Rules:

- Unless otherwise stated, all answers must be mathematically justified.
- Partial answers will be graded.
- Your submission has to be uploaded to Gradescope. In Gradescope, indicate the page on which each problem is written.
- You can work in groups but each student must write his/her/their own solution based on his/her/their own understanding of the problem. Please list on your submission the students you work with for the homework (this will not affect your grade).
- Problems with a ($\star$) are extra credit, they will not (directly) contribute to your score of this homework. However, for every 4 extra credit questions successfully answered your lowest homework score get replaced by a perfect score.
- If you have any questions, feel free to ask them on Ed Discussion (so that everyone can benefit from the answer) or stop at the office hours.

Problem 9.1 (2 points)**.** *Decide if there exists a function $f : \mathbb{R} \to \mathbb{R}$ that satisfies the following properties. If there exists such a function, construct an example and prove that the properties are satisfied; if not, prove that there is a contradiction.*

(a) *$f : \mathbb{R} \to \mathbb{R}$ is both convex and bounded, i.e., $\sup_{x\in\mathbb{R}} |f(x)| \leq C$ for some constant C.*

Consider the function $f(x) = 1, \forall x \in R$, then:
(i): Convexity: $\forall\theta \in [0, 1], x, y \in R, f(\theta x + (1-\theta)y) = 1$

$$\theta f(x) + (1-\theta)f(y) = \theta + 1 - \theta = 1$$
$$\therefore f(\theta x + (1-\theta)y) = \theta f(x) + (1-\theta)f(y)$$
$$\therefore f(\theta x + (1-\theta)y) \leq \theta f(x) + (1-\theta)f(y)$$

(ii): Boundedness:

$$\forall x \in R, \exists M \geqslant 1, \ st$$
$$\sup_{x\in R} |f(x)| = \sup\{1\} = 1 \leq M$$

$\therefore$ Take an arbitrary $M = 2$ would satisfy. $\therefore$ There exists a function that's convex and bounded. Actually, we could prove that any function that's convex and bounded is a constant function. ie. $f(x) = c \quad \forall x \in R$ for some constant $c \in R$. Proof: Suppose f is a convex and bounded function that's not. constant, then. $\exists x, y \in R$, sit $f(x) < f(y)$. $\because f$ is convex $\therefore \forall z \in R$, st. $x < y < z, f(z) \geqslant f(x) + (z-x)\frac{f(y)-f(x)}{y-x}$ This implies that f is unbounded. $\therefore f$ must be constant.

(b) *$f : \mathbb{R} \to \mathbb{R}$ is both strictly convex and bounded.*

Impossible. Suppose we have a function that's strictly convex and bounded, then by strict convexity we have: $\forall x, y, z \in R$, set $x < y < z, f(z) > f(x) + (z-x)\frac{f(y)-f(x)}{y-x}$ (1) which implies that f is unbounded. Since if $f(x)$ is bounded, then $f(x) \leq M, \forall x \in R$, which means $f(x) \leq M, f(y) \leq M, \quad f(z) \leq M$. But (1) means $\forall M > 0$, we can construct a z by setting $f(x) + (z-x)\frac{f(y)-f(x)}{y-x} = M$ and $z = \frac{M-f(x)(y-x)}{f(y)-f(x)} + x$, s.t. $f(z) > M$, which violates the boundedness.
Proof for inequality (1):
Proof. *Proof: Since f is strictly convex, $f(\theta x + (1-\theta)z) < f(x) + (1-\theta)f(z)$ let $y =$*

