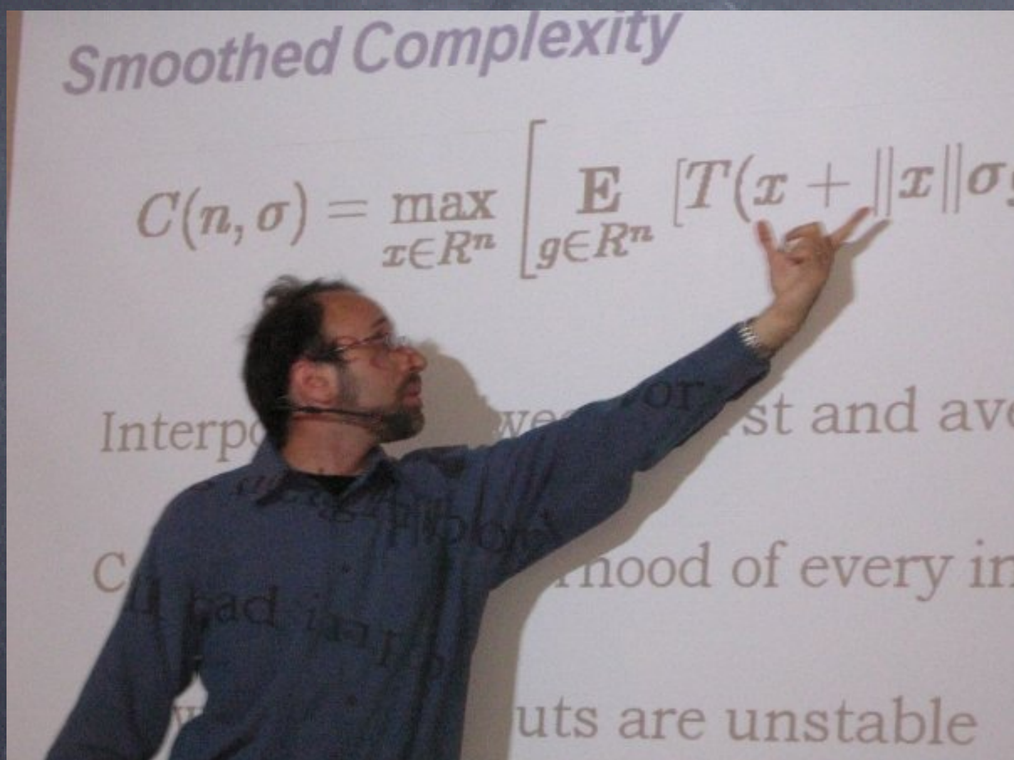
Exact Recovery of Sparsely-Used Dictionaries

