Computing Approximate ℓ_p Sensitivities

Joint work with
David P. Woodruff[2], and Richard Q. Zhang[3]

[2]Carnegie Mellon University; [3]Google DeepMind

October 20, 2023

To appear in NeurIPS 2023

Problem Statement

Motivation

- Approximating a representative set of data points is an important pre-processing step when $n \gg d$

Motivation

- Approximating a representative set of data points is an important pre-processing step when $n \gg d$
- One approach: uniform sampling

Motivation

- Approximating a representative set of data points is an important pre-processing step when $n \gg d$
- One approach: uniform sampling
- Perhaps a more principled approach: importance sampling

Introduction: Sensitivity Scores

- Consider the objective: $\min_{\mathbf{x}\in\mathcal{X}} \sum_{i=1}^{n} f_i(\mathbf{x})$

Introduction: Sensitivity Scores

- Consider the objective: $\min_{\mathbf{x}\in\mathcal{X}} \sum_{i=1}^{n} f_i(\mathbf{x})$

The i^{th} sensitivity score for the above objective is

$$\sigma_i := \max_{\mathbf{x}\in\mathcal{X}} \frac{f_i(\mathbf{x})}{\sum_{j=1}^{n} f_j(\mathbf{x})}$$

Introduction: Sensitivity Scores

- Consider the objective: $\min_{\mathbf{x}\in\mathcal{X}} \sum_{i=1}^{n} f_i(\mathbf{x})$

The i^{th} sensitivity score for the above objective is

$$\sigma_i := \max_{\mathbf{x}\in\mathcal{X}} \frac{f_i(\mathbf{x})}{\sum_{j=1}^{n} f_j(\mathbf{x})}$$

- Sampling $\propto \sigma_i$ preserves the objective in expectation for every $\mathbf{x} \in \mathcal{X}$

Introduction: Sensitivity Scores

- Consider the objective: $\min_{\mathbf{x}\in\mathcal{X}} \sum_{i=1}^{n} f_i(\mathbf{x})$

The i^{th} sensitivity score for the above objective is

$$\sigma_i := \max_{\mathbf{x}\in\mathcal{X}} \frac{f_i(\mathbf{x})}{\sum_{j=1}^{n} f_j(\mathbf{x})}$$

- Sampling $\propto \sigma_i$ preserves the objective in expectation for every $\mathbf{x} \in \mathcal{X}$
- Sampling $\propto \varepsilon^{-2}\mathfrak{S}$ functions, where $\mathfrak{S} := \sum_{i=1}^{n} \sigma_i$, gives a $(1 \pm \varepsilon)$-approximation to the objective (Braverman, Feldman, Lang, Statman, Zhou (2016))

Introduction: Sensitivity Scores

- Consider the objective: $\min_{\mathbf{x}\in\mathcal{X}} \sum_{i=1}^n f_i(\mathbf{x})$

The i^{th} sensitivity score for the above objective is

$$\sigma_i := \max_{\mathbf{x}\in\mathcal{X}} \frac{f_i(\mathbf{x})}{\sum_{j=1}^n f_j(\mathbf{x})}$$

- Sampling $\propto \sigma_i$ preserves the objective in expectation for every $\mathbf{x} \in \mathcal{X}$
- Sampling $\propto \varepsilon^{-2}\mathfrak{S}$ functions, where $\mathfrak{S} := \sum_{i=1}^n \sigma_i$, gives a $(1 \pm \varepsilon)$-approximation to the objective (Braverman, Feldman, Lang, Statman, Zhou (2016))
- Efficient computation of sensitivities less well-studied compared to related quantities like leverage scores and Lewis weights

Related Quantity: Leverage Scores

The i^{th} leverage score of $\mathbf{A} \in \mathbb{R}^{n \times d}$ is defined as

$$\tau(\mathbf{a}_i) := \min_{\mathbf{x} \in \mathbb{R}^n : \mathbf{A}^\top \mathbf{x} = \mathbf{a}_i} \|\mathbf{x}\|_2^2$$

Related Quantity: Leverage Scores

The i^{th} leverage score of $\mathbf{A} \in \mathbb{R}^{n \times d}$ is defined as

$$\tau(\mathbf{a}_i) := \min_{\mathbf{x} \in \mathbb{R}^n : \mathbf{A}^\top \mathbf{x} = \mathbf{a}_i} \|\mathbf{x}\|_2^2$$

- Measures the importance of a row in composing the matrix's rowspace

Related Quantity: Leverage Scores

The i^{th} leverage score of $\mathbf{A} \in \mathbb{R}^{n \times d}$ is defined as

$$\tau(\mathbf{a}_i) := \min_{\mathbf{x} \in \mathbb{R}^n : \mathbf{A}^\top \mathbf{x} = \mathbf{a}_i} \|\mathbf{x}\|_2^2$$

- Measures the importance of a row in composing the matrix's rowspace
- Efficient computation (Cohen, Lee, Musco, Musco, Peng, Sidford (2014))

Related Quantity: Leverage Scores

The i^{th} leverage score of $\mathbf{A} \in \mathbb{R}^{n \times d}$ is defined as

$$\tau(\mathbf{a}_i) := \min_{\mathbf{x} \in \mathbb{R}^n : \mathbf{A}^\top \mathbf{x} = \mathbf{a}_i} \|\mathbf{x}\|_2^2$$

- Measures the importance of a row in composing the matrix's rowspace
- Efficient computation (Cohen, Lee, Musco, Musco, Peng, Sidford (2014))
- Leverage scores give optimal sample complexity for ℓ_2 regression (Mahoney (2011))

Related Quantity: Leverage Scores

The i^{th} leverage score of $\mathbf{A} \in \mathbb{R}^{n \times d}$ is defined as

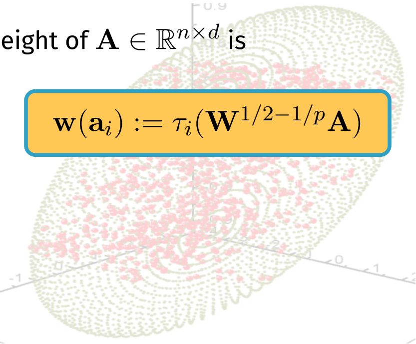
$$\tau(\mathbf{a}_i) := \min_{\mathbf{x} \in \mathbb{R}^n : \mathbf{A}^\top \mathbf{x} = \mathbf{a}_i} \|\mathbf{x}\|_2^2$$

- Measures the importance of a row in composing the matrix's rowspace
- Efficient computation (Cohen, Lee, Musco, Musco, Peng, Sidford (2014))
- Leverage scores give optimal sample complexity for ℓ_2 regression (Mahoney (2011))
- What about for ℓ_p regression?

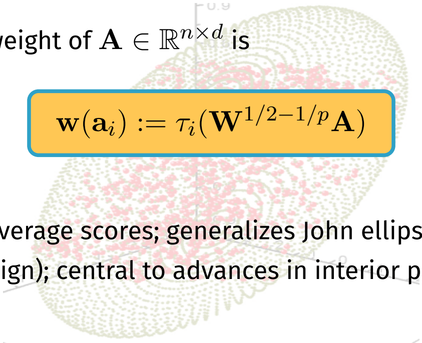
Related Quantity: Lewis Weights

The i^{th} Lewis weight of $\mathbf{A} \in \mathbb{R}^{n \times d}$ is

$$\mathbf{w}(\mathbf{a}_i) := \tau_i(\mathbf{W}^{1/2-1/p}\mathbf{A})$$



Related Quantity: Lewis Weights



The i^{th} Lewis weight of $\mathbf{A} \in \mathbb{R}^{n \times d}$ is

$$\mathbf{w}(\mathbf{a}_i) := \tau_i(\mathbf{W}^{1/2-1/p}\mathbf{A})$$

- Generalizes leverage scores; generalizes John ellipsoid (used in D-optimal design); central to advances in interior point methods

Related Quantity: Lewis Weights

The i^{th} Lewis weight of $\mathbf{A} \in \mathbb{R}^{n \times d}$ is

$$\mathbf{w}(\mathbf{a}_i) := \tau_i(\mathbf{W}^{1/2-1/p}\mathbf{A})$$

- Generalizes leverage scores; generalizes John ellipsoid (used in D-optimal design); central to advances in interior point methods
- Efficient computation (Cohen & Peng (2015), Fazel, Lee, P., Sidford (2022))

Related Quantity: Lewis Weights

The i^{th} Lewis weight of $\mathbf{A} \in \mathbb{R}^{n \times d}$ is

$$\mathbf{w}(\mathbf{a}_i) := \tau_i(\mathbf{W}^{1/2-1/p}\mathbf{A})$$

- Generalizes leverage scores; generalizes John ellipsoid (used in D-optimal design); central to advances in interior point methods
- Efficient computation (Cohen & Peng (2015), Fazel, Lee, P., Sidford (2022))
- Give better sample complexity than leverage scores for ℓ_p regression

Related Quantity: Lewis Weights

The i^{th} Lewis weight of $\mathbf{A} \in \mathbb{R}^{n \times d}$ is

$$\mathbf{w}(\mathbf{a}_i) := \tau_i(\mathbf{W}^{1/2-1/p}\mathbf{A})$$

- Generalizes leverage scores; generalizes John ellipsoid (used in D-optimal design); central to advances in interior point methods
- Efficient computation (Cohen & Peng (2015), Fazel, Lee, P., Sidford (2022))
- Give better sample complexity than leverage scores for ℓ_p regression
- Can we do better in practice?

