1. Consider the following guessing game to find a number in $[0, 1]$: for each guess, the distance to that number is given, but (possibly adversarially) perturbed by 10%. Show that we can get within $\epsilon$ of the true answer in $O(1/\log(1/\epsilon))$ guesses.

**SOLUTION:**
Consider making a guess $x$ and receiving a distance $r$. We know that the target number is within the interval $[x + .9r, x + 1.1r]$ or $[x - 1.1r, x - .9r]$.

We then guess $x + r$, and if the received distance is less than $r$, then the number must be within $[x + .9r, x + 1.1r]$. Otherwise, it is within $[x - 1.1r, x - .9r]$.

Note that we can use iterative refinement with a precision of $\epsilon$, guessing twice at each step, resulting in a complexity of $O(\log(1/\epsilon))$. 

This problem set has a total of 20 written problems on 5 pages. Written solutions should be submitted to Canvas by 2pm Friday Mar 19, 2021.

• If you find the sources of these problems, a citation of the source is sufficient for full points.

• If you choose not to submit a typed write-up, please write neat and legibly.
2. Same as problem 1, but in three dimensions, and distances can be perturbed by a factor of 10. (note that 10 and 3 are both constants: your big-O notation can hide constants depending on them).

**SOLUTION:**
We can use a similar approach as Problem 1, by decreasing $\delta$ (such as setting $\delta = .1$) as the distances are now perturbed by a factor of 10.

Note that we can change the algorithm to operate in 3 dimensions by alternating the dimension that is refined ($x, y, z, x, y, z, \ldots$), resulting in a complexity of $O(\log(1/\epsilon))$. Dimensionality and $\delta$ are hidden within the constants of this complexity.

3. For each positive integer $n$, find the maximum sum of an arithmetic progression $a_1, a_2, \ldots a_{2n+1}$ ($a_i = b + ci$ for some $b$ and $c$) such that $a_1^2 + a_{n+1}^2 \leq 1$.

**SOLUTION:**
First consider the sum of $a_1, \ldots, a_{2n+1}$

$$\frac{(2n + 1)(a_1 + a_{2n+1})}{2} = (2n + 1)a_{n+1}$$

where $a_{n+1}^2 \leq a_1^2 + a_{n+1}^2 \leq 1$. Notice that this means that $a_{n+1} \leq 1$. Thus,

$$\frac{(2n + 1)(a_1 + a_{2n+1})}{2} = (2n + 1)a_{n+1} \leq 2n + 1$$

and equality can be achieved is $a_1 = 0 \rightarrow a_{n+1} = 1$.

4. Prove that for any $a_1 \ldots a_n, b_1 \ldots b_n \in [1, 2]$ such that

$$\sum_i a_i^2 = \sum_i b_i^2$$

we have

$$\sum_i \frac{a_i^3}{b_i} \leq \frac{17}{10} \sum_i b_i^2.$$ 

**HINT:** the ‘book’ solution for this problem involves showing $ab \geq \frac{2}{5}(a^2 + b^2)$ for any $a, b \in [1, 2]$.

**SOLUTION:**
Notice that $(\frac{a}{b} - 2)(\frac{a}{b} - \frac{1}{2}) \leq 0$ as $\frac{1}{2} \leq \frac{a}{b} \leq 2$. Thus, we have
\[
\frac{a_i^2}{b_i^2} - 5 \frac{a_i}{b_i} + 1 \leq 0 \\
\frac{a_i^2}{b_i^2} + 1 \leq 5 \frac{a_i}{b_i} \\
\frac{a_i}{b_i} + \frac{b_i}{a_i} \leq \frac{5}{2} \\
\frac{a_i^3}{b_i} \leq \frac{5}{2} a_i^2 - a_i b_i
\]

