EECS 559 HW 1

QUESTION 1:

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

$$\min_{\boldsymbol{x}\in\mathbb{R}^n} \|\boldsymbol{x}\|_1 \ \ \textbf{s.t.} \ \ \boldsymbol{Ax} = \boldsymbol{y}$$

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

SOLUTION: Firstly, we can show that this is a convex optimization problem by showing that both the objective function and constraints are convex. We start by examining the objective function:

$$\begin{aligned}\|\boldsymbol{x}\|_1 &= \sum_i |x_i| \\ &= \sum_i \max\{x_i, -x_i\}\end{aligned}$$

We know that x_i and $-x_i$ are linear functions, therefore convex, thus the pointwise maximum of these two functions is convex. Since the objective function is a nonnegative sum of these pointwise maxima, it is also convex. The constraint:

$$\boldsymbol{Ax} = \boldsymbol{y}$$

is linear, therefore convex. which means that this *is a convex optimization problem.*

We can now recast this as a linear program in the standard form. We start by defining $\boldsymbol{x}_+$ and $\boldsymbol{x}_-$ as follows:

$$\begin{aligned}\boldsymbol{x}_+ &\triangleq \sum_i x_i \mathbb{1}_{\{x_i\geq 0\}} \boldsymbol{e}_i \\ \boldsymbol{x}_- &\triangleq \sum_i x_i \mathbb{1}_{\{x_i< 0\}} \boldsymbol{e}_i\end{aligned}$$

We can define:

$$\boldsymbol{z} \triangleq \begin{bmatrix}\boldsymbol{x}_+ \\ \boldsymbol{x}_-\end{bmatrix}$$

And clearly:

$$\begin{aligned}\|\boldsymbol{x}\|_1 &= \sum_{i=0}^{n} |x_i| \\ &= \sum_{i=0}^{n} x_i \mathbb{1}_{\{x_i\geq 0\}} + \sum_{i=0}^{n} x_i \mathbb{1}_{\{x_i< 0\}} \\ &= \boldsymbol{1}_n^T \boldsymbol{x}_+ + \boldsymbol{1}_n^T \boldsymbol{x}_- \\ &= \begin{bmatrix}\boldsymbol{1}_n^T & \boldsymbol{1}_n^T\end{bmatrix} \boldsymbol{z}\end{aligned}$$

We can now recast the problem as:

$$\min_{\boldsymbol{x}\in\mathbb{R}^n} \boldsymbol{c}^T\boldsymbol{z} \;\; \textbf{s.t.} \;\; \boldsymbol{a}_i^T\boldsymbol{z} = \boldsymbol{b}_i \;\; \forall i \in [n]$$

Where:

$$\boldsymbol{c} \triangleq \begin{bmatrix} \mathbf{1}_n \\ \mathbf{1}_n \end{bmatrix}, \;\; \boldsymbol{a}_i \triangleq \boldsymbol{M}_{[i,:]}, \;\; \boldsymbol{M} \triangleq \begin{bmatrix} \boldsymbol{A} & -\boldsymbol{a} \end{bmatrix}$$

QUESTION 2: 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).

PROOF: (By contradiction) Let $\boldsymbol{x}_0$ be a local minimizer of f, meaning that $\exists\epsilon > 0$ such that:

$$f(\boldsymbol{x}_0) \leq f(\boldsymbol{x}) \;\; \forall \boldsymbol{x} \in \mathcal{B}(\boldsymbol{x}_0, \epsilon) \triangleq \{\boldsymbol{x} \in \mathbb{R}^n \mid \|\boldsymbol{x}_0 - \boldsymbol{x}\|_2 \leq \epsilon\}$$

Assume $\exists \boldsymbol{x}_* \in \mathbb{R}^n$ such that $f(\boldsymbol{x}_*) < f(\boldsymbol{x}_0)$. $\forall\epsilon$ sufficiently small, $\exists\alpha \in (0, 1)$ such that:

$$\boldsymbol{y} \triangleq \alpha\boldsymbol{x}_0 + (1-\alpha)\boldsymbol{x}_* \in \mathcal{B}(\boldsymbol{x}_0, \epsilon)$$

And by the convexity of f:

$$f(\boldsymbol{y}) \leq \alpha f(\boldsymbol{x}_0) + (1-\alpha)f(\boldsymbol{x}_*)$$

