Near-Linear Sample Complexity for L_p Polynomial Regression

Raphael A. Meyer

Cameron Musco

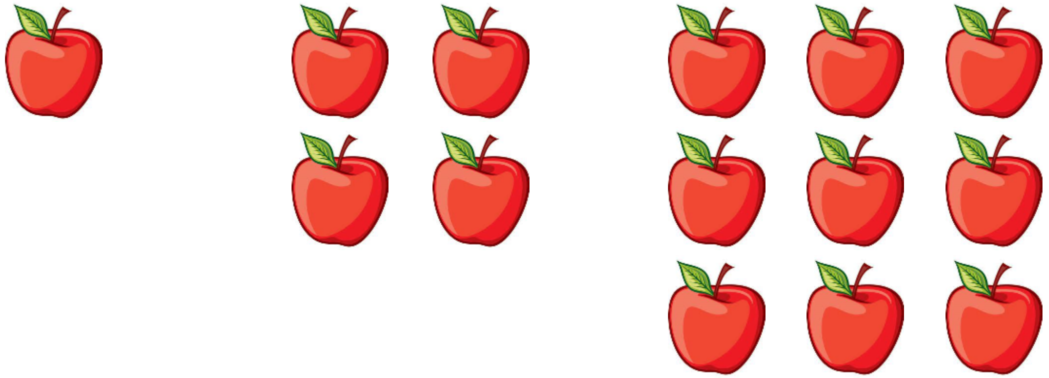
Christopher Musco

David P. Woodruff

Samson Zhou

Over 50% People CAN'T Solve This!!

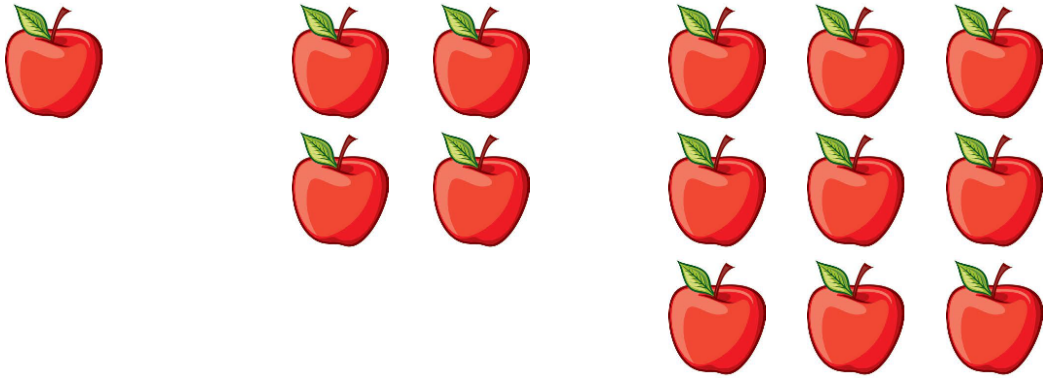
What comes next?



? ?

Over 50% People CAN'T Solve This!!

What comes next?

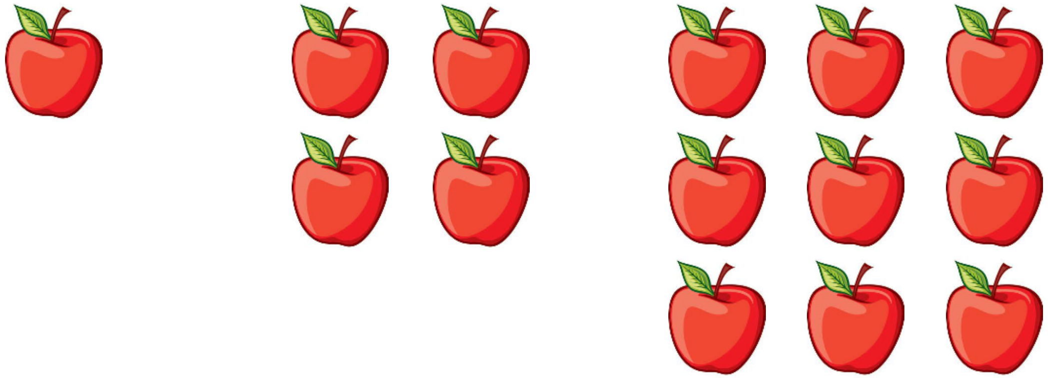


?   ?

1   4   9

Over 50% People CAN'T Solve This!!

What comes next?

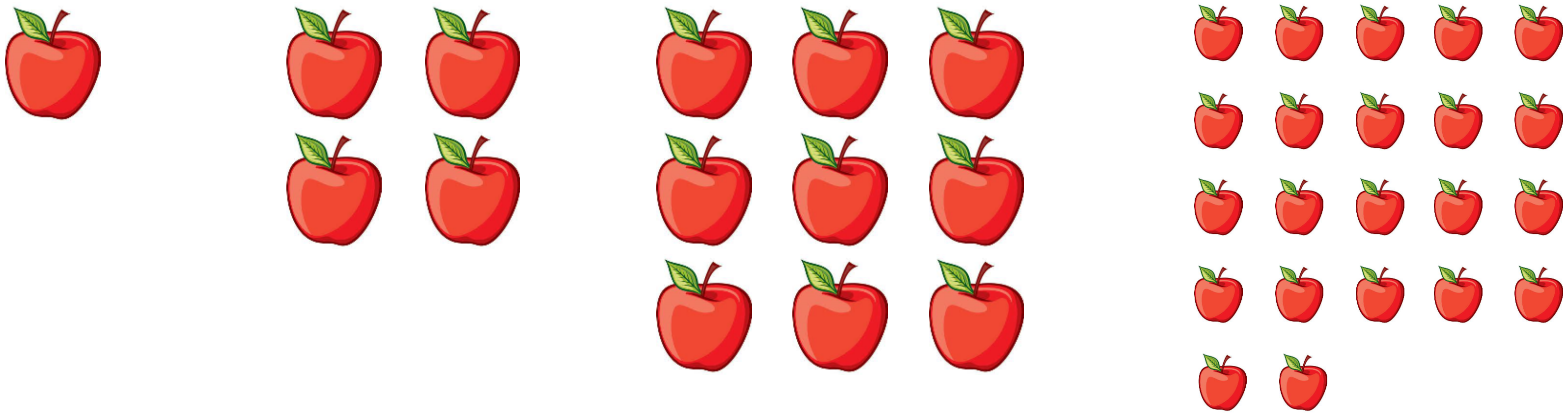


? ?

1 4 9 16 25

Over 50% People CAN'T Solve This!!

What comes next?

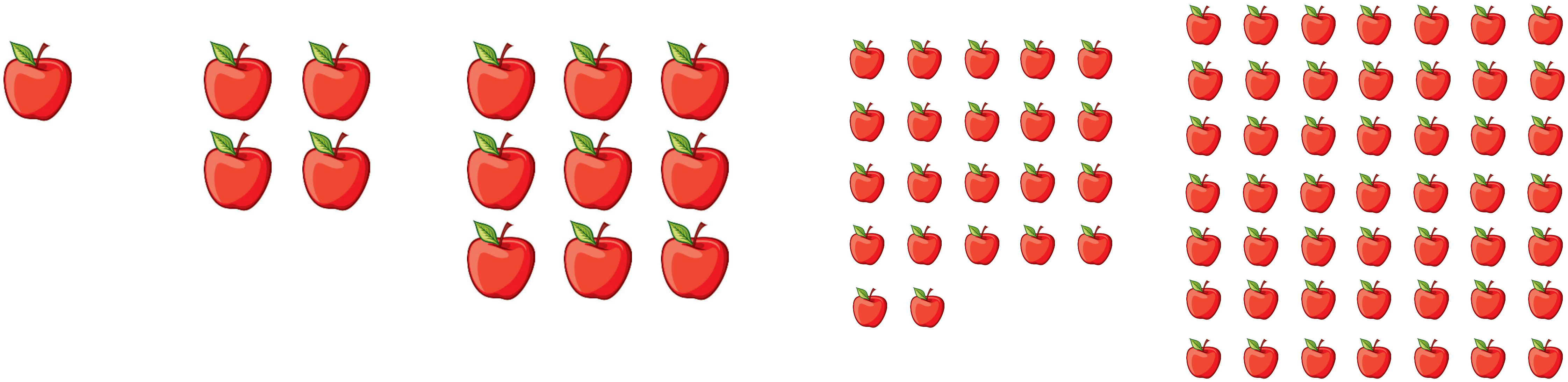


?

1 4 9 22

Over 50% People CAN'T Solve This!!

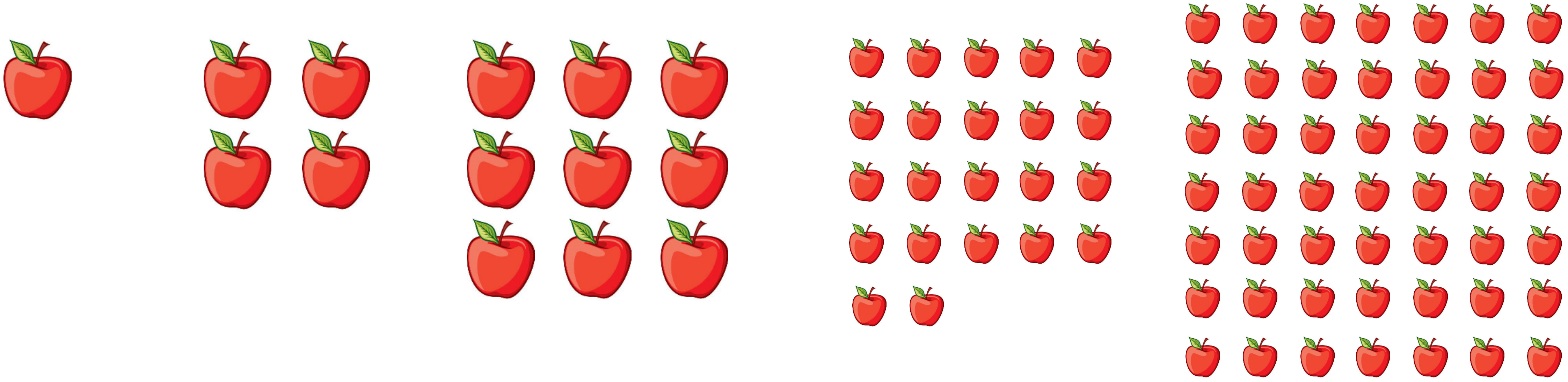
What comes next?