Definition: ℓ_p Sensitivities

Leverage scores may alternately be defined as

$$\tau(\mathbf{a}_i) := \max_{\mathbf{x}:\mathbf{A}\mathbf{x}\neq\mathbf{0}} \frac{|\mathbf{a}_i^\top \mathbf{x}|^2}{\|\mathbf{A}\mathbf{x}\|_2^2}$$

Definition: ℓ_p Sensitivities

Leverage scores may alternately be defined as

$$\tau(\mathbf{a}_i) := \max_{\mathbf{x}:\mathbf{A}\mathbf{x}\neq\mathbf{0}} \frac{|\mathbf{a}_i^\top \mathbf{x}|^2}{\|\mathbf{A}\mathbf{x}\|_2^2}$$

ℓ_p sensitivities are the values of the following convex program

$$\sigma_p(\mathbf{a}_i) := \max_{\mathbf{x}:\mathbf{A}\mathbf{x}\neq\mathbf{0}} \frac{|\mathbf{a}_i^\top \mathbf{x}|^p}{\|\mathbf{A}\mathbf{x}\|_p^p}$$

Definition: ℓ_p Sensitivities

Leverage scores may alternately be defined as

$$\tau(\mathbf{a}_i) := \max_{\mathbf{x}:\mathbf{A}\mathbf{x}\neq\mathbf{0}} \frac{|\mathbf{a}_i^\top \mathbf{x}|^2}{\|\mathbf{A}\mathbf{x}\|_2^2}$$

ℓ_p sensitivities are the values of the following convex program

$$\sigma_p(\mathbf{a}_i) := \max_{\mathbf{x}:\mathbf{A}\mathbf{x}\neq\mathbf{0}} \frac{|\mathbf{a}_i^\top \mathbf{x}|^p}{\|\mathbf{A}\mathbf{x}\|_p^p}$$

- Introduced by Langberg & Schulman (2010)

Definition: ℓ_p Sensitivities

Leverage scores may alternately be defined as

$$\tau(\mathbf{a}_i) := \max_{\mathbf{x}:\mathbf{A}\mathbf{x}\neq\mathbf{0}} \frac{|\mathbf{a}_i^\top \mathbf{x}|^2}{\|\mathbf{A}\mathbf{x}\|_2^2}$$

ℓ_p sensitivities are the values of the following convex program

$$\sigma_p(\mathbf{a}_i) := \max_{\mathbf{x}:\mathbf{A}\mathbf{x}\neq\mathbf{0}} \frac{|\mathbf{a}_i^\top \mathbf{x}|^p}{\|\mathbf{A}\mathbf{x}\|_p^p}$$

- Introduced by Langberg & Schulman (2010)
- Sampling proportional to sensitivities preserves value

Definition: ℓ_p Sensitivities

Leverage scores may alternately be defined as

$$\tau(\mathbf{a}_i) := \max_{\mathbf{x}:\mathbf{A}\mathbf{x}\neq\mathbf{0}} \frac{|\mathbf{a}_i^\top \mathbf{x}|^2}{\|\mathbf{A}\mathbf{x}\|_2^2}$$

ℓ_p sensitivities are the values of the following convex program

$$\sigma_p(\mathbf{a}_i) := \max_{\mathbf{x}:\mathbf{A}\mathbf{x}\neq\mathbf{0}} \frac{|\mathbf{a}_i^\top \mathbf{x}|^p}{\|\mathbf{A}\mathbf{x}\|_p^p}$$

- Introduced by Langberg & Schulman (2010)
- Sampling proportional to sensitivities preserves value
- Total sample complexity proportional to total sensitivity

Applications of Sensitivities

- Applications:

Applications of Sensitivities

- Applications:
  - shape-fitting (Varadarajan & Xiao (2012); Braverman, Feldman, Lang, Statman, Zhou (2021); Tukan, Wu, Zhou, Braverman, Feldman (2022))

Applications of Sensitivities

- Applications:
  - shape-fitting (Varadarajan & Xiao (2012); Braverman, Feldman, Lang, Statman, Zhou (2021); Tukan, Wu, Zhou, Braverman, Feldman (2022))
  - clustering (Feldman & Langberg (2011); Bachem, Lucic, Krause (2015))

Applications of Sensitivities

- Applications:
  - shape-fitting (Varadarajan & Xiao (2012); Braverman, Feldman, Lang, Statman, Zhou (2021); Tukan, Wu, Zhou, Braverman, Feldman (2022))
  - clustering (Feldman & Langberg (2011); Bachem, Lucic, Krause (2015))
  - regression (Huggins, Campbell, Broderick (2016); Munteanu, Schwiegelshohn, Sohler, Woodruff (2018))

Applications of Sensitivities

- Applications:
  - shape-fitting (Varadarajan & Xiao (2012); Braverman, Feldman, Lang, Statman, Zhou (2021); Tukan, Wu, Zhou, Braverman, Feldman (2022))
  - clustering (Feldman & Langberg (2011); Bachem, Lucic, Krause (2015))
  - regression (Huggins, Campbell, Broderick (2016); Munteanu, Schwiegelshohn, Sohler, Woodruff (2018))
  - pruning of deep neural networks (Liebenwein, Baykal, Lang, Feldman, Rus (2019), Tukan, Mualen, Maalouf (2022), Mussay, Feldman, Zhou, Braverman, Osadchy (2021))

Applications of Sensitivities

- Applications:
  - shape-fitting (Varadarajan & Xiao (2012); Braverman, Feldman, Lang, Statman, Zhou (2021); Tukan, Wu, Zhou, Braverman, Feldman (2022))
  - clustering (Feldman & Langberg (2011); Bachem, Lucic, Krause (2015))
  - regression (Huggins, Campbell, Broderick (2016); Munteanu, Schwiegelshohn, Sohler, Woodruff (2018))
  - pruning of deep neural networks (Liebenwein, Baykal, Lang, Feldman, Rus (2019), Tukan, Mualen, Maalouf (2022), Mussay, Feldman, Zhou, Braverman, Osadchy (2021))
- Superior to Lewis weights sampling in practical regimes

Applications of Sensitivities

- Applications:
  - shape-fitting (Varadarajan & Xiao (2012); Braverman, Feldman, Lang, Statman, Zhou (2021); Tukan, Wu, Zhou, Braverman, Feldman (2022))
  - clustering (Feldman & Langberg (2011); Bachem, Lucic, Krause (2015))
  - regression (Huggins, Campbell, Broderick (2016); Munteanu, Schwiegelshohn, Sohler, Woodruff (2018))
  - pruning of deep neural networks (Liebenwein, Baykal, Lang, Feldman, Rus (2019), Tukan, Mualen, Maalouf (2022), Mussay, Feldman, Zhou, Braverman, Osadchy (2021))
- Superior to Lewis weights sampling in practical regimes
  - when the total sensitivity is low (Woodruff & Yasuda (2023))

Applications of Sensitivities

- Applications:
  - shape-fitting (Varadarajan & Xiao (2012); Braverman, Feldman, Lang, Statman, Zhou (2021); Tukan, Wu, Zhou, Braverman, Feldman (2022))
  - clustering (Feldman & Langberg (2011); Bachem, Lucic, Krause (2015))
  - regression (Huggins, Campbell, Broderick (2016); Munteanu, Schwiegelshohn, Sohler, Woodruff (2018))
  - pruning of deep neural networks (Liebenwein, Baykal, Lang, Feldman, Rus (2019), Tukan, Mualen, Maalouf (2022), Mussay, Feldman, Zhou, Braverman, Osadchy (2021))
- Superior to Lewis weights sampling in practical regimes
  - when the total sensitivity is low (Woodruff & Yasuda (2023))
  - structured matrices like sparse/low-rank/combinations (Meyer, Musco, Musco, Woodruff, Zhou (2022))

Our Goal

- Fast algorithms to approximate various functions of sensitivities

Our Goal

additive

- Fast algorithms to approximate various functions of sensitivities

Our Goal

additive

- Fast algorithms to approximate various functions of sensitivities
  - All sensitivities
  - The total sensitivity
  - The maximum sensitivity

Our Goal

additive

- Fast algorithms to approximate various functions of sensitivities
  - All sensitivities
  - The total sensitivity
  - The maximum sensitivity
- Runtime measured in number of sensitivity computations

I. Approximating All ℓ_1 Sensitivities

First Result: Estimating All ℓ_p Sensitivities

Theorem 1: Estimating All ℓ_1 Sensitivities

Given a full-rank $\mathbf{A} \in \mathbb{R}^{m \times n}$ and $\alpha \geq 1$, we compute a vector $\widetilde{\sigma} \in \mathbb{R}^n$ such that, with high probability, for each $i \in [n]$,

$$\sigma_1(\mathbf{a}_i) \leq \widetilde{\sigma}_i \leq O(\sigma_1(\mathbf{a}_i) + \tfrac{\alpha}{n}\mathfrak{S}_1(\mathbf{A})).$$

Our runtime is $\widetilde{O}(\mathbf{nnz}(\mathbf{A}) + \frac{n}{\alpha} \cdot d^{\omega})$.

First Result: Estimating All ℓ_p Sensitivities

Theorem 1: Estimating All ℓ_1 Sensitivities

Given a full-rank $\mathbf{A} \in \mathbb{R}^{m \times n}$ and $\alpha \geq 1$, we compute a vector $\widetilde{\sigma} \in \mathbb{R}^n$ such that, with high probability, for each $i \in [n]$,

$$\sigma_1(\mathbf{a}_i) \leq \widetilde{\sigma}_i \leq O(\sigma_1(\mathbf{a}_i) + \tfrac{\alpha}{n}\mathfrak{S}_1(\mathbf{A})).$$

Our runtime is $\widetilde{O}(\mathbf{nnz}(\mathbf{A}) + \frac{n}{\alpha} \cdot d^{\omega})$.

Extension to $p \geq 1$:

- Guarantee: $\sigma_p(\mathbf{a}_i) \leq \widetilde{\sigma}_i \leq O\left(\alpha^{p-1}\sigma_p(\mathbf{a}_i) + \frac{\alpha^p}{n}\mathfrak{S}_p(\mathbf{A})\right)$
- Cost: $O\left(\mathrm{nnz}(\mathbf{A}) + \frac{n}{\alpha} \cdot \mathcal{C}(d^{p/2}, d, p)\right)$

First Result: Estimating All ℓ_p Sensitivities

Theorem 1: Estimating All ℓ_1 Sensitivities

Given a full-rank $\mathbf{A} \in \mathbb{R}^{m \times n}$ and $\alpha \geq 1$, we compute a vector $\widetilde{\sigma} \in \mathbb{R}^n$ such that, with high probability, for each $i \in [n]$,

$$\sigma_1(\mathbf{a}_i) \leq \widetilde{\sigma}_i \leq O(\sigma_1(\mathbf{a}_i) + \tfrac{\alpha}{n}\mathfrak{S}_1(\mathbf{A})).$$

Our runtime is $\widetilde{O}(\mathbf{nnz}(\mathbf{A}) + \frac{n}{\alpha} \cdot d^\omega)$.