Since $f(\boldsymbol{x}_*) < f(\boldsymbol{x}_0)$, we can write:

$$f(\boldsymbol{y}) < \alpha f(\boldsymbol{x}_0) + (1-\alpha)f(\boldsymbol{x}_0) = f(\boldsymbol{x}_0)$$

Which is when we arrive at our contradiction. Since for any arbitrary $\epsilon > 0$ there is an element $\boldsymbol{y} \in \mathcal{B}(\boldsymbol{x}_0, \epsilon)$ such that $f(\boldsymbol{x}_0) > f(\boldsymbol{y})$, our assumption is false. Therefore, any local minimum is a global minimum. $\square$

Let's now assume that f is *strictly convex*. Assuming it exists, we can show the uniqueness of the global minimizer as follows.

PROOF: (By contradiction) Let $\boldsymbol{x}_0$ is a local minimizer of f. Since strict complexity implies convexity, we know that $\boldsymbol{x}_0$ is also a global minimizer (from the previous argument).

$$f(\boldsymbol{x}_0) \leq f(\boldsymbol{x}) \;\; \forall \boldsymbol{x} \in \mathbb{R}^n$$

Assume $\exists \boldsymbol{x}_* \in \mathbb{R}^n$ such that $f(\boldsymbol{x}_0) = f(\boldsymbol{x}_*)$. By strict convexity of f, $\forall\alpha \in (0, 1)$:

$$f(\alpha\boldsymbol{x}_0 + (1-\alpha)\boldsymbol{x}_*) < \alpha f(\boldsymbol{x}_0) + (1-\alpha)f(\boldsymbol{x}_*) = f(\boldsymbol{x}_0)$$

Therefore there is an element in the domain that achieves a lower value than our global minimizer, which is our contradiction. Therefore our assumption is wrong. Given that f is *strictly convex* the local minimizer $\boldsymbol{x}_0$ is a *unique global* minimizer. $\square$

QUESTION 2: Continued

(b) Furthermore, suppose $f \in \mathcal{C}^1$ (i.e., continuously differentiable). Then show that $\boldsymbol{x}_0$ is a *global* minimizer of f iff:

$$\nabla f(\boldsymbol{x}_0) = \mathbf{0}$$

PROOF: ($\Rightarrow$) (By contradiction) Let $\boldsymbol{x}_0$ be a global minimum of f, we assume $\nabla f(\boldsymbol{x}_0) \neq \mathbf{0}$. Since f is convex, it satisfies the first order condition, $\forall \boldsymbol{y} \in \mathbb{R}^n$:

$$f(\boldsymbol{y}) \geq f(\boldsymbol{x}_0) + (\boldsymbol{y} - \boldsymbol{x}_0)^T \nabla f(\boldsymbol{x}_0)$$

We can specifically choose some arbitrary $\boldsymbol{y}$:

$$\boldsymbol{y}(t) = \boldsymbol{x}_0 + t\nabla f(\boldsymbol{x}_0) \ \ \forall t \in \mathbb{R}$$

We apply this to the first order condition to get:

$$f(\boldsymbol{x}_0 - t\nabla f(\boldsymbol{x}_0)) \geq f(\boldsymbol{x}_0) + t\left\|\nabla f(\boldsymbol{x}_0)\right\|^2$$

$\exists t$ s.t.:

$$f(\boldsymbol{x}_0 - t\nabla f(\boldsymbol{x}_0)) \leq f(\boldsymbol{x}_0) + t\left\|\nabla f(\boldsymbol{x}_0)\right\|^2$$

Unless $\left\|\nabla f(\boldsymbol{x}_0)\right\|^2 = 0$ which is possible iff $\nabla f(\boldsymbol{x}_0) = \mathbf{0}$

PROOF: ($\Leftarrow$) Assume $\nabla f(\boldsymbol{x}_0) = \mathbf{0}$. Since f is convex, it satisfies the first order condition, $\forall \boldsymbol{y} \in \mathbb{R}^n$:

$$\begin{aligned} f(\boldsymbol{y}) &\geq f(\boldsymbol{x}_0) + (\boldsymbol{y} - \boldsymbol{x}_0)^T \nabla f(\boldsymbol{x}_0) \\ &= f(\boldsymbol{x}_0) + (\boldsymbol{y} - \boldsymbol{x}_0)^T \mathbf{0} \\ &= f(\boldsymbol{x}_0) \end{aligned}$$

So $\forall \boldsymbol{y} \in \mathbb{R}^n$, $\nabla f(\boldsymbol{x}_0) = \mathbf{0}$ implies that $\boldsymbol{x}_0$ is a global minimum of f.