Daniel A. Spielman, Huan Wang, John Wright

Yale University
Columbia University



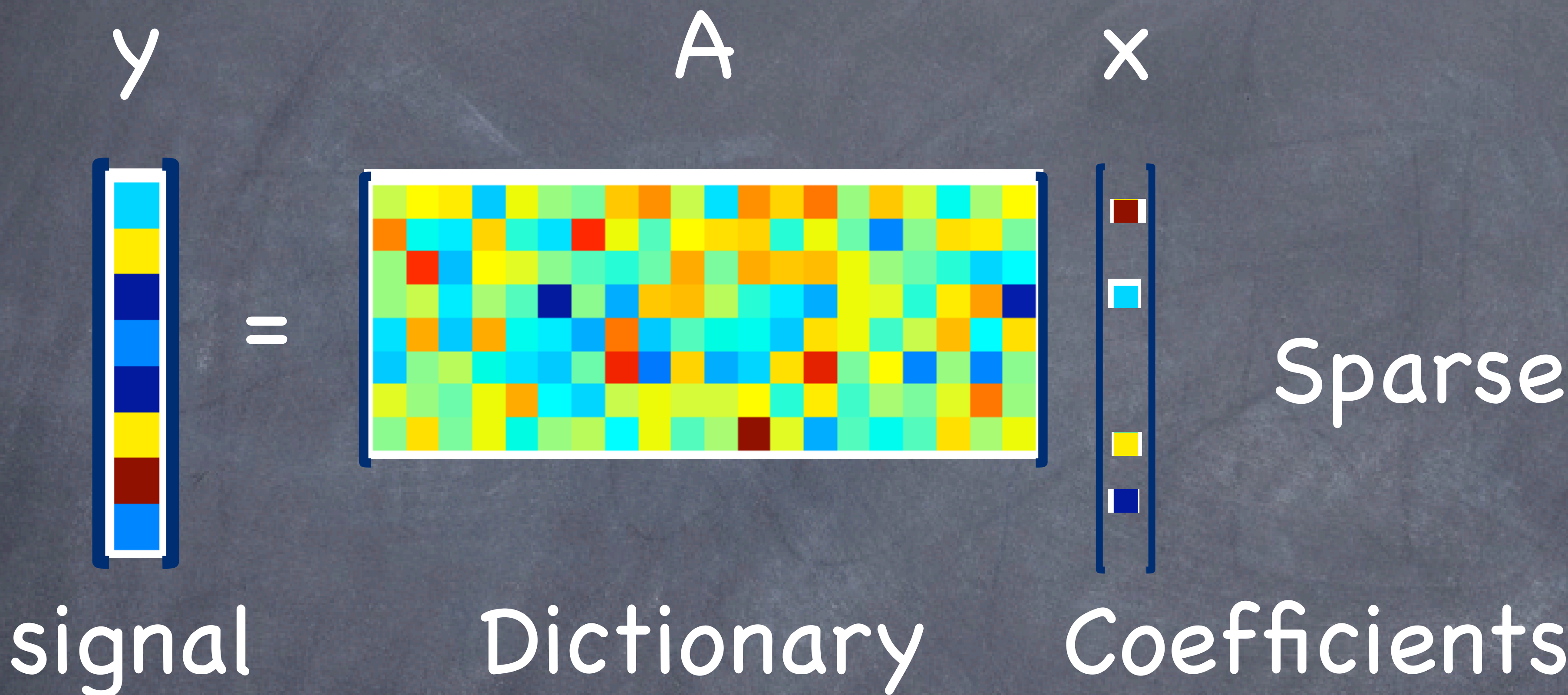
Daniel A. Spielman

Yale University

John Wright

Columbia University

Sparse Approximation



Sparse Assumption

Sparse Approximation

y A x

signal Dictionary Coefficients

Sparse Assumption

Data Dictionary

Dictionary Learning

Given a sample matrix Y (n-by-p), find A (n-by-n) and X(n-by-p), such that

1. Y=AX
2. X is sparse.

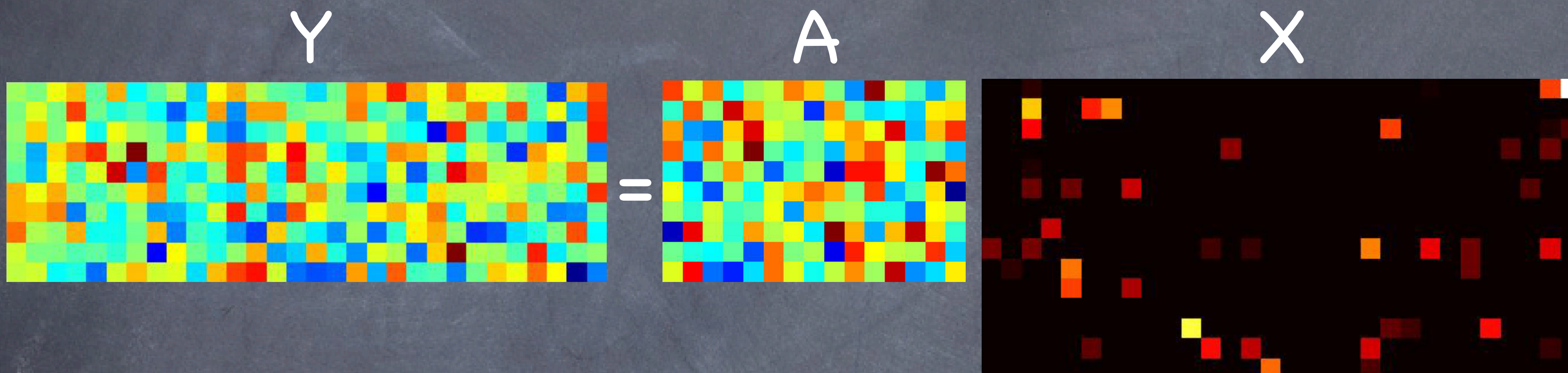
Both A and X are unknown.

Over-Complete Case

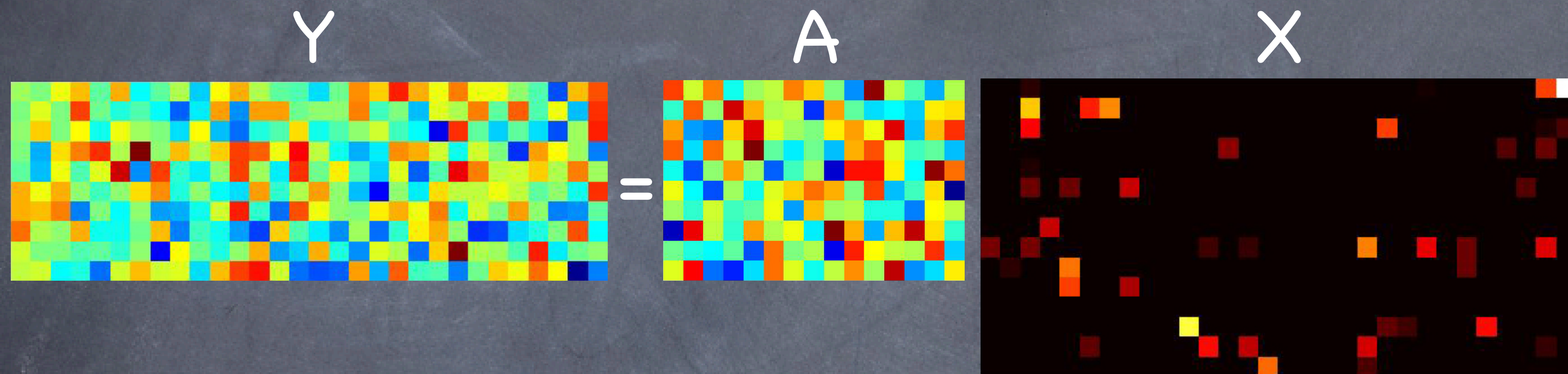
Given a sample matrix Y (m-by-p), and n, find A (m-by-n) and X (n-by-p), such that

1. $Y = AX$
2. X is sparse.
3. $m < n$

Difficulty

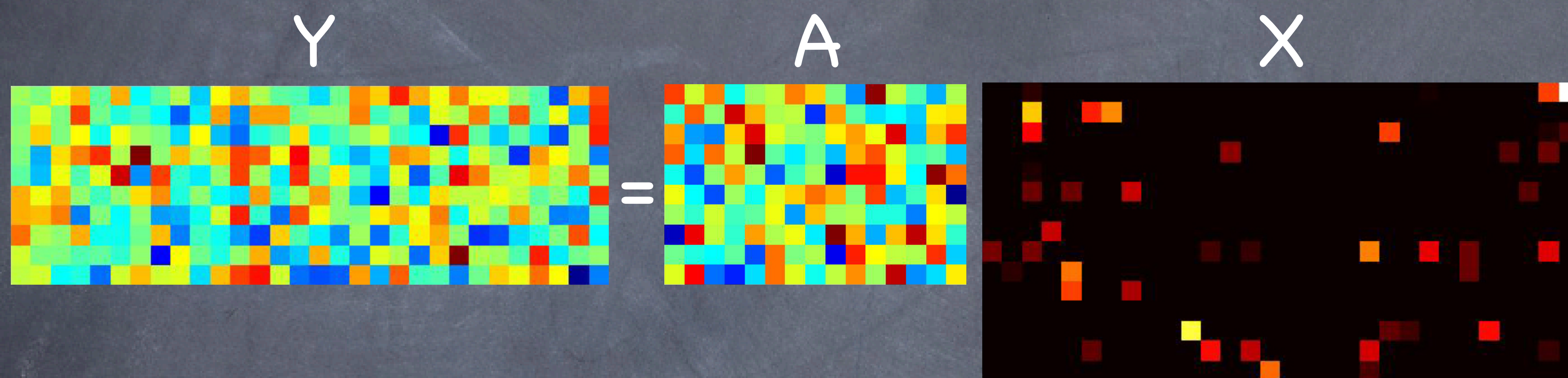


Difficulty



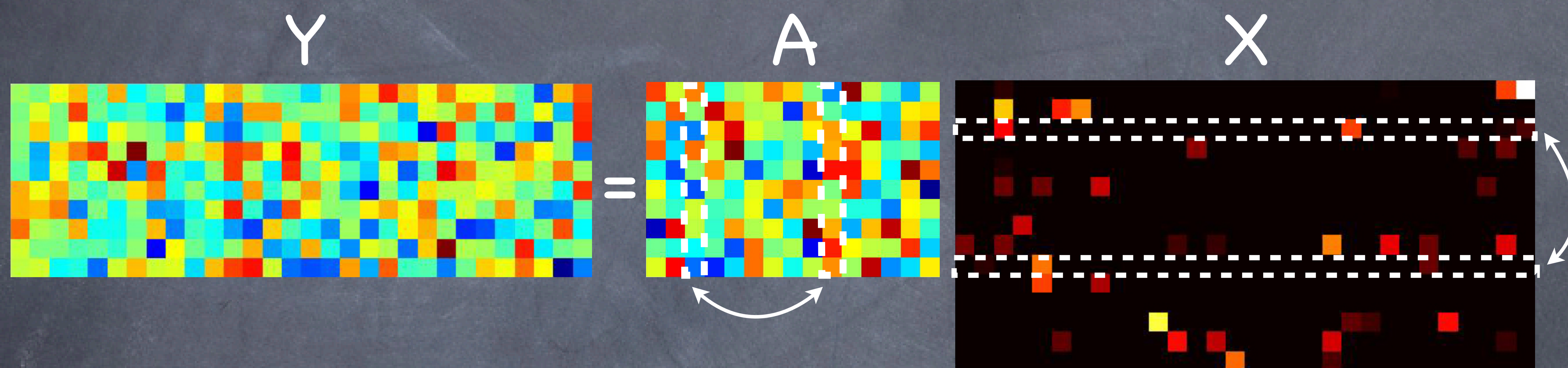
- Ambiguities:

