# Sparse and Overcomplete Data Representation

**Michael Elad**
The CS Department
The Technion – Israel Institute of technology
Haifa 32000, Israel

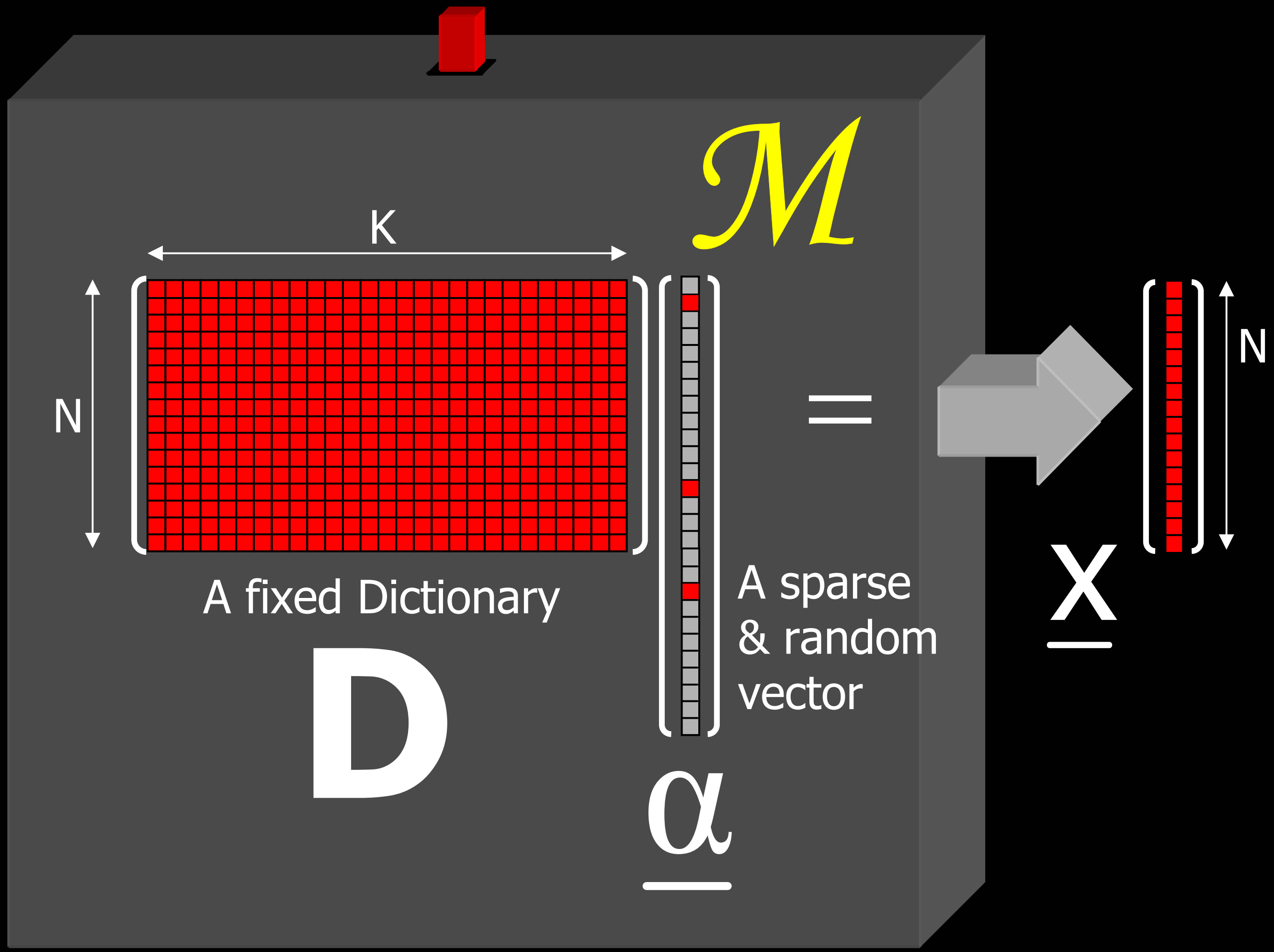
**Israel Statistical Association**
2005 Annual Meeting
Tel-Aviv University (Dan David bldg.)
May 17th, 2005