By rearranging terms from the final line, we also have \( \frac{2}{5}(a_i^2 + b_i^2) \leq a_i b_i \). It follows that

\[
\frac{a_i^3}{b_i} \leq \frac{5}{2} a_i^2 - a_i b_i \\
\frac{a_i^3}{b_i} \leq \frac{5}{2} a_i^2 - \frac{2}{5}(a_i^2 + b_i^2) \\
\frac{a_i^3}{b_i} \leq \frac{21}{10} a_i^2 - \frac{2}{5} b_i^2 \\
\sum_i a_i^3 \quad \leq \sum_i \frac{21}{10} a_i^2 - \sum_i \frac{2}{5} b_i^2 \\
\sum_i a_i^3 \quad \leq \sum_i \frac{17}{10} b_i^2
\]

5. For non-negative reals \( a_1 \ldots a_n \) with sum 1, find the maximum value of

\[
\sum_i (a_i^4 - a_i^5).
\]

**SOLUTION:**

Note that when \( n = 1 \), the maximum value is 0.

Now we consider the case where \( n = 2 \). Define \( f(x) = x^4 - x^5 + (1-x)^4 - (1-x)^5 \). Finding the critical points, we have \( f \) is monotonically increasing between \([0, \frac{1}{2} - \frac{\sqrt{3}}{6}]\) and \([\frac{1}{2}, \frac{1}{2} + \frac{\sqrt{3}}{6}]\) as well as monotonically decreasing between \([\frac{1}{2} - \frac{\sqrt{3}}{6}, \frac{1}{2}]\) and \([\frac{1}{2} + \frac{\sqrt{3}}{6}, 1]\).
We thus can find the maximum sum when \( n = 2 \), \( \frac{1}{12} \).

Now we need to prove that \( \frac{1}{12} \) is the maximum sum for all \( n \geq 2 \). Suppose that this is true for \( n = k \).

Note that we can show

\[
x^4 - x^5 + y^4 - y^5 \leq (x + y)^4 - (x + y)^5
\]
\[
5x^4y + 10x^3y^2 + 10x^2y^3 + 5xy^4 \leq 4x^3y + 6x^2y^2 + 4xy^3
\]
\[
5x^3 + 10x^2y + 10xy^2 + 5y^3 \leq 4x^2 + 6xy + 4y^2
\]
\[
5(x + y)^3 - 5xy(x + y) \leq 4(x + y)^2 - 2xy
\]

as the last inequality is true for \( 0 \leq x + y \leq \frac{4}{5} \).

Now, without loss of generality, let us assign \( a_1 \geq a_2 \ldots \geq a_k \geq a_{k+1} \), and thus \( 0 \leq a_k + a_{k+1} \leq \frac{2}{k+1} \leq \frac{4}{5} \) (as we are dealing with \( k \geq 2 \)). Thus, \( f(a_1 \ldots a_k, a_{k+1}) = \sum_{i=1}^{k-1} a_i^4 - a_k^5 + a_{k+1}^4 - a_{k+1}^5 \leq f(a_1 \ldots a_k + a_{k+1}) = \sum_{i=1}^{k-1} (a_i + a_{k+1})^4 - (a_k + a_{k+1})^5 \), and we have proven the induction step.

6. Prove that for any real numbers \( a_1 \ldots a_n \), each at least 1, we have

\[
\sum_i \frac{1}{a_i + 1} \geq \frac{n}{(\prod a_i)^{1/n} + 1}.
\]

**SOLUTION:**

Consider the function \( f(x) = \frac{1}{e^x + 1} \) for \( x \geq 0 \). Notice that \( f(x) \) is convex as

\[
f''(x) = \frac{e^x(e^x - 1)}{(e^x + 1)^3} \geq 0
\]

