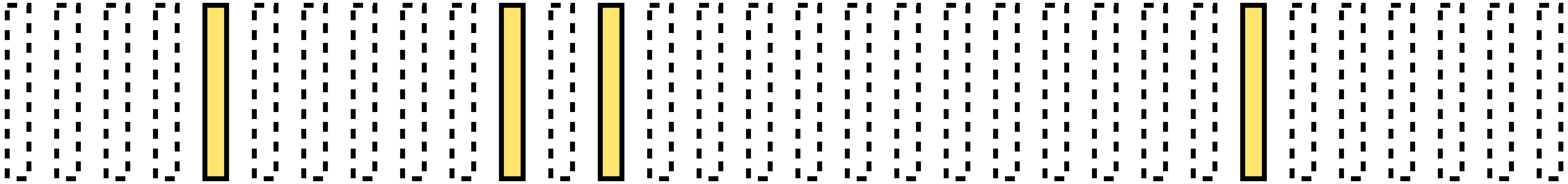


Thesis Proposal:

Advances in Algorithms for Matrix Approximation

via **Sampling** and **Sketching**

Taisuke (Tai) Yasuda

Carnegie Mellon University
Computer Science Department

April 10, 2023

Thesis Committee:

- David P. Woodruff (Carnegie Mellon University, Chair)
- Anupam Gupta (Carnegie Mellon University)
- Richard Peng (Carnegie Mellon University)
- Cameron Musco (University of Massachusetts Amherst)

Matrix Approximation

“Turning big data into tiny data”

Matrix Approximation

“Turning big data into tiny data”

- **Matrix approximation** is the problem of approximating large matrices by smaller matrices

Matrix Approximation

“Turning big data into tiny data”

- **Matrix approximation** is the problem of approximating large matrices by smaller matrices
- Modern large-scale machine learning problems deal with huge matrices!

Matrix Approximation

"Turning big data into tiny data"

- **Matrix approximation** is the problem of approximating large matrices by smaller matrices
- Modern large-scale machine learning problems deal with huge matrices!
  - Billions of training examples and labels

Matrix Approximation

"Turning big data into tiny data"

- **Matrix approximation** is the problem of approximating large matrices by smaller matrices
- Modern large-scale machine learning problems deal with huge matrices!
  - Billions of training examples and labels
  - Thousands of features

Matrix Approximation

"Turning big data into tiny data"

- **Matrix approximation** is the problem of approximating large matrices by smaller matrices
- Modern large-scale machine learning problems deal with huge matrices!
  - Billions of training examples and labels
  - Thousands of features
- Goal: replace a large dataset with a smaller dataset to **improve efficiency of data analytic tasks**

Matrix Approximation

Randomized Numerical Linear Algebra

Matrix Approximation

Randomized Numerical Linear Algebra

- **Numerical linear algebra** [pre 2000's]: deterministic algorithms for solving linear algebra to machine precision

Matrix Approximation

Randomized Numerical Linear Algebra

- **Numerical linear algebra** [pre 2000's]: deterministic algorithms for solving linear algebra to machine precision
- **Randomized numerical linear algebra** [2000's]: randomized approximation algorithms for numerical linear algebra

Matrix Approximation

Randomized Numerical Linear Algebra

- **Numerical linear algebra** [pre 2000's]: deterministic algorithms for solving linear algebra to machine precision
- **Randomized numerical linear algebra** [2000's]: randomized approximation algorithms for numerical linear algebra
  - Randomized: succeed every time → succeed with 99% probability

Matrix Approximation

Randomized Numerical Linear Algebra

- **Numerical linear algebra** [pre 2000's]: deterministic algorithms for solving linear algebra to machine precision
- **Randomized numerical linear algebra** [2000's]: randomized approximation algorithms for numerical linear algebra
  - Randomized: succeed every time → succeed with 99% probability
  - Approximation: solve exactly → solve up to (small) error

Matrix Approximation

Randomized Numerical Linear Algebra

- **Numerical linear algebra** [pre 2000's]: deterministic algorithms for solving linear algebra to machine precision
- **Randomized numerical linear algebra** [2000's]: randomized approximation algorithms for numerical linear algebra
  - Randomized: succeed every time → succeed with 99% probability
  - Approximation: solve exactly → solve up to (small) error
  - **This flexibility leads to extraordinary improvements in efficiency!**

Matrix Approximation

Randomized Numerical Linear Algebra

- **Numerical linear algebra** [pre 2000's]: deterministic algorithms for solving linear algebra to machine precision
- **Randomized numerical linear algebra** [2000's]: randomized approximation algorithms for numerical linear algebra
  - Randomized: succeed every time → succeed with 99% probability
  - Approximation: solve exactly → solve up to (small) error
  - **This flexibility leads to extraordinary improvements in efficiency!**
  - Key techniques: **sampling**, **sketching**, and **optimization**

Matrix Approximation

New Challenges

Matrix Approximation

New Challenges

1. **Tight bounds**: what are the fundamental possibilities and limitations for matrix approximation?

Matrix Approximation

New Challenges

1. **Tight bounds**: what are the fundamental possibilities and limitations for matrix approximation?
2. **Uncertain/resource-limited settings**: are there algorithms for matrix approximation using limited information or computational resources?

Matrix Approximation

New Challenges

1. **Tight bounds**: what are the fundamental possibilities and limitations for matrix approximation?
2. **Uncertain/resource-limited settings**: are there algorithms for matrix approximation using limited information or computational resources?
3. **Generalized loss functions**: are there algorithms for matrix approximation for generalized objectives and loss functions?

Matrix Approximation

New Challenges

1. **Tight bounds**: what are the fundamental possibilities and limitations for matrix approximation?
2. **Uncertain/resource-limited settings**: are there algorithms for matrix approximation using limited information or computational resources?
3. **Generalized loss functions**: are there algorithms for matrix approximation for generalized objectives and loss functions?
4. **Applications**: can techniques for matrix approximation be applied to solve problems in adjacent areas of computer science?

Subspace Embeddings and Linear Regression

Subspace Embeddings and Linear Regression

- Oblivious ℓ_p subspace embeddings: high distortion and low distortion

Subspace Embeddings and Linear Regression

- Oblivious ℓ_p subspace embeddings: high distortion and low distortion
- Non-oblivious subspace embeddings: ℓ_p Lewis weight sampling, general losses

Subspace Embeddings and Linear Regression

- Oblivious ℓ_p subspace embeddings: high distortion and low distortion
- Non-oblivious subspace embeddings: ℓ_p Lewis weight sampling, general losses
- Applications: active learning, streaming computational geometry, low rank approximation

Subspace Embeddings and Linear Regression

Oblivious ℓ_p subspace embeddings: high distortion and low distortion

- Non-oblivious subspace embeddings: ℓ_p Lewis weight sampling, general losses
- Applications: active learning, streaming computational geometry, low rank approximation

Matrix Approximation

Linear Regression

Matrix Approximation

Linear Regression

- **Linear regression**: given matrix $\mathbf{A} \in \mathbb{R}^{n \times d}$ and vector $\mathbf{b} \in \mathbb{R}^n$, solve

Matrix Approximation

Linear Regression

- **Linear regression**: given matrix $\mathbf{A} \in \mathbb{R}^{n \times d}$ and vector $\mathbf{b} \in \mathbb{R}^{n}$, solve

$$\min_{\mathbf{x} \in \mathbb{R}^{d}} \| \mathbf{A}\mathbf{x} - \mathbf{b} \|_2^2$$

Matrix Approximation

Linear Regression

- **Linear regression**: given matrix $\mathbf{A} \in \mathbb{R}^{n \times d}$ and vector $\mathbf{b} \in \mathbb{R}^{n}$, solve

$$\min_{\mathbf{x} \in \mathbb{R}^d} \| \mathbf{A}\mathbf{x} - \mathbf{b} \|_2^2$$

  - Simple model for supervised learning

Matrix Approximation

Linear Regression

- **Linear regression**: given matrix $\mathbf{A} \in \mathbb{R}^{n \times d}$ and vector $\mathbf{b} \in \mathbb{R}^n$, solve

$$\min_{\mathbf{x} \in \mathbb{R}^d} \| \mathbf{A}\mathbf{x} - \mathbf{b} \|_2^2$$

  - Simple model for supervised learning
  - Building block for complex models and algorithms

Matrix Approximation

Linear Regression

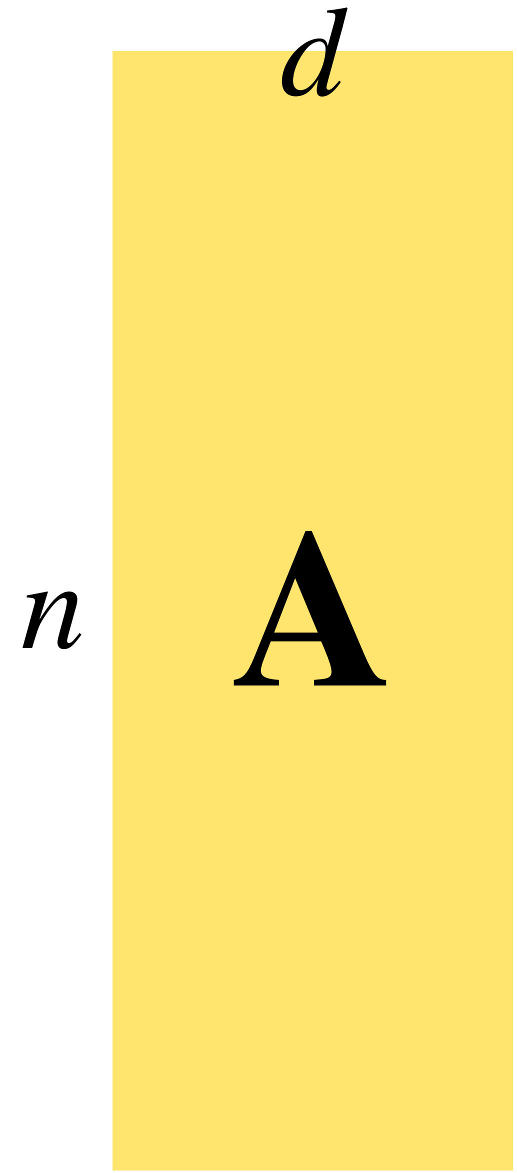
- **Linear regression**: given matrix $\mathbf{A} \in \mathbb{R}^{n \times d}$ and vector $\mathbf{b} \in \mathbb{R}^{n}$, solve

$$\min_{\mathbf{x} \in \mathbb{R}^d} \| \mathbf{A}\mathbf{x} - \mathbf{b} \|_2^2$$

  - Simple model for supervised learning
  - Building block for complex models and algorithms
  - Can we design efficient approximation algorithms for linear regression?

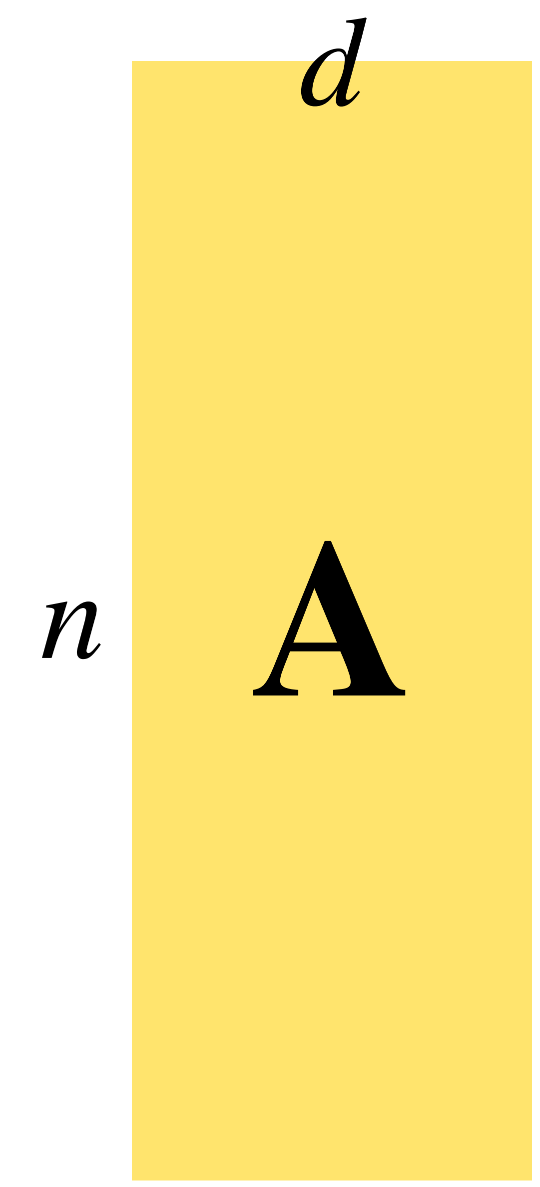
Matrix Approximation

Approximating the Linear Regression Loss Function



$$\min_{\mathbf{x}\in\mathbb{R}^d} \|\mathbf{A}\mathbf{x} - \mathbf{b}\|_2^2$$

Matrix Approximation



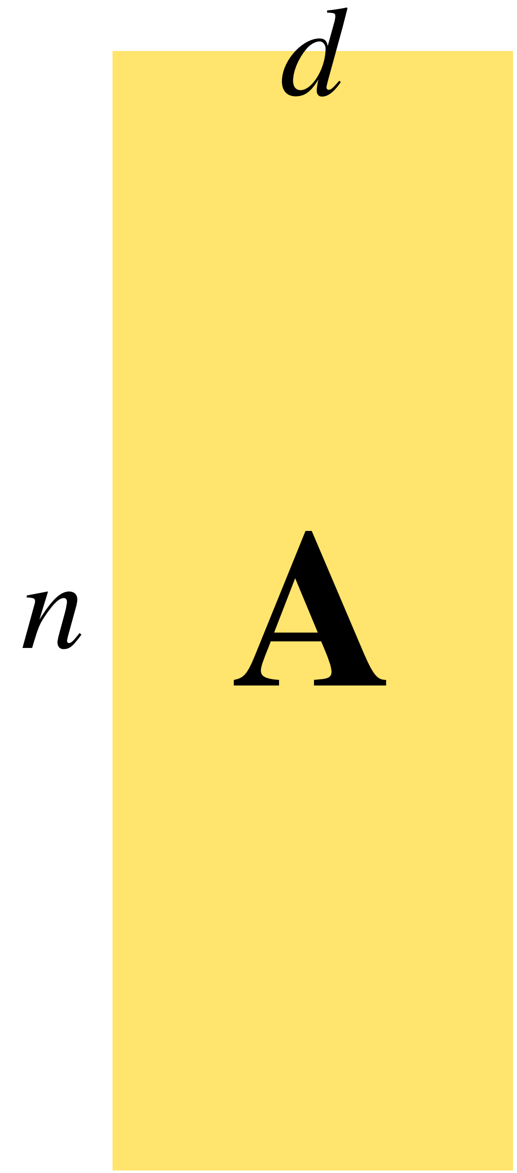
Approximating the Linear Regression Loss Function

$$\min_{\mathbf{x}\in\mathbb{R}^d} \|\mathbf{A}\mathbf{x}-\mathbf{b}\|_2^2$$

- One approach: approximate the original instance by a smaller instance

Matrix Approximation

Approximating the Linear Regression Loss Function



$$\min_{\mathbf{x}\in\mathbb{R}^d} \|\mathbf{A}\mathbf{x}-\mathbf{b}\|_2^2$$

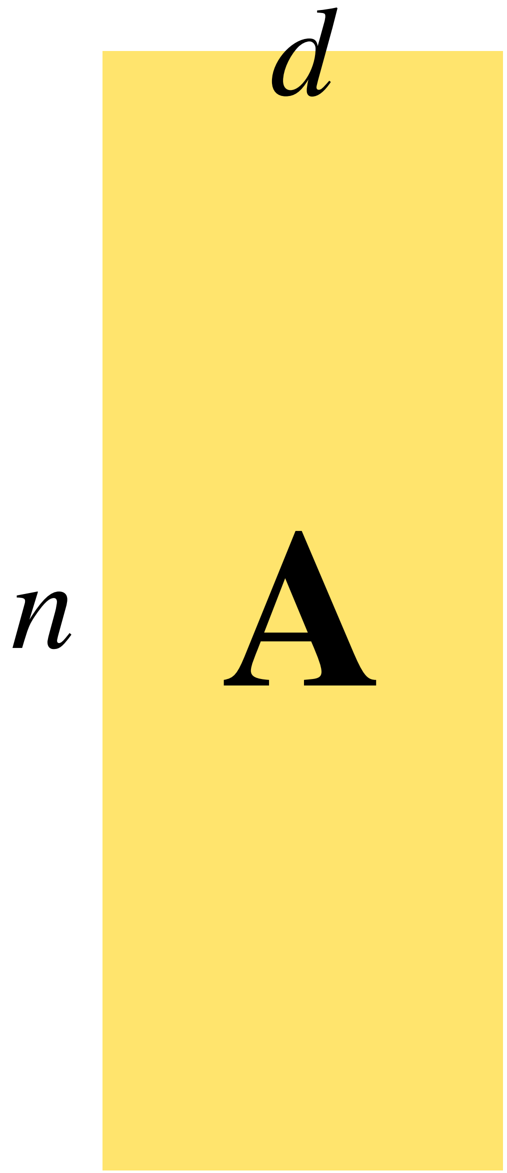
- One approach: approximate the original instance by a smaller instance

Goal: Replace $\mathbf{A}$ and $\mathbf{b}$ by a smaller $\tilde{\mathbf{A}}$ and $\tilde{\mathbf{b}}$

s.t. $\|\mathbf{A}\mathbf{x}-\mathbf{b}\|_2 \approx \|\tilde{\mathbf{A}}\mathbf{x}-\tilde{\mathbf{b}}\|_2$

Matrix Approximation

Approximating the Linear Regression Loss Function



$$\min_{\mathbf{x}\in\mathbb{R}^d} \|\mathbf{A}\mathbf{x}-\mathbf{b}\|_2^2$$

- One approach: approximate the original instance by a smaller instance

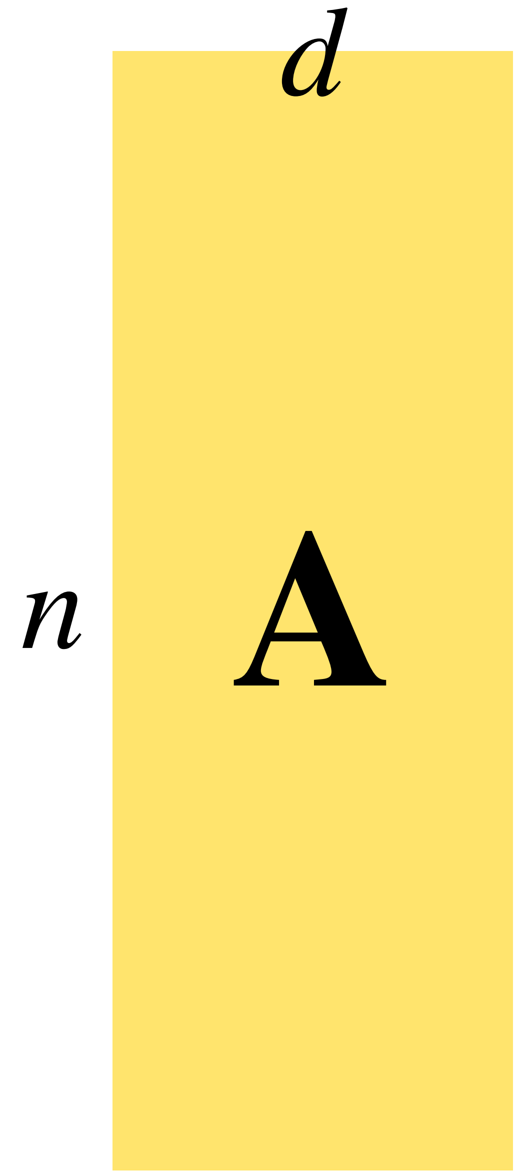
Goal: Replace $\mathbf{A}$ and $\mathbf{b}$ by a smaller $\tilde{\mathbf{A}}$ and $\tilde{\mathbf{b}}$

s.t. $\|\mathbf{A}\mathbf{x}-\mathbf{b}\|_2 \approx \|\tilde{\mathbf{A}}\mathbf{x}-\tilde{\mathbf{b}}\|_2$

- There are many possible ways to choose $\tilde{\mathbf{A}}$ and $\tilde{\mathbf{b}}$!

Matrix Approximation

Approximating the Linear Regression Loss Function



$$\min_{\mathbf{x}\in\mathbb{R}^d} \|\mathbf{A}\mathbf{x}-\mathbf{b}\|_2^2$$

- One approach: approximate the original instance by a smaller instance

Goal: Replace $\mathbf{A}$ and $\mathbf{b}$ by a smaller $\tilde{\mathbf{A}}$ and $\tilde{\mathbf{b}}$

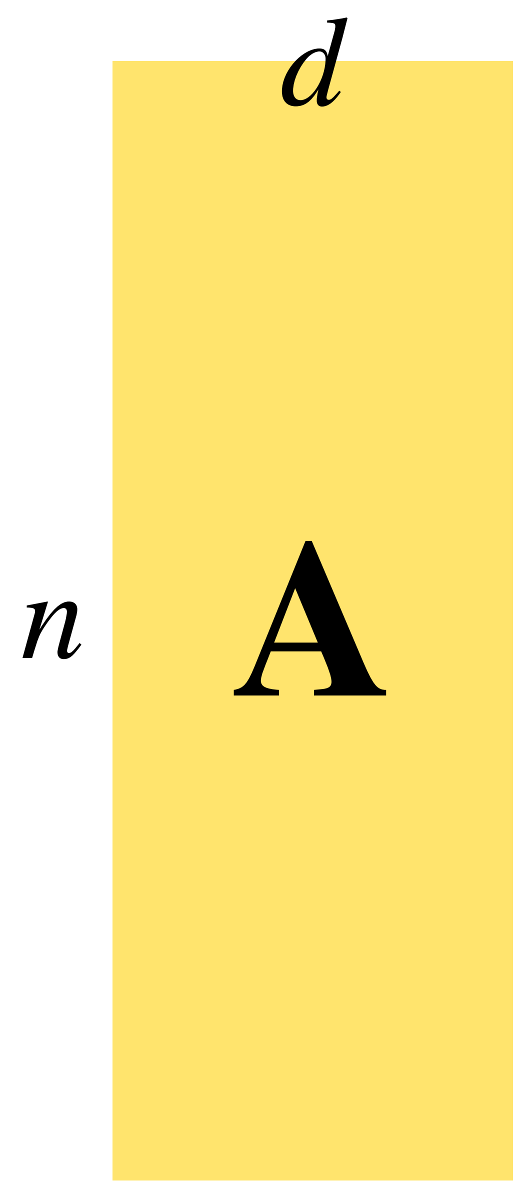
s.t. $\|\mathbf{A}\mathbf{x}-\mathbf{b}\|_2 \approx \|\tilde{\mathbf{A}}\mathbf{x}-\tilde{\mathbf{b}}\|_2$

- There are many possible ways to choose $\tilde{\mathbf{A}}$ and $\tilde{\mathbf{b}}$!

Idea. $\tilde{\mathbf{A}} = \mathbf{S}\mathbf{A}$ and $\tilde{\mathbf{b}} = \mathbf{S}\mathbf{b}$ for some $\mathbf{S} \in \mathbb{R}^{r\times n}$, $r \ll n$

Matrix Approximation

Approximating the Linear Regression Loss Function



$$\min_{\mathbf{x} \in \mathbb{R}^d} \| \mathbf{A}\mathbf{x} - \mathbf{b} \|_2^2$$

- One approach: approximate the original instance by a smaller instance

Goal: Replace $\mathbf{A}$ and $\mathbf{b}$ by a smaller $\tilde{\mathbf{A}}$ and $\tilde{\mathbf{b}}$

s.t. $\| \mathbf{A}\mathbf{x} - \mathbf{b} \|_2 \approx \| \tilde{\mathbf{A}}\mathbf{x} - \tilde{\mathbf{b}} \|_2$

- There are many possible ways to choose $\tilde{\mathbf{A}}$ and $\tilde{\mathbf{b}}$!

Idea. $\tilde{\mathbf{A}} = \mathbf{S}\mathbf{A}$ and $\tilde{\mathbf{b}} = \mathbf{S}\mathbf{b}$ for some $\mathbf{S} \in \mathbb{R}^{r \times n}$, $r \ll n$

“Sketch"

Matrix Approximation

Approximating the Linear Regression Loss Function

Goal: Replace $\mathbf{A}$ and $\mathbf{b}$ by a smaller $\tilde{\mathbf{A}}$ and $\tilde{\mathbf{b}}$ s.t. $\|\mathbf{A}\mathbf{x} - \mathbf{b}\|_2 \approx \|\tilde{\mathbf{A}}\mathbf{x} - \tilde{\mathbf{b}}\|_2$

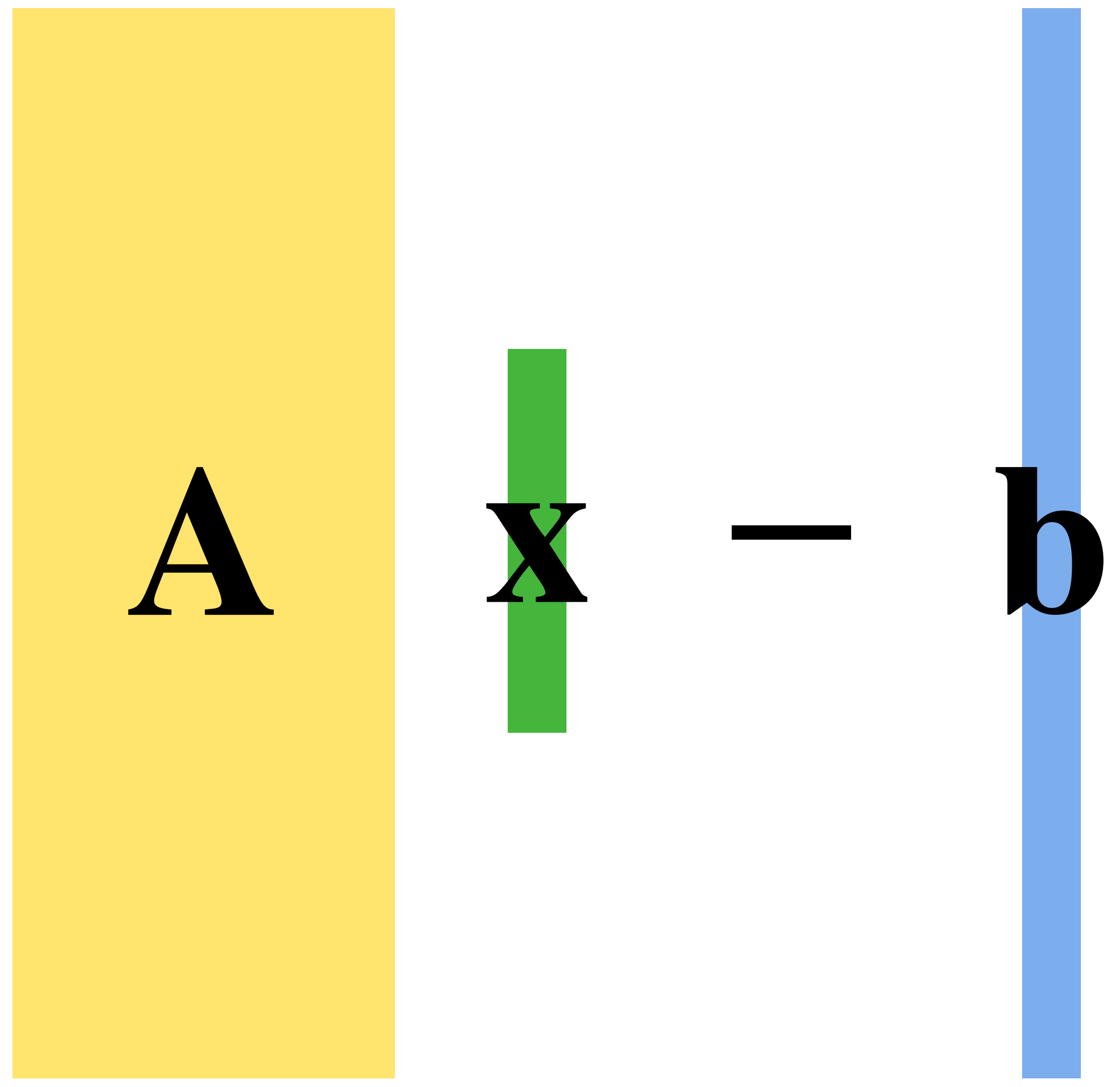
Idea. $\tilde{\mathbf{A}} = \mathbf{S}\mathbf{A}$ and $\tilde{\mathbf{b}} = \mathbf{S}\mathbf{b}$ for some $\mathbf{S} \in \mathbb{R}^{r \times n}$, $r \ll n$

Matrix Approximation

Approximating the Linear Regression Loss Function

Goal: Replace $\mathbf{A}$ and $\mathbf{b}$ by a smaller $\tilde{\mathbf{A}}$ and $\tilde{\mathbf{b}}$ s.t. $\|\mathbf{A}\mathbf{x} - \mathbf{b}\|_2 \approx \|\tilde{\mathbf{A}}\mathbf{x} - \tilde{\mathbf{b}}\|_2$

Idea. $\tilde{\mathbf{A}} = \mathbf{S}\mathbf{A}$ and $\tilde{\mathbf{b}} = \mathbf{S}\mathbf{b}$ for some $\mathbf{S} \in \mathbb{R}^{r \times n}$, $r \ll n$

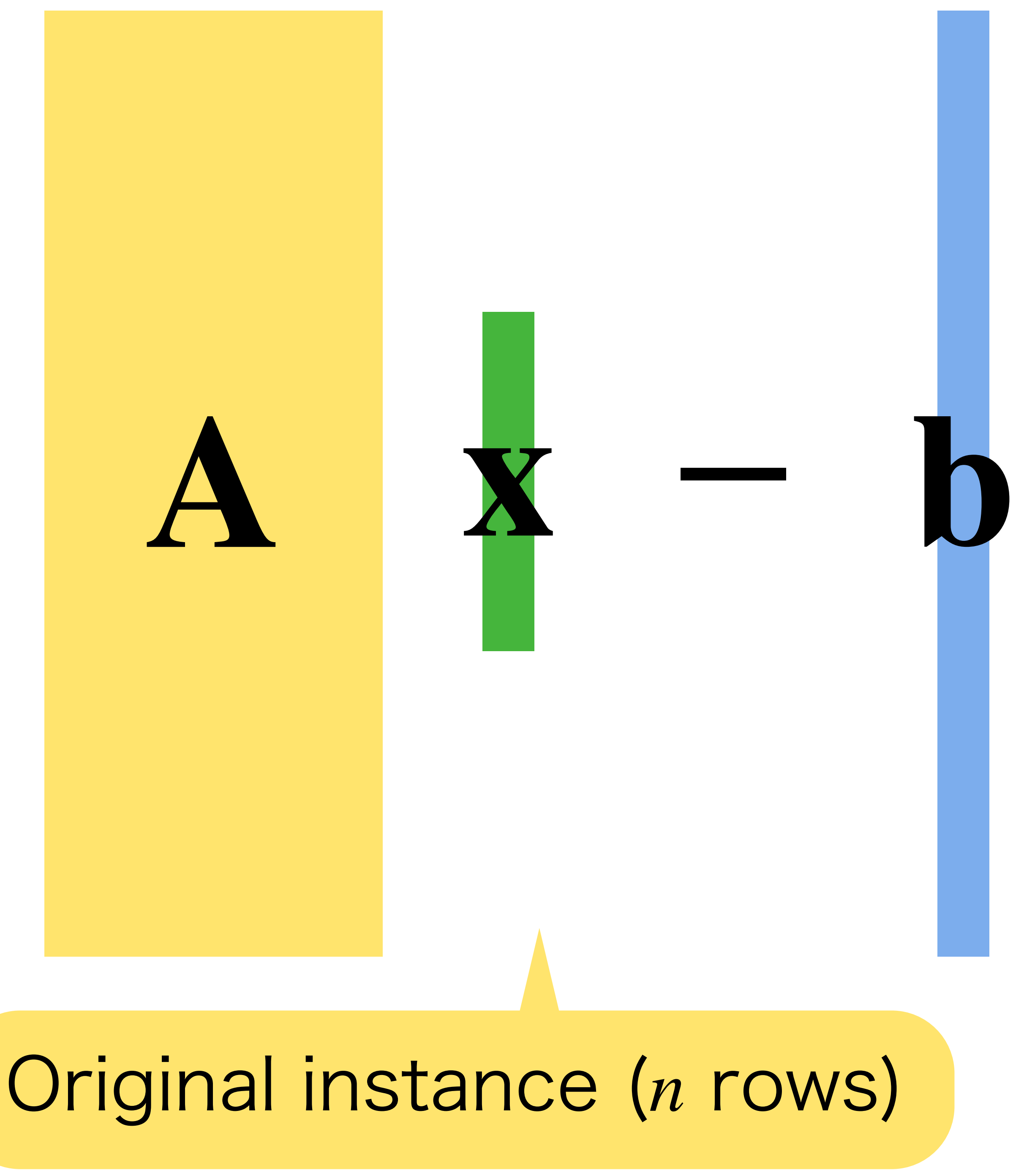


Matrix Approximation

Approximating the Linear Regression Loss Function

Goal: Replace $\mathbf{A}$ and $\mathbf{b}$ by a smaller $\tilde{\mathbf{A}}$ and $\tilde{\mathbf{b}}$ s.t. $\|\mathbf{A}\mathbf{x} - \mathbf{b}\|_2 \approx \|\tilde{\mathbf{A}}\mathbf{x} - \tilde{\mathbf{b}}\|_2$

Idea. $\tilde{\mathbf{A}} = \mathbf{S}\mathbf{A}$ and $\tilde{\mathbf{b}} = \mathbf{S}\mathbf{b}$ for some $\mathbf{S} \in \mathbb{R}^{r \times n}$, $r \ll n$

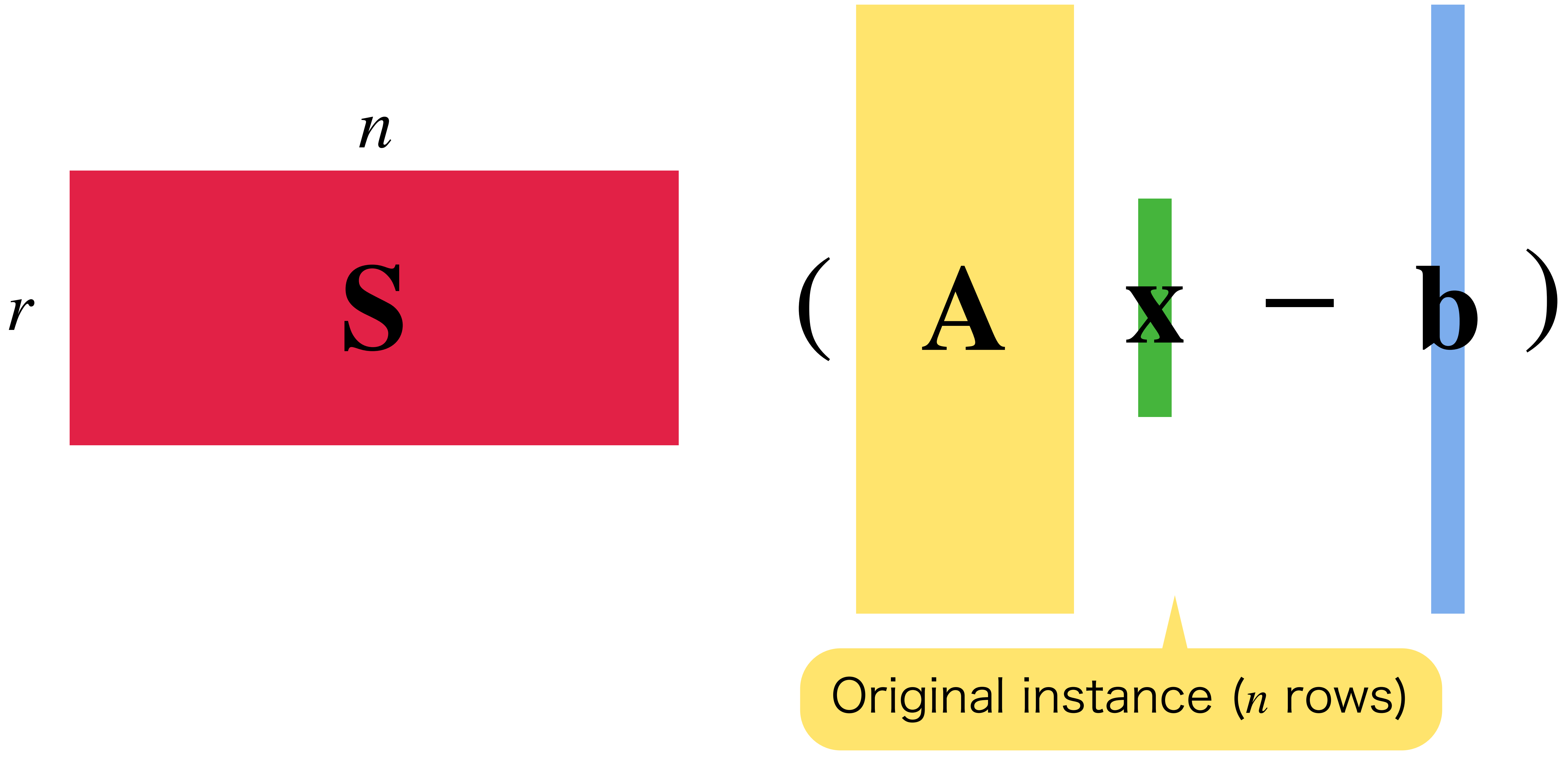


Matrix Approximation

Approximating the Linear Regression Loss Function

Goal: Replace $\mathbf{A}$ and $\mathbf{b}$ by a smaller $\tilde{\mathbf{A}}$ and $\tilde{\mathbf{b}}$ s.t. $\|\mathbf{A}\mathbf{x}-\mathbf{b}\|_2 \approx \|\tilde{\mathbf{A}}\mathbf{x}-\tilde{\mathbf{b}}\|_2$

Idea. $\tilde{\mathbf{A}} = \mathbf{S}\mathbf{A}$ and $\tilde{\mathbf{b}} = \mathbf{S}\mathbf{b}$ for some $\mathbf{S} \in \mathbb{R}^{r \times n}$, $r \ll n$

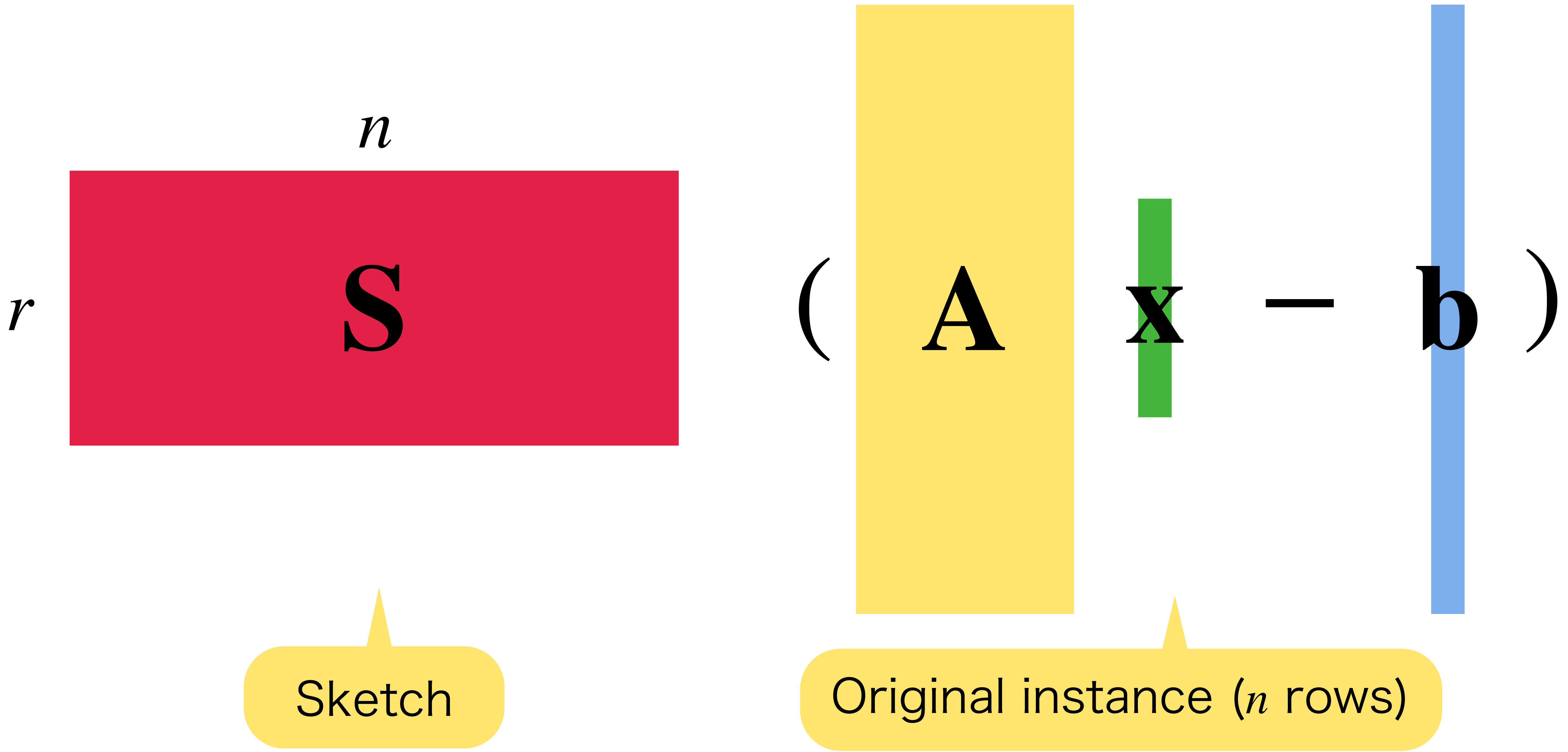


Matrix Approximation

Approximating the Linear Regression Loss Function

Goal: Replace $\mathbf{A}$ and $\mathbf{b}$ by a smaller $\tilde{\mathbf{A}}$ and $\tilde{\mathbf{b}}$ s.t. $\|\mathbf{A}\mathbf{x}-\mathbf{b}\|_2 \approx \|\tilde{\mathbf{A}}\mathbf{x}-\tilde{\mathbf{b}}\|_2$

Idea. $\tilde{\mathbf{A}} = \mathbf{S}\mathbf{A}$ and $\tilde{\mathbf{b}} = \mathbf{S}\mathbf{b}$ for some $\mathbf{S} \in \mathbb{R}^{r \times n}$, $r \ll n$

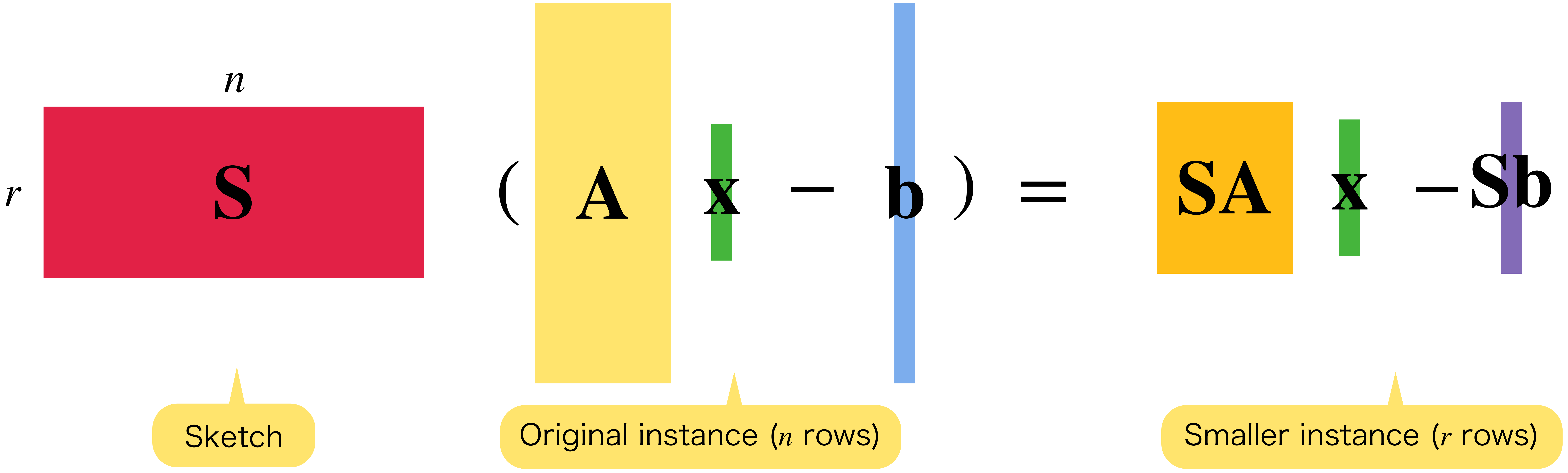


Matrix Approximation

Approximating the Linear Regression Loss Function

Goal: Replace $\mathbf{A}$ and $\mathbf{b}$ by a smaller $\tilde{\mathbf{A}}$ and $\tilde{\mathbf{b}}$ s.t. $\| \mathbf{A}\mathbf{x} - \mathbf{b} \|_2 \approx \| \tilde{\mathbf{A}}\mathbf{x} - \tilde{\mathbf{b}} \|_2$

Idea. $\tilde{\mathbf{A}} = \mathbf{S}\mathbf{A}$ and $\tilde{\mathbf{b}} = \mathbf{S}\mathbf{b}$ for some $\mathbf{S} \in \mathbb{R}^{r \times n}$, $r \ll n$



Matrix Approximation

Approximating the Linear Regression Loss Function

Definition (Sarlos 2006). $\mathbf{S} \in \mathbb{R}^{r \times n}$ is a **subspace embedding** of $\mathbf{A} \in \mathbb{R}^{n \times d}$ if

$$\|\mathbf{Ax}\|_2 \leq \|\mathbf{SAx}\|_2 \leq \kappa \|\mathbf{Ax}\|_2$$

for every $\mathbf{x} \in \mathbb{R}^d$.

Matrix Approximation

Approximating the Linear Regression Loss Function

Definition (Sarlos 2006). $\mathbf{S} \in \mathbb{R}^{r \times n}$ is a **subspace embedding** of $\mathbf{A} \in \mathbb{R}^{n \times d}$ if

$$\|\mathbf{A}\mathbf{x}\|_2 \leq \|\mathbf{S}\mathbf{A}\mathbf{x}\|_2 \leq \kappa \|\mathbf{A}\mathbf{x}\|_2$$

for every $\mathbf{x} \in \mathbb{R}^d$.

Distortion/error

Matrix Approximation

Approximating the Linear Regression Loss Function

Definition (Sarlos 2006). $\mathbf{S} \in \mathbb{R}^{r \times n}$ is a **subspace embedding** of $\mathbf{A} \in \mathbb{R}^{n \times d}$ if

$$\|\mathbf{Ax}\|_2 \leq \|\mathbf{SAx}\|_2 \leq \kappa \|\mathbf{Ax}\|_2$$

for every $\mathbf{x} \in \mathbb{R}^d$.

Distortion/error

Why is this useful?

Matrix Approximation

Approximating the Linear Regression Loss Function

Definition (Sarlos 2006). $\mathbf{S} \in \mathbb{R}^{r \times n}$ is a **subspace embedding** of $\mathbf{A} \in \mathbb{R}^{n \times d}$ if

$$\|\mathbf{Ax}\|_2 \leq \|\mathbf{SAx}\|_2 \leq \kappa \|\mathbf{Ax}\|_2$$

Distortion/error

for every $\mathbf{x} \in \mathbb{R}^d$.

Why is this useful?

- Let $\mathbf{A}' = [\mathbf{A}\ \mathbf{b}] \in \mathbb{R}^{n \times (d+1)}$ and let $\mathbf{S} \in \mathbb{R}^{r \times n}$ be a subspace embedding of $\mathbf{A}'$

Matrix Approximation

Approximating the Linear Regression Loss Function

Definition (Sarlos 2006). $\mathbf{S} \in \mathbb{R}^{r \times n}$ is a **subspace embedding** of $\mathbf{A} \in \mathbb{R}^{n \times d}$ if

$$\|\mathbf{A}\mathbf{x}\|_2 \leq \|\mathbf{S}\mathbf{A}\mathbf{x}\|_2 \leq \kappa \|\mathbf{A}\mathbf{x}\|_2$$

for every $\mathbf{x} \in \mathbb{R}^d$.

Distortion/error

Why is this useful?

- Let $\mathbf{A}' = [\mathbf{A} \ \mathbf{b}] \in \mathbb{R}^{n \times (d+1)}$ and let $\mathbf{S} \in \mathbb{R}^{r \times n}$ be a subspace embedding of $\mathbf{A}'$
- $\mathbf{A}\mathbf{x} - \mathbf{b} = \mathbf{A}'\mathbf{x}'$ for $\mathbf{x}' = [\mathbf{x}; -1] \in \mathbb{R}^{d+1}$

Matrix Approximation

Approximating the Linear Regression Loss Function

Definition (Sarlos 2006). $\mathbf{S} \in \mathbb{R}^{r \times n}$ is a **subspace embedding** of $\mathbf{A} \in \mathbb{R}^{n \times d}$ if

$$\|\mathbf{A}\mathbf{x}\|_2 \leq \|\mathbf{S}\mathbf{A}\mathbf{x}\|_2 \leq \kappa \|\mathbf{A}\mathbf{x}\|_2$$

for every $\mathbf{x} \in \mathbb{R}^d$.

Distortion/error

Why is this useful?

- Let $\mathbf{A}' = [\mathbf{A}\ \ \mathbf{b}] \in \mathbb{R}^{n \times (d+1)}$ and let $\mathbf{S} \in \mathbb{R}^{r \times n}$ be a subspace embedding of $\mathbf{A}'$
- $\mathbf{A}\mathbf{x} - \mathbf{b} = \mathbf{A}'\mathbf{x}'$ for $\mathbf{x}' = [\mathbf{x}; -1] \in \mathbb{R}^{d+1}$
- $\|\mathbf{A}\mathbf{x} - \mathbf{b}\|_2 \leq \|\mathbf{S}\mathbf{A}\mathbf{x} - \mathbf{S}\mathbf{b}\|_2 \leq \kappa \|\mathbf{A}\mathbf{x} - \mathbf{b}\|_2$

Matrix Approximation

Approximating the Linear Regression Loss Function

Definition (Sarlos 2006). $\mathbf{S} \in \mathbb{R}^{r \times n}$ is a **subspace embedding** of $\mathbf{A} \in \mathbb{R}^{n \times d}$ if

$$\|\mathbf{A}\mathbf{x}\|_2 \leq \|\mathbf{S}\mathbf{A}\mathbf{x}\|_2 \leq \kappa \|\mathbf{A}\mathbf{x}\|_2$$

for every $\mathbf{x} \in \mathbb{R}^d$.

Distortion/error

Why is this useful?

- Let $\mathbf{A}' = [\mathbf{A}\ \ \mathbf{b}] \in \mathbb{R}^{n \times (d+1)}$ and let $\mathbf{S} \in \mathbb{R}^{r \times n}$ be a subspace embedding of $\mathbf{A}'$
- $\mathbf{A}\mathbf{x} - \mathbf{b} = \mathbf{A}'\mathbf{x}'$ for $\mathbf{x}' = [\mathbf{x}; -1] \in \mathbb{R}^{d+1}$
- $\|\mathbf{A}\mathbf{x} - \mathbf{b}\|_2 \leq \|\mathbf{S}\mathbf{A}\mathbf{x} - \mathbf{S}\mathbf{b}\|_2 \leq \kappa \|\mathbf{A}\mathbf{x} - \mathbf{b}\|_2$

Cost of $\mathbf{SA}$ and $\mathbf{Sb}$ approximates cost of $\mathbf{A}$ and $\mathbf{b}$

Matrix Approximation

Approximating the Linear Regression Loss Function

Definition (Sarlos 2006). $\mathbf{S} \in \mathbb{R}^{r \times n}$ is a **subspace embedding** of $\mathbf{A} \in \mathbb{R}^{n \times d}$ if

$$\|\mathbf{A}\mathbf{x}\|_2 \leq \|\mathbf{S}\mathbf{A}\mathbf{x}\|_2 \leq \kappa \|\mathbf{A}\mathbf{x}\|_2$$

for every $\mathbf{x} \in \mathbb{R}^d$.

Distortion/error

Why is this useful?

- Let $\mathbf{A}' = [\mathbf{A}\ \ \mathbf{b}] \in \mathbb{R}^{n \times (d+1)}$ and let $\mathbf{S} \in \mathbb{R}^{r \times n}$ be a subspace embedding of $\mathbf{A}'$
- $\mathbf{A}\mathbf{x} - \mathbf{b} = \mathbf{A}'\mathbf{x}'$ for $\mathbf{x}' = [\mathbf{x}; -1] \in \mathbb{R}^{d+1}$
- $\|\mathbf{A}\mathbf{x} - \mathbf{b}\|_2 \leq \|\mathbf{S}\mathbf{A}\mathbf{x} - \mathbf{S}\mathbf{b}\|_2 \leq \kappa \|\mathbf{A}\mathbf{x} - \mathbf{b}\|_2$

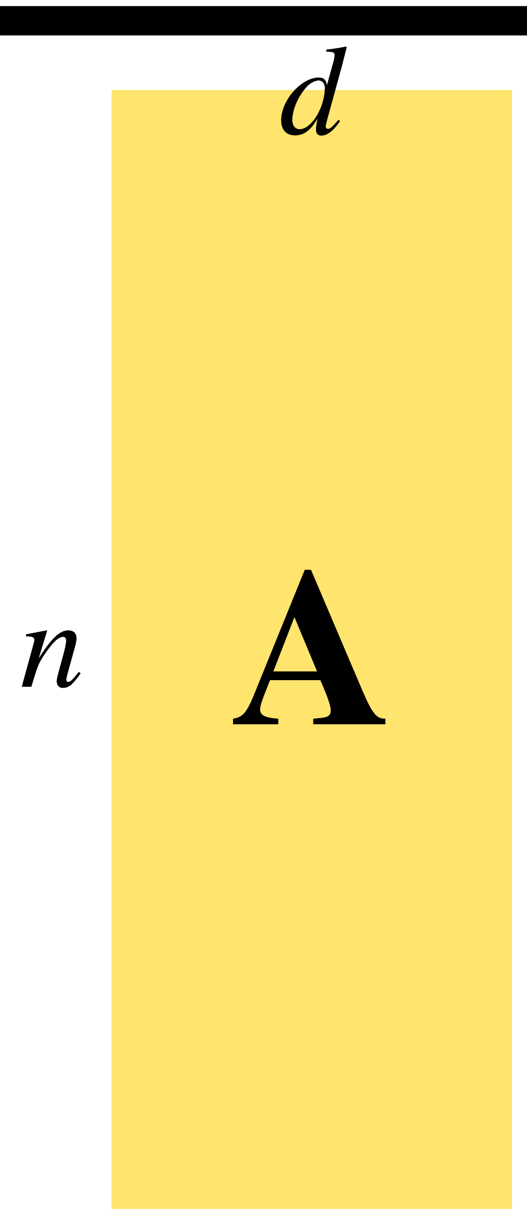
Cost of $\mathbf{SA}$ and $\mathbf{Sb}$ approximates cost of $\mathbf{A}$ and $\mathbf{b}$

- Solve linear regression on $r \times d$ matrix $\mathbf{SA}$ instead of $n \times d$ matrix $\mathbf{A}$!

