dom}(f)$ we have

$$\left(\nabla f(\vec{y}) - \nabla f(\vec{x})\right)^T (\vec{y} - \vec{x}) > 0. \tag{16}$$

Remark: When a function $g : \mathbb{R}^n \to \mathbb{R}^n$ satisfies the condition $\left(g(\vec{y}) - g(\vec{x})\right)^T (\vec{y} - \vec{x}) > 0$ for all $\vec{x} \neq \vec{y} \in \text{dom}(g)$ we say that that g is *strictly monotone*.

Solution: The argument is more or less identical to that in the preceding sub-part of this part of the question, but with strict inequality instead of inequality.

(c) Let $m > 0$. A function $f : \mathbb{R}^n \to \mathbb{R}$ is called m*-strongly convex* if the function

$$h(\vec{x}) := f(\vec{x}) - \frac{m}{2}\|\vec{x}\|_2^2,$$

with $\text{dom}(h) := \text{dom}(f)$, is convex. Let $f : \mathbb{R}^n \to \mathbb{R}$ be a differentiable function with domain $\text{dom}(f)$ (note that this means $\text{dom}(f)$ must be an open set). Given $m > 0$, show that f is m-strongly convex iff it holds that $\text{dom}(f)$ is a convex set and for all $\vec{x}, \vec{y} \in \text{dom}(f)$ we have

$$\left(\nabla f(\vec{y}) - \nabla f(\vec{x})\right)^T (\vec{y} - \vec{x}) \geq m\|\vec{x} - \vec{y}\|_2^2. \tag{17}$$

Remark: When a function $g : \mathbb{R}^n \to \mathbb{R}^n$ satisfies the condition $\left(g(\vec{y}) - g(\vec{x})\right)^T (\vec{y} - \vec{x}) > m\|\vec{x} - \vec{y}\|^2$ for all $\vec{x}, \vec{y} \in \text{dom}(g)$ we say that that g is *strongly monotone* or *coercive* (confusingly, the term "coercive" is also used in a different sense, which we will encounter later). Thus the condition in (17) is saying that ∇f is strongly monotone.

Solution: Suppose f is m-strongly convex. Then, in particular, f is convex, so $\text{dom}(f)$ is a convex set. Let

$$h(\vec{x}) := f(\vec{x}) - \frac{m}{2}\|\vec{x}\|_2^2,$$

with $\text{dom}(h) := \text{dom}(f)$. Then h is convex. By the first sub-part of the first part of this question, we have

$$\left(\nabla h(\vec{y}) - \nabla h(\vec{x})\right)^T (\vec{y} - \vec{x}) \geq 0,$$

for all $\vec{x}, \vec{y} \in \text{dom}(h) = \text{dom}(f)$. Since

$$\nabla h(\vec{x}) = \nabla f(\vec{x}) - m\vec{x}, \quad \nabla h(\vec{y}) = \nabla f(\vec{y}) - m\vec{y},$$

this gives (17), as desired.

Conversely, suppose $\text{dom}(f)$ is convex and (17) holds for all $\vec{x}, \vec{y} \in \text{dom}(f)$. If we define

$$h(\vec{z}) := f(\vec{z}) - \frac{m}{2}\|\vec{z}\|_2^2,$$

for all $\vec{z} \in \text{dom}(f)$, with $\text{dom}(h) := \text{dom}(f)$, then we have

$$\nabla h(\vec{z}) = \nabla f(\vec{z}) - m\vec{z},$$

and so from (17) we have

$$\left(\nabla h(\vec{y}) - \nabla h(\vec{x})\right)^T (\vec{y} - \vec{x}) \geq 0,$$

from which it follows, since this holds for all $\vec{x} \neq \vec{y} \in \text{dom}(h) = \text{dom}(f)$, that h is convex (as shown in the first sub-part of the first part of this question) and hence that f is m-strongly convex.

5. **Homework Process**

With whom did you work on this homework? List the names and SIDs of your group members.

NOTE: If you didn't work with anyone, you can put "none" as your answer.