By Jensen’s inequality

\[
\frac{1}{n} \sum_i f(\log a_i) \geq f\left(\frac{1}{n} \sum \log a_i\right)
\]
\[
\frac{1}{n} \sum_i \frac{1}{a_i + 1} \geq \frac{1}{(\prod a_i)^{1/n} + 1}
\]
\[
\sum_i \frac{1}{a_i + 1} \geq \frac{n}{(\prod a_i)^{1/n} + 1}
\]
7. Prove for any $n$ and any $3n^2$ variables $f_{ij}, g_{ij}, h_{ij}$ (for $1 \leq i, j \leq n$) we have

$$\left( \sum_{i,j,k} f_{ij}g_{jk}h_{ki} \right)^2 \leq \left( \sum_{ij} f_{ij}^2 \right) \left( \sum_{ij} g_{ij}^2 \right) \left( \sum_{ij} h_{ij}^2 \right)$$

**SOLUTION:**

Applying Cauchy on the entries $f_{ij}$ and $\sum_k g_{ik}h_{kj}$ gives

$$\left( \sum_{i,j,k} f_{ij}g_{jk}h_{ki} \right)^2 \leq \left( \sum_{ij} f_{ij}^2 \right) \left( \sum_{ik} \left( \sum_k g_{jk}h_{ki} \right)^2 \right)$$

So it suffices to show that the terms inside the brackets is at most the squared Frobenius norms of $g$ and $h$. Expanding gives

$$\sum_{ij} \left( \sum_k g_{jk}h_{ki} \right)^2 = \sum_{ijk_1k_2} g_{jk_1}h_{k_1i}g_{jk_2}h_{k_2i} \leq \frac{1}{2} \left( \sum_{ijk_1k_2} g_{jk_1}^2 h_{k_2i}^2 + h_{k_1i}^2 g_{jk_2}^2 \right)$$

Now $k_1$ and $k_2$ are independent, both products factor exactly into $(\sum_{ik_1} g_{ik_1}^2)(\sum_{jk_2} h_{jk_2}^2)$.

The next three problems are about generalizations of iterative methods to non-linear settings. As we will work with multi-variate functions, we will use the $\vec{x}$ notation to denote vectors.

The most important definition here is the Bregman divergence, which is the difference against the gradient extrapolation:

$$D_f(\vec{x}_1, \vec{x}_2) := f(\vec{x}_1) - f(\vec{x}_2) - \nabla f(\vec{x}_2)^\top (\vec{x}_1 - \vec{x}_2).$$

Note that for a convex function, the Bregman divergence is always non-negative.

8. Consider the single variate function $f(x) = x^4$. Prove that its Bregman divergence is bounded (within constants) by $x^2\delta^2 + \delta^4$, that is

$$0.01 \left( x^2\delta^2 + \delta^4 \right) \leq D_{f(x) = x^4}(x + \delta, x) \leq 100 \left( x^2\delta^2 + \delta^4 \right).$$

**SOLUTION:**

Notice that
\[ D_{f(x)} = x^4(x + \delta, x) = (x + \delta)^4 - x^4 - 4x^3\delta = 6x^2\delta^2 + 4x\delta^3 + \delta^4 \]

Thus, we note that since by AM-GM
\[ 5x^2\delta^2 + \frac{4}{5}\delta^4 \geq 2\sqrt{4x^2\delta^6} = 4|x\delta^3| \]

we have
\[ -5x^2\delta^2 - \frac{4}{5}\delta^4 \leq 4x\delta^3 \leq 5x^2\delta^2 + \frac{4}{5}\delta^4 \]
\[ x^2\delta^2 + \frac{1}{5}\delta^4 \leq 6x^2\delta^2 + 4x\delta^3 + \delta^4 \leq 11x^2\delta^2 + \frac{9}{5}\delta^4 \]
\[ .01(x^2\delta^2 + \delta^4) \leq x^2\delta^2 + \frac{1}{5}\delta^4 \leq 6x^2\delta^2 + 4x\delta^3 + \delta^4 \leq 11x^2\delta^2 + \frac{9}{5}\delta^4 \leq 100(x^2\delta^2 + \delta^4) \]