Matrix Approximation

Oblivious Subspace Embeddings

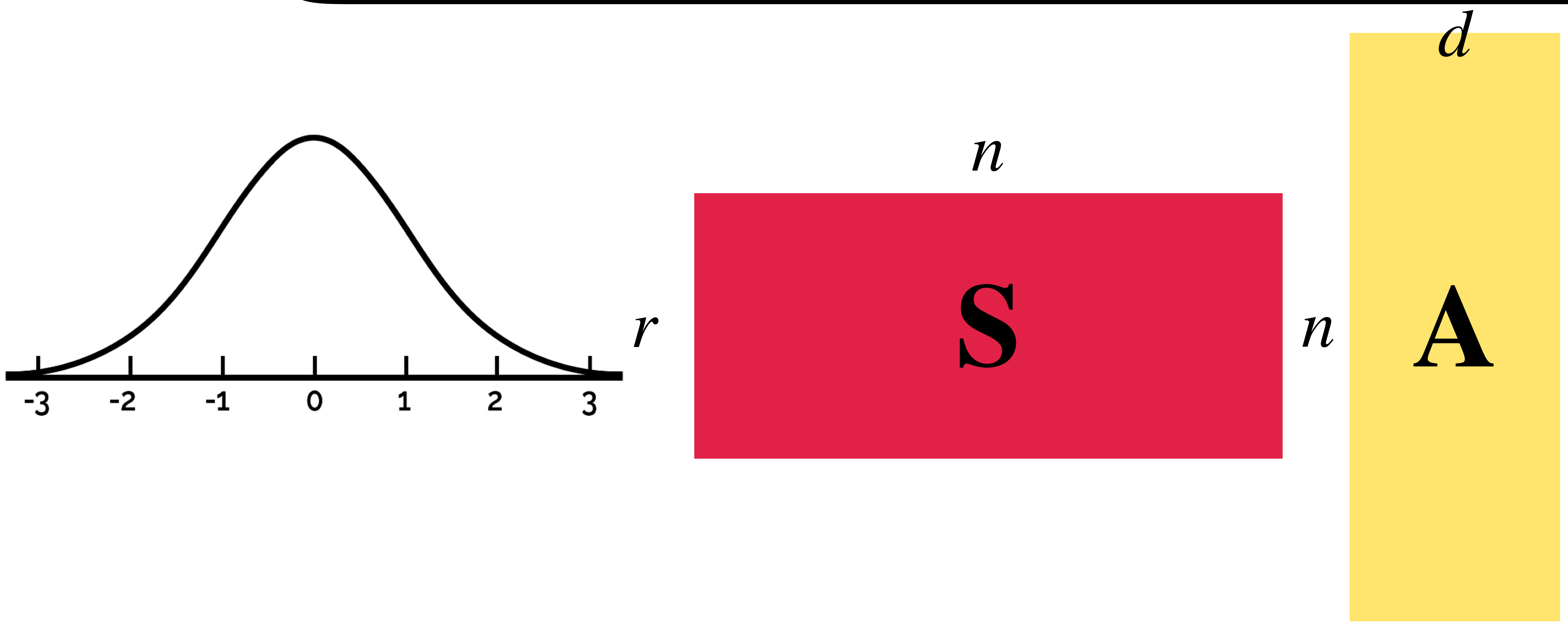
Theorem (Sarlos 2006). Let $\kappa = (1+\varepsilon)$. Let $r = \tilde{O}(\varepsilon^{-2}d)$. If $\mathbf{S}$ is an $r \times n$ Gaussian matrix, then $\mathbf{S}$ is a subspace embedding for any $\mathbf{A}$ with distortion κ, with probability 99%.



Matrix Approximation

Oblivious Subspace Embeddings

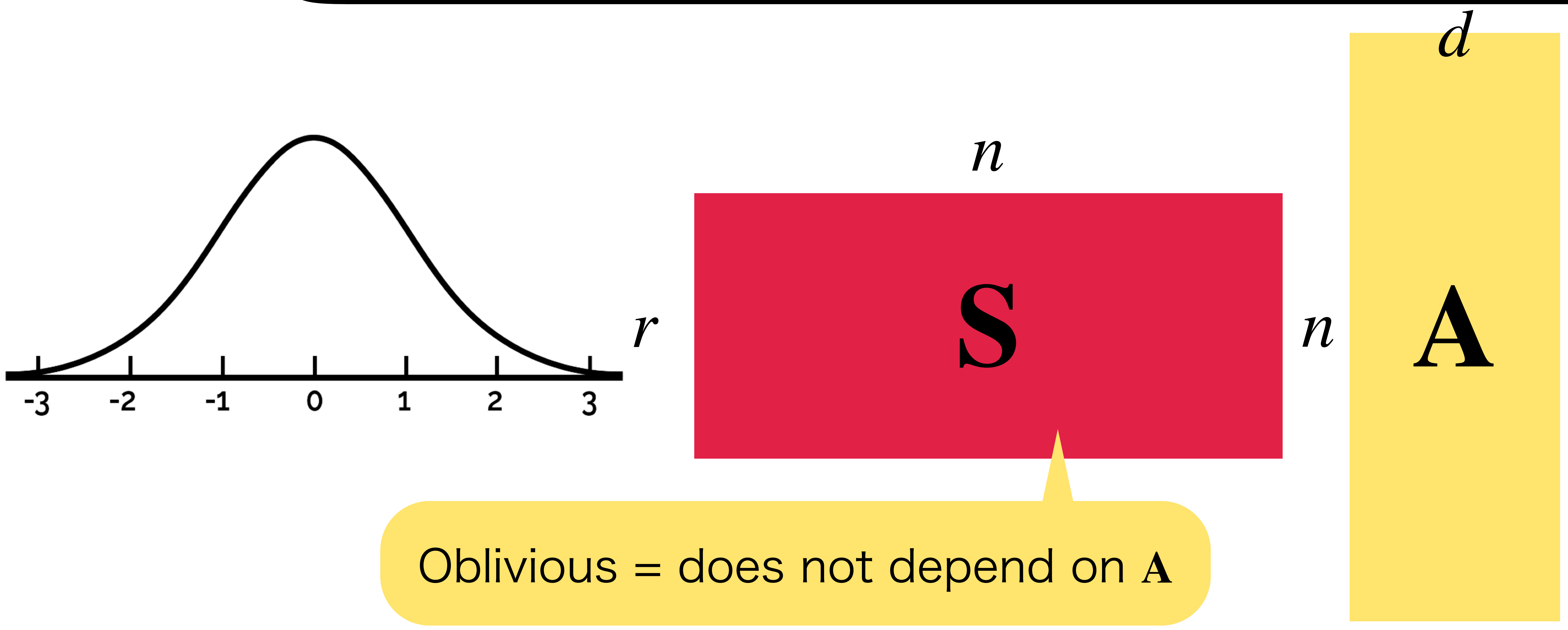
Theorem (Sarlos 2006). Let $\kappa = (1 + \varepsilon)$. Let $r = \tilde{O}(\varepsilon^{-2} d)$. If $\mathbf{S}$ is an $r \times n$ Gaussian matrix, then $\mathbf{S}$ is a subspace embedding for any $\mathbf{A}$ with distortion κ, with probability 99%.



Matrix Approximation

Oblivious Subspace Embeddings

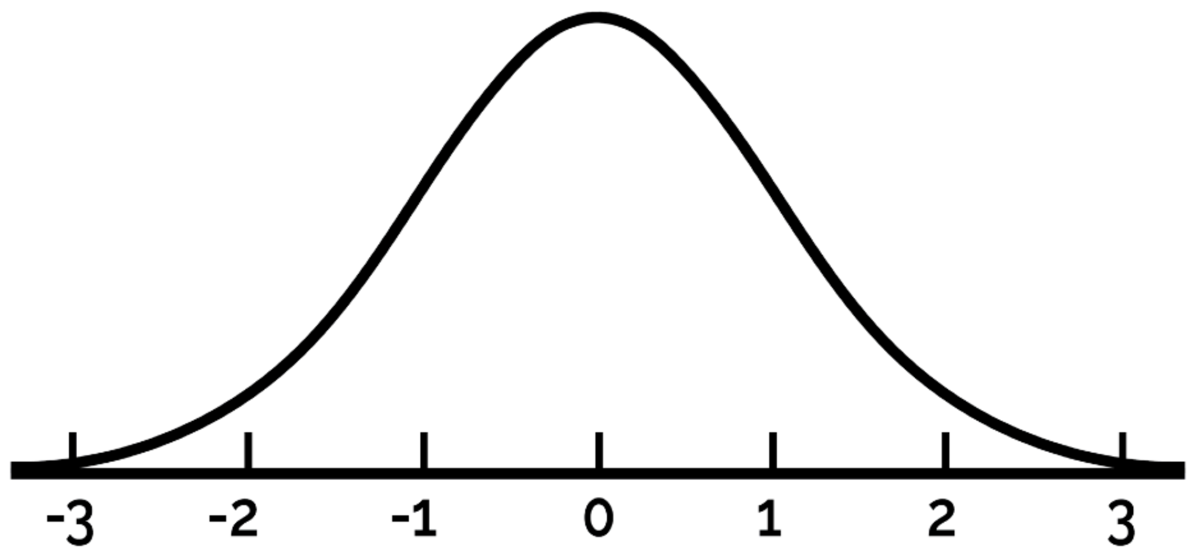
Theorem (Sarlos 2006). Let $\kappa = (1 + \varepsilon)$. Let $r = \tilde{O}(\varepsilon^{-2} d)$. If $\mathbf{S}$ is an $r \times n$ Gaussian matrix, then $\mathbf{S}$ is a subspace embedding for any $\mathbf{A}$ with distortion κ, with probability 99%.

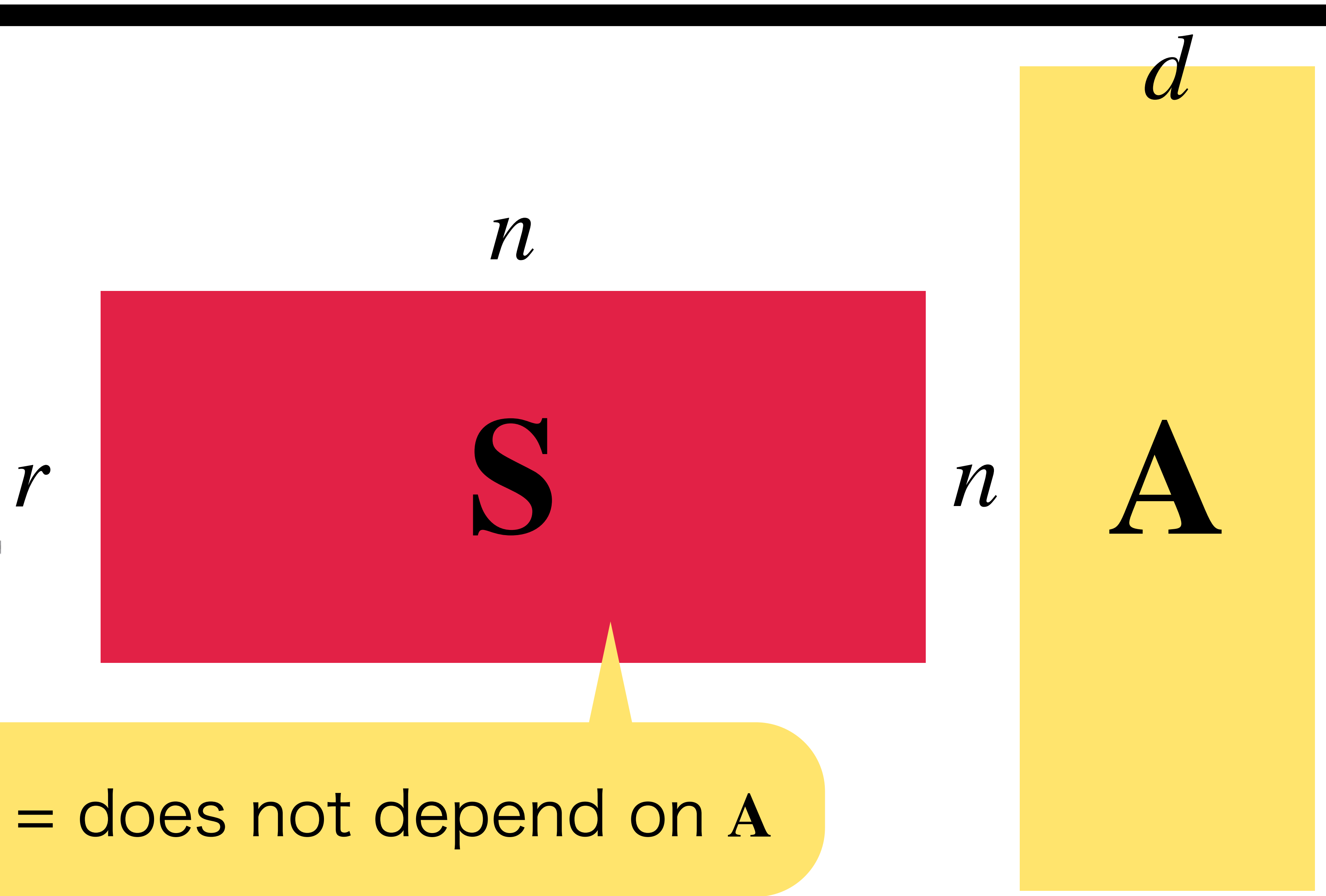


Matrix Approximation

Oblivious Subspace Embeddings

Theorem (Sarlos 2006). Let $\kappa = (1+\varepsilon)$. Let $r = \tilde{O}(\varepsilon^{-2}d)$. If $\mathbf{S}$ is an $r \times n$ Gaussian matrix, then $\mathbf{S}$ is a subspace embedding for any $\mathbf{A}$ with distortion κ, with probability 99%.





Oblivious = does not depend on $\mathbf{A}$

So ℓ_2 regression is resolved.
What's next?

Matrix Approximation

ℓ_p Linear Regression

Matrix Approximation

ℓ_p Linear Regression

ℓ_2 linear regression

$$\min_{\mathbf{x}\in\mathbb{R}^d} \|\mathbf{A}\mathbf{x}-\mathbf{b}\|_2^2$$

- Minimize the **sum of squares** of errors

Matrix Approximation

ℓ_p Linear Regression

ℓ_1 linear regression

$$\min_{\mathbf{x}\in\mathbb{R}^d} \| \mathbf{A}\mathbf{x} - \mathbf{b} \|_1$$

- Minimize the **average** error
- **Robust** loss function

ℓ_2 linear regression

$$\min_{\mathbf{x}\in\mathbb{R}^d} \| \mathbf{A}\mathbf{x} - \mathbf{b} \|_2^2$$

- Minimize the **sum of squares** of errors

Matrix Approximation

ℓ_p Linear Regression

ℓ_1 linear regression

$$\min_{\mathbf{x}\in\mathbb{R}^d} \|\mathbf{Ax}-\mathbf{b}\|_1$$

- Minimize the **average** error
- **Robust** loss function

ℓ_2 linear regression

$$\min_{\mathbf{x}\in\mathbb{R}^d} \|\mathbf{Ax}-\mathbf{b}\|_2^2$$

- Minimize the **sum of squares** of errors

ℓ_∞ linear regression

$$\min_{\mathbf{x}\in\mathbb{R}^d} \|\mathbf{Ax}-\mathbf{b}\|_\infty$$

- Minimize the **worst-case** error
- **Sensitive** loss function

Matrix Approximation

ℓ_p Linear Regression

ℓ_1 linear regression

$$\min_{\mathbf{x}\in\mathbb{R}^d} \|\mathbf{A}\mathbf{x}-\mathbf{b}\|_1$$

- Minimize the **average** error
- **Robust** loss function

ℓ_2 linear regression

$$\min_{\mathbf{x}\in\mathbb{R}^d} \|\mathbf{A}\mathbf{x}-\mathbf{b}\|_2^2$$

- Minimize the **sum of squares** of errors

ℓ_∞ linear regression

$$\min_{\mathbf{x}\in\mathbb{R}^d} \|\mathbf{A}\mathbf{x}-\mathbf{b}\|_\infty$$

- Minimize the **worst-case** error
- **Sensitive** loss function

ℓ_p linear regression

$$\min_{\mathbf{x}\in\mathbb{R}^d} \|\mathbf{A}\mathbf{x}-\mathbf{b}\|_p^p$$

Matrix Approximation

ℓ_p Linear Regression

ℓ_1 linear regression

$$\min_{\mathbf{x}\in\mathbb{R}^d} \|\mathbf{Ax}-\mathbf{b}\|_1$$

- Minimize the **average** error
- **Robust** loss function

ℓ_2 linear regression

$$\min_{\mathbf{x}\in\mathbb{R}^d} \|\mathbf{Ax}-\mathbf{b}\|_2^2$$

- Minimize the **sum of squares** of errors

ℓ_∞ linear regression

$$\min_{\mathbf{x}\in\mathbb{R}^d} \|\mathbf{Ax}-\mathbf{b}\|_\infty$$

- Minimize the **worst-case** error
- **Sensitive** loss function

ℓ_p linear regression

$$\min_{\mathbf{x}\in\mathbb{R}^d} \|\mathbf{Ax}-\mathbf{b}\|_p^p$$

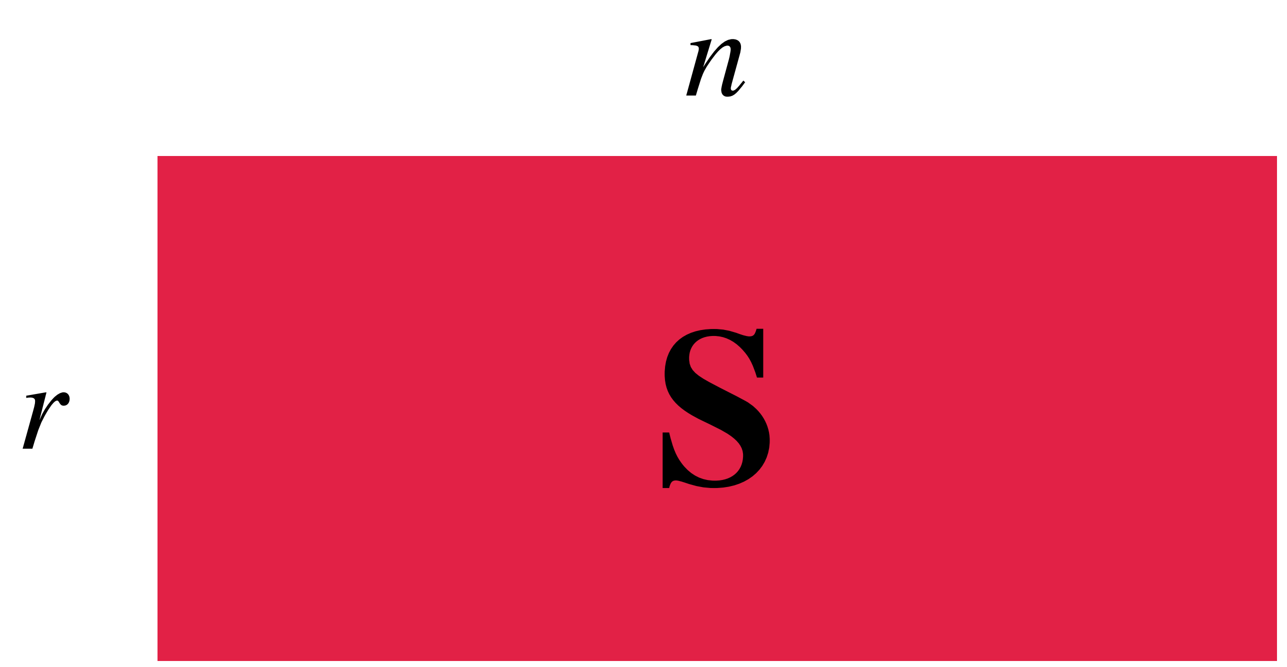
Question. What trade-offs are possible for oblivious subspace embeddings under the ℓ_p loss?

Matrix Approximation

Oblivious ℓ_p Subspace Embeddings

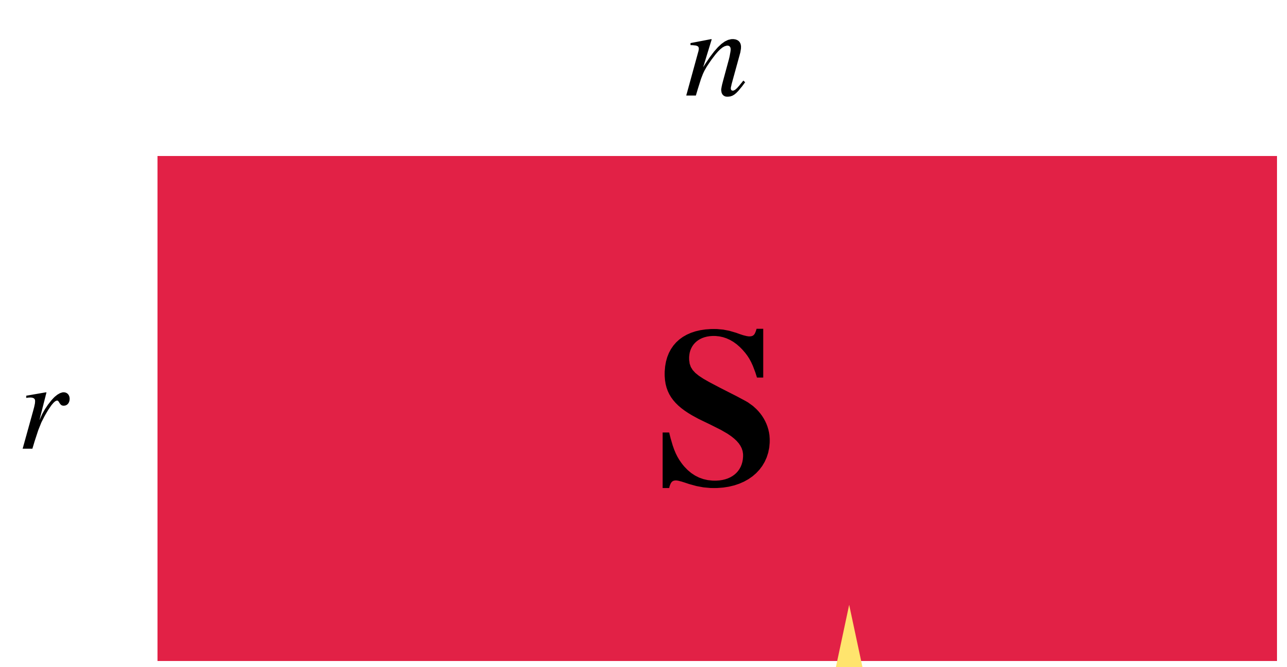
Matrix Approximation

Oblivious ℓ_p Subspace Embeddings



Matrix Approximation

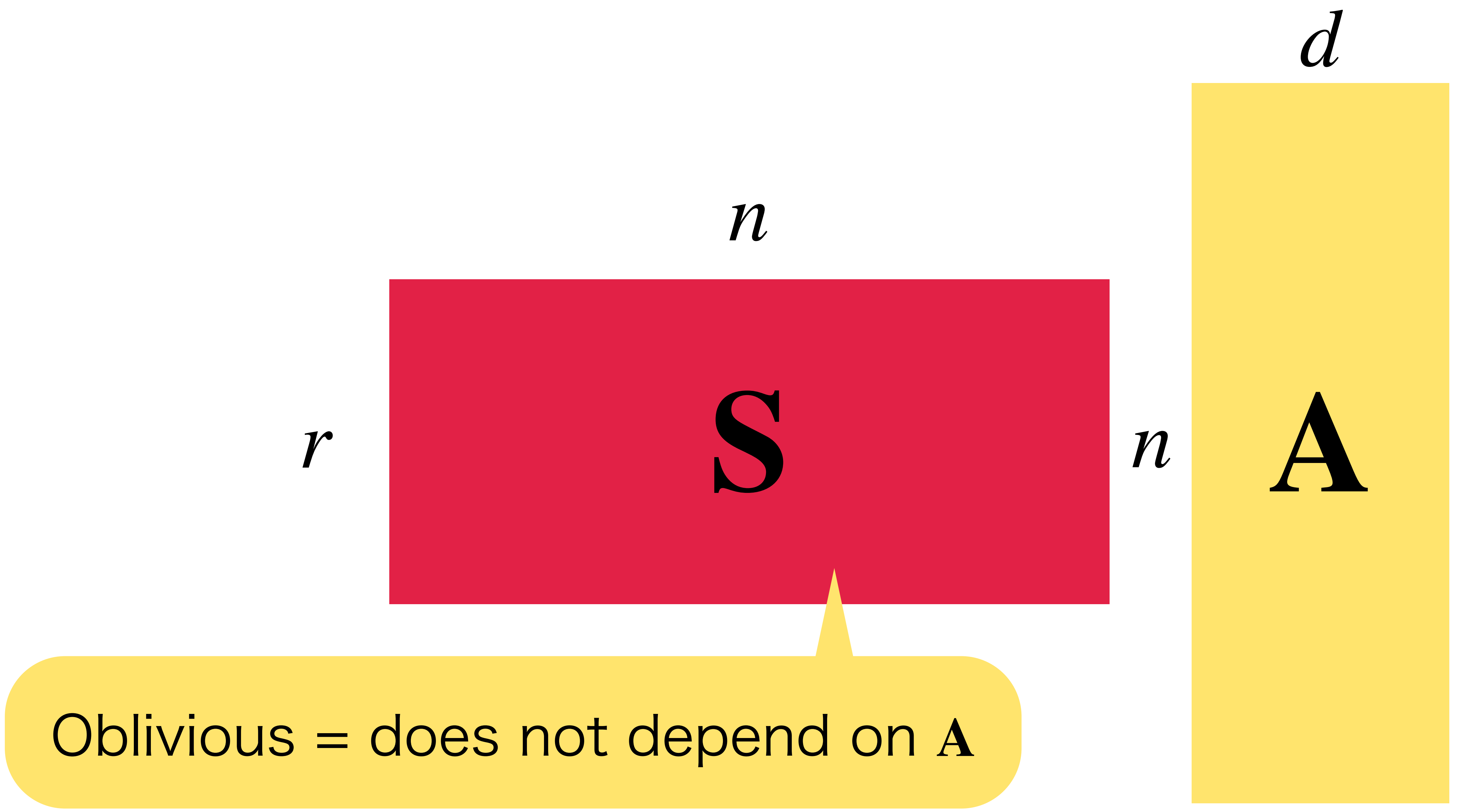
Oblivious ℓ_p Subspace Embeddings



Oblivious = does not depend on $\mathbf{A}$

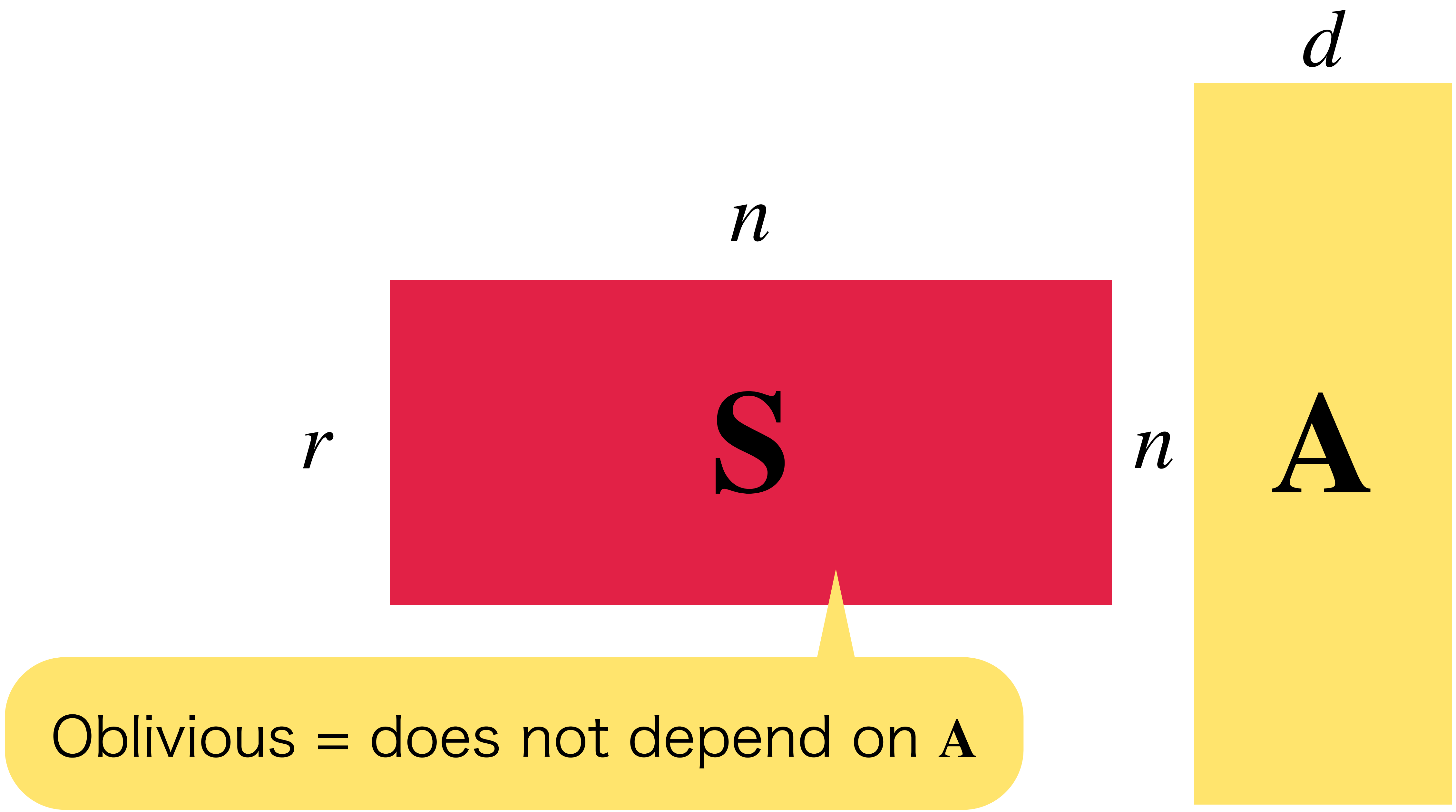
Matrix Approximation

Oblivious ℓ_p Subspace Embeddings



Matrix Approximation

Oblivious ℓ_p Subspace Embeddings



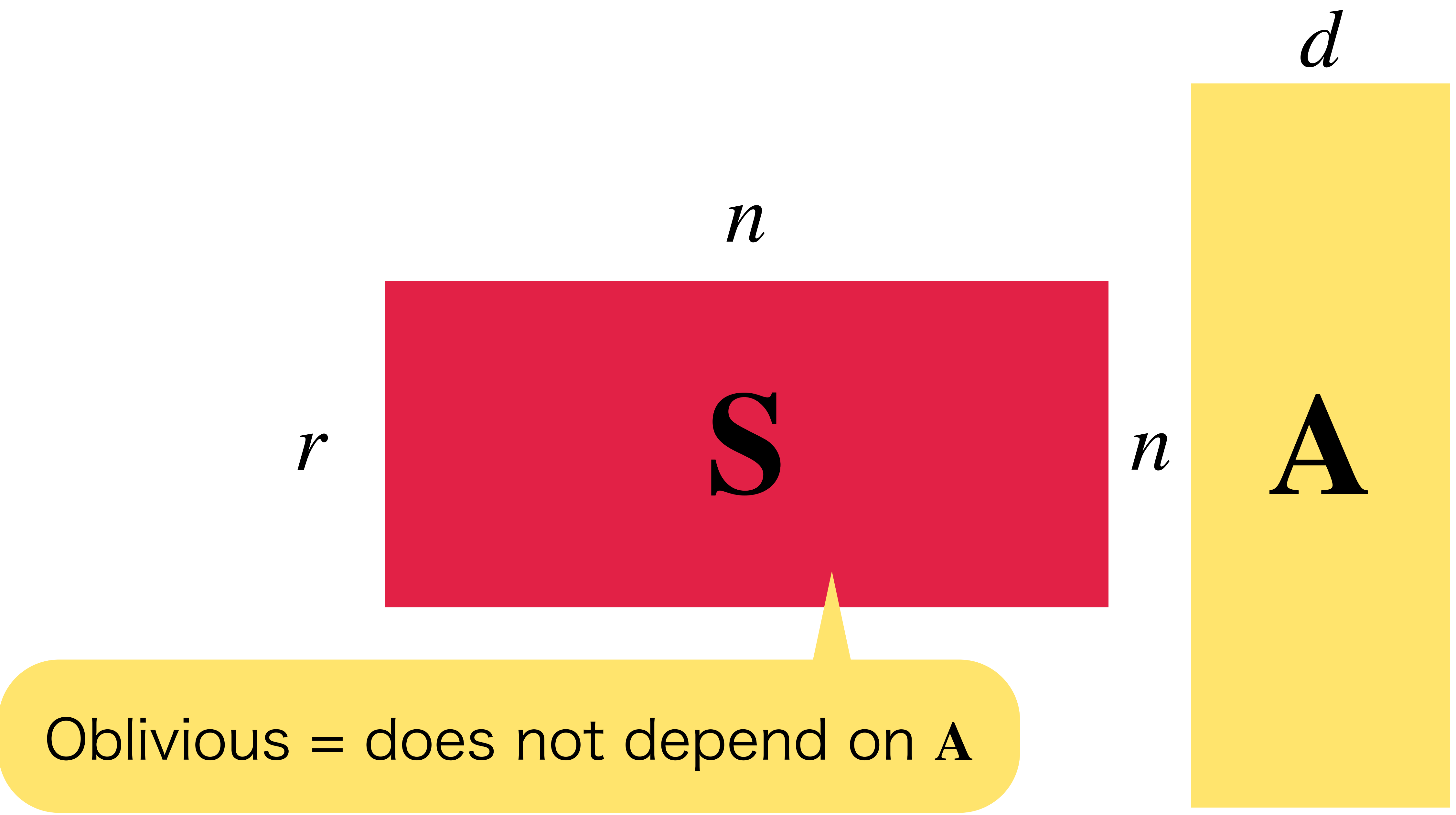
$$\|\mathbf{Ax}\|_p \leq \|\mathbf{SAx}\|_p \leq \kappa\|\mathbf{Ax}\|_p$$

for every $\mathbf{x} \in \mathbb{R}^d$

Matrix Approximation

Oblivious ℓ_p Subspace Embeddings

Two regimes for $1 \leq p < 2$



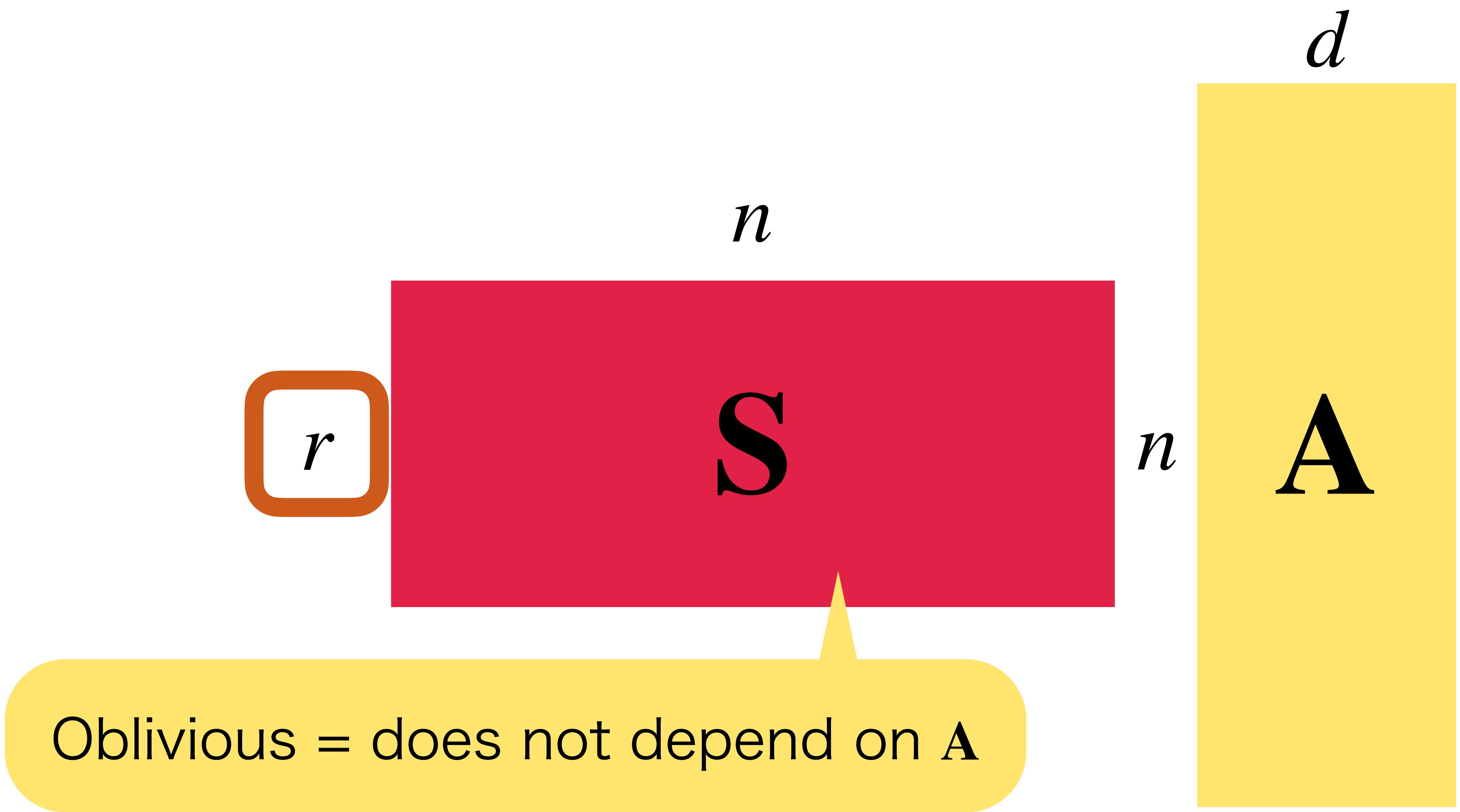
$$\|\mathbf{Ax}\|_p \leq \|\mathbf{SAx}\|_p \leq \kappa \|\mathbf{Ax}\|_p$$

for every $\mathbf{x} \in \mathbb{R}^d$

Matrix Approximation

Oblivious ℓ_p Subspace Embeddings

Two regimes for $1 \leq p < 2$



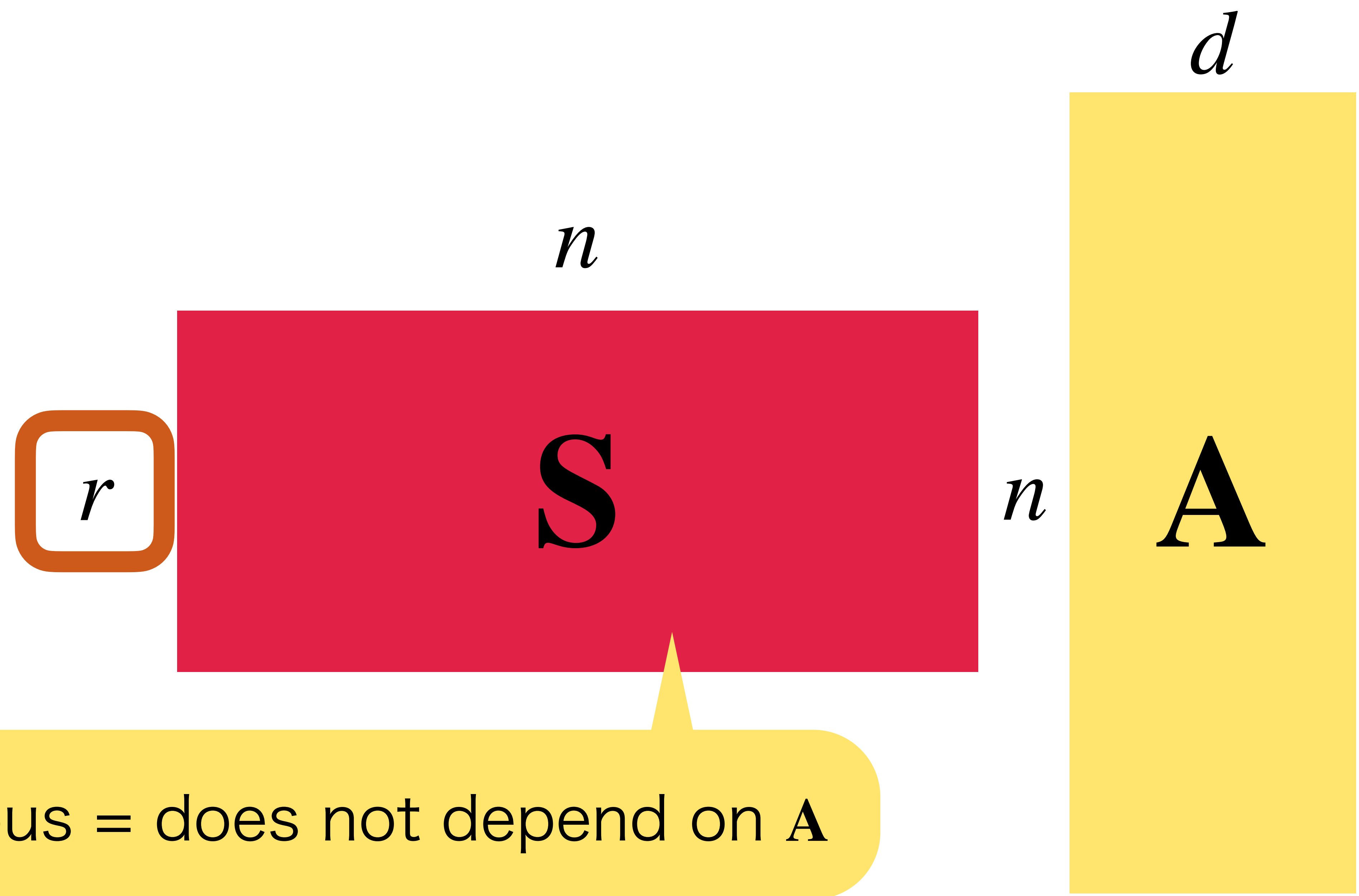
$$\|\mathbf{A}\mathbf{x}\|_p \leq \|\mathbf{S}\mathbf{A}\mathbf{x}\|_p \leq \kappa\|\mathbf{A}\mathbf{x}\|_p$$

for every $\mathbf{x} \in \mathbb{R}^d$

Matrix Approximation

Oblivious ℓ_p Subspace Embeddings

Two regimes for $1 \leq p < 2$



Oblivious = does not depend on $\mathbf{A}$

$$\|\mathbf{Ax}\|_p \leq \|\mathbf{SAx}\|_p \leq \boxed{\kappa} \|\mathbf{Ax}\|_p$$

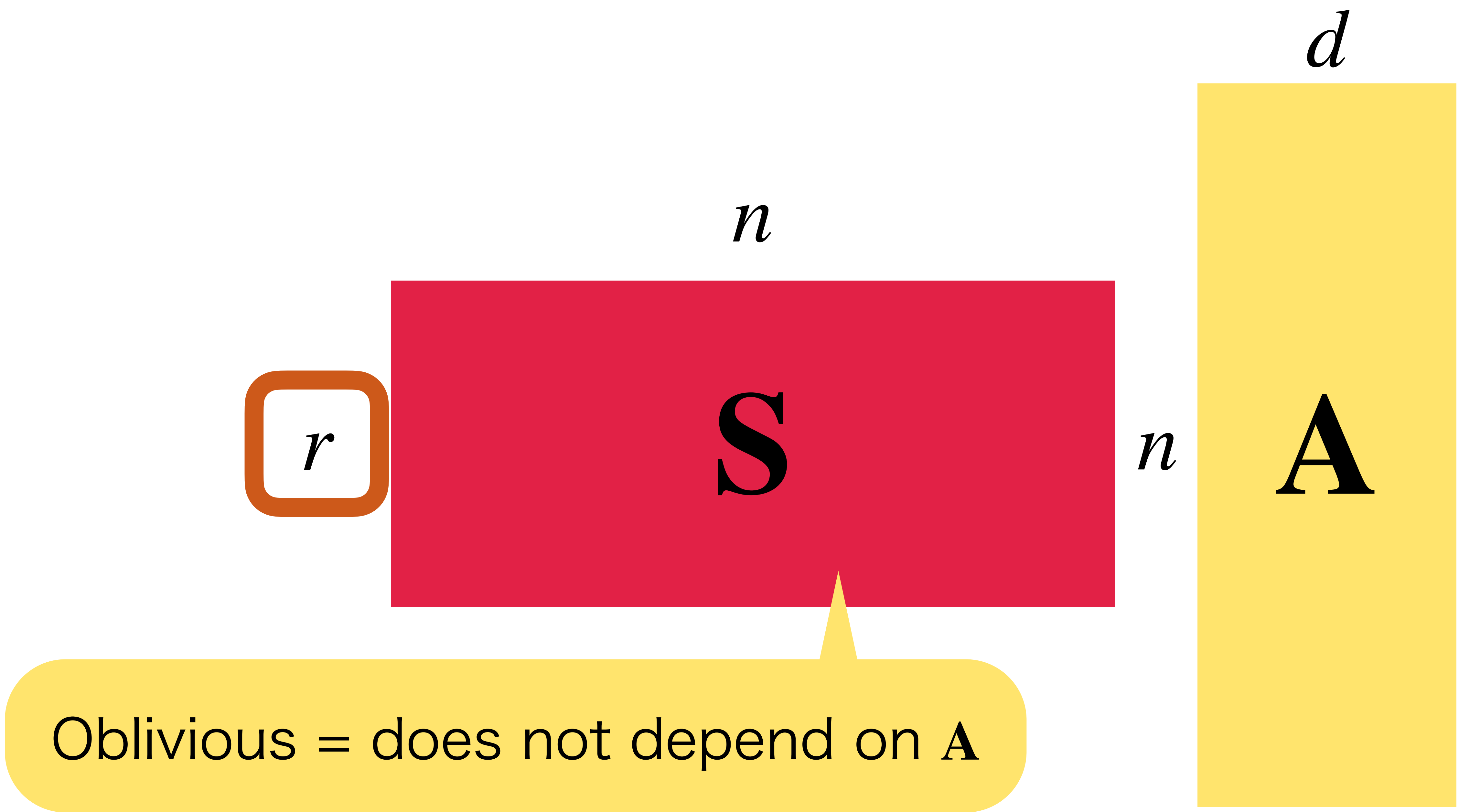
for every $\mathbf{x} \in \mathbb{R}^d$

Matrix Approximation

Oblivious ℓ_p Subspace Embeddings

Two regimes for $1 \leq p < 2$

- **High distortion**: sketch $\mathbf{S}$ has...



$$\|\mathbf{Ax}\|_p \leq \|\mathbf{SAx}\|_p \leq \boxed{\kappa}\|\mathbf{Ax}\|_p$$

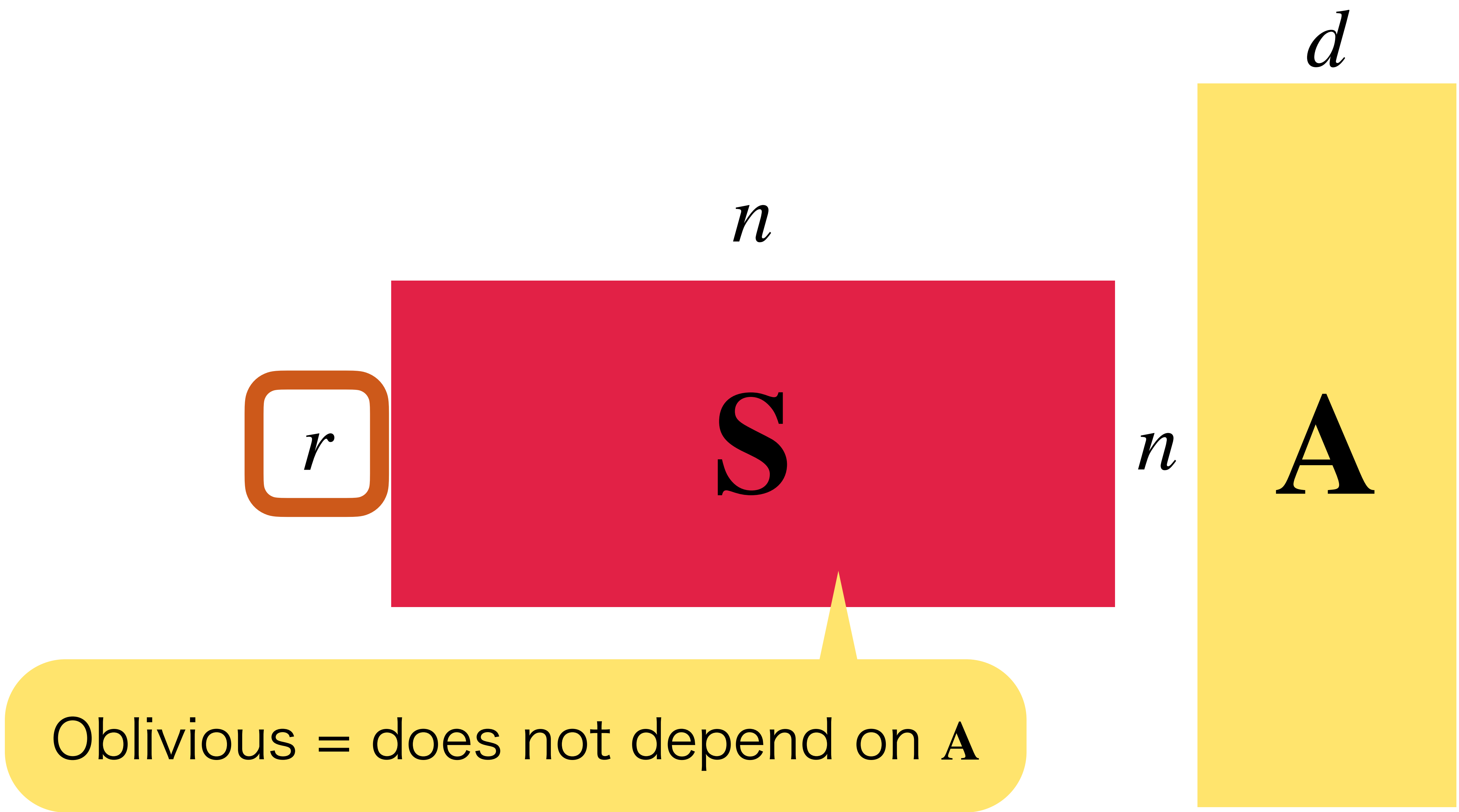
for every $\mathbf{x} \in \mathbb{R}^d$

Matrix Approximation

Oblivious ℓ_p Subspace Embeddings

Two regimes for $1 \le p < 2$

- **High distortion**: sketch $\mathbf{S}$ has…



- **Low distortion**: sketch $\mathbf{S}$ has…

$$\|\mathbf{Ax}\|_p \le \|\mathbf{SAx}\|_p \le \boxed{\kappa}\|\mathbf{Ax}\|_p$$

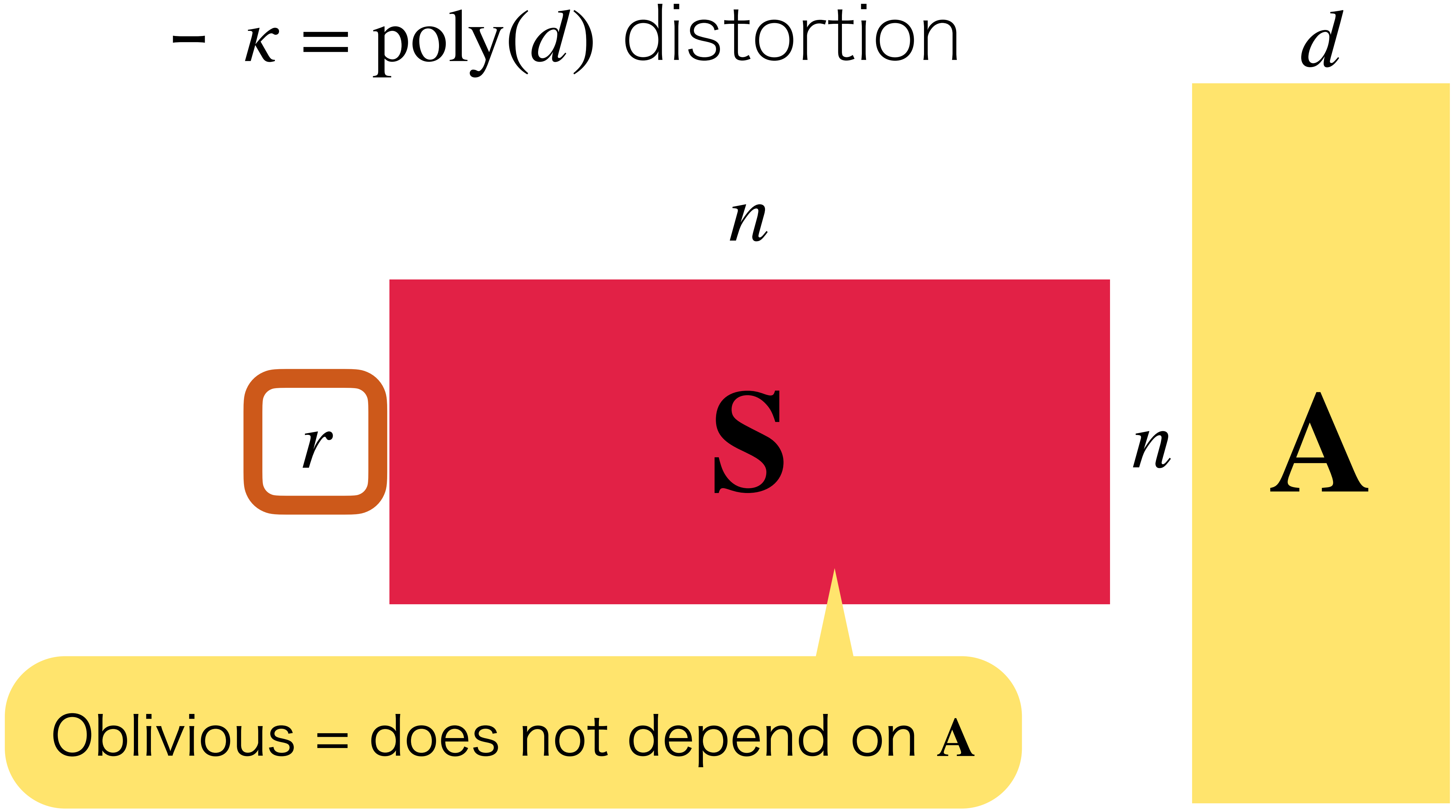
for every $\mathbf{x} \in \mathbb{R}^d$

Matrix Approximation

Oblivious ℓ_p Subspace Embeddings

Two regimes for $1 \leq p < 2$

- **High distortion**: sketch $\mathbf{S}$ has...
  - $r = \text{poly}(d)$ rows
  - $\kappa = \text{poly}(d)$ distortion



- **Low distortion**: sketch $\mathbf{S}$ has...

$$\|\mathbf{Ax}\|_p \leq \|\mathbf{SAx}\|_p \leq \boxed{\kappa}\|\mathbf{Ax}\|_p$$

for every $\mathbf{x} \in \mathbb{R}^d$

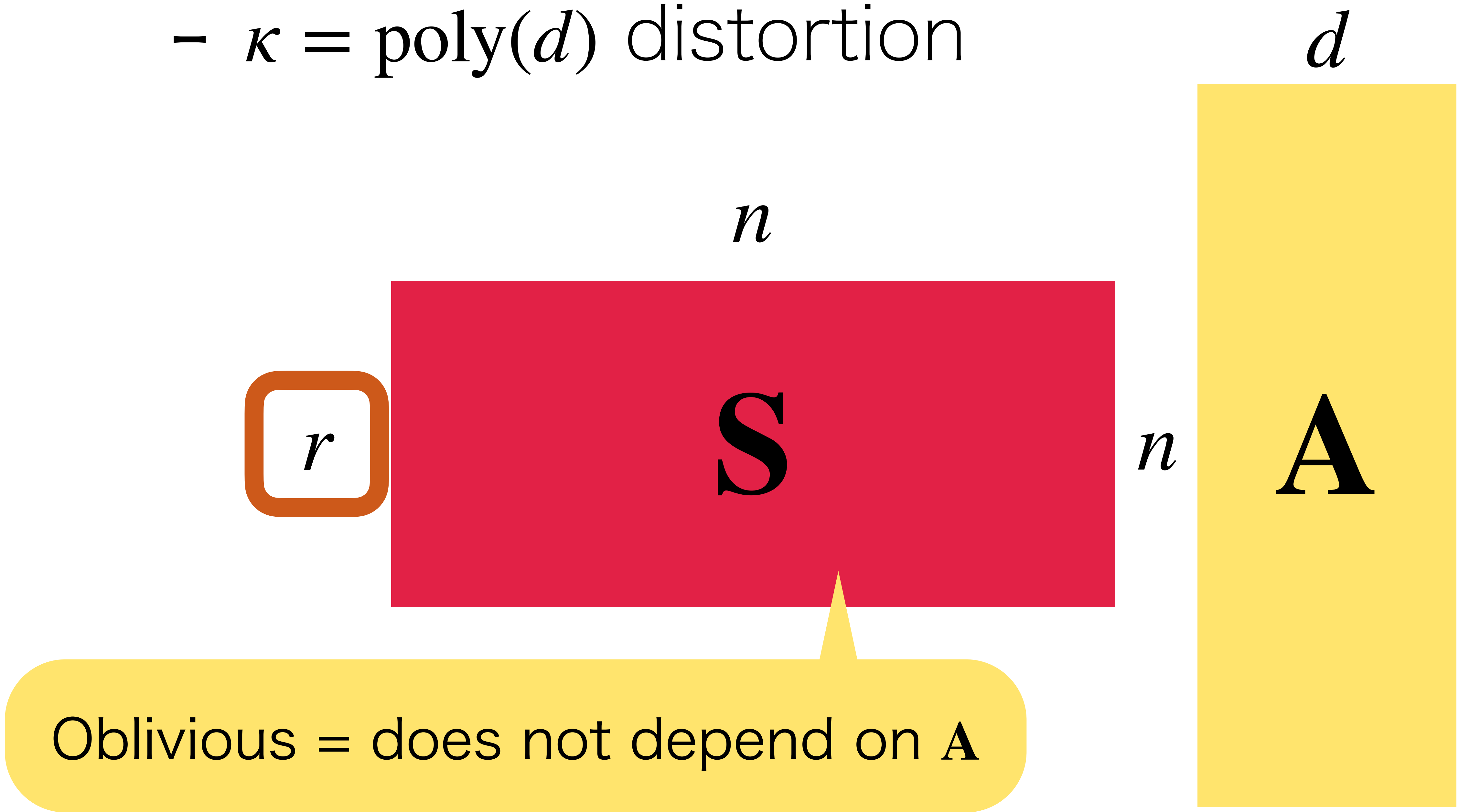
Matrix Approximation

Oblivious ℓ_p Subspace Embeddings

Two regimes for $1 \le p < 2$

- **High distortion**: sketch $\mathbf{S}$ has...
  - $r = \mathrm{poly}(d)$ rows
  - $\kappa = \mathrm{poly}(d)$ distortion

- **Low distortion**: sketch $\mathbf{S}$ has...
  - $r \gg \mathrm{poly}(d)$ rows
  - $\kappa = (1 + \varepsilon)$ distortion



$$\|\mathbf{Ax}\|_p \le \|\mathbf{SAx}\|_p \le \boxed{\kappa}\|\mathbf{Ax}\|_p$$

for every $\mathbf{x} \in \mathbb{R}^d$

Matrix Approximation

Oblivious ℓ_p Subspace Embeddings

Two regimes for $1 \leq p < 2$

- **High distortion**: sketch $\mathbf{S}$ has...
  - $r = \mathrm{poly}(d)$ rows
  - $\kappa = \mathrm{poly}(d)$ distortion

- **Low distortion**: sketch $\mathbf{S}$ has...
  - $r \gg \mathrm{poly}(d)$ rows
  - $\kappa = (1 + \varepsilon)$ distortion

Matrix Approximation

Oblivious ℓ_p Subspace Embeddings

Two regimes for $1 \leq p < 2$

- **High distortion**: sketch $\mathbf{S}$ has...
  - $r = \mathrm{poly}(d)$ rows
  - $\kappa = \mathrm{poly}(d)$ distortion

- **Low distortion**: sketch $\mathbf{S}$ has...
  - $r \gg \mathrm{poly}(d)$ rows
  - $\kappa = (1+\varepsilon)$ distortion

Best of both worlds is not possible! (Wang—Woodruff 2019)

Matrix Approximation

Oblivious ℓ_p Subspace Embeddings

Two regimes for $1 \leq p < 2$

- **High distortion**: sketch $\mathbf{S}$ has...
  - $r = \mathrm{poly}(d)$ rows
  - $\kappa = \mathrm{poly}(d)$ distortion

- **Low distortion**: sketch $\mathbf{S}$ has...
  - $r \gg \mathrm{poly}(d)$ rows
  - $\kappa = (1 + \varepsilon)$ distortion

Best of both worlds is not possible! (Wang—Woodruff 2019)

- **High distortion**: sketch $\mathbf{S}$ has...
  - $r = \mathrm{poly}(d) \Rightarrow \kappa \gtrsim d^{1/p}$

Matrix Approximation

Oblivious ℓ_p Subspace Embeddings

Two regimes for $1 \leq p < 2$

- **High distortion**: sketch $\mathbf{S}$ has...
  - $r = \mathrm{poly}(d)$ rows
  - $\kappa = \mathrm{poly}(d)$ distortion

- **Low distortion**: sketch $\mathbf{S}$ has...
  - $r \gg \mathrm{poly}(d)$ rows
  - $\kappa = (1+\varepsilon)$ distortion

Best of both worlds is not possible! (Wang—Woodruff 2019)

- **High distortion**: sketch $\mathbf{S}$ has...
  - $r = \mathrm{poly}(d) \Rightarrow \kappa \gtrsim d^{1/p}$

- **Low distortion**: sketch $\mathbf{S}$ has...
  - $\kappa = O(1)$ distortion $\Rightarrow r \gtrsim \exp(\sqrt{d})$

Matrix Approximation

Oblivious ℓ_p Subspace Embeddings

Best of both worlds is not possible! (Wang—Woodruff 2019)

- **High distortion**: sketch $\mathbf{S}$ has...
  - $r = \mathrm{poly}(d) \Rightarrow \kappa \gtrsim d^{1/p}$

- **Low distortion**: sketch $\mathbf{S}$ has...
  - $\kappa = O(1)$ distortion $\Rightarrow r \gtrsim \exp(\sqrt{d})$

Matrix Approximation

Oblivious ℓ_p Subspace Embeddings

Best of both worlds is not possible! (Wang—Woodruff 2019)

- **High distortion**: sketch $\mathbf{S}$ has...
  - $r = \mathrm{poly}(d) \Rightarrow \kappa \gtrsim d^{1/p}$
- **Low distortion**: sketch $\mathbf{S}$ has...
  - $\kappa = O(1)$ distortion $\Rightarrow r \gtrsim \exp(\sqrt{d})$

Question. Are the lower bounds of Wang—Woodruff 2019 tight?