Extension to $p \geq 1$:

- Guarantee: $\sigma_p(\mathbf{a}_i) \leq \widetilde{\sigma}_i \leq O\left(\alpha^{p-1}\sigma_p(\mathbf{a}_i) + \frac{\alpha^p}{n}\mathfrak{S}_p(\mathbf{A})\right)$
- Cost: $O\left(\mathrm{nnz}(\mathbf{A}) + \frac{n}{\alpha} \cdot \mathcal{C}(d^{p/2}, d, p)\right)$

cost of ℓ_p regression on a $d^{p/2} \times d$ matrix

Our Algorithm: All ℓ_1 Sensitivities

1. Compute $\mathbf{SA} \in \mathbb{R}^{\widetilde{O}(d)\times d}$, an ℓ_1 subspace embedding of $\mathbf{A}$
2. Partition $\mathbf{A}$ into $\frac{n}{\alpha}$ random blocks $\mathbf{B}_1, \mathbf{B}_2, \ldots, \mathbf{B}_{n/\alpha}$
3. Hash each block $\mathbf{B}_i$ into 100 rows
4. Let $\mathbf{P} \in \mathbb{R}^{100\frac{n}{\alpha}\times d}$ be the matrix of the all the n/α hashed rows from step 4. Compute $\sigma_1^{\mathbf{SA}}(\mathbf{P})$
5. For $i = 1, 2, 3, \ldots, n$ iterations, do:
   - Let J be the rows in $\mathbf{P}$ that $\mathbf{a}_i$ is mapped to in step 4
   - Set $\widetilde{\sigma}_i := \max_{j\in J} \sigma_1^{\mathbf{SA}}(\mathbf{p}_j)$

   $\max_{\mathbf{x}:\mathbf{Ax}\neq\mathbf{0}} \frac{|\mathbf{a}_i^\top \mathbf{x}|}{\|\mathbf{SAx}\|_1}$
6. Return $\widetilde{\sigma}$

Proof Sketch: Estimating All ℓ_1 Sensitivities

Theorem 1 (Informal): Estimating All ℓ_1 Sensitivities

Our output $\widetilde{\sigma}$ satisfies $\sigma_1(\mathbf{a}_i) \leq \widetilde{\sigma}_i \leq O(\sigma_1(\mathbf{a}_i) + \frac{\alpha}{n}\mathfrak{S}_1(\mathbf{A}))$ for all $i \in [n]$. Our runtime is $O\left(\mathrm{nnz}(\mathbf{A}) + \frac{n}{\alpha} \cdot d^{\omega}\right)$.

Proof Sketch: Estimating All ℓ_1 Sensitivities

Theorem 1 (Informal): Estimating All ℓ_1 Sensitivities

Our output $\widetilde{\sigma}$ satisfies $\sigma_1(\mathbf{a}_i) \leq \widetilde{\sigma}_i \leq O(\sigma_1(\mathbf{a}_i) + \frac{\alpha}{n}\mathfrak{S}_1(\mathbf{A}))$ for all $i \in [n]$. Our runtime is $O\left(\mathrm{nnz}(\mathbf{A}) + \frac{n}{\alpha} \cdot d^{\omega}\right)$.

Proof. We have,

Proof Sketch: Estimating All ℓ_1 Sensitivities

Theorem 1 (Informal): Estimating All ℓ_1 Sensitivities

Our output $\widetilde{\sigma}$ satisfies $\sigma_1(\mathbf{a}_i) \leq \widetilde{\sigma}_i \leq O(\sigma_1(\mathbf{a}_i) + \frac{\alpha}{n}\mathfrak{S}_1(\mathbf{A}))$ for all $i \in [n]$. Our runtime is $O\left(\mathrm{nnz}(\mathbf{A}) + \frac{n}{\alpha} \cdot d^{\omega}\right)$.

Proof. We have,

$$\sigma_1^{\mathbf{SA}}(\mathbf{p}_j) = \max_{\mathbf{x} \in \mathbb{R}^d} \frac{|\mathbf{r}_k^\top \mathbf{B}_\ell \mathbf{x}|}{\|\mathbf{SAx}\|_1}$$

by definition

Proof Sketch: Estimating All ℓ_1 Sensitivities

Theorem 1 (Informal): Estimating All ℓ_1 Sensitivities

Our output $\widetilde{\sigma}$ satisfies $\sigma_1(\mathbf{a}_i) \leq \widetilde{\sigma}_i \leq O(\sigma_1(\mathbf{a}_i) + \frac{\alpha}{n}\mathfrak{S}_1(\mathbf{A}))$ for all $i \in [n]$. Our runtime is $O\left(\mathrm{nnz}(\mathbf{A}) + \frac{n}{\alpha} \cdot d^\omega\right)$.

Proof. We have,

$$\sigma_1^{\mathbf{SA}}(\mathbf{p}_j) = \max_{\mathbf{x}\in\mathbb{R}^d} \frac{|\mathbf{r}_k^\top \mathbf{B}_\ell \mathbf{x}|}{\|\mathbf{SAx}\|_1} \geq \frac{|\mathbf{r}_k^\top \mathbf{B}_\ell \mathbf{x}^*|}{\|\mathbf{SAx}^*\|_1}$$

specific choice of $\mathbf{x}^*$

Proof Sketch: Estimating All ℓ_1 Sensitivities

Theorem 1 (Informal): Estimating All ℓ_1 Sensitivities

Our output $\widetilde{\sigma}$ satisfies $\sigma_1(\mathbf{a}_i) \leq \widetilde{\sigma}_i \leq O(\sigma_1(\mathbf{a}_i) + \frac{\alpha}{n}\mathfrak{S}_1(\mathbf{A}))$ for all $i \in [n]$. Our runtime is $O\left(\mathrm{nnz}(\mathbf{A}) + \frac{n}{\alpha} \cdot d^\omega\right)$.

Proof. We have,

$$\sigma_1^{\mathbf{SA}}(\mathbf{p}_j) = \max_{\mathbf{x}\in\mathbb{R}^d} \frac{|\mathbf{r}_k^\top \mathbf{B}_\ell \mathbf{x}|}{\|\mathbf{SAx}\|_1} \geq \frac{|\mathbf{r}_k^\top \mathbf{B}_\ell \mathbf{x}^*|}{\|\mathbf{SAx}^*\|_1} \approx \frac{|\mathbf{r}_k^\top \mathbf{B}_\ell \mathbf{x}^*|}{\|\mathbf{Ax}^*\|_1}$$

since $\mathbf{SA}$ is a subspace embedding

Proof Sketch: Estimating All ℓ_1 Sensitivities

Theorem 1 (Informal): Estimating All ℓ_1 Sensitivities

Our output $\widetilde{\sigma}$ satisfies $\sigma_1(\mathbf{a}_i) \leq \widetilde{\sigma}_i \leq O(\sigma_1(\mathbf{a}_i) + \frac{\alpha}{n}\mathfrak{S}_1(\mathbf{A}))$ for all $i \in [n]$. Our runtime is $O\left(\mathrm{nnz}(\mathbf{A}) + \frac{n}{\alpha} \cdot d^\omega\right)$.

Proof. We have,

$$\sigma_1^{\mathbf{SA}}(\mathbf{p}_j) = \max_{\mathbf{x}\in\mathbb{R}^d} \frac{|\mathbf{r}_k^\top \mathbf{B}_\ell \mathbf{x}|}{\|\mathbf{SAx}\|_1} \geq \frac{|\mathbf{r}_k^\top \mathbf{B}_\ell \mathbf{x}^*|}{\|\mathbf{SAx}^*\|_1} \approx \frac{|\mathbf{r}_k^\top \mathbf{B}_\ell \mathbf{x}^*|}{\|\mathbf{Ax}^*\|_1} \geq \Theta(1) \frac{|\mathbf{a}_i^\top \mathbf{x}^*|}{\|\mathbf{Ax}^*\|_1}$$

$$|\mathbf{r}_k^\top \mathbf{B}_\ell \mathbf{x}^*| = |\mathbf{r}_{k,i} \mathbf{a}_i^\top \mathbf{x}^* + \textstyle\sum_{j\neq i} \mathbf{r}_{k,j} (\mathbf{B}_\ell \mathbf{x}^*)_j|$$

$\geq |\mathbf{a}_i^\top \mathbf{x}^*|$ with a probability at least $1/2$

Proof Sketch: Estimating All ℓ_1 Sensitivities

Theorem 1 (Informal): Estimating All ℓ_1 Sensitivities

Our output $\widetilde{\sigma}$ satisfies $\sigma_1(\mathbf{a}_i) \leq \widetilde{\sigma}_i \leq O(\sigma_1(\mathbf{a}_i) + \frac{\alpha}{n}\mathfrak{S}_1(\mathbf{A}))$ for all $i \in [n]$. Our runtime is $O\left(\mathrm{nnz}(\mathbf{A}) + \frac{n}{\alpha} \cdot d^\omega\right)$.

Proof. We have,

$$\sigma_1^{\mathbf{SA}}(\mathbf{p}_j) = \max_{\mathbf{x}\in\mathbb{R}^d} \frac{|\mathbf{r}_k^\top \mathbf{B}_\ell \mathbf{x}|}{\|\mathbf{SAx}\|_1} \geq \frac{|\mathbf{r}_k^\top \mathbf{B}_\ell \mathbf{x}^*|}{\|\mathbf{SAx}^*\|_1} \approx \frac{|\mathbf{r}_k^\top \mathbf{B}_\ell \mathbf{x}^*|}{\|\mathbf{Ax}^*\|_1} \geq \Theta(1)\frac{|\mathbf{a}_i^\top \mathbf{x}^*|}{\|\mathbf{Ax}^*\|_1} \geq \sigma_1(\mathbf{a}_i)$$

by definition

Proof Sketch: Estimating All ℓ_1 Sensitivities

Theorem 1 (Informal): Estimating All ℓ_1 Sensitivities

Our output $\widetilde{\sigma}$ satisfies $\sigma_1(\mathbf{a}_i) \leq \widetilde{\sigma}_i \leq O(\sigma_1(\mathbf{a}_i) + \frac{\alpha}{n}\mathfrak{S}_1(\mathbf{A}))$ for all $i \in [n]$. Our runtime is $O\left(\mathrm{nnz}(\mathbf{A}) + \frac{n}{\alpha} \cdot d^{\omega}\right)$.

Proof. We have,

$$\sigma_1^{\mathbf{SA}}(\mathbf{p}_j) = \max_{\mathbf{x} \in \mathbb{R}^d} \frac{|\mathbf{r}_k^\top \mathbf{B}_\ell \mathbf{x}|}{\|\mathbf{SAx}\|_1} \geq \frac{|\mathbf{r}_k^\top \mathbf{B}_\ell \mathbf{x}^*|}{\|\mathbf{SAx}^*\|_1} \approx \frac{|\mathbf{r}_k^\top \mathbf{B}_\ell \mathbf{x}^*|}{\|\mathbf{Ax}^*\|_1} \geq \Theta(1) \frac{|\mathbf{a}_i^\top \mathbf{x}^*|}{\|\mathbf{Ax}^*\|_1} \geq \sigma_1(\mathbf{a}_i)$$

Proof Sketch: Estimating All ℓ_1 Sensitivities

Theorem 1 (Informal): Estimating All ℓ_1 Sensitivities

Our output $\widetilde{\sigma}$ satisfies $\sigma_1(\mathbf{a}_i) \leq \widetilde{\sigma}_i \leq O(\sigma_1(\mathbf{a}_i) + \frac{\alpha}{n}\mathfrak{S}_1(\mathbf{A}))$ for all $i \in [n]$. Our runtime is $O\left(\mathrm{nnz}(\mathbf{A}) + \frac{n}{\alpha} \cdot d^\omega\right)$.