9. Recall the (preconditioned) iterative methods for solving linear systems discussed in class. Suppose for the convex function \( f \) and current value \( \bar{x} \) (possibly high dimensional), there is another function \( g \) such that:

(a) the gradient is exactly preserved: \( \nabla g(\bar{x}) = \nabla f(\bar{x}) \),

(b) the Bregman divergence is well approximated for any \( \Delta \):
\[ \alpha^{-1}D_f(\bar{x} + \Delta, \bar{x}) \leq D_g(\bar{x} + \Delta, \bar{x}) \leq \alpha D_f(\bar{x} + \Delta, \bar{x}) \quad \forall \Delta, \]

and let
\[ \Delta^* = \arg \min_{\Delta} g(\bar{x} + \Delta), \]

show that we can use \( \bar{x} \) and \( \Delta^* \) to obtain another vector \( \bar{y} \) such that
\[ f(\bar{y}) - OPT \leq \left( 1 - \frac{1}{100 \cdot \alpha^{16}} \right) (f(\bar{x}) - OPT), \]

where \( OPT \) is the minimum of \( f \).

**SOLUTION:**

We can use iterative refinement to approximate a \( \bar{y} \) that meets the conditions.
Note that since $\nabla g(\vec{x}) = \nabla f(\vec{x})$, we have

$$\vec{\Delta}^* = \arg\min_{\Delta} g\left(\vec{x} + \Delta\right) \implies \nabla g(\vec{x} + \vec{\Delta}^*) = 0 \implies \nabla f(\vec{x} + \vec{\Delta}^*) = 0$$

Thus, the Bregman divergence is given by

$$D_f\left(\vec{x}, \vec{x} + \vec{\Delta}^*\right) = f(\vec{x}) - f(\vec{x} + \vec{\Delta}^*) = f(\vec{x}) - \text{OPT}$$

Starting with $\vec{y} = \vec{x}$ and iterating $\vec{y}$, we have

$$\vec{y} \leftarrow \vec{y} + \vec{\Delta}^*$$

$$f(\vec{y}) \leftarrow f(\vec{y} + \vec{\Delta}^*) = f(\vec{y}) - (f(\vec{y}) - f(\vec{y} + \vec{\Delta}^*)) = f(\vec{y}) - D_g\left(\vec{y}, \vec{y} + \vec{\Delta}^*\right)$$

where $f(\vec{y})$ becomes an approximation of $\text{OPT}$.

10. Show that given an algorithm that computes 2-approximations to $\min_{\vec{x}}$:

$$A\vec{x} = \vec{b} \sum_i \vec{r}_i \vec{x}_i^2 + \vec{s}_i \vec{x}_i^4$$

for any matrix $A$, vector $\vec{b}$, and weights $\vec{r}$, $\vec{s}$, we can call this algorithm $O(\log(1/\epsilon))$ times (plus overheads of $O(m)$) to obtain an $(1 + \epsilon)$-approximate solution to any 4-norm minimization problem.

**SOLUTION:**
Choosing $\vec{r}_i = 0$ and $\vec{s}_i = 1$ allows the algorithm to find 2-approximations of $\min_{\vec{x}: A\vec{x} = \vec{b}} \sum \vec{x}_i^4$. Note that we can set $A$ to be the same matrix as in the target 4-norm minimization problem.

Thus, we use iterative refinement on

$$\vec{x} \leftarrow \vec{x} + \delta \text{ALG}(\vec{b} - A\vec{x})$$

where $\text{ALG}$ is the given algorithm, to find a $(1 + \epsilon)$-approximate solution in $O(\log(1/\epsilon))$ time.

The next three problems use the iterative methods above to solve $p$-norm regression problems, which have the format

$$\min_{\vec{x}: A\vec{x} = b} \|\vec{x}\|_p$$

where $\|\vec{x}\|_p = (\sum_i |x_i|^p)^{1/p}$. 