Matrix Approximation

Oblivious ℓ_p Subspace Embeddings

Best of both worlds is not possible! (Wang—Woodruff 2019)

- **High distortion**: sketch $\mathbf{S}$ has...
  - $r = \mathrm{poly}(d) \Rightarrow \kappa \gtrsim d^{1/p}$

- **Low distortion**: sketch $\mathbf{S}$ has...
  - $\kappa = O(1)$ distortion $\Rightarrow r \gtrsim \exp(\sqrt{d})$

Question. Are the lower bounds of Wang—Woodruff 2019 tight?

Yes!

- **Woodruff—Y 2023**
  - $r = \tilde{O}(d)$, $\kappa = \tilde{O}(d^{1/p})$

Matrix Approximation

Oblivious ℓ_p Subspace Embeddings

Best of both worlds is not possible! (Wang—Woodruff 2019)

- **High distortion**: sketch $\mathbf{S}$ has...
  - $r = \mathrm{poly}(d) \Rightarrow \kappa \gtrsim d^{1/p}$

- **Low distortion**: sketch $\mathbf{S}$ has...
  - $\kappa = O(1)$ distortion $\Rightarrow r \gtrsim \exp(\sqrt{d})$

Question. Are the lower bounds of Wang—Woodruff 2019 tight?

Yes!

- **Woodruff—Y 2023**
  - $r = \tilde{O}(d)$, $\kappa = \tilde{O}(d^{1/p})$
- Prior best: Wang—Woodruff 2019
  - $r = \tilde{O}(d)$, $\kappa = \tilde{O}(d)$

Matrix Approximation

Oblivious ℓ_p Subspace Embeddings

Best of both worlds is not possible! (Wang—Woodruff 2019)

- **High distortion**: sketch $\mathbf{S}$ has...
  - $r = \mathrm{poly}(d) \Rightarrow \kappa \gtrsim d^{1/p}$

- **Low distortion**: sketch $\mathbf{S}$ has...
  - $\kappa = O(1)$ distortion $\Rightarrow r \gtrsim \exp(\sqrt{d})$

Question. Are the lower bounds of Wang—Woodruff 2019 tight?

Yes!

- **Woodruff—Y 2023**
  - $r = \tilde{O}(d)$, $\kappa = \tilde{O}(d^{1/p})$
- Prior best: Wang—Woodruff 2019
  - $r = \tilde{O}(d)$, $\kappa = \tilde{O}(d)$

Almost!

Matrix Approximation

Oblivious ℓ_p Subspace Embeddings

Best of both worlds is not possible! (Wang—Woodruff 2019)

- **High distortion**: sketch $\mathbf{S}$ has…
  - $r = \mathrm{poly}(d) \Rightarrow \kappa \gtrsim d^{1/p}$
- **Low distortion**: sketch $\mathbf{S}$ has…
  - $\kappa = O(1)$ distortion $\Rightarrow r \gtrsim \exp(\sqrt{d})$

Question. Are the lower bounds of Wang—Woodruff 2019 tight?

Yes!

- **Woodruff—Y 2023**
  - $r = \tilde{O}(d)$, $\kappa = \tilde{O}(d^{1/p})$
- Prior best: Wang—Woodruff 2019
  - $r = \tilde{O}(d)$, $\kappa = \tilde{O}(d)$

Almost!

- **Li—Woodruff—Y 2021**
  - $r = \exp(\varepsilon^{-1} d)$, $\kappa = (1 + \varepsilon)$

Matrix Approximation

Oblivious ℓ_p Subspace Embeddings

Best of both worlds is not possible! (Wang—Woodruff 2019)

- **High distortion**: sketch $\mathbf{S}$ has…
  - $r = \mathrm{poly}(d) \Rightarrow \kappa \gtrsim d^{1/p}$

- **Low distortion**: sketch $\mathbf{S}$ has…
  - $\kappa = O(1)$ distortion $\Rightarrow r \gtrsim \exp(\sqrt{d})$

Question. Are the lower bounds of Wang—Woodruff 2019 tight?

Yes!

- **Woodruff—Y 2023**
  - $r = \tilde{O}(d)$, $\kappa = \tilde{O}(d^{1/p})$
- Prior best: Wang—Woodruff 2019
  - $r = \tilde{O}(d)$, $\kappa = \tilde{O}(d)$

Almost!

- **Li—Woodruff—Y 2021**
  - $r = \exp(\varepsilon^{-1} d)$, $\kappa = (1 + \varepsilon)$
- Prior best: Wang—Woodruff 2019
  - $r = \exp(\exp(\varepsilon^{-2} d))$, $\kappa = (1 + \varepsilon)$

Matrix Approximation

Oblivious ℓ_p Subspace Embeddings

Best of both worlds is not possible! (Wang—Woodruff 2019)

- **High distortion**: sketch $\mathbf{S}$ has...
  - $r = \mathrm{poly}(d) \Rightarrow \kappa \gtrsim d^{1/p}$

- **Low distortion**: sketch $\mathbf{S}$ has...
  - $\kappa = O(1)$ distortion $\Rightarrow r \gtrsim \exp(\sqrt{d})$

Question. Are the lower bounds of Wang—Woodruff 2019 tight?

Yes!

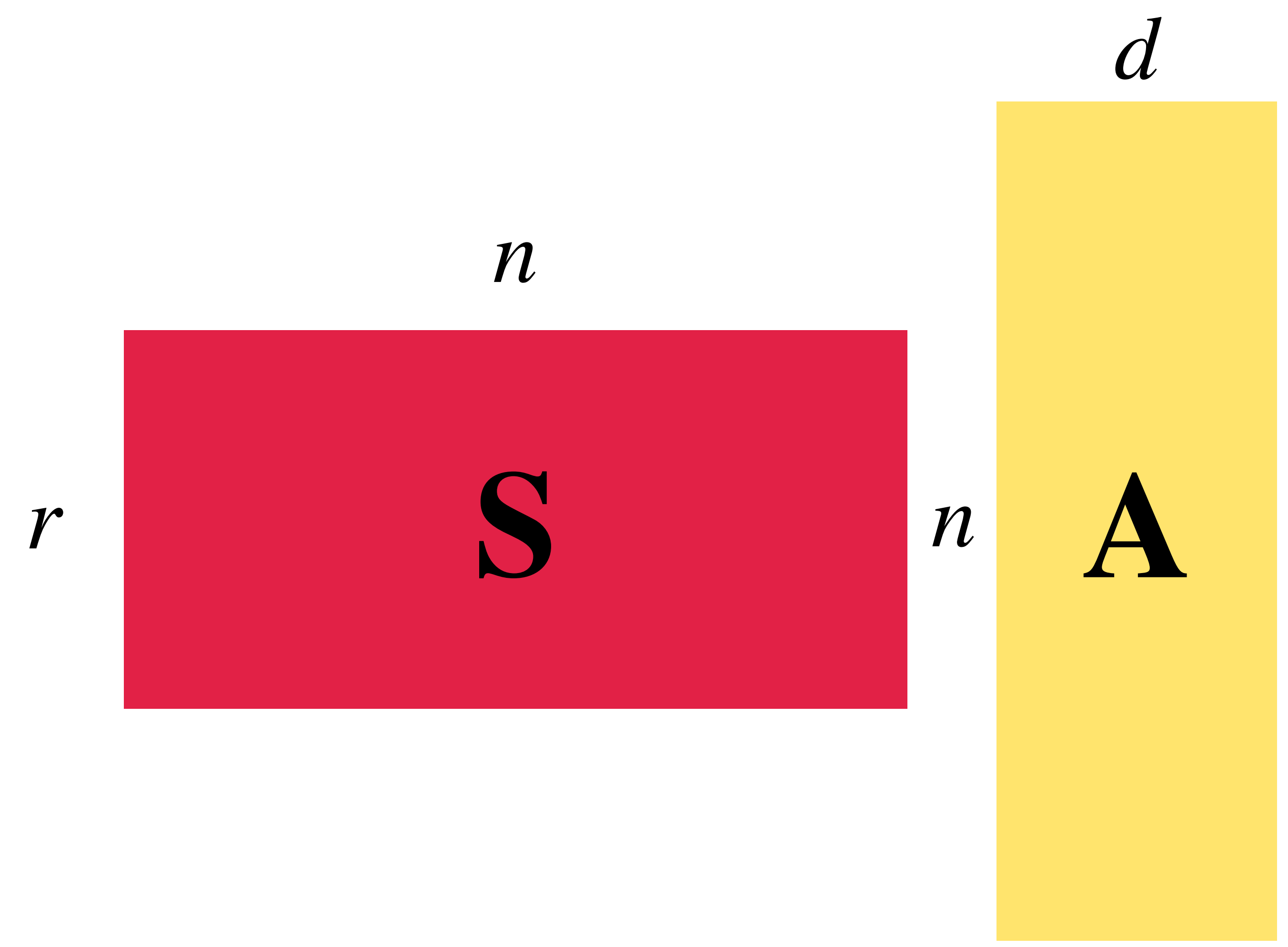
- **Woodruff—Y 2023**
  - $r = \tilde{O}(d),\ \kappa = \tilde{O}(d^{1/p})$
- Prior best: Wang—Woodruff 2019
  - $r = \tilde{O}(d),\ \kappa = \tilde{O}(d)$

Almost!

- **Li—Woodruff—Y 2021**
  - $r = \exp(\varepsilon^{-1} d),\ \kappa = (1+\varepsilon)$
- Prior best: Wang—Woodruff 2019
  - $r = \exp(\exp(\varepsilon^{-2} d)),\ \kappa = (1+\varepsilon)$

Matrix Approximation

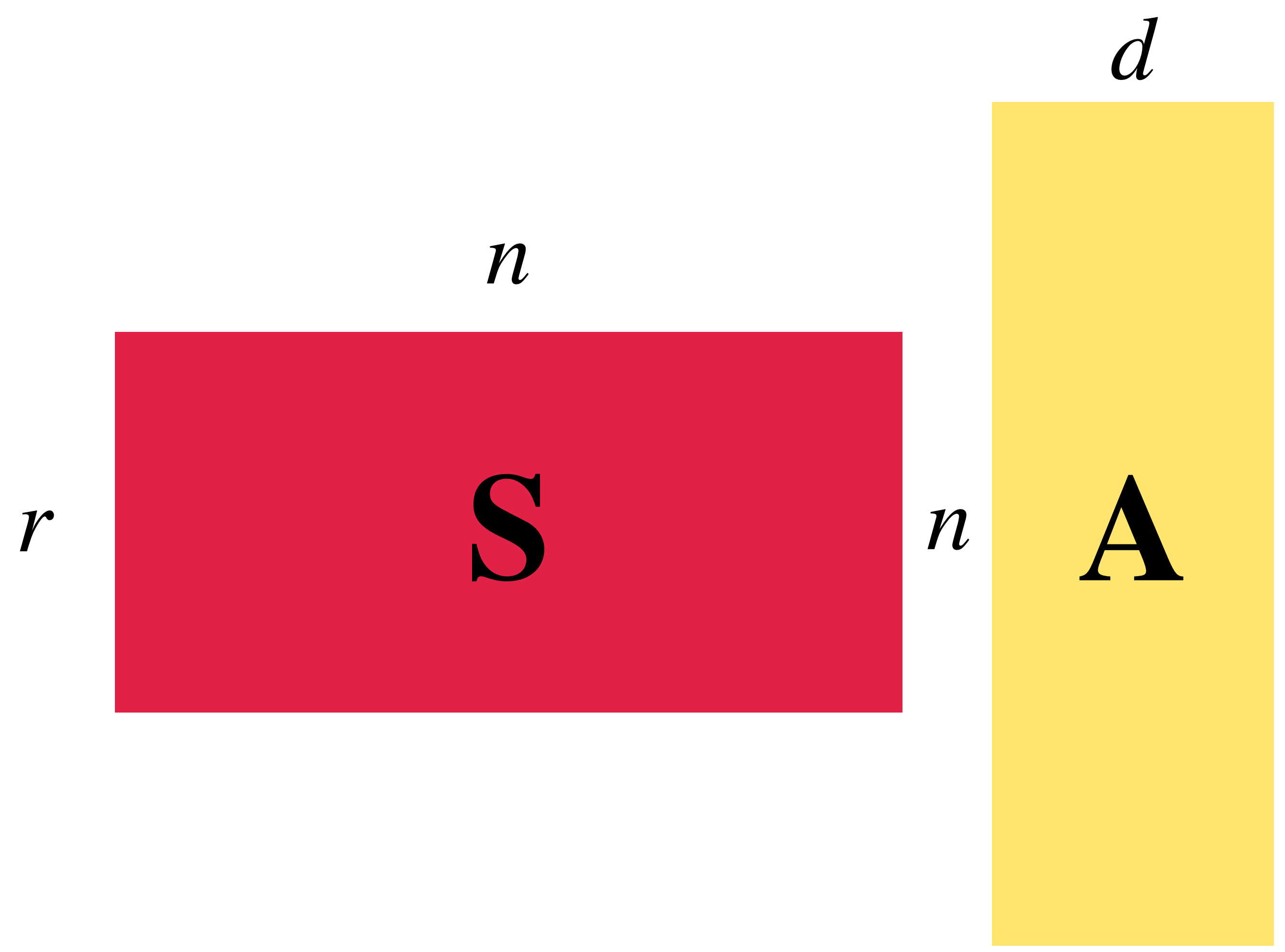
Oblivious ℓ_p Subspace Embeddings



Matrix Approximation

Oblivious ℓ_p Subspace Embeddings

Fact. Oblivious ℓ_p subspace embeddings reduce to constructing **well-conditioned bases** for subspaces

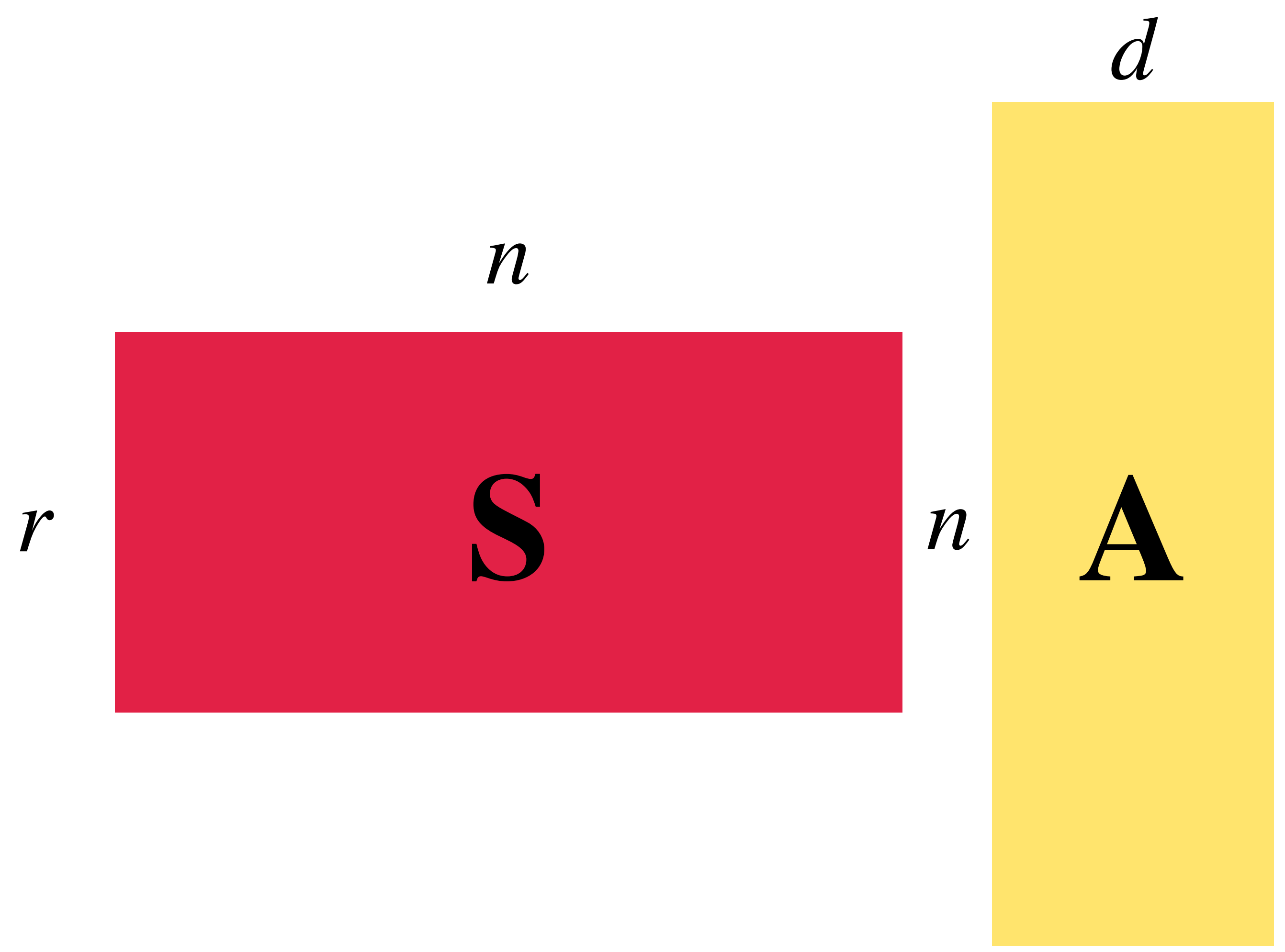


Matrix Approximation

Oblivious ℓ_p Subspace Embeddings

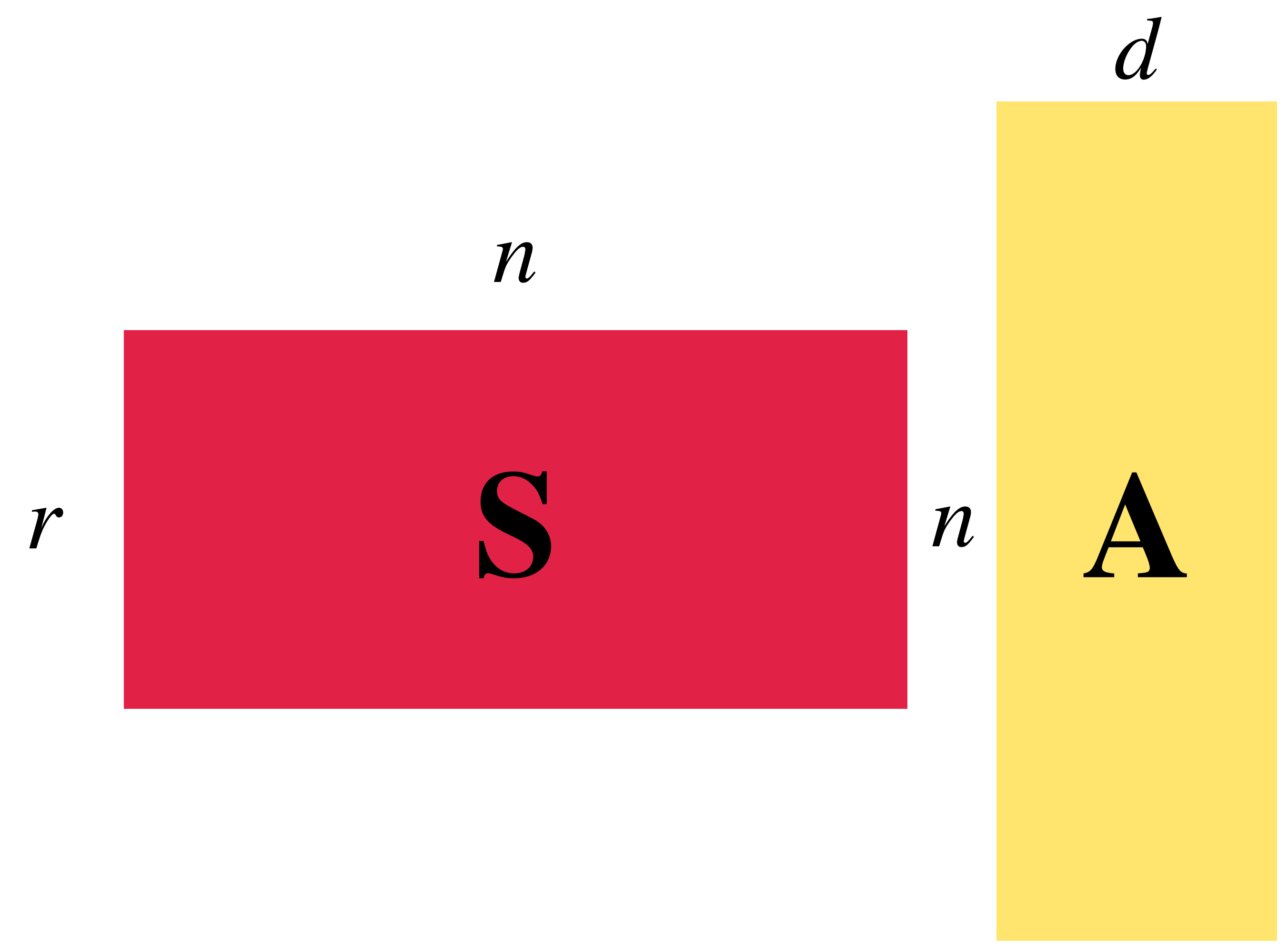
Fact. Oblivious ℓ_p subspace embeddings reduce to constructing **well-conditioned bases** for subspaces

≈ orthonormal bases for ℓ_p norms



Matrix Approximation

Oblivious ℓ_p Subspace Embeddings



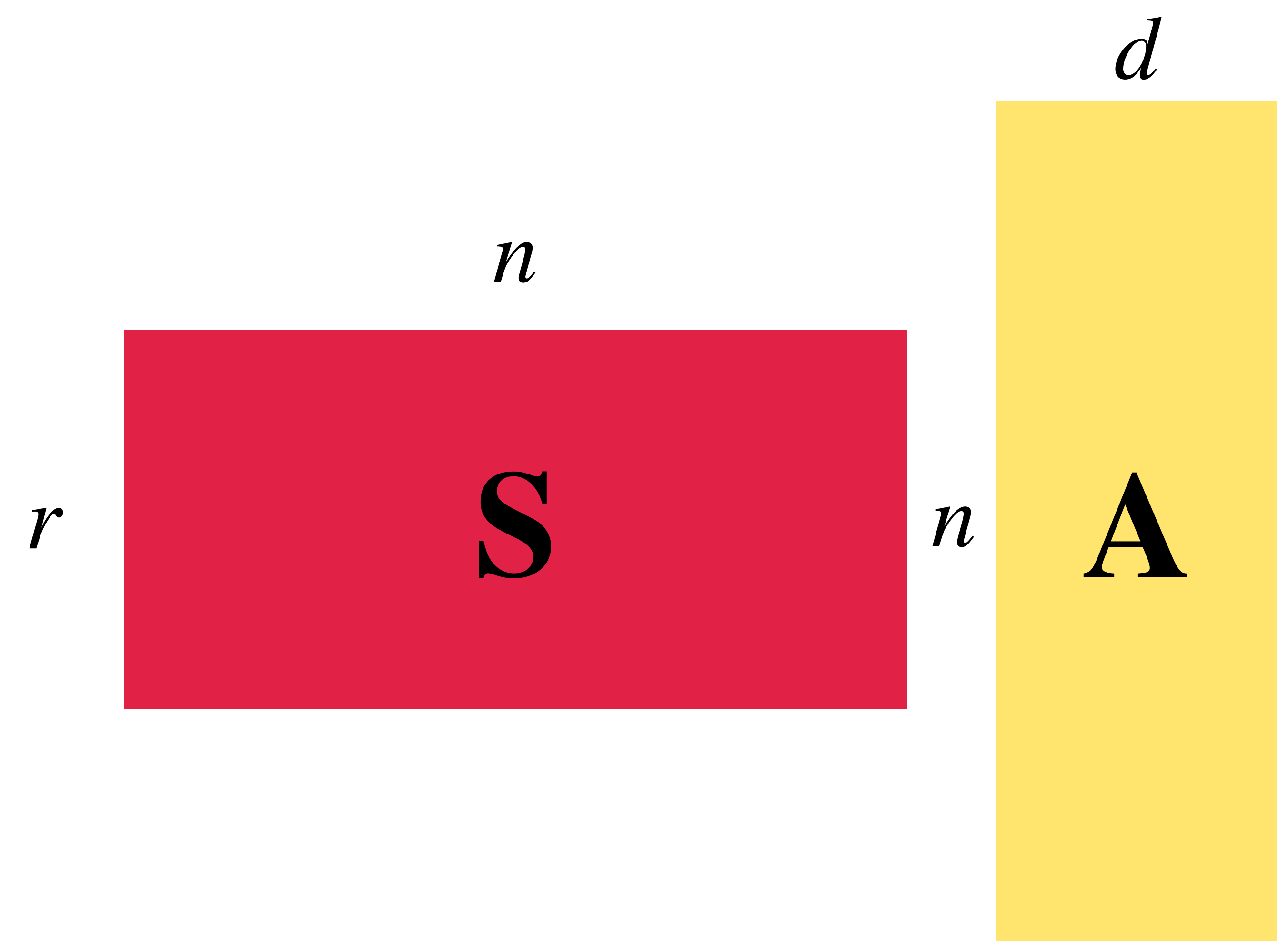
Fact. Oblivious ℓ_p subspace embeddings reduce to constructing **well-conditioned bases** for subspaces

≈ orthonormal bases for ℓ_p norms

- Let $\mathbf{U}$ be an orthonormal basis for $\mathbf{A}$

Matrix Approximation

Oblivious ℓ_p Subspace Embeddings



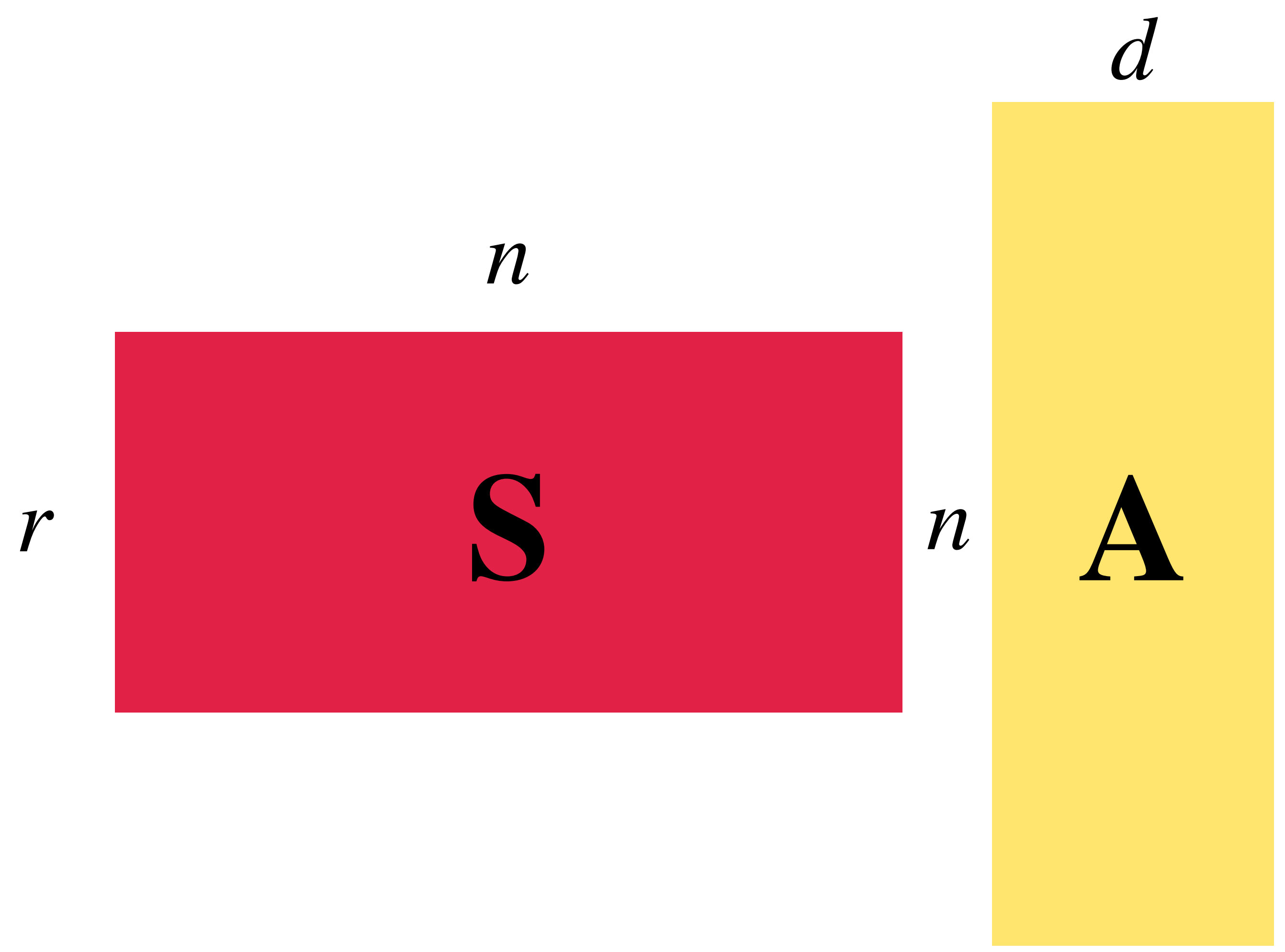
Fact. Oblivious ℓ_p subspace embeddings reduce to constructing **well-conditioned bases** for subspaces

≈ orthonormal bases for ℓ_p norms

- Let $\mathbf{U}$ be an orthonormal basis for $\mathbf{A}$
  - $\|\mathbf{U}\|_F \leq d^{1/2}$ (with equality)

Matrix Approximation

Oblivious ℓ_p Subspace Embeddings



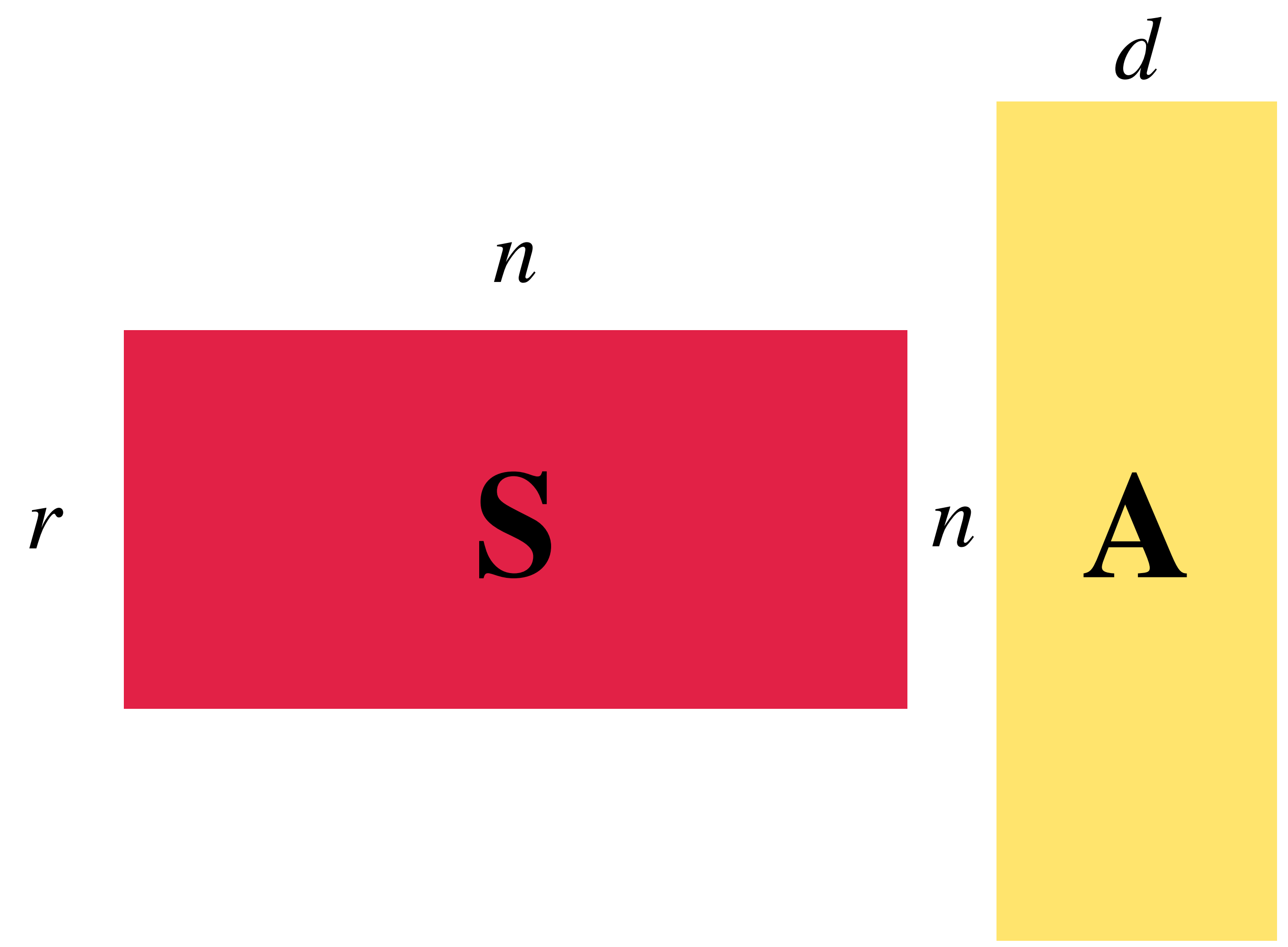
Fact. Oblivious ℓ_p subspace embeddings reduce to constructing **well-conditioned bases** for subspaces

≈ orthonormal bases for ℓ_p norms

- Let $\mathbf{U}$ be an orthonormal basis for $\mathbf{A}$
  - $\|\mathbf{U}\|_F \leq d^{1/2}$ (with equality)
  - $\|\mathbf{x}\|_2 \leq \|\mathbf{U}\mathbf{x}\|_2$ for every $\mathbf{x} \in \mathbb{R}^d$ (with equality)

Matrix Approximation

Oblivious ℓ_p Subspace Embeddings



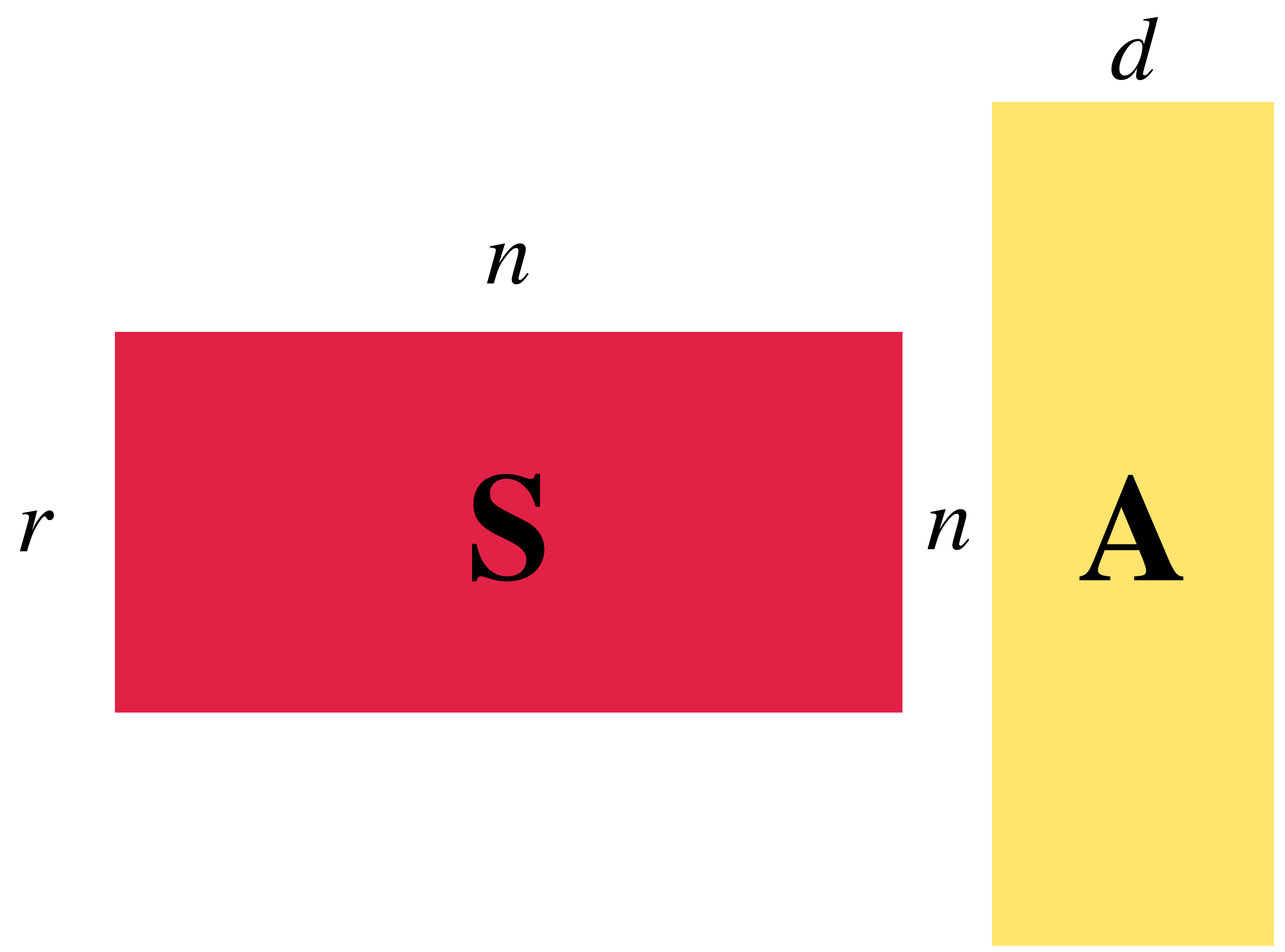
Fact. Oblivious ℓ_p subspace embeddings reduce to constructing **well-conditioned bases** for subspaces

≈ orthonormal bases for ℓ_p norms

- Let $\mathbf{U}$ be an ~~orthonormal basis~~ well-conditioned basis for $\mathbf{A}$
  - $\|\mathbf{U}\|_F \leq d^{1/2}$ (with equality)
  - $\|\mathbf{x}\|_2 \leq \|\mathbf{U}\mathbf{x}\|_2$ for every $\mathbf{x} \in \mathbb{R}^d$ (with equality)

Matrix Approximation

Oblivious ℓ_p Subspace Embeddings



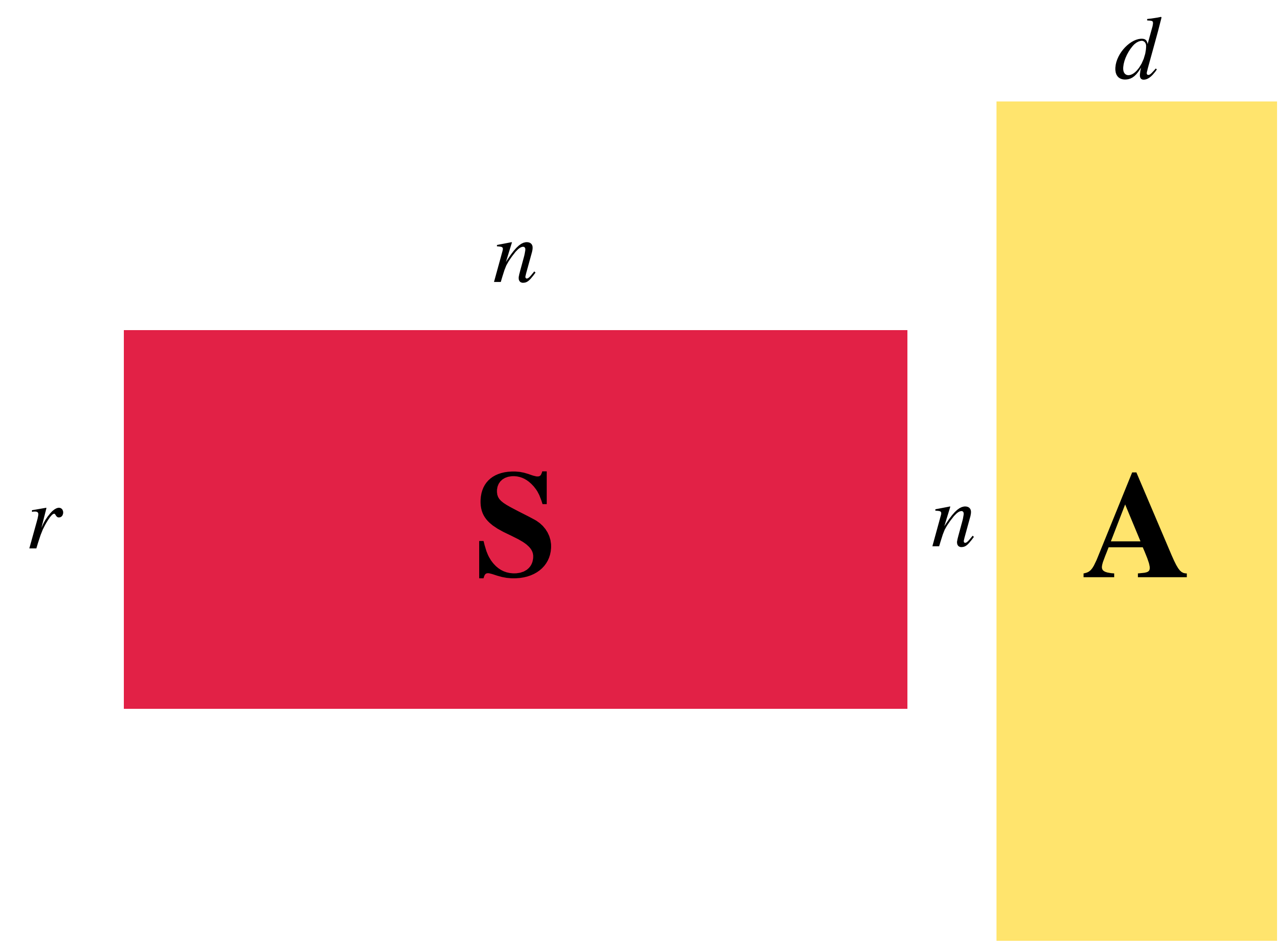
Fact. Oblivious ℓ_p subspace embeddings reduce to constructing **well-conditioned bases** for subspaces

≈ orthonormal bases for ℓ_p norms

- Let $\mathbf{U}$ be an ~~orthonormal basis~~ well-conditioned basis for $\mathbf{A}$
  - ~~$\|\mathbf{U}\|_F \leq d^{1/2}$ (with equality)~~ $\|\mathbf{U}\|_{p,p} \leq \alpha$ entrywise ℓ_p norm
  - $\|\mathbf{x}\|_2 \leq \|\mathbf{U}\mathbf{x}\|_2$ for every $\mathbf{x} \in \mathbb{R}^d$ (with equality)

Matrix Approximation

Oblivious ℓ_p Subspace Embeddings



Fact. Oblivious ℓ_p subspace embeddings reduce to constructing **well-conditioned bases** for subspaces

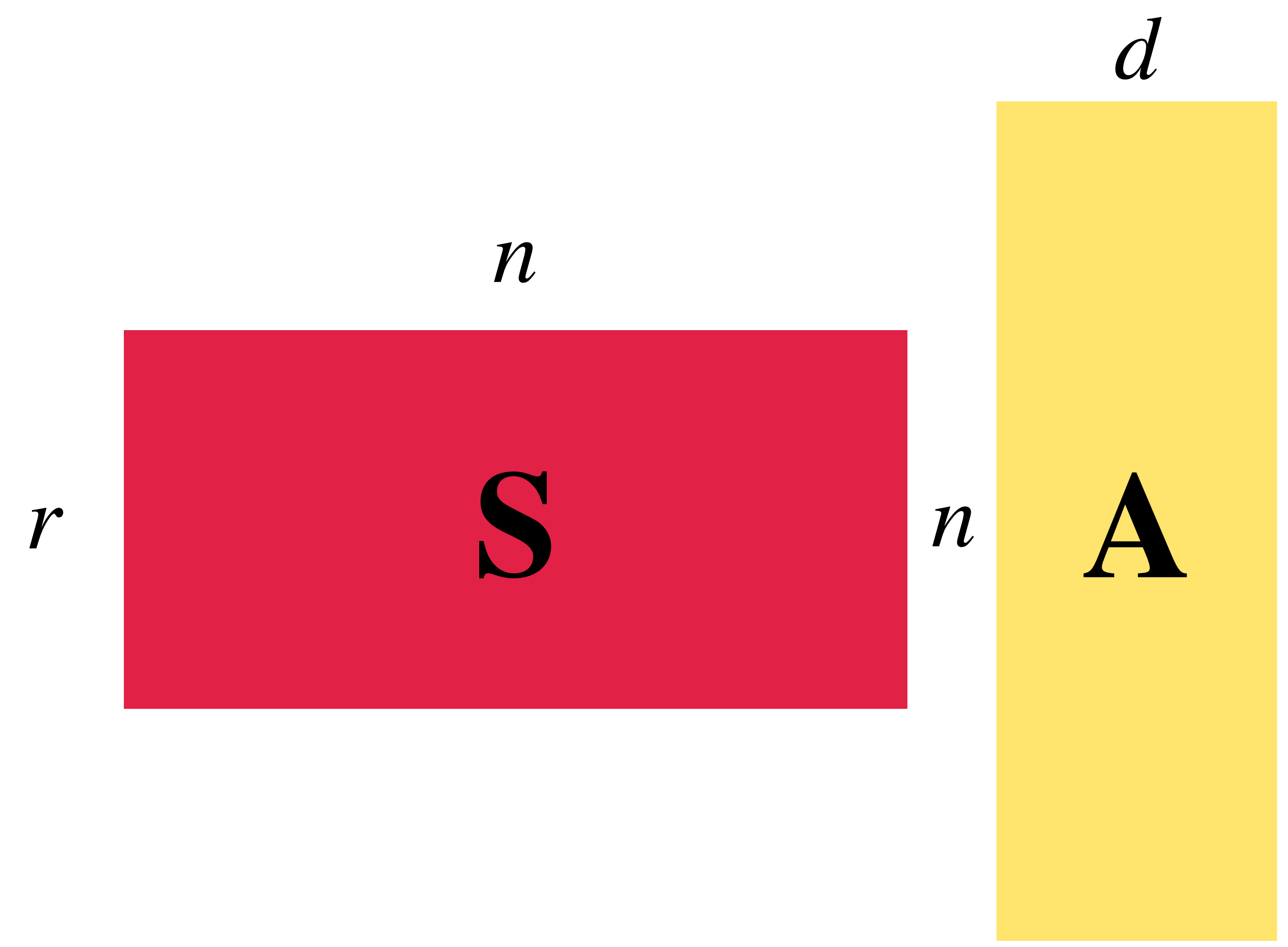
$\approx$ orthonormal bases for ℓ_p norms

- Let $\mathbf{U}$ be an ~~orthonormal basis~~ well-conditioned basis for $\mathbf{A}$
  - $\cancel{\|\mathbf{U}\|_F \leq d^{1/2}}$ ~~(with equality)~~
    $\|\mathbf{U}\|_{p,p} \leq \alpha$ entrywise ℓ_p norm
  - $\cancel{\|\mathbf{x}\|_2 \leq \|\mathbf{U}\mathbf{x}\|_2}$ ~~for every~~ $\cancel{\mathbf{x} \in \mathbb{R}^d}$ ~~(with equality)~~

    $\|\mathbf{x}\|_q \leq \|\mathbf{U}\mathbf{x}\|_p$ for every $\mathbf{x} \in \mathbb{R}^d$

Matrix Approximation

Oblivious ℓ_p Subspace Embeddings



Fact. Oblivious ℓ_p subspace embeddings reduce to constructing **well-conditioned bases** for subspaces

≈ orthonormal bases for ℓ_p norms

- Let $\mathbf{U}$ be an ~~orthonormal basis~~ for $\mathbf{A}$
  well-conditioned basis
  - $\cancel{\|\mathbf{U}\|_F \leq d^{1/2}}$ ~~(with equality)~~
    $\|\mathbf{U}\|_{p,p} \leq \alpha$ entrywise ℓ_p norm
  - $\cancel{\|\mathbf{x}\|_2 \leq \|\mathbf{U}\mathbf{x}\|_2}$ ~~for every~~ $\cancel{\mathbf{x} \in \mathbb{R}^d}$ ~~(with equality)~~

    $\|\mathbf{x}\|_q \leq \|\mathbf{U}\mathbf{x}\|_p$ for every $\mathbf{x} \in \mathbb{R}^d$

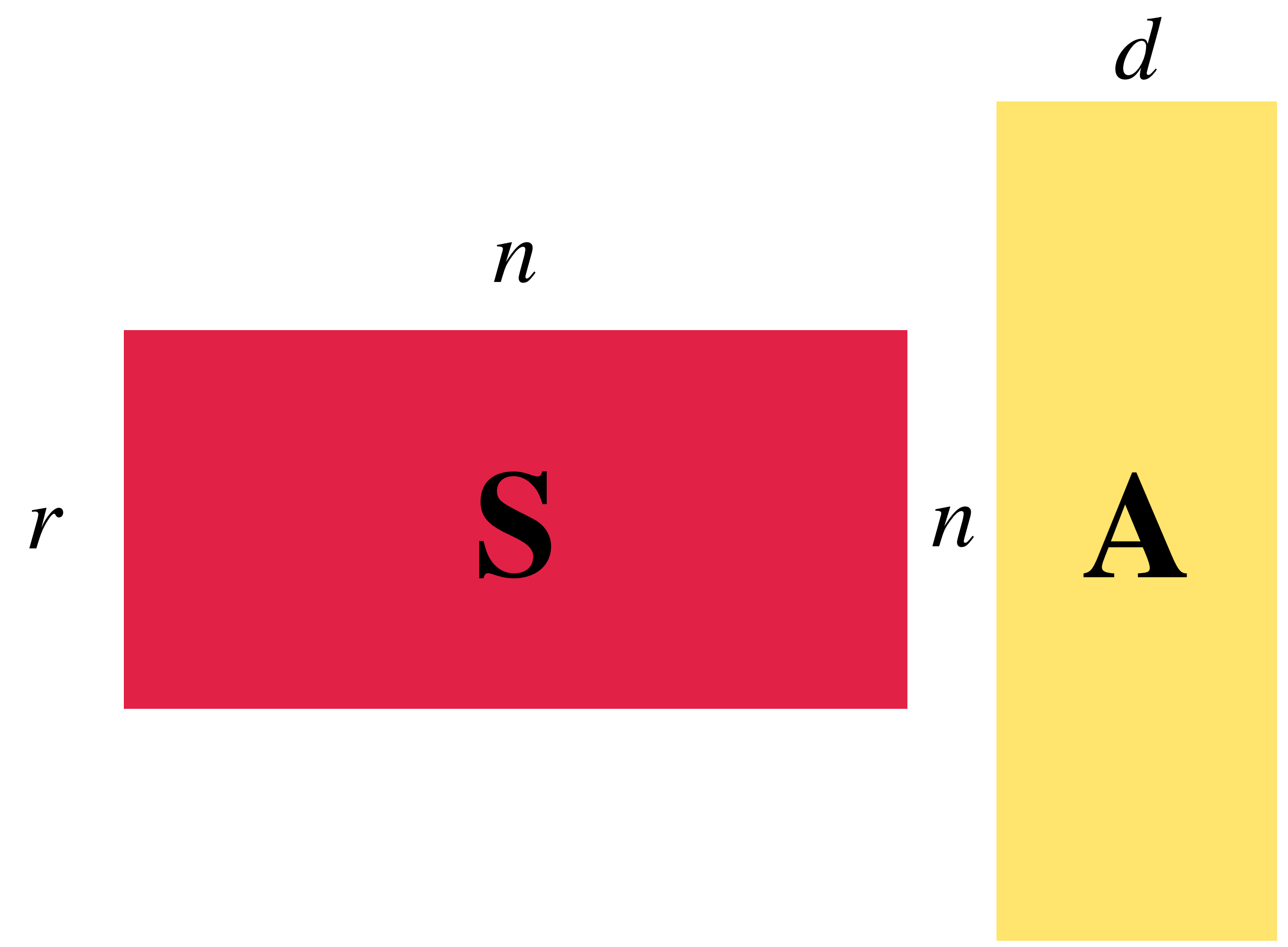
Hölder conjugate, $\frac{1}{p} + \frac{1}{q} = 1$

Matrix Approximation

Oblivious ℓ_p Subspace Embeddings

Fact. Oblivious ℓ_p subspace embeddings reduce to constructing **well-conditioned bases** for subspaces

≈ orthonormal bases for ℓ_p norms



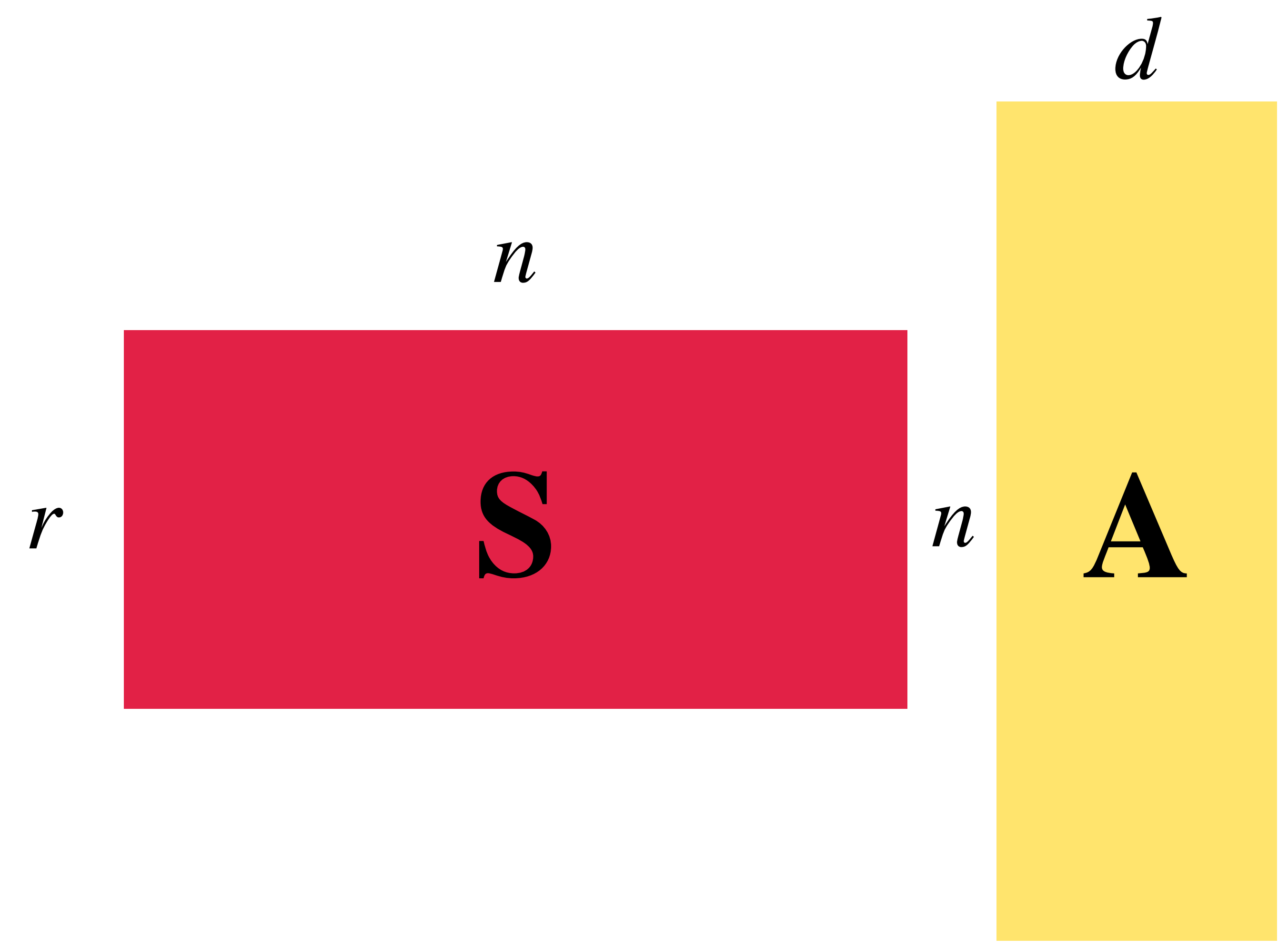
- Let $\mathbf{U}$ be an ~~orthonormal basis~~ for $\mathbf{A}$
  well-conditioned basis
  - $\cancel{\|\mathbf{U}\|_F \leq d^{1/2}}$ ~~(with equality)~~
    $\|\mathbf{U}\|_{p,p} \leq \alpha$ entrywise ℓ_p norm
  - $\cancel{\|\mathbf{x}\|_2 \leq \|\mathbf{U}\mathbf{x}\|_2}$ ~~for every~~ $\cancel{\mathbf{x} \in \mathbb{R}^d}$ ~~(with equality)~~
    $\|\mathbf{x}\|_q \leq \|\mathbf{U}\mathbf{x}\|_p$ for every $\mathbf{x} \in \mathbb{R}^d$

Hölder conjugate, $\frac{1}{p} + \frac{1}{q} = 1$

Theorem (Auerbach 1930). For any $\mathbf{A}$, there is $\mathbf{U}$ with $\alpha = d$.