1 4 9 22 49

Over 50% People CAN'T Solve This!!

What comes next?



1 4 9 22 49

$$q(x) = x^3 - 5x^2 + 11x - 6$$

Polynomial Fitting

- Given $q(t_1), \dots, q(t_m)$, recover the polynomial $q(x)$

- For a degree d polynomial $q(x)$, must have $m \geq d + 1$ samples to recover $q(x)$

Polynomial Fitting

$$q(x) = a_d x^d + \cdots + a_1 x + a_0$$

$$\begin{gathered} q(t_1) = a_d {t_1}^d + \cdots + a_1 t_1 + a_0 \\ q(t_2) = a_d {t_2}^d + \cdots + a_1 t_2 + a_0 \\ \vdots \\ q(t_m) = a_d {t_m}^d + \cdots + a_1 t_m + a_0 \end{gathered}$$

Polynomial Fitting

- For $m \geq d + 1$, any choice of distinct $t_1, \ldots, t_m$ can recover $q(x)$

- Solve the linear system, Lagrangian interpolation, etc.

Polynomial Regression

- For a signal f, recover the degree d polynomial $q(x)$ that is the "best fit" to f

- What does best fit mean?

Polynomial Regression

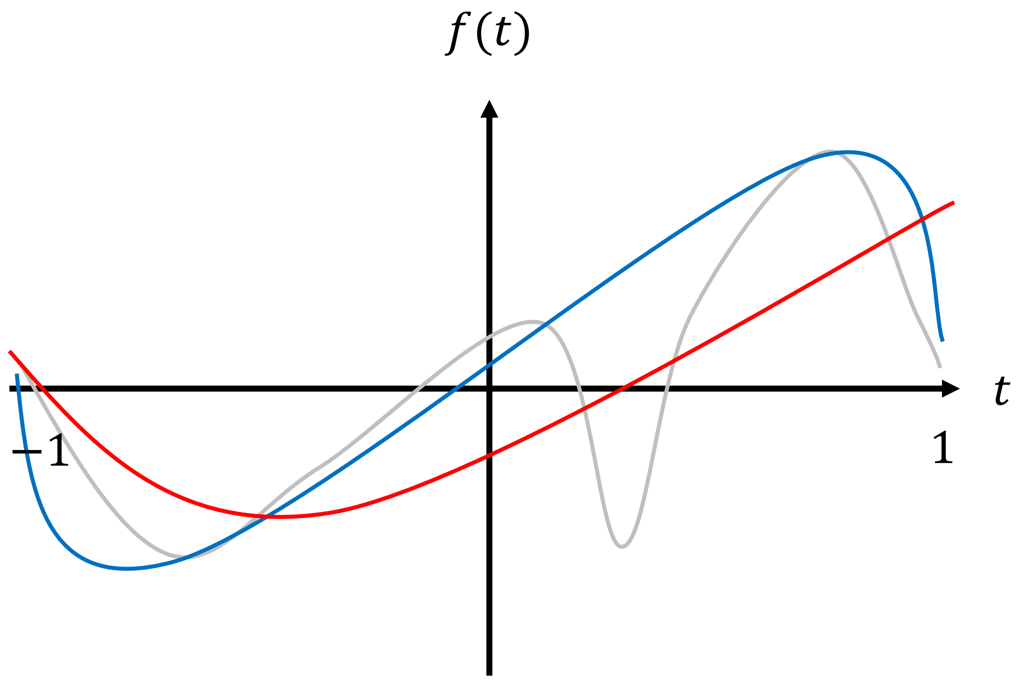
- $\|f-q\|_p = \left(\int_{-1}^{1} |f(t)-q(t)|^p \, dt\right)^{1/p}$
- $\|f-q\|_\infty = \max_{t\in[-1,1]} |f(t)-q(t)|$

- Polynomial regression: Given $\varepsilon > 0$ and $p \in [1,\infty]$, output $\widehat{q(t)}$ such that

$$\|f-\hat{q}\|_p \leq (1+\varepsilon)\left(\min_{\deg(q)\leq d} \|f-q\|_p\right)$$

Polynomial Regression

- $\|f-q\|_p = \left(\int_{-1}^{1} |f(t)-q(t)|^p \, dt\right)^{1/p}$
- $\|f-q\|_\infty = \max_{t\in[-1,1]} |f(t)-q(t)|$



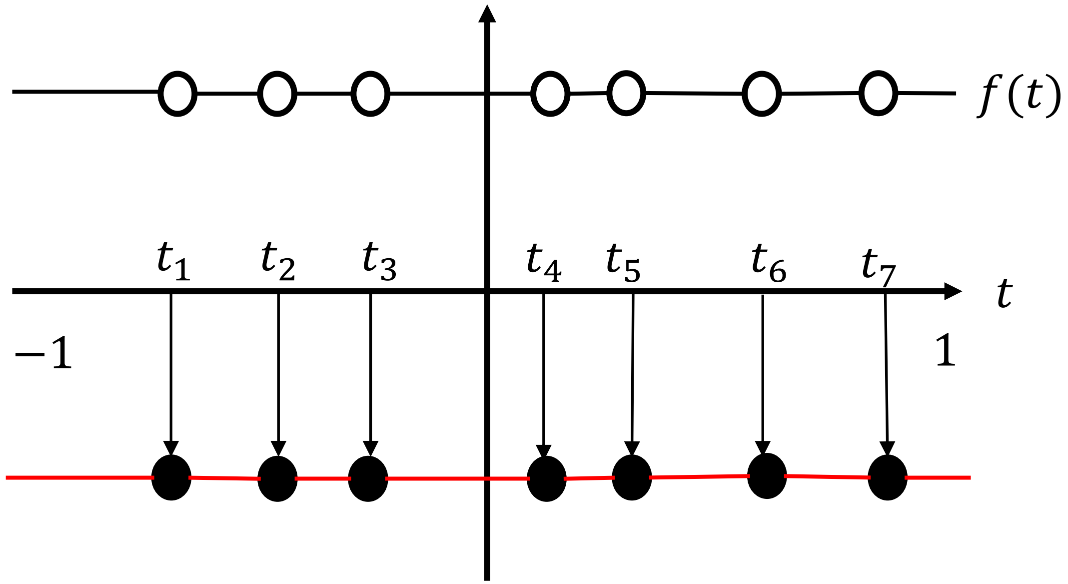
Sample Complexity

- **Sample complexity:** Number m of locations $t_1, \dots, t_m$ at which the signal f is read

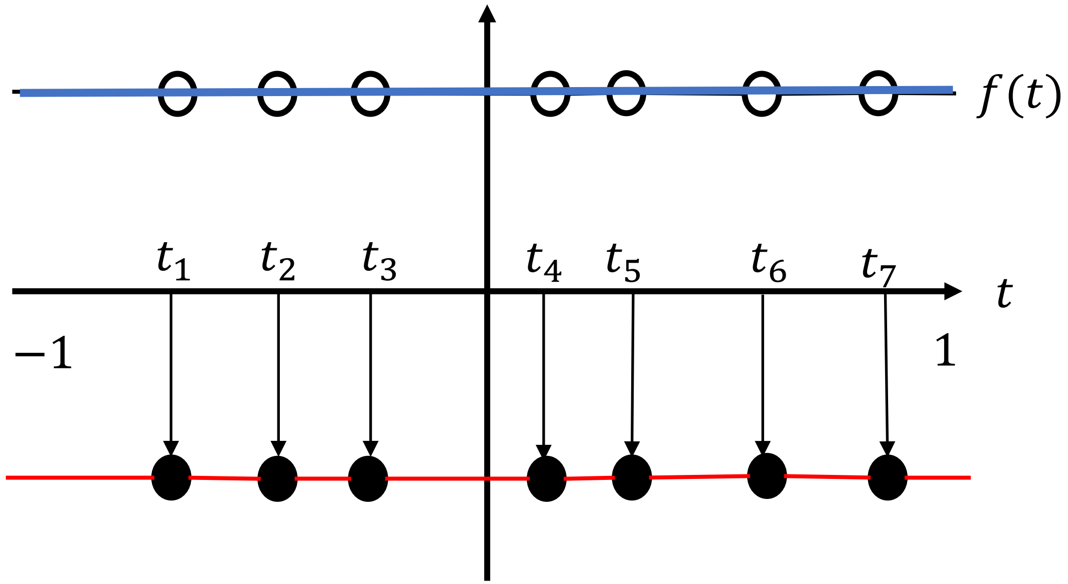
- Sample complexity of polynomial fitting is $m = d + 1$

- What is the sample complexity of polynomial regression?