$\theta x + (1-\theta)z$, solve for θ we have $\theta = \frac{y-z}{x-z}$, then we have the following:

$$
\begin{aligned}
&f(y) < \frac{y-z}{x-z}f(x) + \frac{x-y}{x-z}f(z) \\
&(x-z)f(y) > (y-z)f(x) + (x-y)f(z) \\
&\frac{x-z}{x-y}f(y) - \frac{y-z}{x-y}f(x) < f(z) \\
&\frac{xf(y) - zf(y) - yf(x) + zf(x)}{x-y} < f(z) \\
&f(z) > \frac{xf(y) - yf(x) + xf(x) - xf(x) + z(f(x) - f(y)}{x-y} \\
&f(z) > \frac{x(f(y) - f(x)) + (x-y)f(x) + z(f(x) - f(y))}{x-y} \\
&f(z) > \frac{(z-x)(f(x) - f(y)}{x-y} + f(x) \\
&f(z) > f(x) + \frac{z-x}{y-x}(f(y) - f(x))
\end{aligned}
$$

□

Problem 9.2 (3 points). *Let $A \in \mathbb{R}^{n\times m}$ and $y \in \mathbb{R}^n$. For $x \in \mathbb{R}^m$ we define*

$$f(x) = \|Ax - y\|^2.$$

We will see a lot of this function when we discuss linear regression next lecture. Here we prove some useful properties of this function.

(**a**) *Compute the gradient $\nabla f(x)$.*

$$f(\vec{x}) : \vec{x}^\top A^\top A\vec{x} - 2\vec{b}^\top A\vec{x} + \|\vec{y}\|_2^2$$
$$\therefore \nabla f(\vec{x}) = \left(A^\top A + \left(A^\top A\right)^\top\right)\vec{x} - 2A^\top\vec{b} = 2A^\top A\vec{x} - 2A^\top\vec{b}$$

(**b**) *Compute the Hessian $H_f(x)$.*

$$\begin{aligned} H_f(\vec{x}) &= \nabla(\nabla f(\vec{x})) \\ &= \nabla\left(2A^\top A\vec{x} - 2A^\top\vec{b}\right) \\ &= 2A^\top A \end{aligned}$$

(**c**) *Show that f is convex.*

We will prove a stronger argument:
Claim: For any quadratic function $f : R^m \to R$, given by $f(\vec{x}) = \vec{x}^\top A\vec{x} + \vec{b}^\top\vec{x} + r$, where $A \in S^m, \vec{b} \in R^m, r \in R$

$$f(\vec{x})\textit{is convex} \Leftrightarrow A \succeq 0\,(A \in S_+^n)$$
$$f(\vec{x})\textit{is strictly convex} \Leftrightarrow A \succ 0\,(A \in S_{++}^n)$$

Proof for (1): *Since $f(\vec{x})$ is convex and twice differentible, by definition $H_f = A \geqslant 0$. conversely, if $A \geq 0$, then $h(\vec{x}) = \vec{x}^\top A\vec{x}$ is convex by definition. We now prove that $g(\bar{x}) = \vec{b}^\top\bar{x} + r$*

$$\forall\theta \in [0,1], \bar{x}, \vec{y} \in R^m,$$
$$\begin{aligned} g(\theta\vec{x} + (1-\theta)\vec{y}) = \vec{b}^\top(\theta\vec{x} + (1-\theta)\vec{y}) + r &= \theta\left(\vec{b}^-\vec{x} + r\right) + (1-\theta)(\vec{b}\vec{y} + r) \\ &\leqslant \theta g(\vec{x}) + (1-\theta)g(\vec{y}) \end{aligned}$$

Finally we prove that for $f_1, f_2, \cdots f_n : R^m \to R, \omega_i \geq 0\forall i$, the nonnegative weighted sum defined $l_y t(\vec{x}) = \sum_{i=1}^n \omega_i f_i(\vec{x})$ is convex.

$$\begin{aligned} &\forall\theta \in [0,1], \vec{x}, \vec{y} \in R^m \\ t(\theta\vec{x} + (1-\theta)\vec{y}) &= \sum_{i=1}^n \omega_i f_i(\theta\vec{x} + (1-\theta)\vec{y}) \\ &\leqslant \sum_{i=1}^n \omega_i\left(\theta f_i(\vec{x}) + (1-\theta)f_i(\vec{y})\right) \\ &= \theta\sum_{i=1}^m \omega_i f_i(\vec{x}) + (1-\theta)\sum_{i=1}^n \omega_i f_i(\vec{y}). \\ &= \theta t(\vec{x}) + (1-\theta)t(\vec{y}) \end{aligned}$$