Proof. We have,

$$\sigma_1^{\mathbf{SA}}(\mathbf{p}_j) = \max_{\mathbf{x}\in\mathbb{R}^d} \frac{|\mathbf{r}_k^\top \mathbf{B}_\ell \mathbf{x}|}{\|\mathbf{SAx}\|_1} \geq \frac{|\mathbf{r}_k^\top \mathbf{B}_\ell \mathbf{x}^*|}{\|\mathbf{SAx}^*\|_1} \approx \frac{|\mathbf{r}_k^\top \mathbf{B}_\ell \mathbf{x}^*|}{\|\mathbf{Ax}^*\|_1} \geq \Theta(1) \frac{|\mathbf{a}_i^\top \mathbf{x}^*|}{\|\mathbf{Ax}^*\|_1} \geq \sigma_1(\mathbf{a}_i)$$

$$\sigma_1^{\mathbf{SA}}(\mathbf{p}_j) = \max_{\mathbf{x}\in\mathbb{R}^d} \frac{|\mathbf{r}_k^\top \mathbf{B}_\ell \mathbf{x}|}{\|\mathbf{SAx}\|_1}$$

by definition

Proof Sketch: Estimating All ℓ_1 Sensitivities

Theorem 1 (Informal): Estimating All ℓ_1 Sensitivities

Our output $\widetilde{\sigma}$ satisfies $\sigma_1(\mathbf{a}_i) \leq \widetilde{\sigma}_i \leq O(\sigma_1(\mathbf{a}_i) + \frac{\alpha}{n}\mathfrak{S}_1(\mathbf{A}))$ for all $i \in [n]$. Our runtime is $O\left(\mathrm{nnz}(\mathbf{A}) + \frac{n}{\alpha} \cdot d^{\omega}\right)$.

Proof. We have,

$$\sigma_1^{\mathbf{SA}}(\mathbf{p}_j) = \max_{\mathbf{x}\in\mathbb{R}^d} \frac{|\mathbf{r}_k^\top \mathbf{B}_\ell \mathbf{x}|}{\|\mathbf{SAx}\|_1} \geq \frac{|\mathbf{r}_k^\top \mathbf{B}_\ell \mathbf{x}^*|}{\|\mathbf{SAx}^*\|_1} \approx \frac{|\mathbf{r}_k^\top \mathbf{B}_\ell \mathbf{x}^*|}{\|\mathbf{Ax}^*\|_1} \geq \Theta(1) \frac{|\mathbf{a}_i^\top \mathbf{x}^*|}{\|\mathbf{Ax}^*\|_1} \geq \sigma_1(\mathbf{a}_i)$$

$$\sigma_1^{\mathbf{SA}}(\mathbf{p}_j) = \max_{\mathbf{x}\in\mathbb{R}^d} \frac{|\mathbf{r}_k^\top \mathbf{B}_\ell \mathbf{x}|}{\|\mathbf{SAx}\|_1} \leq \max_{\mathbf{x}\in\mathbb{R}^d} \frac{\|\mathbf{B}_\ell \mathbf{x}\|_1}{\|\mathbf{SAx}\|_1}$$

Hölder inequality

Proof Sketch: Estimating All ℓ_1 Sensitivities

Theorem 1 (Informal): Estimating All ℓ_1 Sensitivities

Our output $\widetilde{\sigma}$ satisfies $\sigma_1(\mathbf{a}_i) \leq \widetilde{\sigma}_i \leq O(\sigma_1(\mathbf{a}_i) + \frac{\alpha}{n}\mathfrak{S}_1(\mathbf{A}))$ for all $i \in [n]$. Our runtime is $O\left(\mathrm{nnz}(\mathbf{A}) + \frac{n}{\alpha} \cdot d^{\omega}\right)$.

Proof. We have,

$$\sigma_1^{\mathbf{SA}}(\mathbf{p}_j) = \max_{\mathbf{x}\in\mathbb{R}^d} \frac{|\mathbf{r}_k^\top \mathbf{B}_\ell \mathbf{x}|}{\|\mathbf{SAx}\|_1} \geq \frac{|\mathbf{r}_k^\top \mathbf{B}_\ell \mathbf{x}^*|}{\|\mathbf{SAx}^*\|_1} \approx \frac{|\mathbf{r}_k^\top \mathbf{B}_\ell \mathbf{x}^*|}{\|\mathbf{Ax}^*\|_1} \geq \Theta(1) \frac{|\mathbf{a}_i^\top \mathbf{x}^*|}{\|\mathbf{Ax}^*\|_1} \geq \sigma_1(\mathbf{a}_i)$$

$$\sigma_1^{\mathbf{SA}}(\mathbf{p}_j) = \max_{\mathbf{x}\in\mathbb{R}^d} \frac{|\mathbf{r}_k^\top \mathbf{B}_\ell \mathbf{x}|}{\|\mathbf{SAx}\|_1} \leq \max_{\mathbf{x}\in\mathbb{R}^d} \frac{\|\mathbf{B}_\ell \mathbf{x}\|_1}{\|\mathbf{SAx}\|_1} \approx \max_{\mathbf{x}\in\mathbb{R}^d} \frac{\|\mathbf{B}_\ell \mathbf{x}\|_1}{\|\mathbf{Ax}\|_1}$$

since $\mathbf{SA}$ is a subspace embedding

Proof Sketch: Estimating All ℓ_1 Sensitivities

Theorem 1 (Informal): Estimating All ℓ_1 Sensitivities

Our output $\widetilde{\sigma}$ satisfies $\sigma_1(\mathbf{a}_i) \leq \widetilde{\sigma}_i \leq O(\sigma_1(\mathbf{a}_i) + \frac{\alpha}{n}\mathfrak{S}_1(\mathbf{A}))$ for all $i \in [n]$. Our runtime is $O\left(\mathrm{nnz}(\mathbf{A}) + \frac{n}{\alpha} \cdot d^{\omega}\right)$.

Proof. We have,

$$\sigma_1^{\mathbf{SA}}(\mathbf{p}_j) = \max_{\mathbf{x}\in\mathbb{R}^d} \frac{|\mathbf{r}_k^\top \mathbf{B}_\ell \mathbf{x}|}{\|\mathbf{SAx}\|_1} \geq \frac{|\mathbf{r}_k^\top \mathbf{B}_\ell \mathbf{x}^*|}{\|\mathbf{SAx}^*\|_1} \approx \frac{|\mathbf{r}_k^\top \mathbf{B}_\ell \mathbf{x}^*|}{\|\mathbf{Ax}^*\|_1} \geq \Theta(1) \frac{|\mathbf{a}_i^\top \mathbf{x}^*|}{\|\mathbf{Ax}^*\|_1} \geq \sigma_1(\mathbf{a}_i)$$

$$\sigma_1^{\mathbf{SA}}(\mathbf{p}_j) = \max_{\mathbf{x}\in\mathbb{R}^d} \frac{|\mathbf{r}_k^\top \mathbf{B}_\ell \mathbf{x}|}{\|\mathbf{SAx}\|_1} \leq \max_{\mathbf{x}\in\mathbb{R}^d} \frac{\|\mathbf{B}_\ell \mathbf{x}\|_1}{\|\mathbf{SAx}\|_1} \approx \max_{\mathbf{x}\in\mathbb{R}^d} \frac{\|\mathbf{B}_\ell \mathbf{x}\|_1}{\|\mathbf{Ax}\|_1} \leq \sum_{j:\mathbf{a}_j \in \mathbf{B}_\ell} \sigma_1(\mathbf{a}_i)$$

expanding $\|\mathbf{B}_\ell \mathbf{x}\|_1$

Proof Sketch: Estimating All ℓ_1 Sensitivities

Theorem 1 (Informal): Estimating All ℓ_1 Sensitivities

Our output $\widetilde{\sigma}$ satisfies $\sigma_1(\mathbf{a}_i) \leq \widetilde{\sigma}_i \leq O(\sigma_1(\mathbf{a}_i) + \frac{\alpha}{n}\mathfrak{S}_1(\mathbf{A}))$ for all $i \in [n]$. Our runtime is $O\left(\text{nnz}(\mathbf{A}) + \frac{n}{\alpha} \cdot d^{\omega}\right)$.

Proof. We have,

$$\sigma_1^{\mathbf{SA}}(\mathbf{p}_j) = \max_{\mathbf{x}\in\mathbb{R}^d} \frac{|\mathbf{r}_k^\top \mathbf{B}_\ell \mathbf{x}|}{\|\mathbf{SAx}\|_1} \geq \frac{|\mathbf{r}_k^\top \mathbf{B}_\ell \mathbf{x}^*|}{\|\mathbf{SAx}^*\|_1} \approx \frac{|\mathbf{r}_k^\top \mathbf{B}_\ell \mathbf{x}^*|}{\|\mathbf{Ax}^*\|_1} \geq \Theta(1) \frac{|\mathbf{a}_i^\top \mathbf{x}^*|}{\|\mathbf{Ax}^*\|_1} \geq \sigma_1(\mathbf{a}_i)$$

$$\sigma_1^{\mathbf{SA}}(\mathbf{p}_j) = \max_{\mathbf{x}\in\mathbb{R}^d} \frac{|\mathbf{r}_k^\top \mathbf{B}_\ell \mathbf{x}|}{\|\mathbf{SAx}\|_1} \leq \max_{\mathbf{x}\in\mathbb{R}^d} \frac{\|\mathbf{B}_\ell \mathbf{x}\|_1}{\|\mathbf{SAx}\|_1} \approx \max_{\mathbf{x}\in\mathbb{R}^d} \frac{\|\mathbf{B}_\ell \mathbf{x}\|_1}{\|\mathbf{Ax}\|_1} \leq \sum_{j:\mathbf{a}_j\in\mathbf{B}_\ell} \sigma_1(\mathbf{a}_i)$$

apply Markov to finish

Proof Sketch: Estimating All ℓ_1 Sensitivities

Theorem 1 (Informal): Estimating All ℓ_1 Sensitivities

Our output $\widetilde{\sigma}$ satisfies $\sigma_1(\mathbf{a}_i) \leq \widetilde{\sigma}_i \leq O(\sigma_1(\mathbf{a}_i) + \frac{\alpha}{n}\mathfrak{S}_1(\mathbf{A}))$ for all $i \in [n]$. Our runtime is $O\left(\text{nnz}(\mathbf{A}) + \frac{n}{\alpha} \cdot d^\omega\right)$.