Deterministic Algorithms Do Not Work



Deterministic Algorithms Do Not Work



Previous Work for L_2 Regression

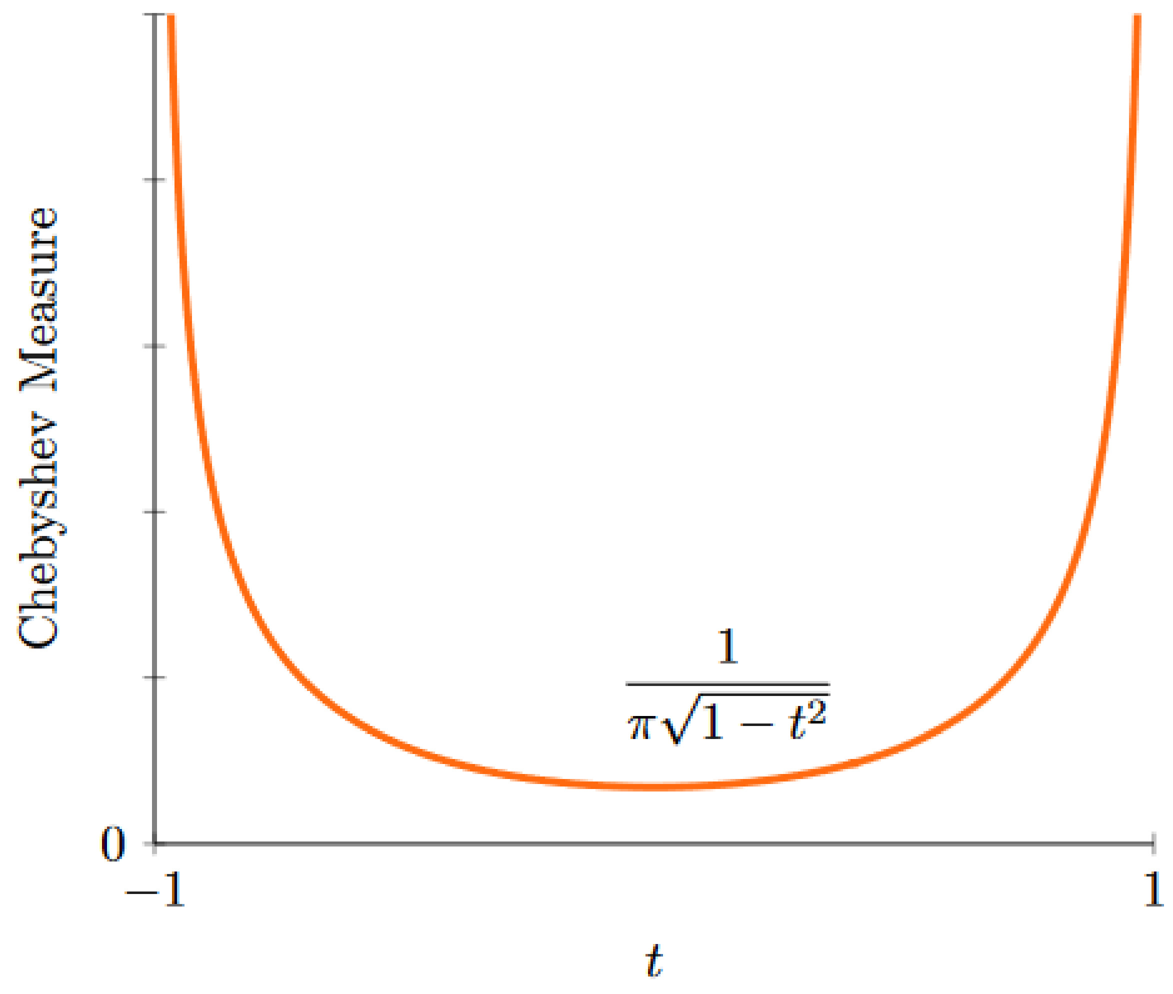
- $(1+\varepsilon)$-approximation to L_2 regression with $O\left(\frac{d \log d}{\varepsilon}\right)$ queries [RauhutWard12, CohenDavenportLeviatan13, CohenMigliorati13]

- $(1+\varepsilon)$-approximation to L_2 regression with $O\left(\frac{d}{\varepsilon}\right)$ queries [ChenPrice19]

Previous Work for L_∞ Regression

- $O(\log d)$-approximation to L_∞ regression with $O(d \log d)$ queries [Trefethen12]

- Constant factor approximation to L_∞ regression with $O(d \log d)$ queries [KaneKarmalkarPrice17]


Chebyshev Measure
$\frac{1}{\pi\sqrt{1-t^2}}$
0
−1
1
t

Our Results (I)

- $(1+\varepsilon)$-approximation to L_p regression with $dp\left(\frac{\log^{O(p)} d}{\varepsilon^{O(p)}}\right)$ queries from the Chebyshev density for all $p \geq 1$

- Upper bound shows separation in the degree d between polynomial L_p regression and matrix L_p regression, which requires $\Omega\left(d^{p/2}\right)$ samples [LiWangWoodruff20]

Our Results (II)

- $\Omega\left(\frac{1}{\varepsilon^{p-1}}\right)$ queries are necessary for $(1+\varepsilon)$-approximation to L_p regression

- Proof recovers a result by [KaneKarmalkarPrice17] showing impossibility of $(2-\varepsilon)$-approximation to L_∞ regression

Approach	Sample Complexity	Approximation
L_p sensitivity sampling ([MMWY21] + Theorem 5.3)	$d^2 p \left(\frac{\log d}{\varepsilon}\right)^{O(1)}$	$(1+\varepsilon)$
L_p sensitivity + Lewis weight sampling [MMWY21]	$d^{\max(1,p/2)} \left(\frac{\log d}{\varepsilon}\right)^{O(1)}$	$(1+\varepsilon)$
L_1 Lewis weight sampling [MMM+22]	$dp \log d$	$O(1)$
Chebyshev measure sampling for all $p \geq 1$ (our results)	$dp \left(\frac{\log d}{\varepsilon}\right)^{O(p)}$	$(1+\varepsilon)$

Algorithm

1. Sample with respect to Chebyshev density on $[-1,1]$
2. Return approximately optimal solution on sketched instance

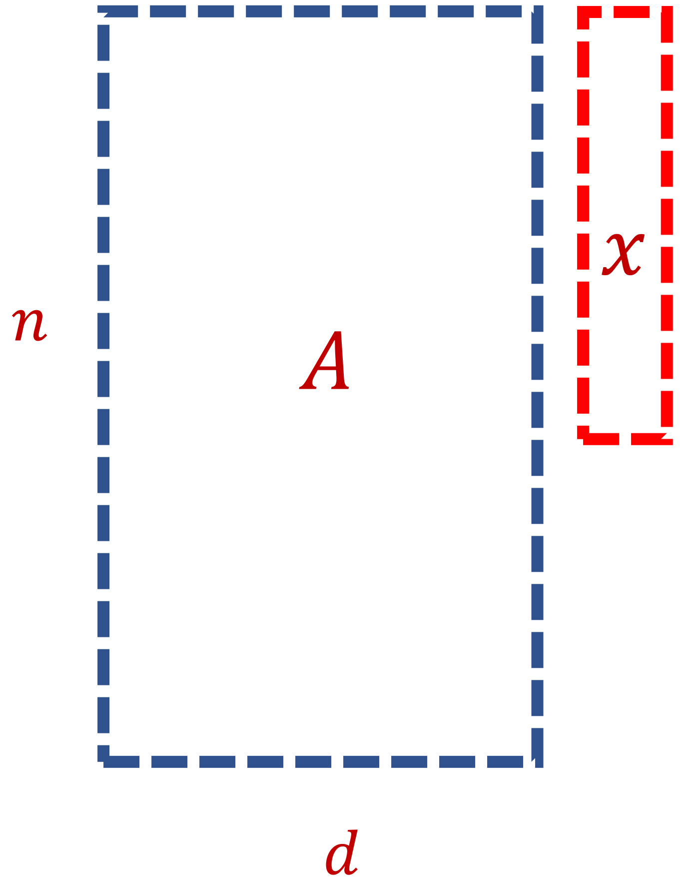
Questions?

- Part 1: Background
- Part 2: Subspace Embeddings
- Part 3: Lewis Weights
- Part 4: Algorithm

Format

1. Show Chebyshev density are the L_p sensitivities
2. Show Chebyshev density are the Lewis weights
3. Uniform sampling + Lewis weight sampling for $p \in [1,2]$
4. Tensor trick + compact net for $p > 2$

Subspace Embedding



- Subspace embedding: Given $\varepsilon > 0$ and $A \in R^{n \times d}$, find matrix $T \in R^{m \times d}$ with $m \ll n$, such that for *every* $x \in R^d$,

$$(1-\varepsilon)\|Ax\|_p \leq \|Tx\|_p \leq (1+\varepsilon)\|Ax\|_p$$

Subspace Embedding

$$n \quad \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \quad A$$

$$d$$

- If the rows of A are “roughly” uniform, could uniformly sample a small number of rows of A and rescale them to form subspace embedding T

Leverage Scores

- Intuition: how "important" a row is (importance sampling)
- $\tau_i(A) = \max \frac{\langle a_i, x\rangle^2}{\|Ax\|_2^2}$ are the *leverage scores* of A (in this case of row a_i)

$$\begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}$$

- $\tau_i(A) = \max \frac{\langle a_i, x\rangle^2}{\|Ax\|_2^2} = \max \frac{\langle a_i, x\rangle^2}{\sum_{i=1}^{n} \langle a_i, x\rangle^2} \leq 1$

Leverage Scores

- $\tau_i(A) = \max \frac{\langle a_i, x\rangle^2}{\|Ax\|_2^2} = \max \frac{\langle a_i, x\rangle^2}{\sum_{i=1}^{n} \langle a_i, x\rangle^2}$
- For $x = (1 \quad -1)$:

$$\begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ -1 \end{bmatrix}$$

- $\langle a_1, x\rangle^2 = (1+0)^2 = 1$ and $\langle a_2, x\rangle^2 = (1-1)^2 = 0$
- $\frac{\langle a_1, x\rangle^2}{\langle a_1, x\rangle^2 + \langle a_2, x\rangle^2} = \frac{1}{1} = 1$, so $\tau_1 = 1$

Leverage Scores

- $\tau_i(A) = \max \frac{\langle a_i, x \rangle^2}{\|Ax\|_2^2} = \max \frac{\langle a_i, x \rangle^2}{\sum_{i=1}^{n} \langle a_i, x \rangle^2}$
- For $x = (1 \quad -1)$:

$$\begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ -1 \end{bmatrix}$$

- $\langle a_1, x \rangle^2 = (1+0)^2 = 1$ and $\langle a_2, x \rangle^2 = (1-1)^2 = 0$
- $\frac{\langle a_2, x \rangle^2}{\langle a_1, x \rangle^2 + \langle a_2, x \rangle^2} = \frac{0}{1} = 0$

