Please provide complete and well-written solutions to the following exercises.

Due October 11, in the discussion section.

Homework 2

Exercise 1. Let $f\colon \mathbf{R}^n \to \mathbf{R}$ be a strictly convex function. Show that f has at most one global minimum.

Then, find a convex set $K \subseteq \mathbf{R}$ and a strictly convex function $f\colon K \to \mathbf{R}$ such that f does not have a global minimum.

Exercise 2. Let $f\colon \mathbf{R}^n \to \mathbf{R}$ be a convex function. Let $x \in \mathbf{R}^n$ be a local minimum of f. Show that x is in fact a global minimum of f.

Now suppose additionally that $f \in C^1$, and $x \in \mathbf{R}^n$ satisfies $\nabla f(x) = 0$. Show that x is a global minimum of f.

Exercise 3. In statistics and other applications, we can be presented with data points $(x_1, y_1), \ldots, (x_n, y_n)$. We would like to find the line $y = mx + b$ which lies "closest" to all of these data points. Such a line is known as a **linear regression**. There are many ways to define the "closest" such line. The standard method is to use **least squares minimization**. A line which lies close to all of the data points should make the quantities $(y_i - mx_i - b)$ all very small. We would like to find numbers m, b such that the following quantity is minimized:

$$f(m,b) = \sum_{i=1}^{n} (y_i - mx_i - b)^2.$$

Show that the global minimum value of f is achieved when

$$m = \frac{\left(\sum_{i=1}^n x_i\right)\left(\sum_{j=1}^n y_j\right) - n\left(\sum_{k=1}^n x_k y_k\right)}{\left(\sum_{i=1}^n x_i\right)^2 - n\left(\sum_{j=1}^n x_j^2\right)}.$$

$$b = \frac{1}{n}\left(\sum_{i=1}^n y_i - m\sum_{j=1}^n x_j\right).$$

(In probabilistic terminlogy, $-m$ is a covariance divided by a variance.)

Exercise 4. Find a function $f\colon \mathbf{R} \to \mathbf{R}$ such that no local or global maximum of f exists, and no local or global minimum of f exists.

Exercise 5. Let $f\colon \mathbf{R}^n \to \mathbf{R}$. A version of Taylor's Theorem for functions on $\mathbf{R}^n$ follows from Taylor's Theorem for functions on $\mathbf{R}$ in the following way. (For simplicity, we look at the Taylor expansion of f at $x = 0$.) Let $y \in \mathbf{R}^n$, let $t \in \mathbf{R}$, and define $g\colon \mathbf{R} \to \mathbf{R}$ by $g(t) = f(ty)$. Then Taylor's Theorem for g holds. Using the Chain rule, what are the first two or three terms in the Taylor expansion of g, in terms of derivatives of f at $x = 0$?

Exercise 6. Maximize $f(x,y)=x^2+y^2$ subject to the constraint $x^2+2y^2=1$.

Exercise 7. Suppose that we have a probability distribution on the set $\{1,\ldots,n\}$, i.e. a sequence $p=(p_1,\ldots,p_n)$ of probabilities in the set $\overline{\mathcal{P}_n}$, where

$$\overline{\mathcal{P}_n} := \left\{p\in[0,1]^n\colon\ \sum_{i=1}^n p_i = 1\right\},\qquad \mathcal{P}_n := \left\{p\in(0,1)^n\colon\ \sum_{i=1}^n p_i = 1\right\}.$$

A fundamental quantity for a probability distribution p is its *entropy*

$$S(p) := -\sum_{i=1}^n p_i\log p_i\,.$$

(We extend the function $x\log x$ to 0 by continuity, so that $0\log 0 := 0$.) The entropy of p measures the disorder or lack of information in p.

(i) Using Lagrange multipliers, find the local maximum q of S on the set $\mathcal{P}_n$. Compute the value of S at q.
(ii) Prove that S reaches its maximum on $\overline{\mathcal{P}_n}$ at q.

Exercise 8. Let A be a real symmetric positive definite $n\times n$ matrix. Let $b\in\mathbf{R}^n$. Define $f\colon\mathbf{R}^n\to\mathbf{R}$ so that, for any $y\in\mathbf{R}^n$,

$$f(y) := \frac{1}{2}y^TAy - b^Ty$$

Show that f is strictly convex. Conclude that f has exactly one global minimum. (Recall that strict convexity alone does not guarantee that a global minimum exists.)

More generally, let $1\leq k\leq n-1$, let $H\subseteq\mathbf{R}^n$ be a k-dimensional subspace of $\mathbf{R}^n$, let $x^{(0)}\in\mathbf{R}^n$ and let

$$K := \{x^{(0)}+h\colon h\in H\}.$$

Let $f_K\colon K\to\mathbf{R}$ by $f_K(y) := \frac{1}{2}y^TAy - b^Ty$, $\forall\ y\in K$. Then f_K also has exactly one global minimum $x_K\in K$. Moreover, $\nabla f(x_K) = Ax_K - b$ is orthogonal to H. Conversely, if $x_K\in K$ satisfies Ax_K-b is orthogonal to H, then x_K is the unique global minimum of f on K.

Exercise 9. Give an example of a function $f\colon\mathbf{R}^2\to\mathbf{R}$ such that $\nabla f(0)=0$, all eigenvalues of $D^2f(0)$ are nonnegative, but f does not have a local minimum at 0.

Exercise 10. Let $f\colon\mathbf{R}^2\to\mathbf{R}$ so that $f(x,y)=x^2+y^2(1+x)^3$. Show that f has one critical point which is a local minimum, but f has no global maximum, and f has no global minimum.

That is, having only one critical point which is a local minimum does not imply that this point is a global minimum.