Difficulty



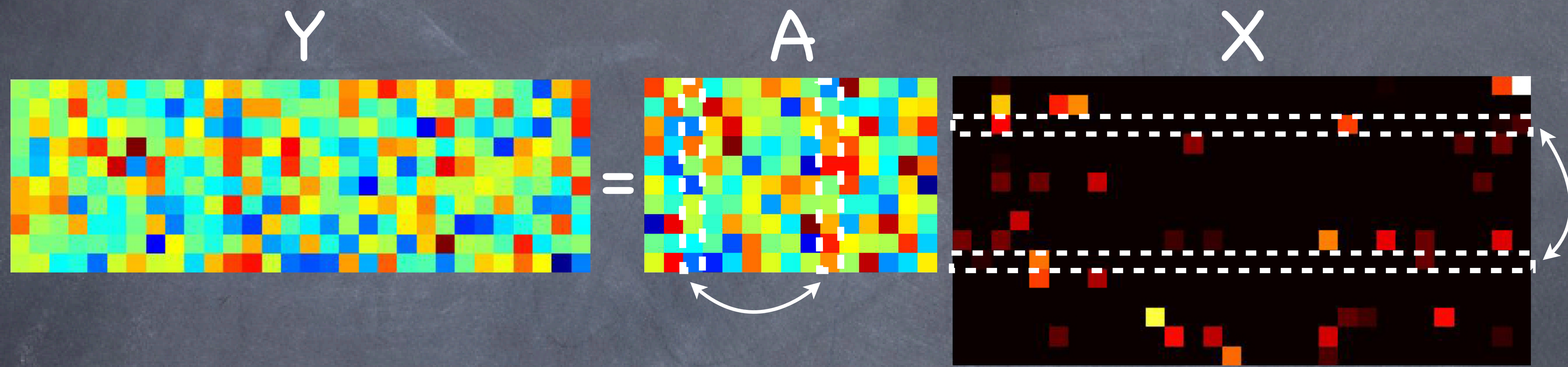
- Ambiguities: (A, X) *or* $(A\Pi\Lambda, \Lambda^{-1}\Pi^T X)$

Difficulty



- Ambiguities: (A, X) *or* $(A\Pi\Lambda, \Lambda^{-1}\Pi^{T} X)$

Difficulty



- Ambiguities: (A, X) *or* $(A\Pi\Lambda, \Lambda^{-1}\Pi^T X)$

- Non-Convexity: Bilinear Form

Previous Works:

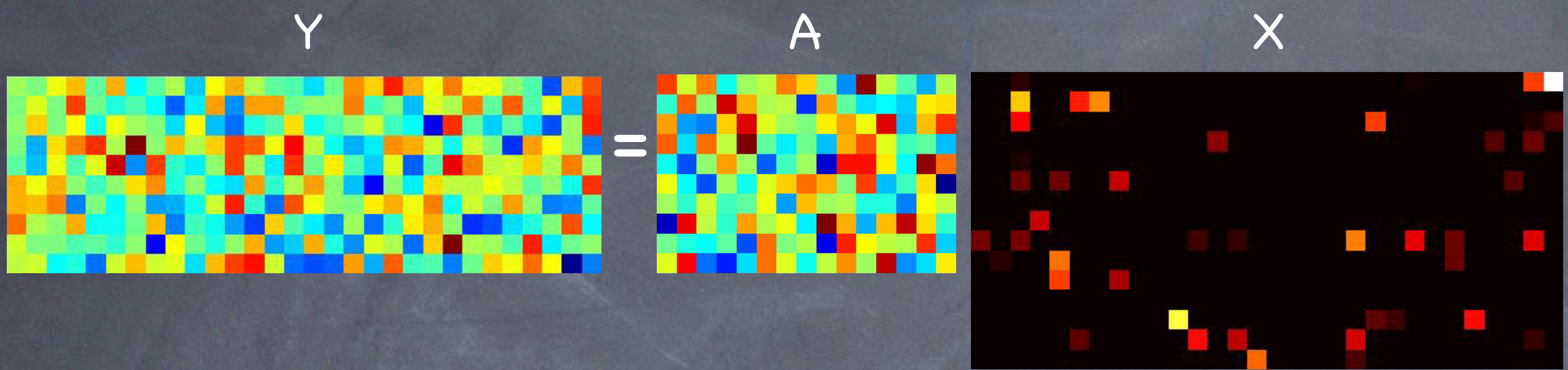
Aharon, Elad, and Bruckstein: K-SVD

Mairal, Bach, Ponce, and Sapiro: Online Dictionary Learning

Geng, and Wright: Local Analysis

Vainsencher, Mannor and Bruckstein: Generalization Bound

A ladder to Global Optimum



- Model:
  - A is non-singular, and square.
  - X is Bernoulli-Gaussian.
  - p is large ($p \sim n \log n$)

Bernoulli-Gaussian

Prob. NZ: θ

NZ entry: $N(0,1)$

The Algorithm

For i=1 to p

$$w_i = \arg\min_w \| w^T Y \|_1 \text{ s.t. } \quad (Ye_i)^T w = 1$$

$$x_i = w_i^T Y$$

end

Advantages and Future Work

- Provable performance guarantee.
- Higher Accuracy

- Over-complete
- Noise

Results

Is the Answer Unique?

$$\min_{A,X} \|X\|_0$$

s.t. $$Y = AX$$

- Unique up to scaling, and permutation when:
  - $1/n < \theta \leq 1/4$
  - the sample #: p>cnlogn

n: dimension

Performance Guarantee

$$\min_w \| w^T Y \|_1 \qquad \text{s.t.} \qquad r^T w = 1$$

- In each Iteration: use a column of Y as r.

- Correctly recover all rows of X when
  - $\frac{2}{n} \le \theta \le \frac{\alpha}{\sqrt{n \log n}}$
  - sample # $p > cn^2 \log^2 n$

n: dimension

Performance Guarantee

$$\min_w \| w^T Y \|_1 \qquad \text{s.t.} \qquad r^T w = 1$$

- In each Iteration: use a column of Y as r.

r=Y(:,i)+Y(:,j)

- Correctly recover all rows of X when
  - $\frac{2}{n} \le \theta \le \frac{\alpha}{\sqrt{n \log n}}$
  - sample # $p > cn^2 \log^2 n$

n: dimension

Performance Guarantee

$$\min_w \| w^T Y \|_1 \qquad \text{s.t.} \qquad r^T w = 1$$

- In each Iteration: use a column of Y as r.

r=Y(:,i)+Y(:,j)

- Correctly recover all rows of X when
  - $\frac{2}{n} \le \theta \le \frac{\alpha}{\sqrt{n \log n}}$
  - sample # $p > cn^2 \log^2 n$

$\frac{2}{n} \le \theta \le \frac{\alpha}{\sqrt{n}}$

n: dimension

Intuition of Proof

Intuition-Is the Answer Unique?

$$Y=AX=A'X'$$

Intuition-Is the Answer Unique?

$$Y = AX = A'X'$$

- when p is large, rank(Y)=rank(X)=n.

Intuition-Is the Answer Unique?

$$Y = AX = A'X'$$

- when p is large, $\text{rank}(Y) = \text{rank}(X) = n$.
- A' has to be nonsingular.