QUESTION 3: Given $f(\boldsymbol{x}), f_1(\boldsymbol{x}), f_2(\boldsymbol{x}), \ldots, f_m(\boldsymbol{x})$ are convex on $\mathbb{R}^m$, show that:

(a) If $\alpha_i \geq 0 \;\; \forall i \in [n]$, then $g(\boldsymbol{x}) \triangleq \sum \alpha_i f_i(\boldsymbol{x})$ is a convex function.

SOLUTION: Consider some $\beta \in [0,1]$ and $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$. By the confexity of f_i:

$$\beta f_i(\boldsymbol{x}) + (1-\beta) f_i(\boldsymbol{y}) \geq f_i(\beta \boldsymbol{x} + (1-\beta)\boldsymbol{y})$$

Which implies:

$$\begin{aligned} \sum \alpha_i \left[\beta f_i(\boldsymbol{x}) + (1-\beta) f_i(\boldsymbol{y})\right] &\geq \sum \alpha_i f_i(\beta \boldsymbol{x} + (1-\beta)\boldsymbol{y}) \\ \beta \sum \alpha_i f_i(\boldsymbol{x}) + (1-\beta) \sum \alpha_i f_i(\boldsymbol{y}) &\geq g(\beta \boldsymbol{x} + (1-\beta)\boldsymbol{y}) \\ \beta g(\boldsymbol{x}) + (1-\beta) g(\boldsymbol{y}) &\geq g(\beta \boldsymbol{x} + (1-\beta)\boldsymbol{y}) \;\; \forall \beta \in [0,1] \;\; \forall \boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n \end{aligned}$$

Therefore g is a convex function.

(b) $g(\boldsymbol{x}) \triangleq \max\{f_1(\boldsymbol{x}), \ldots, f_n(\boldsymbol{x})\}$ is convex:

SOLUTION: Consider some $\beta \in [0,1]$ and $\boldsymbol{x}_1, \boldsymbol{x}_2 \in \mathbb{R}^n$.

$$\begin{aligned} f_i(\beta \boldsymbol{x}_1 + (1-\beta)\boldsymbol{x}_2) &\leq \beta f_i(\boldsymbol{x}_1) + (1-\beta) f_i(\boldsymbol{x}_2) \\ \max_i f_i(\beta \boldsymbol{x}_1 + (1-\beta)\boldsymbol{x}_2) &\leq \max_i \left[\beta f_i(\boldsymbol{x}_1) + (1-\beta) f_i(\boldsymbol{x}_2)\right] \end{aligned}$$

Notice that $\forall a_i, b_i \in \mathbb{R}$, $\max\{a_i + b_i\} \leq \max\{a_i\} + \max\{b_i\}$. Therefore:

$$\begin{aligned} \max_i f_i(\beta \boldsymbol{x}_1 + (1-\beta)\boldsymbol{x}_2) &\leq \beta \max_i f_i(\boldsymbol{x}_1) + (1-\beta) \max_i f_i(\boldsymbol{x}_2) \\ g(\beta \boldsymbol{x}_1 + (1-\beta)\boldsymbol{x}_2) &\leq \beta g(\boldsymbol{x}_1) + (1-\beta) g(\boldsymbol{x}_2) \end{aligned}$$

Therefore the convexity of f implies the convexity of g.

(c) Given $\boldsymbol{A} \in \mathbb{R}^{n \times m}$, $\boldsymbol{x} \in \mathbb{R}^m$ and $\boldsymbol{b} \in \mathbb{R}^n$, $g(\boldsymbol{x}) \triangleq f(\boldsymbol{A}\boldsymbol{x} + \boldsymbol{b})$ is convex.