# Agenda

1. **A Visit to *Sparseland***
   **Motivating Sparsity & Overcompleteness**

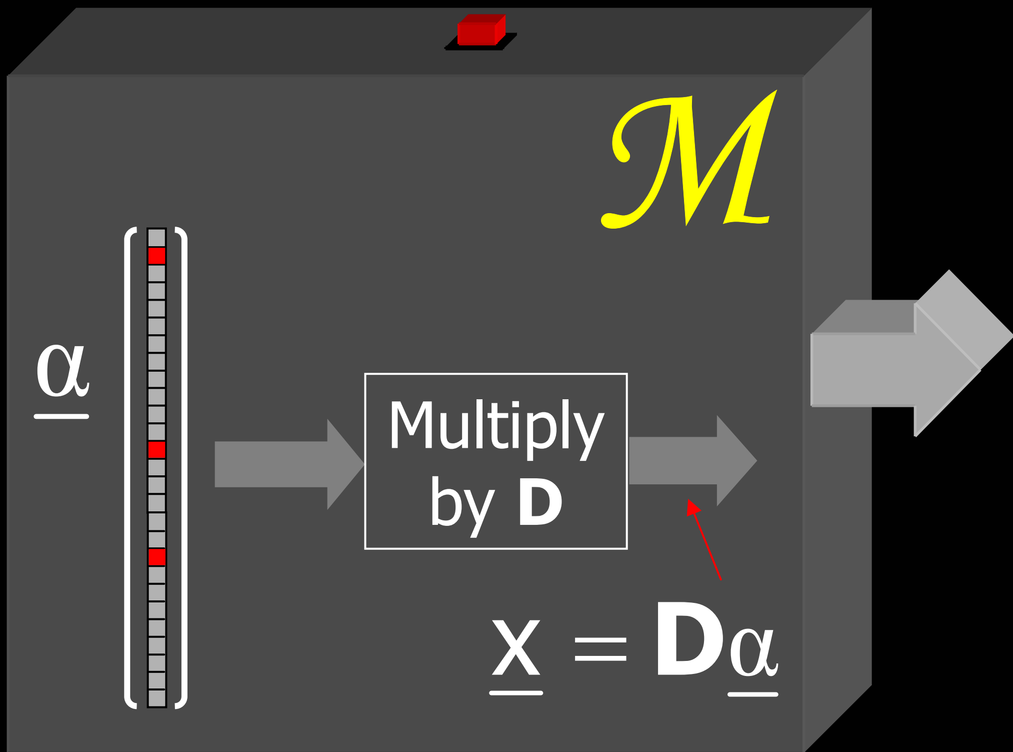
2. Answering the 4 Questions
   How & why should this work?

# Data Synthesis in *Sparseland*




- Every column in **D** (dictionary) is a prototype data vector (Atom).
- The vector $\underline{\alpha}$ is generated randomly with few non-zeros in random locations and random values.