Intuition-Is the Answer Unique?

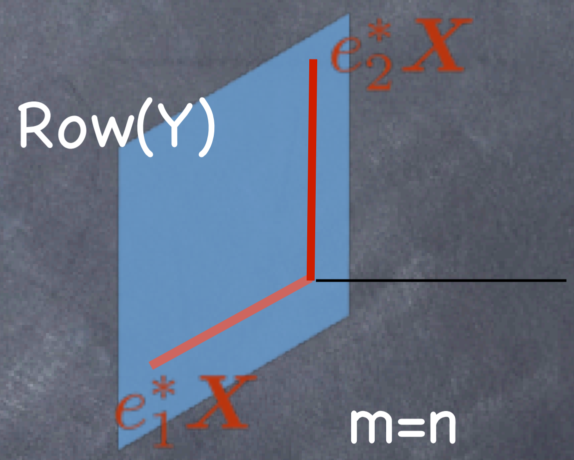
$$Y = AX = A'X'$$

- when p is large, $\mathrm{rank}(Y) = \mathrm{rank}(X) = n$.
- A' has to be nonsingular.
- $\mathrm{span}(Y) = \mathrm{span}(X) = \mathrm{span}(X')$

Intuition-Is the Answer Unique?

Y=AX=A′X′

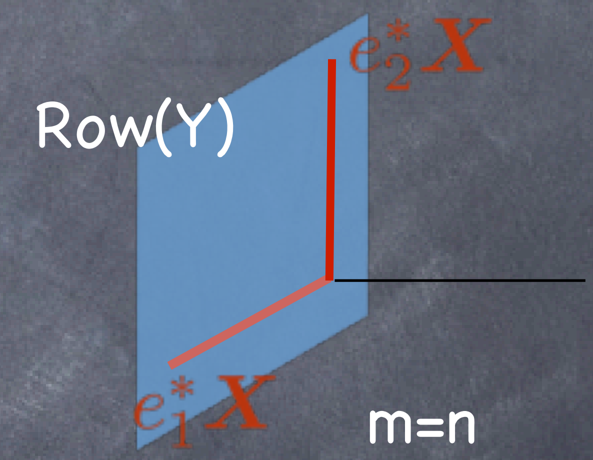
- when p is large, rank(Y)=rank(X)=n.
- A′ has to be nonsingular.
- span(Y)=span(X)=span(X′)



Intuition-Is the Answer Unique?

$Y=AX=A'X'$

- when p is large, rank(Y)=rank(X)=n.
- A' has to be nonsingular.
- span(Y)=span(X)=span(X′)



Rows of X are the only sparse vectors in span(Y)

Intuition

x	x

0	0	x	0	0	x	x
x	0	0	0	x	0	x

=

x	0	x	0	x	x	x

Intuition

x	x

0	0	x	0	0	x	x
x	0	0	0	x	0	x

=

x	0	x	0	x	x	x

Rigorous when X is Bernoulli-Gaussian.

Intuition-Algorithm

Rows of X are the only sparse vectors in span(Y)

Intuition-Algorithm

Rows of X are the only sparse vectors in span(Y)

Seek sparse vectors in span(Y)

Intuition-Algorithm

Rows of X are the only sparse vectors in span(Y)

Seek sparse vectors in span(Y)

$$\min_w \| w^T Y \|_0$$

Intuition-Algorithm

Rows of X are the only sparse vectors in span(Y)

Seek sparse vectors in span(Y)

$$\min_w \| w^T Y \|_0 \quad \text{s.t.} \quad w \neq 0$$

Intuition-Algorithm

Rows of X are the only sparse vectors in span(Y)

Seek sparse vectors in span(Y)

$$\min_w \| w^T Y \|_0 \quad \text{s.t.} \quad w \neq 0$$

Not Convex!

Intuition-Algorithm

Rows of X are the only sparse vectors in span(Y)

Seek sparse vectors in span(Y)

$$\min_w \|w^T Y\|_0 \quad \text{s.t.} \quad w \neq 0$$

Not Convex!

$$\min_w \|w^T Y\|_1$$

Intuition-Algorithm

Rows of X are the only sparse vectors in span(Y)

Seek sparse vectors in span(Y)

$$\min_w \| w^T Y \|_0 \quad \text{s.t.} \quad w \neq 0$$

Not Convex!

$$\min_w \| w^T Y \|_1 \quad \text{s.t.} \quad \mathbf{r}^T w = 1$$

Intuition-Algorithm

Rows of X are the only sparse vectors in span(Y)

Seek sparse vectors in span(Y)

$$\min_w \| w^T Y \|_0 \quad \text{s.t.} \quad w \neq 0$$

Not Convex!

$$\min_w \| w^T Y \|_1 \quad \text{s.t.} \quad \mathbf{r}^T w = 1$$

Linear Programming

How to Choose r?

$$\min_w \| w^T Y \|_1 \quad \text{s.t.} \quad \mathbf{r}^T w = 1$$

How to Choose r?

$$\min_w \| w^T Y \|_1 \quad \text{s.t.} \quad \mathbf{r}^T w = 1$$

Since Y=AX, let $v = A^T w$

$$\min_v \| \mathbf{v}^T X \|_1 \quad \text{s.t.} \quad b^T v = 1$$

where $b = A^{-1} r$

How to Choose r?

$$\min_w \| w^T Y \|_1 \quad \text{s.t.} \quad \mathbf{r}^T w = 1$$

Since Y=AX, let $v = A^T w$

only for analysis

$$\min_v \| \mathbf{v}^T X \|_1 \quad \text{s.t.} \quad b^T v = 1$$

where $b = A^{-1} r$

How to Choose b?

$$\min_v \| \mathbf{v}^T X \|_1 \quad \text{s.t.} \quad b^T v = 1$$

If b is an all-one vector?

How to Choose b?

$$\min_{v} \| \mathbf{v}^T X \|_1 \quad \text{s.t.} \quad b^T v = 1$$

If b is an all-one vector?

Symmetric, no preference which row to pick up

How to Choose b?

$$\min_v \| \mathbf{v}^T X \|_1 \quad \text{s.t.} \quad b^T v = 1$$

If b is an all-one vector?

Symmetric, no preference which row to pick up

Combination of Rows

How to Choose b?

$$\min_v \| \mathbf{v}^T X \|_1 \quad \text{s.t.} \quad b^T v = 1$$

$$\left\| \begin{bmatrix} 1/2 & 1/2 \end{bmatrix} \times \begin{bmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & -1 \end{bmatrix} \right\|_1 = 1$$

$$\left\| \begin{bmatrix} 1 & 0 \end{bmatrix} \times \begin{bmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & -1 \end{bmatrix} \right\|_1 = 2$$

How to Choose b?

$$\min_v \| \mathbf{v}^T X \|_1 \quad \text{s.t.} \quad b^T v = 1$$

If b is e_1 ?

1	0	0	0	0

How to Choose b?

$$\min_v \| \mathbf{v}^T X \|_1 \quad \text{s.t.} \quad b^T v = 1$$

If b is e_1 ?

1	0	0	0	0

We will get the first row of X!

How to Choose b?

$$\min_v \| \mathbf{v}^T X \|_1 \quad \text{s.t.} \quad b^T v = 1$$

If b is e_1 ?

