# Multiplicative Rank-1 Approximation using Length-Squared Sampling

Ragesh Jaiswal

CSE, IIT Delhi

SOSA Talk, January 06, 2020

[Joint work with Amit Kumar (IIT Delhi)]

# Main Result

## Main result (informal)

The span of $\Omega(\frac{1}{\varepsilon^4})$ rows of any matrix $A \subset \mathbb{R}^{n \times d}$ sampled according to the length-squared distribution, contains a rank-1 matrix that gives a $(1 + \varepsilon)$ **multiplicative** approximation under the Frobenius norm.

- Length-squared distribution: The probability of sampling the $i^{th}$ row, $\mathbf{A}^{(i)}$, is proportional to $\|\mathbf{A}^{(i)}\|^2$.

# Main Result

**Main result (informal)**

The span of $\Omega(\frac{1}{\varepsilon^4})$ rows of any matrix $A \subset \mathbb{R}^{n \times d}$ sampled according to the length-squared distribution, contains a rank-1 matrix that gives a $(1+\varepsilon)$ **multiplicative** approximation under the Frobenius norm.

- Length-squared distribution: The probability of sampling the $i^{th}$ row, $\mathbf{A}^{(i)}$, is proportional to $\|\mathbf{A}^{(i)}\|^2$.

**Theorem (Main Theorem)**

*Let $0 < \varepsilon < 1$. Let $S$ be a sample of $s$ rows of an $n \times d$ matrix $A$, each chosen independently with length-squared distribution. If $s = \Omega(\frac{1}{\varepsilon^4})$, then the span of $S$ contains a matrix $\tilde{A}$ of rank-1 such that:*

$$\mathbf{E}[\|A - \tilde{A}\|_F^2] \leq (1+\varepsilon) \cdot \|A - \pi_1(A)\|_F^2, \quad \textit{where}$$

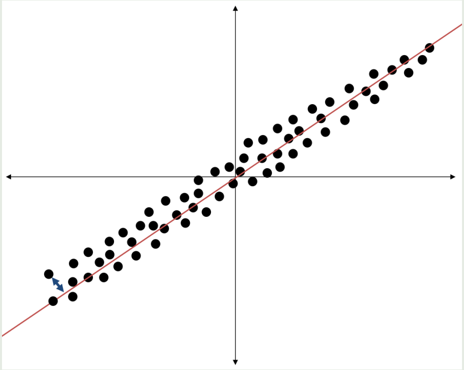
$$\pi_1(A) = \underset{X:rank(X)=1}{\operatorname{argmin}} \|A - X\|_F^2$$

# Discussion: Best Fit Line

## Main result (informal)

The span of $\Omega(\frac{1}{\varepsilon^4})$ rows of any matrix $A \subset \mathbb{R}^{n \times d}$ sampled according to the length-squared distribution, contains a rank-1 matrix that gives a $(1+\varepsilon)$ **multiplicative** approximation under the Frobenius norm.

## Best Fit Line Problem (interpret rows as points)

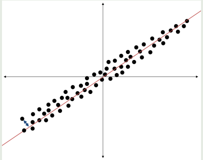


# Discussion: Best Fit Line

## Main result (informal)

The span of $\Omega(\frac{1}{\varepsilon^4})$ rows of any matrix $A \subset \mathbb{R}^{n \times d}$ sampled according to the length-squared distribution, contains a rank-1 matrix that gives a $(1+\varepsilon)$ **multiplicative** approximation under the Frobenius norm.

## Best Fit Line Problem (interpret rows as points)



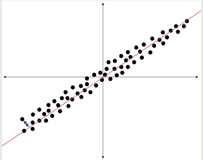
- The problem is **not** hard.
- The solution is the first singular vector and can be found using SVD.
- This discussion is about simple sampling based technique.

# Discussion: Best Fit Line

## Main result (informal)

The span of $\Omega(\frac{1}{\varepsilon^4})$ rows of any matrix $A \subset \mathbb{R}^{n \times d}$ sampled according to the length-squared distribution, contains a rank-1 matrix that gives a $(1+\varepsilon)$ **multiplicative** approximation under the Frobenius norm.