Leverage Scores

- $\tau_i(A) = \max \frac{\langle a_i, x\rangle^2}{\|Ax\|_2^2} = \max \frac{\langle a_i, x\rangle^2}{\sum_{i=1}^n \langle a_i, x\rangle^2}$
- For $x = (0 \;\; 1)$:

$$\begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

- $\langle a_1, x\rangle^2 = (0+0)^2 = 0$ and $\langle a_2, x\rangle^2 = (0+1)^2 = 1$
- $\frac{\langle a_2, x\rangle^2}{\langle a_1, x\rangle^2 + \langle a_2, x\rangle^2} = \frac{1}{1} = 1$, so $\tau_2 = 1$

Leverage Scores

- $\tau_i(A) = \max \frac{\langle a_i, x\rangle^2}{\|Ax\|_2^2} = \max \frac{\langle a_i, x\rangle^2}{\sum_{i=1}^n \langle a_i, x\rangle^2}$
- For $x = (1 \;\; 0)$:

$$\begin{pmatrix} 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

- $\frac{\langle a_1, x\rangle^2}{\sum_{i=1}^n \langle a_i, x\rangle^2} = \frac{1}{5}$ and in fact $\tau_1 = \frac{1}{5}$

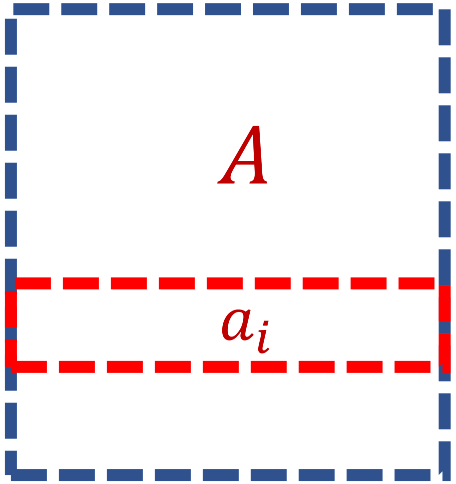
Leverage Scores

- Intuition: how "important" a row is (importance sampling)
- $\tau_i(A) = \max \frac{\langle a_i, x \rangle^2}{\|Ax\|_2^2}$ are the *leverage scores* of A (in this case of row a_i)

$$\begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}$$

- Take $x = (1 \quad -1)$ to see that $\tau_1 = 1$
- Take $x = (0 \quad 1)$ to see that $\tau_2 = 1$

- $\tau_i(A) = a_i (A^\top A)^{-1} a_i^\top, \ \sum \tau_i = d$

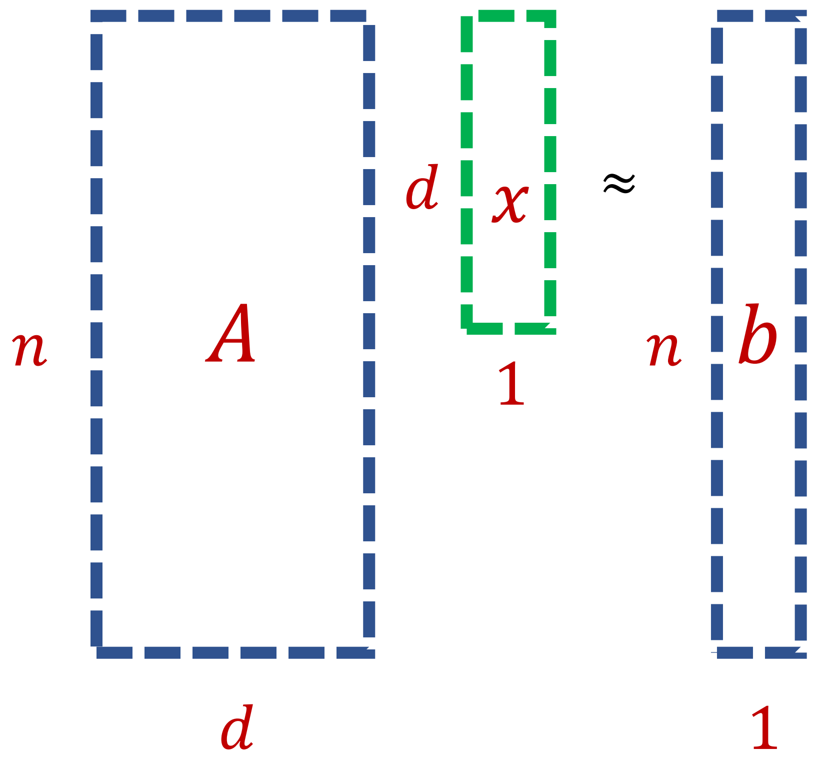


Leverage Scores

- Leverage score sampling: Sample $O\left(\frac{d \log d}{\varepsilon^2}\right)$ rows of A with probability proportional to leverage score $\tau_i(A) = \max \frac{\langle a_i, x \rangle^2}{\|Ax\|_2^2}$
- Rescale sampled rows to form subspace embedding T

$$(1-\varepsilon)\|Ax\|_2 \le \|Tx\|_2 \le (1+\varepsilon)\|Ax\|_2$$

Linear Regression

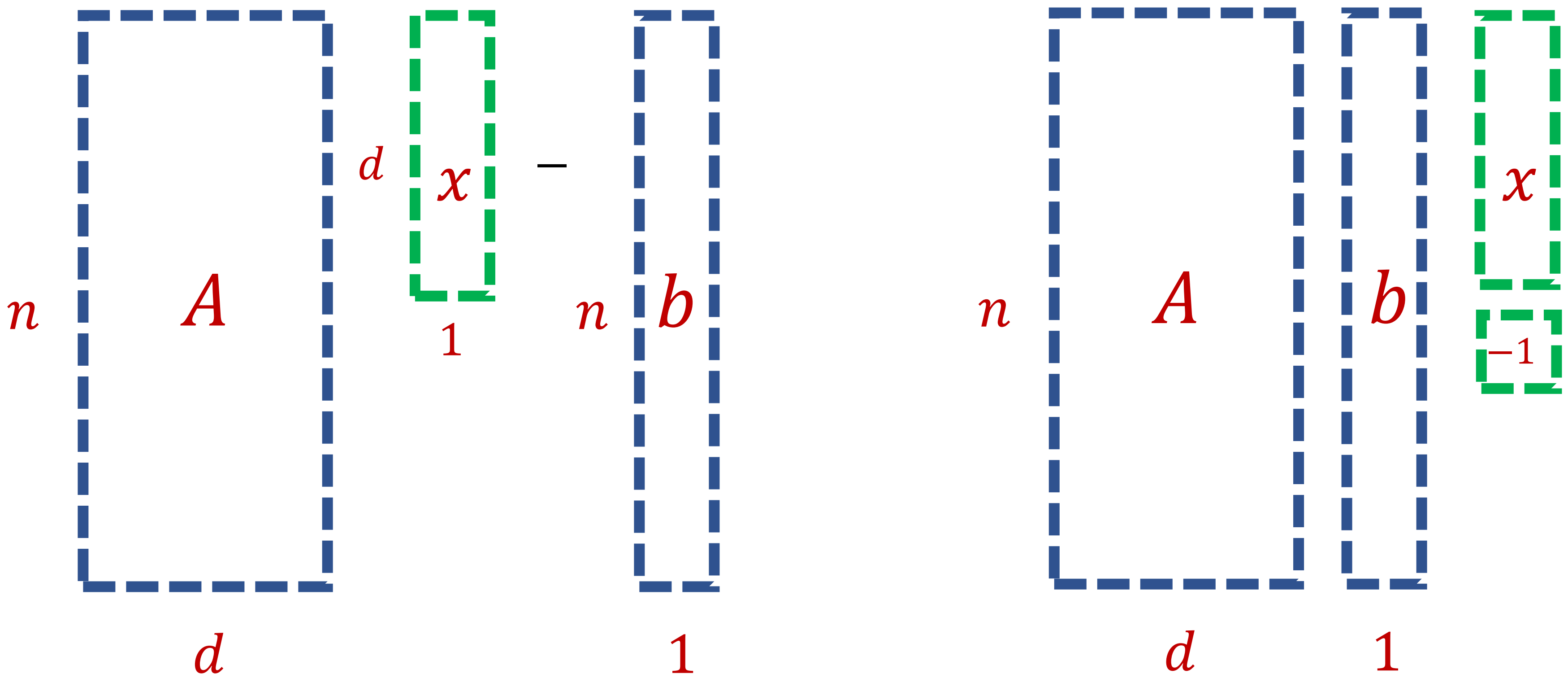


- Find the vector x that minimizes $\|Ax - b\|_2$
- "Least squares" optimization

- Find a vector $\hat{x}$ with $\|A\hat{x} - b\|_2 \leq (1 + \varepsilon)(\min\|Ax - b\|_2)$