7
11. By incorporating Holder’s inequality, given an algorithm for solving 4-norm minimization to relative error $1/poly(n)$ by solving $\tilde{O}(n^{1/4})$ systems of linear equations.

12. Show that for any $p \geq 5$, any $x$, and any $\delta$, we have

$$D_{|.|^p} \left( x - \frac{\delta}{100p}, x \right) \leq 10 \cdot D_{|.|^p} (x + \delta, x).$$

13. For any $p \geq 2$, give a $p$-norm regression algorithm that solves $\tilde{O}(pn^{1/2})$ linear systems. Note here $p$ is treated as a non-constant parameter.

14. Define $F(a, b, c)$ to the max magnitude of the monoic cubic polynomial with coefficients $a, b,$ and $c$ on the range $[1, 3]$:

$$F(a, b, c) := \max_{1 \leq x \leq 3} \left| x^3 + ax^2 + bx + c \right|$$

find the minimum of $F(a, b, c)$ over all reals $a, b,$ and $c$.

**SOLUTION:**

Note that we intuitively want to find a pair of $a, b$ that minimizes the distance between $f_{\text{max}}$ and $f_{\text{min}}$, setting $c = \frac{f_{\text{max}} + f_{\text{min}}}{2}$.

Thus, we need $x = 2$ to be the inflection point of the cubic function, and as a result, $a = -6$. Similarly, $f_{\text{max}} = f\left(\frac{6 - \sqrt{36 - 36}}{3}\right) = f(3)$.

Solving for $b$ and $c$, we get $F(-6, \frac{45}{4}, -\frac{13}{2}) = \frac{1}{4}$ as the minimum.

15. Find the minimum and maximum of

$$\cos x \cdot \sin y \cdot \cos z$$

over all $x \geq y \geq z \geq \frac{\pi}{12}$ and $x + y + z = \frac{\pi}{2}$.

**SOLUTION:**

As $\cos z = \sin(x + y)$, we are solving for $f(x, y) = \cos x \sin y \sin(x + y)$.

We note that the max and the min must either lie at stationary points of the function $f$ or at the bounds of the domain.

Let us first check the stationary points. We have that $f'_x = f'_y = 0$ and thus

$$f'_x - f'_y = (\cos x \cos y + \sin x \sin y) \sin(x + y) = \cos(x - y) \sin(x + y) = 0$$

As this has no solutions within the given domain, we move on to check the bounds of the domain.
• Consider the case in which \( z = \frac{\pi}{12} \). We then are maximizing
\[
g(x) = \cos \frac{\pi}{12} \cos x \sin \left( \frac{5\pi}{12} - x \right)
\]
over \( \frac{5\pi}{24} \leq x \leq \frac{\pi}{3} \). We find the maximum at \((x, g(x)) = \left( \frac{5\pi}{24}, \frac{2 + \sqrt{3}}{8} \right)\) and the minimum at \((x, g(x)) = \left( \frac{\pi}{3}, \frac{1}{8} \right)\).

• Consider the case in which \( z = \frac{\pi}{6} \). We then have \( x = y = z = \frac{\pi}{6} \) and
\[
g(x) = \cos \frac{\pi}{6} \cos \frac{\pi}{6} \sin \frac{\pi}{6} = \frac{3}{8}
\]
Therefore, the minimum of the function is \( \frac{1}{8} \) at \((x, y, z) = \left( \frac{\pi}{3}, \frac{\pi}{12}, \frac{\pi}{12} \right)\) and the maximum is \( \frac{2 + \sqrt{3}}{8} \) at \((x, y, z) = \left( \frac{5\pi}{24}, \frac{5\pi}{24}, \frac{\pi}{12} \right)\).