1	0	0	0	0

We will get the first row of X!

Unbalanced b is used to break the symmetry.

How to Choose b?

$$\| (v_1 + v_{2:})^T X \|_1 \geq \| v_1^T X \|_1$$

How to Choose b?

$$\| (v_1 + v_{2:})^T X \|_1 \geq \| v_1^T X \|_1 + (|T| - |S|) E(| v_{2:}^T X(:,1) |)$$

How to Choose b?

$$\| (v_1 + v_{2:})^T X \|_1 \geq \| v_1^T X \|_1 + (| T | - | S |) E(| v_{2:}^T X(:,1) |)$$

How to Choose b?

$$\|(v_1 + v_{2:})^T X\|_1 \geq \|v_1^T X\|_1 + (|T| - |S|)E(|v_{2:}^T X(:,1)|)$$

As long as |T|>|S|

v_1 is the solution.

How to Choose b?

$$\min_v \| \mathbf{v}^T X \|_1 \quad \text{s.t.} \quad b^T v = 1$$

Suppose $j = \arg\max_i | b(i) |$

We will get the jth row of X if there is a large gap between |b(j)| and the rest.

$$\hat{v} = [0,0,\ldots,1/b_j,0,\ldots,0]^T$$

How to Choose b?

We prefer a sparse b. If $r = Ye_i$

$$\min_v \| \mathbf{v}^T X \|_1 \quad \text{s.t.} \quad b^T v = 1$$

$$\min_w \| w^T Y \|_1 \quad \text{s.t.} \quad \mathbf{r}^T w = 1$$

$$b = A^{-1} r = A^{-1} Y e_i = A^{-1} A X e_i = X(:,i)$$

Two-Step Argument

$$\min_v \| \mathbf{v}^T X \|_1 \quad \text{s.t.} \quad b^T v = 1$$

1. v supports only on the non-zero entries of b.

b

1.25	0	0	0	-0.3

v

*	*	*	*	*

Two-Step Argument

$$\min_v \| \mathbf{v}^T X \|_1 \quad \text{s.t.} \quad b^T v = 1$$

1. v supports only on the non-zero entries of b.

b

1.25	0	0	0	-0.3

v

*	*	*	*	*

b

1.25	0	0	0	-0.3

$\hat{v}$

*	0	0	0	*

Two-Step Argument

$$\min_v \| \mathbf{v}^T X \|_1 \quad \text{s.t.} \quad b^T v = 1$$

2. v has 1 nonzero.

b

1.25	0	0	0	-0.3

$\hat{v}$

*	0	0	0	*

Two-Step Argument

$$\min_v \| \mathbf{v}^T X \|_1 \quad \text{s.t.} \quad b^T v = 1$$

2. v has 1 nonzero.