Linear Regression

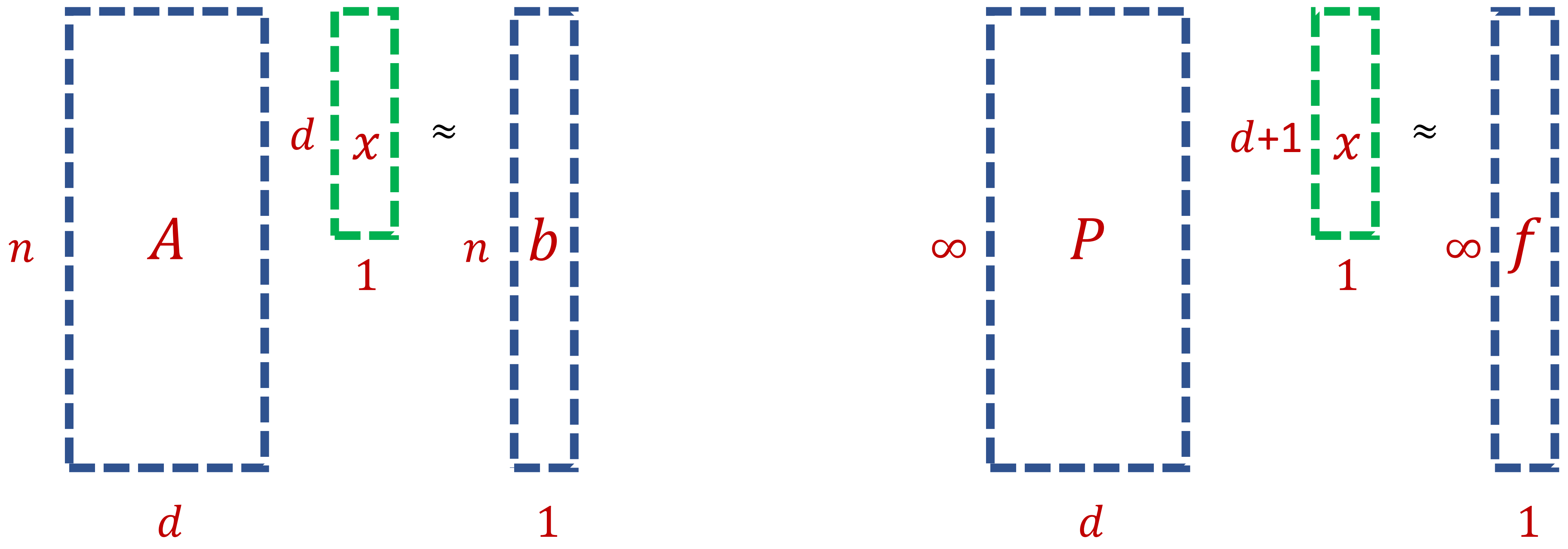
❖ If $B = [A; b]$ and $y = [x; -1]$, then $Ax - b = By$



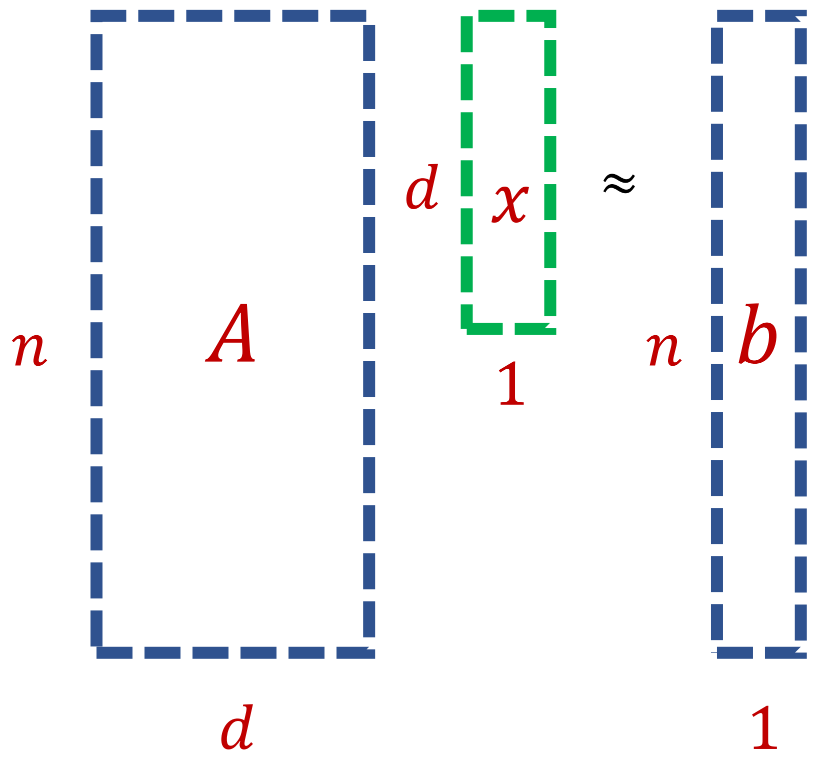
Linear Regression

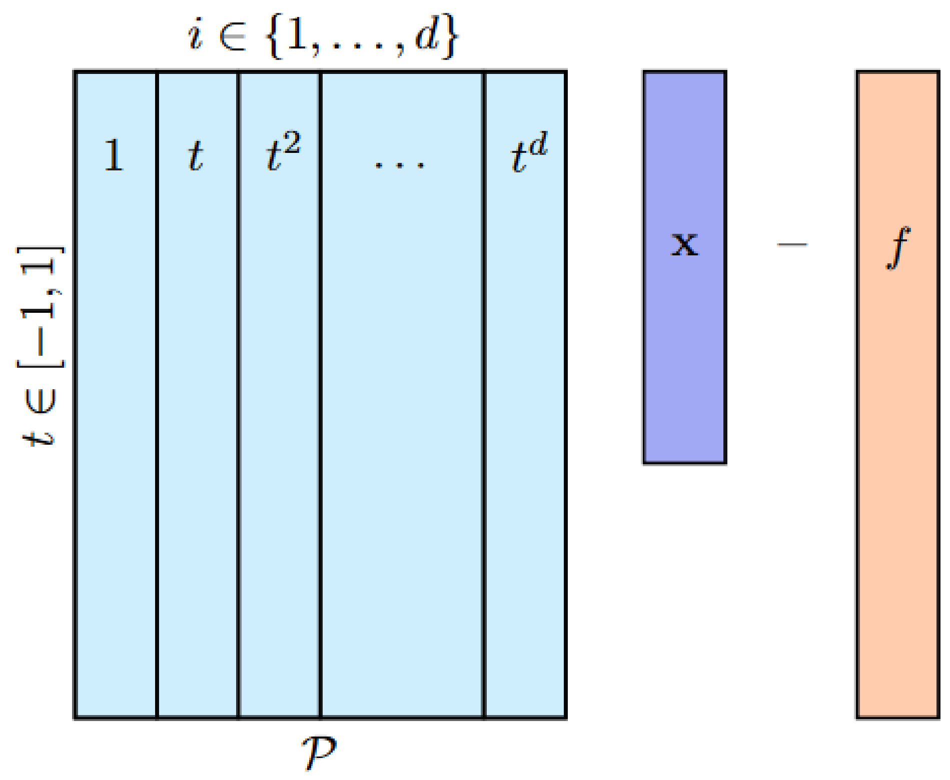
- If $B = [A; b]$ and $y = [x; -1]$, then $Ax - b = By$
- If "free" access to all entries of $B = [A; b]$, suffices to find a subspace embedding for B and then minimize $\|By\|_2$

Linear Regression to Polynomial Regression



Linear Regression to Polynomial Regression





L_2 Polynomial Regression

- Leverage score for matrices: $\tau_i = \max \frac{\langle a_i, x \rangle^2}{\|Ax\|_2^2}$
- Leverage function for operators: $\tau(t) = \max\limits_{\deg(q) \leq d} \frac{|q(t)|^2}{\|q\|_2^2}$
- Can show $\tau(t) \leq O\left(\frac{d}{\sqrt{1-t^2}}\right)$, so roughly $O\left(\frac{d \log^2 d}{\varepsilon^2}\right)$ samples from the Chebyshev density suffice

Toward General p

- Analog of leverage score for general p?

- Previous L_2 leverage scores: $\tau_i(A) = \max \frac{\langle a_i, x\rangle^2}{\|Ax\|_2^2}$

L_p Sensitivities

- L_p sensitivites: $\tau_i^{(p)}(A) = \max \frac{|\langle a_i, x\rangle|^p}{\|Ax\|_p^p}$
- Sample each row a_i with probability $p_i \propto \tau_i^{(p)}(A)$ gives L_p subspace embedding

- Pros: Easy to understand, generalize, i.e., "importance sampling"
- Cons: Gives suboptimal bounds, e.g., $\tilde{O}(d^2)$ samples for $p \in [1,2)$

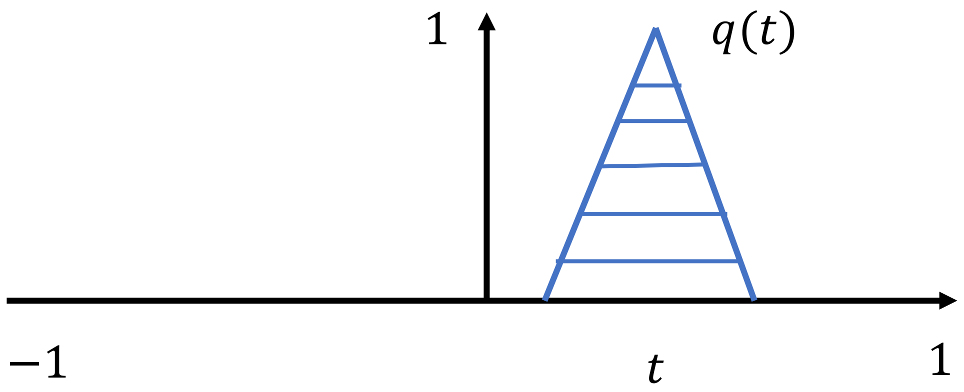
L_p Sensitivities

- L_p sensitivities for matrices: $\tau_i^{(p)}(A) = \max \frac{|\langle a_i, x\rangle|^p}{\|Ax\|_p^p}$
- L_p sensitivities for operators: $\tau^{(p)}(t) = \max_{\deg(q) \leq d} \frac{|q(t)|^p}{\|q\|_p^p}$
- Want to bound $\tau^{(p)}(t)$

Upper Bound for L_p Sensitivities

- Structural result: $\tau^{(p)}(t) = \max_{\deg(q) \le d} \frac{|q(t)|^p}{\|q\|_p^p} \le O\left(\min\left(\frac{dp \log d}{\sqrt{1-t^2}}, d^2 p\right)\right)$

- Normalize $q(t) = 1$, how small can $\|q\|_p^p$ be?