SOLUTION: Consider $\boldsymbol{x}_1$, $\boldsymbol{x}_2 \in \mathbb{R}^m$, and $\alpha \in [0,1]$

$$f(\boldsymbol{A}(\alpha \boldsymbol{x}_1 + (1-\alpha)\boldsymbol{x}_2) + \boldsymbol{b}) = f(\alpha(\boldsymbol{A}\boldsymbol{x}_1 + \boldsymbol{b}) + (1-\alpha)(\boldsymbol{A}\boldsymbol{x}_2 + \boldsymbol{b}))$$

By the convexity of f:

$$\begin{aligned} f(\boldsymbol{A}(\alpha \boldsymbol{x}_1 + (1-\alpha)\boldsymbol{x}_2) + \boldsymbol{b}) &\leq \alpha f(\boldsymbol{A}\boldsymbol{x}_1 + \boldsymbol{b}) + (1-\alpha) f(\boldsymbol{A}\boldsymbol{x}_2 + \boldsymbol{b}) \\ g(\alpha \boldsymbol{x}_1 + (1-\alpha)\boldsymbol{x}_2) &\leq \alpha g(\boldsymbol{x}_1) + (1-\alpha) g(\boldsymbol{x}_2) \end{aligned}$$

Therefore the convexity of f implies the convexity of g.

QUESTION 4: Show that $f : \mathbb{R}^n \mapsto \mathbb{R}$ is convex iff for all integers $m \geq 2$:

$$f\left(\sum_{i=1}^{m} \lambda_i \boldsymbol{x}_i\right) \leq \sum_{i=1}^{m} \lambda_i f(\boldsymbol{x}_i) \tag{$\star$}$$

Where $x_i \in \mathbb{R}^n$ and $\lambda_i \geq 0\ \forall i \in [m]$ satisfying $\sum \lambda_i = 1$.

PROOF: (by induction) Consider the base case ($m = 2$). The expression above boils down to the following:

$$f(\lambda_1 \boldsymbol{x}_1 + \lambda_2 \boldsymbol{x}_2) \leq \lambda_1 f(\boldsymbol{x}_1) + \lambda_2 f(\boldsymbol{x}_2)$$

Notice that $\lambda_1 + \lambda_2 = 1$ is equivalent to $\lambda_2 = 1 - \lambda_1$. And since $\lambda_2 \geq 0$ must be true, we know that $\lambda_1 \in [0, 1]$. After substituting we get the definition of convexity:

$$f(\lambda_1 \boldsymbol{x}_1 + (1 - \lambda_1)\boldsymbol{x}_2) \leq \lambda_1 f(\boldsymbol{x}_1) + (1 - \lambda_1) f(\boldsymbol{x}_2)$$

So the base case ($m = 2$) the inequality ($\star$) is true iff f is convex.

Now let's assume ($\star$) is true for some fixed value ($m = k \geq 2$):

$$f\left(\sum_{i=1}^{k} \lambda_i \boldsymbol{x}_i\right) \leq \sum_{i=1}^{k} \lambda_i f(\boldsymbol{x}_i)$$

W.l.o.g., we can express $\boldsymbol{x}_k$ as follows:

$$\boldsymbol{x}_k = \alpha \boldsymbol{y}_k + (1 - \alpha)\boldsymbol{y}_{k+1} \ \text{ where } \ \alpha \in [0, 1]$$

Where $\boldsymbol{y}_k$ and $\boldsymbol{y}_{k+1}$ are arbitrary vectors in $\mathbb{R}^n$. We can rewrite ($\star$) as follows:

$$f\left(\sum_{i=1}^{k-1} \lambda_i \boldsymbol{x}_i + \lambda_k \alpha \boldsymbol{y}_k + \lambda_k (1 - \alpha)\boldsymbol{y}_{k+1}\right) \leq \sum_{i=1}^{k-1} \lambda_i f(\boldsymbol{x}_i) + \lambda_k f(\alpha \boldsymbol{y}_k + (1 - \alpha)\boldsymbol{y}_{k+1})$$

And by the convexity of f:

$$\lambda_k f(\alpha \boldsymbol{y}_k + (1 - \alpha)\boldsymbol{y}_{k+1}) \leq \lambda_k \alpha f(\boldsymbol{y}_k) + \lambda_k (1 - \alpha) f(\boldsymbol{y}_{k+1})$$

Which leaves us with:

$$f\left(\sum_{i=1}^{k-1} \lambda_i \boldsymbol{x}_i + \lambda_k \alpha \boldsymbol{y}_k + \lambda_k (1 - \alpha)\boldsymbol{y}_{k+1}\right) \leq \sum_{i=1}^{k-1} \lambda_i f(\boldsymbol{x}_i) + \lambda_k \alpha f(\boldsymbol{y}_k) + \lambda_k (1 - \alpha) f(\boldsymbol{y}_{k+1})$$