Now, notice that our quadratic function is composed of three convex functions $h(\vec{x}) = \vec{x}^\top A\vec{x}, g(\vec{x}) = \vec{b}^\top\vec{x} + r$ $\therefore$ If $A \succeq 0$, $f(\vec{x}) = h(\vec{x}) + g(\vec{x})$ is convex If $f(\vec{x})$ is convex, then

$H_f(\bar{x}) = A \geq 0$.
Proof for (2): *The $\Leftarrow$ direction follows naturally from the above proof. To check the forward direction we just need to show If $h(\vec{x}) = \vec{x}^\top A\vec{x}$ is strictly convex, then $A \succ 0$. Assume $h(\vec{x})$ is strictly convex, then by definition.*

$$h(\vec{x}) > h(\vec{y}) + \nabla f(\vec{y})^\top (\vec{x} - \vec{y})$$

By second order Taylor expansion we have.

second order Taylor expansion we have.
$h(\vec{x}) = h(\vec{y}) + \nabla f(\vec{y})^\top (\vec{x} - \vec{y}) + \frac{1}{2}(\vec{x} - \vec{y})^\top \nabla^2 f(\vec{z})(\vec{x} - \vec{y})$ *for some* $\vec{z} = \theta\vec{x} + (1-2)\vec{y}$
$\therefore \frac{1}{2}(\vec{x} - \vec{y})^\top \nabla^2 f(\vec{z})(\vec{x} - \vec{y}) > 0, \forall \vec{x} \neq \vec{y}$
$\therefore \frac{1}{2}(\vec{x} - \vec{y})^\top A(\vec{x} - \vec{y}) > 0, \forall \vec{x} \neq \vec{y}$
$\therefore A > 0$ $\square$

$\therefore$ *Here* $f(\vec{x}) = \vec{x}^\top A^\top A\vec{x} - \vec{b}^\top A\vec{x} + \|\vec{y}\|_2^2$ *where* $A^\top A \in S_+^m$ *we can denote* $f(\bar{x}) = \vec{x}^\top S\vec{x} + \vec{p}^\prime\dot{x} + r$ *where* $S \in S_+^m, \vec{p} \in R^m, r \in R$*, which is a convex function.*

Problem 9.3 (3 points).

(**a**) *Show that the function $f(x) = \log(x)$ is concave, that is, $-f(x)$ is convex.*

Note that $\mathrm{dam}\, f = \{x \in R : x > 0\}$*, which is definitely convex. We then use second order condition where* $(-f(x))'' = \frac{1}{x^2} > 0 \forall x \in dom \therefore -f(x)$ *is convex* $\therefore f(x)$ *is concave.*

(**b**) *Use the concavity of* $\log$*, prove that for $p, q > 1, \frac{1}{p} + \frac{1}{q} = 1$, and any non-negative reals $x, y \in \mathbb{R}_+$, we have the inequality*

$$xy \leq \frac{x^p}{p} + \frac{y^q}{q}.$$

Hint: evaluate the logarithm of the above expression.

Given $x, y \in R_+$*, we have* $x^n, y^n > 0 \therefore x^n, y^n \in \mathrm{dom}\, f$. $\because f(x) = \log(x)$ *is concave* $\therefore f\left(\frac{1}{p}x^p + \frac{1}{q}y^q\right) \geq \frac{1}{p}f\left(x^p\right) + \frac{1}{q}f\left(y^q\right) \therefore \log\left(\frac{1}{p}x^p + \frac{1}{q}y^q\right) \geq \frac{1}{p}\log\left(x^p\right) + \frac{1}{q}\log\left(y^q\right)$ $\therefore \log\left(\frac{1}{p}x^p + \frac{1}{q}yq\right) \geq \log(xy) \because f'(x) = \frac{1}{x} > 0 \quad \forall x \in \mathrm{dom}\, f \therefore f(x)$ *is monotonically increasing* $\therefore xy \leq \frac{x^p}{p} + \frac{y^q}{q}$