Proof. We have,

$$\sigma_1^{\mathbf{SA}}(\mathbf{p}_j) = \max_{\mathbf{x}\in\mathbb{R}^d} \frac{|\mathbf{r}_k^\top \mathbf{B}_\ell \mathbf{x}|}{\|\mathbf{SAx}\|_1} \geq \frac{|\mathbf{r}_k^\top \mathbf{B}_\ell \mathbf{x}^*|}{\|\mathbf{SAx}^*\|_1} \approx \frac{|\mathbf{r}_k^\top \mathbf{B}_\ell \mathbf{x}^*|}{\|\mathbf{Ax}^*\|_1} \geq \Theta(1) \frac{|\mathbf{a}_i^\top \mathbf{x}^*|}{\|\mathbf{Ax}^*\|_1} \geq \sigma_1(\mathbf{a}_i)$$

$$\sigma_1^{\mathbf{SA}}(\mathbf{p}_j) = \max_{\mathbf{x}\in\mathbb{R}^d} \frac{|\mathbf{r}_k^\top \mathbf{B}_\ell \mathbf{x}|}{\|\mathbf{SAx}\|_1} \leq \max_{\mathbf{x}\in\mathbb{R}^d} \frac{\|\mathbf{B}_\ell \mathbf{x}\|_1}{\|\mathbf{SAx}\|_1} \approx \max_{\mathbf{x}\in\mathbb{R}^d} \frac{\|\mathbf{B}_\ell \mathbf{x}\|_1}{\|\mathbf{Ax}\|_1} \leq \sum_{j:\mathbf{a}_j \in \mathbf{B}_\ell} \sigma_1(\mathbf{a}_i)$$

Runtime: cost of computing n/α sensitivities w.r.t. $\mathbf{SA} \in \mathbb{R}^{d\times d}$.

Estimating All ℓ_p Sensitivities: Key Takeaway

- Our computed sensitivities are *approximate*

Estimating All ℓ_p Sensitivities: Key Takeaway

- Our computed sensitivities are *approximate*
- Still, when compared to *true* sensitivities, they preserve ℓ_p regression approximation guarantees well enough while increasing sample complexity by only a small amount

Estimating All ℓ_p Sensitivities: Key Takeaway

- Our computed sensitivities are *approximate*
- Still, when compared to *true* sensitivities, they preserve ℓ_p regression approximation guarantees well enough while increasing sample complexity by only a small amount
- Further, the increased sample complexity (due to *approximate* sensitivities) is still much lower than that due to Lewis weights

II. Approximating the Sum of ℓ_p Sensitivities

Estimating the Total ℓ_p Sensitivity

Theorem 2: Estimating Total ℓ_p Sensitivity

Given a full-rank $\mathbf{A} \in \mathbb{R}^{m \times n}$ and $\gamma \in (0,1)$, we compute a scalar $\widetilde{\sigma}$ such that, with high probability,

$$\mathfrak{S}_p(\mathbf{A}) \leq \widetilde{\sigma} \leq (1 + O(\gamma))\mathfrak{S}_p(\mathbf{A}).$$

Our runtime is $\widetilde{O}\left(\mathbf{nnz}(\mathbf{A}) + \frac{1}{\gamma^2} \cdot d^{|1-p/2|} \cdot \mathcal{C}(d^{\max(1,p/2)}, d, p)\right)$.

Estimating the Total ℓ_p Sensitivity

Theorem 2: Estimating Total ℓ_p Sensitivity

Given a full-rank $\mathbf{A} \in \mathbb{R}^{m \times n}$ and $\gamma \in (0,1)$, we compute a scalar $\widetilde{\sigma}$ such that, with high probability,

$$\mathfrak{S}_p(\mathbf{A}) \leq \widetilde{\sigma} \leq (1 + O(\gamma))\mathfrak{S}_p(\mathbf{A}).$$

Our runtime is $\widetilde{O}\left(\mathbf{nnz}(\mathbf{A}) + \frac{1}{\gamma^2} \cdot d^{|1-p/2|} \cdot \mathcal{C}(d^{\max(1,p/2)}, d, p)\right)$.

- Techniques used: importance sampling of ℓ_p Lewis weights

Estimating the Total ℓ_p Sensitivity

Theorem 2: Estimating Total ℓ_p Sensitivity

Given a full-rank $\mathbf{A} \in \mathbb{R}^{m \times n}$ and $\gamma \in (0,1)$, we compute a scalar $\widetilde{\sigma}$ such that, with high probability,

$$\mathfrak{S}_p(\mathbf{A}) \leq \widetilde{\sigma} \leq (1 + O(\gamma))\mathfrak{S}_p(\mathbf{A}).$$

Our runtime is $\widetilde{O}\left(\mathbf{nnz}(\mathbf{A}) + \frac{1}{\gamma^2} \cdot d^{|1-p/2|} \cdot \mathcal{C}(d^{\max(1,p/2)}, d, p)\right)$.

- Techniques used: importance sampling of ℓ_p Lewis weights
- For $p = 1$, we have a recursive algorithm using only leverage scores

Our Algorithm: Total ℓ_p Sensitivity

1. Compute $\mathbf{w}_p(\mathbf{A})$, the ℓ_p Lewis weights of $\mathbf{A}$
2. Define the sampling vector $v \in \mathbb{R}^n_{\geq 0}$ such that $v_i = \frac{\mathbf{w}_p(\mathbf{a}_i)}{d}$
3. Sample $m = O(d^{|1-p/2|})$ rows with replacement, where we pick the i^{th} row with a probability of v_i
4. Construct an ℓ_p sampling matrix $\mathbf{S}_p\mathbf{A}$ with $\{v_i\}_{i=1}^n$
5. For each sampled row i_j (where $j \in [m]$)
   - Compute $r_j = \frac{\sigma_p^{\mathbf{S}_p\mathbf{A}}(\mathbf{A})}{v_{i_j}}$
6. Return $\frac{1}{m}\sum_{j=1}^m r_j$

Theorem 2 (Informal): Estimating Total ℓ_p Sensitivity

Our output $\widetilde{\sigma}$ satisfies $\mathfrak{S}_p(\mathbf{A}) \leq \widetilde{\sigma} \leq (1 + O(\gamma))\mathfrak{S}_p(\mathbf{A}))$. Our runtime is $O\left(\mathrm{nnz}(\mathbf{A}) + \frac{1}{\gamma^2} \cdot d^{|1-p/2|} \cdot \mathcal{C}(d^{\max(1,p/2)}, d, p)\right)$.

Theorem 2 (Informal): Estimating Total ℓ_p Sensitivity

Our output $\widetilde{\sigma}$ satisfies $\mathfrak{S}_p(\mathbf{A}) \leq \widetilde{\sigma} \leq (1+O(\gamma))\mathfrak{S}_p(\mathbf{A}))$. Our runtime is $O\left(\mathrm{nnz}(\mathbf{A}) + \frac{1}{\gamma^2} \cdot d^{|1-p/2|} \cdot \mathcal{C}(d^{\max(1,p/2)}, d, p)\right)$.

Proof. Our estimate is unbiased; the variance satisfies:

Theorem 2 (Informal): Estimating Total ℓ_p Sensitivity

Our output $\widetilde{\sigma}$ satisfies $\mathfrak{S}_p(\mathbf{A}) \leq \widetilde{\sigma} \leq (1 + O(\gamma))\mathfrak{S}_p(\mathbf{A}))$. Our runtime is $O\left(\mathrm{nnz}(\mathbf{A}) + \frac{1}{\gamma^2} \cdot d^{|1-p/2|} \cdot \mathcal{C}(d^{\max(1,p/2)}, d, p)\right)$.

Proof. Our estimate is unbiased; the variance satisfies:

$$\mathrm{Var}\left(\frac{1}{m}\sum_{j\in[m]} r_j\right) \leq \frac{1}{m}\sum_{i=1}^{n} \frac{\sigma_p^{\mathbf{S}_p\mathbf{A}}(\mathbf{a}_i)^2}{v_i}$$

by definition

Theorem 2 (Informal): Estimating Total ℓ_p Sensitivity

Our output $\widetilde{\sigma}$ satisfies $\mathfrak{S}_p(\mathbf{A}) \leq \widetilde{\sigma} \leq (1 + O(\gamma))\mathfrak{S}_p(\mathbf{A}))$. Our runtime is $O\left(\mathrm{nnz}(\mathbf{A}) + \frac{1}{\gamma^2} \cdot d^{|1-p/2|} \cdot \mathcal{C}(d^{\max(1,p/2)}, d, p)\right)$.

Proof. Our estimate is unbiased; the variance satisfies:

$$\mathrm{Var}\left(\frac{1}{m}\sum_{j\in[m]} r_j\right) \leq \frac{1}{m}\sum_{i=1}^{n}\frac{\sigma_p^{\mathbf{S}_p\mathbf{A}}(\mathbf{a}_i)^2}{v_i} = \frac{d}{m}\cdot\sum_{i=1}^{n}\frac{\sigma_p^{\mathbf{S}_p\mathbf{A}}(\mathbf{a}_i)^2}{\mathbf{w}_p(\mathbf{a}_i)}$$

our choice of v_i

Theorem 2 (Informal): Estimating Total ℓ_p Sensitivity

Our output $\tilde{\sigma}$ satisfies $\mathfrak{S}_p(\mathbf{A}) \leq \tilde{\sigma} \leq (1 + O(\gamma))\mathfrak{S}_p(\mathbf{A}))$. Our runtime is $O\left(\mathrm{nnz}(\mathbf{A}) + \frac{1}{\gamma^2} \cdot d^{|1-p/2|} \cdot \mathcal{C}(d^{\max(1,p/2)}, d, p)\right)$.

Proof. Our estimate is unbiased; the variance satisfies:

$$\mathrm{Var}\left(\frac{1}{m}\sum_{j\in[m]} r_j\right) \leq \frac{1}{m}\sum_{i=1}^{n} \frac{\sigma_p^{\mathbf{S}_p\mathbf{A}}(\mathbf{a}_i)^2}{v_i} = \frac{d}{m} \cdot \sum_{i=1}^{n} \frac{\sigma_p^{\mathbf{S}_p\mathbf{A}}(\mathbf{a}_i)^2}{\mathbf{w}_p(\mathbf{a}_i)}$$

When $p \geq 2$, we have

$$\frac{d}{m} \cdot \sum_{i=1}^{n} \frac{\sigma_p^{\mathbf{S}_p\mathbf{A}}(\mathbf{a}_i)^2}{\mathbf{w}_p(\mathbf{a}_i)} \leq \frac{d}{m} \cdot \sum_{i=1}^{n} \sigma_p^{\mathbf{S}_p\mathbf{A}}(\mathbf{a}_i) \cdot d^{p/2-1}$$

$\sigma_p(\mathbf{a}_i) \leq d^{p/2-1}\mathbf{w}_p(\mathbf{a}_i)$

Theorem 2 (Informal): Estimating Total ℓ_p Sensitivity

Our output $\widetilde{\sigma}$ satisfies $\mathfrak{S}_p(\mathbf{A}) \leq \widetilde{\sigma} \leq (1 + O(\gamma))\mathfrak{S}_p(\mathbf{A}))$. Our runtime is $O\left(\mathrm{nnz}(\mathbf{A}) + \frac{1}{\gamma^2} \cdot d^{|1-p/2|} \cdot \mathcal{C}(d^{\max(1,p/2)}, d, p)\right)$.

