Gradescope: To submit this homework assignment, use Gradescope accessed through Canvas. Please allow some time in advance of the deadline to become familiar with Gradescope.

Problem 1 (Norms, 5 points each) For any $\boldsymbol{x} \in \mathbb{R}^n$, let us introduce

$$\|\boldsymbol{x}\|_p \;=\; \left(\sum_{i=1}^{n} |x_i|^p \right)^{1/p}.$$

We let $\|\boldsymbol{x}\|_0 = \lim_{p \searrow 0} \|\boldsymbol{x}\|_p = \sum_{i=1}^n \mathbb{1}_{x_i \neq 0}$, where $\mathbb{1}_{\cdot}$ is an indicator function.

(a) Show that the for $p \geq 1$, the following inequality holds for any $\boldsymbol{x}, \boldsymbol{y}$ such that $\|\boldsymbol{x}\|_p = \|\boldsymbol{y}\|_p = 1$ and $\lambda \in [0, 1]$:

$$||\lambda \boldsymbol{x} + (1-\lambda)\boldsymbol{y}||_p \leq 1$$

(b) Show that the function $f(\boldsymbol{x}) = \|\boldsymbol{x}\|_p$ is a norm for $p \geq 1$.
(Hint: to show triangle inequality, use homogeneous property to reduce it to part a)

(c) Show that $\|\boldsymbol{x}\|_p$ is *not* a norm for $0 \leq p < 1$.
(Hint: it suffices to find counter examples)

(d) Show that for $1 \leq p < q$, we have $\|\boldsymbol{x}\|_p \geq \|\boldsymbol{x}\|_q$. For what $\boldsymbol{x}$ is equality obtained ($\|\boldsymbol{x}\|_p = \|\boldsymbol{x}\|_q$)?
(Hint: since both functions are homogeneous, it suffice to show the case when $\|\boldsymbol{x}\|_p = 1$)

Problem 2 (Linear Program, 5 points each)

(a) Given $\boldsymbol{A} \in \mathbb{R}^{m \times n}$ and $\boldsymbol{y} \in \mathbb{R}^m$, show that the basis pursuit problem

$$\min_{\boldsymbol{x}} \|\boldsymbol{x}\|_1, \quad \text{s.t.} \quad \boldsymbol{A}\boldsymbol{x} \;=\; \boldsymbol{y}.$$

is a convex problem, and recast it as a *linear program* in the standard form.

(b) Rewrite the following optimization problem in form that is solvable as a linear programming problem :

$$\min_{\boldsymbol{x} \in \mathbb{R}^n} \|\boldsymbol{A}\boldsymbol{x} - \boldsymbol{y}\|_\infty, \quad \text{s.t.} \quad \|\boldsymbol{x}\|_\infty \;\leq\; b.$$

where $\boldsymbol{A} \in \mathbb{R}^{m \times n}$ and $\boldsymbol{y} \in \mathbb{R}^m$ and $b > 0$.

Problem 3 (Global Optimality of Convex Problems, 5 points each) Consider the following unconstrained optimization problem

$$\min_{\boldsymbol{x} \in \mathbb{R}^n} \; f(\boldsymbol{x}),$$

where $f : \mathbb{R}^n \mapsto \mathbb{R}$ is a convex function.

(a) Show that if $\boldsymbol{x}_0$ is a local minimizer of f then it is also a global solution (need not to be unique). Moreover, if f is *strictly* convex, then show that the solution is *unique* (if exists).

(b) Furthermore, suppose $f \in \mathscr{C}^1$ (i.e., continuously differentiable). Then show that $\boldsymbol{x}_\star$ is a *global* solution of $f(\boldsymbol{x})$ if and only if

$$\nabla f(\boldsymbol{x}_\star) \;=\; \boldsymbol{0}.$$

Problem 4 (Operations that Preserves Convexity, 2.5 points each)

Given $f(\boldsymbol{x})$, $f_1(\boldsymbol{x}), \cdots, f_n(\boldsymbol{x})$ are convex function defined on $\mathbb{R}^n$ show that:

(a) (Nonnegative weighted sum)
If $\alpha_i \geq 0$ for all $i = 1, 2, \cdots, n$, then show that $g(\boldsymbol{x}) := \sum_{i=1}^n \alpha_i f_i(\boldsymbol{x})$ is a convex function.

(b) (Pointwise maximum)
$g(\boldsymbol{x}) := \max_{i=1,\ldots,n} \{f_1(\boldsymbol{x}), \cdots, f_n(\boldsymbol{x})\}$ is convex.

(c) (Composition with an affine mapping)
Given a matrix $\boldsymbol{A} \in \mathbb{R}^{n\times m}$ and a vector $\boldsymbol{b} \in \mathbb{R}^n$, show $g(\boldsymbol{x}) := f(\boldsymbol{A}\boldsymbol{x} + \boldsymbol{b})$ is convex with respect to $\boldsymbol{x} \in \mathbb{R}^m$.