Upper Bound for L_p Sensitivities

- Bernstein's inequality: If q is a polynomial with degree d and $|q(t)| \leq 1$ for $t \in [-1,1]$, then $|q'(t)| \leq \frac{d}{\sqrt{1-t^2}}$ for all $t \in [-1,1]$

- Markov brothers' inequality: If q is a polynomial with degree d and $|q(t)| \leq 1$ for $t \in [-1,1]$, then $|q'(t)| \leq d^2$ for all $t \in [-1,1]$

L_p Sensitivities

- If $|q|$ achieves maximum at t, then $\|q\|_p^p \geq \Omega\left(\max\left(\frac{\sqrt{1-t^2}}{dp}, \frac{1}{d^2 p}\right)\right)$

- Otherwise, show there exists a degree $O(d \log d)$ polynomial r that achieves maximum "near" t and $\left|\|r\|_p^p - \|q\|_p^p\right| \leq \frac{1}{d^3}$

L_p Sensitivities

- Structural result: $\tau^{(p)}(t) = \max_{\deg(q) \leq d} \frac{|q(t)|^p}{\|q\|_p^p} \leq O\left(\min\left(\frac{dp \log d}{\sqrt{1-t^2}}, d^2 p\right)\right)$

- Constant factor approximation to L_p regression with $\text{poly}(d, p)$ queries from the Chebyshev density for all $p \geq 1$, showing separation between polynomial L_p regression and matrix L_p regression, which requires $\Omega\left(d^{p/2}\right)$ samples [LiWangWoodruff20]

Questions?

- Part 1: Background
- Part 2: Subspace Embeddings
- Part 3: Lewis Weights
- Part 4: Algorithm

Format

1. Show Chebyshev density are the L_p sensitivities
2. Show Chebyshev density are the Lewis weights
3. Uniform sampling + Lewis weight sampling for $p \in [1,2]$
4. Tensor trick + compact net for $p > 2$

L_p Sensitivities

- L_p sensitivites: $\tau_i^{(p)}(A) = \max \frac{|\langle a_i, x\rangle|^p}{\|Ax\|_p^p}$
- Sample each row a_i with probability $p_i \propto \tau_i^{(p)}(A)$ gives L_p subspace embedding

- Pros: Easy to understand, generalize, i.e., "importance sampling"
- Cons: Gives suboptimal bounds, e.g., $\tilde{O}(d^2)$ samples for $p \in [1,2)$

L_p Lewis Weights

- L_p Lewis weights [CohenPeng15]: $w_i = \tau_i\left(W^{\frac{1}{2}-\frac{1}{p}}A\right)$
- Sample each row a_i with probability $p_i \propto w_i$ gives L_p subspace embedding

- Pros: Gives near-optimal bounds, e.g., $\tilde{O}(d)$ samples for $p \in [1,2)$
- Cons: Difficult to understand, generalize, i.e., “reweighted importance sampling”

Properties L_p Lewis Weights

- L_p Lewis weights can be approximated by iteratively computing $\tau_i\left(W^{\frac{1}{2}-\frac{1}{p}}A\right)$ after initializing $W = I_n$

- If $\frac{1}{C} \leq \frac{\tau_i\left(W^{\frac{1}{2}-\frac{1}{p}}A\right)}{w_i} \leq C$, then W is a C-approximation to the L_p Lewis weights, for $p \in [1,2]$

L_1 Lewis Weight Fixed Point Ratio

- Goal: Show $\frac{1}{C} \leq \frac{\tau\left(W^{-1/2}\, P\right)}{w(t)} \leq C$, where τ is the leverage score function, $w(t) = \frac{d}{\sqrt{1-t^2}}$ is the Chebyshev density, and P is the polynomial operator

- Change of basis to Chebyshev polynomials of the second kind, which are orthogonal under the inner product

$$\int_{-1}^{1} f(t)g(t)\sqrt{1-t^2}dt$$

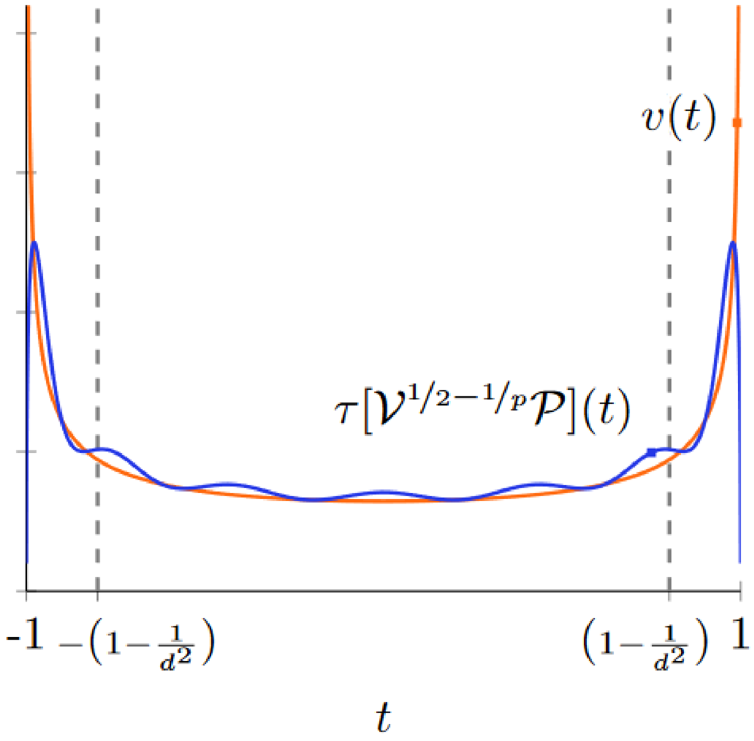
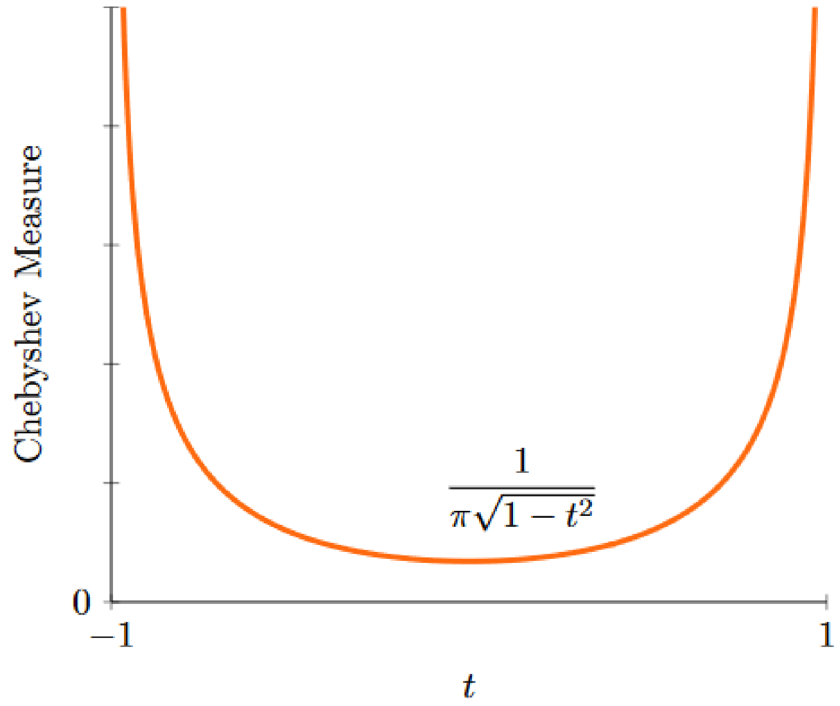


Figure 6: Plot of the scaled Chebyshev Measure (—) and corresponding reweighted leverage function $\tau[\mathcal{V}^{\frac{1}{2}-\frac{1}{p}}\mathcal{P}](t)$ (—) on $[-1, 1]$ for $d = 6$, $p = 1$. For most values of t both curves are close, but for $|t| > 1 - \frac{1}{d^2}$ the curves diverge. This means that the Chebyshev density itself does not directly approximate the L_p Lewis weights, motivating our study of a clipped version of the measure, denoted $w(t)$.

L_1 Lewis Weight Fixed Point Ratio

- Goal: Show $\frac{1}{C} \leq \frac{\tau(W^{-1/2}P)}{w(t)} \leq C$, where τ is the leverage score function, $w(t) = \frac{d}{\sqrt{1-t^2}}$ is the Chebyshev density, and P is the polynomial operator



- NOT TRUE!