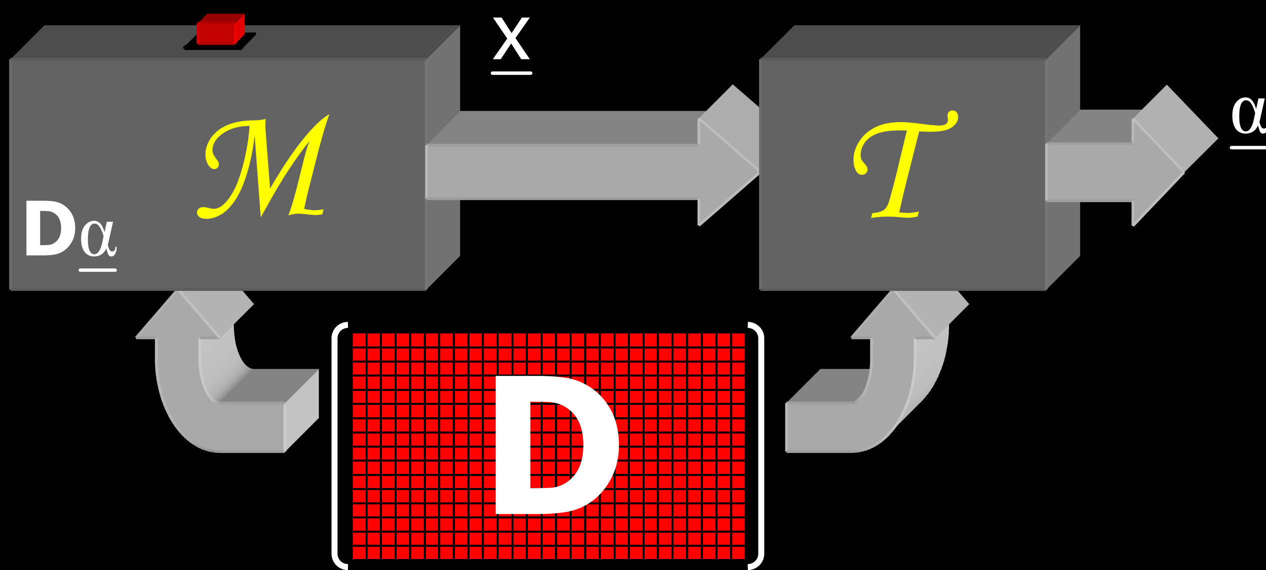
# *Sparseland* Data is Special



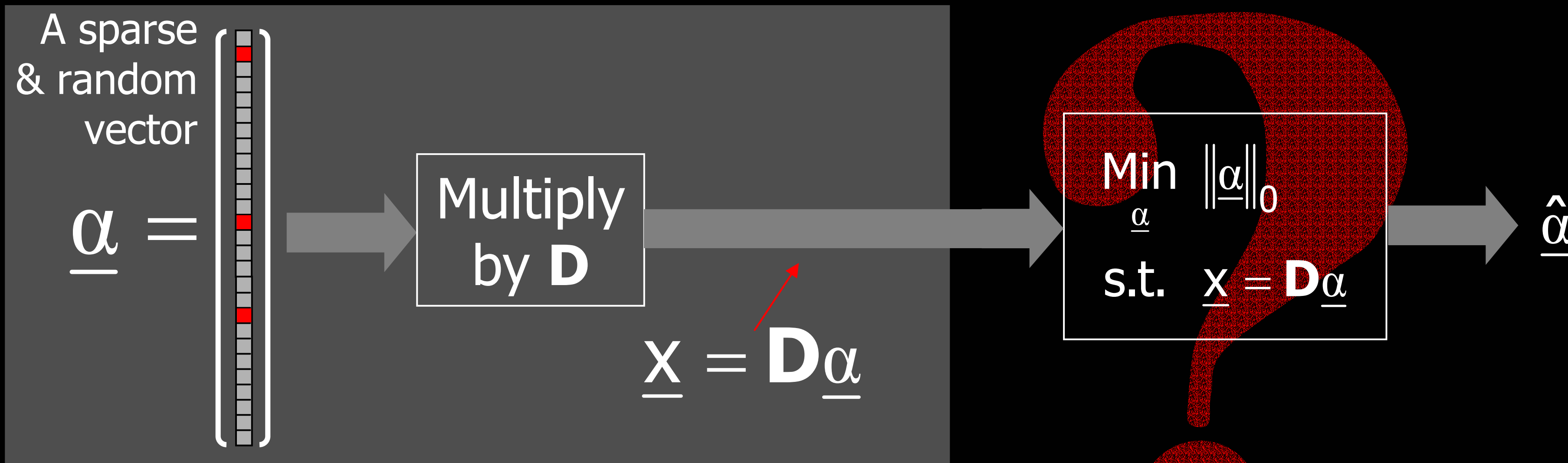

- **Simple:** Every generated vector is built as a linear combination of <u>**few**</u> atoms from our dictionary **D**
- **Rich:** A general model: the obtained vectors are a special type mixture-of-Gaussians (or Laplacians).

# Transforms in *Sparseland* ?

- Assume that x is known to emerge from $\mathcal{M}$.
- We desire simplicity, independence, and expressiveness.
- How about "Given x, find the α that generated it in $\mathcal{M}$" ?



# Difficulties with the Transform




**4 Major Questions**

- Is $\hat{\underline{\alpha}} = \underline{\alpha}$ ? Under which conditions?
- Are there practical ways to get $\hat{\underline{\alpha}}$ ?
- How effective are those ways?
- How would we get **D**?

# Why Is It Interesting?

Several recent trends from signal/image processing worth looking at:

- JPEG to JPEG2000 - From (L ... let and non-linear approximati...
- From ... (Fourier) to $L_1$. ... Sparsity.
- From u... ift-invariar... ompleteness.
- Approxi... near approximation ➡ Spa... Overcompleteness.
- ICA and related models ➡ Independence and Sparsity.