(d) (Restriction to a line)
Fix some vector $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$, define $g : \mathbb{R} \to \mathbb{R}$ as $g(t) := f(\boldsymbol{x} + t\boldsymbol{y})$. Then g is convex with respect to t.

Problem 5 (Matrix Norms and Approximations, 5 points each)

(a) Show that the Schatten p-norm is *unitarily invariant* for any $p > 0$ even though it is *only* truly a norm for $p \geq 1$.

(b) (5 points) Let $\boldsymbol{A} \in \mathbb{R}^{m\times n}$ ($n > m$) with compact SVD $\boldsymbol{A} = \sum_{i=1}^m \sigma_i \boldsymbol{u}_i \boldsymbol{v}_i^\top$. For the rank-$r$ ($r < m$) approximation problem:

$$\min_{\boldsymbol{X}} \|\boldsymbol{X} - \boldsymbol{A}\|_F, \quad \text{s.t.} \quad \operatorname{rank}(\boldsymbol{X}) \leq r,$$

show that one optimal solution is the truncated SVD of the following form

$$\boldsymbol{A}_r = \sum_{i=1}^r \sigma_i \boldsymbol{u}_i \boldsymbol{v}_i^\top.$$

Moreover, show that the solution is unique whenever $\sigma_r > \sigma_{r+1}$.

Problem 6 (Taylor Approximation & Lipschitz Property, 5 points each) Let $\mathscr{S} \subset \mathbb{R}^n$ and $f : \mathscr{S} \mapsto \mathbb{R}$. Let $\boldsymbol{x} \in \mathscr{S}$ and $\boldsymbol{s} \in \mathbb{R}^n$, such that the line segment $[\boldsymbol{x}, \boldsymbol{x} + \boldsymbol{s}] \in \mathscr{S}$.

(a) Suppose $f \in \mathscr{C}^1$, and the gradient $g(\boldsymbol{x})$ of the function f is Lipschitz continuous with constant γ^L on $\mathscr{S}$. By defining $\phi(\alpha) = f(\boldsymbol{x} + \alpha\boldsymbol{s})$ and using the Fundamental Theorem of Calculus:

$$\phi(1) = \phi(0) + \int_0^1 \phi'(\alpha) d\alpha,$$

show that

$$|f(\boldsymbol{x} + \boldsymbol{s}) - f(\boldsymbol{x}) - \langle g(\boldsymbol{x}), \boldsymbol{s}\rangle| \leq \frac{\gamma^L}{2} \|\boldsymbol{s}\|_2^2.$$

(b) Suppose $f \in \mathscr{C}^2$ and the Hessian $H(\boldsymbol{x})$ of $f(\boldsymbol{x})$ is Lipschitz continuous with constant γ^Q on $\mathscr{S}$. Based on **(a)**, justify the following formula:

$$\phi(1) = \phi(0) + \phi'(0) + \int_0^1 \int_0^\alpha \phi''(t) dt d\alpha.$$

Using the formula above, prove the following:

$$\left| f(\boldsymbol{x} + \boldsymbol{s}) - f(\boldsymbol{x}) - \langle g(\boldsymbol{x}), \boldsymbol{s}\rangle - \frac{1}{2}\boldsymbol{s}^\top H(\boldsymbol{x})\boldsymbol{s} \right| \leq \frac{\gamma^Q}{6} \|\boldsymbol{s}\|_2^3.$$

(c) Consider the regularized least-squares cost function

$$h(\boldsymbol{x}) \;=\; \frac{1}{2}\,\|\boldsymbol{A}\boldsymbol{x}-\boldsymbol{y}\|_2^2 \;+\; \frac{\lambda}{2}\,\|\boldsymbol{T}\boldsymbol{x}\|_2^2\,,$$

where $\lambda > 0$. Find the Lipschitz constants for the gradient of $h(\boldsymbol{x})$, expressed in terms of the singular values (or spectral norm) of $\boldsymbol{A}$ and $\boldsymbol{T}$.

Problem 7 (Optimality, 5 points each)

(a) Define $f(x, y) = x^2y$, show that $(0, 0)$ is a stationary point of the function, but it's not the global minimizer.

(b) Define $f(x, y) = x^2 + y^2 + xy^2$, show that (0,0) satisfies the 2nd-order optimality condition, but it's not a global minimizer.

Problem 8 (Rate of Convergence, 5 points each)

Recommended reading:"rate of convergence" under appendix of the numerical optimization textbook by Nocedal and Wright

(a) Show that the sequence $x_k = 1 + \left(\frac{1}{2}\right)^{2^k}$ is Q-quadratically convergent to 1.

(b) Does the sequence $x_k = 1/k!$ converge Q-superlinearly? Q-quadratically?

(c) Consider the sequence $\{x_k\}$ defined by

$$x_k \;=\; \begin{cases} \left(\frac{1}{4}\right)^{2^k}, & k \text{ even,} \\ x_{k-1}/k, & k \text{ odd.} \end{cases}$$

Is this sequence Q-superlinearly convergent? Q-quadratically convergent? R-quadratically convergent?