Matrix Approximation

Oblivious ℓ_p Subspace Embeddings



Fact. Oblivious ℓ_p subspace embeddings reduce to constructing **well-conditioned bases** for subspaces

≈ orthonormal bases for ℓ_p norms

- Let $\mathbf{U}$ be an ~~orthonormal basis~~ for $\mathbf{A}$
  well-conditioned basis
  - ~~$\|\mathbf{U}\|_F \leq d^{1/2}$ (with equality)~~
    $\|\mathbf{U}\|_{p,p} \leq \alpha$ entrywise ℓ_p norm
  - ~~$\|\mathbf{x}\|_2 \leq \|\mathbf{U}\mathbf{x}\|_2$ for every $\mathbf{x} \in \mathbb{R}^d$ (with equality)~~
    $\|\mathbf{x}\|_q \leq \|\mathbf{U}\mathbf{x}\|_p$ for every $\mathbf{x} \in \mathbb{R}^d$

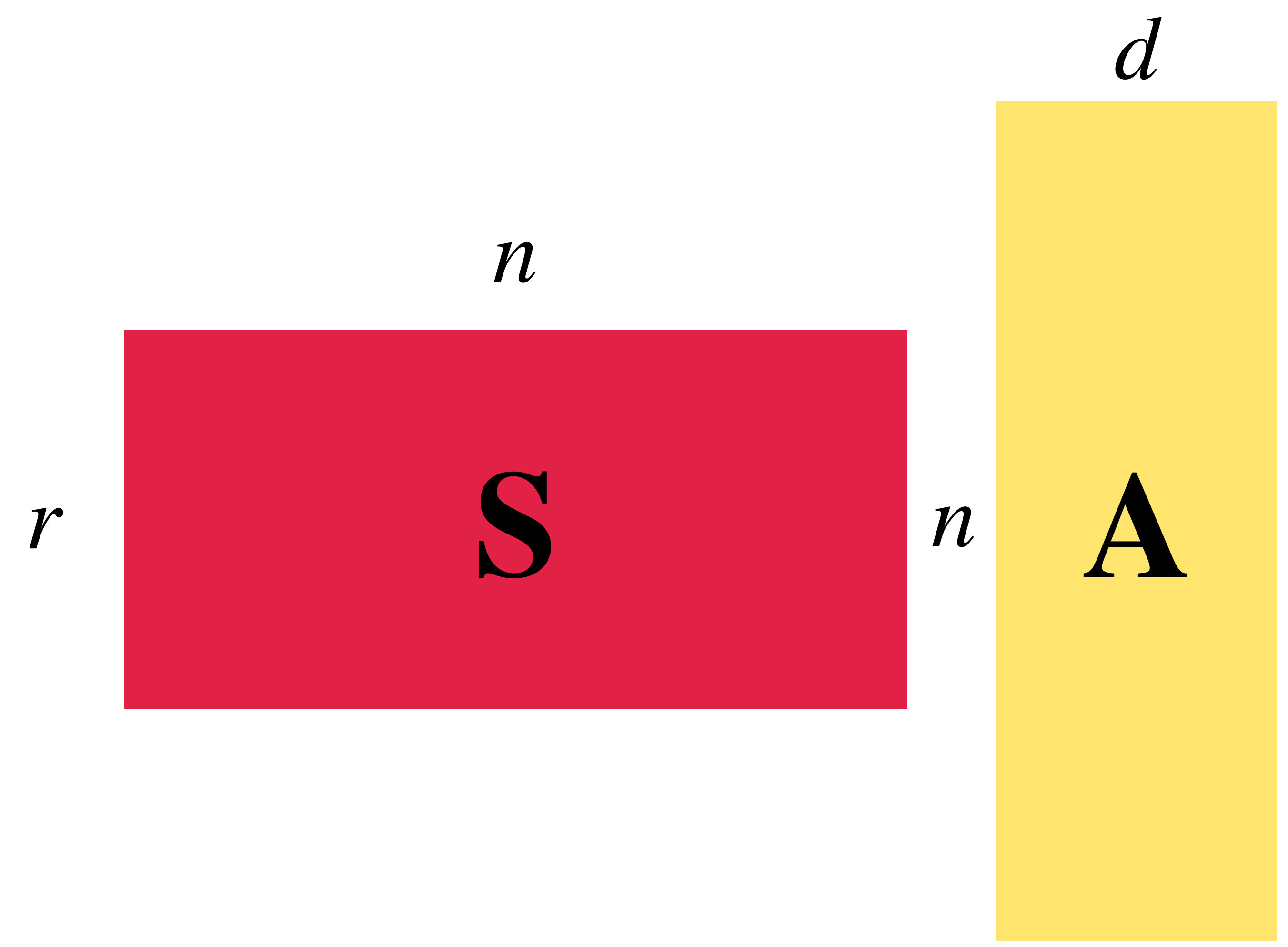
Hölder conjugate, $\frac{1}{p} + \frac{1}{q} = 1$

Theorem (Auerbach 1930). For any $\mathbf{A}$, there is $\mathbf{U}$ with $\alpha = d$.

Conjecture. For any $\mathbf{A}$, there is $\mathbf{U}$ with $\alpha = d^{1/p}$.

Matrix Approximation

Oblivious ℓ_p Subspace Embeddings



Fact. Oblivious ℓ_p subspace embeddings reduce to constructing **well-conditioned bases** for subspaces

≈ orthonormal bases for ℓ_p norms

- Let $\mathbf{U}$ be an ~~orthonormal basis~~ well-conditioned basis for $\mathbf{A}$
  - $\cancel{\|\mathbf{U}\|_F \leq d^{1/2}}$ ~~(with equality)~~
    $\|\mathbf{U}\|_{p,p} \leq \alpha$ — entrywise ℓ_p norm
  - $\cancel{\|\mathbf{x}\|_2 \leq \|\mathbf{U}\mathbf{x}\|_2}$ ~~for every~~ $\cancel{\mathbf{x} \in \mathbb{R}^d}$ ~~(with equality)~~
    $\|\mathbf{x}\|_q \leq \|\mathbf{U}\mathbf{x}\|_p$ for every $\mathbf{x} \in \mathbb{R}^d$

Hölder conjugate, $\frac{1}{p} + \frac{1}{q} = 1$

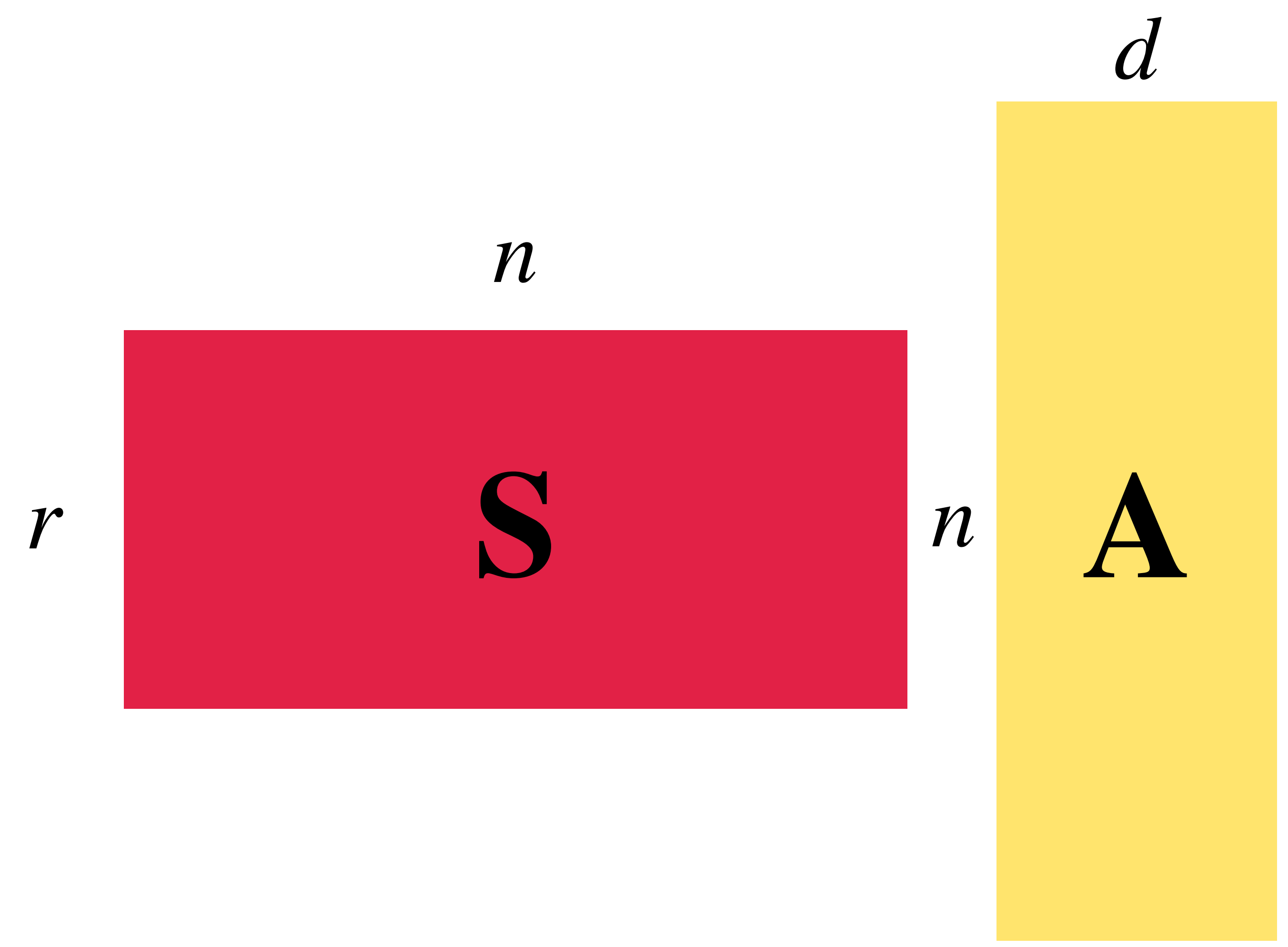
Theorem (Auerbach 1930). For any $\mathbf{A}$, there is $\mathbf{U}$ with $\alpha = d$.

Conjecture. For any $\mathbf{A}$, there is $\mathbf{U}$ with $\alpha = d^{1/p}$.

No success in showing this conjecture! ☹️

Matrix Approximation

Oblivious ℓ_p Subspace Embeddings



Fact. Oblivious ℓ_p subspace embeddings reduce to constructing **well-conditioned bases** for subspaces

$\approx$ orthonormal bases for ℓ_p norms

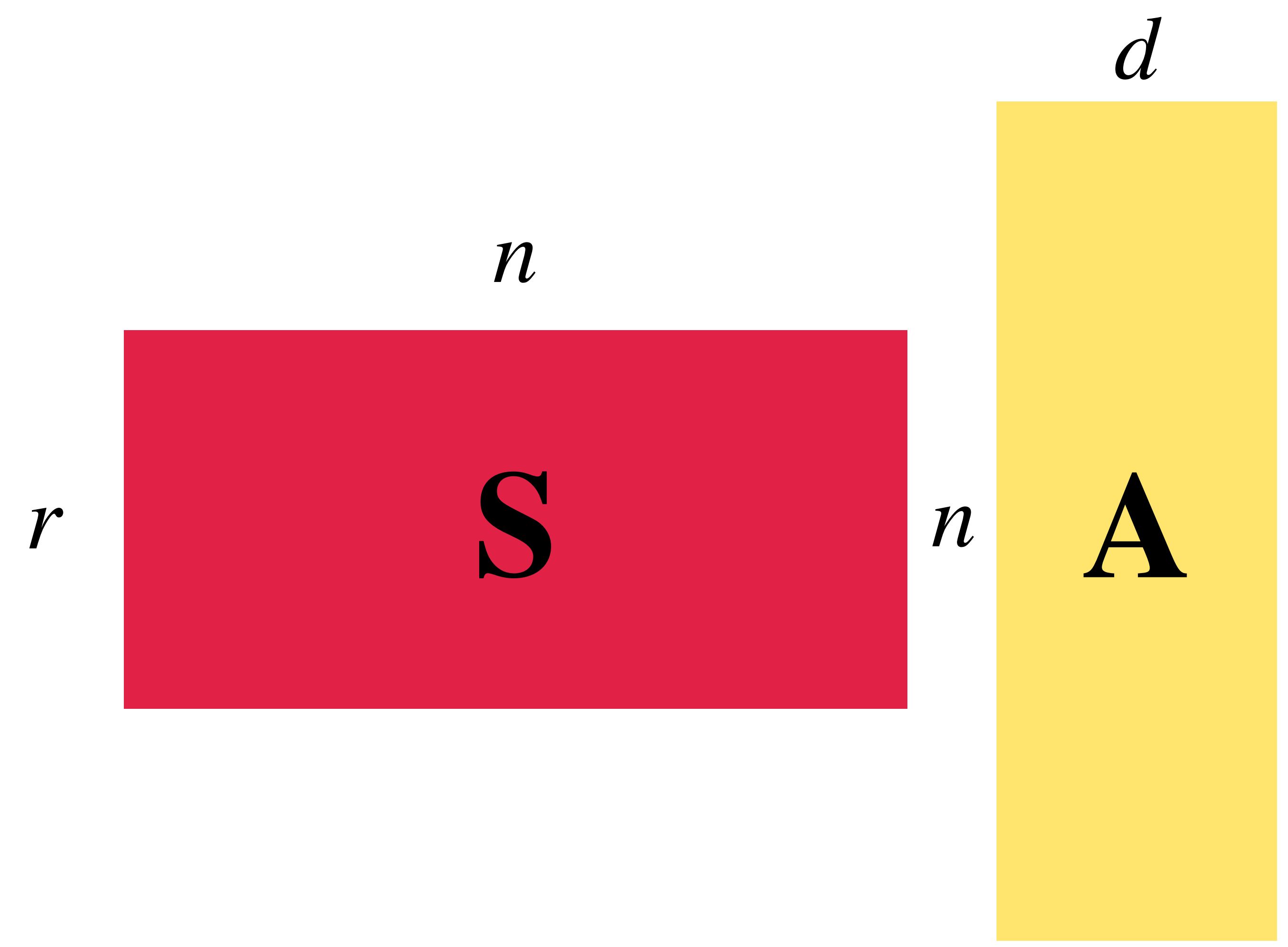
Theorem (Auerbach 1930). For any $\mathbf{A}$, there is $\mathbf{U}$ with $\alpha = d$.

Conjecture. For any $\mathbf{A}$, there is $\mathbf{U}$ with $\alpha = d^{1/p}$.

No success in showing this conjecture! ☹️

Matrix Approximation

Oblivious ℓ_p Subspace Embeddings



Fact. Oblivious ℓ_p subspace embeddings reduce to constructing **well-conditioned bases** for subspaces

$\approx$ orthonormal bases for ℓ_p norms

- **Idea.** What if we relax well-conditioned **bases** to well-conditioned **spanning sets**?

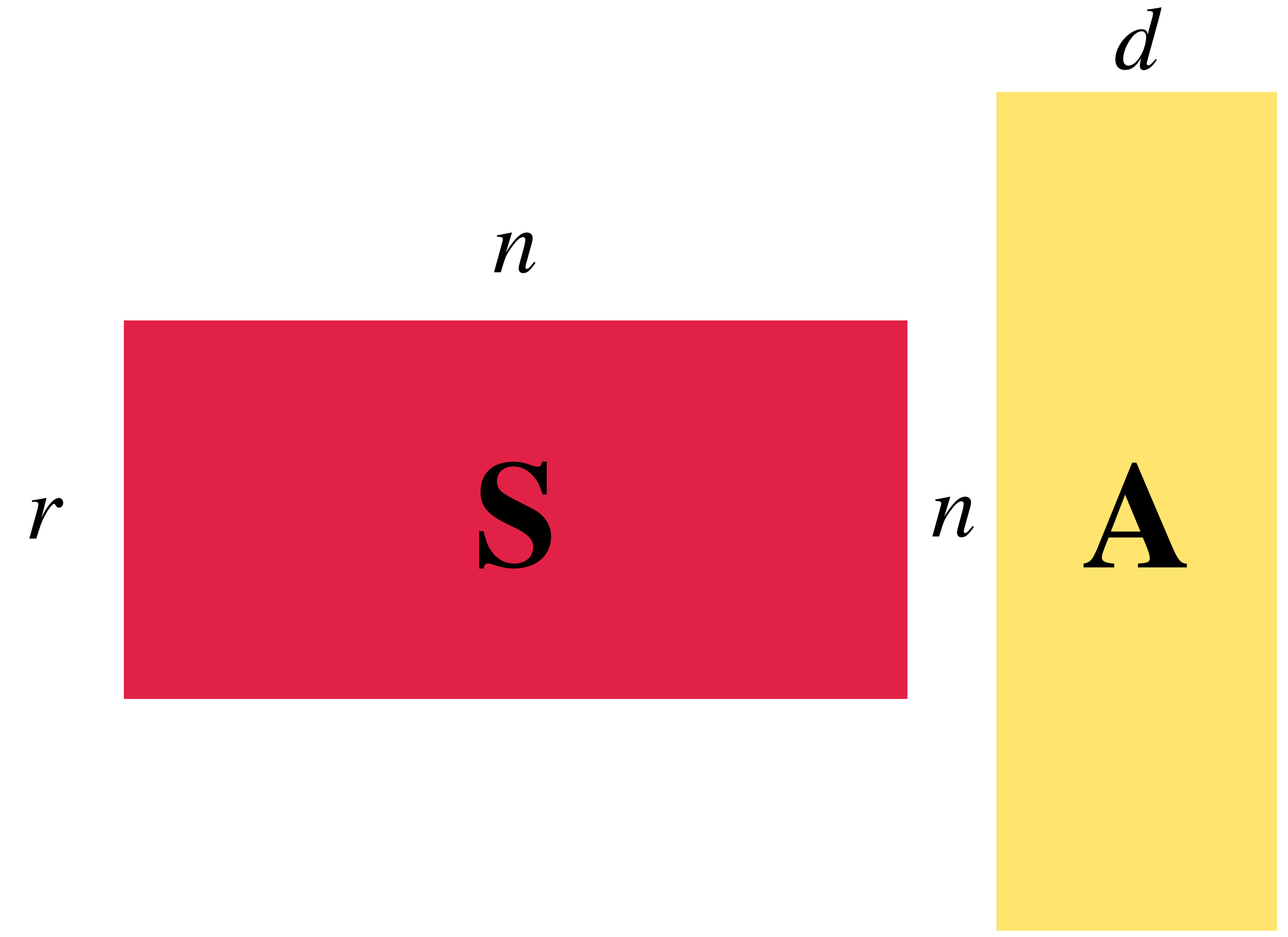
Theorem (Auerbach 1930). For any $\mathbf{A}$, there is $\mathbf{U}$ with $\alpha = d$.

Conjecture. For any $\mathbf{A}$, there is $\mathbf{U}$ with $\alpha = d^{1/p}$.

No success in showing this conjecture! ☹️

Matrix Approximation

Oblivious ℓ_p Subspace Embeddings



Fact. Oblivious ℓ_p subspace embeddings reduce to constructing **well-conditioned bases** for subspaces

$\approx$ orthonormal bases for ℓ_p norms

- **Idea.** What if we relax well-conditioned **bases** to well-conditioned **spanning sets**?

Theorem (Woodruff—Y 2023). For any $\mathbf{A}$, there is $\mathbf{U} \in \mathbb{R}^{n \times s}$ for $s = O(d \log \log d)$ such that

- $\|\mathbf{U}\|_{p,p} \leq s^{1/p}$
- for every $\mathbf{z} \in \mathbb{R}^d$, there is $\mathbf{x} \in \mathbb{R}^s$ such that $\mathbf{A}\mathbf{z} = \mathbf{U}\mathbf{x}$ and $\|\mathbf{x}\|_2 \leq \|\mathbf{U}\mathbf{x}\|_p$

Theorem (Auerbach 1930). For any $\mathbf{A}$, there is $\mathbf{U}$ with $\alpha = d$.

Conjecture. For any $\mathbf{A}$, there is $\mathbf{U}$ with $\alpha = d^{1/p}$.

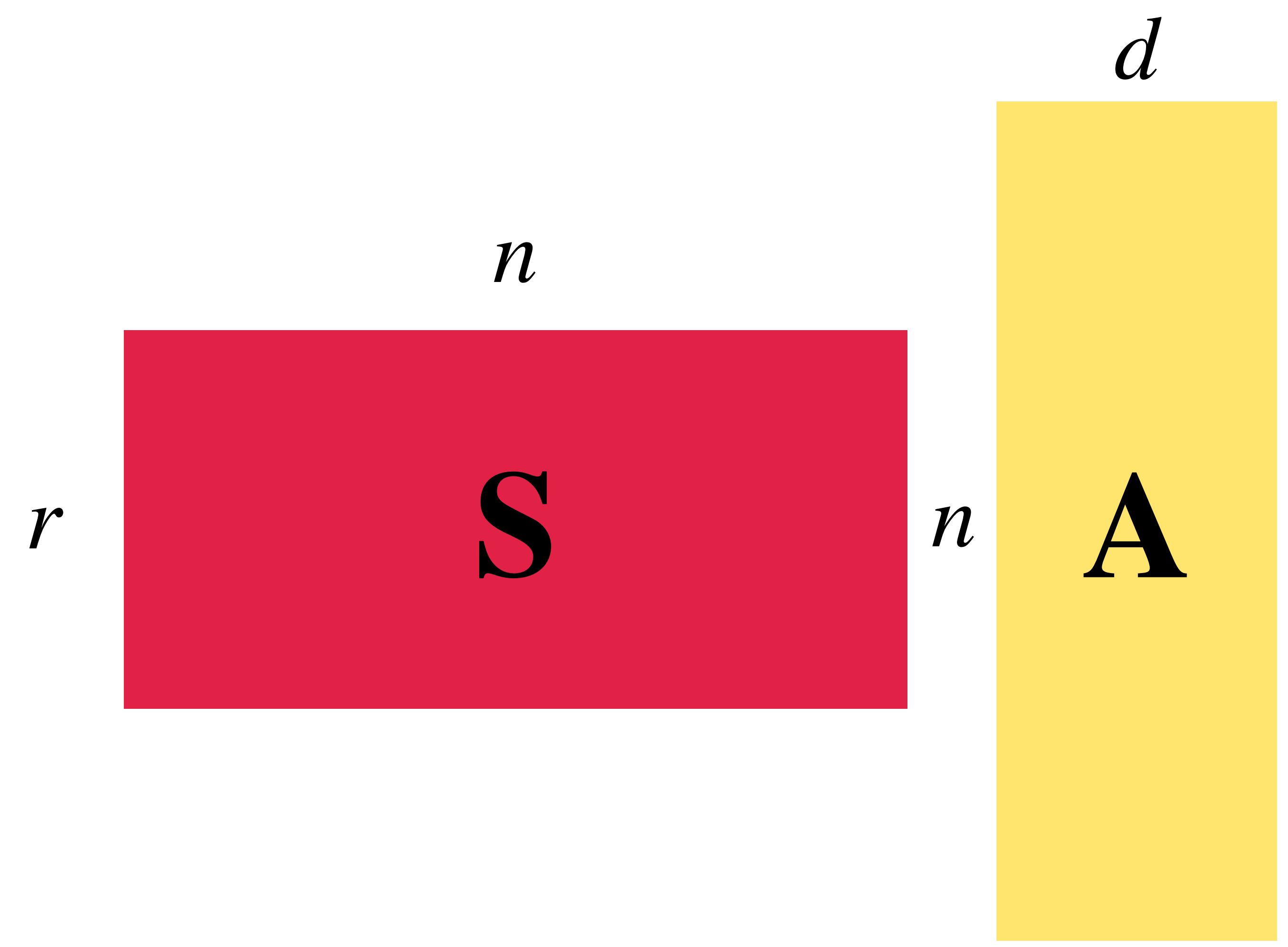
No success in showing this conjecture! ☹️

Matrix Approximation

Oblivious ℓ_p Subspace Embeddings

Fact. Oblivious ℓ_p subspace embeddings reduce to constructing **well-conditioned bases** for subspaces

$\approx$ orthonormal bases for ℓ_p norms



- **Idea.** What if we relax well-conditioned **bases** to well-conditioned **spanning sets**?

Theorem (Woodruff—Y 2023). For any $\mathbf{A}$, there is $\mathbf{U} \in \mathbb{R}^{n \times s}$ for $\boxed{s = O(d \log \log d)}$ such that

- $\|\mathbf{U}\|_{p,p} \leq s^{1/p}$
- for every $\mathbf{z} \in \mathbb{R}^d$, there is $\mathbf{x} \in \mathbb{R}^s$ such that $\mathbf{A}\mathbf{z} = \mathbf{U}\mathbf{x}$ and $\|\mathbf{x}\|_2 \leq \|\mathbf{U}\mathbf{x}\|_p$

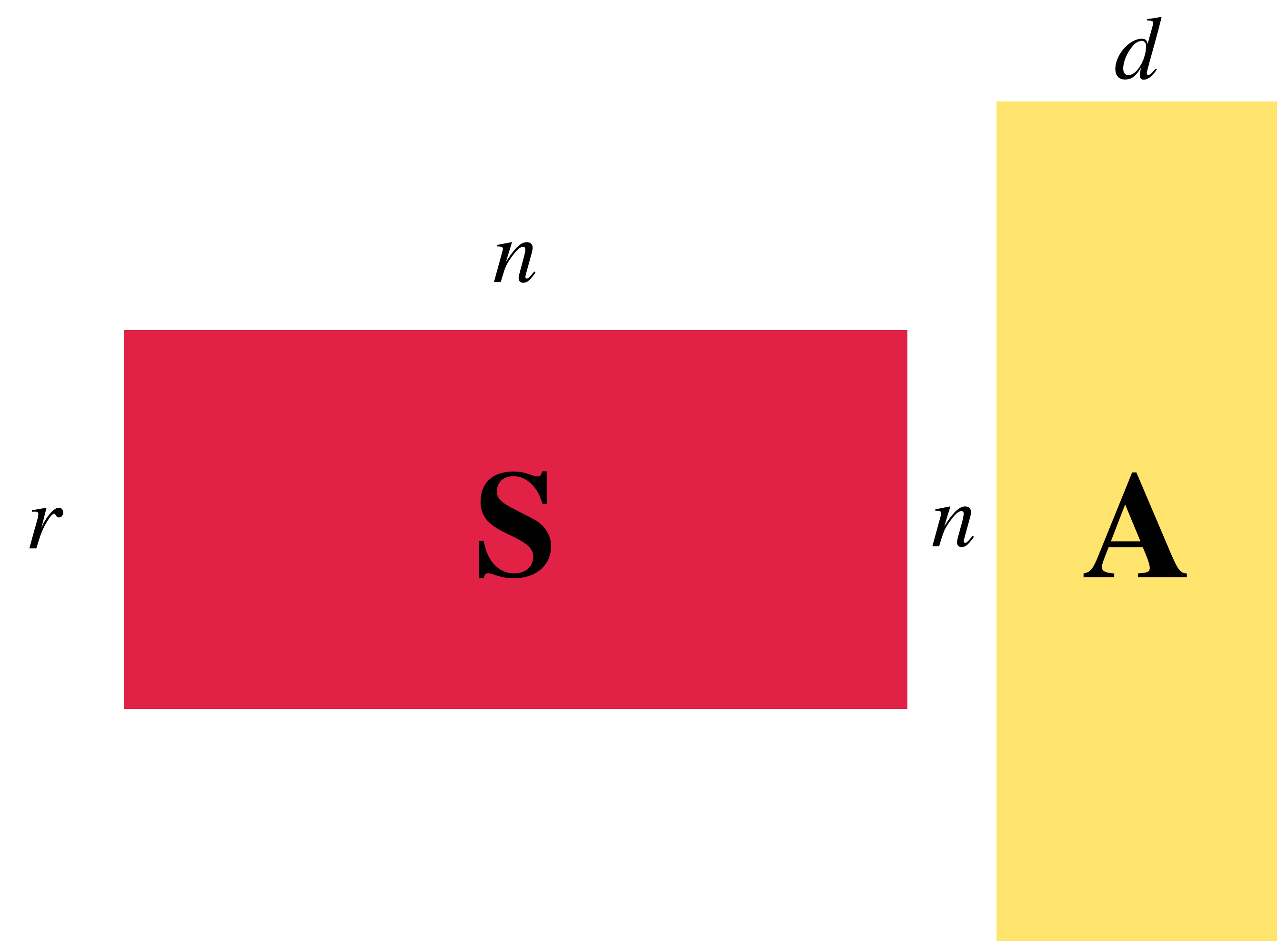
Theorem (Auerbach 1930). For any $\mathbf{A}$, there is $\mathbf{U}$ with $\alpha = d$.

Conjecture. For any $\mathbf{A}$, there is $\mathbf{U}$ with $\alpha = d^{1/p}$.

No success in showing this conjecture! ☹️

Matrix Approximation

Oblivious ℓ_p Subspace Embeddings



Fact. Oblivious ℓ_p subspace embeddings reduce to constructing **well-conditioned bases** for subspaces

≈ orthonormal bases for ℓ_p norms

- **Idea.** What if we relax well-conditioned **bases** to well-conditioned **spanning sets**?

Theorem (Woodruff—Y 2023). For any $\mathbf{A}$, there is $\mathbf{U} \in \mathbb{R}^{n \times s}$ for $\boxed{s = O(d \log \log d)}$ such that

- $\boxed{\|\mathbf{U}\|_{p,p} \leq s^{1/p}}$
- for every $\mathbf{z} \in \mathbb{R}^d$, there is $\mathbf{x} \in \mathbb{R}^s$ such that $\mathbf{A}\mathbf{z} = \mathbf{U}\mathbf{x}$ and $\|\mathbf{x}\|_2 \leq \|\mathbf{U}\mathbf{x}\|_p$

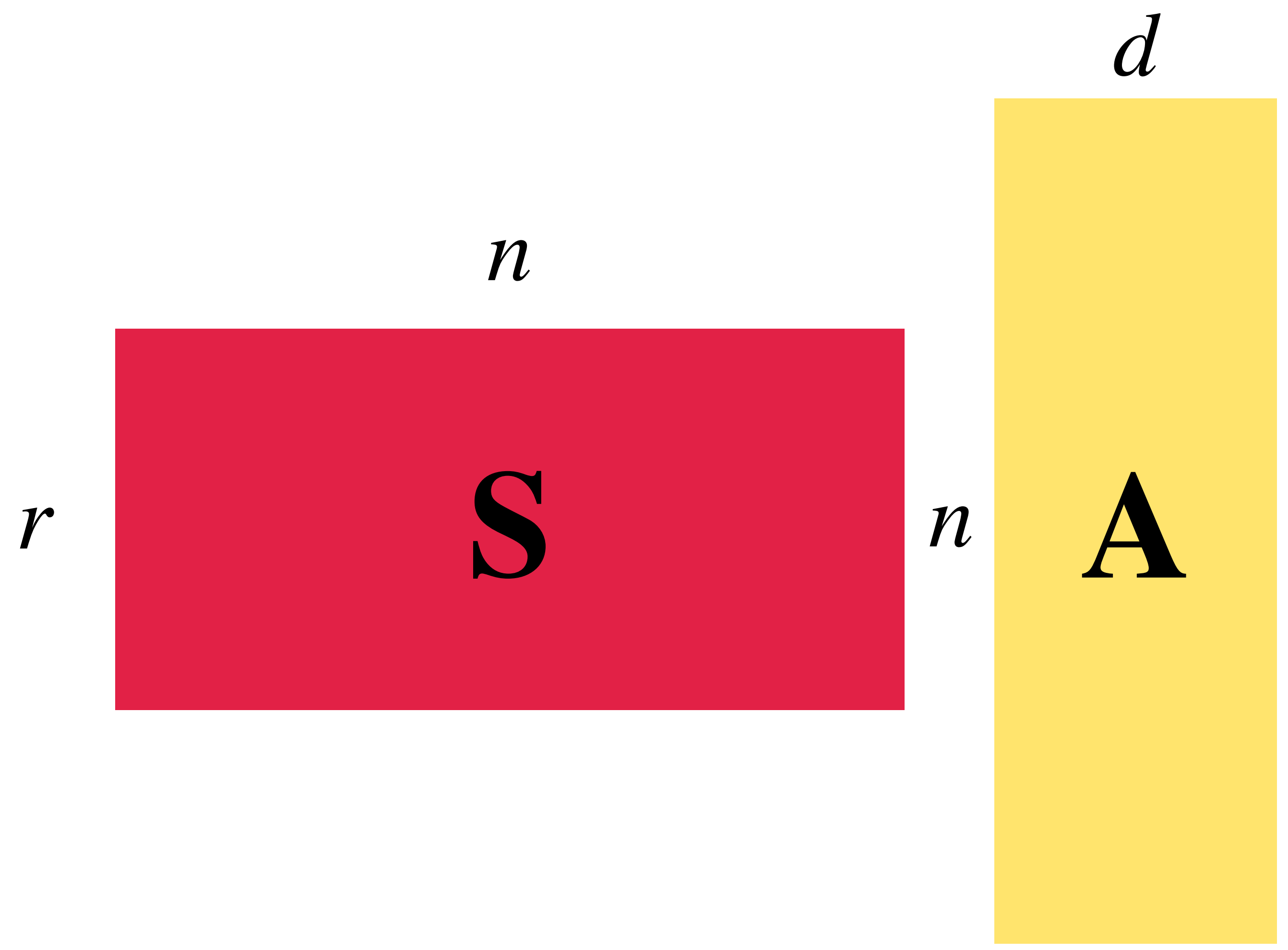
Theorem (Auerbach 1930). For any $\mathbf{A}$, there is $\mathbf{U}$ with $\alpha = d$.

Conjecture. For any $\mathbf{A}$, there is $\mathbf{U}$ with $\alpha = d^{1/p}$.

No success in showing this conjecture! ☹️

Matrix Approximation

Oblivious ℓ_p Subspace Embeddings



Fact. Oblivious ℓ_p subspace embeddings reduce to constructing **well-conditioned bases** for subspaces

$\approx$ orthonormal bases for ℓ_p norms

- **Idea.** What if we relax well-conditioned **bases** to well-conditioned **spanning sets**?

Theorem (Woodruff—Y 2023). For any $\mathbf{A}$, there is $\mathbf{U} \in \mathbb{R}^{n \times s}$ for $\boxed{s = O(d \log \log d)}$ such that

- $\boxed{\|\mathbf{U}\|_{p,p} \leq s^{1/p}}$
- for every $\mathbf{z} \in \mathbb{R}^d$, there is $\mathbf{x} \in \mathbb{R}^s$ such that $\mathbf{Az} = \mathbf{Ux}$ and $\boxed{\|\mathbf{x}\|_2 \leq \|\mathbf{Ux}\|_p}$

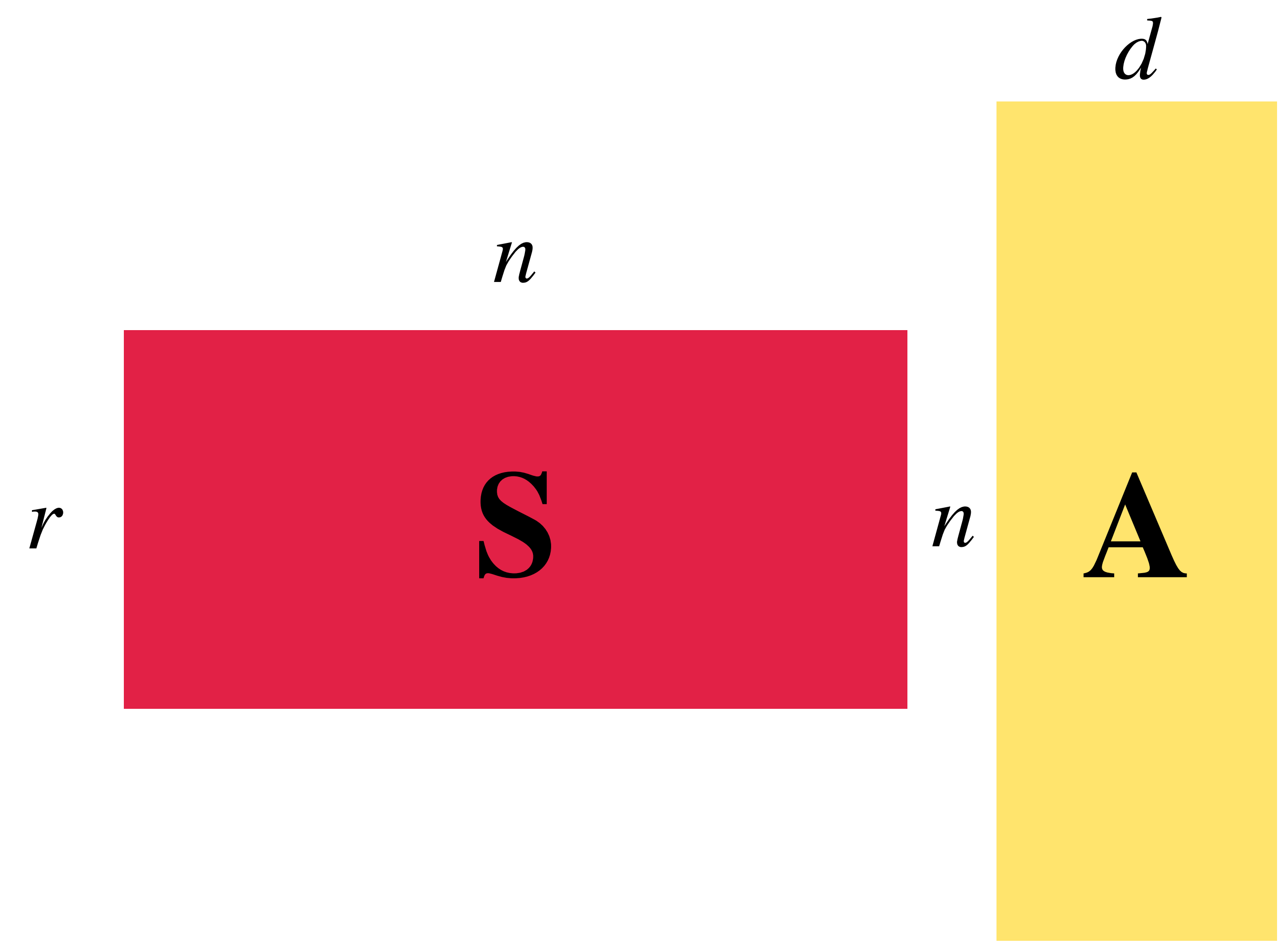
Theorem (Auerbach 1930). For any $\mathbf{A}$, there is $\mathbf{U}$ with $\alpha = d$.

Conjecture. For any $\mathbf{A}$, there is $\mathbf{U}$ with $\alpha = d^{1/p}$.

No success in showing this conjecture! ☹️

Matrix Approximation

Oblivious ℓ_p Subspace Embeddings



Fact. Oblivious ℓ_p subspace embeddings reduce to constructing **well-conditioned bases** for subspaces

≈ orthonormal bases for ℓ_p norms

- **Idea.** What if we relax well-conditioned **bases** to well-conditioned **spanning sets**?

Theorem (Auerbach 1930). For any $\mathbf{A}$, there is $\mathbf{U}$ with $\alpha = d$.

Conjecture. For any $\mathbf{A}$, there is $\mathbf{U}$ with $\alpha = d^{1/p}$.

No success in showing this conjecture! ☹️

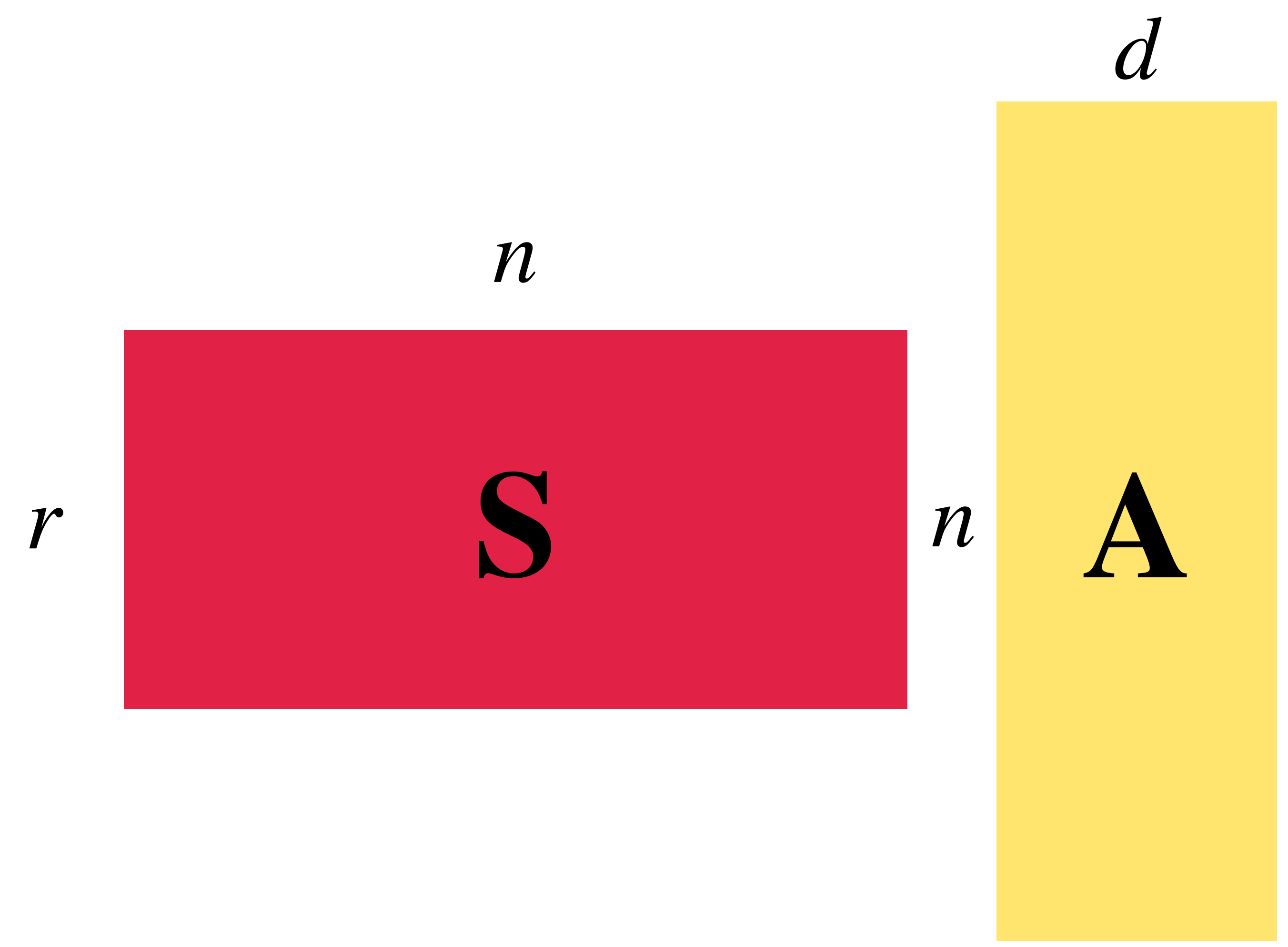
Theorem (Woodruff—Y 2023). For any $\mathbf{A}$, there is $\mathbf{U} \in \mathbb{R}^{n \times s}$ for $\boxed{s = O(d \log \log d)}$ such that

- $\boxed{\|\mathbf{U}\|_{p,p} \leq s^{1/p}}$
- for every $\mathbf{z} \in \mathbb{R}^d$, there is $\mathbf{x} \in \mathbb{R}^s$ such that $\mathbf{Az} = \mathbf{Ux}$ and $\boxed{\|\mathbf{x}\|_2 \leq \|\mathbf{Ux}\|_p}$

≈ as good as showing $\alpha = s^{1/p}$

Matrix Approximation

Oblivious ℓ_p Subspace Embeddings



Fact. Oblivious ℓ_p subspace embeddings reduce to constructing **well-conditioned bases** for subspaces

$\approx$ orthonormal bases for ℓ_p norms

- **Idea.** What if we relax well-conditioned **bases** to well-conditioned **spanning sets**?

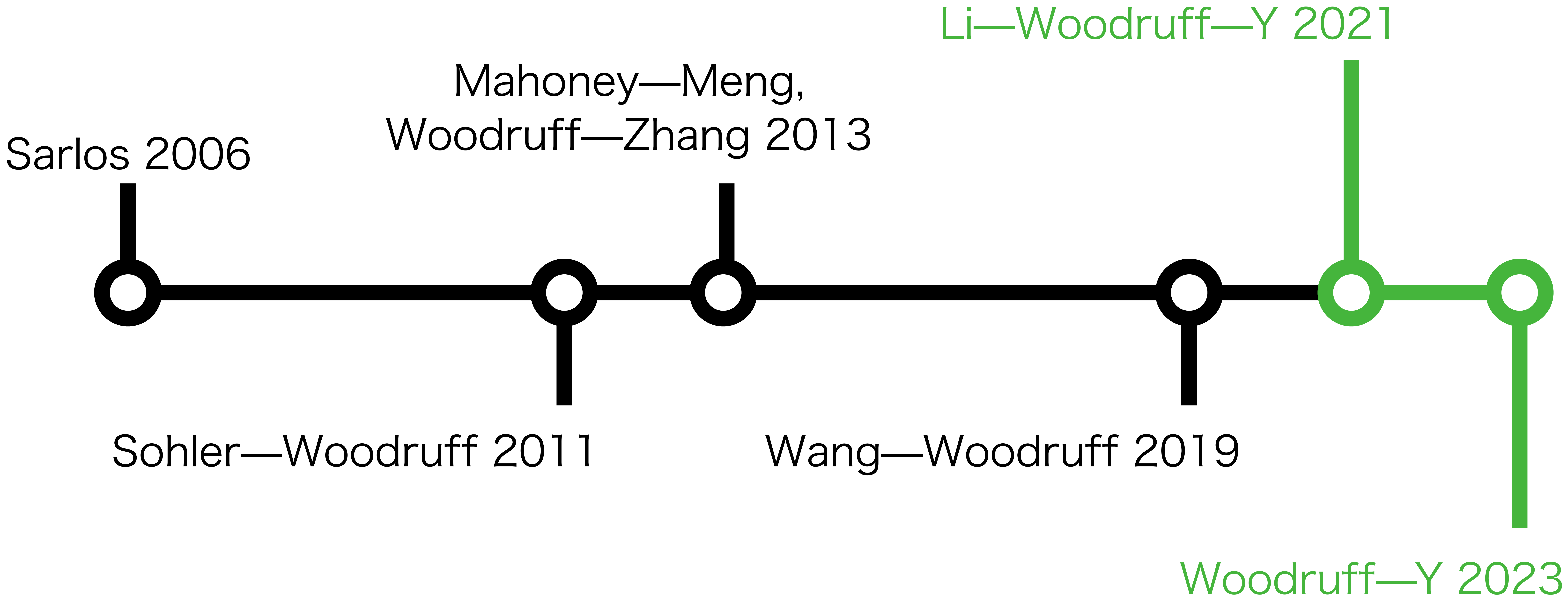
Theorem (Woodruff—Y 2023). For any $\mathbf{A}$, there is $\mathbf{U} \in \mathbb{R}^{n \times s}$ for $\boxed{s = O(d \log \log d)}$ such that

- $\boxed{\|\mathbf{U}\|_{p,p} \leq s^{1/p}}$
- for every $\mathbf{z} \in \mathbb{R}^d$, there is $\mathbf{x} \in \mathbb{R}^s$ such that $\mathbf{Az} = \mathbf{Ux}$ and $\boxed{\|\mathbf{x}\|_2 \leq \|\mathbf{Ux}\|_p}$

Theorem (Woodruff—Y 2023). There are oblivious ℓ_p subspace embeddings with $r = \tilde{O}(d)$ and $\kappa = \tilde{O}(d^{1/p})$, which is **nearly optimal**.