**Sparseland is HERE !**

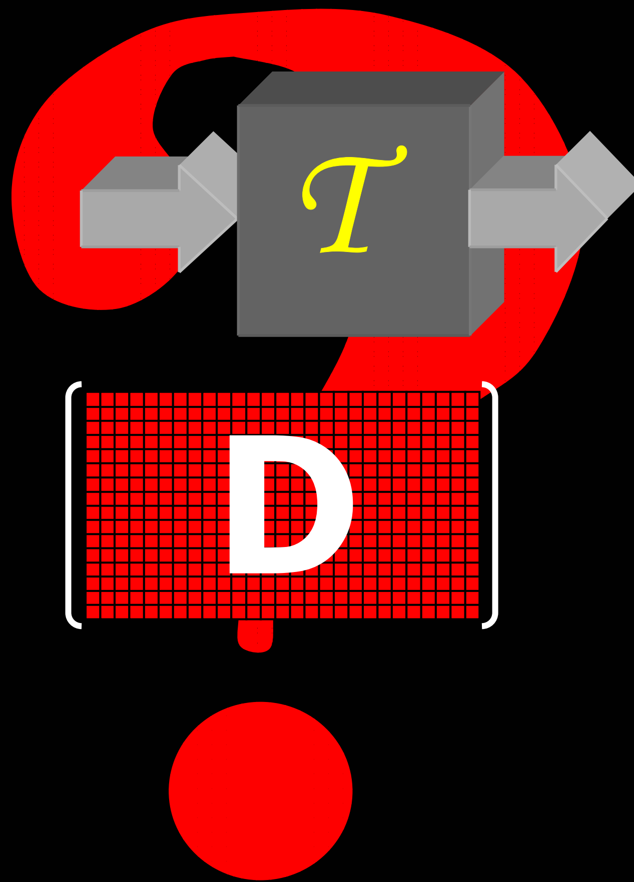
# Agenda

1. A Visit to *Sparseland*

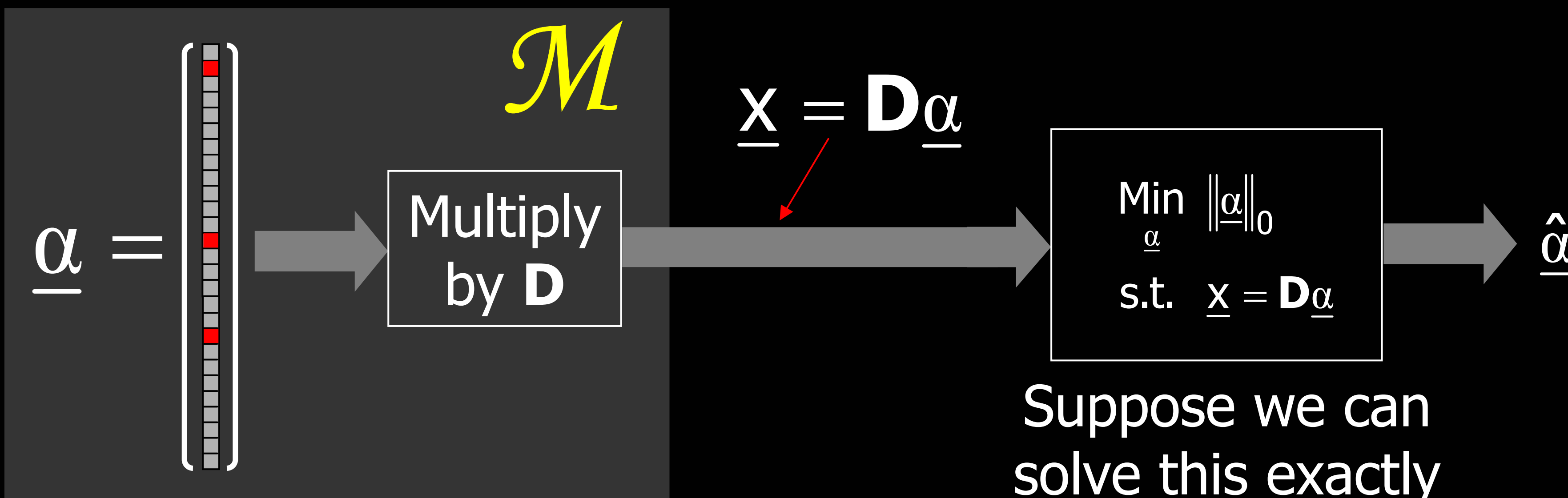
   Motivating Sparsity & Overcompleteness

2. **Answering the 4 Questions**

   **How & why should this work?**



# Question 1 – Uniqueness?




**Why should we necessarily get $\hat{\underline{\alpha}} = \underline{\alpha}$?**

**It might happen that eventually $\|\hat{\underline{\alpha}}\|_0 < \|\underline{\alpha}\|_0$.**

# Matrix "Spark"

**Definition:** Given a matrix **D**, $\sigma$=Spark{**D**} is the smallest number of columns that are linearly dependent.

Donoho & Elad ('02)

**Example:**

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

**Rank = 4**

**Spark = 3**

# Uniqueness Rule

Suppose this problem has been solved somehow

$$\underset{\underline{\alpha}}{\text{Min}} \left\| \underline{\alpha} \right\|_0 \quad \text{s.t.} \quad \underline{x} = \mathbf{D}\underline{\alpha}$$

**Uniqueness**

Donoho & Elad ('02)