And since $\sum_{i=1}^{k-1} \lambda_i + \lambda_k(\alpha + 1 - \alpha) = 1$, the $m = k$ case being true implies that the $m = k + 1$ case is aslo true. Therefor by mathematical induction, $\forall m \geq 2$, $\forall i \in [m]$, $\forall \lambda_i \geq 0$, $\forall \boldsymbol{x}_i \in \mathbb{R}^n$:

$$f : \mathbb{R}^n \mapsto \mathbb{R} \text{ convex} \quad \Longleftrightarrow \quad f\left(\sum_{i=1}^{m} \lambda_i \boldsymbol{x}_i\right) \leq \sum_{i=1}^{m} \lambda_i f(\boldsymbol{x}_i) \ \text{ where } \ \sum \lambda_i = 1$$

$\square$

QUESTION 5:

(a) Show that the sequence $x_k = 1 + \left(\frac{1}{2}\right)^{2^k}$ is Q-quadratically convergent to 1.
SOLUTION: We know that $\lim_{k \mapsto \inf} x_k = 1$, consider:

$$\begin{aligned}
\left\| 1 + \left(\frac{1}{2}\right)^{2^k} - 1 \right\|_2 &\leq \gamma \left\| 1 + \left(\frac{1}{2}\right)^{2^{k-1}} - 1 \right\|_2^2 \\
\left(\frac{1}{2}\right)^{2^k} &\leq \gamma \left[\left(\frac{1}{2}\right)^{2^{k-1}} \right]^2 \\
\left(\frac{1}{2}\right)^{2^k} &\leq \gamma \left(\frac{1}{2}\right)^{2^k}
\end{aligned}$$

Which is true $\forall \gamma \leq 1$, $\forall k \geq 1$, therefore the sequence $\{x_k\}$ is Q-quadratically convergent.

(b) Does the sequence $x_k = \frac{1}{k!}$ converge Q-superlinearly? Q-quadratically?
SOLUTION: We know that $\lim_{k \mapsto \inf} x_k = 0$, consider:

$$\begin{aligned}
\left\| \frac{1}{k!} \right\|_2 &\leq \gamma \left\| \frac{1}{(k-1)!} \right\|_2^p \\
\frac{1}{k!} &\leq \gamma \left[\frac{1}{(k-1)!} \right]^p \\
\frac{((k-1)!)^{p-1}}{k} &\leq \gamma \\
(p-1)\log((k-1)!) - \log(k) &\leq \log(\gamma) \\
(p-1)\sum_{i=1}^{k-1} \log(i) - \log(k) &\leq \log(\gamma)
\end{aligned}$$

The sequence above is clearly only upper bounded if $p - 1 \leq 0$, therefore $\{x_k\}$ is Q-convergent with $p \leq 1$. So the sequence is neither Q-quadratically nor Q-superlinearly convergent.

QUESTION 5: Continued

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

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

Is this sequence Q-superlinearly convergent? Q-quadratically convergent? R-quadratically convergent? Justify your answer.

SOLUTION: We know that $\lim_{k \mapsto \inf} x_k = 0$, consider the case where k is odd:

$$\begin{aligned} \|x_k\|_2 &\leq \gamma \|x_{k-1}\|_2^p \\ \frac{x_{k-1}}{k} &\leq \gamma x_{k-1}^p \\ \frac{1}{k}\left(\frac{1}{4}\right)^{(1-p)2^k} &\leq \gamma \end{aligned}$$

There can only be such γ if $p \leq 1$, therefore $\{x_k\}$ is neither Q-quadratically nor Q-superlinearly convergent. Now consider the sequence $\{\rho_k\}$ defined as follows:

$$\rho_k = \left(\frac{1}{4}\right)^{2^k} \quad \forall k \in \mathbb{N}$$

Notice that:

$$\begin{aligned} \rho_k &\geq x_k \ \ \forall k \in \mathbb{N} \\ \rho_k &\geq \|x_k - 0\|_2 \end{aligned}$$

We know from (a) that $\{\rho_k\}$ is Q-quadratically convergent, therefore $\{x_k\}$ is R-quadratically convergent.