b

1.25	0	0	0	-0.3

$\hat{v}$

*	0	0	0	*

b

1.25	0	0	0	-0.3

$\hat{v}$

0.8	0	0	0	0

Two-Step Argument

$$\min_v \| \mathbf{v}^T X \|_1 \quad \text{s.t.} \quad b^T v = 1$$

$\hat{v}$

*	0	0	0	*

X

X		X				X			
			X					X	

Two-Step Argument

$$\min_v \|\mathbf{v}^T X\|_1 \quad \text{s.t.} \quad b^T v = 1$$

$\hat{v}$

*	0	0	0	*

X

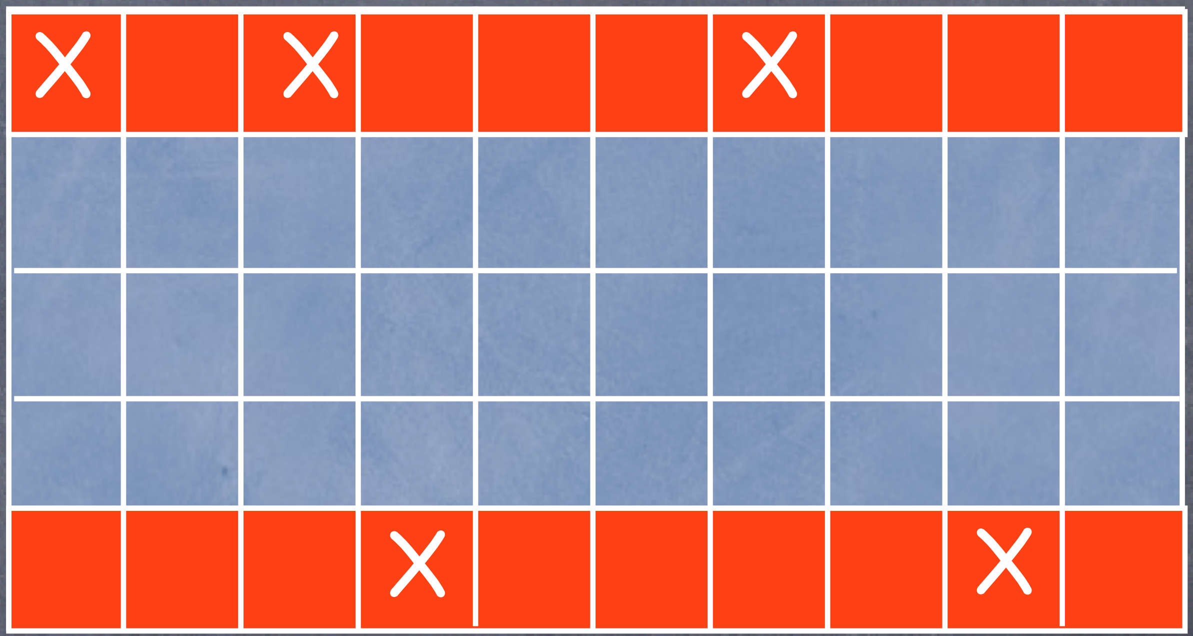
X		X				X			
			X					X	

Two-Step Argument

$$\min_v \|\mathbf{v}^T X\|_1 \quad \text{s.t.} \quad b^T v = 1$$

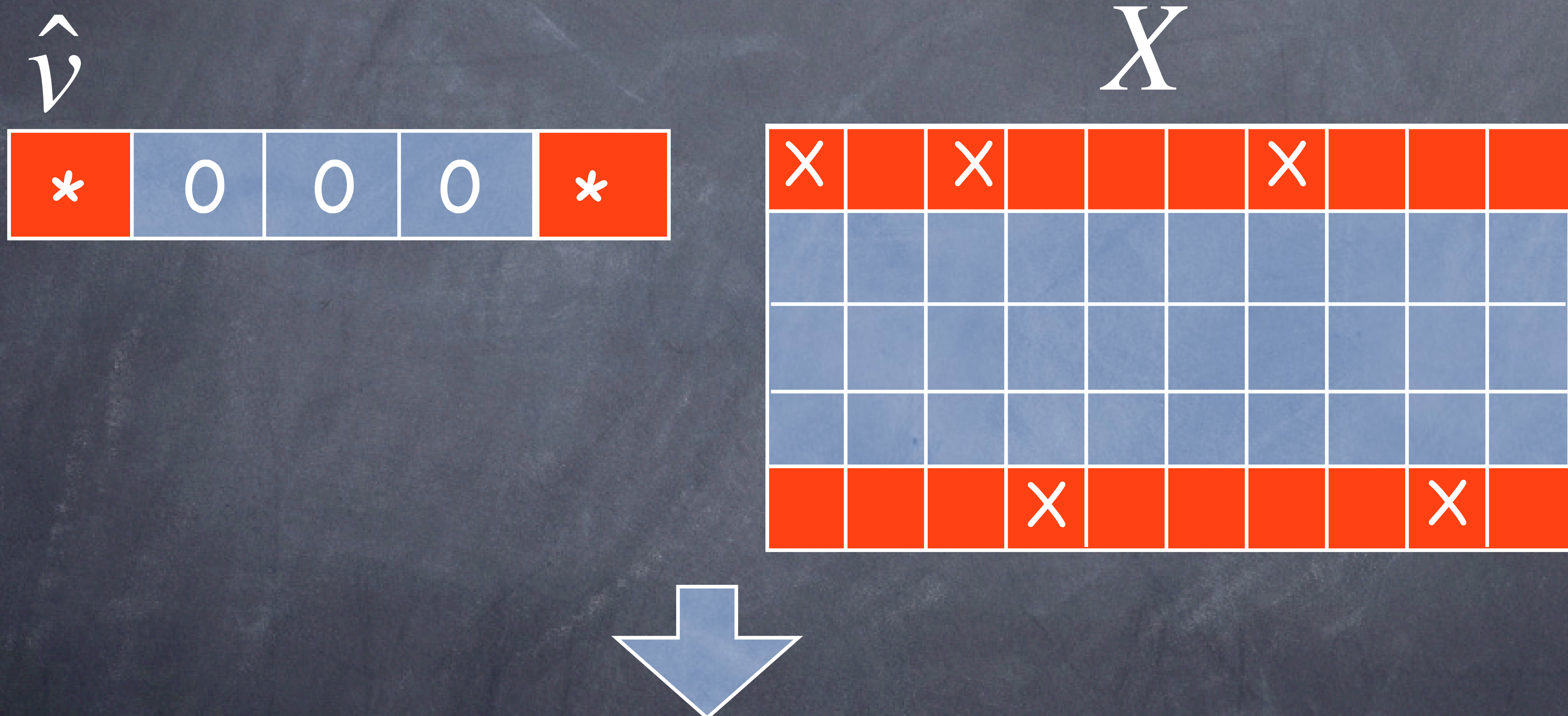
$\hat{v}$

X



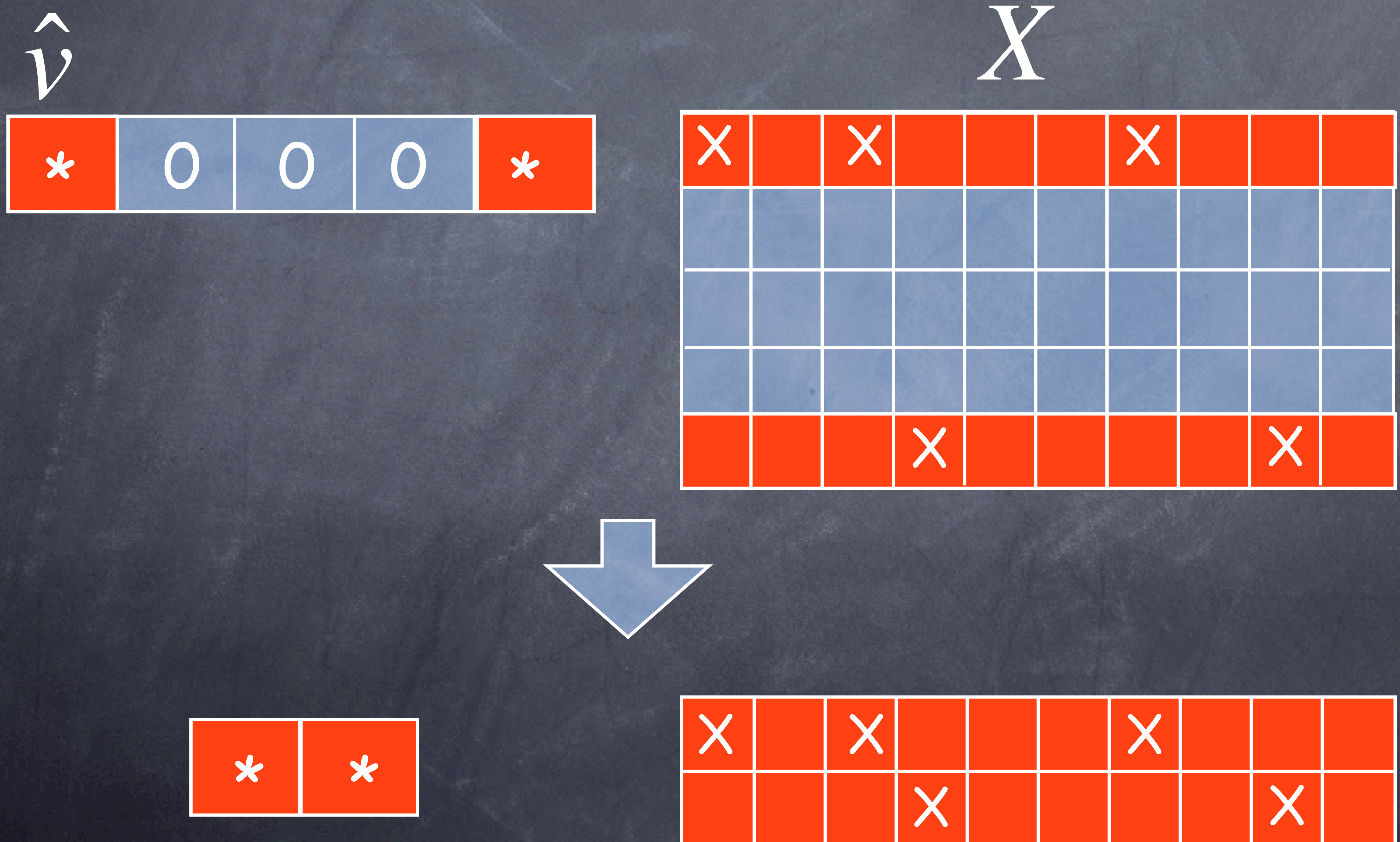
Two-Step Argument