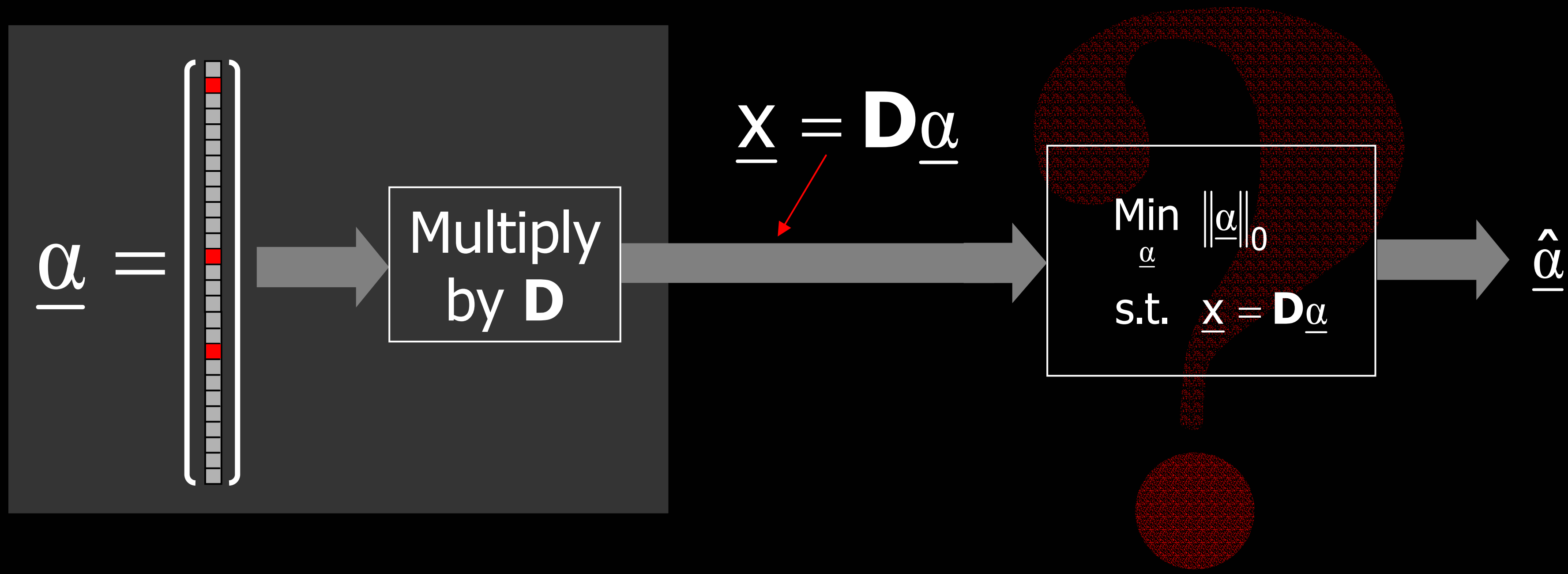
If we found a representation that satisfy

$$\frac{\sigma}{2} > \left\| \underline{\alpha} \right\|_0$$

then necessarily it is unique (the sparsest).

This result implies that if $\mathcal{M}$ generates vectors using "sparse enough" $\underline{\alpha}$, the solution of the above will find it exactly.

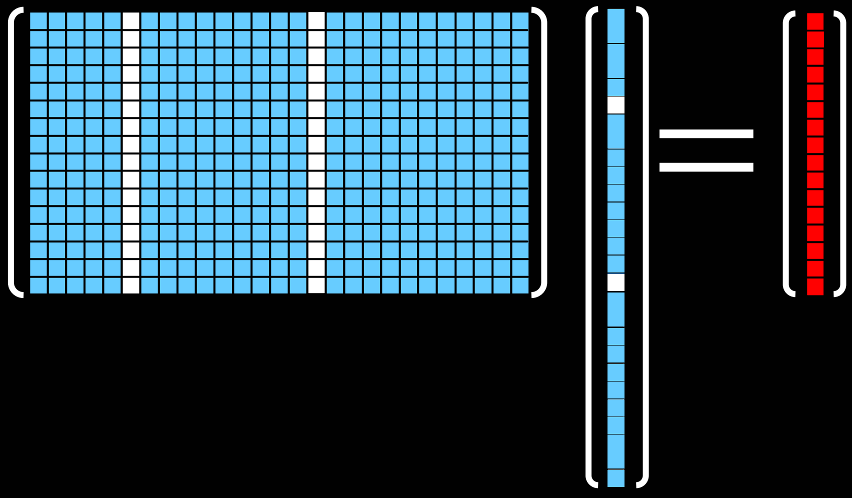
# Question 2 – Practical $P_0$ Solver?




**Are there reasonable ways to find $\hat{\underline{\alpha}}$ ?**

# Matching Pursuit (MP) Mallat & Zhang (1993)

- The MP is a greedy algorithm that finds one atom at a time.
- Step 1: find the one atom that best matches the signal.
- Next steps: given the previously found atoms, find the next **one** to best fit …