Proof. Our estimate is unbiased; the variance satisfies:

$$\mathrm{Var}\left(\frac{1}{m}\sum_{j\in[m]} r_j\right) \leq \frac{1}{m}\sum_{i=1}^{n} \frac{\sigma_p^{\mathbf{S}_p\mathbf{A}}(\mathbf{a}_i)^2}{v_i} = \frac{d}{m} \cdot \sum_{i=1}^{n} \frac{\sigma_p^{\mathbf{S}_p\mathbf{A}}(\mathbf{a}_i)^2}{\mathbf{w}_p(\mathbf{a}_i)}$$

When $p \geq 2$, we have

$$\frac{d}{m} \cdot \sum_{i=1}^{n} \frac{\sigma_p^{\mathbf{S}_p\mathbf{A}}(\mathbf{a}_i)^2}{\mathbf{w}_p(\mathbf{a}_i)} \leq \frac{d}{m} \cdot \sum_{i=1}^{n} \sigma_p^{\mathbf{S}_p\mathbf{A}}(\mathbf{a}_i) \cdot d^{p/2-1} \leq \frac{d^{p/2}}{m} \mathfrak{S}_p^{\mathbf{S}_p\mathbf{A}}(\mathbf{A})$$

by definition

Theorem 2 (Informal): Estimating Total ℓ_p Sensitivity

Our output $\widetilde{\sigma}$ satisfies $\mathfrak{S}_p(\mathbf{A}) \leq \widetilde{\sigma} \leq (1 + O(\gamma))\mathfrak{S}_p(\mathbf{A}))$. Our runtime is $O\left(\mathrm{nnz}(\mathbf{A}) + \frac{1}{\gamma^2} \cdot d^{|1-p/2|} \cdot \mathcal{C}(d^{\max(1,p/2)}, d, p)\right)$.

Proof. Our estimate is unbiased; the variance satisfies:

$$\mathrm{Var}\left(\frac{1}{m}\sum_{j\in[m]} r_j\right) \leq \frac{1}{m}\sum_{i=1}^{n} \frac{\sigma_p^{\mathbf{S}_p\mathbf{A}}(\mathbf{a}_i)^2}{v_i} = \frac{d}{m} \cdot \sum_{i=1}^{n} \frac{\sigma_p^{\mathbf{S}_p\mathbf{A}}(\mathbf{a}_i)^2}{\mathbf{w}_p(\mathbf{a}_i)}$$

When $p \geq 2$, we have

$$\frac{d}{m} \cdot \sum_{i=1}^{n} \frac{\sigma_p^{\mathbf{S}_p\mathbf{A}}(\mathbf{a}_i)^2}{\mathbf{w}_p(\mathbf{a}_i)} \leq \frac{d}{m} \cdot \sum_{i=1}^{n} \sigma_p^{\mathbf{S}_p\mathbf{A}}(\mathbf{a}_i) \cdot d^{p/2-1} \leq \frac{d^{p/2}}{m}\mathfrak{S}_p^{\mathbf{S}_p\mathbf{A}}(\mathbf{A})$$

Theorem 2 (Informal): Estimating Total ℓ_p Sensitivity

Our output $\widetilde{\sigma}$ satisfies $\mathfrak{S}_p(\mathbf{A}) \leq \widetilde{\sigma} \leq (1 + O(\gamma))\mathfrak{S}_p(\mathbf{A}))$. Our runtime is $O\left(\operatorname{nnz}(\mathbf{A}) + \frac{1}{\gamma^2} \cdot d^{|1-p/2|} \cdot \mathcal{C}(d^{\max(1,p/2)}, d, p)\right)$.

Proof. Our estimate is unbiased; the variance satisfies:

$$\operatorname{Var}\left(\frac{1}{m}\sum_{j\in[m]} r_j\right) \leq \frac{1}{m}\sum_{i=1}^{n}\frac{\sigma_p^{\mathbf{S}_p\mathbf{A}}(\mathbf{a}_i)^2}{v_i} = \frac{d}{m}\cdot\sum_{i=1}^{n}\frac{\sigma_p^{\mathbf{S}_p\mathbf{A}}(\mathbf{a}_i)^2}{\mathbf{w}_p(\mathbf{a}_i)}$$

When $p \geq 2$, we have

$$\frac{d}{m}\cdot\sum_{i=1}^{n}\frac{\sigma_p^{\mathbf{S}_p\mathbf{A}}(\mathbf{a}_i)^2}{\mathbf{w}_p(\mathbf{a}_i)} \leq \frac{d}{m}\cdot\sum_{i=1}^{n}\sigma_p^{\mathbf{S}_p\mathbf{A}}(\mathbf{a}_i)\cdot d^{p/2-1} \leq \frac{d^{p/2}}{m}\mathfrak{S}_p^{\mathbf{S}_p\mathbf{A}}(\mathbf{A})$$

By Chebyshev, pick $m = d^{p/2}/\mathfrak{S}_p^{\mathbf{S}_p\mathbf{A}}(\mathbf{A}) \approx d^{p/2}/\mathfrak{S}_p(\mathbf{A}) \leq d^{p/2-1}$.

Theorem 2 (Informal): Estimating Total ℓ_p Sensitivity

Our output $\widetilde{\sigma}$ satisfies $\mathfrak{S}_p(\mathbf{A}) \leq \widetilde{\sigma} \leq (1 + O(\gamma))\mathfrak{S}_p(\mathbf{A}))$. Our runtime is $O\left(\mathrm{nnz}(\mathbf{A}) + \frac{1}{\gamma^2} \cdot d^{|1-p/2|} \cdot \mathcal{C}(d^{\max(1,p/2)}, d, p)\right)$.

Proof. Our estimate is unbiased; the variance satisfies:

$$\mathrm{Var}\left(\frac{1}{m}\sum_{j\in[m]} r_j\right) \leq \frac{1}{m}\sum_{i=1}^{n} \frac{\sigma_p^{\mathbf{S}_p\mathbf{A}}(\mathbf{a}_i)^2}{v_i} = \frac{d}{m} \cdot \sum_{i=1}^{n} \frac{\sigma_p^{\mathbf{S}_p\mathbf{A}}(\mathbf{a}_i)^2}{\mathbf{w}_p(\mathbf{a}_i)}$$

When $p \geq 2$, we have

$$\frac{d}{m} \cdot \sum_{i=1}^{n} \frac{\sigma_p^{\mathbf{S}_p\mathbf{A}}(\mathbf{a}_i)^2}{\mathbf{w}_p(\mathbf{a}_i)} \leq \frac{d}{m} \cdot \sum_{i=1}^{n} \sigma_p^{\mathbf{S}_p\mathbf{A}}(\mathbf{a}_i) \cdot d^{p/2-1} \leq \frac{d^{p/2}}{m} \mathfrak{S}_p^{\mathbf{S}_p\mathbf{A}}(\mathbf{A})$$

for $p \geq 2$, we have $\mathfrak{S}_p(\mathbf{A}) \geq d$

By Chebyshev, pick $m = d^{p/2}/\mathfrak{S}_p^{\mathbf{S}_p\mathbf{A}}(\mathbf{A}) \approx d^{p/2}/\mathfrak{S}_p(\mathbf{A}) \leq d^{p/2-1}$.

Theorem 2 (Informal): Estimating Total ℓ_p Sensitivity

Our output $\widetilde{\sigma}$ satisfies $\mathfrak{S}_p(\mathbf{A}) \leq \widetilde{\sigma} \leq (1+O(\gamma))\mathfrak{S}_p(\mathbf{A}))$. Our runtime is $O\left(\mathrm{nnz}(\mathbf{A}) + \frac{1}{\gamma^2} \cdot d^{|1-p/2|} \cdot \mathcal{C}(d^{\max(1,p/2)}, d, p)\right)$.

Proof. Our estimate is unbiased; the variance satisfies:

$$\mathrm{Var}\left(\frac{1}{m}\sum_{j\in[m]} r_j\right) \leq \frac{1}{m}\sum_{i=1}^{n} \frac{\sigma_p^{\mathbf{S}_p\mathbf{A}}(\mathbf{a}_i)^2}{v_i} = \frac{d}{m}\cdot\sum_{i=1}^{n}\frac{\sigma_p^{\mathbf{S}_p\mathbf{A}}(\mathbf{a}_i)^2}{\mathbf{w}_p(\mathbf{a}_i)}$$

When $p \geq 2$, we have

$$\frac{d}{m}\cdot\sum_{i=1}^{n}\frac{\sigma_p^{\mathbf{S}_p\mathbf{A}}(\mathbf{a}_i)^2}{\mathbf{w}_p(\mathbf{a}_i)} \leq \frac{d}{m}\cdot\sum_{i=1}^{n}\sigma_p^{\mathbf{S}_p\mathbf{A}}(\mathbf{a}_i)\cdot d^{p/2-1} \leq \frac{d^{p/2}}{m}\mathfrak{S}_p^{\mathbf{S}_p\mathbf{A}}(\mathbf{A})$$

Runtime: cost of Lewis weights and $d^{p/2-1}$ sensitivities w.r.t. $\mathbf{S}_p\mathbf{A}$

Estimating Total ℓ_p Sensitivity: Key Takeaway

- We can compute the total sensitivity up to a constant accuracy by only $\mathrm{poly}(d)$ sensitivity computations
- Our main technique is importance sampling using Lewis weights, which are in turn cheap to compute

III. Approximating the Maximum of ℓ_p Sensitivities

Estimating the Maximum ℓ_p Sensitivity

Theorem 3: Estimating Maximum ℓ_p Sensitivity

Given a full-rank $\mathbf{A} \in \mathbb{R}^{m \times n}$, we compute a scalar $\widetilde{\sigma}$ such that

$$\|\sigma_p(\mathbf{A})\|_\infty \leq \widetilde{\sigma} \leq O(\sqrt{d}\|\sigma_1(\mathbf{A})\|_\infty.$$

Our runtime is $\widetilde{O}\left(\mathbf{nnz}(\mathbf{A}) + d^{\max(1,p/2)} \cdot \mathcal{C}(d^{\max(1,p/2)}, d, p)\right)$.

Estimating the Maximum ℓ_p Sensitivity

Theorem 3: Estimating Maximum ℓ_p Sensitivity

Given a full-rank $\mathbf{A} \in \mathbb{R}^{m \times n}$, we compute a scalar $\widetilde{\sigma}$ such that

$$\|\sigma_p(\mathbf{A})\|_\infty \le \widetilde{\sigma} \le O(\sqrt{d}\|\sigma_1(\mathbf{A})\|_\infty.$$

Our runtime is $\widetilde{O}\left(\mathbf{nnz}(\mathbf{A}) + d^{\max(1,p/2)} \cdot \mathcal{C}(d^{\max(1,p/2)}, d, p)\right)$.