$$\min_v \|\mathbf{v}^T X\|_1 \quad \text{s.t.} \quad b^T v = 1$$



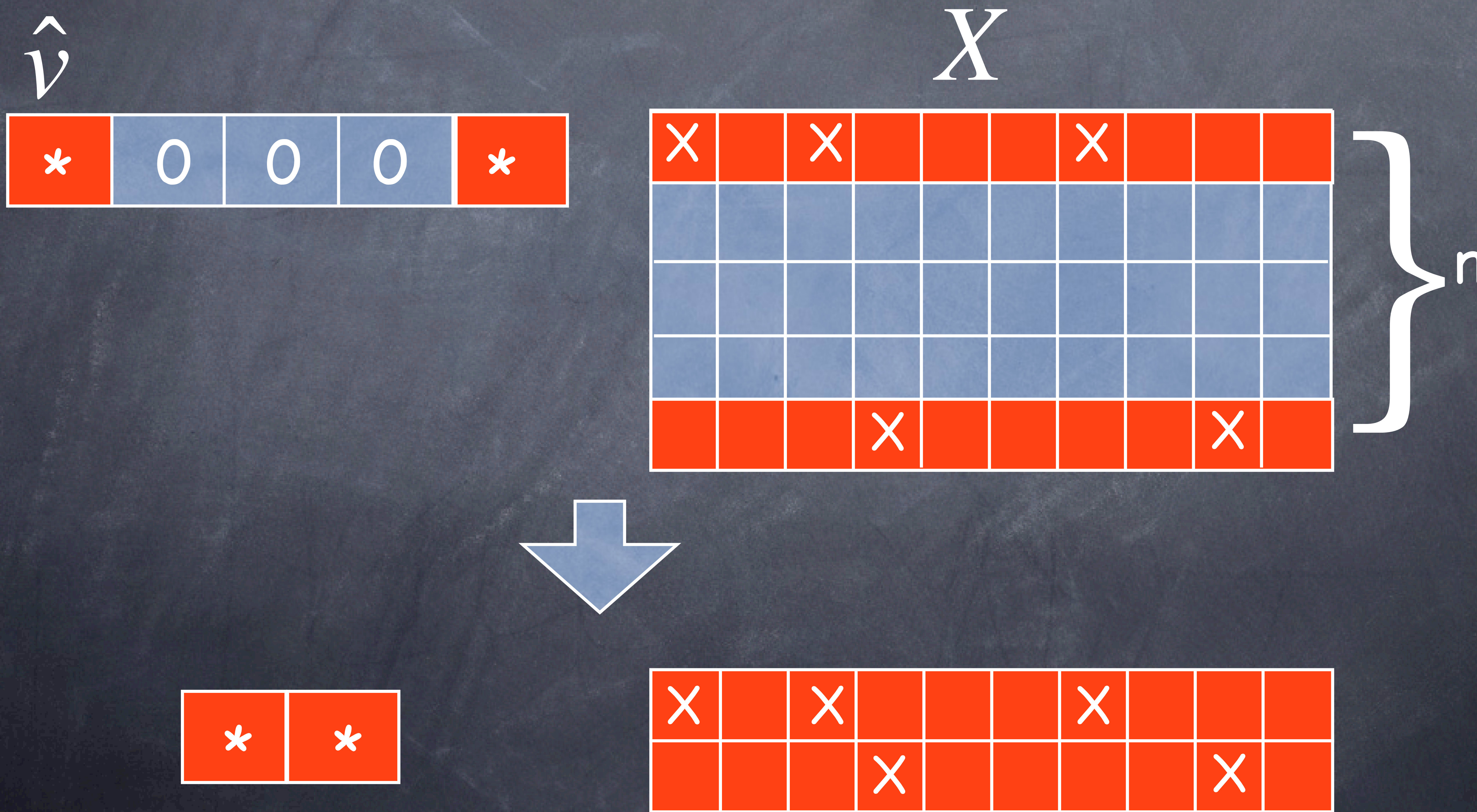
Two-Step Argument

$$\min_v \| \mathbf{v}^T X \|_1 \quad \text{s.t.} \quad b^T v = 1$$



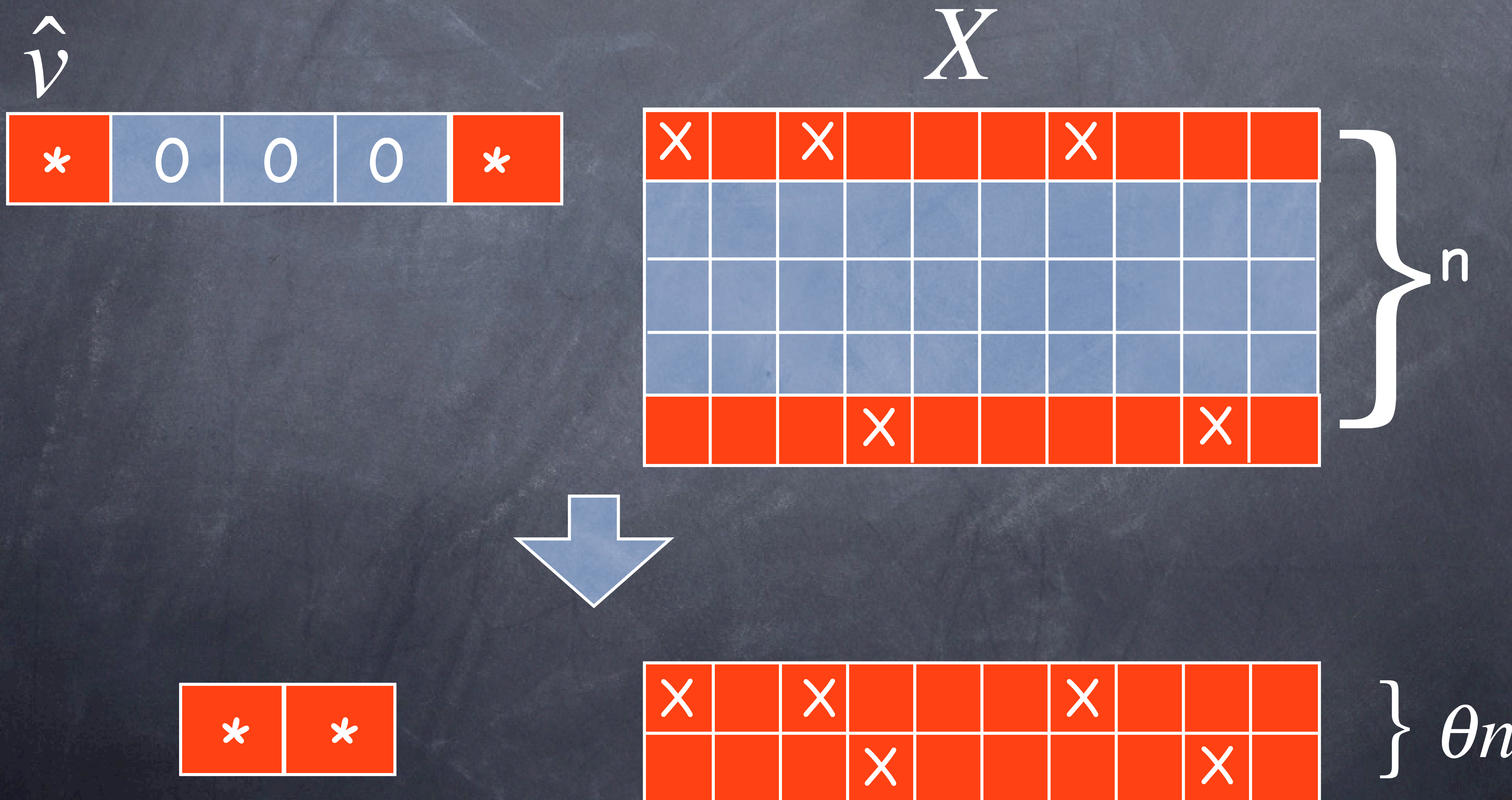
Two-Step Argument

$$\min_v \| \mathbf{v}^T X \|_1 \quad \text{s.t.} \quad b^T v = 1$$



Two-Step Argument

$$\min_v \| \mathbf{v}^T X \|_1 \quad \text{s.t.} \quad b^T v = 1$$



Two-Step Argument

$$\min_v \| \mathbf{v}^T X \|_1 \quad \text{s.t.} \quad b^T v = 1$$

Expected # of nonzeros per column:

$$(\theta n)\theta \le \left(\frac{c}{\sqrt{n \log n}}\right)^2 n = \frac{c^2}{\log\ n} < 1$$