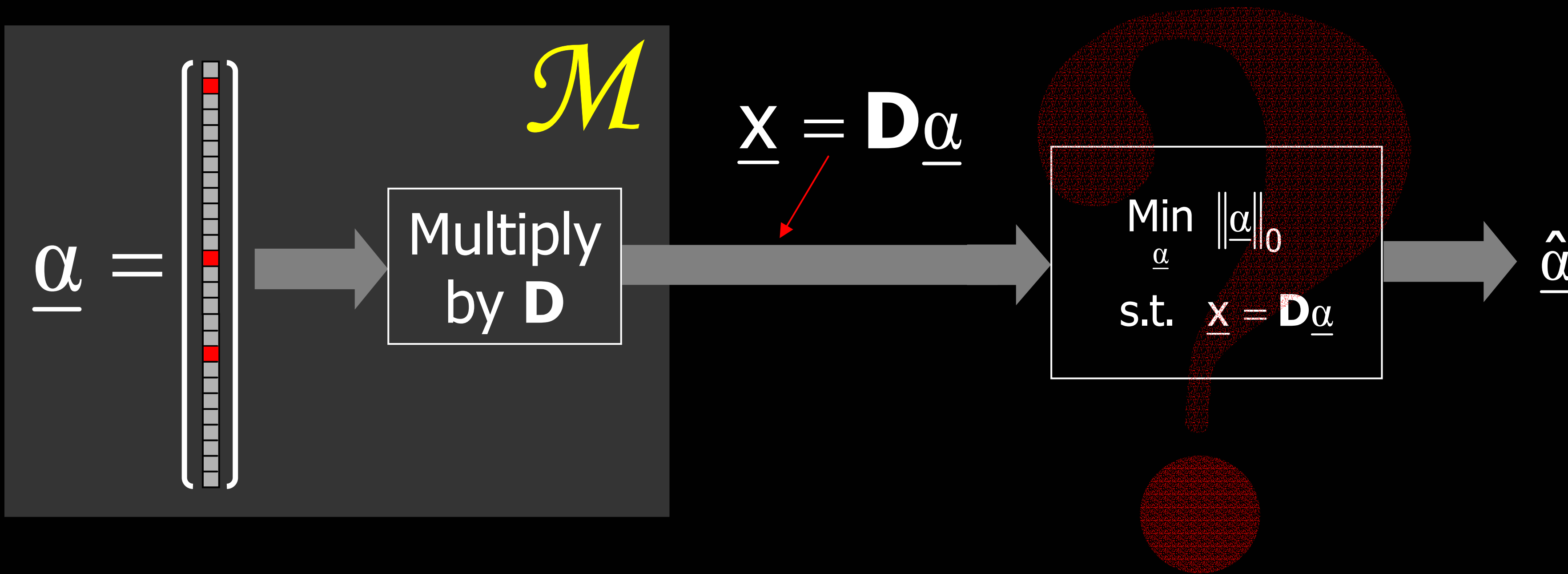
- The Orthogonal MP (OMP) is an improved version that re-evaluates the coefficients after each round.

# Basis Pursuit (BP) Chen, Donoho, & Saunders (1995)

| Instead of solving | | Solve Instead |
|---|---|---|
| $\underset{\underline{\alpha}}{\text{Min}} \left\| \underline{\alpha} \right\|_0 \quad \text{s.t.} \quad \underline{x} = \mathbf{D}\underline{\alpha}$ | → | $\underset{\underline{\alpha}}{\text{Min}} \left\| \underline{\alpha} \right\|_1 \quad \text{s.t.} \quad \underline{x} = \mathbf{D}\underline{\alpha}$ |

- The newly defined problem is convex.
- It has a Linear Programming structure.
- Very efficient solvers can be deployed:
  - Interior point methods [Chen, Donoho, & Saunders (`95)] ,
  - Sequential shrinkage for union of ortho-bases [Bruce et.al. (`98)],
  - If computing $\mathbf{D}\underline{x}$ and $\mathbf{D}^T\underline{\alpha}$ are fast, based on shrinkage [Elad (`05)].

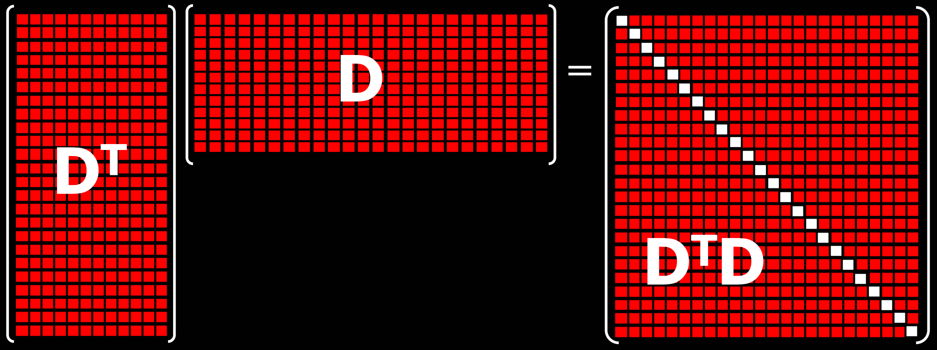
# Question 3 – Approx. Quality?




**How effective are the MP/BP in finding $\hat{\underline{\alpha}}$ ?**

# Mutual Coherence

- Compute




Assume normalized columns

- The Mutual Coherence $\mu$ is the largest entry in absolute value outside the main diagonal of $\mathbf{D}^T\mathbf{D}$.

- The Mutual Coherence is a property of the dictionary (just like the "Spark"). The smaller it is, the better the dictionary.