Matrix Approximation

Oblivious ℓ_p Subspace Embeddings



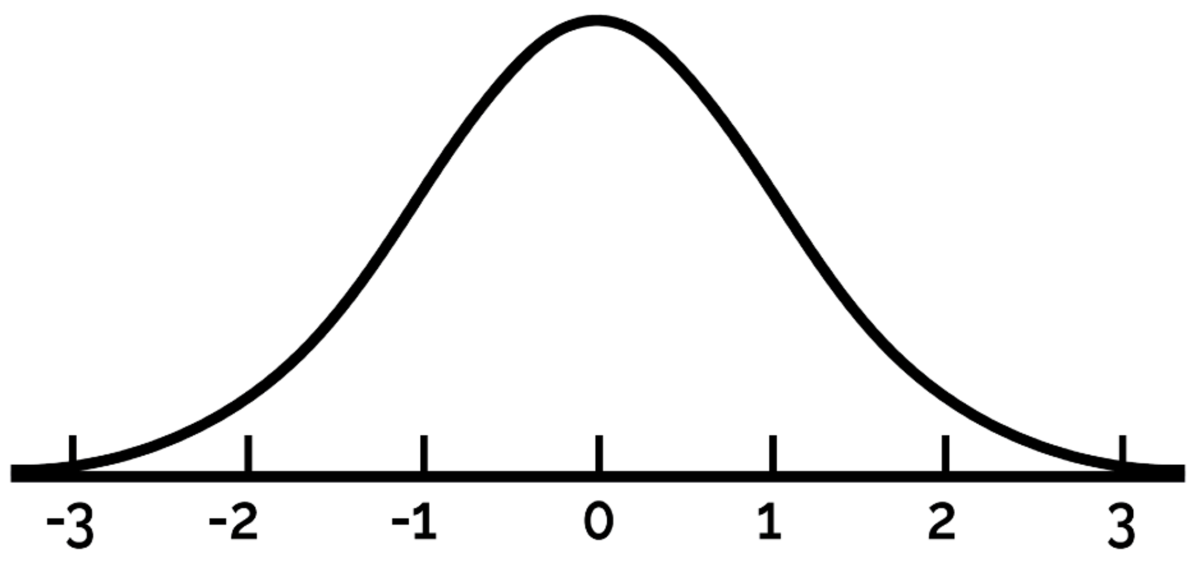
Subspace Embeddings and Linear Regression

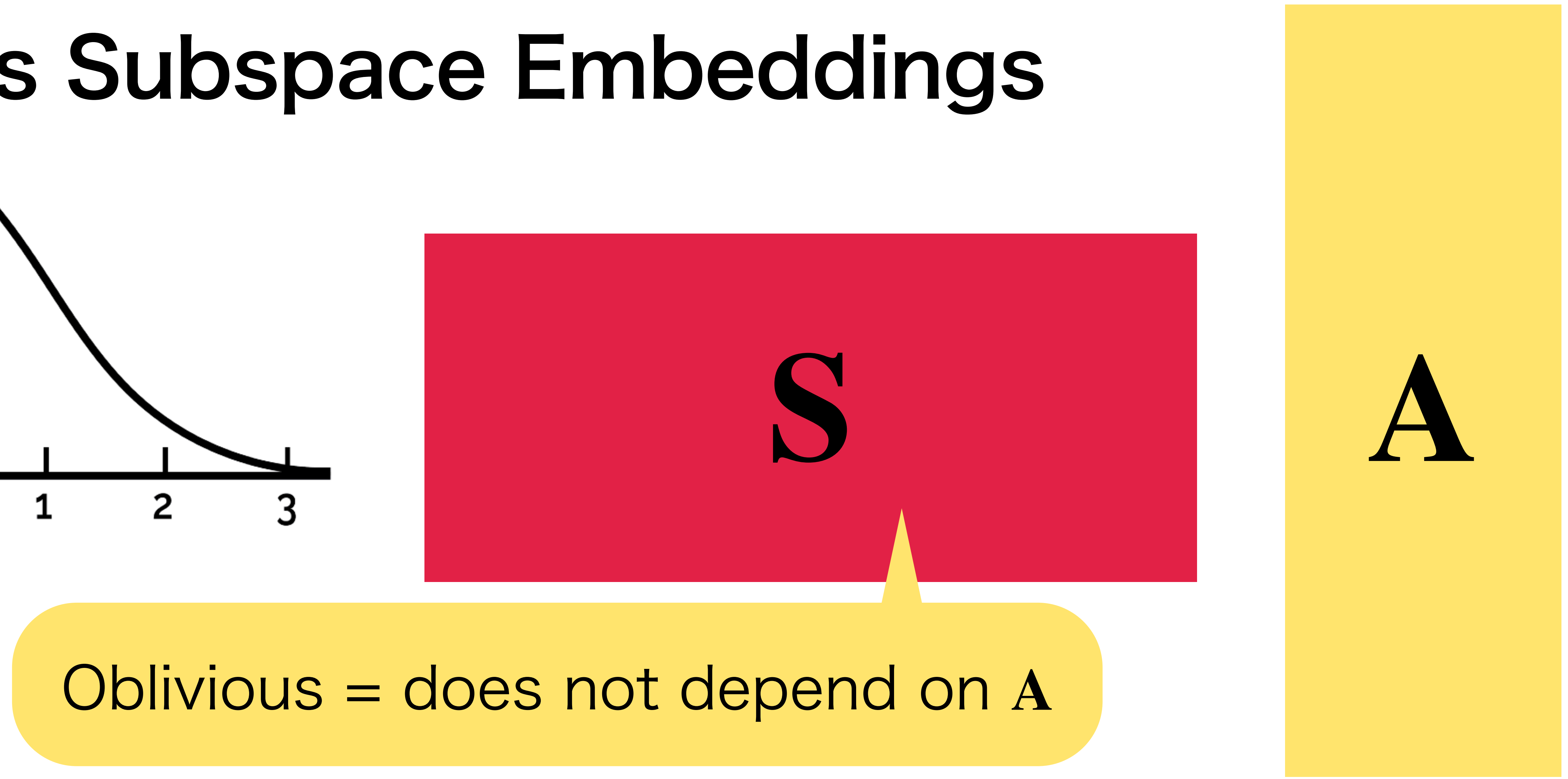
- Oblivious ℓ_p subspace embeddings: high distortion and low distortion
- $\boxed{\text{Non-oblivious subspace embeddings: } \ell_p \text{ Lewis weight sampling, general losses}}$
- Applications: active learning, streaming computational geometry, low rank approximation

Matrix Approximation

Non-oblivious Subspace Embeddings

Oblivious

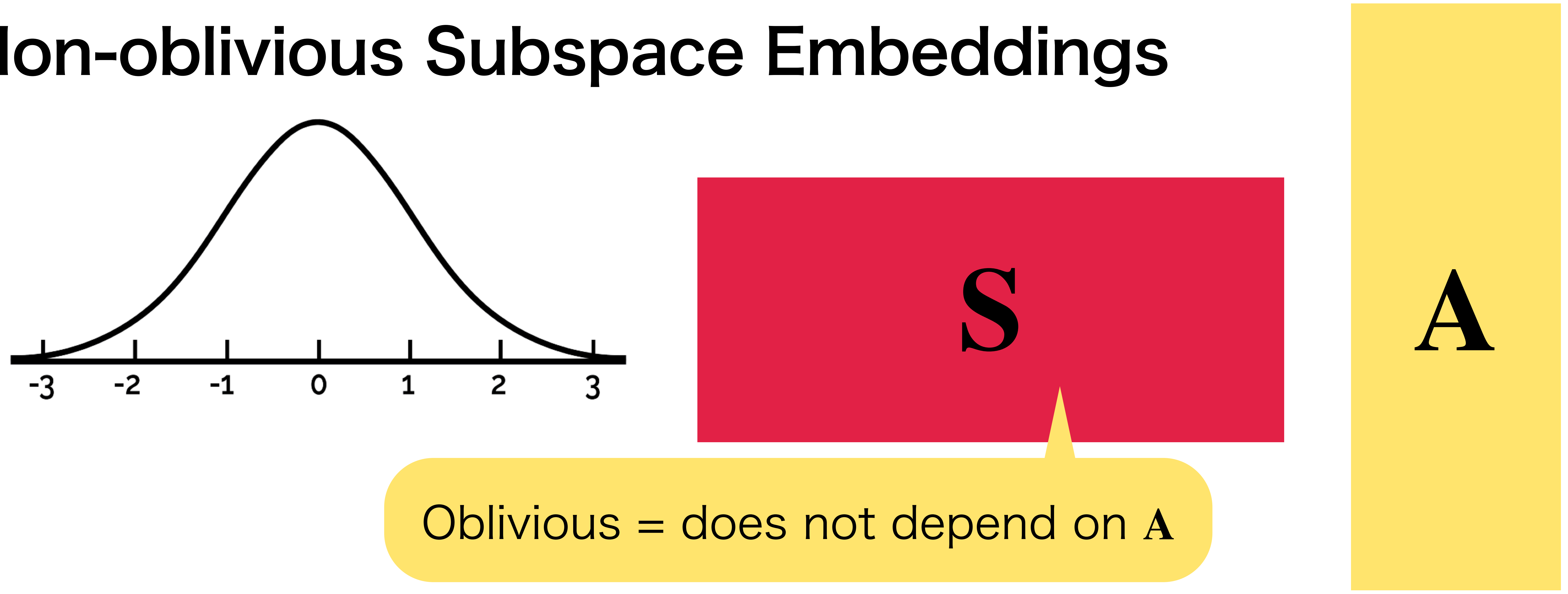




Matrix Approximation

Non-oblivious Subspace Embeddings

Oblivious

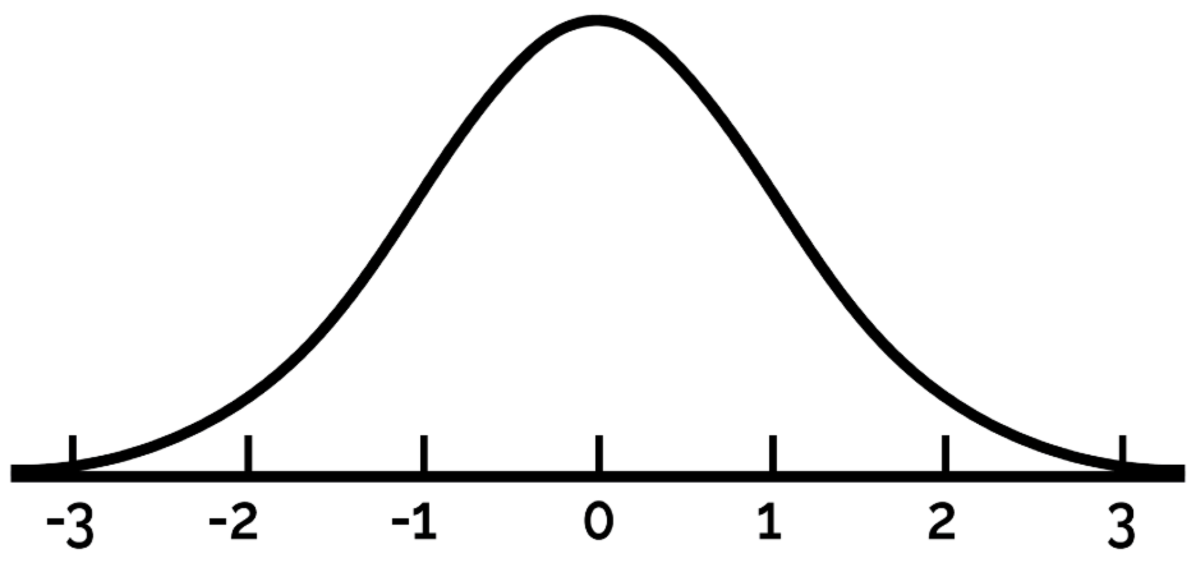


Non-oblivious/ Sampling

Matrix Approximation

Non-oblivious Subspace Embeddings

Oblivious



$\mathbf{S}$

$\mathbf{A}$

Oblivious = does not depend on $\mathbf{A}$

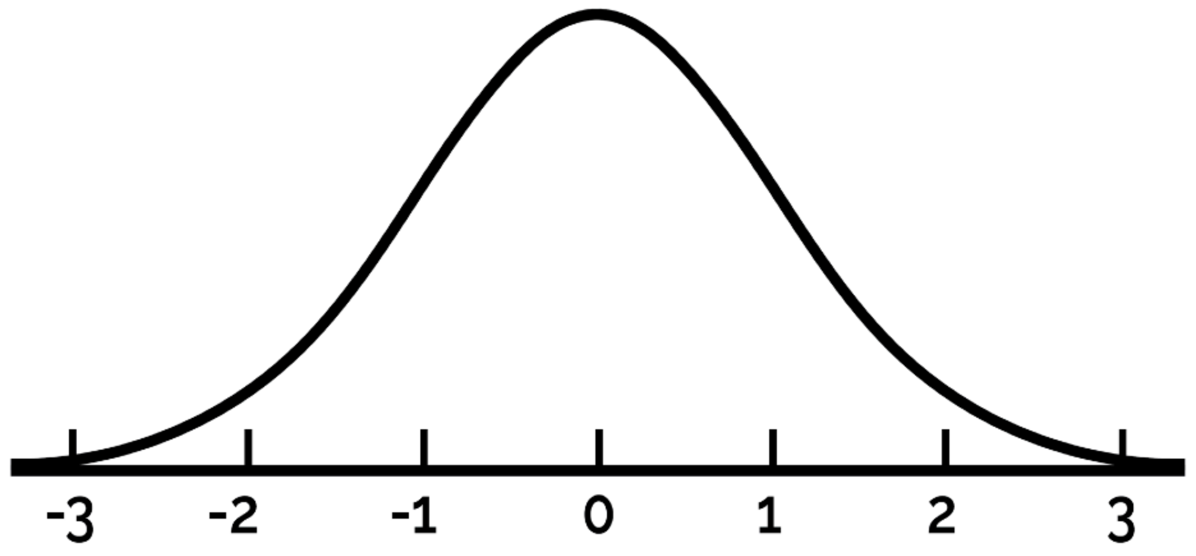
Non-oblivious/ Sampling

$\mathbf{A}$

Matrix Approximation

Non-oblivious Subspace Embeddings

Oblivious



$\mathbf{S}$

$\mathbf{A}$

Oblivious = does not depend on $\mathbf{A}$

Non-oblivious/ Sampling

Step 1. Compute "importance scores" for the rows of $\mathbf{A}$

q_1

q_2

$\vdots$

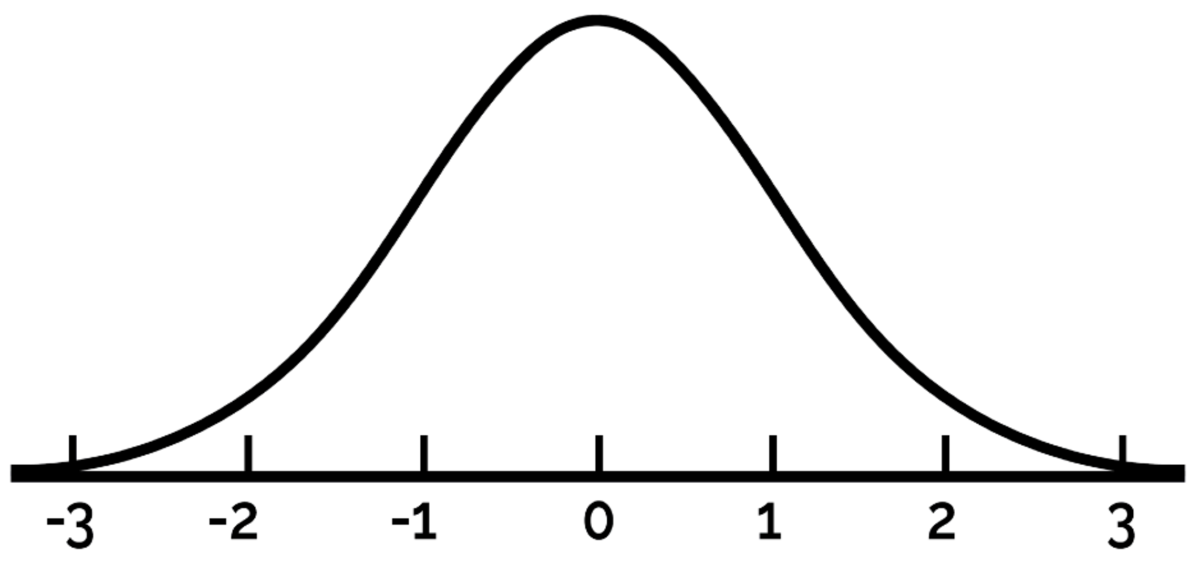
q_n

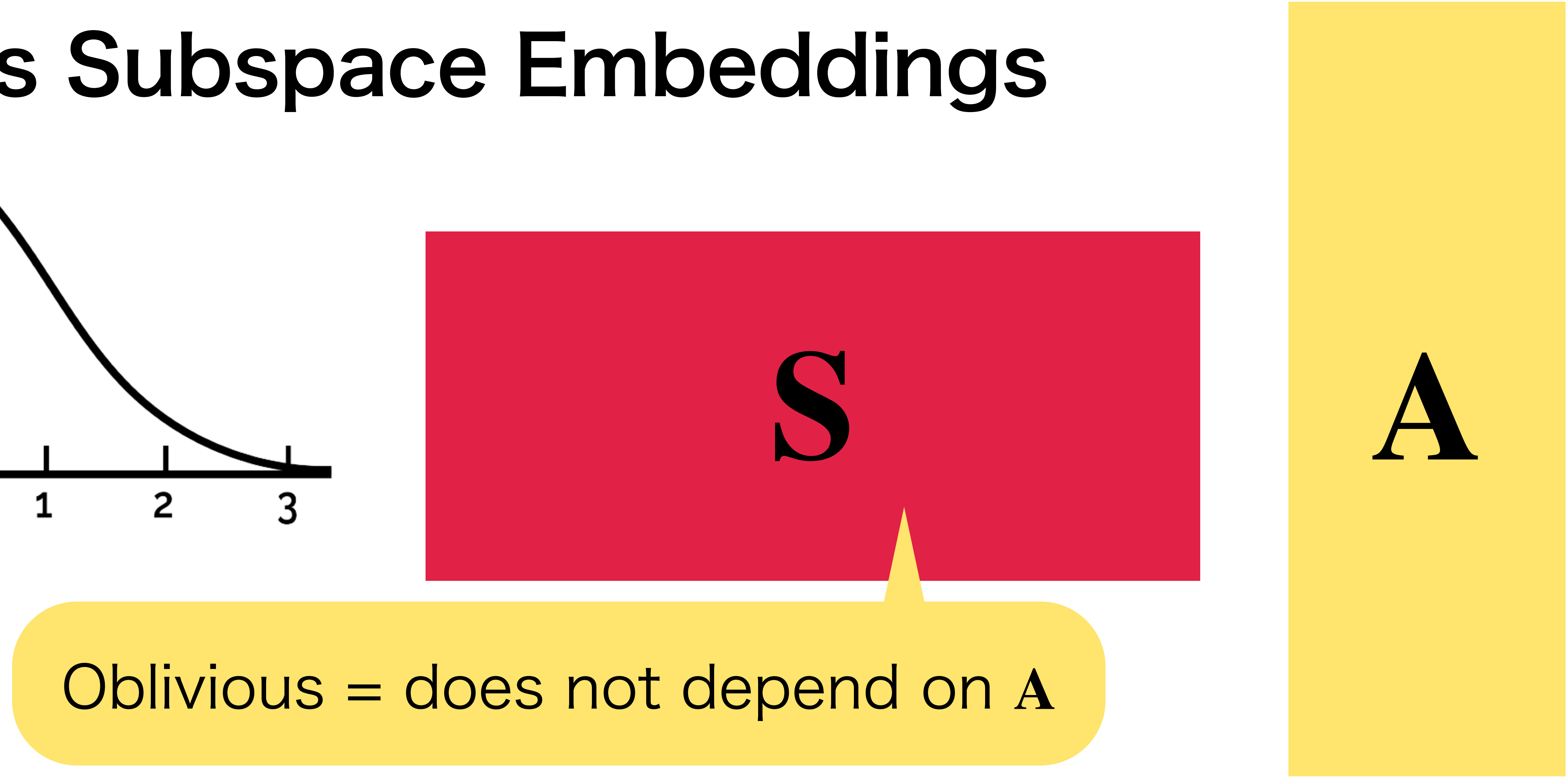
$\mathbf{A}$

Matrix Approximation

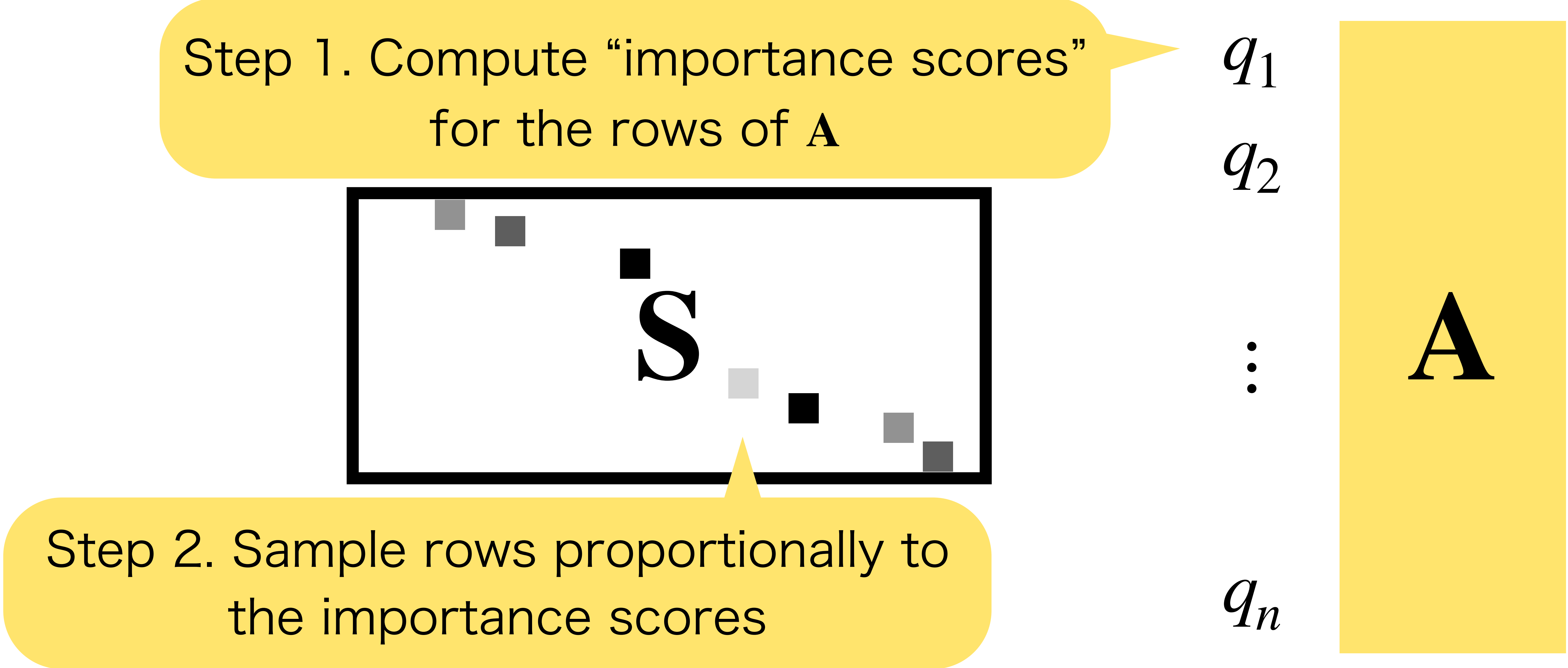
Non-oblivious Subspace Embeddings

Oblivious



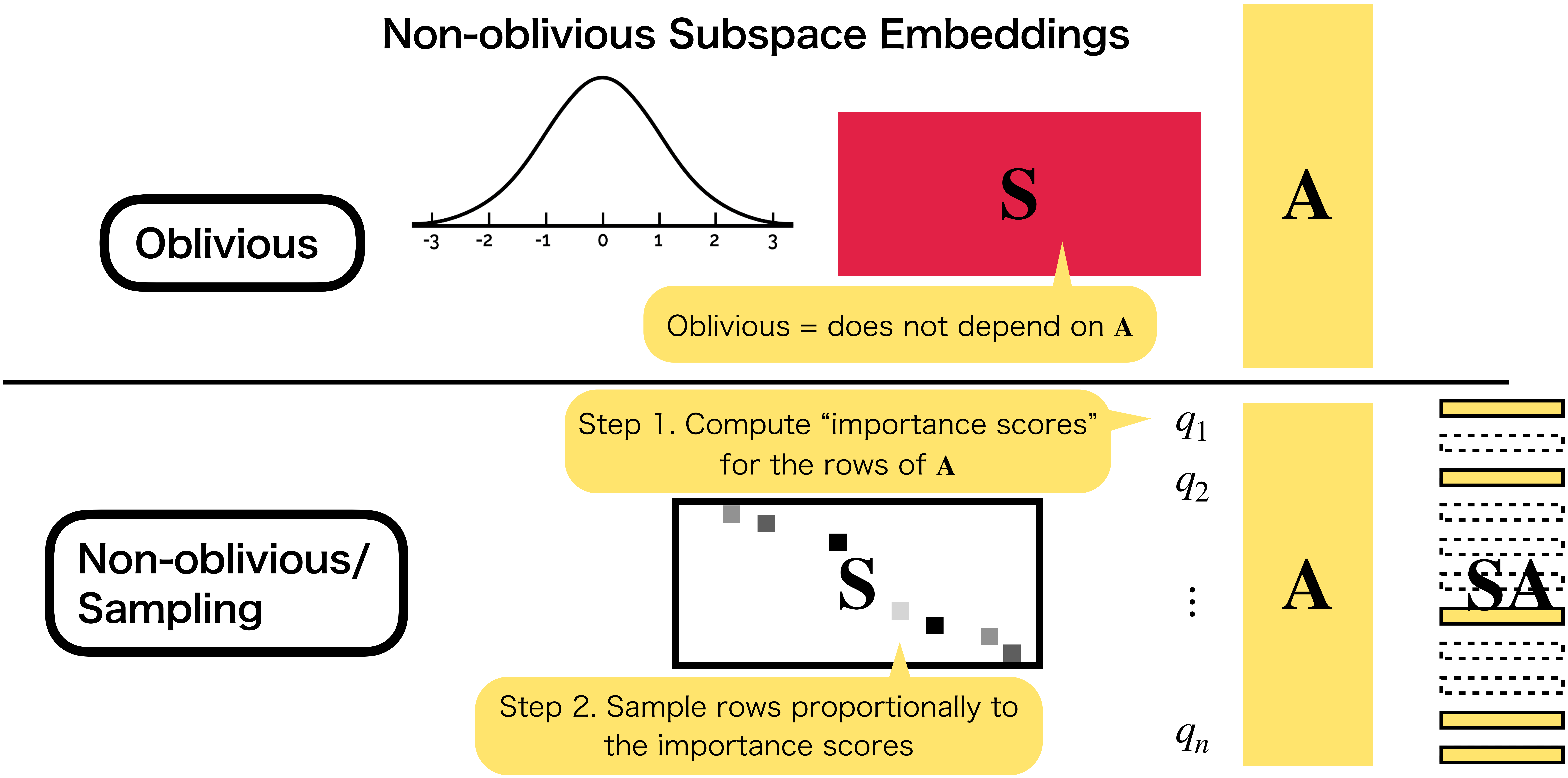


Non-oblivious/ Sampling



Matrix Approximation

Non-oblivious Subspace Embeddings

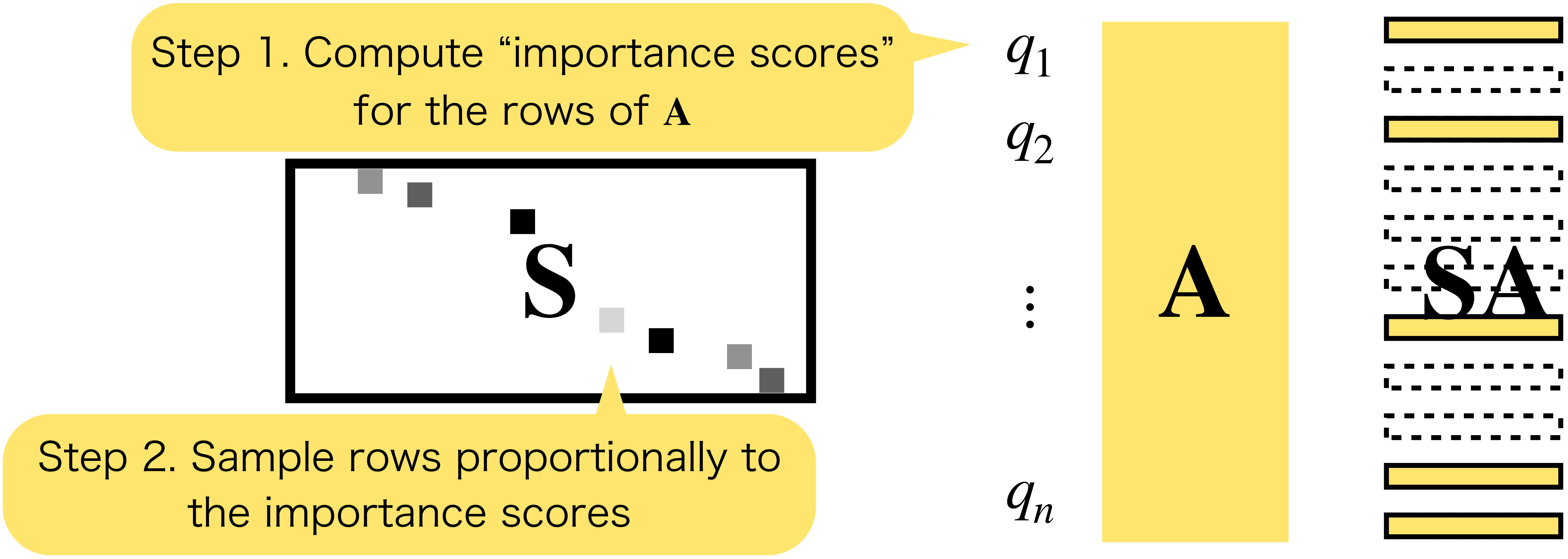


Matrix Approximation

Non-oblivious Subspace Embeddings

Theorem (Leverage score sampling). For any $\mathbf{A} \in \mathbb{R}^{n \times d}$, there are probabilities $q_1, q_2, \ldots q_n$ that sample $r = \tilde{O}(\varepsilon^{-2} d)$ rows of $\mathbf{A}$ that forms an ℓ_2 subspace embedding with distortion $\kappa = (1 + \varepsilon)$, with probability 99%.

Non-oblivious/ Sampling

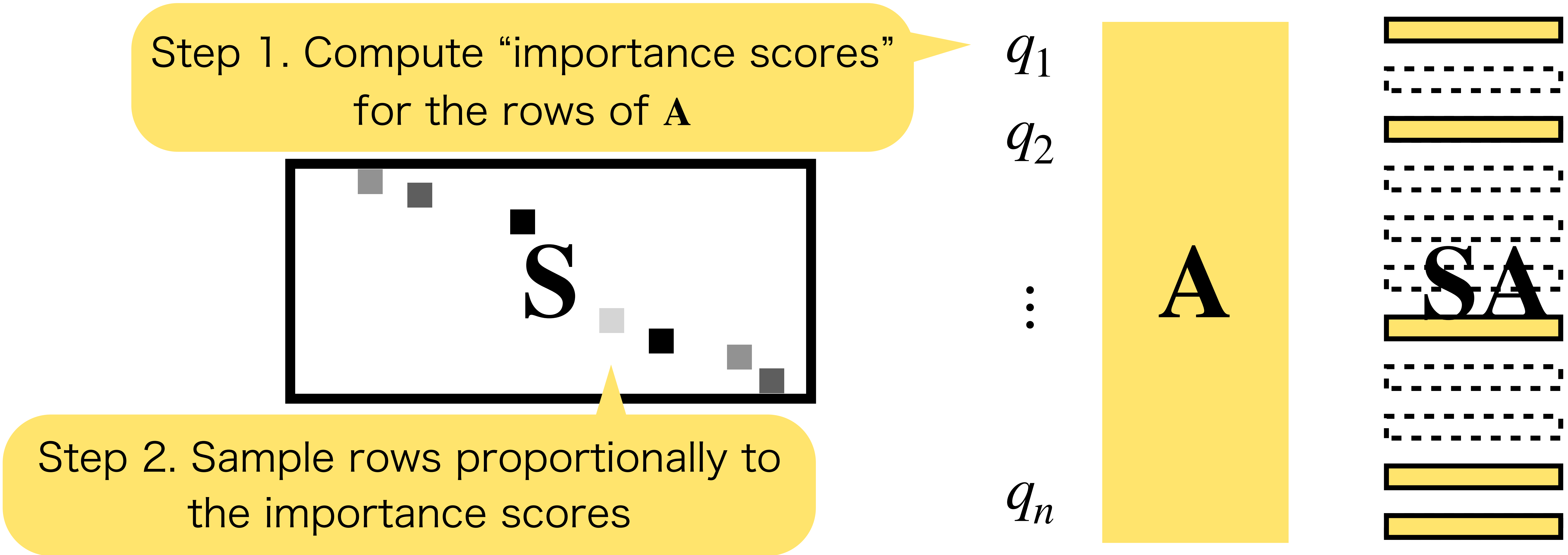


Matrix Approximation

Non-oblivious Subspace Embeddings

Theorem (Leverage score sampling). For any $\mathbf{A} \in \mathbb{R}^{n \times d}$, there are probabilities $q_1, q_2, \ldots q_n$ that sample $\boxed{r = \tilde{O}(\varepsilon^{-2} d)}$ rows of $\mathbf{A}$ that forms an ℓ_2 subspace embedding with distortion $\boxed{\kappa = (1 + \varepsilon)}$, with probability 99%.

Non-oblivious/ Sampling



Matrix Approximation

Non-oblivious Subspace Embeddings

Theorem (Leverage score sampling). For any $\mathbf{A} \in \mathbb{R}^{n \times d}$, there are probabilities $q_1, q_2, \ldots q_n$ that sample $\boxed{r = \tilde{O}(\varepsilon^{-2} d)}$ rows of $\mathbf{A}$ that forms an ℓ_2 subspace embedding with distortion $\boxed{\kappa = (1 + \varepsilon)}$, with probability 99%.

Matrix Approximation

Non-oblivious Subspace Embeddings

Theorem (Leverage score sampling). For any $\mathbf{A} \in \mathbb{R}^{n \times d}$, there are probabilities $q_1, q_2, \ldots q_n$ that sample $\boxed{r = \tilde{O}(\varepsilon^{-2} d)}$ rows of $\mathbf{A}$ that forms an ℓ_2 subspace embedding with distortion $\boxed{\kappa = (1 + \varepsilon)}$, with probability 99%.

- Same row count (r) vs distortion (κ) trade-off as the oblivious case…

Matrix Approximation

Non-oblivious Subspace Embeddings

Theorem (Leverage score sampling). For any $\mathbf{A} \in \mathbb{R}^{n \times d}$, there are probabilities $q_1, q_2, \ldots q_n$ that sample $\boxed{r = \tilde{O}(\varepsilon^{-2} d)}$ rows of $\mathbf{A}$ that forms an ℓ_2 subspace embedding with distortion $\boxed{\kappa = (1 + \varepsilon)}$, with probability 99%.

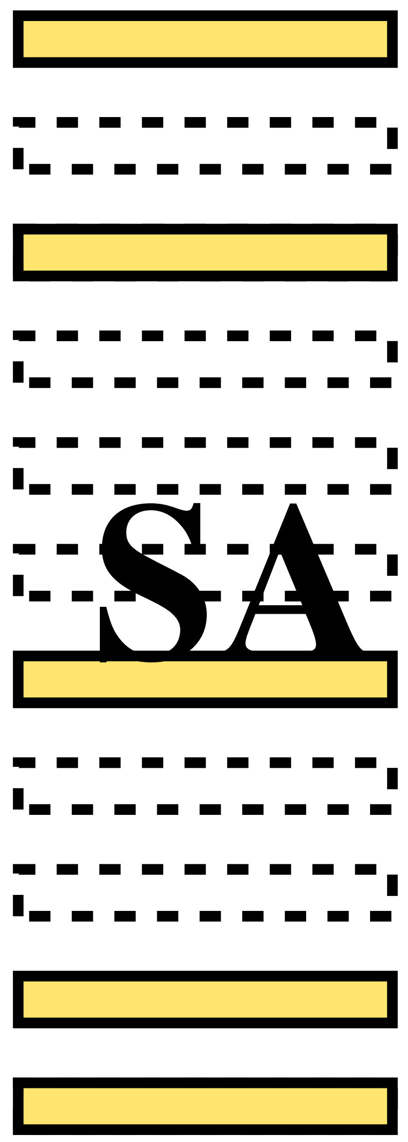
- Same row count (r) vs distortion (κ) trade-off as the oblivious case…
- Generalizes to much better trade-offs for ℓ_p norms with $p \neq 2$!

Matrix Approximation

Non-oblivious Subspace Embeddings

Theorem (Leverage score sampling). For any $\mathbf{A} \in \mathbb{R}^{n \times d}$, there are probabilities $q_1, q_2, \ldots q_n$ that sample $\boxed{r = \tilde{O}(\varepsilon^{-2} d)}$ rows of $\mathbf{A}$ that forms an ℓ_2 subspace embedding with distortion $\boxed{\kappa = (1 + \varepsilon)}$, with probability 99%.

- Same row count (r) vs distortion (κ) trade-off as the oblivious case…
- Generalizes to much better trade-offs for ℓ_p norms with $p \neq 2$!
  - Oblivious: either $\kappa = \mathrm{poly}(d)$ or $r \gg \mathrm{poly}(d)$, and only for $p \leq 2$…

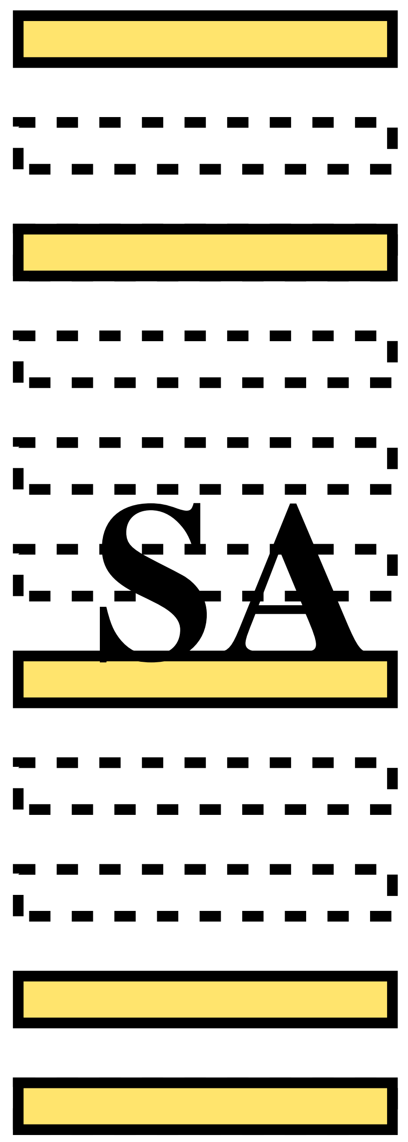


Matrix Approximation

Non-oblivious Subspace Embeddings

Theorem (Leverage score sampling). For any $\mathbf{A} \in \mathbb{R}^{n \times d}$, there are probabilities $q_1, q_2, \ldots q_n$ that sample $\boxed{r = \tilde{O}(\varepsilon^{-2} d)}$ rows of $\mathbf{A}$ that forms an ℓ_2 subspace embedding with distortion $\boxed{\kappa = (1 + \varepsilon)}$, with probability 99%.

- Same row count (r) vs distortion (κ) trade-off as the oblivious case...
- Generalizes to much better trade-offs for ℓ_p norms with $p \neq 2$!
  - Oblivious: either $\kappa = \mathrm{poly}(d)$ or $r \gg \mathrm{poly}(d)$, and only for $p \leq 2$...
  - Non-oblivious: $\kappa = (1 + \varepsilon)$ and $r = \mathrm{poly}(d)$ for any fixed p!



Matrix Approximation

Non-oblivious Subspace Embeddings

Definition (Leverage scores). For $\mathbf{A} \in \mathbb{R}^{n \times d}$ and $i \in [n]$, the ith leverage score is

$$\tau_i(\mathbf{A}) = \sup_{\mathbf{A}\mathbf{x} \neq 0} \frac{\langle \mathbf{a}_i, \mathbf{x} \rangle^2}{\|\mathbf{A}\mathbf{x}\|_2^2}$$

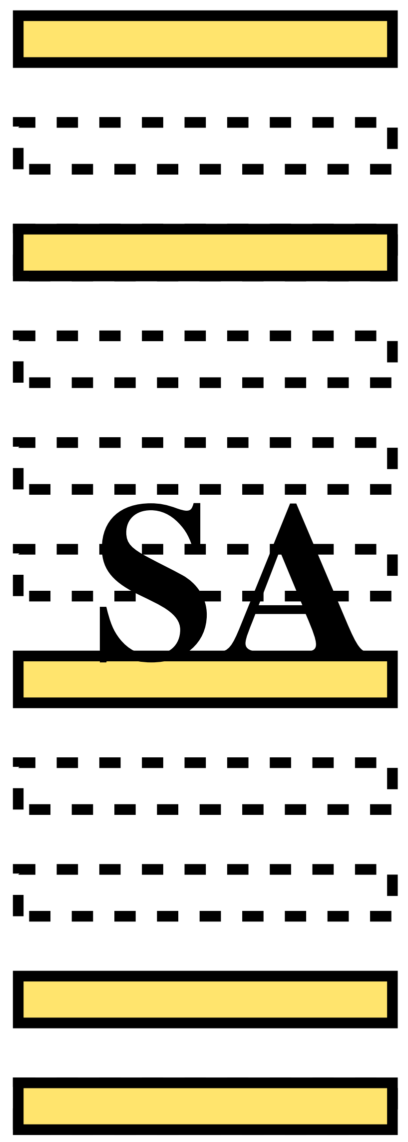
Matrix Approximation

Non-oblivious Subspace Embeddings

Definition (Leverage scores). For $\mathbf{A} \in \mathbb{R}^{n \times d}$ and $i \in [n]$, the ith leverage score is

$$\tau_i(\mathbf{A}) = \sup_{\mathbf{Ax} \neq 0} \frac{\langle \mathbf{a}_i, \mathbf{x} \rangle^2}{\|\mathbf{Ax}\|_2^2}$$

ℓ_2 norm of $\mathbf{Ax}$



Matrix Approximation

Non-oblivious Subspace Embeddings

Definition (Leverage scores). For $\mathbf{A} \in \mathbb{R}^{n \times d}$ and $i \in [n]$, the ith leverage score is

$$\tau_i(\mathbf{A}) = \sup_{\mathbf{Ax} \neq 0} \frac{\langle \mathbf{a}_i, \mathbf{x} \rangle^2}{\|\mathbf{Ax}\|_2^2}$$

Contribution of the ith coordinate

ℓ_2 norm of $\mathbf{Ax}$

Matrix Approximation

Non-oblivious Subspace Embeddings

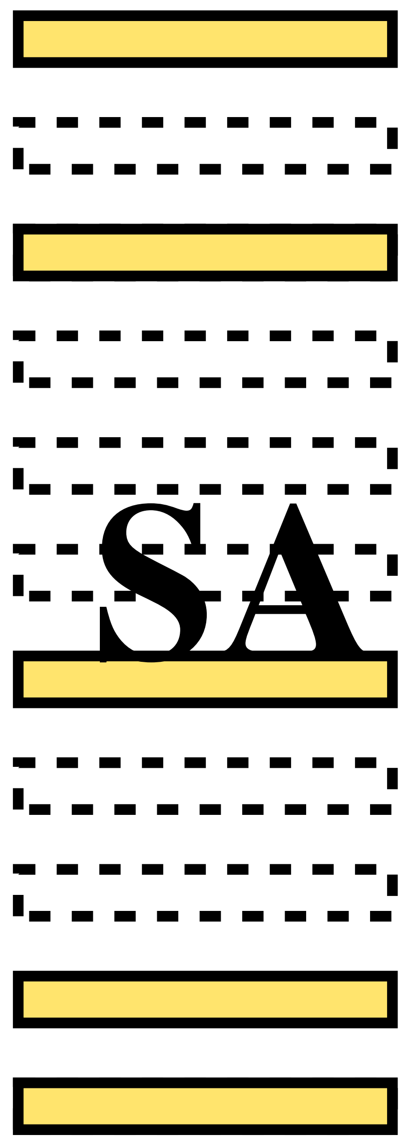
Definition (Leverage scores). For $\mathbf{A} \in \mathbb{R}^{n \times d}$ and $i \in [n]$, the ith leverage score is

$$\tau_i(\mathbf{A}) = \sup_{\mathbf{Ax} \neq 0} \frac{\langle \mathbf{a}_i, \mathbf{x} \rangle^2}{\|\mathbf{Ax}\|_2^2}$$

Contribution of the ith coordinate

ℓ_2 norm of $\mathbf{Ax}$

“Importance score”: largest fraction of ℓ_2 norm occupied by the ith coordinate



Matrix Approximation

Non-oblivious Subspace Embeddings

Definition (Leverage scores). For $\mathbf{A} \in \mathbb{R}^{n \times d}$ and $i \in [n]$, the ith leverage score is

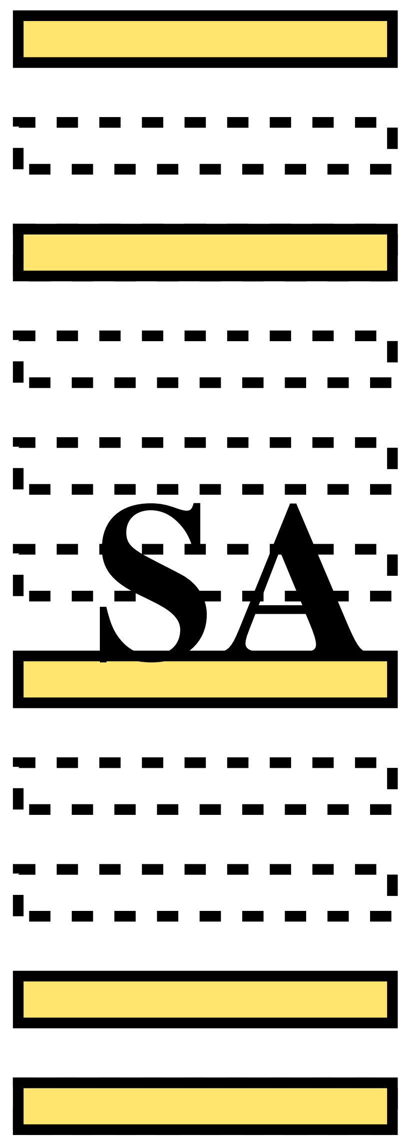
$$\tau_i(\mathbf{A}) = \sup_{\mathbf{Ax} \neq 0} \frac{\langle \mathbf{a}_i, \mathbf{x} \rangle^2}{\|\mathbf{Ax}\|_2^2}$$

Contribution of the ith coordinate

ℓ_2 norm of $\mathbf{Ax}$

"Importance score": largest fraction of ℓ_2 norm occupied by the ith coordinate

- Many generalizations:



Matrix Approximation

Non-oblivious Subspace Embeddings

Definition (Leverage scores). For $\mathbf{A} \in \mathbb{R}^{n \times d}$ and $i \in [n]$, the ith leverage score is

$$\tau_i(\mathbf{A}) = \sup_{\mathbf{Ax} \neq 0} \frac{\langle \mathbf{a}_i, \mathbf{x} \rangle^2}{\|\mathbf{Ax}\|_2^2}$$

Contribution of the ith coordinate

ℓ_2 norm of $\mathbf{Ax}$

"Importance score": largest fraction of ℓ_2 norm occupied by the ith coordinate

- Many generalizations:
  - ℓ_p sensitivity scores (Langberg—Schulman 2010)

Matrix Approximation

Non-oblivious Subspace Embeddings

Definition (Leverage scores). For $\mathbf{A} \in \mathbb{R}^{n \times d}$ and $i \in [n]$, the ith leverage score is

$$\tau_i(\mathbf{A}) = \sup_{\mathbf{A}\mathbf{x} \neq 0} \frac{\langle \mathbf{a}_i, \mathbf{x} \rangle^2}{\|\mathbf{A}\mathbf{x}\|_2^2}$$

Contribution of the ith coordinate

ℓ_2 norm of $\mathbf{Ax}$

“Importance score”: largest fraction of ℓ_2 norm occupied by the ith coordinate

- Many generalizations:
  - ℓ_p sensitivity scores (Langberg—Schulman 2010)
  - ℓ_p Lewis weights (Lewis 1978)

Matrix Approximation

Non-oblivious Subspace Embeddings

Definition (Leverage scores). For $\mathbf{A} \in \mathbb{R}^{n \times d}$ and $i \in [n]$, the ith leverage score is

$$\tau_i(\mathbf{A}) = \sup_{\mathbf{Ax} \neq 0} \frac{\langle \mathbf{a}_i, \mathbf{x} \rangle^2}{\|\mathbf{Ax}\|_2^2}$$

Contribution of the ith coordinate

ℓ_2 norm of $\mathbf{Ax}$

"Importance score": largest fraction of ℓ_2 norm occupied by the ith coordinate

- Many generalizations:
  - ℓ_p sensitivity scores (Langberg—Schulman 2010)
  - ℓ_p Lewis weights (Lewis 1978)
  - Online leverage scores (Cohen—Musco—Pachocki 2016)

Matrix Approximation

Non-oblivious Subspace Embeddings

Definition (Leverage scores). For $\mathbf{A} \in \mathbb{R}^{n\times d}$ and $i \in [n]$, the ith leverage score is

$$\tau_i(\mathbf{A}) = \sup_{\mathbf{A}\mathbf{x}\neq 0} \frac{\langle \mathbf{a}_i, \mathbf{x}\rangle^2}{\|\mathbf{A}\mathbf{x}\|_2^2}$$

Contribution of the ith coordinate

ℓ_2 norm of $\mathbf{Ax}$

"Importance score": largest fraction of ℓ_2 norm occupied by the ith coordinate

- Many generalizations:
  - ℓ_p sensitivity scores (Langberg—Schulman 2010)
  - ℓ_p Lewis weights (Lewis 1978)
  - Online leverage scores (Cohen—Musco—Pachocki 2016)

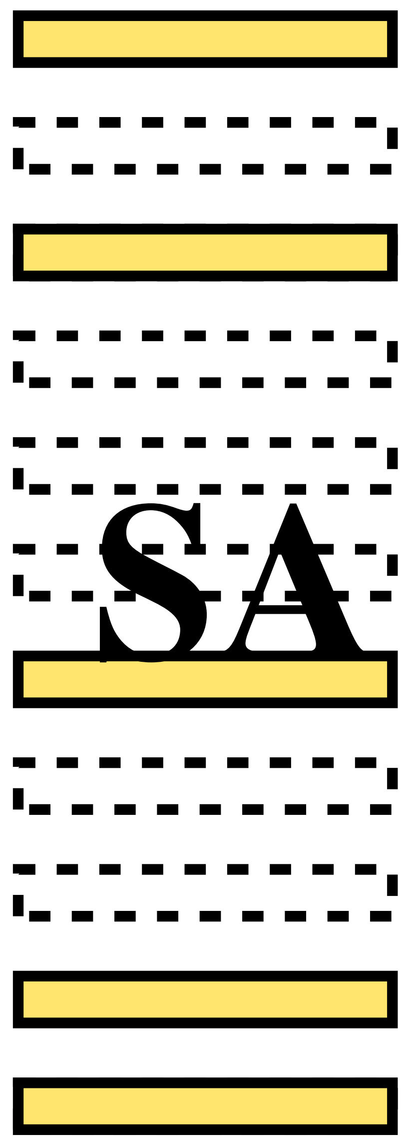
Matrix Approximation

Non-oblivious Subspace Embeddings

Theorem (Lewis weight sampling, Cohen—Peng 2015). For any $\mathbf{A} \in \mathbb{R}^{n \times d}$, there are probabilities $q_1, q_2, \ldots q_n$ that sample

$$r = \begin{cases} \tilde{O}(\varepsilon^{-2} d) & p < 2 \\ \tilde{O}(\varepsilon^{-5} d^{p/2}) & p > 2 \end{cases}$$

rows of $\mathbf{A}$ that forms an ℓ_p subspace embedding with distortion $\kappa = (1 + \varepsilon)$, with probability 99%.



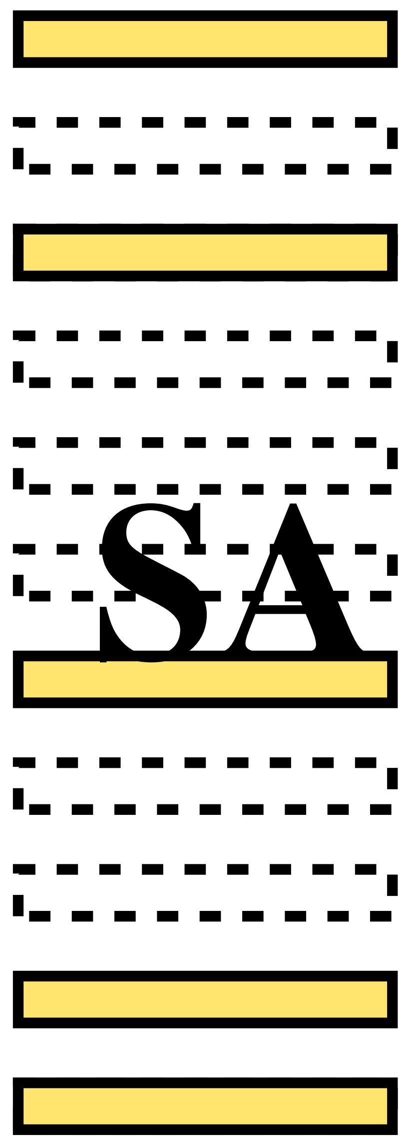
Matrix Approximation

Non-oblivious Subspace Embeddings

Theorem (Lewis weight sampling, Cohen—Peng 2015). For any $\mathbf{A} \in \mathbb{R}^{n \times d}$, there are probabilities $q_1, q_2, \ldots q_n$ that sample

$$\boxed{r = \begin{cases} \tilde{O}(\varepsilon^{-2} d) & p < 2 \\ \tilde{O}(\varepsilon^{-5} d^{p/2}) & p > 2 \end{cases}}$$

rows of $\mathbf{A}$ that forms an ℓ_p subspace embedding with distortion $\boxed{\kappa = (1 + \varepsilon)}$ with probability 99%.



Matrix Approximation

Non-oblivious Subspace Embeddings

$$\boxed{r = \begin{cases} \tilde{O}(\varepsilon^{-2} d) & p < 2 \\ \tilde{O}(\varepsilon^{-5} d^{p/2}) & p > 2 \end{cases}}$$

Matrix Approximation

Non-oblivious Subspace Embeddings

Two questions:

$$r = \begin{cases} \tilde{O}(\varepsilon^{-2} d) & p < 2 \\ \tilde{O}(\varepsilon^{-5} d^{p/2}) & p > 2 \end{cases}$$

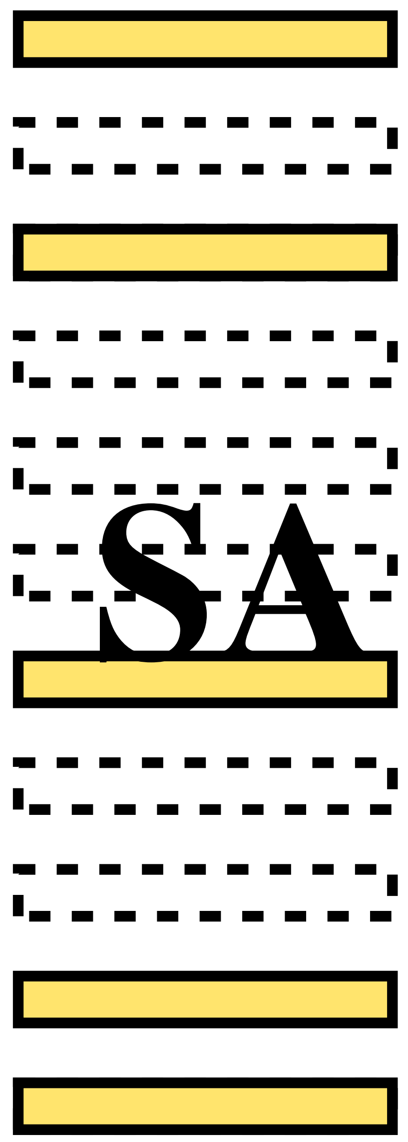
Matrix Approximation

Non-oblivious Subspace Embeddings

Two questions:

Question 1. Can these bounds be improved?

$$r = \begin{cases} \tilde{O}(\varepsilon^{-2} d) & p < 2 \\ \tilde{O}(\varepsilon^{-5} d^{p/2}) & p > 2 \end{cases}$$



Matrix Approximation

Non-oblivious Subspace Embeddings

Two questions:

Question 1. Can these bounds be improved?

$$r = \begin{cases} \tilde{O}(\varepsilon^{-2} d) & p < 2 \\ \tilde{O}(\varepsilon^{-5} d^{p/2}) & p > 2 \end{cases}$$

Question 2. What if the rows of $\mathbf{A}$ arrive one by one in a stream?

Matrix Approximation

Non-oblivious Subspace Embeddings

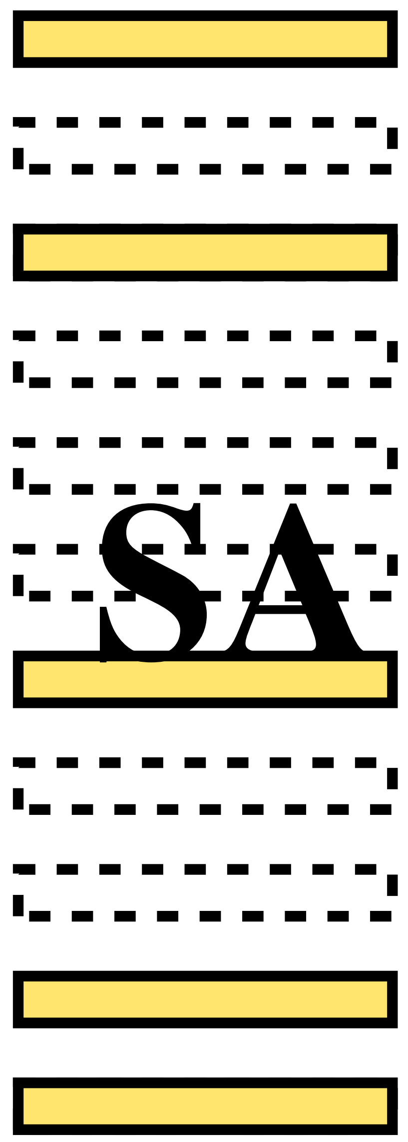
Two questions:

Question 1. Can these bounds be improved?

$$r = \begin{cases} \tilde{O}(\varepsilon^{-2} d) & p < 2 \\ \tilde{O}(\varepsilon^{-5} d^{p/2}) & p > 2 \end{cases}$$

Nearly optimal (Li—Wang—Woodruff 2020)

Question 2. What if the rows of $\mathbf{A}$ arrive one by one in a stream?