Zero Columns+Unique Columns

Recovery of all rows

- Recover all rows of X, when p is large.
- Rows of X are the only sparse vectors in span(Y).
- Greedy algorithm

The Algorithm

Initialize: $X(1,:) = \arg\min_{x_i} \| x_i \|_0$

For i=2:n

$$X(i,:) = \arg\min_{x_i} \| x_i \|_0 \quad \text{s.t.} \quad x_i \notin span(X)$$

end

The Algorithm

For i=1 to n

For j=1 to p

$$r = P_{W^c} Y e_j$$

$$w_i = \arg\min_w \| w^T Y \|_1 \text{ s.t. } r^T w = 1$$

end

$$W(:,i) = \arg\min_{w_i} \| w_i^T Y \|_0$$

$$X(i,:) = W(:,i)^T Y$$

end

Simulations

Measure: $$error = \frac{\min_{\Pi,\Lambda} \| \hat{A}\Lambda\Pi - A \|_F}{\| A \|_F}$$

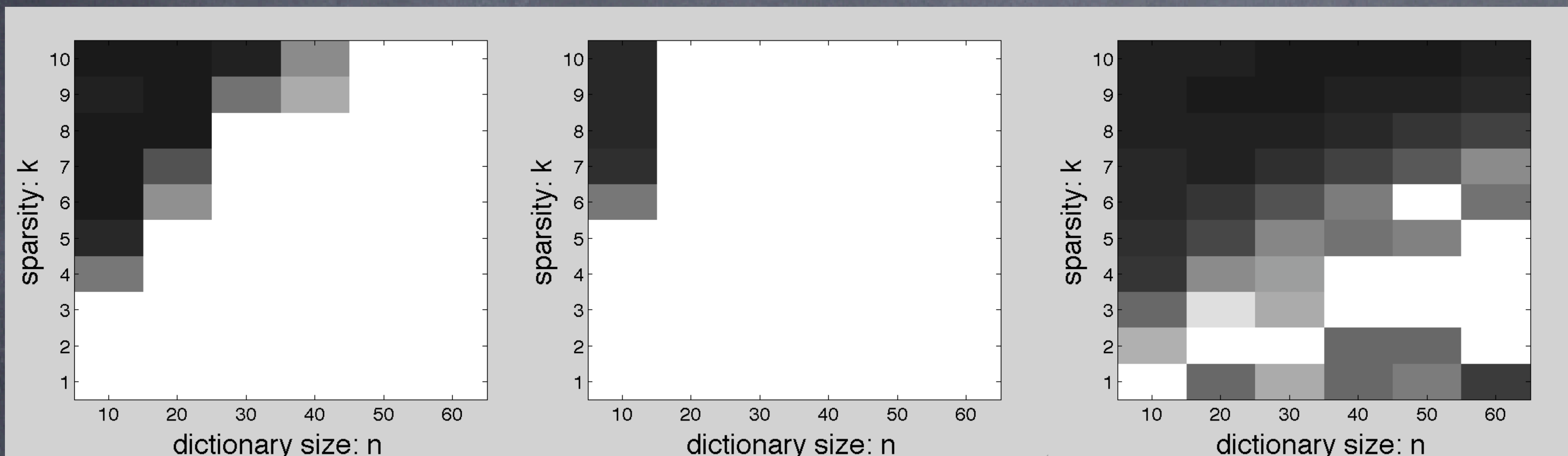
k=#nnz per column

k=1:10 10 trials for each configuration

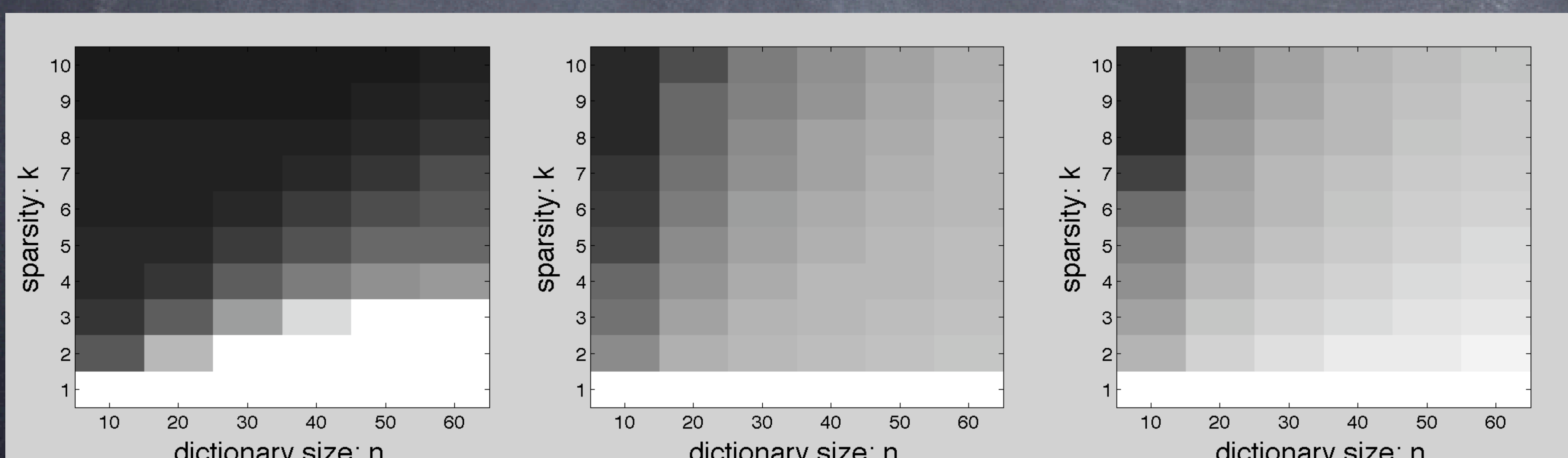
n=10:10:60

p=5nlogn

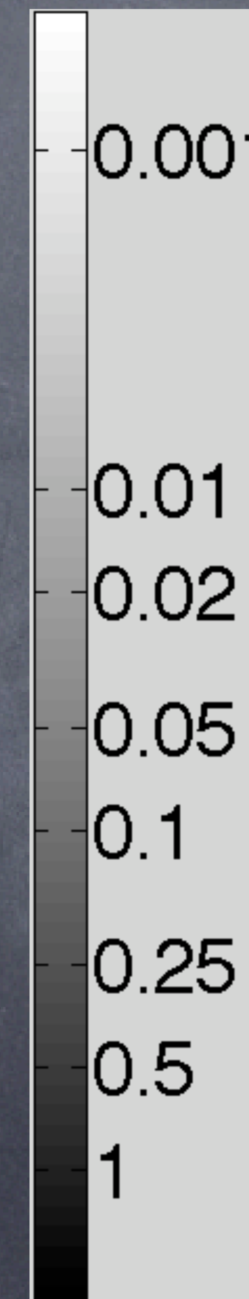
Simulations



SPUD(greedy) SPUD(proj) SIV



KSVD Online Learning Rel. Newton



Thank you!