## Best Fit Line Problem (interpret rows as points)



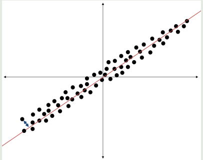
- This discussion is about simple sampling based technique.
  - Question: Can we approximate the best-fit line using a few samples?

# Discussion: Best Fit Line

## Main result (informal)

The span of $\Omega(\frac{1}{\varepsilon^4})$ rows of any matrix $A \subset \mathbb{R}^{n \times d}$ sampled according to the length-squared distribution, contains a rank-1 matrix that gives a $(1+\varepsilon)$ **multiplicative** approximation under the Frobenius norm.

## Best Fit Line Problem (interpret rows as points)



- Question: Can we approximate the best-fit line using a few samples?
  - Non-constructive: Yes using $O(\frac{1}{\varepsilon} \log \frac{1}{\varepsilon})$ samples [SV12].

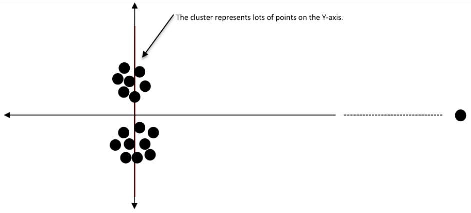
# Discussion: Best Fit Line

## Main result (informal)

The span of $\Omega(\frac{1}{\varepsilon^4})$ rows of any matrix $A \subset \mathbb{R}^{n \times d}$ sampled according to the length-squared distribution, contains a rank-1 matrix that gives a $(1+\varepsilon)$ **multiplicative** approximation under the Frobenius norm.

## Best Fit Line Problem (interpret rows as points)

- Question: Can we approximate the best-fit line using a few **uniformly** sampled points?
  - No. See example below.



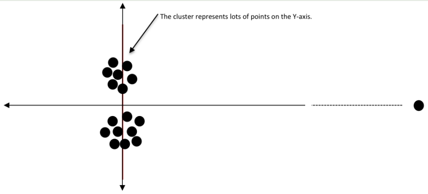
# Discussion: Best Fit Line

## Main result (informal)

The span of $\Omega(\frac{1}{\varepsilon^4})$ rows of any matrix $A \subset \mathbb{R}^{n \times d}$ sampled according to the length-squared distribution, contains a rank-1 matrix that gives a $(1+\varepsilon)$ **multiplicative** approximation under the Frobenius norm.

## Best Fit Line Problem (interpret rows as points)

- Question: Can we approximate the best-fit line using a few uniformly sampled points?
  - No. See example below.
  - This motivates length-squared sampling since the distance of a point from origin is relevant.



# Discussion: Known Results

**Main result (informal)**

The span of $\Omega(\frac{1}{\varepsilon^4})$ rows of any matrix $A \subset \mathbb{R}^{n \times d}$ sampled according to the length-squared distribution, contains a rank-1 matrix that gives a $(1+\varepsilon)$ **multiplicative** approximation under the Frobenius norm.

- Length-squared distribution: The probability of sampling the $i^{th}$ row, $\mathbf{A}^{(i)}$, is proportional to $\|\mathbf{A}^{(i)}\|^2$.
- Question: What was known about length-squared sampling in the current context?

# Discussion: Known Results

**Main result (informal)**

The span of $\Omega(\frac{1}{\varepsilon^4})$ rows of any matrix $A \subset \mathbb{R}^{n \times d}$ sampled according to the length-squared distribution, contains a rank-1 matrix that gives a $(1+\varepsilon)$ **multiplicative** approximation under the Frobenius norm.