Matrix Approximation

Non-oblivious Subspace Embeddings

Two questions:

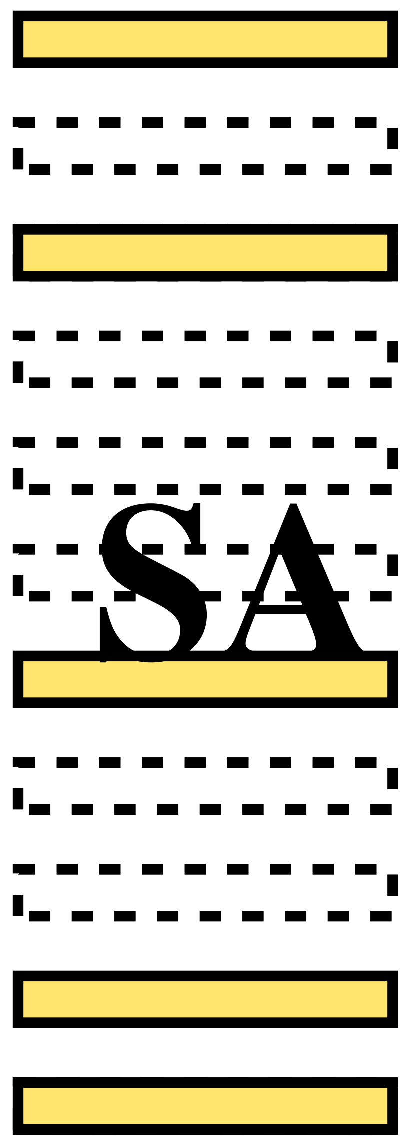
Question 1. Can these bounds be improved?

$$r = \begin{cases} \tilde{O}(\varepsilon^{-2} d) & p < 2 \\ \tilde{O}(\varepsilon^{-5} d^{p/2}) & p > 2 \end{cases}$$

Nearly optimal (Li—Wang—Woodruff 2020)

$d^{p/2}$ is nearly optimal for $\varepsilon = O(1)$
(Li—Wang—Woodruff 2020)

Question 2. What if the rows of $\mathbf{A}$ arrive one by one in a stream?



Matrix Approximation

Non-oblivious Subspace Embeddings

Two questions:

Question 1. Can these bounds be improved?

$$r = \begin{cases} \tilde{O}(\varepsilon^{-2}d) & p < 2 \\ \tilde{O}(\varepsilon^{-5}d^{p/2}) & p > 2 \end{cases}$$

Nearly optimal (Li—Wang—Woodruff 2020)

ε^{-2} should be possible here!

$d^{p/2}$ is nearly optimal for $\varepsilon = O(1)$ (Li—Wang—Woodruff 2020)

Question 2. What if the rows of $\mathbf{A}$ arrive one by one in a stream?

Matrix Approximation

Non-oblivious Subspace Embeddings

Question 1. Can these bounds be improved?

$$r = \begin{cases} \tilde{O}(\varepsilon^{-2} d) & p < 2 \\ \tilde{O}(\varepsilon^{-5} d^{p/2}) & p > 2 \end{cases}$$

Question 2. What if the rows of $\mathbf{A}$ arrive one by one in a stream?

Matrix Approximation

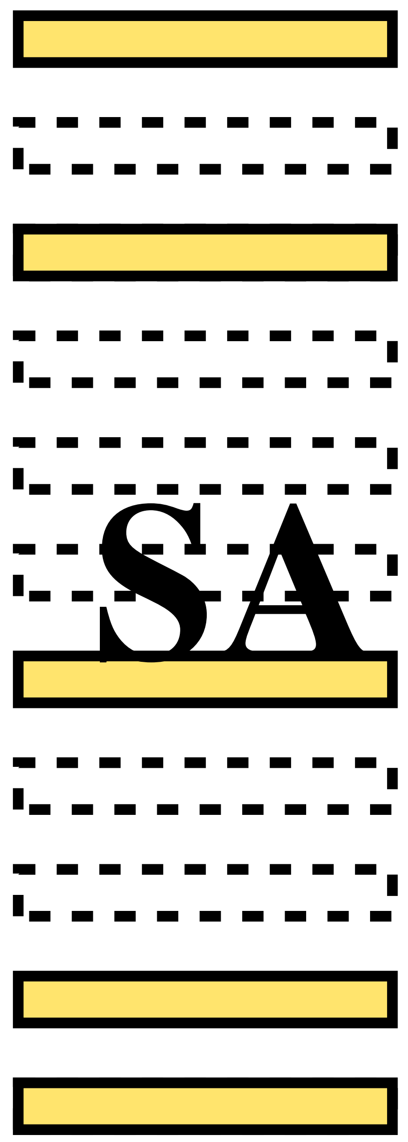
Non-oblivious Subspace Embeddings

Woodruff—Y 2023

Question 1. Can these bounds be improved?

$$r = \begin{cases} \tilde{O}(\varepsilon^{-2} d) & p < 2 \\ \tilde{O}(\varepsilon^{-5} d^{p/2}) & p > 2 \end{cases}$$

Question 2. What if the rows of $\mathbf{A}$ arrive one by one in a stream?



Matrix Approximation

Non-oblivious Subspace Embeddings

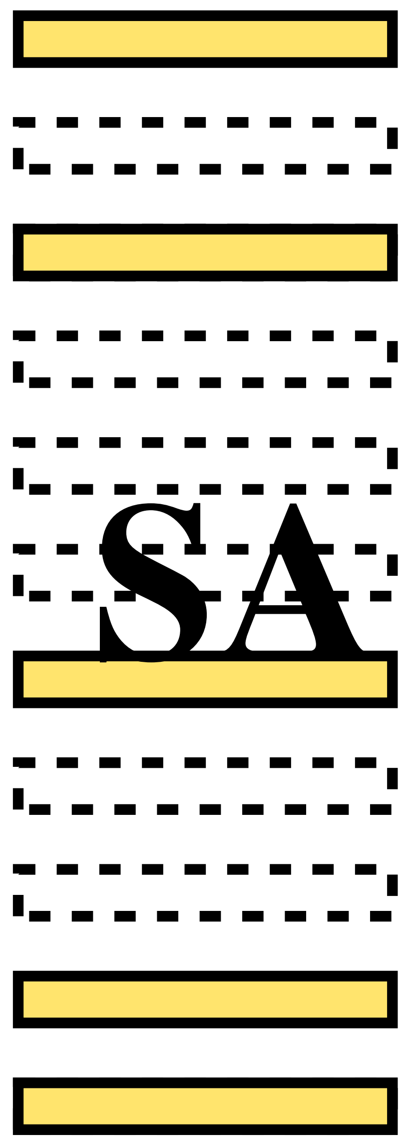
Woodruff—Y 2023

Question 1. Can these bounds be improved?

Questions 1 and 2 can be addressed using related techniques!

$$r = \begin{cases} \tilde{O}(\varepsilon^{-2} d) & p < 2 \\ \tilde{O}(\varepsilon^{-5} d^{p/2}) & p > 2 \end{cases}$$

Question 2. What if the rows of $\mathbf{A}$ arrive one by one in a stream?



Matrix Approximation

Non-oblivious Subspace Embeddings

Woodruff—Y 2023

Question 1. Can these bounds be improved?

Questions 1 and 2 can be addressed using related techniques!

$$r = \begin{cases} \tilde{O}(\varepsilon^{-2} d) & p < 2 \\ \tilde{O}(\varepsilon^{-5} d^{p/2}) & p > 2 \end{cases}$$

Question 2. What if the rows of $\mathbf{A}$ arrive one by one in a stream?

Handling addition of rows in a stream → better sampling bounds

Matrix Approximation

Non-oblivious Subspace Embeddings

Woodruff—Y 2023

Question 1. Can these bounds be improved?

Questions 1 and 2 can be addressed using related techniques!

$$r = \begin{cases} \tilde{O}(\varepsilon^{-2}d) & p < 2 \\ \tilde{O}(\cancel{\varepsilon^{-\text{[illegible]}}d^{p/2}}) & p > 2 \end{cases}$$

Question 2. What if the rows of $\mathbf{A}$ arrive one by one in a stream?

Handling addition of rows in a stream → better sampling bounds

Matrix Approximation

Non-oblivious Subspace Embeddings

Woodruff—Y 2023

Question 1. Can these bounds be improved?

Questions 1 and 2 can be addressed using related techniques!

$$r = \begin{cases} \tilde{O}(\varepsilon^{-2}d) & p < 2 \\ \cancel{\tilde{O}(\varepsilon^{-2}d^{p/2})} & p > 2 \end{cases}$$

$\tilde{O}(\varepsilon^{-2}d^{p/2})$ (Woodruff—Y 2023)

Question 2. What if the rows of $\mathbf{A}$ arrive one by one in a stream?

Handling addition of rows in a stream → better sampling bounds

Matrix Approximation

Non-oblivious Subspace Embeddings

Woodruff—Y 2023

Question 1. Can these bounds be improved?

These bounds also hold in the streaming setting (Woodruff—Y 2023)

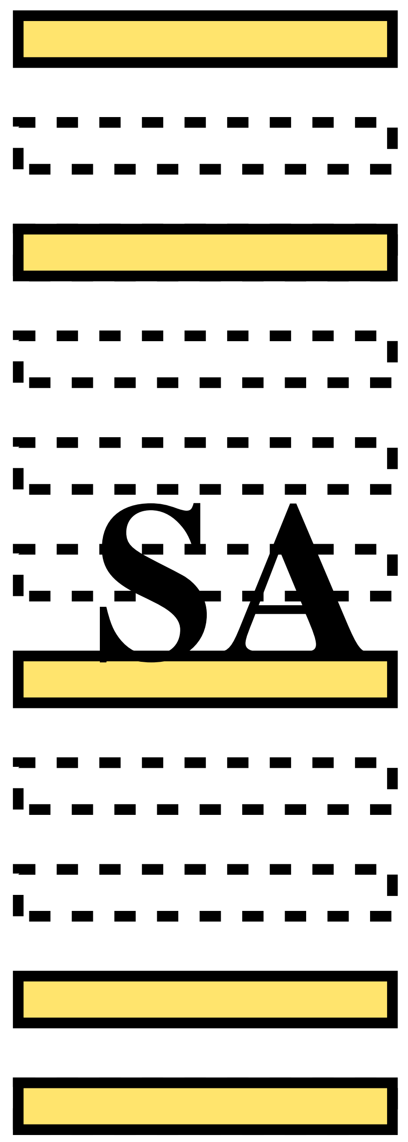
Questions 1 and 2 can be addressed using related techniques!

$$r = \begin{cases} \tilde{O}(\varepsilon^{-2}d) & p < 2 \\ \cancel{\tilde{O}(\varepsilon^{-2}d^{p/2})} & p > 2 \end{cases}$$

$\tilde{O}(\varepsilon^{-2}d^{p/2})$ (Woodruff—Y 2023)

Question 2. What if the rows of $\mathbf{A}$ arrive one by one in a stream?

Handling addition of rows in a stream → better sampling bounds



Matrix Approximation

Non-oblivious Subspace Embeddings

Question 3. Are there similar results for other loss functions?

Matrix Approximation

Non-oblivious Subspace Embeddings

Question 3. Are there similar results for other loss functions?

$$\|\mathbf{Ax}\|_g := \sum_{i=1}^{n} g([\mathbf{Ax}](i))$$

Matrix Approximation

Non-oblivious Subspace Embeddings

Question 3. Are there similar results for other loss functions?

$$\|\mathbf{Ax}\|_g := \sum_{i=1}^{n} g([\mathbf{Ax}](i))$$

- There are many other loss functions used for linear regression…

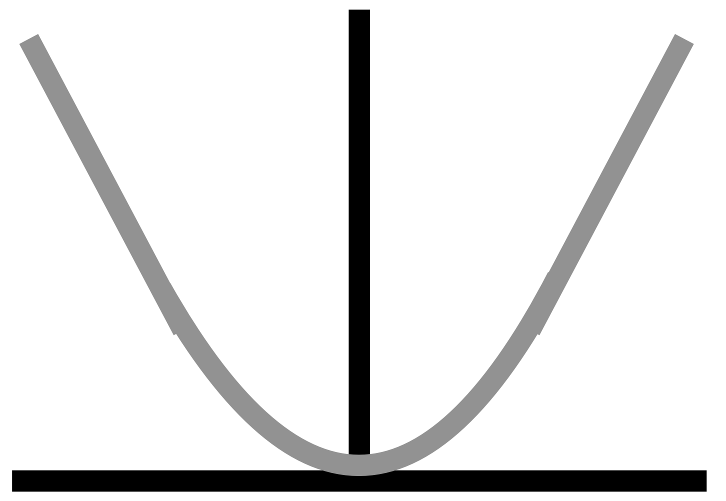
Matrix Approximation

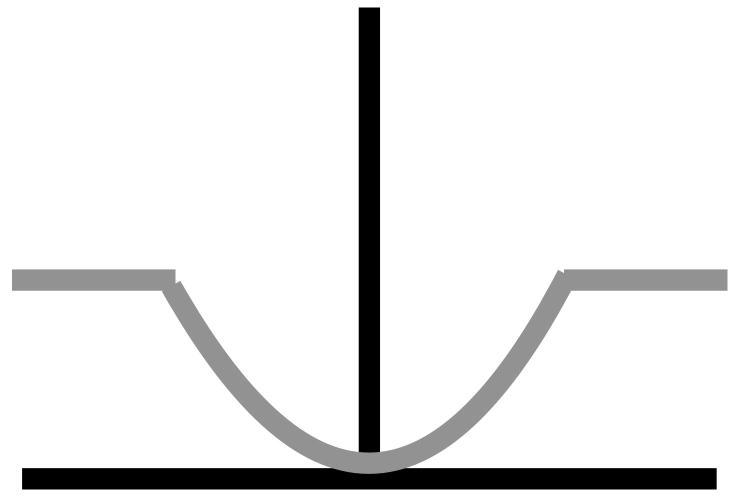
Non-oblivious Subspace Embeddings

Question 3. Are there similar results for other loss functions?

$$\|\mathbf{A}\mathbf{x}\|_g := \sum_{i=1}^{n} g([\mathbf{A}\mathbf{x}](i))$$

- There are many other loss functions used for linear regression…
  - Huber/Tukey loss for robust statistics





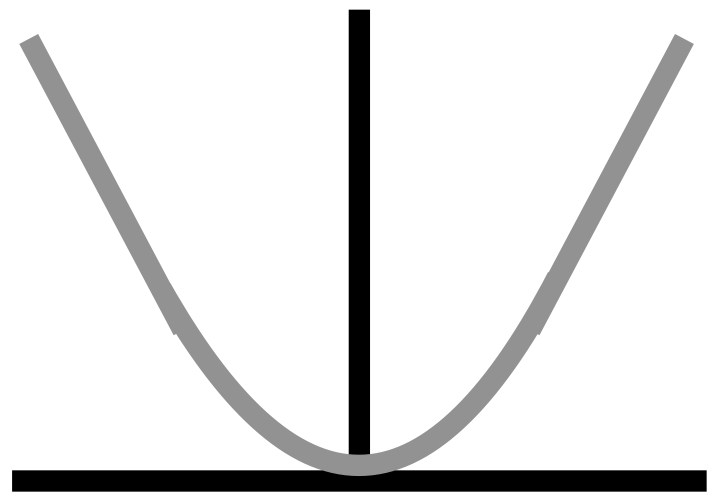
Matrix Approximation

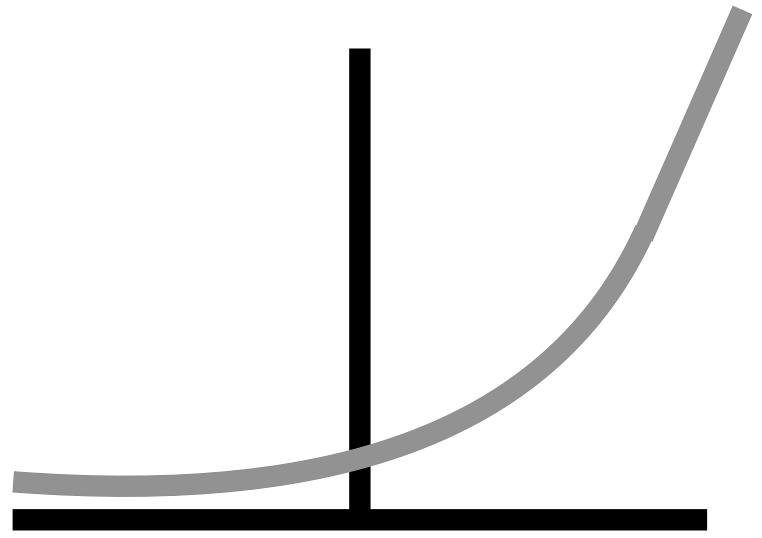
Non-oblivious Subspace Embeddings

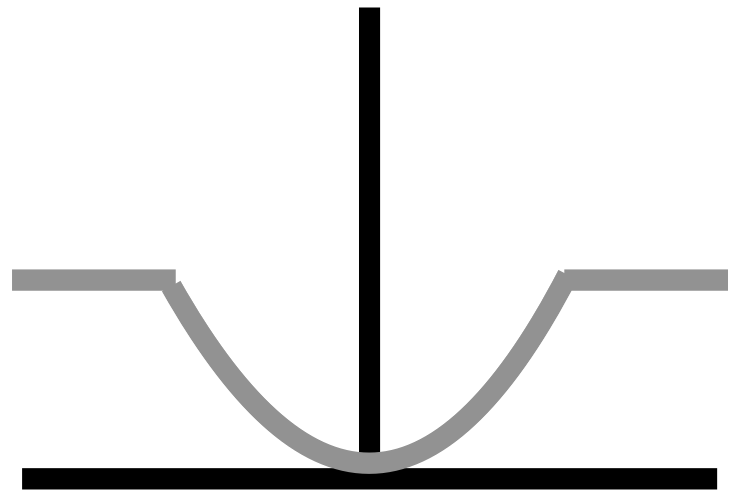
Question 3. Are there similar results for other loss functions?

$$\|\mathbf{Ax}\|_g := \sum_{i=1}^{n} g([\mathbf{Ax}](i))$$

- There are many other loss functions used for linear regression…
  - Huber/Tukey loss for robust statistics
  - Logistic regression for classification





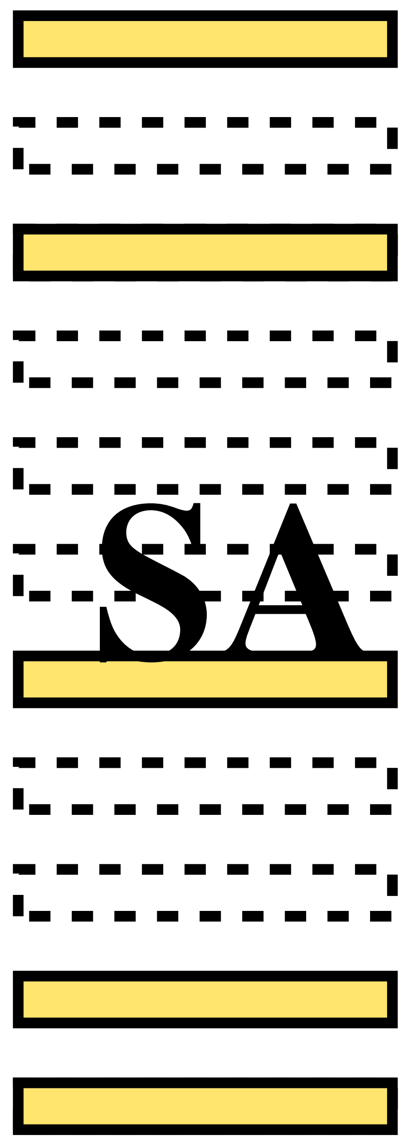


Matrix Approximation

Non-oblivious Subspace Embeddings

Question 3. Are there similar results for other loss functions?

$$\|\mathbf{Ax}\|_g := \sum_{i=1}^{n} g([\mathbf{Ax}](i))$$



Matrix Approximation

Non-oblivious Subspace Embeddings

Question 3. Are there similar results for other loss functions?

$$\|\mathbf{A}\mathbf{x}\|_g := \sum_{i=1}^{n} g([\mathbf{A}\mathbf{x}](i))$$

Musco—Musco—Woodruff—Y 2022

Matrix Approximation

Non-oblivious Subspace Embeddings

Question 3. Are there similar results for other loss functions?

$$\|\mathbf{Ax}\|_g := \sum_{i=1}^{n} g([\mathbf{Ax}](i))$$

Musco—Musco—Woodruff—Y 2022

General losses with quadratic growth

Matrix Approximation

Non-oblivious Subspace Embeddings

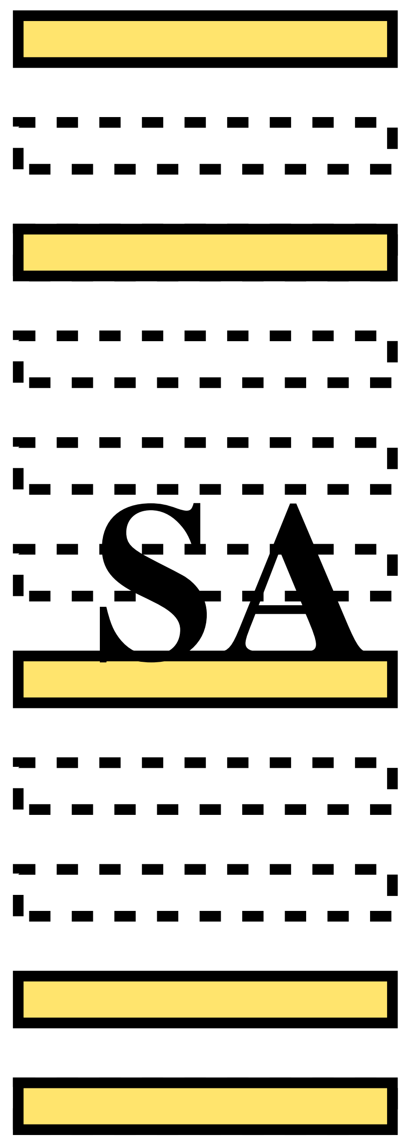
Question 3. Are there similar results for other loss functions?

$$\|\mathbf{Ax}\|_g := \sum_{i=1}^{n} g([\mathbf{Ax}](i))$$

Musco—Musco—Woodruff—Y 2022

General losses with quadratic growth

$$r = \tilde{O}(\varepsilon^{-2} d^2)$$



Matrix Approximation

Non-oblivious Subspace Embeddings

Question 3. Are there similar results for other loss functions?

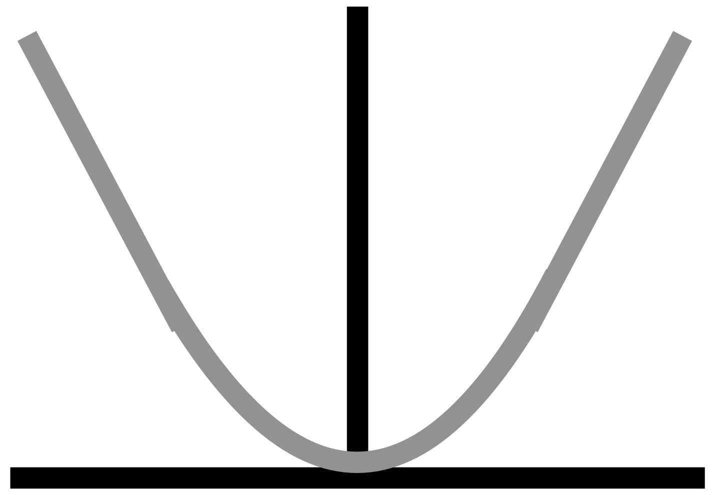
$$\|\mathbf{Ax}\|_g := \sum_{i=1}^{n} g([\mathbf{Ax}](i))$$

Musco—Musco—Woodruff—Y 2022

General losses with quadratic growth

$r = \tilde{O}(\varepsilon^{-2} d^2)$

Huber loss



Matrix Approximation

Non-oblivious Subspace Embeddings

Question 3. Are there similar results for other loss functions?

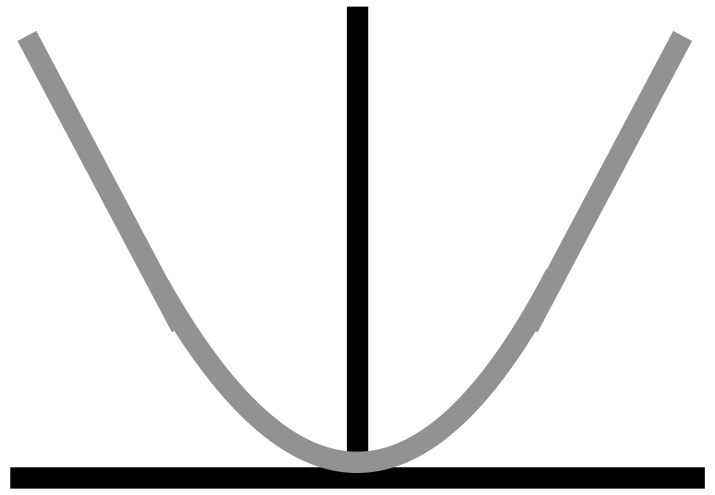
$$\|\mathbf{Ax}\|_g := \sum_{i=1}^{n} g([\mathbf{Ax}](i))$$

Musco—Musco—Woodruff—Y 2022

General losses with quadratic growth

$r = \tilde{O}(\varepsilon^{-2} d^2)$

Huber loss



$r = \mathrm{poly}(\varepsilon^{-1}) d^{4-2\sqrt{2}} \approx d^{1.172}$

Matrix Approximation

Non-oblivious Subspace Embeddings

Question 3. Are there similar results for other loss functions?

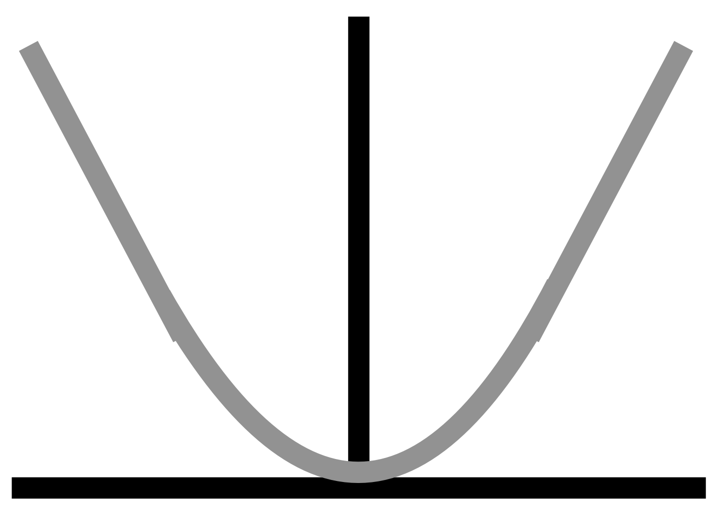
$$\|\mathbf{Ax}\|_g := \sum_{i=1}^{n} g([\mathbf{Ax}](i))$$

Musco—Musco—Woodruff—Y 2022

General losses with quadratic growth

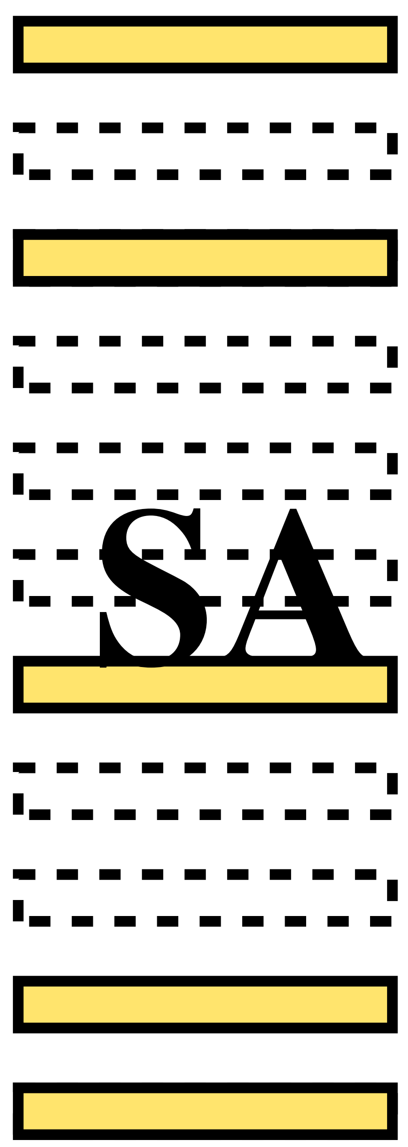
$$r = \tilde{O}(\varepsilon^{-2} d^2)$$

Huber loss



$$r = \mathrm{poly}(\varepsilon^{-1}) d^{4-2\sqrt{2}} \approx d^{1.172}$$

I am very interested in improving this to d



Subspace Embeddings and Linear Regression

- Oblivious ℓ_p subspace embeddings: high distortion and low distortion
- Non-oblivious subspace embeddings: ℓ_p Lewis weight sampling, general losses

$\boxed{\text{Applications: active learning, streaming computational geometry, low rank approximation}}$

Matrix Approximation

Active ℓ_p Linear Regression

$$\min_{\mathbf{x} \in \mathbb{R}^d} \| \mathbf{A}\mathbf{x} - \mathbf{b} \|_p^p$$

Matrix Approximation

Active ℓ_p Linear Regression

$$\min_{\mathbf{x}\in\mathbb{R}^d} \|\mathbf{Ax}-\mathbf{b}\|_p^p$$

- Active learning: machine learning, where **label acquisition** is the most expensive resource

Matrix Approximation

Active ℓ_p Linear Regression

$$\min_{\mathbf{x}\in\mathbb{R}^d} \|\mathbf{A}\mathbf{x}-\mathbf{b}\|_p^p$$

- Active learning: machine learning, where **label acquisition** is the most expensive resource
  - Labeling could require manual labor

Matrix Approximation

Active ℓ_p Linear Regression

$$\min_{\mathbf{x}\in\mathbb{R}^d} \|\mathbf{A}\mathbf{x}-\mathbf{b}\|_p^p$$

- Active learning: machine learning, where **label acquisition** is the most expensive resource
  - Labeling could require manual labor
  - Labeling could require purchasing information

Matrix Approximation

Active ℓ_p Linear Regression

$$\min_{\mathbf{x}\in\mathbb{R}^d} \|\mathbf{A}\mathbf{x}-\mathbf{b}\|_p^p$$

- Active learning: machine learning, where **label acquisition** is the most expensive resource
  - Labeling could require manual labor
  - Labeling could require purchasing information
  - Labeling could require involve an invasive medical procedure

Matrix Approximation

Active ℓ_p Linear Regression

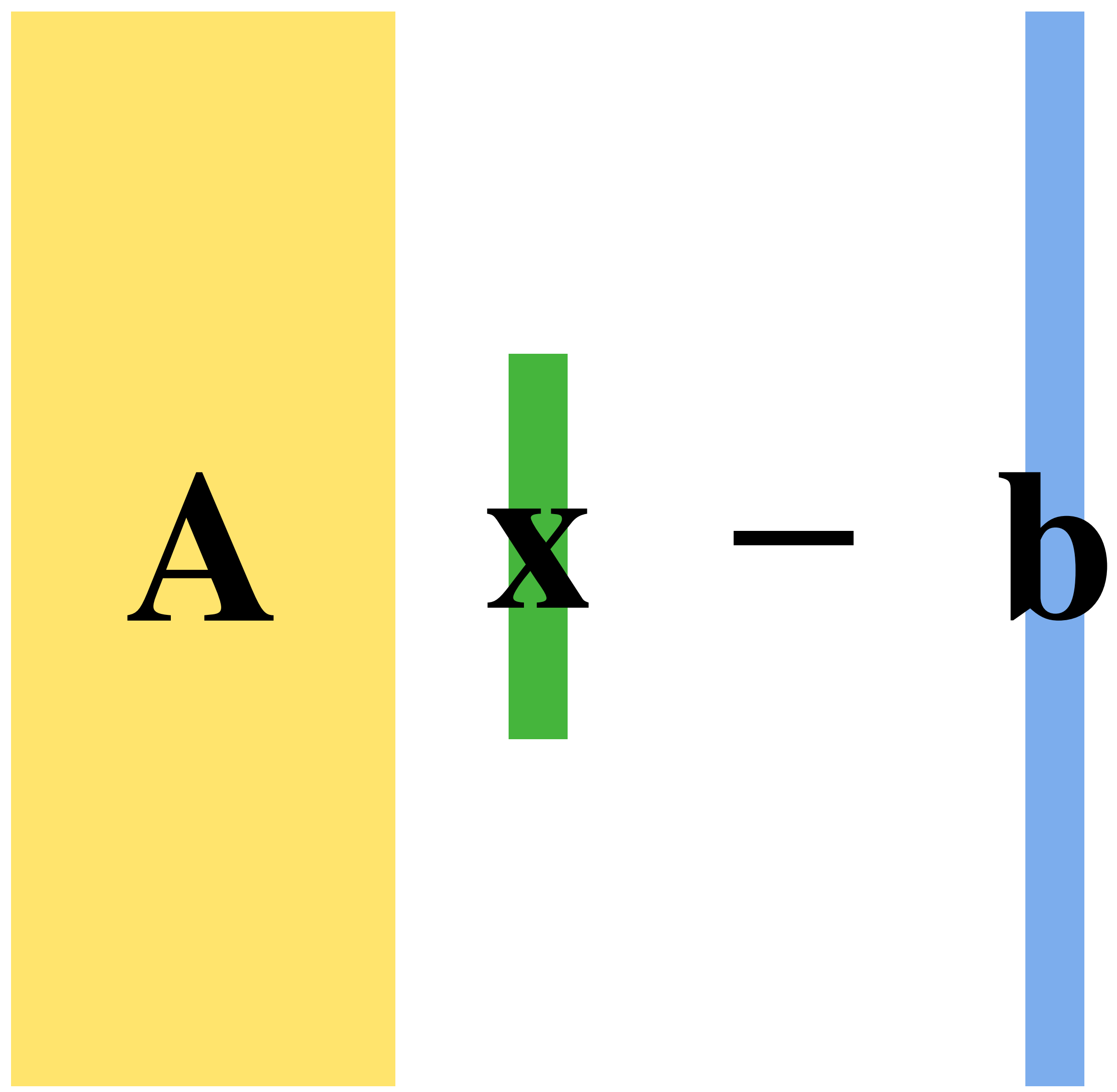
$$\min_{\mathbf{x}\in\mathbb{R}^d} \|\mathbf{A}\mathbf{x}-\mathbf{b}\|_p^p$$

- Active learning: machine learning, where **label acquisition** is the most expensive resource
  - Labeling could require manual labor
  - Labeling could require purchasing information
  - Labeling could require involve an invasive medical procedure
- Goal: **minimize the number of label entries** that are read

Matrix Approximation

Active ℓ_p Linear Regression

$$\min_{\mathbf{x}\in\mathbb{R}^d} \| \mathbf{A}\mathbf{x} - \mathbf{b} \|_p^p$$

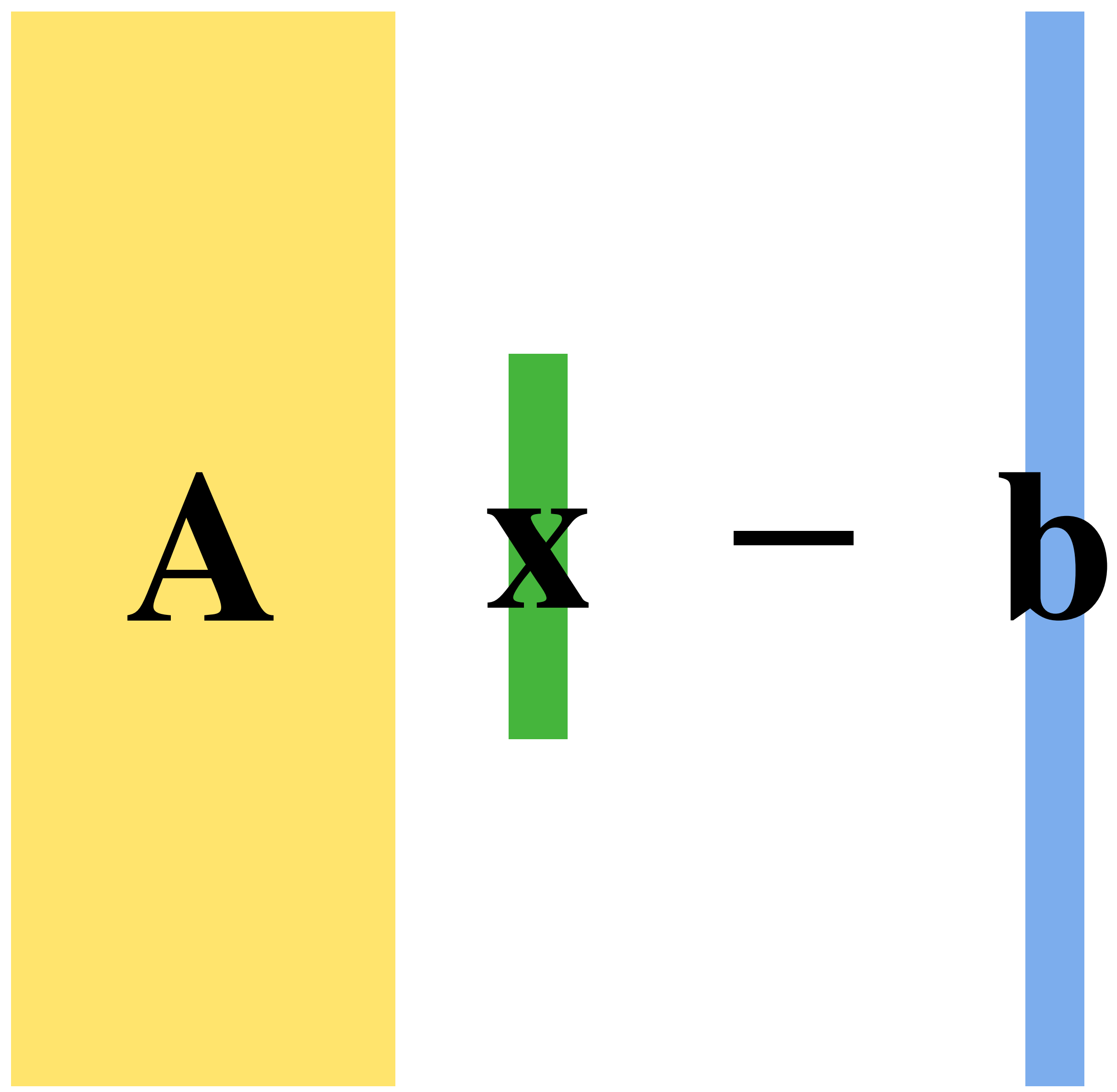


Matrix Approximation

Active ℓ_p Linear Regression

$$\min_{\mathbf{x}\in\mathbb{R}^d} \|\mathbf{A}\mathbf{x}-\mathbf{b}\|_p^p$$

The algorithm has full access to $\mathbf{A}$

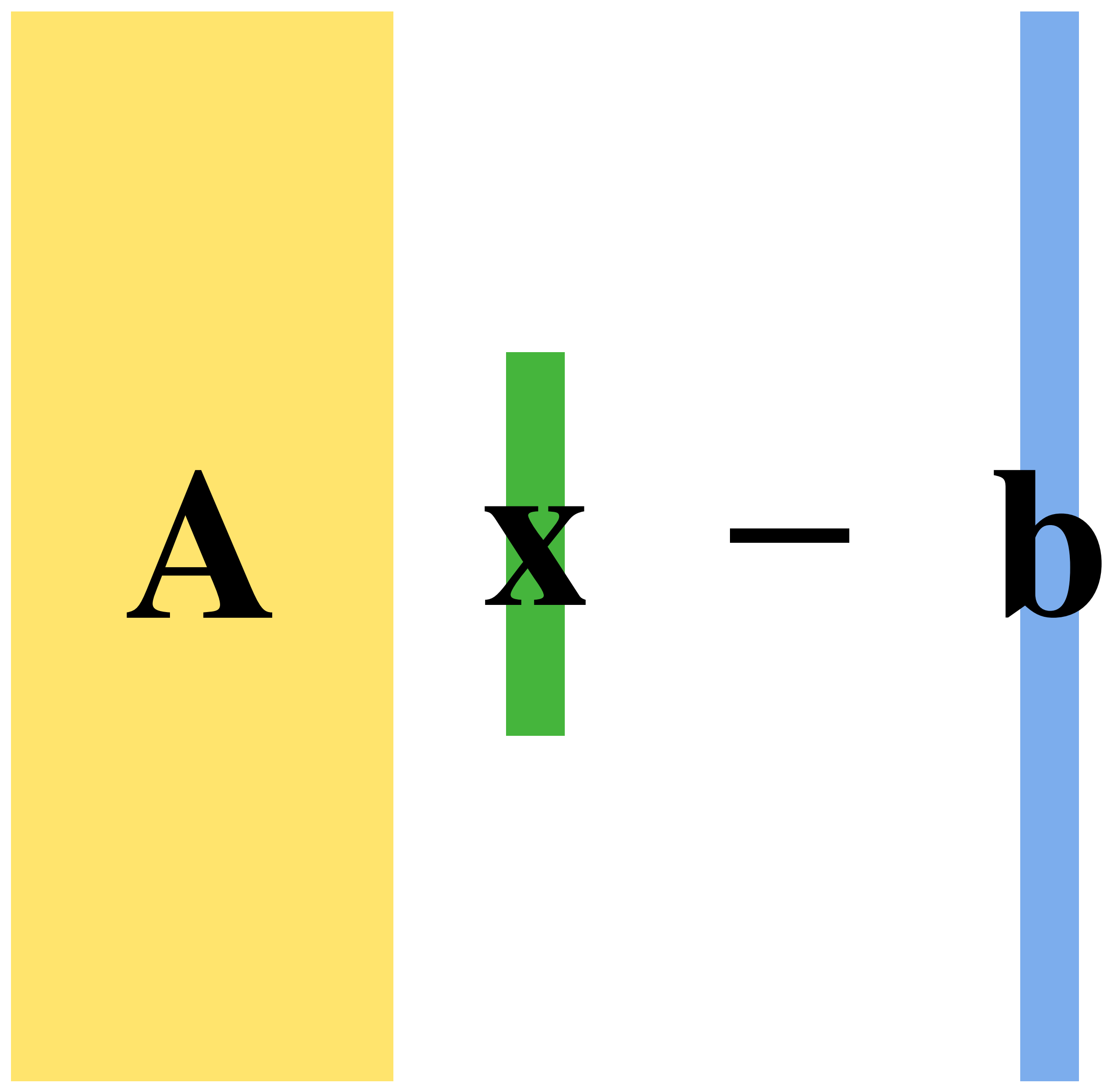


Matrix Approximation

Active ℓ_p Linear Regression

$$\min_{\mathbf{x}\in\mathbb{R}^d} \|\mathbf{A}\mathbf{x}-\mathbf{b}\|_p^p$$

The algorithm has full access to $\mathbf{A}$



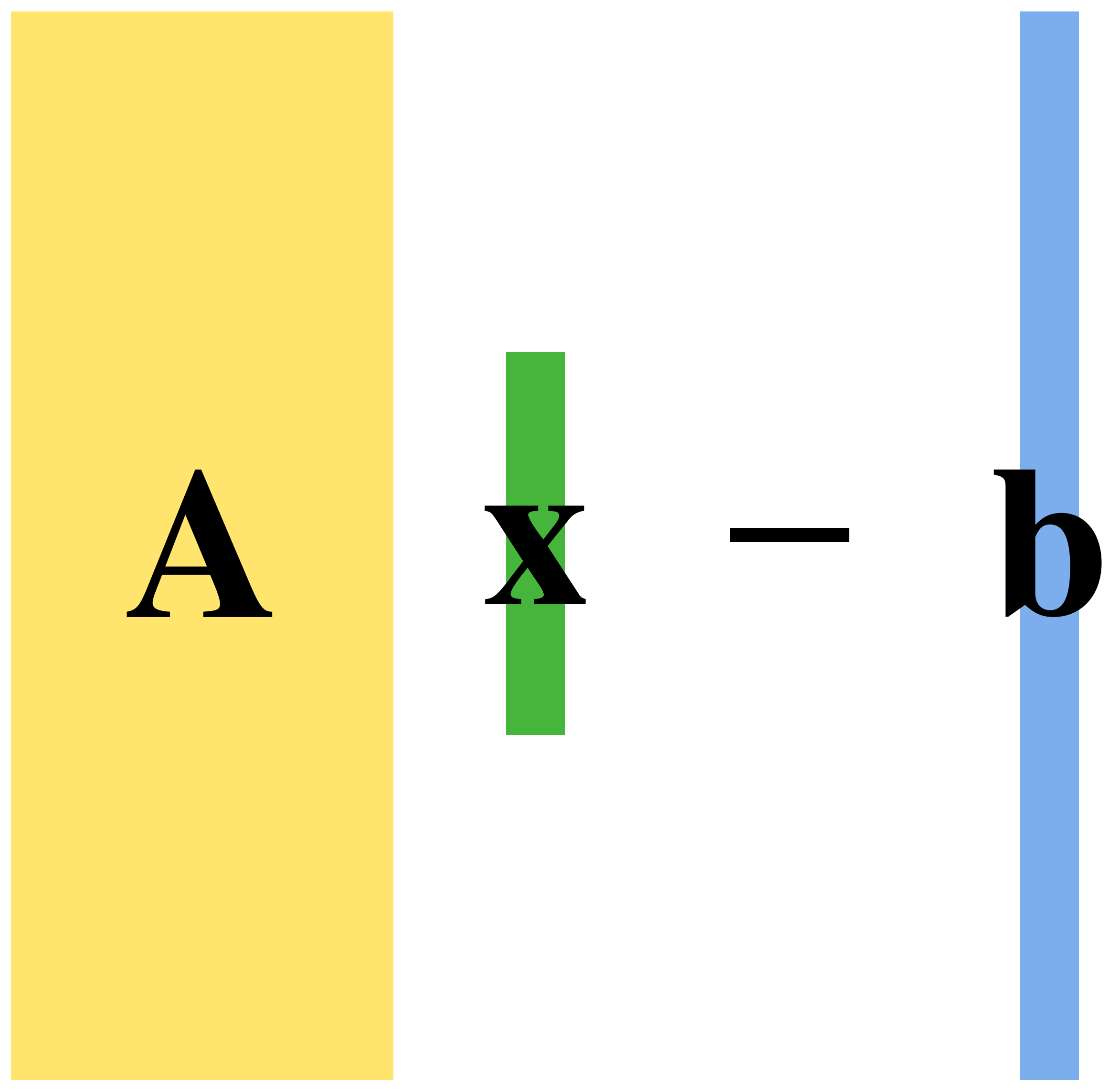
$\mathbf{b}$ vector is hidden, and the algorithm has query access to it

Matrix Approximation

Active ℓ_p Linear Regression

$$\min_{\mathbf{x}\in\mathbb{R}^d} \|\mathbf{A}\mathbf{x}-\mathbf{b}\|_p^p$$

The algorithm has full access to $\mathbf{A}$



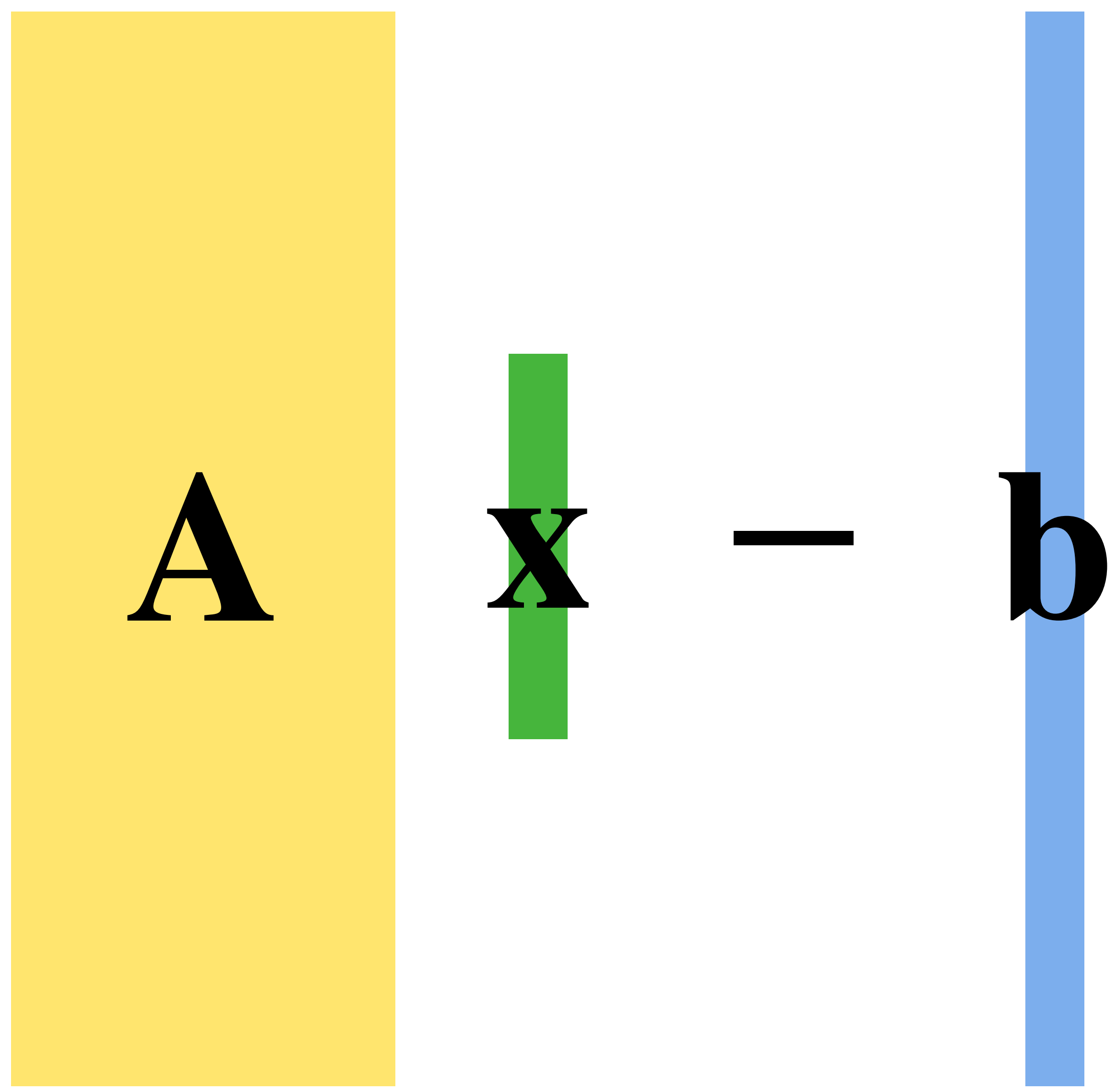
$\mathbf{b}$ vector is hidden, and the algorithm has query access to it

Question. How many entries of $\mathbf{b}$ need to be read?

Matrix Approximation

Active ℓ_p Linear Regression

$$\min_{\mathbf{x}\in\mathbb{R}^d} \| \mathbf{A}\mathbf{x} - \mathbf{b} \|_p^p$$



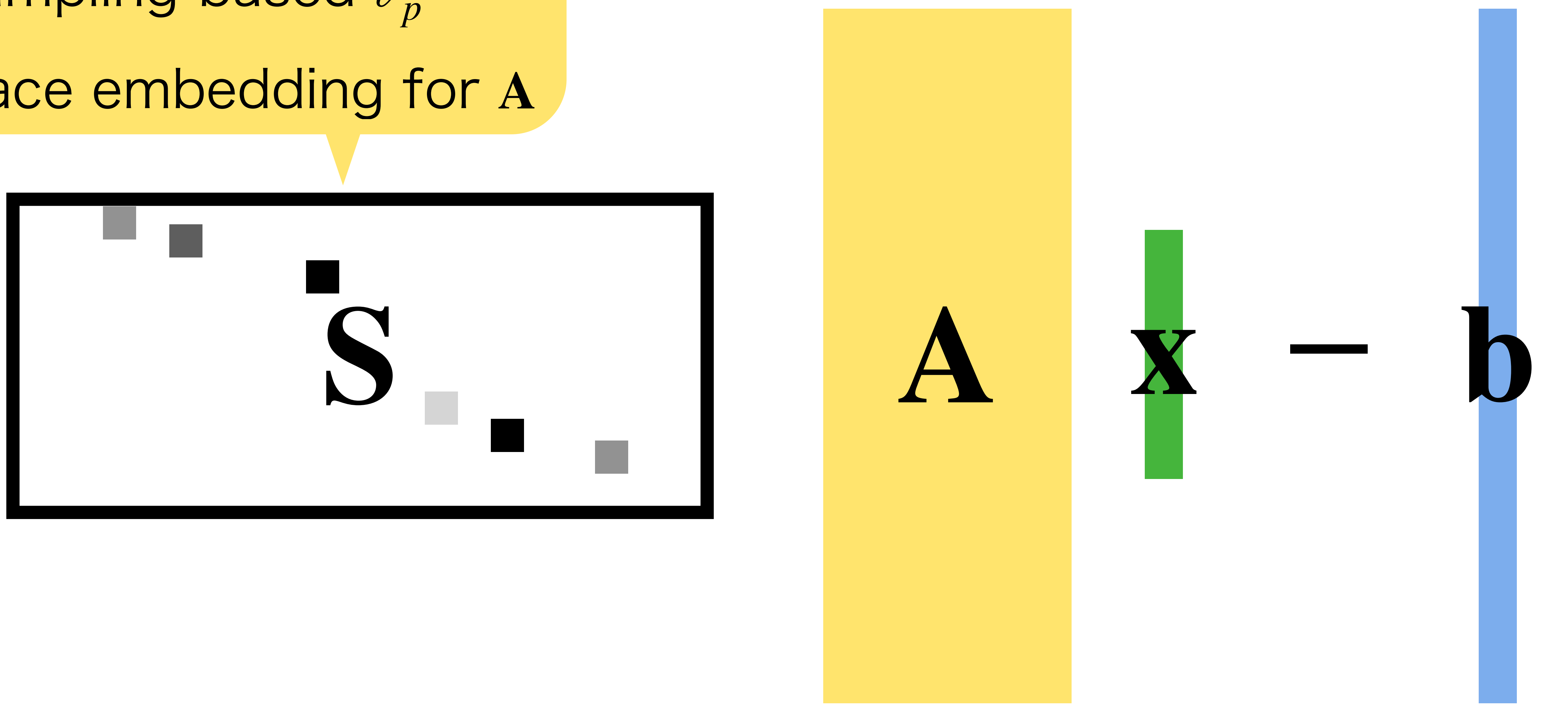
Question. How many entries of $\mathbf{b}$ need to be read?

Matrix Approximation

Active ℓ_p Linear Regression

$$\min_{\mathbf{x}\in\mathbb{R}^d} \| \mathbf{A}\mathbf{x} - \mathbf{b} \|_p^p$$

Sampling-based ℓ_p subspace embedding for $\mathbf{A}$

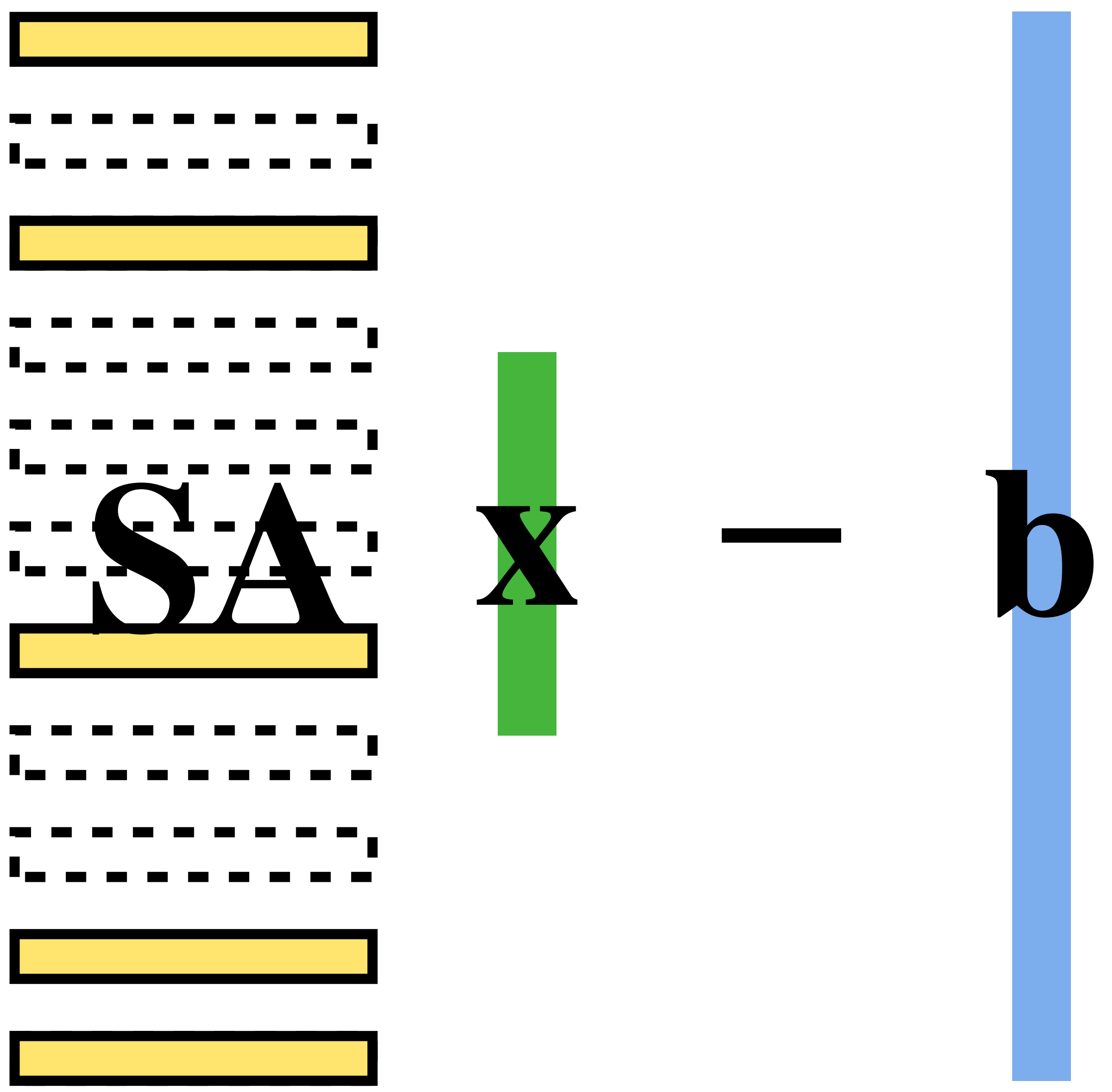


Question. How many entries of $\mathbf{b}$ need to be read?

