All hw should be uploaded to canvas as a *pdf*. Make sure that if you scan your handwritten notes that they are legible and appropriately oriented. If you use an online resource to solve any problem, please appropriately cite that source.

Problem 1. (Convex sets.) For each of the sets described below, either show that the set is convex, or provide a counter example (for example, give 2 points in the set that violate convexity). To show convexity, you can either (1) verify the definition of convexity by checking the line segments connecting every pair of points in the set, or (2) use convexity of known sets shown in the lecture (hyperplanes, halfspaces, norm balls, ellipsoids, norm cones), and properties that preserve convexity.

a. *Voronoi region.* Let $x_0, \ldots, x_M \in \mathbf{R}^n$; the set of all points that are closer in Euclidean norm to x_0 than any other x_k, i.e., $\{x \in \mathbf{R}^n \mid \|x - x_0\|_2 \le \|x - x_k\|_2,\ k = 1, \ldots, M\}$. *Hint:* Use the intersection property.

b. *Union of intervals.* $S = \{x \in \mathbf{R} \mid x < 0\} \bigcup \{x \in \mathbf{R} \mid 0 < x < 1\}$.

c. *Simplex.* $\{x \in \mathbf{R}^n \mid x \succeq 0, \mathbf{1}^T x \le 1\}$.

d. $\{(x, y) \in \mathbf{R}^2 \mid y \ge \frac{1}{x}, x > 0\}$. *Hint:* For this set, it is enough to draw it and inspect visually.

e. $\{x \in \mathbf{R}^n \mid x^T A x \ge 0\}$, where we assume A is a positive semidefinite matrix ($A \succeq 0$).

f. $\{x \in \mathbf{R}^n \mid \sum_{i=1}^{n-1}(x_{i+1} - x_i)^2 \le 1\}$. *Hint:* Recall the 'difference matrix' applied to vector x and use it to describe this set. Then use one of the properties that preserve convexity of sets.

Problem 2. (Norm cones.) Show that the ℓ_∞-norm cone, defined as $C = \{(x, t) \mid \|x\|_\infty \le t\}$ with $x \in \mathbf{R}^n$ and $t \in \mathbf{R}$, is a convex set, using the definition—that is,

$$\forall \begin{bmatrix} x_1 \\ t_1 \end{bmatrix}, \begin{bmatrix} x_2 \\ t_2 \end{bmatrix} \in C \text{ and } \ 0 \le \theta \le 1,$$

show that

$$\theta \begin{bmatrix} x_1 \\ t_1 \end{bmatrix} + (1 - \theta) \begin{bmatrix} x_2 \\ t_2 \end{bmatrix} \in C.$$

Problem 3. (Shrunk set.) For $a \ge 0$ we define $S_{-a} = \{x \mid B(x, a) \subseteq S\}$, where $B(x, a)$ is a norm ball (in a general norm $\|\cdot\|$) centered at x, with radius a; we refer to S_{-a} as S *shrunk* by a. Suppose S is convex, show that S_{-a} is also convex by showing for any two points in the set S_{-a}, their convex combination should also belong to S_{-a}.

Problem 4. (Convex functions.) For each of the following functions determine whether it is convex, concave, or neither (and say why). *Hint:* Compute the Hessian first.

a. $f(x_1, x_2) = x_1 x_2$ on $\mathbf{R}^2_{++}$.

b. $f(x_1, x_2) = x_1 / x_2$ on $\mathbf{R}^2_{++}$.

Problem 5. (Quadratic functions.)

a. Consider the function $f(x_1, x_2) = x_1^2 + x_2^2 - ax_1x_2$ on $\mathbf{R}^2$. For what values of the parameter a is this function convex?

b. Consider the function $f(x) = x^T(A + \alpha I)x$, where A is a general symmetric $n \times n$ matrix. For what values of α is this a convex function? (Your answer for the value of α should be in terms of quantities related to matrix A).

Problem 6. (log-sum-exp/softmax function.) Consider the function $f(z) = \log\left(\sum_{i=1}^n e^{z_i}\right)$ (also known as the *softmax* function in ML). Oftenm the convexity of this function is proven by deriving the Hessian. In this problem, you will show that f is convex using another method.

a. Show that, for any given $s > 0$, we have

$$\log s = -1 + \min_v \left(se^v - v\right).$$

b. Show that

$$f(z) = -1 + \min_v \left(\sum_{i=1}^n e^{z_i + v} - v\right).$$

c. Prove convexity of f based on the above result. *Hint:* Use the convexity-preserving property called "partial minimization", this is Module 4 lecture 2 slides, p. 16 (which we'll cover on Monday).

Problem 7. (Questions related to gradient descent notebook.) Refer to the Python notebook `M4-N1.ipynb` under module 4 on the course website, and answer the following questions about your observations. You do not need to turn in the notebook. You simply need to provide an answer to each of these questions.

a. Relating to stepsize:

   i. Describe what happens if the step size is too small? Try 0.0001.
   ii. Describe what happens if the step size is too big? Try 1.001.
   iii. What is a step size that converges in one step?
   iv. What is the Lipschitz constant (as defined in the notebook) of the gradient? Describe what happens if you set the step size to $1/\beta$ or $1/\beta \pm 0.1$?

b. Relating to initial guess:

   i. Trying different values of 'x0', describe the relationship you observe between the shape of the function, the initial guess, and the number of iterations.
   ii. Trying different values of $c \in (-2, 2)$, describe a relationship between the shape of the function, the initial guess, and the number of iterations.