L_1 Lewis Weight Fixed Point Ratio

- Goal: Show $\frac{1}{C} \leq \frac{\tau\left(U^{-1/2}\,P\right)}{u(t)} \leq C$, where $u(t) = \min\left(\frac{d}{\sqrt{1-t^2}}, d^2\right)$ is the clipped Chebyshev density

- Behavior in the "middle" of $u(t)$ is similar to $w(t)$
- Upper bounding the ratio in the "endcaps" from upper bounding the numerator
- Lower bounding the ratio in the "endcaps" by evaluating the numerator for a low-degree approximation of a high-degree polynomial

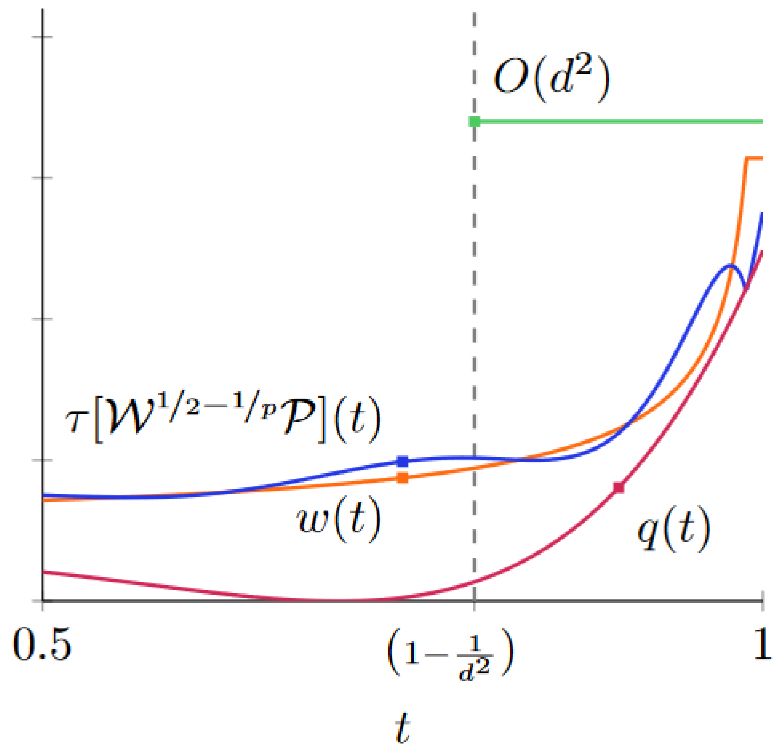


Figure 7: Plot of the clipped Chebyshev Measure (—) and corresponding reweighted leverage function (—) for $t \in [0.5, 1]$ and $d = 6$, $p = 1$. As proven in Theorem 2.2, these functions are within a constant factor for all t, so we can claim that the clipped measure approximates the L_p Lewis weights. We also visualize the "spike" polynomial $q(t)$ (—) and upper bound (—) used in the proof of Theorem 2.2.

L_p Lewis Weight Fixed Point Ratio

- Structural result for $p = 1$: $\frac{1}{\text{polylog}(d)} \leq \frac{\tau(U^{-1/2} P)}{u(t)} \leq \text{polylog}(d)$

- By using Jacobi polynomials instead: $\frac{1}{\text{polylog}(d)} \leq \frac{\tau\left(U^{\frac{1}{2}-\frac{1}{p}} P\right)}{u(t)} \leq \text{polylog}(d)$ for $p \in [1,2]$

L_p Lewis Weights Challenges

- There are no L_p known Lewis weights for operators

- …no approximate Lewis weight theorem!

Questions?

- Part 1: Background
- Part 2: Subspace Embeddings
- Part 3: Lewis Weights
- Part 4: Algorithm

Format

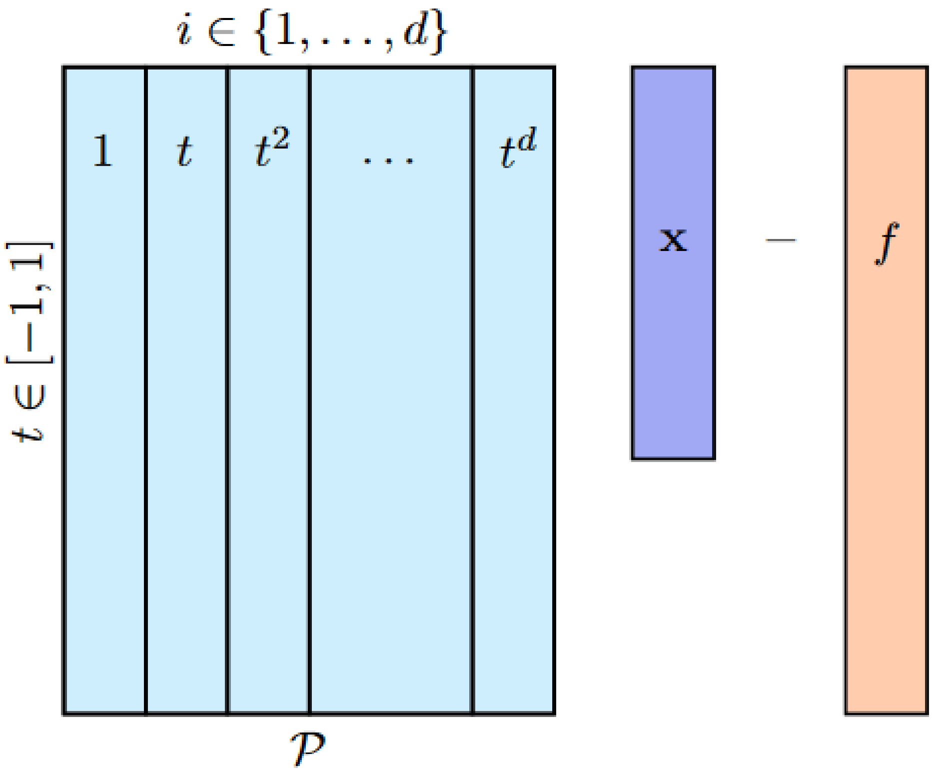
1. Show Chebyshev density are the L_p sensitivities
2. Show Chebyshev density are the Lewis weights
3. Uniform sampling + Lewis weight sampling for $p \in [1,2]$
4. Tensor trick + compact net for $p > 2$

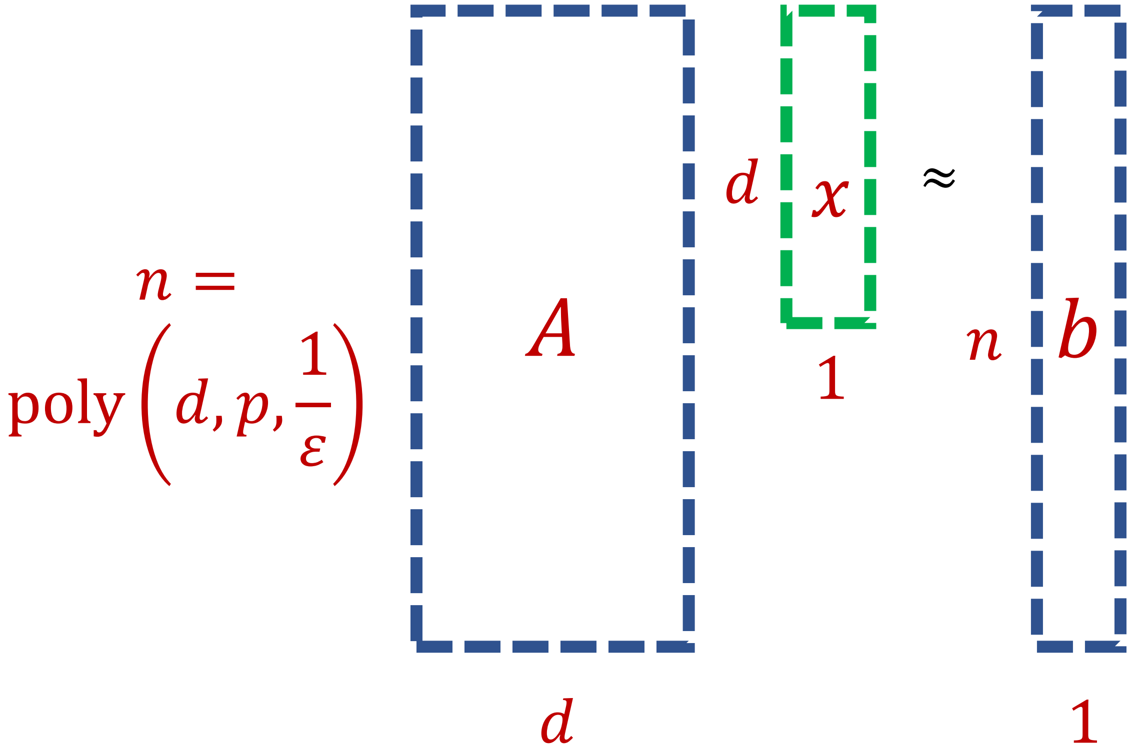
Uniform Sampling

- Sample $\mathrm{poly}\left(d, p, \frac{1}{\varepsilon}\right)$ points uniformly at random from $[-1,1]$ and form a matrix A from these points
- Let b be the corresponding measurements of the signal f

- $\|Ax - b\|_p \approx \|Px - f\|_p = \left(\int_{-1}^{1} |Px(t) - f(t)|^p \, dt\right)^{1/p}$