Matrix Approximation

Active ℓ_p Linear Regression

$$\min_{\mathbf{x}\in\mathbb{R}^d} \|\mathbf{A}\mathbf{x}-\mathbf{b}\|_p^p$$

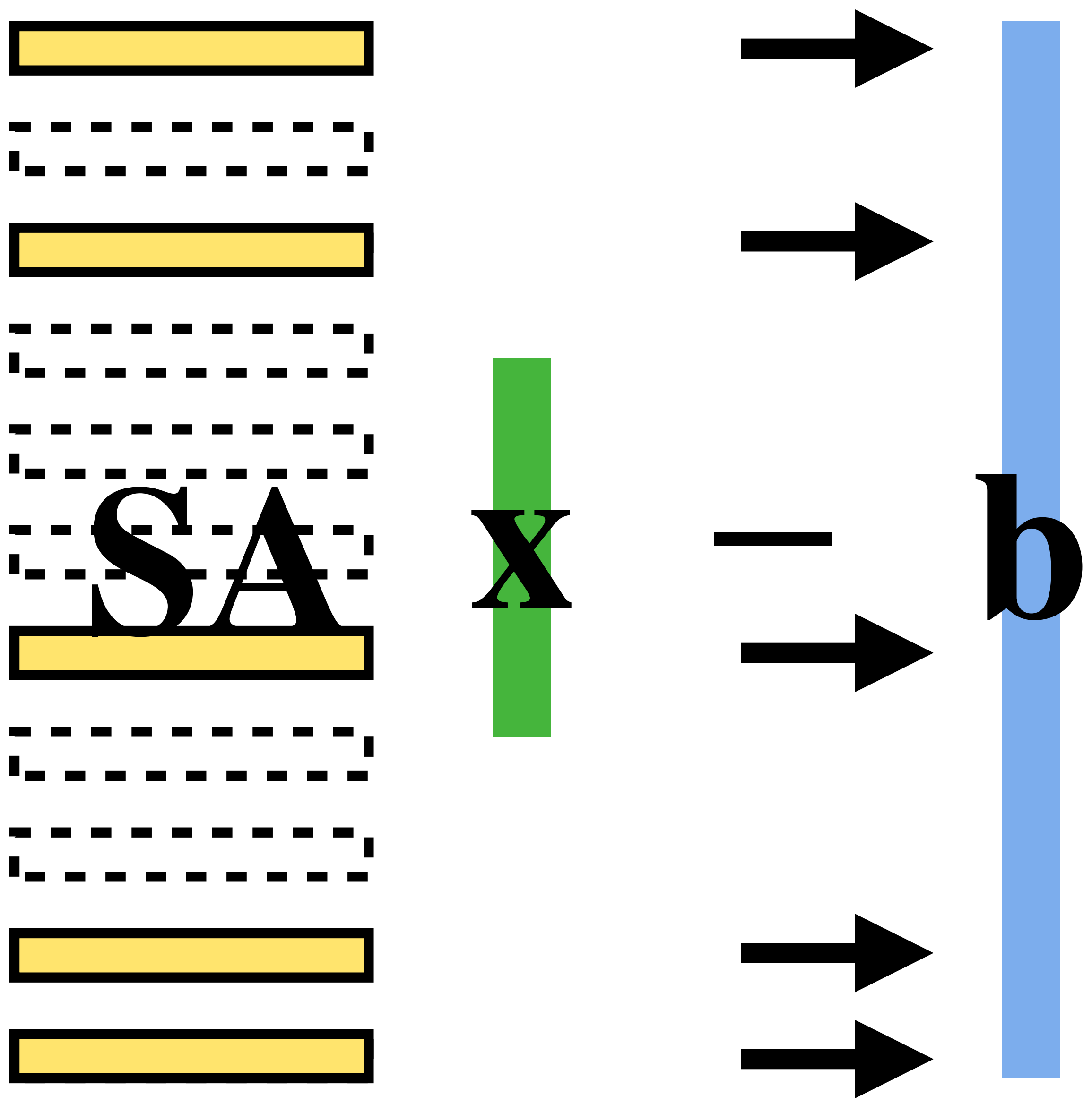


Question. How many entries of $\mathbf{b}$ need to be read?

Matrix Approximation

Active ℓ_p Linear Regression

$$\min_{\mathbf{x}\in\mathbb{R}^d} \|\mathbf{A}\mathbf{x}-\mathbf{b}\|_p^p$$

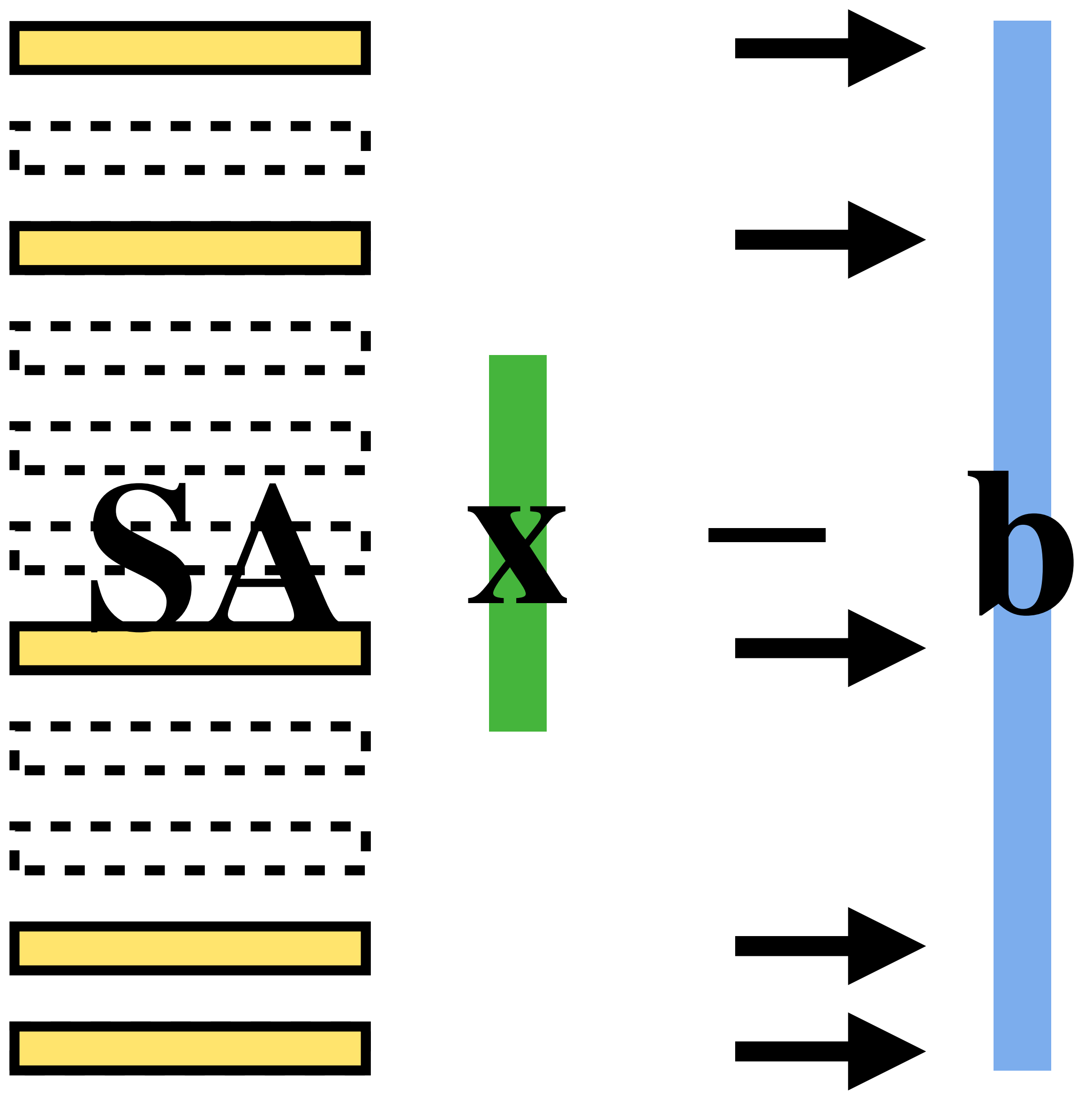


Question. How many entries of $\mathbf{b}$ need to be read?

Matrix Approximation

Active ℓ_p Linear Regression

$$\min_{\mathbf{x}\in\mathbb{R}^d} \| \mathbf{Ax} - \mathbf{b} \|_p^p$$

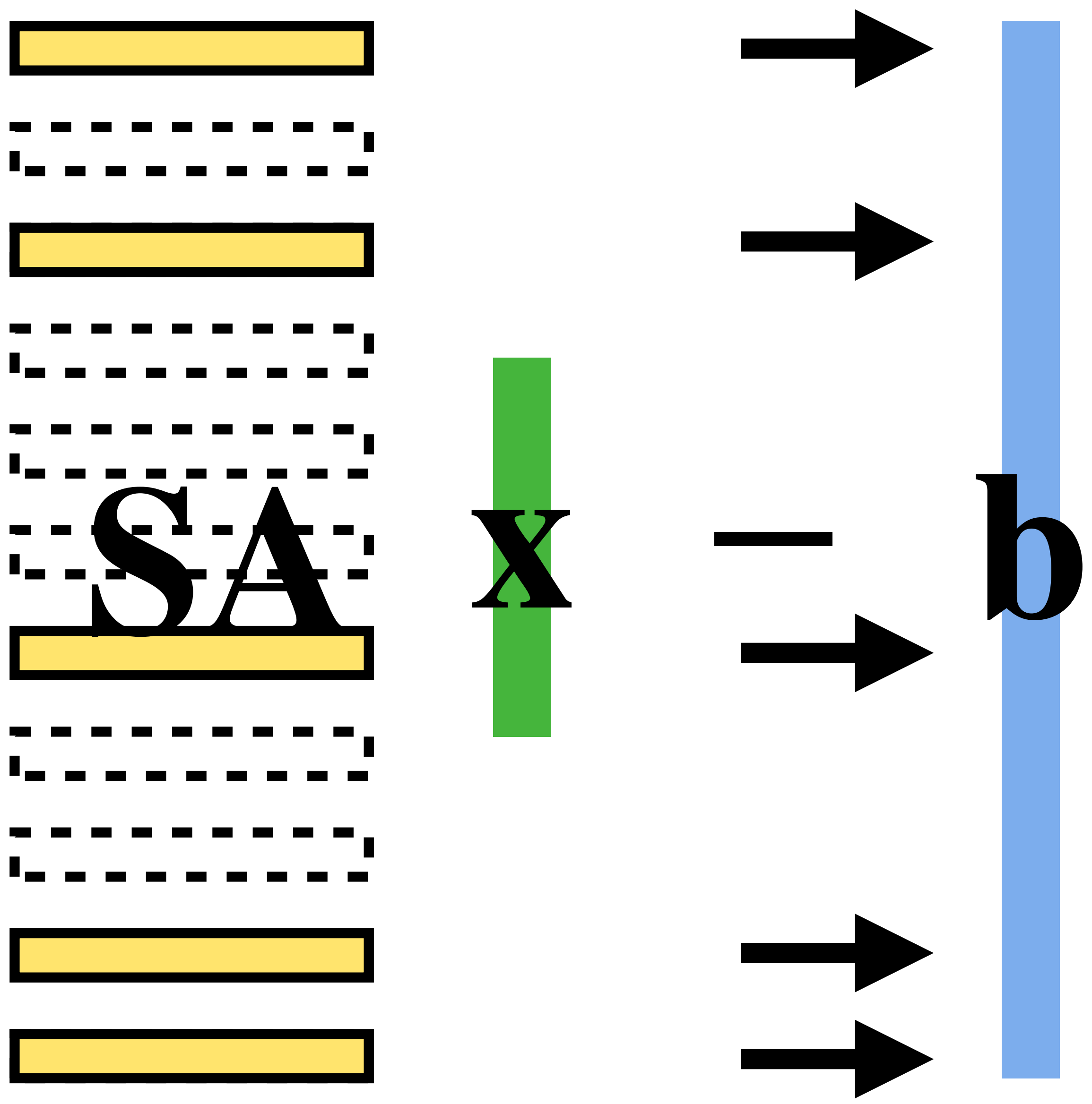


Question. How many entries of $\mathbf{b}$ need to be read?

Matrix Approximation

Active ℓ_p Linear Regression

$$\min_{\mathbf{x}\in\mathbb{R}^d} \| \mathbf{A}\mathbf{x} - \mathbf{b} \|_p^p$$



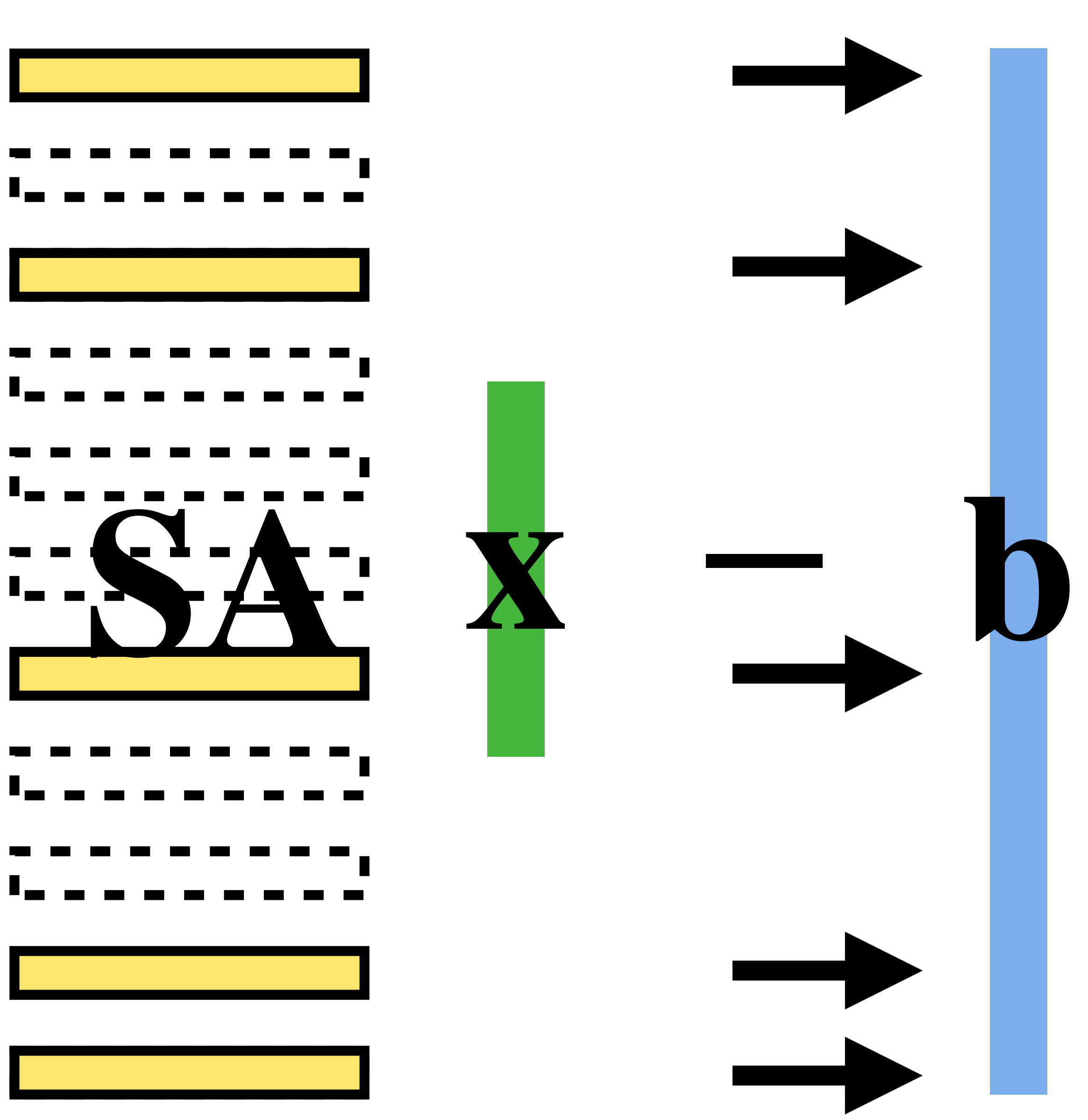
- $p = 2$: $\Theta(\varepsilon^{-1}d)$ (Chen—Price 2019)

Question. How many entries of $\mathbf{b}$ need to be read?

Matrix Approximation

Active ℓ_p Linear Regression

$$\min_{\mathbf{x}\in\mathbb{R}^d} \| \mathbf{Ax} - \mathbf{b} \|_p^p$$



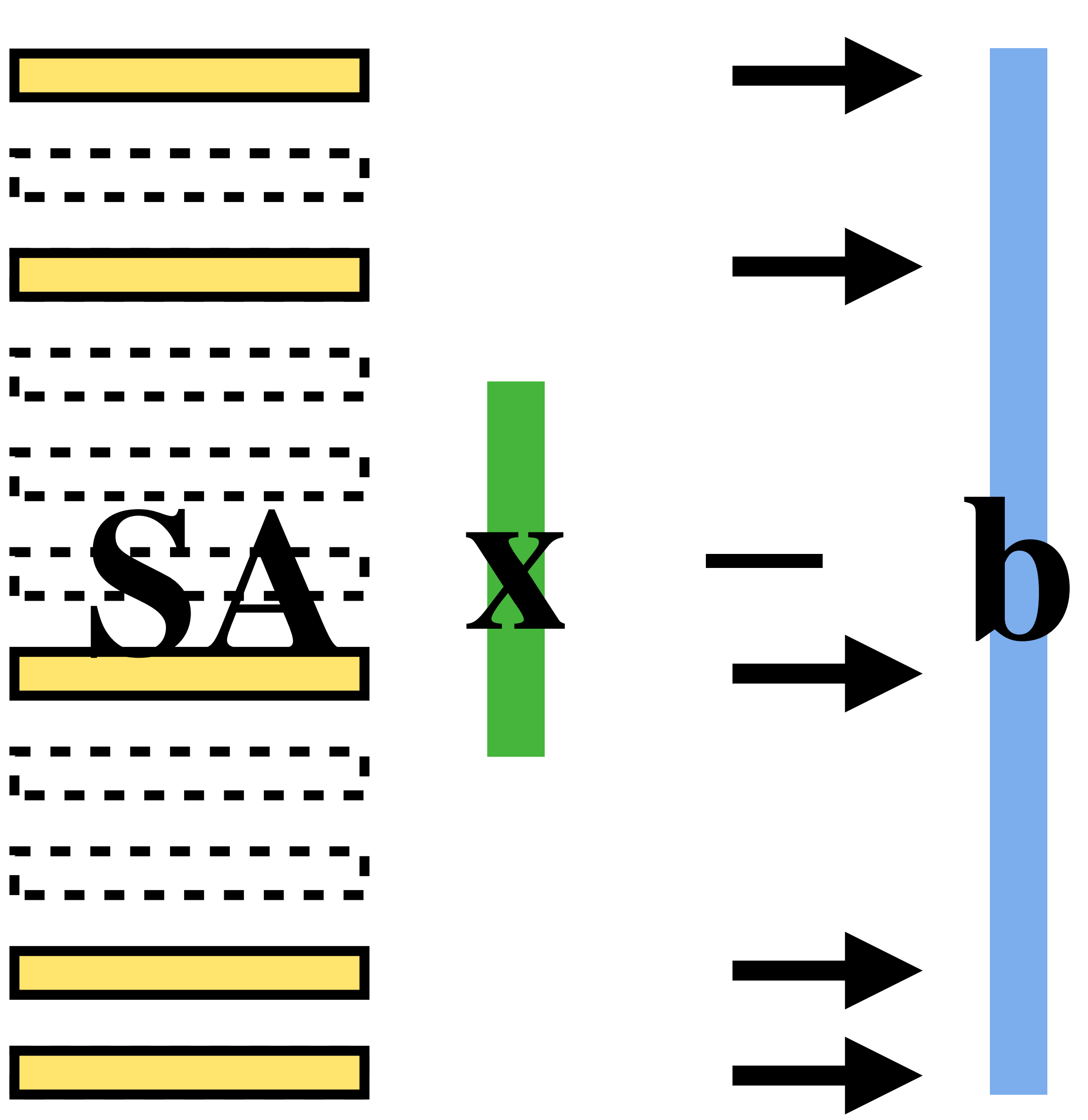
- $p = 2$: $\Theta(\varepsilon^{-1}d)$ (Chen—Price 2019)
- $p = 1$: $\tilde{\Theta}(\varepsilon^{-2}d)$ (Chen—Derezinski, Parulekar—Parulekar—Price 2021)

Question. How many entries of $\mathbf{b}$ need to be read?

Matrix Approximation

Active ℓ_p Linear Regression

$$\min_{\mathbf{x}\in\mathbb{R}^d} \|\mathbf{A}\mathbf{x} - \mathbf{b}\|_p^p$$



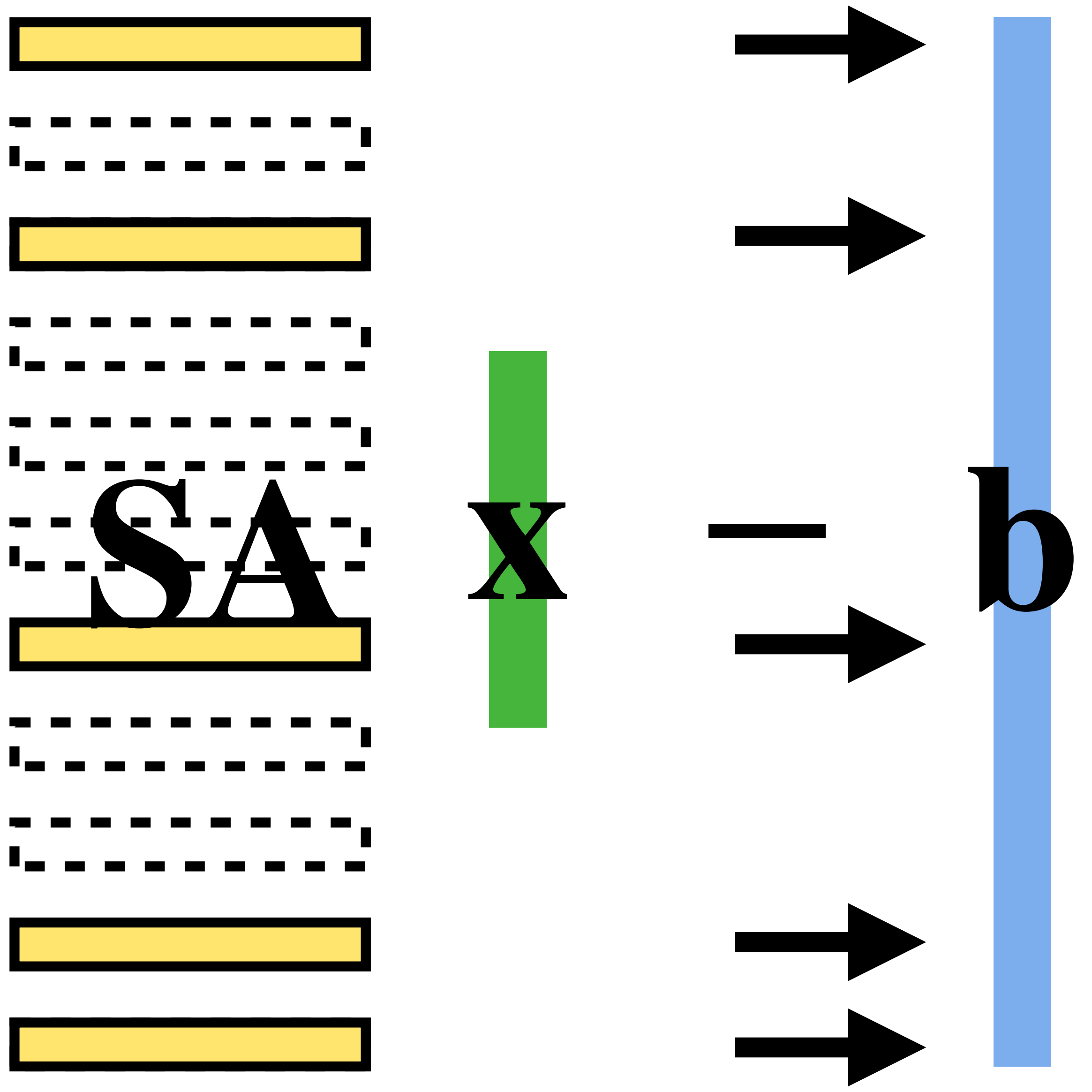
- $p = 2$: $\Theta(\varepsilon^{-1}d)$ (Chen—Price 2019)
- $p = 1$: $\tilde{\Theta}(\varepsilon^{-2}d)$ (Chen—Derezinski, Parulekar—Parulekar—Price 2021)
- $1 < p < 2$: $\tilde{O}(\varepsilon^{-2}d^2)$ (Chen—Derezinski 2021)

Question. How many entries of $\mathbf{b}$ need to be read?

Matrix Approximation

Active ℓ_p Linear Regression

$$\min_{\mathbf{x}\in\mathbb{R}^d} \| \mathbf{Ax} - \mathbf{b} \|_p^p$$



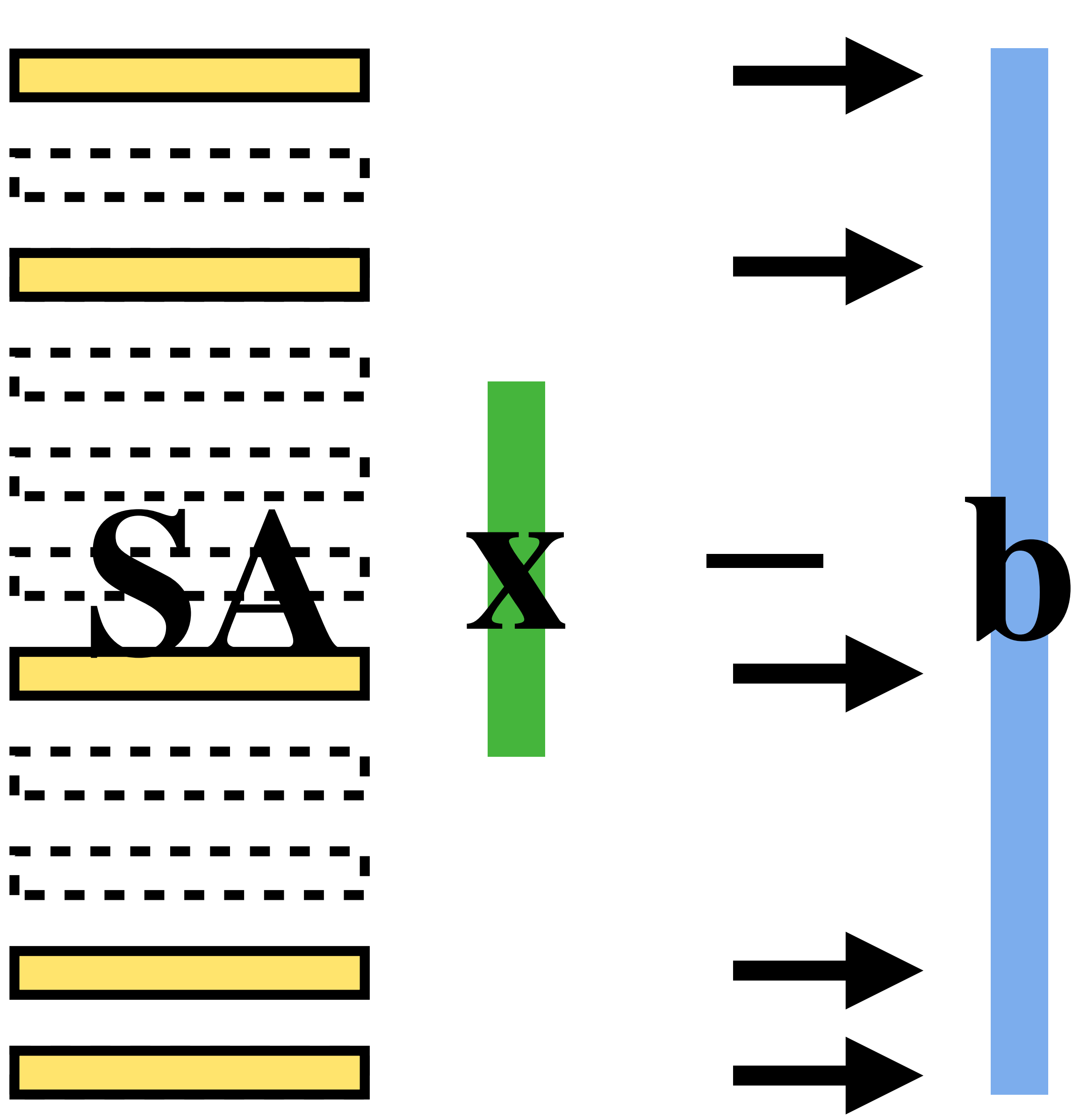
- $p = 2$: $\Theta(\varepsilon^{-1}d)$ (Chen—Price 2019)
- $p = 1$: $\tilde{\Theta}(\varepsilon^{-2}d)$ (Chen—Derezinski, Parulekar—Parulekar—Price 2021)
- $1 < p < 2$: $\tilde{O}(\varepsilon^{\cancel{-2}}d^2)$ (Chen—Derezinski 2021)

Question. How many entries of $\mathbf{b}$ need to be read?

Matrix Approximation

Active ℓ_p Linear Regression

$$\min_{\mathbf{x}\in\mathbb{R}^d} \| \mathbf{A}\mathbf{x} - \mathbf{b} \|_p^p$$



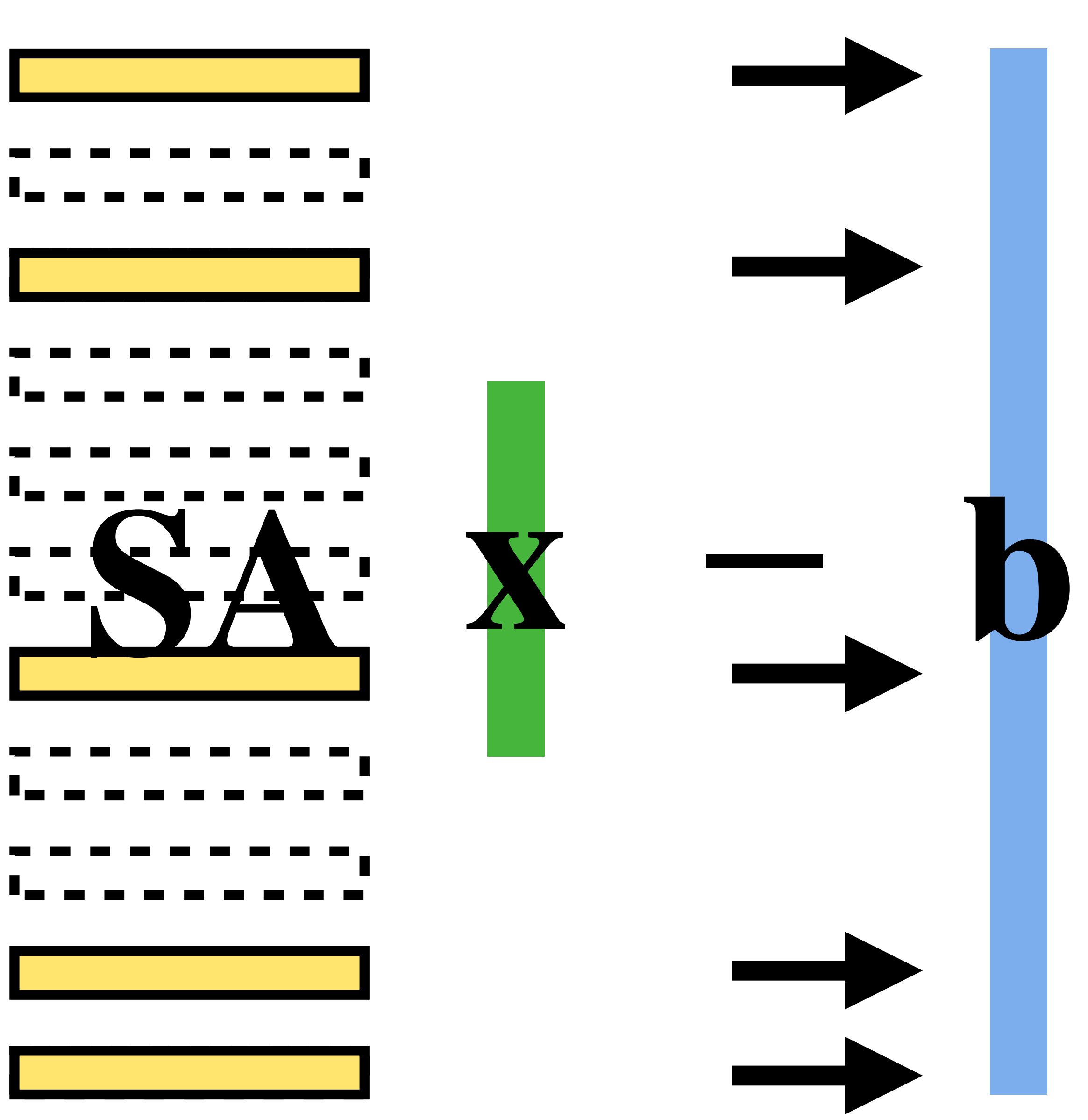
- $p = 2$: $\Theta(\varepsilon^{-1}d)$ (Chen—Price 2019)
- $p = 1$: $\tilde{\Theta}(\varepsilon^{-2}d)$ (Chen—Derezinski, Parulekar—Parulekar—Price 2021)
- $1 < p < 2$: $\tilde{O}(\cancel{\varepsilon^{-2}d^2})$ (Chen—Derezinski 2021)
  $\tilde{\Theta}(\varepsilon^{-1}d)$ (Musco—Musco—Woodruff—Y 2022)

Question. How many entries of $\mathbf{b}$ need to be read?

Matrix Approximation

Active ℓ_p Linear Regression

$$\min_{\mathbf{x}\in\mathbb{R}^d} \| \mathbf{Ax} - \mathbf{b} \|_p^p$$



- $p = 2$: $\Theta(\varepsilon^{-1}d)$ (Chen—Price 2019)
- $p = 1$: $\tilde{\Theta}(\varepsilon^{-2}d)$ (Chen—Derezinski, Parulekar—Parulekar—Price 2021)
- $1 < p < 2$: $\tilde{O}(\cancel{\varepsilon^{-2}d^2})$ (Chen—Derezinski 2021)
  $\tilde{\Theta}(\varepsilon^{-1}d)$ (Musco—Musco—Woodruff—Y 2022)
- $2 < p < \infty$: $\tilde{\Theta}(\varepsilon^{1-p}d^{p/2})$ (Woodruff—Y 2023)

Question. How many entries of $\mathbf{b}$ need to be read?

Matrix Approximation

Streaming Löwner—John Ellipsoids

Matrix Approximation

Streaming Löwner—John Ellipsoids

Input: symmetric polytope with
$2n$ faces in d dims

Matrix Approximation

Streaming Löwner—John Ellipsoids

Input: symmetric polytope with
$2n$ faces in d dims

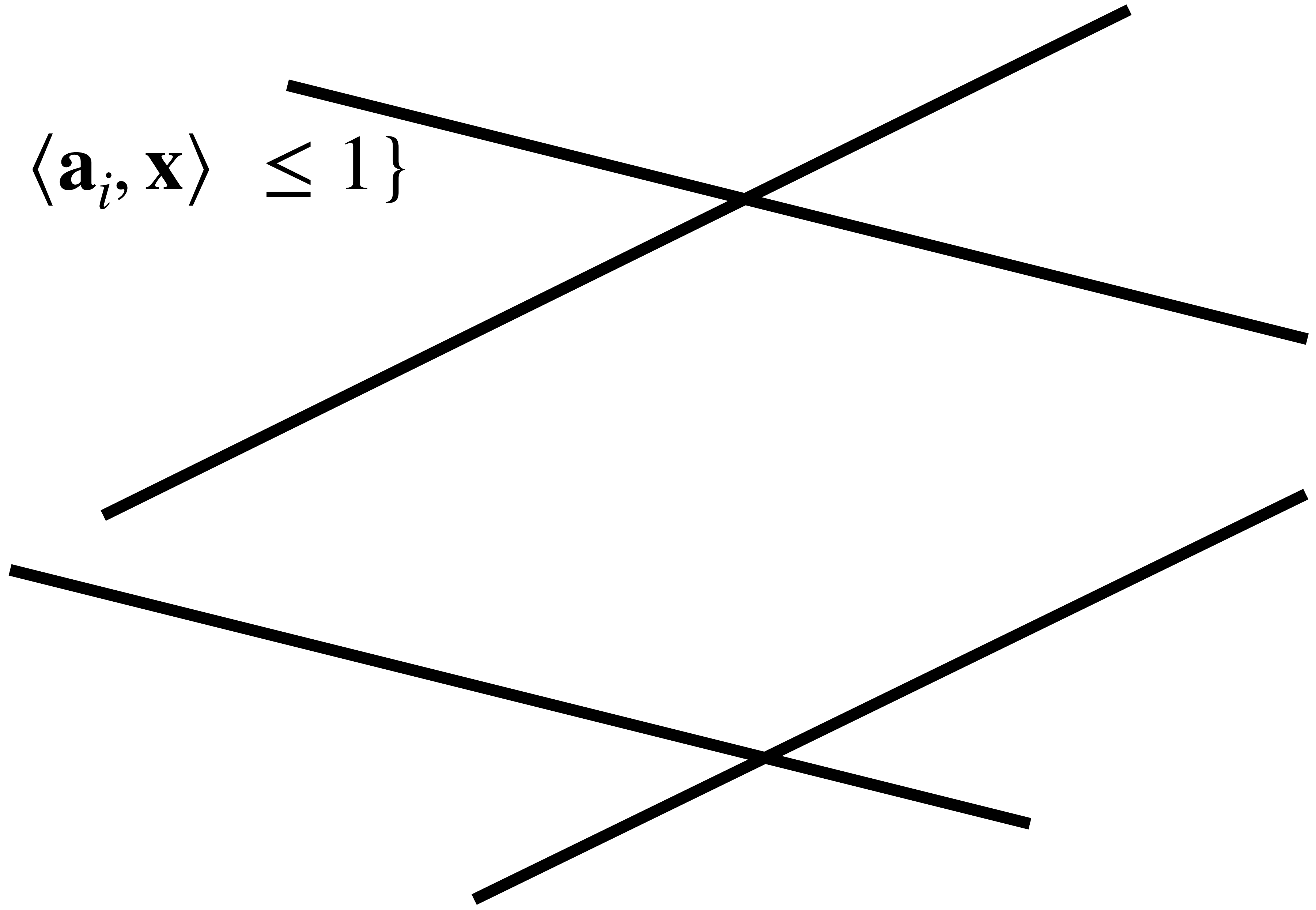
$\{\mathbf{x} : \langle \mathbf{a}_i, \mathbf{x} \rangle \leq 1\}$

Matrix Approximation

Streaming Löwner—John Ellipsoids

Input: symmetric polytope with
$2n$ faces in d dims

$\{\mathbf{x} : \langle \mathbf{a}_i, \mathbf{x} \rangle \leq 1\}$

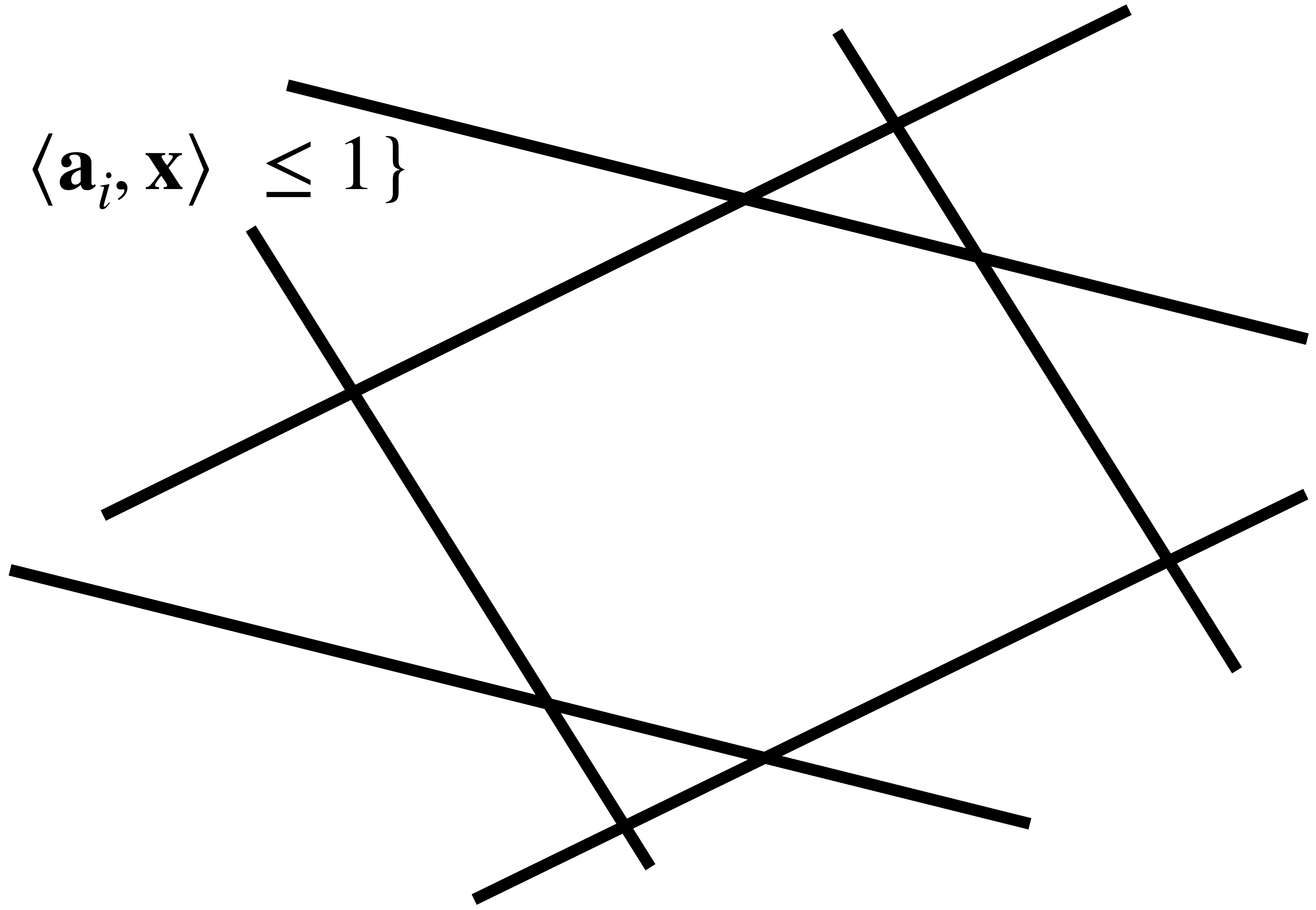


Matrix Approximation

Streaming Löwner—John Ellipsoids

Input: symmetric polytope with
$2n$ faces in d dims

$\{\mathbf{x} : \langle \mathbf{a}_i, \mathbf{x} \rangle \leq 1\}$

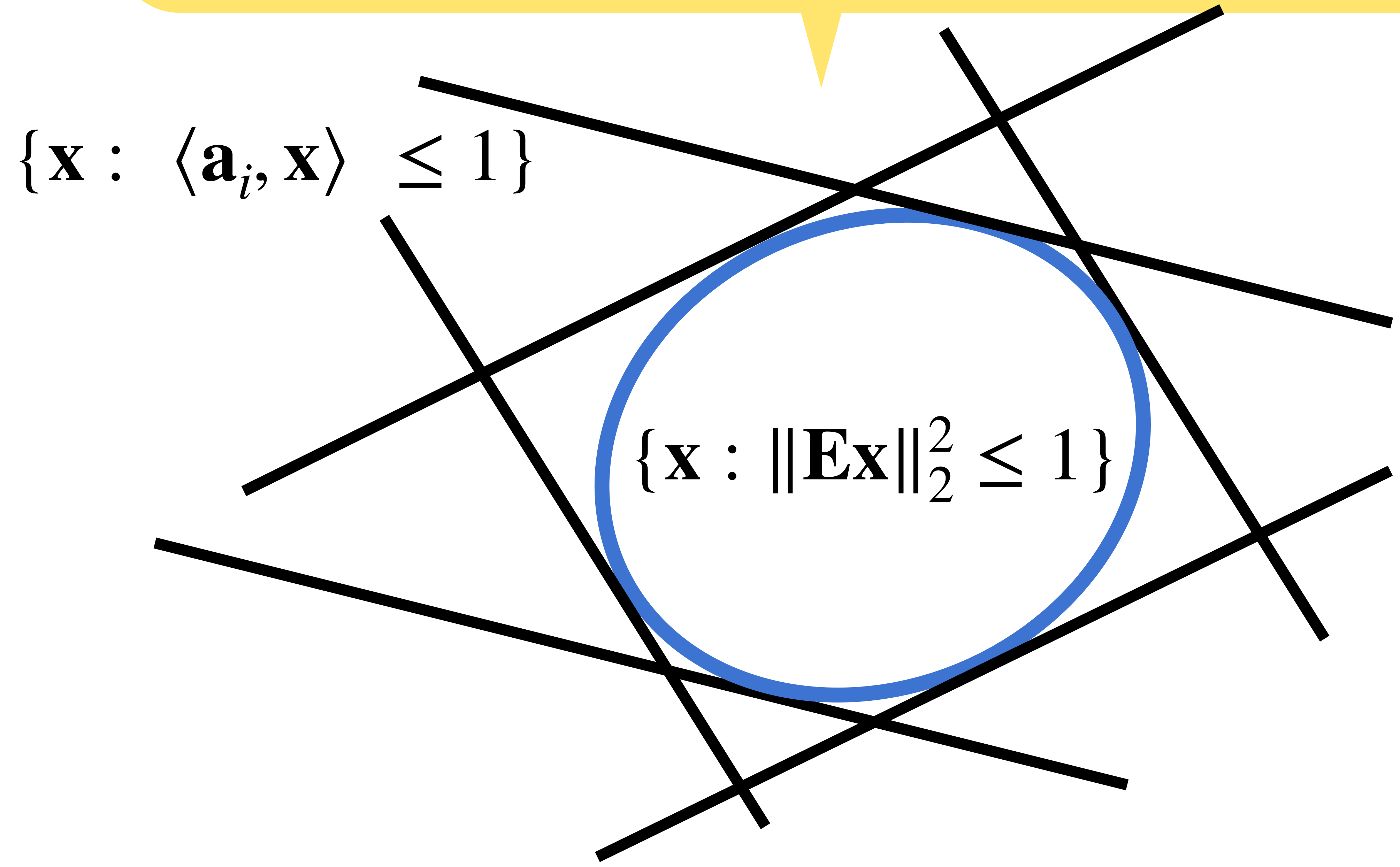


Matrix Approximation

Streaming Löwner—John Ellipsoids

Input: symmetric polytope with
$2n$ faces in d dims

Want the "largest" ellipsoid enclosed in polytope

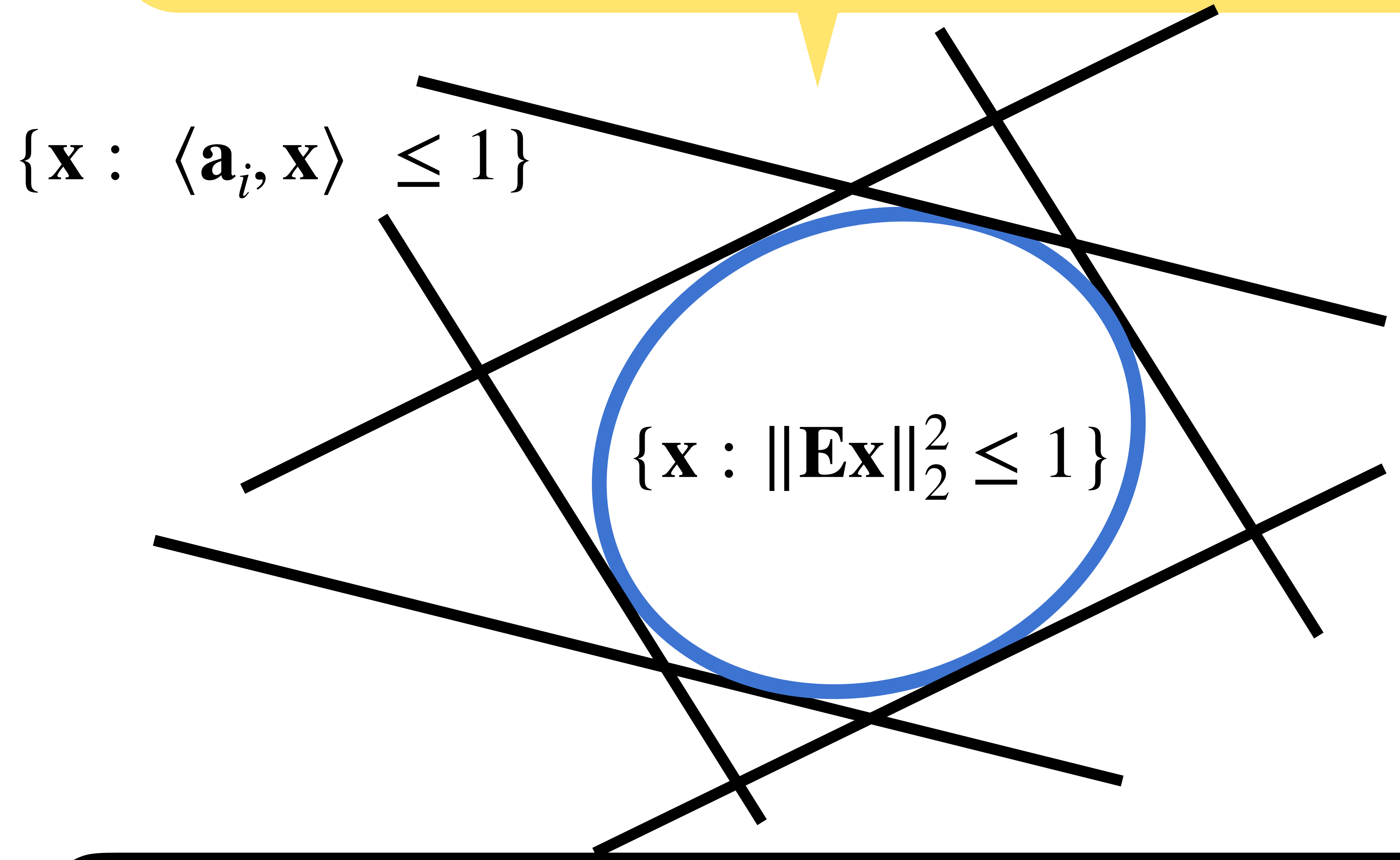


Matrix Approximation

Streaming Löwner—John Ellipsoids

Input: symmetric polytope with
$2n$ faces in d dims

Want the "largest" ellipsoid enclosed in polytope



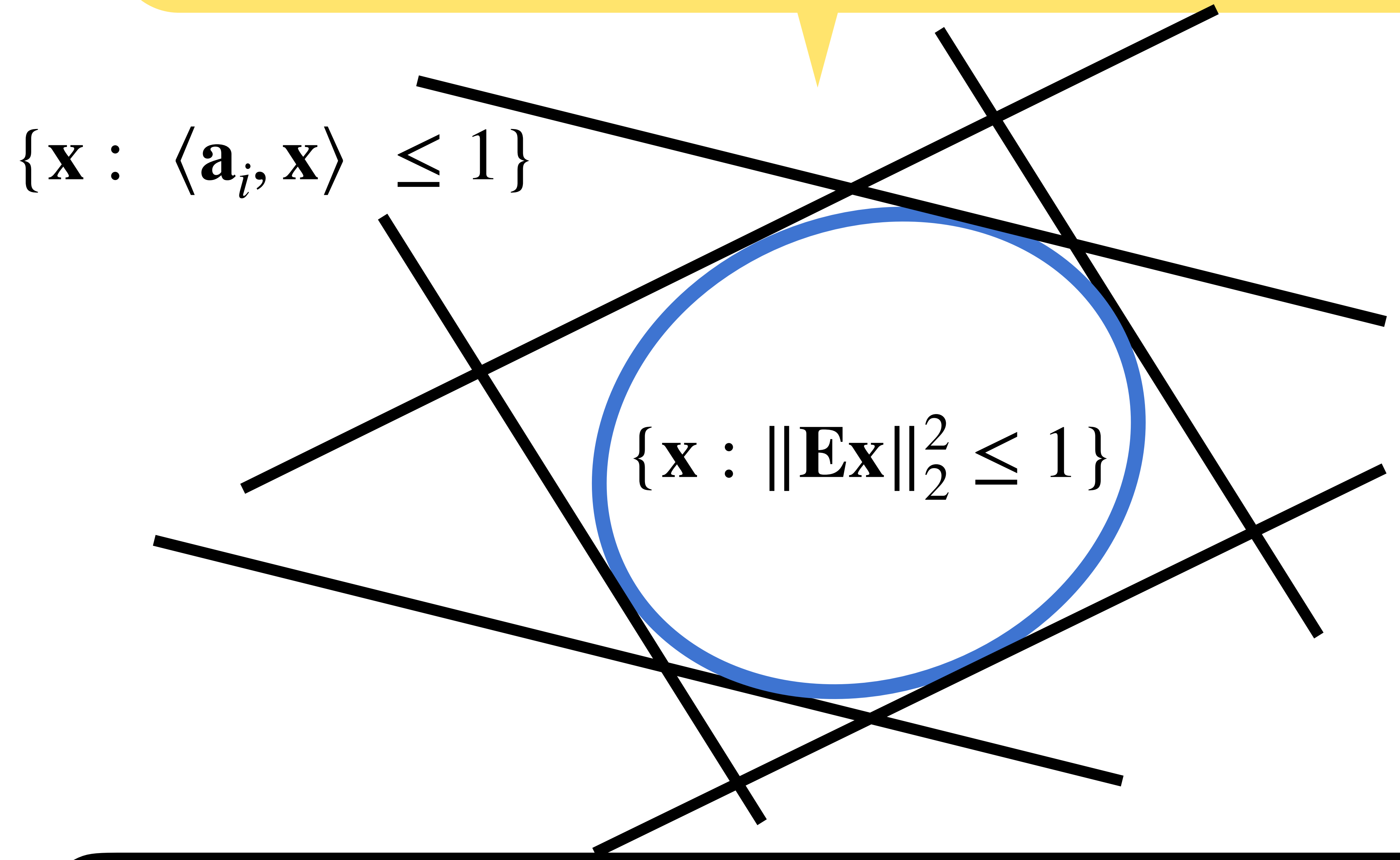
Question. Can Löwner—John ellipsoids be maintained in $\mathrm{poly}(d, \log n)$ bits of space?