- Key technique used: new results in sample-efficient ℓ_∞ subspace embeddings

Our Algorithm: Maximum ℓ_1 Sensitivity

1. Compute an ℓ_∞ subspace embedding $\mathbf{S}_\infty\mathbf{A}$ such that it is a subset of rows of $\mathbf{A}$ (Woodruff & Yasuda (2022))
2. Compute an ℓ_1 subspace embedding $\mathbf{S}_1\mathbf{A}$ of $\mathbf{A}$
3. Return $\sqrt{d}\|\sigma_1^{\mathbf{S}_1\mathbf{A}}(\mathbf{S}_\infty\mathbf{A})\|_\infty$

Theorem 3 (Informal): Estimating Maximum ℓ_1 Sensitivity

Our output $\widetilde{\sigma}$ satisfies $\|\sigma_1(\mathbf{A})\|_\infty \leq \widetilde{\sigma} \leq O(\sqrt{d})\|\sigma_1(\mathbf{A})\|_\infty$. Our runtime is $O\left(\mathrm{nnz}(\mathbf{A}) + d^{\omega+1}\right)$.

Theorem 3 (Informal): Estimating Maximum ℓ_1 Sensitivity

Our output $\widetilde{\sigma}$ satisfies $\|\sigma_1(\mathbf{A})\|_\infty \leq \widetilde{\sigma} \leq O(\sqrt{d})\|\sigma_1(\mathbf{A})\|_\infty$. Our runtime is $O\left(\mathrm{nnz}(\mathbf{A}) + d^{\omega+1}\right)$.

Proof. Let $\mathbf{x}^*, i^* := \arg\max_{\mathbf{x}, i \in [n]} \frac{|\mathbf{a}_i^\top \mathbf{x}|}{\|\mathbf{A}\mathbf{x}\|_1}$. Suppose $\mathbf{a}_{i^*} \notin \mathbf{S}_\infty \mathbf{A}$. Then,

Theorem 3 (Informal): Estimating Maximum ℓ_1 Sensitivity

Our output $\widetilde{\sigma}$ satisfies $\|\sigma_1(\mathbf{A})\|_\infty \leq \widetilde{\sigma} \leq O(\sqrt{d})\|\sigma_1(\mathbf{A})\|_\infty$. Our runtime is $O\left(\mathrm{nnz}(\mathbf{A}) + d^{\omega+1}\right)$.

Proof. Let $\mathbf{x}^*, i^* := \arg\max_{\mathbf{x}, i \in [n]} \frac{|\mathbf{a}_i^\top \mathbf{x}|}{\|\mathbf{A}\mathbf{x}\|_1}$. Suppose $\mathbf{a}_{i^*} \notin \mathbf{S}_\infty \mathbf{A}$. Then,

$$\|\sigma_1^{\mathbf{S}_1\mathbf{A}}(\mathbf{S}_\infty \mathbf{A})\|_\infty = \max_{\substack{\mathbf{x} \in \mathbb{R}^d, \\ \mathbf{c}_j \in \mathbf{S}_\infty \mathbf{A}}} \frac{|\mathbf{c}_j^\top \mathbf{x}|}{\|\mathbf{S}_1 \mathbf{A}\mathbf{x}\|_1}$$

by definition over $\mathbf{c}_j \in \mathbf{S}_\infty \mathbf{A}$

Theorem 3 (Informal): Estimating Maximum ℓ_1 Sensitivity

Our output $\widetilde{\sigma}$ satisfies $\|\sigma_1(\mathbf{A})\|_\infty \leq \widetilde{\sigma} \leq O(\sqrt{d})\|\sigma_1(\mathbf{A})\|_\infty$. Our runtime is $O\left(\mathrm{nnz}(\mathbf{A}) + d^{\omega+1}\right)$.

Proof. Let $\mathbf{x}^*, i^* := \arg\max_{\mathbf{x}, i \in [n]} \frac{|\mathbf{a}_i^\top \mathbf{x}|}{\|\mathbf{A}\mathbf{x}\|_1}$. Suppose $\mathbf{a}_{i^*} \notin \mathbf{S}_\infty \mathbf{A}$. Then,

$$\|\sigma_1^{\mathbf{S}_1\mathbf{A}}(\mathbf{S}_\infty\mathbf{A})\|_\infty = \max_{\substack{\mathbf{x}\in\mathbb{R}^d, \\ \mathbf{c}_j \in \mathbf{S}_\infty \mathbf{A}}} \frac{|\mathbf{c}_j^\top \mathbf{x}|}{\|\mathbf{S}_1\mathbf{A}\mathbf{x}\|_1} \geq \max_{\mathbf{x}\in\mathbb{R}^d} \frac{\|\mathbf{S}_\infty\mathbf{A}\mathbf{x}\|_\infty}{\|\mathbf{S}_1\mathbf{A}\mathbf{x}\|_1}$$

specific choice of numerator

Theorem 3 (Informal): Estimating Maximum ℓ_1 Sensitivity

Our output $\widetilde{\sigma}$ satisfies $\|\sigma_1(\mathbf{A})\|_\infty \leq \widetilde{\sigma} \leq O(\sqrt{d})\|\sigma_1(\mathbf{A})\|_\infty$. Our runtime is $O\left(\operatorname{nnz}(\mathbf{A}) + d^{\omega+1}\right)$.

Proof. Let $\mathbf{x}^*, i^* := \arg\max_{\mathbf{x}, i \in [n]} \frac{|\mathbf{a}_i^\top \mathbf{x}|}{\|\mathbf{A}\mathbf{x}\|_1}$. Suppose $\mathbf{a}_{i^*} \notin \mathbf{S}_\infty \mathbf{A}$. Then,

$$\|\sigma_1^{\mathbf{S}_1 \mathbf{A}}(\mathbf{S}_\infty \mathbf{A})\|_\infty = \max_{\substack{\mathbf{x} \in \mathbb{R}^d, \\ \mathbf{c}_j \in \mathbf{S}_\infty \mathbf{A}}} \frac{|\mathbf{c}_j^\top \mathbf{x}|}{\|\mathbf{S}_1 \mathbf{A}\mathbf{x}\|_1} \geq \max_{\mathbf{x} \in \mathbb{R}^d} \frac{\|\mathbf{S}_\infty \mathbf{A}\mathbf{x}\|_\infty}{\|\mathbf{S}_1 \mathbf{A}\mathbf{x}\|_1} \geq \max_{\mathbf{x} \in \mathbb{R}^d} \frac{\|\mathbf{S}_\infty \mathbf{A}\mathbf{x}\|_\infty}{\|\mathbf{A}\mathbf{x}\|_1}$$

since $\mathbf{S}_1 \mathbf{A}$ is an ℓ_1 subspace embedding

Theorem 3 (Informal): Estimating Maximum ℓ_1 Sensitivity

Our output $\widetilde{\sigma}$ satisfies $\|\sigma_1(\mathbf{A})\|_\infty \leq \widetilde{\sigma} \leq O(\sqrt{d})\|\sigma_1(\mathbf{A})\|_\infty$. Our runtime is $O\left(\mathrm{nnz}(\mathbf{A}) + d^{\omega+1}\right)$.

Proof. Let $\mathbf{x}^*, i^* := \arg\max_{\mathbf{x}, i\in[n]} \frac{|\mathbf{a}_i^\top \mathbf{x}|}{\|\mathbf{A}\mathbf{x}\|_1}$. Suppose $\mathbf{a}_{i^*} \notin \mathbf{S}_\infty \mathbf{A}$. Then,

$$\|\sigma_1^{\mathbf{S}_1\mathbf{A}}(\mathbf{S}_\infty \mathbf{A})\|_\infty = \max_{\substack{\mathbf{x}\in\mathbb{R}^d,\\ \mathbf{c}_j\in\mathbf{S}_\infty\mathbf{A}}} \frac{|\mathbf{c}_j^\top \mathbf{x}|}{\|\mathbf{S}_1\mathbf{A}\mathbf{x}\|_1} \geq \max_{\mathbf{x}\in\mathbb{R}^d} \frac{\|\mathbf{S}_\infty\mathbf{A}\mathbf{x}\|_\infty}{\|\mathbf{S}_1\mathbf{A}\mathbf{x}\|_1} \geq \max_{\mathbf{x}\in\mathbb{R}^d} \frac{\|\mathbf{S}_\infty\mathbf{A}\mathbf{x}\|_\infty}{\|\mathbf{A}\mathbf{x}\|_1}$$

$$\max_{\mathbf{x}\in\mathbb{R}^d} \frac{\|\mathbf{S}_\infty\mathbf{A}\mathbf{x}\|_\infty}{\|\mathbf{A}\mathbf{x}\|_1} \geq \frac{\|\mathbf{S}_\infty\mathbf{A}\mathbf{x}^*\|_\infty}{\|\mathbf{A}\mathbf{x}^*\|_1}$$

specific choice of $\mathbf{x}$

Theorem 3 (Informal): Estimating Maximum ℓ_1 Sensitivity

Our output $\widetilde{\sigma}$ satisfies $\|\sigma_1(\mathbf{A})\|_\infty \leq \widetilde{\sigma} \leq O(\sqrt{d})\|\sigma_1(\mathbf{A})\|_\infty$. Our runtime is $O\left(\mathrm{nnz}(\mathbf{A}) + d^{\omega+1}\right)$.

Proof. Let $\mathbf{x}^*, i^* := \arg\max_{\mathbf{x}, i \in [n]} \frac{|\mathbf{a}_i^\top \mathbf{x}|}{\|\mathbf{A}\mathbf{x}\|_1}$. Suppose $\mathbf{a}_{i^*} \notin \mathbf{S}_\infty \mathbf{A}$. Then,

$$\|\sigma_1^{\mathbf{S}_1\mathbf{A}}(\mathbf{S}_\infty \mathbf{A})\|_\infty = \max_{\substack{\mathbf{x}\in\mathbb{R}^d,\\ \mathbf{c}_j \in \mathbf{S}_\infty \mathbf{A}}} \frac{|\mathbf{c}_j^\top \mathbf{x}|}{\|\mathbf{S}_1\mathbf{A}\mathbf{x}\|_1} \geq \max_{\mathbf{x}\in\mathbb{R}^d} \frac{\|\mathbf{S}_\infty\mathbf{A}\mathbf{x}\|_\infty}{\|\mathbf{S}_1\mathbf{A}\mathbf{x}\|_1} \geq \max_{\mathbf{x}\in\mathbb{R}^d} \frac{\|\mathbf{S}_\infty\mathbf{A}\mathbf{x}\|_\infty}{\|\mathbf{A}\mathbf{x}\|_1}$$

$$\max_{\mathbf{x}\in\mathbb{R}^d} \frac{\|\mathbf{S}_\infty\mathbf{A}\mathbf{x}\|_\infty}{\|\mathbf{A}\mathbf{x}\|_1} \geq \frac{\|\mathbf{S}_\infty\mathbf{A}\mathbf{x}^*\|_\infty}{\|\mathbf{A}\mathbf{x}^*\|_1} \geq \frac{\|\mathbf{A}\mathbf{x}^*\|_\infty}{\sqrt{d}\|\mathbf{A}\mathbf{x}^*\|_1}$$

$\mathbf{S}_\infty$ is an ℓ_∞ subspace embedding

Theorem 3 (Informal): Estimating Maximum ℓ_1 Sensitivity

Our output $\widetilde{\sigma}$ satisfies $\|\sigma_1(\mathbf{A})\|_\infty \leq \widetilde{\sigma} \leq O(\sqrt{d})\|\sigma_1(\mathbf{A})\|_\infty$. Our runtime is $O\left(\text{nnz}(\mathbf{A}) + d^{\omega+1}\right)$.