16. Show that for any 13 real numbers \( x_1 \ldots x_{13} \), there are two numbers \( x_i \) and \( x_j \) such that
\[
0 < \frac{x_i - x_j}{1 + x_i x_j} < \sqrt{\frac{2 - \sqrt{3}}{2 + \sqrt{3}}}.
\]

**SOLUTION:**

Notice that
\[
\sqrt{\frac{2 - \sqrt{3}}{2 + \sqrt{3}}} = \sqrt{\left( \frac{2 - \sqrt{3}}{2 + \sqrt{3}} \right)^2} = 2 - \sqrt{3} = \tan 15^\circ
\]

We need to show that an inequality exists for some pair of real numbers when we are given 13. This hints us towards pigeonhole principle, however the current domain of \( x_i \) is \(( -\infty, \infty )\) which is very large. We can reduce this domain by transforming it in terms of the \( \tan \) function. We rewrite our 13 real numbers in terms of the angles \( \theta_1, \theta_2, \ldots, \theta_{13} \) where \( \tan \theta_i = x_i \). We can now rewrite the desired inequality as:
\[
\frac{x_i - x_j}{1 + x_i x_j} = \frac{\tan \theta_i - \tan \theta_j}{1 + \tan \theta_i \tan \theta_j} = \tan (\theta_i - \theta_j) < \tan 15^\circ
\]

Since \( \tan \theta \in (-\infty, \infty) \) for \( \theta \in [0^\circ, 180^\circ] \), we can restrict the values of our angles to be \( \theta_i \in [0^\circ, 180^\circ] \). Now we can divide this region \([0^\circ, 180^\circ]\) into 12 parts of 15\(^\circ\) (i.e. the 12 regions \([15a, 15(a + 1)]\) for \( a \in [0, 11] \)). By the pigeonhole principle, 2 of the 13 angles will fall in the same hole, which implies that there exists \( i \) and \( j \) such that \( \theta_i - \theta_j < 15^\circ \). Thus, we have shown the inequality.
17. Use multiplicative weights updates, and the fact that $s$-$t$ shortest paths can be computed in $\tilde{O}(m)$ time (via Dijkstra’s algorithm), give an algorithm for approximating min-cost flows to $m^{-10}$ additive error in a graph with $m$ edges, costs and capacities in the range $[1, m^{10}]$ in $\tilde{O}(m^2)$ time.

**SOLUTION:**

18. Suppose we have access to some linear operator for producing approximate solutions to 4-norm minimization. That is, for any vector $b$, the solution $x \leftarrow Rb$ satisfies

(a) feasibility: $ARb = b$,
(b) small objective: $\|Rb\|_4 \leq 100 \cdot \min_{x: A x = b} \|x\|_4$

Show that we can compute an $1/poly(n)$ relative error solution to $\min_{x: A x = b} \|x\|_2$ by evaluating $R$ on inputs chosen by the algorithm $\tilde{O}(n^{1/2})$ times, plus operations whose total cost is $\tilde{O}(n^{3/2})$.

**OPEN**, but unintendedly. The originally intended solution required adjusting weights on the $\delta_i^4$ terms based on the current value of $x$.

19. Show that the following iterative reweighted least squares algorithm (https://en.wikipedia.org/wiki/Iteratively_reweighted_least_squares) for $\ell_4$ norm regression ($\min_{x: A x = b} \|x\|_4$) converges to $1/poly(n)$ relative error in $\tilde{O}(n^{1/4})$ iterations (for some choices of $\alpha$ and $\beta$ that you can compute adaptively):

| (a) Initialize $x \leftarrow \arg \min_{A x = b} \|x\|_2$. |
| (b) Repeat |
| i. Compute $w_e = x_e^2 + \alpha \|x\|_2$. |
| ii. Compute $\delta \leftarrow \arg \min_{A \delta = b} \sum_i w_e x_e^2$. |
| iii. $x \leftarrow x + \beta \delta$. |

Note the algorithm is different than the more general version in that it’s not allowed to generate a new gradient at each iteration.

**SOLUTION:**

20. Show an iterated reweighted least squares algorithm for 4-norm regression whose iteration count is $\tilde{O}(m^{0.24})$ or better.

**OPEN**