(**c**) *With the above result, show that for $p, q > 1, \frac{1}{p} + \frac{1}{q} = 1$, and any $x, y \in \mathbb{R}^n$, we have*

$$\sum_{i=1}^{n} |x_i y_i| \leq \|x\|_p \|y\|_q,$$

where $\|x\|_p = (\sum_{i=1}^n |x_i|^p)^{1/p}$ is the ℓ_p norm.
Hint: apply the inequality in part (b) to the product $\frac{|x_i|}{\|x\|_p}\frac{|y_i|}{\|y\|_q}$.

We have the following:

$$\begin{aligned}
&\because xy \leq \frac{x^p}{p} + \frac{y^q}{q} \\
&\therefore \sum_{i=1}^{n} \frac{|x_i|}{\|x\|_p}\frac{|y_i|}{\|y\|_q} \leqslant \sum_{i=1}^{n} \frac{|x_i|^p}{\|x\|_p^p \cdot p} + \sum_{i=1}^{n} \frac{|y_i|^q}{\|y\|_q^q \cdot q} \\
&= \frac{\|x\|_p^p}{\|x\|_p^p \cdot p} + \frac{\|y\|_q^q}{\|y\|_q^q \cdot q} \\
&= \frac{1}{p} + \frac{1}{q} \\
&= 1 \\
&\therefore \sum_{i=1}^{n} |x_i|\,|y_i| \leqslant \|x\|_p \|y\|_q
\end{aligned}$$

Problem 9.4 (2 points)**.** *Let $f : \mathbb{R}^n \to \mathbb{R}$ be a convex function. Assume that the minimum $m \stackrel{\text{def}}{=} \min_{x\in\mathbb{R}^n} f(x)$ of f on $\mathbb{R}^n$ is finite, and that the set of minimizers of f*

$$\mathcal{M} \stackrel{\text{def}}{=} \{v \in \mathbb{R}^n \mid f(v) = m\}$$

is non-empty.

(a) *Show that $\mathcal{M}$ is a convex set.*

$\forall \vec{x}, \vec{y} \in M$, we have $f(\vec{x}) = m, f(\vec{y}) = m$. Since f is convex, $\forall\theta \in [0, 1]$, we have $f(\theta\vec{x}+(1-\theta)\vec{y}) \leqslant \theta f(\vec{x})+(1-\theta)f(\vec{y}) \therefore f(\theta\vec{x}+(1-\theta)\vec{y}) \leqslant \theta\cdot m+(1-\theta)\cdot m = m \because f(\theta\vec{x}+(1-\theta)\vec{y}) \geqslant m$ by definition of m.

$$\therefore f(\theta\vec{x} + (1-\theta)\vec{y}) = m$$
$$\therefore \theta\vec{x} + (1-\theta)\vec{y} \in M$$

$\therefore M$ is a convex set.

(b) *Show that if f is strictly convex, then $\mathcal{M}$ has only one element.*

Suppose $\exists\vec{x} \neq \vec{y} \in M$, we have $f'(\vec{x}) = f(\vec{y}) = m$. $\because f$ is strictly convex $\therefore \forall\theta \in [0, 1]$,

$$\begin{aligned} f(\theta\vec{x} + (1-\theta)\vec{y}) &< \theta f(\vec{x}) + (1-\theta)f(\vec{y}) \\ &= \theta\cdot m + (1-\theta)\cdot m \\ &= m \end{aligned}$$

$\because$ By definition: $f(\theta\vec{x}+(1-\theta)\vec{y}) \geqslant m$. $\therefore$ The above statement is a contradiction. $\therefore M$ has unique element in it.