Uniform Sampling

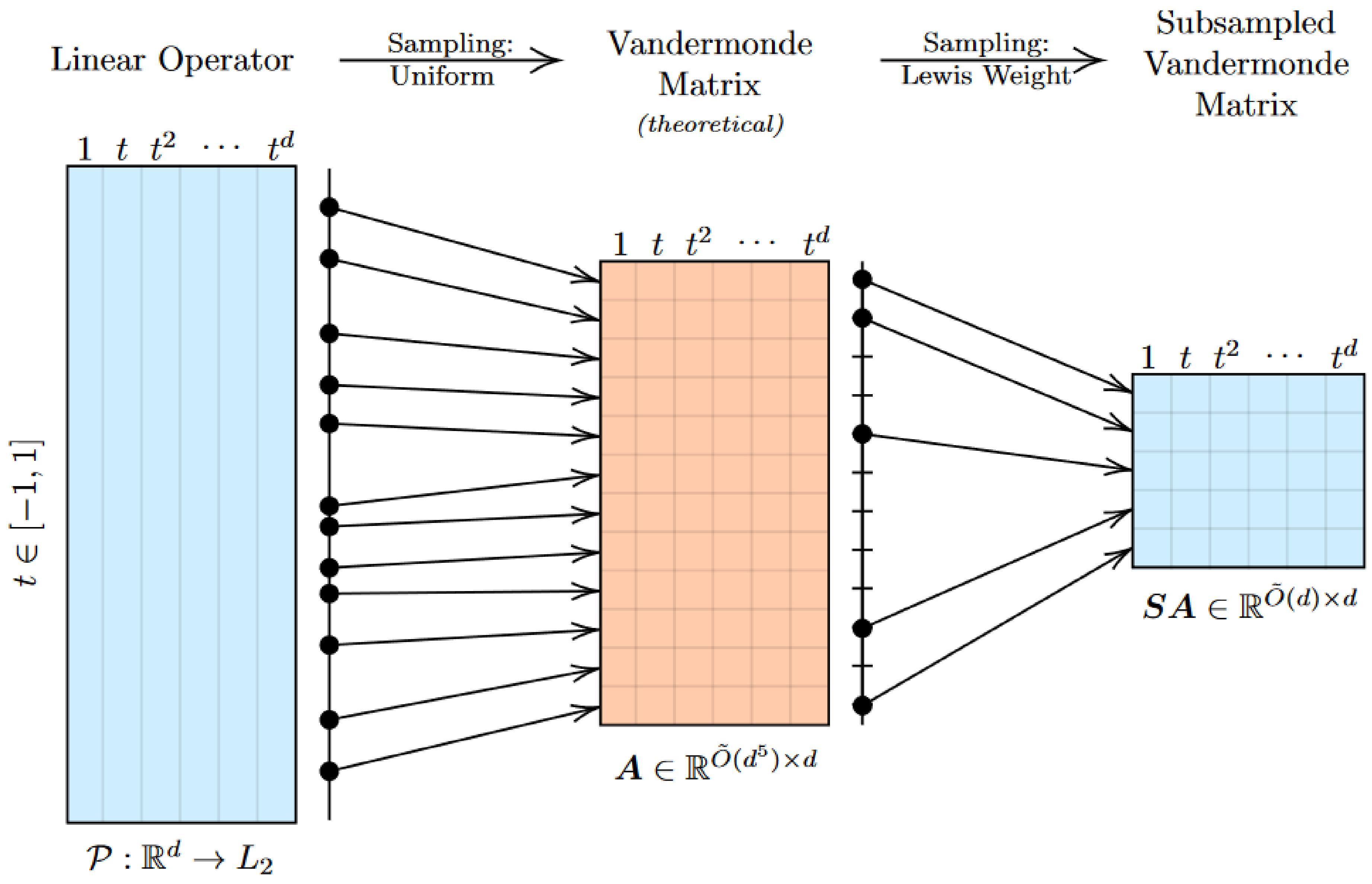




Uniform Sampling Preserves Fixed Point Ratio

- $\frac{\tau(U^{-1/2} P)}{u(t)} \approx \frac{\tau(W^{-1/2} A)}{w_i(A)}$

- Only $\tilde{O}\left(\frac{d}{\varepsilon^{O(1)}}\right)$ samples with L_p Lewis weights needed for $(1+\varepsilon)$-approximation to L_p regression with $p \in [1,2]$ [ChenDerezinski21, ParulekarParulekarPrice21, MuscoMuscoWoodruffYasuda22]


Linear Operator
Sampling: Uniform
Vandermonde Matrix (theoretical)
Sampling: Lewis Weight
Subsampled Vandermonde Matrix
$1 \quad t \quad t^2 \quad \cdots \quad t^d$
$t \in [-1, 1]$
$\mathcal{P} : \mathbb{R}^d \to L_2$
$\boldsymbol{A} \in \mathbb{R}^{\tilde{O}(d^5) \times d}$
$\boldsymbol{S}\boldsymbol{A} \in \mathbb{R}^{\tilde{O}(d) \times d}$

(Simplified) Algorithm

1. Uniform sample $n = \text{poly}\left(d, p, \frac{1}{\varepsilon}\right)$ points from $[-1,1]$ form a matrix A from these points
2. Perform L_p Lewis weight sampling on A
3. Return approximately optimal solution on sketched instance

1. Sample with respect to Chebyshev density on $[-1,1]$
2. Return approximately optimal solution on sketched instance

Challenges for $p > 2$

- Do not have structural property relating Chebyshev density with L_p Lewis weights for $p > 2$

- L_p Lewis weights use $O\left(d^{p/2}\right)$ samples

L_p Regression for $p > 2$

- L_p sensitivities are upper bounded by Chebyshev density

- Use tensoring trick of [MeyerMuscoMuscoWoodruffZhou22] to union bound over a smaller net

(Simplified) Algorithm for $p > 2$

1. Uniform sample $n = \text{poly}\left(d, p, \frac{1}{\varepsilon}\right)$ points from $[-1,1]$ form a matrix A from these points
2. Perform L_p sensitivity sampling on A
3. Return approximately optimal solution on sketched instance

1. Sample with respect to Chebyshev density on $[-1,1]$
2. Return approximately optimal solution on sketched instance

Lower Bound

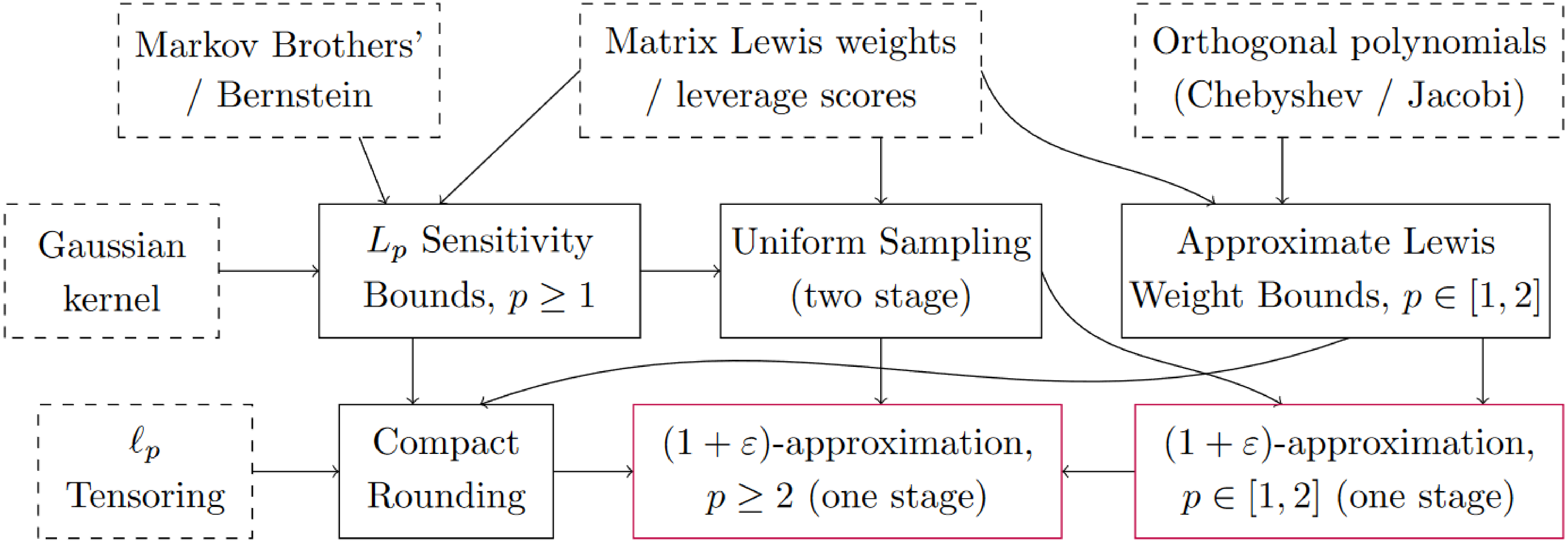
- $\Omega\left(\frac{1}{\varepsilon^{p-1}}\right)$ queries are necessary for $(1+\varepsilon)$-approximation to L_p regression
- Let $n = \frac{1}{\varepsilon^{p-1}}$ and I be an interval of length $\frac{n}{100}$ from $[-1,1]$ so that with probability $\frac{2}{3}$, no query lands in I
- Define $f_+ = \frac{2^{\frac{1}{p}}}{\varepsilon}$ on I and 0 elsewhere, define $f_- = -\frac{2^{\frac{1}{p}}}{\varepsilon}$ on I and 0 elsewhere
- $\|q - f_+\|_p^p = \left(1 - O(\varepsilon)\right) \|f_+\|_p^p$ for $q(t) = 1$

Summary

- $(1+\varepsilon)$-approximation to L_p regression with $dp\left(\frac{\log^{O(p)} d}{\varepsilon^{O(p)}}\right)$ queries from the Chebyshev density for all $p \geq 1$
- $\Omega\left(\frac{1}{\varepsilon^{p-1}}\right)$ queries are necessary for $(1+\varepsilon)$-approximation to L_p regression
- Structural result: $\tau^{(p)}(t) = \max_{\deg(q) \leq d} \frac{|q(t)|^p}{\|q\|_p^p} \leq O\left(\min\left(\frac{dp \log d}{\sqrt{1-t^2}}, d^2 p\right)\right)$
- Structural result: $\frac{1}{\mathrm{polylog}(d)} \leq \frac{\tau\left(U^{\frac{1}{2}-\frac{1}{p}} P\right)}{u(t)} \leq \mathrm{polylog}(d)$ for $p \in [1,2]$

Summary

- $(1+\varepsilon)$-approximation to L_p regression with $dp\left(\frac{\log^{O(p)} d}{\varepsilon^{O(p)}}\right)$ queries from the Chebyshev density for all $p \geq 1$
- $\Omega\left(\frac{1}{\varepsilon^{p-1}}\right)$ queries are necessary for $(1+\varepsilon)$-approximation to L_p regression
- Question: Other loss functions?
- Question: Sparse Fourier regression [ChenKanePriceSong16, AvronKapralovMuscoMuscoVelingkerZandieh19]


Markov Brothers' / Bernstein
Matrix Lewis weights / leverage scores
Orthogonal polynomials (Chebyshev / Jacobi)
Gaussian kernel
L_p Sensitivity Bounds, $p \geq 1$
Uniform Sampling (two stage)
Approximate Lewis Weight Bounds, $p \in [1, 2]$
ℓ_p Tensoring
Compact Rounding
$(1+\varepsilon)$-approximation, $p \geq 2$ (one stage)
$(1+\varepsilon)$-approximation, $p \in [1, 2]$ (one stage)