EECS 127/227AT Optimization Models in Engineering UC Berkeley Spring 2024
Homework 7

This homework is due at 11 PM on March 1, 2024.

Submission Format: Your homework submission should consist of a single PDF file that contains all of your answers (any handwritten answers should be scanned).

1. **Midsemester Survey**

   Please complete this mid-semester survey at the following link: link. You will get a code at the end of the survey; write it in as the solution for this problem.

2. 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 $\mathbb{R}$.

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

(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:

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

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

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

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

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

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*

$$w\left(\lambda(x_1, y_1) + (1-\lambda)(x_2, y_2)\right) \le \lambda w((x_1, y_1)) + (1-\lambda) w((x_2, y_2)). \tag{4}$$

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.

(d) $f(x) = \log(1 + \mathrm{e}^x)$. 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?

3. 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:

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

and use the fact that the supremum of a family of convex functions is convex. This question proves that this norm is convex, so you may not use the fact that norms are convex.

4. 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

$$\min_{\vec{x} \in \mathbb{R}^n} f(\vec{x}) \tag{5}$$

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.

(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 set, 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, and $\vec{\tilde{x}}$ is a global minimum. Use the definition of the convexity of a function to prove by contradiction.

5. Convexity of Rank 1 Matrices

In this problem, we explore the effect of rank constraints on the convexity of matrix sets.

First, consider the set of all 2×2 matrices with diagonal elements $(1, 2)$, which we can write explicitly as

$$\mathcal{S} = \left\{ \begin{bmatrix} 1 & x \\ y & 2 \end{bmatrix} \middle| \, x, y \in \mathbb{R} \right\}. \tag{6}$$

(a) Is set $\mathcal{S}$ convex? If so, provide a proof, and if not, provide a counterexample.

(b) Suppose we now wish to define $\mathcal{S}_1 \subset \mathcal{S}$, the set of all rank-1 matrices in $\mathcal{S}$. Write out conditions on x and y (i.e. equation constraints that x and y must satisfy) to define $\mathcal{S}_1$ explicitly.

(c) Is set $\mathcal{S}_1$ convex? If so, provide a proof, and if not, provide a counterexample.

(d) In this class, we will sometimes pose optimization problems in which we optimize over sets of matrices. Since low-dimensional models are often easier to interpret, it would be nice to impose rank constraints on these solution matrices. Suppose we wish to solve the optimization problem

$$\min_{A \in \mathcal{S}_1} \|A\|_F^2 \tag{7}$$

which is equivalent to

$$\min_{A \in \mathcal{S}} \quad \|A\|_F^2 \tag{8}$$

$$\text{s.t.} \quad \operatorname{rank}(A) = 1. \tag{9}$$

Is this optimization problem convex?

6. Convexity

(a) Show the conservation of convexity through affine transformation, i.e., prove that if $S \subseteq \mathbb{R}^n$ is convex, then the image of S under an affine function f,

$$f(S) = \{f(\vec{x}) \mid \vec{x} \in S\}, \tag{10}$$

is convex.

(b) Show that a function $f : \mathbb{R}^n \to \mathbb{R}$ is convex if and only if its epigraph, defined as $\operatorname{epi}(f) = \{(\vec{x}, t) \mid \vec{x} \in \operatorname{dom}(f), f(\vec{x}) \leq t\}$, is convex.

7. Gradient Descent Algorithm

Given a continuous and differentiable function $f : \mathbb{R}^n \to \mathbb{R}$, the gradient of f at any point $\vec{x}$, $\nabla f(\vec{x})$, is orthogonal to the level curve of f at point $\vec{x}$, and it points in the increasing direction of f. In other words, moving from point $\vec{x}$ in the direction $\nabla f(\vec{x})$ leads to an increase in the value of f, while moving in the direction of $-\nabla f(\vec{x})$ decreases the value of f. This idea gives an iterative algorithm to minimize the function f: the gradient descent algorithm.

(a) Consider $f(x) = \frac{1}{2}(x-2)^2$, and assume that we use the gradient descent algorithm:

$$x_{k+1} = x_k - \eta \nabla f(x_k) \quad \forall k \geq 0, \tag{11}$$

with some random initialization x_0, where $\eta > 0$ is the step size (or the learning rate) of the algorithm. Write $(x_k - 2)$ in terms of $(x_0 - 2)$, and show that x_k converges to 2, which is the unique minimizer of f, when $\eta = 0.2$.

(b) Let $\alpha, \beta \in \mathbb{R}$ and that for all η such that $\alpha < \eta < \beta$, the gradient descent algorithm converges to 2 from all possible initializations in $\mathbb{R}$. What are the smallest α and the largest β? After you determine α, β, answer the following. What happens when we set (i) $\eta = \alpha$ (ii) $\eta = \beta$ (iii) $\eta > \beta$?

(c) Now assume that we use the gradient descent algorithm to minimize $f(\vec{x}) = \frac{1}{2}\left\|A\vec{x} - \vec{b}\right\|_2^2$ for some $A \in \mathbb{R}^{m \times n}$ and $\vec{b} \in \mathbb{R}^m$, where A has full column rank. First compute $\nabla f(\vec{x})$. Note that $(A^\top A)^{-1}A^\top \vec{b}$ is the solution to the least-squares problem, and $(\vec{x}_k - (A^\top A)^{-1}A^\top \vec{b})$ is the distance from the solution at time k. Write $(\vec{x}_k - (A^\top A)^{-1}A^\top \vec{b})$ in terms of $(\vec{x}_0 - (A^\top A)^{-1}A^\top b)$.

(d) Now consider $f(\vec{x}) = \frac{1}{2}\left\|A\vec{x} - \vec{b}\right\|_2^2 + \frac{1}{2}\lambda \|\vec{x}\|_2^2$ for some $A \in \mathbb{R}^{m \times n}$ and $\vec{b} \in \mathbb{R}^m$, where A has full column rank. Suppose we solve this problem via gradient descent with step-size $\eta = \frac{1}{\sigma_1^2 + \lambda}$, where σ_1 is the maximum singular value of A. Show the gradient descent converges.

8. 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.