Problem 9.5 (*). *Given positive definite matrix $A \in \mathbb{R}^{n\times n}$, consider the following non-negative cost function $f : \mathbb{R}^n \to \mathbb{R}$,*

$$f(x) = \|A - xx^\top\|_F^2, \quad x \in \mathbb{R}^n,$$

where the Frobenius norm of a matrix M is defined as $\|M\|_F = \sqrt{\mathrm{Tr}(M^\top M)}$.
Prove or disprove: f is convex.

Proof.

$$\begin{aligned}
f(\vec{x}) &= \mathrm{Tr}\left[\left(A - \vec{x}\vec{x}^\top\right)^\top \left(A - \vec{x}\vec{x}^\top\right)\right] \\
&= \mathrm{Tr}\left[A^\top A - A\vec{x}\vec{x}^\top - \vec{x}^\top A + \vec{x}\vec{x}^\top \|\vec{x}\|_2^2\right] \\
&= \mathrm{Tr}\left(A^\top A\right) - \mathrm{Tr}\left(A\vec{x}\vec{x}^\top\right) - \mathrm{Tr}\left(\vec{x}\vec{x}^\top A\right) + \|\vec{x}\|_2^2 \,\mathrm{Tr}\left(\vec{x}\vec{x}^\top\right) \\
&= \|A\|_F^2 - 2\,\mathrm{Tr}\left(A\vec{x}\vec{x}^\top\right) + \|\vec{x}\|_2^2 \cdot \|\vec{x}\|_2^2 \\
&= \|A\|_F^2 - 2\,\mathrm{Tr}\left(\vec{x}^\top A\vec{x}\right) + \|\vec{x}\|_2^4 \\
&= \|\vec{x}\|_2^4 - 2\vec{x}^\top A\vec{x} + \|A\|_F^2
\end{aligned}$$

For $h(\vec{x}) = \|\vec{x}\|_2^*$, we compute its hessian.

$$\begin{aligned}
\nabla h(\vec{x}) = &\frac{\partial\left(\vec{x}^T\vec{x}\right)^2}{\partial \vec{x}} = 2\left(\vec{x}^T\vec{x}\right)\cdot 2\vec{x} \\
&= 4\left(\vec{x}^\top\vec{x}\right)\vec{x} \\
\nabla^2 h(\vec{x}) &= 4\left(\vec{x}^\top\vec{x}\right) I = 4\|\vec{x}\|_2^2 \cdot I \\
\therefore \nabla^2 f(\vec{x}) &= 4\|\vec{x}\|_2^2 \cdot I - 4A \\
&= 4\left(\|\vec{x}\|_2^2 \cdot I - A\right) \\
\therefore \lambda_i\left(\nabla^2 f(\vec{x})\right) &= 4\left(\|\vec{x}\|_2^2 - \lambda_i(A)\right)
\end{aligned}$$

Since the smallest eigenvalues depend on the choice of $\vec{x}$, we cannot grantee that $\nabla^2 h(\vec{x}) \succeq 0 \quad \forall \vec{x} \in R^n$ $\therefore$ f isn't convex A counterexample would be letting

$$\begin{aligned}
&A = [5] \quad \vec{x} = [-2] \quad \vec{y} = [0] \quad \theta = 0.5 \\
&f(\vec{x}) = \left\|[5] - [-2][-2]^\top\right\|_F^2 = 1 \\
&f(\vec{y}) = \left\|[5] - [0][0]^\top\right\|_F^2 = 25 \\
&f(\theta\vec{x} + (1-\theta)\vec{y}) = \left\|[5] - [-1][-1]^\top\right\|_F^2 = 16. \\
&\theta f(\vec{x}) + (1-\theta) f(\vec{y}) = \frac{25+1}{2} = 13 \\
&\therefore \theta f(\vec{x}) + (1-\theta) f(\vec{y}) < f(\theta\vec{x} + (1-\theta)\vec{y})
\end{aligned}$$

Which violates the convexity. □