Proof. Let $\mathbf{x}^*, i^* := \arg\max_{\mathbf{x}, i \in [n]} \frac{|\mathbf{a}_i^\top \mathbf{x}|}{\|\mathbf{A}\mathbf{x}\|_1}$. Suppose $\mathbf{a}_{i^*} \notin \mathbf{S}_\infty \mathbf{A}$. Then,

$$\|\sigma_1^{\mathbf{S}_1\mathbf{A}}(\mathbf{S}_\infty \mathbf{A})\|_\infty = \max_{\substack{\mathbf{x}\in\mathbb{R}^d, \\ \mathbf{c}_j \in \mathbf{S}_\infty \mathbf{A}}} \frac{|\mathbf{c}_j^\top \mathbf{x}|}{\|\mathbf{S}_1\mathbf{A}\mathbf{x}\|_1} \geq \max_{\mathbf{x}\in\mathbb{R}^d} \frac{\|\mathbf{S}_\infty\mathbf{A}\mathbf{x}\|_\infty}{\|\mathbf{S}_1\mathbf{A}\mathbf{x}\|_1} \geq \max_{\mathbf{x}\in\mathbb{R}^d} \frac{\|\mathbf{S}_\infty\mathbf{A}\mathbf{x}\|_\infty}{\|\mathbf{A}\mathbf{x}\|_1}$$

$$\max_{\mathbf{x}\in\mathbb{R}^d} \frac{\|\mathbf{S}_\infty\mathbf{A}\mathbf{x}\|_\infty}{\|\mathbf{A}\mathbf{x}\|_1} \geq \frac{\|\mathbf{S}_\infty\mathbf{A}\mathbf{x}^*\|_\infty}{\|\mathbf{A}\mathbf{x}^*\|_1} \geq \frac{\|\mathbf{A}\mathbf{x}^*\|_\infty}{\sqrt{d}\|\mathbf{A}\mathbf{x}^*\|_1} \geq \frac{|\mathbf{a}_{i^*}^\top \mathbf{x}^*|}{\sqrt{d}\|\mathbf{A}\mathbf{x}^*\|_1}$$

since $a_{i^*} \in \mathbf{A}$

Theorem 3 (Informal): Estimating Maximum ℓ_1 Sensitivity

Our output $\widetilde{\sigma}$ satisfies $\|\sigma_1(\mathbf{A})\|_\infty \leq \widetilde{\sigma} \leq O(\sqrt{d})\|\sigma_1(\mathbf{A})\|_\infty$. Our runtime is $O\left(\mathrm{nnz}(\mathbf{A}) + d^{\omega+1}\right)$.

Proof. Let $\mathbf{x}^*, i^* := \arg\max_{\mathbf{x}, i \in [n]} \frac{|\mathbf{a}_i^\top \mathbf{x}|}{\|\mathbf{A}\mathbf{x}\|_1}$. Suppose $\mathbf{a}_{i^*} \notin \mathbf{S}_\infty \mathbf{A}$. Then,

$$\|\sigma_1^{\mathbf{S}_1\mathbf{A}}(\mathbf{S}_\infty\mathbf{A})\|_\infty = \max_{\substack{\mathbf{x}\in\mathbb{R}^d,\\ \mathbf{c}_j\in\mathbf{S}_\infty\mathbf{A}}} \frac{|\mathbf{c}_j^\top \mathbf{x}|}{\|\mathbf{S}_1\mathbf{A}\mathbf{x}\|_1} \geq \max_{\mathbf{x}\in\mathbb{R}^d} \frac{\|\mathbf{S}_\infty\mathbf{A}\mathbf{x}\|_\infty}{\|\mathbf{S}_1\mathbf{A}\mathbf{x}\|_1} \geq \max_{\mathbf{x}\in\mathbb{R}^d} \frac{\|\mathbf{S}_\infty\mathbf{A}\mathbf{x}\|_\infty}{\|\mathbf{A}\mathbf{x}\|_1}$$

$$\max_{\mathbf{x}\in\mathbb{R}^d} \frac{\|\mathbf{S}_\infty\mathbf{A}\mathbf{x}\|_\infty}{\|\mathbf{A}\mathbf{x}\|_1} \geq \frac{\|\mathbf{S}_\infty\mathbf{A}\mathbf{x}^*\|_\infty}{\|\mathbf{A}\mathbf{x}^*\|_1} \geq \frac{\|\mathbf{A}\mathbf{x}^*\|_\infty}{\sqrt{d}\|\mathbf{A}\mathbf{x}^*\|_1} \geq \frac{|\mathbf{a}_{i^*}^\top \mathbf{x}^*|}{\sqrt{d}\|\mathbf{A}\mathbf{x}^*\|_1} \geq \frac{1}{\sqrt{d}}\|\sigma_1(\mathbf{A})\|_\infty$$

by definition of i^*

Theorem 3 (Informal): Estimating Maximum ℓ_1 Sensitivity

Our output $\widetilde{\sigma}$ satisfies $\|\sigma_1(\mathbf{A})\|_\infty \leq \widetilde{\sigma} \leq O(\sqrt{d})\|\sigma_1(\mathbf{A})\|_\infty$. Our runtime is $O\left(\mathrm{nnz}(\mathbf{A}) + d^{\omega+1}\right)$.

Proof. Let $\mathbf{x}^*, i^* := \arg\max_{\mathbf{x}, i\in[n]} \frac{|\mathbf{a}_i^\top \mathbf{x}|}{\|\mathbf{A}\mathbf{x}\|_1}$. Suppose $\mathbf{a}_{i^*} \notin \mathbf{S}_\infty\mathbf{A}$. Then,

$$\|\sigma_1^{\mathbf{S}_1\mathbf{A}}(\mathbf{S}_\infty\mathbf{A})\|_\infty = \max_{\substack{\mathbf{x}\in\mathbb{R}^d,\\ \mathbf{c}_j\in\mathbf{S}_\infty\mathbf{A}}} \frac{|\mathbf{c}_j^\top \mathbf{x}|}{\|\mathbf{S}_1\mathbf{A}\mathbf{x}\|_1} \geq \max_{\mathbf{x}\in\mathbb{R}^d} \frac{\|\mathbf{S}_\infty\mathbf{A}\mathbf{x}\|_\infty}{\|\mathbf{S}_1\mathbf{A}\mathbf{x}\|_1} \geq \max_{\mathbf{x}\in\mathbb{R}^d} \frac{\|\mathbf{S}_\infty\mathbf{A}\mathbf{x}\|_\infty}{\|\mathbf{A}\mathbf{x}\|_1}$$

$$\max_{\mathbf{x}\in\mathbb{R}^d} \frac{\|\mathbf{S}_\infty\mathbf{A}\mathbf{x}\|_\infty}{\|\mathbf{A}\mathbf{x}\|_1} \geq \frac{\|\mathbf{S}_\infty\mathbf{A}\mathbf{x}^*\|_\infty}{\|\mathbf{A}\mathbf{x}^*\|_1} \geq \frac{\|\mathbf{A}\mathbf{x}^*\|_\infty}{\sqrt{d}\|\mathbf{A}\mathbf{x}^*\|_1} \geq \frac{|\mathbf{a}_{i^*}^\top \mathbf{x}^*|}{\sqrt{d}\|\mathbf{A}\mathbf{x}^*\|_1} \geq \frac{1}{\sqrt{d}}\|\sigma_1(\mathbf{A})\|_\infty$$

Theorem 3 (Informal): Estimating Maximum ℓ_1 Sensitivity

Our output $\widetilde{\sigma}$ satisfies $\|\sigma_1(\mathbf{A})\|_\infty \leq \widetilde{\sigma} \leq O(\sqrt{d})\|\sigma_1(\mathbf{A})\|_\infty$. Our runtime is $O\left(\mathrm{nnz}(\mathbf{A}) + d^{\omega+1}\right)$.

Proof. Let $\mathbf{x}^*, i^* := \arg\max_{\mathbf{x}, i \in [n]} \frac{|\mathbf{a}_i^\top \mathbf{x}|}{\|\mathbf{A}\mathbf{x}\|_1}$. Suppose $\mathbf{a}_{i^*} \notin \mathbf{S}_\infty \mathbf{A}$. Then,

$$\|\sigma_1^{\mathbf{S}_1\mathbf{A}}(\mathbf{S}_\infty \mathbf{A})\|_\infty = \max_{\substack{\mathbf{x}\in\mathbb{R}^d,\\ \mathbf{c}_j \in \mathbf{S}_\infty\mathbf{A}}} \frac{|\mathbf{c}_j^\top \mathbf{x}|}{\|\mathbf{S}_1\mathbf{A}\mathbf{x}\|_1} \geq \max_{\mathbf{x}\in\mathbb{R}^d} \frac{\|\mathbf{S}_\infty\mathbf{A}\mathbf{x}\|_\infty}{\|\mathbf{S}_1\mathbf{A}\mathbf{x}\|_1} \geq \max_{\mathbf{x}\in\mathbb{R}^d} \frac{\|\mathbf{S}_\infty\mathbf{A}\mathbf{x}\|_\infty}{\|\mathbf{A}\mathbf{x}\|_1}$$

$$\max_{\mathbf{x}\in\mathbb{R}^d} \frac{\|\mathbf{S}_\infty\mathbf{A}\mathbf{x}\|_\infty}{\|\mathbf{A}\mathbf{x}\|_1} \geq \frac{\|\mathbf{S}_\infty\mathbf{A}\mathbf{x}^*\|_\infty}{\|\mathbf{A}\mathbf{x}^*\|_1} \geq \frac{\|\mathbf{A}\mathbf{x}^*\|_\infty}{\sqrt{d}\|\mathbf{A}\mathbf{x}^*\|_1} \geq \frac{|\mathbf{a}_{i^*}^\top\mathbf{x}^*|}{\sqrt{d}\|\mathbf{A}\mathbf{x}^*\|_1} \geq \frac{1}{\sqrt{d}}\|\sigma_1(\mathbf{A})\|_\infty$$

Runtime: cost of computing $\mathbf{S}_\infty$ and d of ℓ_1 sensitivities w.r.t. $\mathbf{S}_1\mathbf{A}$.

Concluding Thoughts

- Can we efficiently approximate sensitivities for other functions?

Thank You!