# BP and MP Equivalence

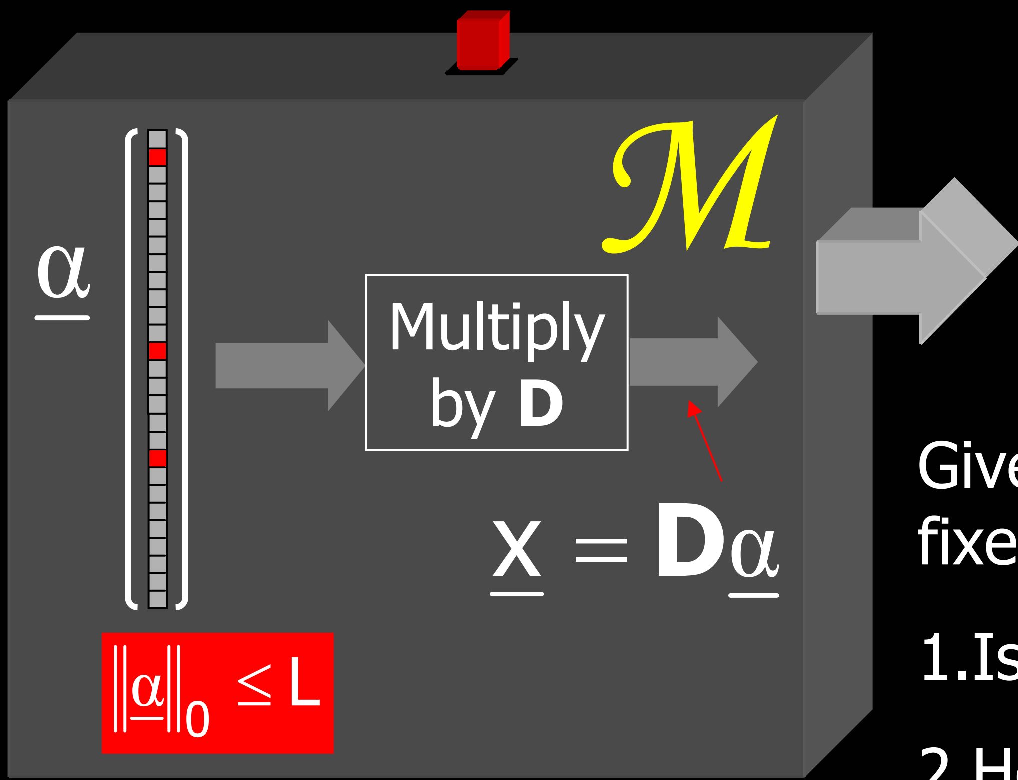
**Equivalence**

Donoho & Elad ('02)
Gribonval & Nielsen ('03)
Tropp ('03)
Temlyakov ('03)

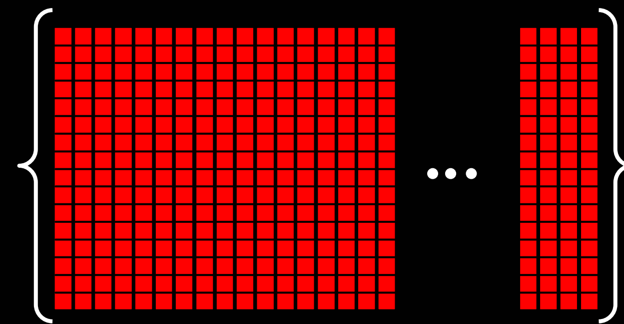
Given a vector $\underline{x}$ with a representation $\underline{x} = \mathbf{D}\underline{\alpha}$, Assuming that $\|\underline{\alpha}\|_0 < 0.5(1+1/\mu)$, BP and MP are **Guaranteed** to find the sparsest solution.

- MP is typically inferior to BP!
- The above result corresponds to the worst-case.
- Average performance results are available too, showing much better bounds [Donoho (`04), Candes et.al. (`04), Elad and Zibulevsky (`04)].

# Question 4 – Finding D?




$$\left\{ \underline{X}_j \right\}_{j=1}^{P}$$



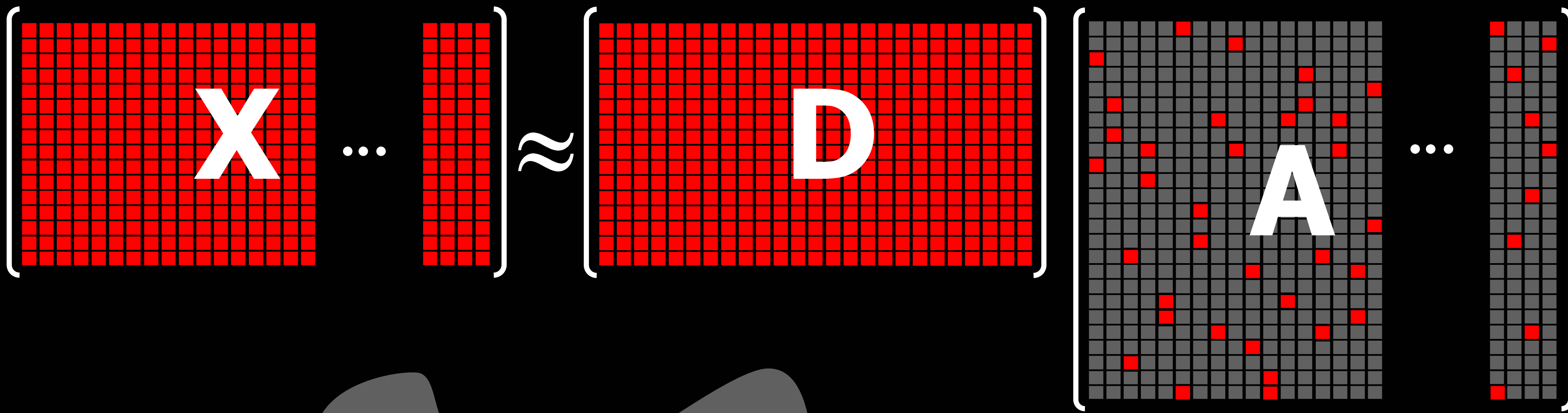
Given these P examples and a fixed size [N×K] dictionary **D**:

1. Is **D** unique? **(Yes)**
2. How to find **D**?

Train!! The **K-SVD** algorithm

Skip?

# Training: The Objective




$$\min_{\mathbf{D},\mathbf{A}} \sum_{j=1}^{P} \left\| \mathbf{D}\underline{\alpha}_j - \underline{x}_j \right\|_2^2 \quad \text{s.t.} \quad \forall j, \left\| \underline{\alpha}_j \right\|_0 \le L$$

Each example is a linear combination of atoms from **D**

Each example has a sparse representation with no more than L atoms

# The K–SVD Algorithm

Aharon, Elad, & Bruckstein (`04)

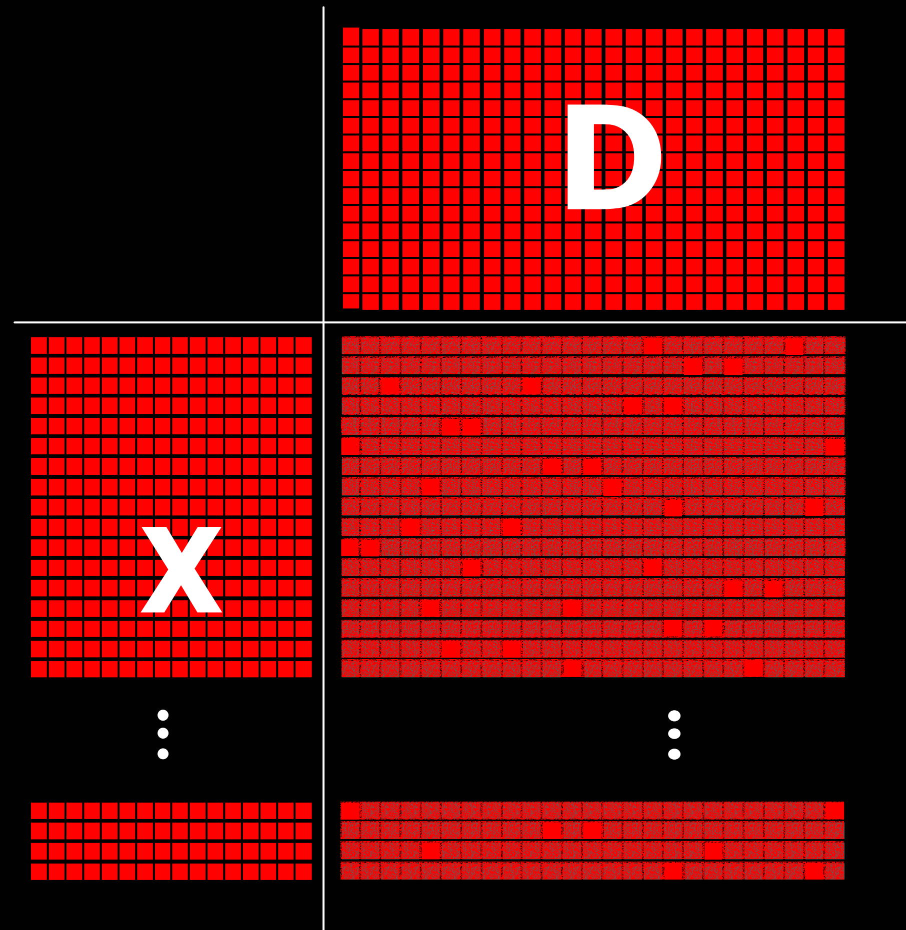
Initialize
**D**

Sparse Coding

Use MP or BP

Dictionary Update

Column-by-Column by **SVD computation**



# Today We Discussed

1. **A Visit to** *Sparseland*

   **Motivating Sparsity & Overcompleteness**

2. **Answering the 4 Questions**

   **How & why should this work?**

# Summary

Sparsity and Over-completeness are important ideas that can be used in designing better tools in data/signal/image processing

There are difficulties in using them!

We are working on resolving those difficulties:

- Performance of pursuit alg.
- Speedup of those methods,
- Training the dictionary,
- Demonstrating applications,
- …

The dream?

Future transforms and regularizations will be data-driven, non-linear, overcomplete, and promoting sparsity.