Matrix Approximation

Streaming Löwner—John Ellipsoids

Input: symmetric polytope with
$2n$ faces in d dims

Want the "largest" ellipsoid enclosed in polytope



- Agarwal—Har-Peled—Varadarajan 2004: $\exp(\Theta(d))$ bits

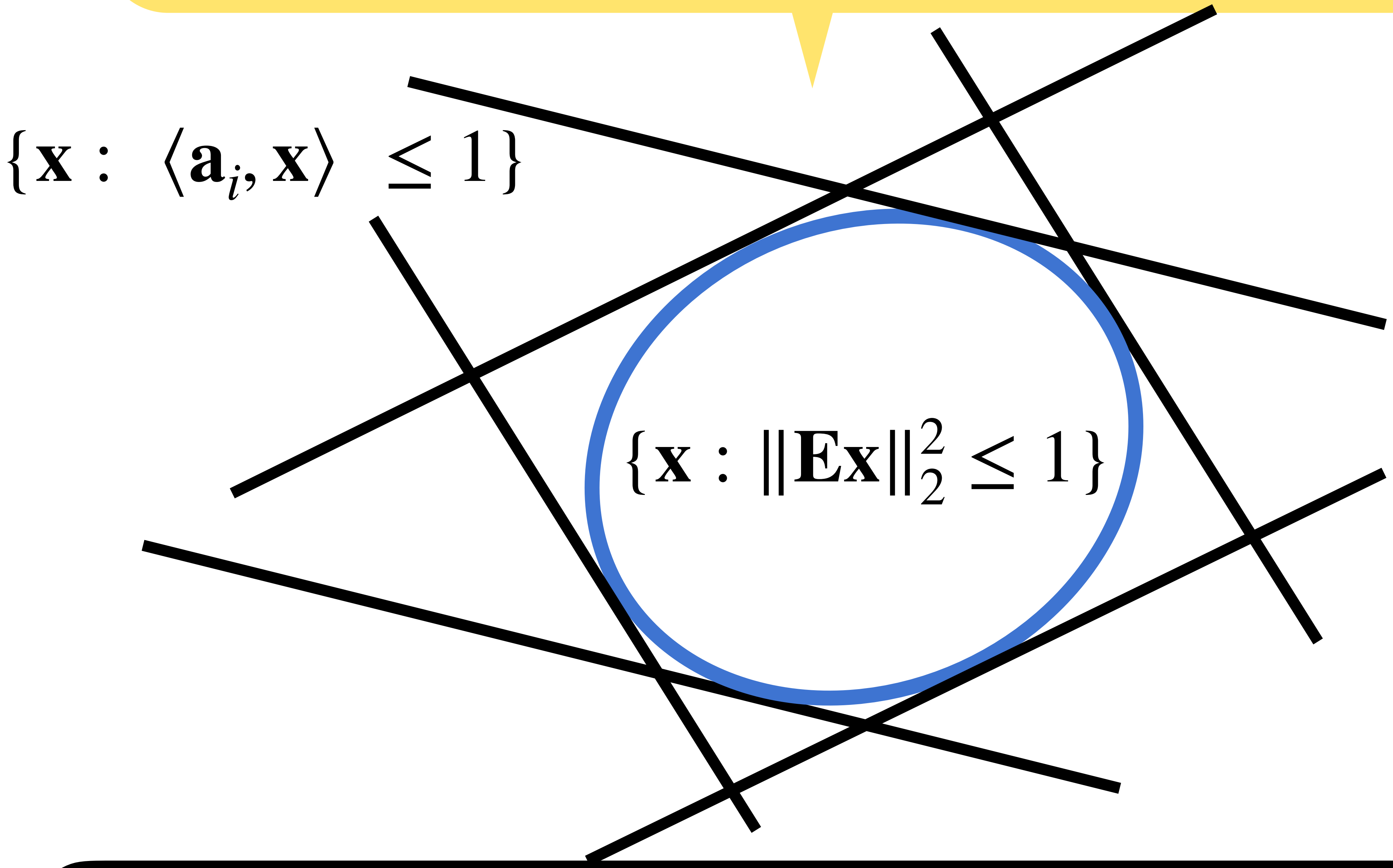
Question. Can Löwner—John ellipsoids be maintained in $\mathrm{poly}(d, \log n)$ bits of space?

Matrix Approximation

Streaming Löwner—John Ellipsoids

Input: symmetric polytope with $2n$ faces in d dims

Want the "largest" ellipsoid enclosed in polytope



Question. Can Löwner—John ellipsoids be maintained in $\mathrm{poly}(d, \log n)$ bits of space?

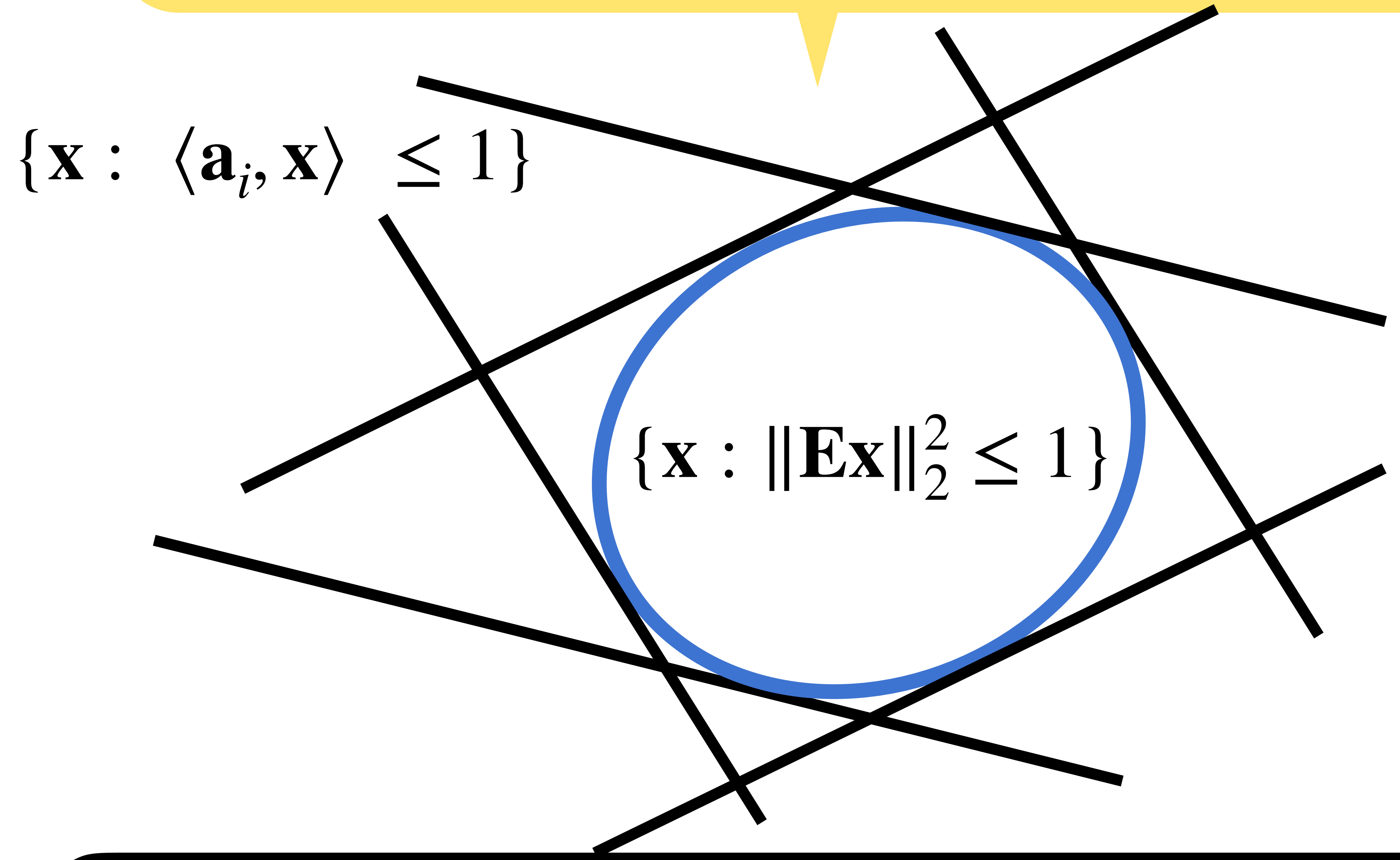
- Agarwal—Har-Peled—Varadarajan 2004: $\exp(\Theta(d))$ bits
- Woodruff—Y 2022: $O(d^2 \log^2 n)$ bits

Matrix Approximation

Streaming Löwner—John Ellipsoids

Input: symmetric polytope with $2n$ faces in d dims

Want the “largest” ellipsoid enclosed in polytope



- Agarwal—Har-Peled—Varadarajan 2004: $\exp(\Theta(d))$ bits
- Woodruff—Y 2022: $O(d^2 \log^2 n)$ bits
- Approach: select a subset of constraints

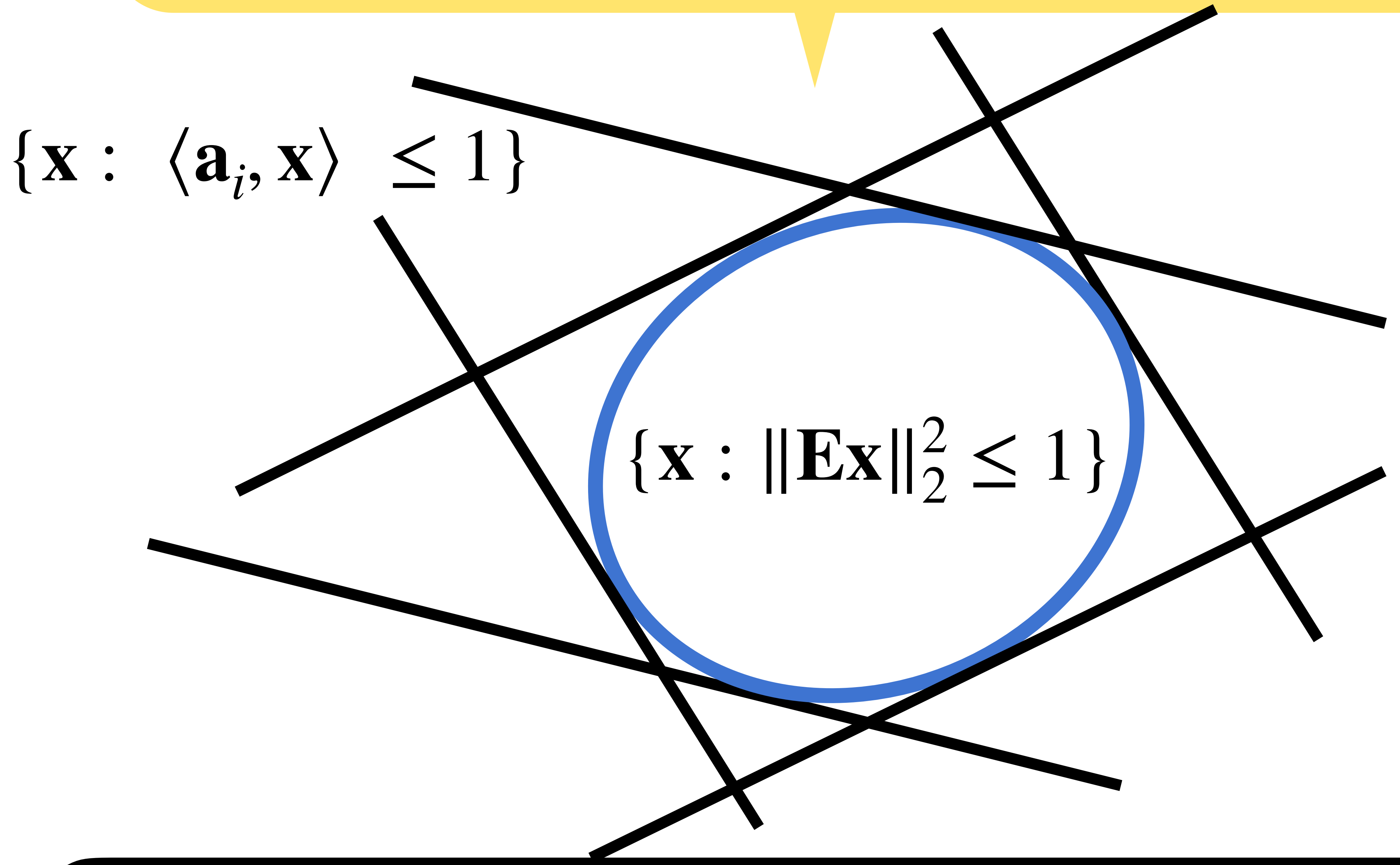
Question. Can Löwner—John ellipsoids be maintained in $\mathrm{poly}(d, \log n)$ bits of space?

Matrix Approximation

Streaming Löwner—John Ellipsoids

Input: symmetric polytope with $2n$ faces in d dims

Want the "largest" ellipsoid enclosed in polytope



- Agarwal—Har-Peled—Varadarajan 2004: $\exp(\Theta(d))$ bits
- Woodruff—Y 2022: $O(d^2 \log^2 n)$ bits
- Approach: select a subset of constraints
- **Techniques:** use leverage scores to…

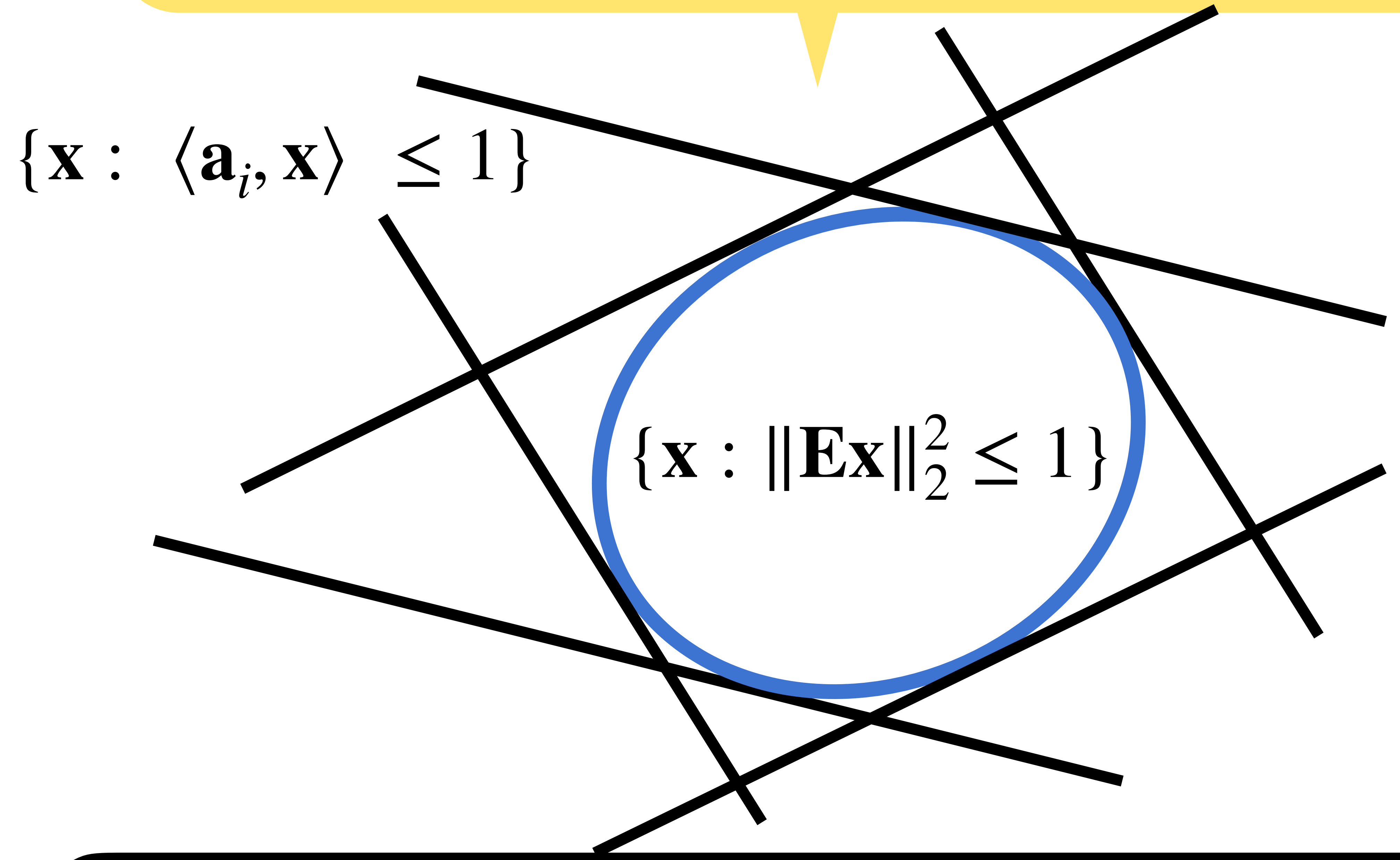
Question. Can Löwner—John ellipsoids be maintained in $\mathrm{poly}(d, \log n)$ bits of space?

Matrix Approximation

Streaming Löwner—John Ellipsoids

Input: symmetric polytope with $2n$ faces in d dims

Want the "largest" ellipsoid enclosed in polytope



- Agarwal—Har-Peled—Varadarajan 2004: $\exp(\Theta(d))$ bits
- Woodruff—Y 2022: $O(d^2 \log^2 n)$ bits
- Approach: select a subset of constraints
- **Techniques:** use leverage scores to...
  - test if ellipsoid respects a new constraint $\mathbf{a}_i$

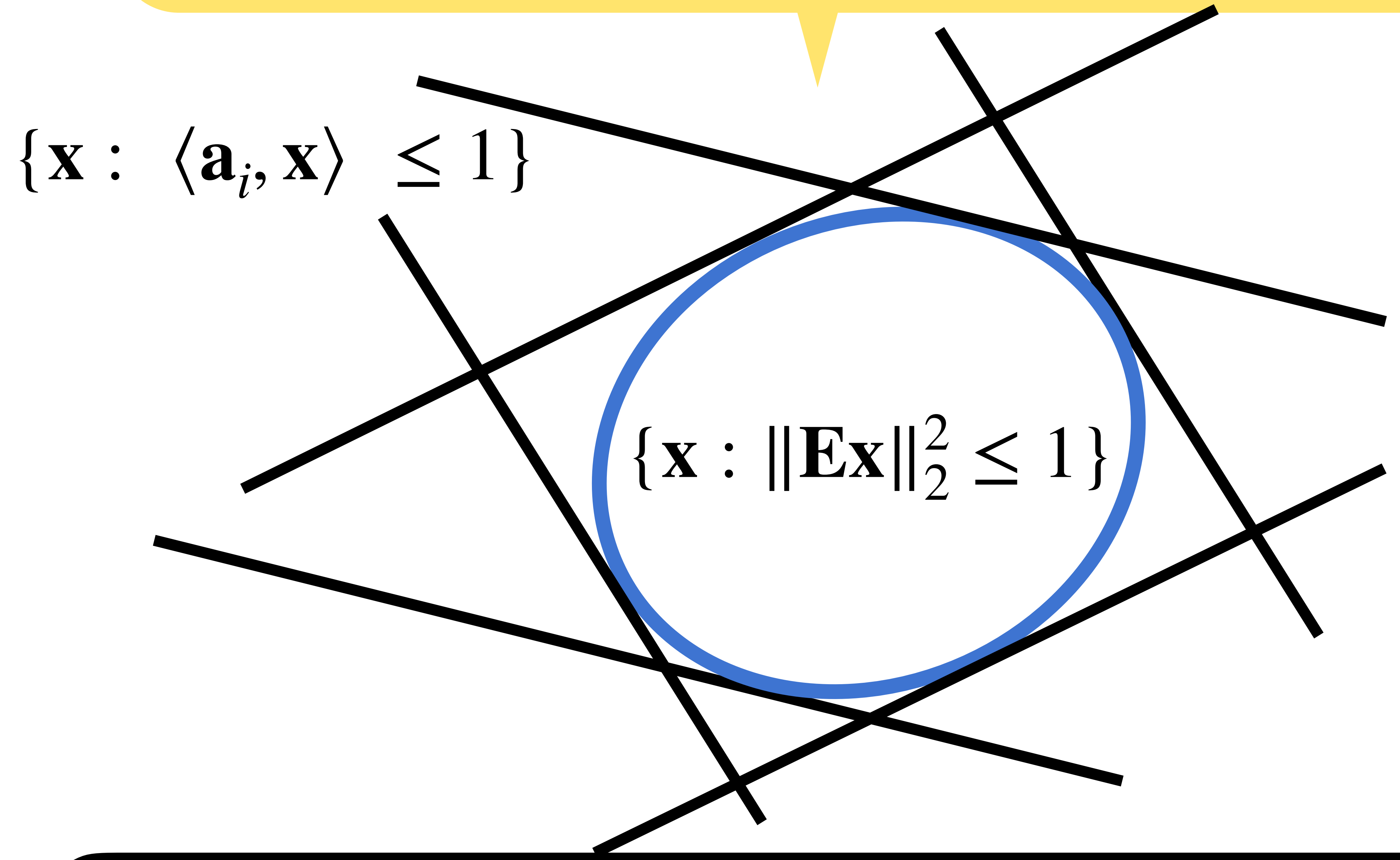
Question. Can Löwner—John ellipsoids be maintained in $\mathrm{poly}(d, \log n)$ bits of space?

Matrix Approximation

Streaming Löwner—John Ellipsoids

Input: symmetric polytope with $2n$ faces in d dims

Want the "largest" ellipsoid enclosed in polytope



- Agarwal—Har-Peled—Varadarajan 2004: $\exp(\Theta(d))$ bits
- Woodruff—Y 2022: $O(d^2 \log^2 n)$ bits

- Approach: select a subset of constraints
- **Techniques:** use leverage scores to…
  - test if ellipsoid respects a new constraint $\mathbf{a}_i$
  - bound # of times we keep a new constraint $\mathbf{a}_i$

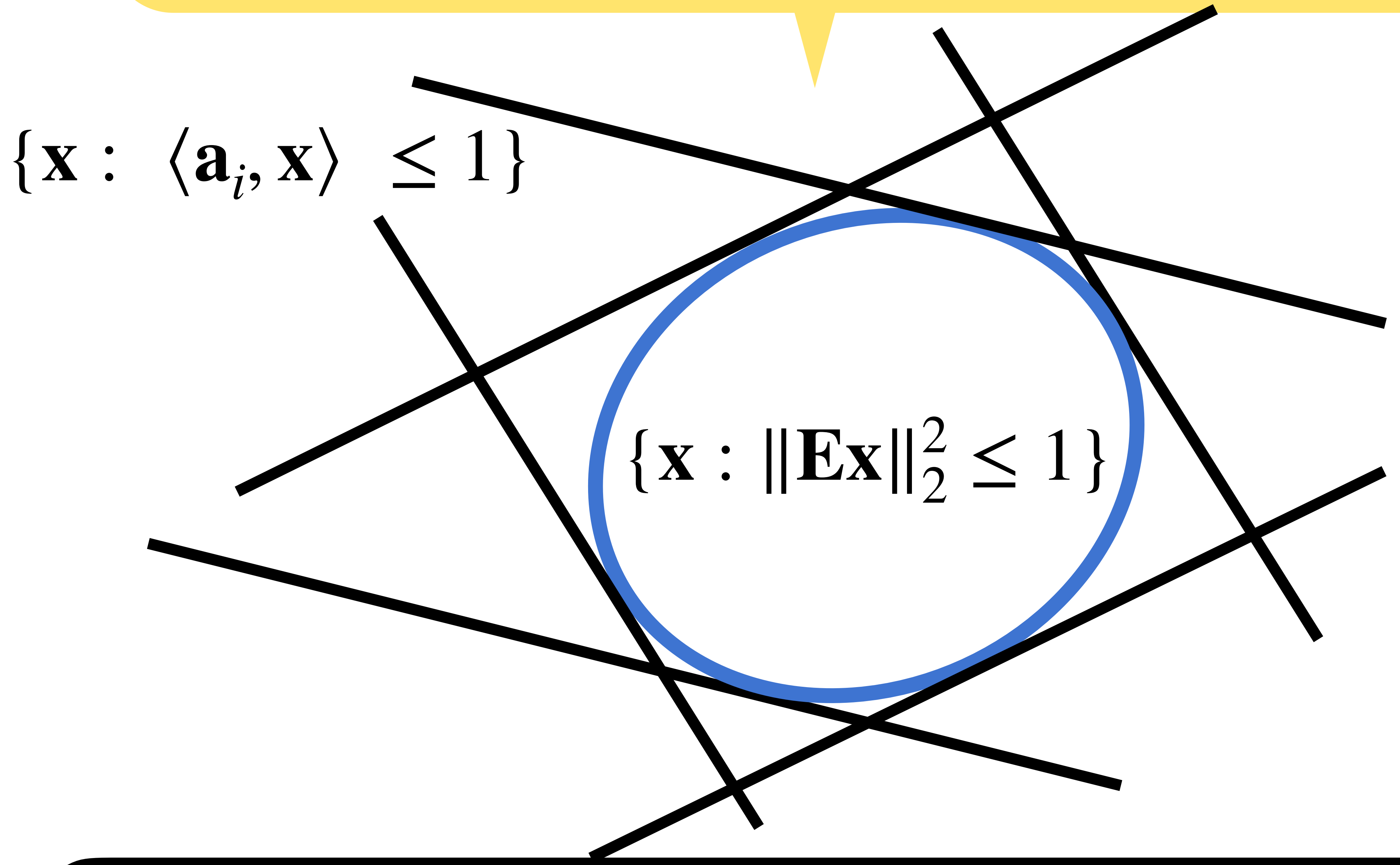
Question. Can Löwner—John ellipsoids be maintained in $\mathrm{poly}(d, \log n)$ bits of space?

Matrix Approximation

Streaming Löwner—John Ellipsoids

Input: symmetric polytope with
$2n$ faces in d dims

Want the "largest" ellipsoid enclosed in polytope



- Agarwal—Har-Peled—Varadarajan 2004: $\exp(\Theta(d))$ bits
- Woodruff—Y 2022: $O(d^2 \log^2 n)$ bits
- Approach: select a subset of constraints
- **Techniques:** use leverage scores to...
  - test if ellipsoid respects a new constraint $\mathbf{a}_i$
  - bound # of times we keep a new constraint $\mathbf{a}_i$

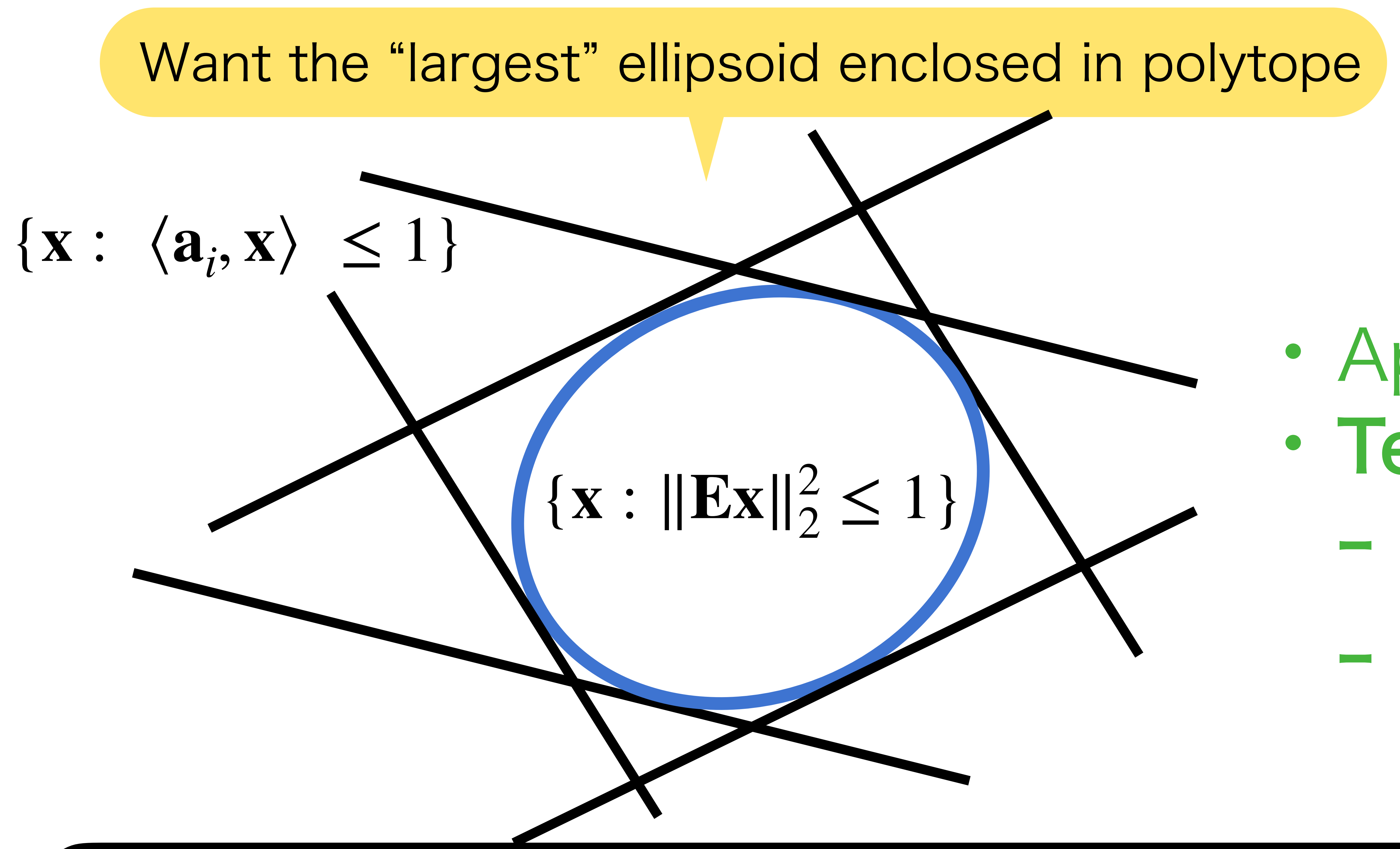
$$\sup_{\mathbf{E}\mathbf{x} \neq 0} \frac{\langle \mathbf{a}_i, \mathbf{x} \rangle^2}{\|\mathbf{E}\mathbf{x}\|_2^2}$$

Question. Can Löwner—John ellipsoids be maintained in $\mathrm{poly}(d, \log n)$ bits of space?

Matrix Approximation

Streaming Löwner—John Ellipsoids

Input: symmetric polytope with
$2n$ faces in d dims



- Agarwal—Har-Peled—Varadarajan 2004: $\exp(\Theta(d))$ bits
- Woodruff—Y 2022: $O(d^2 \log^2 n)$ bits

- Approach: select a subset of constraints
- **Techniques:** use leverage scores to...
  - test if ellipsoid respects a new constraint $\mathbf{a}_i$
  - bound # of times we keep a new constraint $\mathbf{a}_i$

≤ 1 iff ellipsoid $\mathbf{E}$ respects constraint $\mathbf{a}_i$

$$\sup_{\mathbf{E}\mathbf{x} \neq 0} \frac{\langle \mathbf{a}_i, \mathbf{x} \rangle^2}{\|\mathbf{E}\mathbf{x}\|_2^2}$$

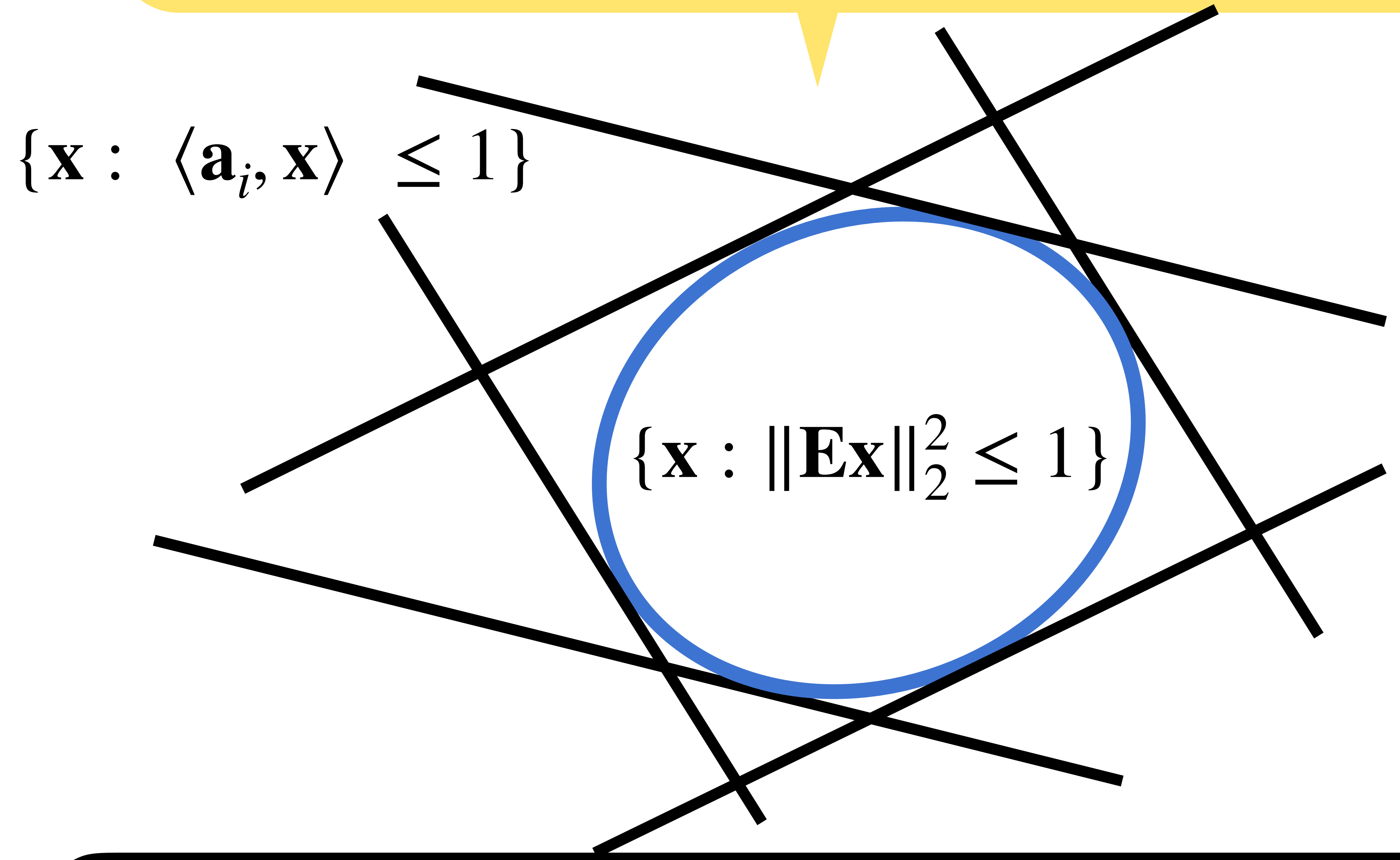
Question. Can Löwner—John ellipsoids be maintained in $\mathrm{poly}(d, \log n)$ bits of space?

Matrix Approximation

Streaming Löwner—John Ellipsoids

Input: symmetric polytope with $2n$ faces in d dims

Want the "largest" ellipsoid enclosed in polytope



Question. Can Löwner—John ellipsoids be maintained in $\mathrm{poly}(d, \log n)$ bits of space?

- Agarwal—Har-Peled—Varadarajan 2004: $\exp(\Theta(d))$ bits
- Woodruff—Y 2022: $O(d^2 \log^2 n)$ bits

- Approach: select a subset of constraints
- **Techniques:** use leverage scores to…
  - test if ellipsoid respects a new constraint $\mathbf{a}_i$
  - bound # of times we keep a new constraint $\mathbf{a}_i$

≤ 1 iff ellipsoid $\mathbf{E}$ respects constraint $\mathbf{a}_i$

$$\sup_{\mathbf{E}\mathbf{x} \neq 0} \frac{\langle \mathbf{a}_i, \mathbf{x} \rangle^2}{\|\mathbf{E}\mathbf{x}\|_2^2}$$

Leverage scores: $\tau_i(\mathbf{A}) = \sup_{\mathbf{A}\mathbf{x} \neq 0} \frac{\langle \mathbf{a}_i, \mathbf{x} \rangle^2}{\|\mathbf{A}\mathbf{x}\|_2^2}$

Matrix Approximation

Low Rank Approximation with General Losses

Matrix Approximation

Low Rank Approximation with General Losses

$$\min_{\mathbf{U}\in\mathbb{R}^{n\times k},\mathbf{V}\in\mathbb{R}^{k\times d}} \| \mathbf{A} - \mathbf{UV} \|_F^2$$

Matrix Approximation

Low Rank Approximation with General Losses

$$\min_{\mathbf{U}\in\mathbb{R}^{n\times k},\mathbf{V}\in\mathbb{R}^{k\times d}} \|\mathbf{A}-\mathbf{UV}\|_F^2$$

$$\min_{\mathbf{U}\in\mathbb{R}^{n\times k},\mathbf{V}\in\mathbb{R}^{k\times d}} \|\mathbf{A}-\mathbf{UV}\|_{p,p}^p = \sum_{i,j} (\mathbf{A}-\mathbf{UV})_{i,j}{}^p$$

Matrix Approximation

Low Rank Approximation with General Losses

$$\min_{\mathbf{U}\in\mathbb{R}^{n\times k},\mathbf{V}\in\mathbb{R}^{k\times d}} \| \mathbf{A} - \mathbf{U}\mathbf{V} \|_F^2$$

$$\min_{\mathbf{U}\in\mathbb{R}^{n\times k},\mathbf{V}\in\mathbb{R}^{k\times d}} \| \mathbf{A} - \mathbf{U}\mathbf{V} \|_{p,p}^p = \sum_{i,j} (\mathbf{A} - \mathbf{U}\mathbf{V})_{i,j}^{\ p}$$

$$\min_{\mathbf{U}\in\mathbb{R}^{n\times k},\mathbf{V}\in\mathbb{R}^{k\times d}} \| \mathbf{A} - \mathbf{U}\mathbf{V} \|_g = \sum_{i,j} g(\ (\mathbf{A} - \mathbf{U}\mathbf{V})_{i,j}\ )$$

Matrix Approximation

Low Rank Approximation with General Losses

$$\min_{\mathbf{U}\in\mathbb{R}^{n\times k},\mathbf{V}\in\mathbb{R}^{k\times d}} \| \mathbf{A} - \mathbf{UV} \|_F^2$$

$$\min_{\mathbf{U}\in\mathbb{R}^{n\times k},\mathbf{V}\in\mathbb{R}^{k\times d}} \| \mathbf{A} - \mathbf{UV} \|_{p,p}^p = \sum_{i,j} (\mathbf{A} - \mathbf{UV})_{i,j}^{\ p}$$

$$\min_{\mathbf{U}\in\mathbb{R}^{n\times k},\mathbf{V}\in\mathbb{R}^{k\times d}} \| \mathbf{A} - \mathbf{UV} \|_g = \sum_{i,j} g((\mathbf{A} - \mathbf{UV})_{i,j})$$

- NP-hard… → need approximation/bicriteria algorithms

Matrix Approximation

Low Rank Approximation with General Losses

$$\min_{\mathbf{U}\in\mathbb{R}^{n\times k},\mathbf{V}\in\mathbb{R}^{k\times d}} \| \mathbf{A} - \mathbf{U}\mathbf{V} \|_F^2$$

$$\min_{\mathbf{U}\in\mathbb{R}^{n\times k},\mathbf{V}\in\mathbb{R}^{k\times d}} \| \mathbf{A} - \mathbf{U}\mathbf{V} \|_{p,p}^p = \sum_{i,j} (\mathbf{A} - \mathbf{U}\mathbf{V})_{i,j}^{\ p}$$

$$\min_{\mathbf{U}\in\mathbb{R}^{n\times k},\mathbf{V}\in\mathbb{R}^{k\times d}} \| \mathbf{A} - \mathbf{U}\mathbf{V} \|_g = \sum_{i,j} g((\mathbf{A} - \mathbf{U}\mathbf{V})_{i,j})$$

- NP-hard… → need approximation/bicriteria algorithms
- Focus on **column subset selection** algorithms: $\mathbf{U}$ is a set of $\approx k$ columns of $\mathbf{A}$

Matrix Approximation

Low Rank Approximation with General Losses

$$\min_{\mathbf{U}\in\mathbb{R}^{n\times k},\mathbf{V}\in\mathbb{R}^{k\times d}} \| \mathbf{A} - \mathbf{UV} \|_F^2$$

$$\rightarrow \quad \min_{\mathbf{U}\in\mathbb{R}^{n\times k},\mathbf{V}\in\mathbb{R}^{k\times d}} \| \mathbf{A} - \mathbf{UV} \|_{p,p}^p = \sum_{i,j} (\mathbf{A} - \mathbf{UV})_{i,j}{}^p$$

$$\rightarrow \quad \min_{\mathbf{U}\in\mathbb{R}^{n\times k},\mathbf{V}\in\mathbb{R}^{k\times d}} \| \mathbf{A} - \mathbf{UV} \|_g = \sum_{i,j} g((\mathbf{A} - \mathbf{UV})_{i,j})$$

- NP-hard… → need approximation/bicriteria algorithms
- Focus on **column subset selection** algorithms: $\mathbf{U}$ is a set of $\approx k$ columns of $\mathbf{A}$
  - Useful for unsupervised feature selection

Matrix Approximation

Low Rank Approximation with General Losses

$$\min_{\mathbf{U}\in\mathbb{R}^{n\times k},\mathbf{V}\in\mathbb{R}^{k\times d}} \| \mathbf{A} - \mathbf{UV} \|_F^2$$

$$\min_{\mathbf{U}\in\mathbb{R}^{n\times k},\mathbf{V}\in\mathbb{R}^{k\times d}} \| \mathbf{A} - \mathbf{UV} \|_{p,p}^p = \sum_{i,j} (\mathbf{A} - \mathbf{UV})_{i,j}^{\ p}$$

$$\min_{\mathbf{U}\in\mathbb{R}^{n\times k},\mathbf{V}\in\mathbb{R}^{k\times d}} \| \mathbf{A} - \mathbf{UV} \|_g = \sum_{i,j} g(\, (\mathbf{A} - \mathbf{UV})_{i,j}\,)$$

- NP-hard… → need approximation/bicriteria algorithms
- Focus on **column subset selection** algorithms: $\mathbf{U}$ is a set of $\approx k$ columns of $\mathbf{A}$
  - Useful for unsupervised feature selection
  - This framework gives the best known algorithms for this problem!

Matrix Approximation

Low Rank Approximation with General Losses

$$\min_{\mathbf{U}\in\mathbb{R}^{n\times k},\mathbf{V}\in\mathbb{R}^{k\times d}} \| \mathbf{A} - \mathbf{UV} \|_g = \sum_{i,j} g((\mathbf{A} - \mathbf{UV})_{i,j})$$

Matrix Approximation

Low Rank Approximation with General Losses

$$\min_{\mathbf{U}\in\mathbb{R}^{n\times k},\mathbf{V}\in\mathbb{R}^{k\times d}} \| \mathbf{A} - \mathbf{U}\mathbf{V} \|_g = \sum_{i,j} g(\ (\mathbf{A} - \mathbf{U}\mathbf{V})_{i,j}\ )$$

$$\| \mathbf{A} - \hat{\mathbf{U}}\hat{\mathbf{V}} \|_g \leq \kappa \min_{\mathbf{U}\in\mathbb{R}^{n\times k},\mathbf{V}\in\mathbb{R}^{k\times d}} \| \mathbf{A} - \mathbf{U}\mathbf{V} \|_g$$

Matrix Approximation

Low Rank Approximation with General Losses

$$\min_{\mathbf{U}\in\mathbb{R}^{n\times k},\mathbf{V}\in\mathbb{R}^{k\times d}} \|\mathbf{A}-\mathbf{U}\mathbf{V}\|_g = \sum_{i,j} g(\,(\mathbf{A}-\mathbf{U}\mathbf{V})_{i,j}\,)$$

$$\|\mathbf{A}-\hat{\mathbf{U}}\hat{\mathbf{V}}\|_g \le \kappa \min_{\mathbf{U}\in\mathbb{R}^{n\times k},\mathbf{V}\in\mathbb{R}^{k\times d}} \|\mathbf{A}-\mathbf{U}\mathbf{V}\|_g$$

Subset of $\approx k$ columns of $\mathbf{A}$

Matrix Approximation

Low Rank Approximation with General Losses

$$\min_{\mathbf{U}\in\mathbb{R}^{n\times k},\mathbf{V}\in\mathbb{R}^{k\times d}} \| \mathbf{A} - \mathbf{U}\mathbf{V} \|_g = \sum_{i,j} g(\, (\mathbf{A} - \mathbf{U}\mathbf{V})_{i,j} \,)$$

$$\| \mathbf{A} - \hat{\mathbf{U}}\hat{\mathbf{V}} \|_g \leq \kappa \min_{\mathbf{U}\in\mathbb{R}^{n\times k},\mathbf{V}\in\mathbb{R}^{k\times d}} \| \mathbf{A} - \mathbf{U}\mathbf{V} \|_g$$

Subset of $\approx k$ columns of $\mathbf{A}$

Approximation factor κ

Matrix Approximation

Low Rank Approximation with General Losses

$$\min_{\mathbf{U}\in\mathbb{R}^{n\times k},\mathbf{V}\in\mathbb{R}^{k\times d}} \| \mathbf{A} - \mathbf{UV} \|_g = \sum_{i,j} g(\, (\mathbf{A} - \mathbf{UV})_{i,j}\,)$$

$$\| \mathbf{A} - \hat{\mathbf{U}}\hat{\mathbf{V}} \|_g \leq \kappa \min_{\mathbf{U}\in\mathbb{R}^{n\times k},\mathbf{V}\in\mathbb{R}^{k\times d}} \| \mathbf{A} - \mathbf{UV} \|_g$$

Subset of $\approx k$ columns of $\mathbf{A}$

Approximation factor κ

- ℓ_p, $p < 2$: $\kappa \approx k^{1/p-1/2}$ (Mahankali—Woodruff 2021)
- $\ell_p, p > 2$: $\kappa \approx k^{1-1/p}$ (Dan—Wang—Zhang—Zhou—Ravikumar 2019)

Matrix Approximation

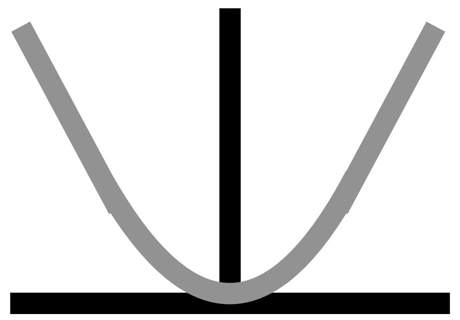
Low Rank Approximation with General Losses

$$\min_{\mathbf{U}\in\mathbb{R}^{n\times k},\mathbf{V}\in\mathbb{R}^{k\times d}} \| \mathbf{A} - \mathbf{U}\mathbf{V} \|_g = \sum_{i,j} g((\mathbf{A} - \mathbf{U}\mathbf{V})_{i,j})$$

$$\| \mathbf{A} - \hat{\mathbf{U}}\hat{\mathbf{V}} \|_g \leq \kappa \min_{\mathbf{U}\in\mathbb{R}^{n\times k},\mathbf{V}\in\mathbb{R}^{k\times d}} \| \mathbf{A} - \mathbf{U}\mathbf{V} \|_g$$

Subset of $\approx k$ columns of $\mathbf{A}$

Approximation factor κ

- ℓ_p, $p < 2$: $\kappa \approx k^{1/p-1/2}$ (Mahankali—Woodruff 2021)
- $\ell_p, p > 2$: $\kappa \approx k^{1-1/p}$ (Dan—Wang—Zhang—Zhou—Ravikumar 2019)
- Huber: $\kappa \approx k^2$ (Song—Woodruff—Zhong 2019) 

Matrix Approximation

Low Rank Approximation with General Losses

$$\min_{\mathbf{U}\in\mathbb{R}^{n\times k},\mathbf{V}\in\mathbb{R}^{k\times d}} \| \mathbf{A}-\mathbf{U}\mathbf{V} \|_g = \sum_{i,j} g((\mathbf{A}-\mathbf{U}\mathbf{V})_{i,j})$$

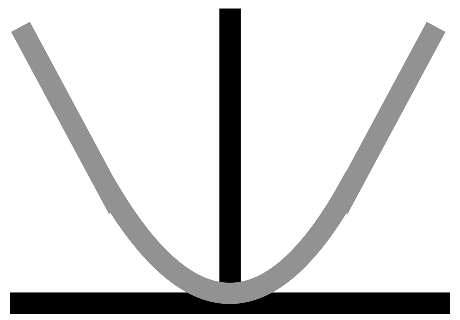
$$\| \mathbf{A}-\hat{\mathbf{U}}\hat{\mathbf{V}} \|_g \le \kappa \min_{\mathbf{U}\in\mathbb{R}^{n\times k},\mathbf{V}\in\mathbb{R}^{k\times d}} \| \mathbf{A}-\mathbf{U}\mathbf{V} \|_g$$

Subset of $\approx k$ columns of $\mathbf{A}$

Approximation factor κ

- ℓ_p, $p < 2$: $\kappa \approx k^{1/p-1/2}$ (Mahankali—Woodruff 2021)
- $\ell_p, p > 2$: $\kappa \approx k^{1-1/p}$ (Dan—Wang—Zhang—Zhou—Ravikumar 2019)

  $\kappa \approx k^{1/2-1/p}$ (Woodruff—Y 2023)
- Huber: $\kappa \approx k^2$ (Song—Woodruff—Zhong 2019)



Matrix Approximation

Low Rank Approximation with General Losses

$$\min_{\mathbf{U}\in\mathbb{R}^{n\times k},\mathbf{V}\in\mathbb{R}^{k\times d}} \| \mathbf{A} - \mathbf{UV} \|_g = \sum_{i,j} g(\ (\mathbf{A} - \mathbf{UV})_{i,j}\ )$$

$$\| \mathbf{A} - \hat{\mathbf{U}}\hat{\mathbf{V}} \|_g \leq \kappa \min_{\mathbf{U}\in\mathbb{R}^{n\times k},\mathbf{V}\in\mathbb{R}^{k\times d}} \| \mathbf{A} - \mathbf{UV} \|_g$$

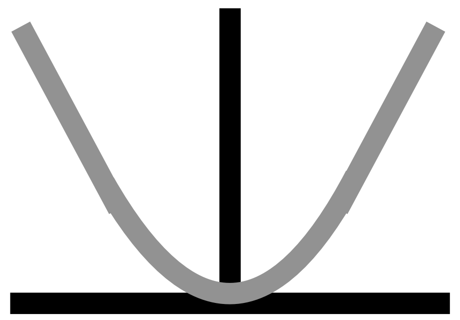
Subset of $\approx k$ columns of $\mathbf{A}$

Approximation factor κ

- ℓ_p, $p < 2$: $\kappa \approx k^{1/p-1/2}$ (Mahankali—Woodruff 2021)
- $\ell_p, p > 2$: $\kappa \approx k^{1-1/p}$ (Dan—Wang—Zhang—Zhou—Ravikumar 2019)

  $\kappa \approx k^{1/2-1/p}$ (Woodruff—Y 2023)
- Huber: $\kappa \approx k^2$ (Song—Woodruff—Zhong 2019)

  $\kappa \approx k$ (Woodruff—Y 2023)



Matrix Approximation

Low Rank Approximation with General Losses

$$\min_{\mathbf{U}\in\mathbb{R}^{n\times k},\mathbf{V}\in\mathbb{R}^{k\times d}} \| \mathbf{A} - \mathbf{U}\mathbf{V} \|_g = \sum_{i,j} g(\ (\mathbf{A} - \mathbf{U}\mathbf{V})_{i,j}\ )$$

$$\| \mathbf{A} - \hat{\mathbf{U}}\hat{\mathbf{V}} \|_g \leq \kappa \min_{\mathbf{U}\in\mathbb{R}^{n\times k},\mathbf{V}\in\mathbb{R}^{k\times d}} \| \mathbf{A} - \mathbf{U}\mathbf{V} \|_g$$

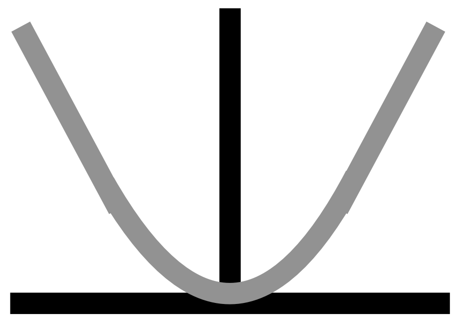
Subset of $\approx k$ columns of $\mathbf{A}$

Approximation factor κ

- ℓ_p, $p < 2$: $\kappa \approx k^{1/p-1/2}$ (Mahankali—Woodruff 2021)
- $\ell_p, p > 2$: $\kappa \approx k^{1-1/p}$ (Dan—Wang—Zhang—Zhou—Ravikumar 2019)

  $\kappa \approx k^{1/2-1/p}$ (Woodruff—Y 2023)
- Huber: $\kappa \approx k^2$ (Song—Woodruff—Zhong 2019)

  $\kappa \approx k$ (Woodruff—Y 2023)



Techniques: well-conditioned spanning sets, ℓ_p Lewis weights, ...

Matrix Approximation

Summary

Matrix Approximation

Summary

- We develop a wide variety of techniques for matrix approximation based on sampling and sketching techniques

Matrix Approximation

Summary

- We develop a wide variety of techniques for matrix approximation based on sampling and sketching techniques
  - **Tight bounds**: we obtain nearly optimal bounds for fundamental problems for matrix approximation

Matrix Approximation

Summary

- We develop a wide variety of techniques for matrix approximation based on sampling and sketching techniques
  - **Tight bounds**: we obtain nearly optimal bounds for fundamental problems for matrix approximation
  - **Uncertain/resource-limited settings**: we develop the first algorithms for matrix approximation in online, streaming, and other models of computation

Matrix Approximation

Summary

- We develop a wide variety of techniques for matrix approximation based on sampling and sketching techniques
  - **Tight bounds**: we obtain nearly optimal bounds for fundamental problems for matrix approximation
  - **Uncertain/resource-limited settings**: we develop the first algorithms for matrix approximation in online, streaming, and other models of computation
  - **Generalized loss functions**: we obtain new and improved algorithms for matrix approximation under ℓ_p, Huber, and even further general loss functions

Matrix Approximation

Summary

- We develop a wide variety of techniques for matrix approximation based on sampling and sketching techniques
  - **Tight bounds**: we obtain nearly optimal bounds for fundamental problems for matrix approximation
  - **Uncertain/resource-limited settings**: we develop the first algorithms for matrix approximation in online, streaming, and other models of computation
  - **Generalized loss functions**: we obtain new and improved algorithms for matrix approximation under ℓ_p, Huber, and even further general loss functions
  - **Applications**: we apply algorithmic techniques from matrix approximation to solve fundamental problems in machine learning and computational geometry

Matrix Approximation

Open Problems

Matrix Approximation

Open Problems

Open Problem 1. Is there a $\kappa = O(1)$ distortion subspace embedding for the Huber loss with $r \approx d$ rows?

Matrix Approximation

Open Problems

Open Problem 1. Is there a $\kappa = O(1)$ distortion subspace embedding for the Huber loss with $r \approx d$ rows?

Current best: $r \approx d^{4-2\sqrt{2}}$ (Musco—Musco—Woodruff—Y 2022)

Matrix Approximation

Open Problems

Open Problem 1. Is there a $\kappa = O(1)$ distortion subspace embedding for the Huber loss with $r \approx d$ rows?

Current best: $r \approx d^{4-2\sqrt{2}}$ (Musco—Musco—Woodruff—Y 2022)

Open Problem 2. Is $r \approx \varepsilon^{-2} d^{p/2}$ optimal for $\kappa = (1+\varepsilon)$ distortion ℓ_p subspace embeddings for $p > 2$?

Matrix Approximation

Open Problems

Open Problem 1. Is there a $\kappa = O(1)$ distortion subspace embedding for the Huber loss with $r \approx d$ rows?

Current best: $r \approx d^{4-2\sqrt{2}}$ (Musco—Musco—Woodruff—Y 2022)

Open Problem 2. Is $r \approx \varepsilon^{-2} d^{p/2}$ optimal for $\kappa = (1+\varepsilon)$ distortion ℓ_p subspace embeddings for $p > 2$?

Current best: $r \gtrsim \varepsilon^{-2} + d^{p/2}$ (Li—Wang—Woodruff 2020)

Matrix Approximation

Other works

- Subspace Embeddings + Applications:
  - **Sharper Bounds for ℓ_p Sensitivity Sampling [preprint]**
  - New Subset Selection Algorithms for Low Rank Approximation: Offline and Online [STOC'23]
  - Online Lewis Weight Sampling [SODA'23]
  - High-Dimensional Geometric Streaming in Polynomial Space [FOCS'22]
  - Active Linear Regression for ℓ_p Norms and Beyond [FOCS'22]
  - Exponentially Improved Dimensionality Reduction for ℓ_1: Subspace Embeddings and Independence Testing [COLT'21]
- Low Rank Approximation
  - New Subset Selection Algorithms for Low Rank Approximation: Offline and Online [STOC'23]
  - **Improved Algorithms for Low Rank Approximation from Sparsity [SODA'22]**
- **Sequential Attention for Feature Selection [ICLR'23]**