- Length-squared distribution: The probability of sampling the $i^{th}$ row, $\mathbf{A}^{(i)}$, is proportional to $\|\mathbf{A}^{(i)}\|^2$.
- Question: What was known about length-squared sampling in the current context? Additive approximation

**Theorem (Freize, Kannan, and Vempala [FKV04])**

*Let $0 < \varepsilon < 1$. Let $S$ be a sample of $s$ rows of an $n \times d$ matrix $A$, each chosen independently with length-squared distribution. If $s = \Omega(\frac{1}{\varepsilon})$, then the span of $S$ contains a matrix $\tilde{A}$ of rank-1 such that:*

$$\mathbf{E}[\|A - \tilde{A}\|_F^2] \leq \|A - \pi_1(A)\|_F^2 \; + \; \varepsilon \cdot \|A\|_F^2.$$

# Discussion: Known Results

**Main result (informal)**

The span of $\Omega(\frac{1}{\varepsilon^4})$ rows of any matrix $A \subset \mathbb{R}^{n \times d}$ sampled according to the length-squared distribution, contains a rank-1 matrix that gives a $(1+\varepsilon)$ **multiplicative** approximation under the Frobenius norm.

- Length-squared distribution: The probability of sampling the $i^{th}$ row, $\mathbf{A}^{(i)}$, is proportional to $\|\mathbf{A}^{(i)}\|^2$.
- Question: What was known about length-squared sampling in the current context? Additive approximation
- Question: Does some other sampling technique give a multiplicative approximation?

# Discussion: Known Results

**Main result (informal)**

The span of $\Omega(\frac{1}{\varepsilon^4})$ rows of any matrix $A \subset \mathbb{R}^{n \times d}$ sampled according to the length-squared distribution, contains a rank-1 matrix that gives a $(1 + \varepsilon)$ **multiplicative** approximation under the Frobenius norm.

- Length-squared distribution: The probability of sampling the $i^{th}$ row, $\mathbf{A}^{(i)}$, is proportional to $\|\mathbf{A}^{(i)}\|^2$.
- Question: What was known about length-squared sampling in the current context? Additive approximation
- Question: Does some other sampling technique give a multiplicative approximation?
  - Yes. *Adaptive length-squared sampling* along with *volume sampling* gives multiplicative approximation with $O(\frac{1}{\varepsilon})$ samples [DV06, DRVW06].
  - The above even works for rank-$k$ approximation.

# Discussion: Our Result

**Main result (informal)**

The span of $\Omega(\frac{1}{\varepsilon^4})$ rows of any matrix $A \subset \mathbb{R}^{n\times d}$ sampled according to the length-squared distribution, contains a rank-1 matrix that gives a $(1+\varepsilon)$ **multiplicative** approximation under the Frobenius norm.

- Length-squared distribution: The probability of sampling the $i^{th}$ row, $\mathbf{A}^{(i)}$, is proportional to $\|\mathbf{A}^{(i)}\|^2$.
- Question: What was known about length-squared sampling in the current context? Additive approximation
- Question: Does some other sampling technique give a multiplicative approximation?
    - Yes. *Adaptive length-squared sampling* along with *volume sampling* gives multiplicative approximation with $O(\frac{1}{\varepsilon})$ samples [DV06, DRVW06].
    - The above even works for rank-$k$ approximation.
- **This work**: Does length-squared sampling suffice for multiplicative rank-1 approximation? Yes with $O(\frac{1}{\varepsilon^4})$ samples.

# Discussion: Main Ideas

**Theorem (Main Theorem)**

*Let $0 < \varepsilon < 1$. Let $S$ be a sample of $s$ rows of an $n \times d$ matrix $A$, each chosen independently with length-squared distribution. If $s = \Omega(\frac{1}{\varepsilon^4})$, then the span of $S$ contains a matrix $\tilde{A}$ of rank-1 such that:*

$$\mathbf{E}[\|A-\tilde{A}\|_F^2] \leq (1+\varepsilon)\cdot\|A-\pi_1(A)\|_F^2, \text{ where } \pi_1(A) = \underset{X:rank(X)=1}{\text{argmin}} \|A-X\|_F^2.$$

- Let $\sigma^2 \equiv \|\pi_1(A)\|_F^2$ and $r^2 \equiv \|A - \pi_1(A)\|_F^2$.
- By a suitable rotation, we can assume that $\pi_1(A) = \begin{pmatrix} \sigma u_1, 0, \dots, 0 \\ \vdots \\ \sigma u_2, 0, \dots, 0 \end{pmatrix}$
- We do a case analysis:
  - Case 1: $(r^2 > \varepsilon^3 \sigma^2)$: Apply additive approximation of Freize, Kannan and Vempala [FKV04].
  - Case 2: $(r^2 \leq \varepsilon^3 \sigma^2)$:

**Theorem (Freize, Kannan, and Vempala [FKV04])**

*Let $0 < \varepsilon < 1$. Let $S$ be a sample of $s$ rows of an $n \times d$ matrix $A$, each chosen independently with length-squared distribution. If $s = \Omega(\frac{1}{\varepsilon})$, then the span of $S$ contains a matrix $\tilde{A}$ of rank-1 such that:*

$$\mathbf{E}[\|A - \tilde{A}\|_F^2] \leq \|A - \pi_1(A)\|_F^2 \ + \ \varepsilon \cdot \|A\|_F^2.$$

# Discussion: Main Ideas

**Theorem (Main Theorem)**

*Let $0 < \varepsilon < 1$. Let $S$ be a sample of $s$ rows of an $n \times d$ matrix $A$, each chosen independently with length-squared distribution. If $s = \Omega(\frac{1}{\varepsilon^4})$, then the span of $S$ contains a matrix $\tilde{A}$ of rank-1 such that:*

$$\mathbf{E}[\|A - \tilde{A}\|_F^2] \leq (1+\varepsilon) \cdot \|A - \pi_1(A)\|_F^2, \text{ where } \pi_1(A) = \operatorname*{argmin}_{X:rank(X)=1} \|A - X\|_F^2.$$

- Let $\sigma^2 \equiv \|\pi_1(A)\|_F^2$ and $r^2 \equiv \|A - \pi_1(A)\|_F^2$.
- By a suitable rotation, we can assume that $\pi_1(A) = \begin{pmatrix} \sigma u_1, 0, ..., 0 \\ \vdots \\ \sigma u_2, 0, ..., 0 \end{pmatrix}$
- We do a case analysis:
  - Case 1: $(r^2 > \varepsilon^3 \sigma^2)$: Apply additive approximation of Freize, Kannan and Vempala [FKV04].
  - Case 2: $(r^2 \leq \varepsilon^3 \sigma^2)$: Lots of careful calculations!

# Possible extensions and applications

- Projective clustering: Fitting $k$, $j$-dimensional flats to a given dataset.
  - Note that fitting $k$, 0-dimensional flats is the classical $k$-means problem and fitting $k$, 1-dimensional flats is the $k$-lines problem.
  - Idea: Extending the sampling based ideas, being developed for $k$-means, to projective clustering.
  - Issues: Not clear if sampling helps in fitting even 1, 2-dimensional flat. The sampling based analysis breaks down even for fitting $k$, 1-dimensional flats (i.e., the $k$-lines problem).
- Streaming setting: Much better *sketching* based algorithms exist for rank-$k$ approximation.

# Observations and open problems

- Observation 1: A single length-squared sampled point gives a 2 factor approximation in expectation.
- Observation 2: Length-squared sampling does not work for rank-$k$ approximation for $k > 1$.
- Open question 1: Our multiplicative approximation uses $O(\frac{1}{\varepsilon^4})$ samples. Are these many samples necessary?
- Open question 2: Can we get multiplicative approximation using just adaptive length-squared sampling?

# References I


Amit Deshpande, Luis Rademacher, Santosh Vempala, and Grant Wang, *Matrix approximation and projective clustering via volume sampling*, Theory of Computing **2** (2006), no. 12, 225–247.

Amit Deshpande and Santosh Vempala, *Adaptive sampling and fast low-rank matrix approximation*, Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques (Berlin, Heidelberg) (Josep Díaz, Klaus Jansen, José D. P. Rolim, and Uri Zwick, eds.), Springer Berlin Heidelberg, 2006, pp. 292–303.

Alan Frieze, Ravi Kannan, and Santosh Vempala, *Fast monte-carlo algorithms for finding low-rank approximations*, J. ACM **51** (2004), no. 6, 1025–1041.

Nariankadu D. Shyamalkumar and Kasturi Varadarajan, *Efficient subspace approximation algorithms*, Discrete & Computational Geometry **47** (2012), no. 1, 44–63.

Thank you