The K–SVD

Design of Dictionaries for Redundant and Sparse Representation of Signals *

Michael Elad
The Computer Science Department
The Technion – Israel Institute of technology
Haifa 32000, Israel

SPIE – Wavelets XI
San-Diego: August 2nd, 2005

* Joint work with

Michal Aharon

Freddy Bruckstein

Agenda

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

   Motivating redundancy & Sparsity

2. The Quest for a Dictionary – Fundamentals

   Common Approaches?

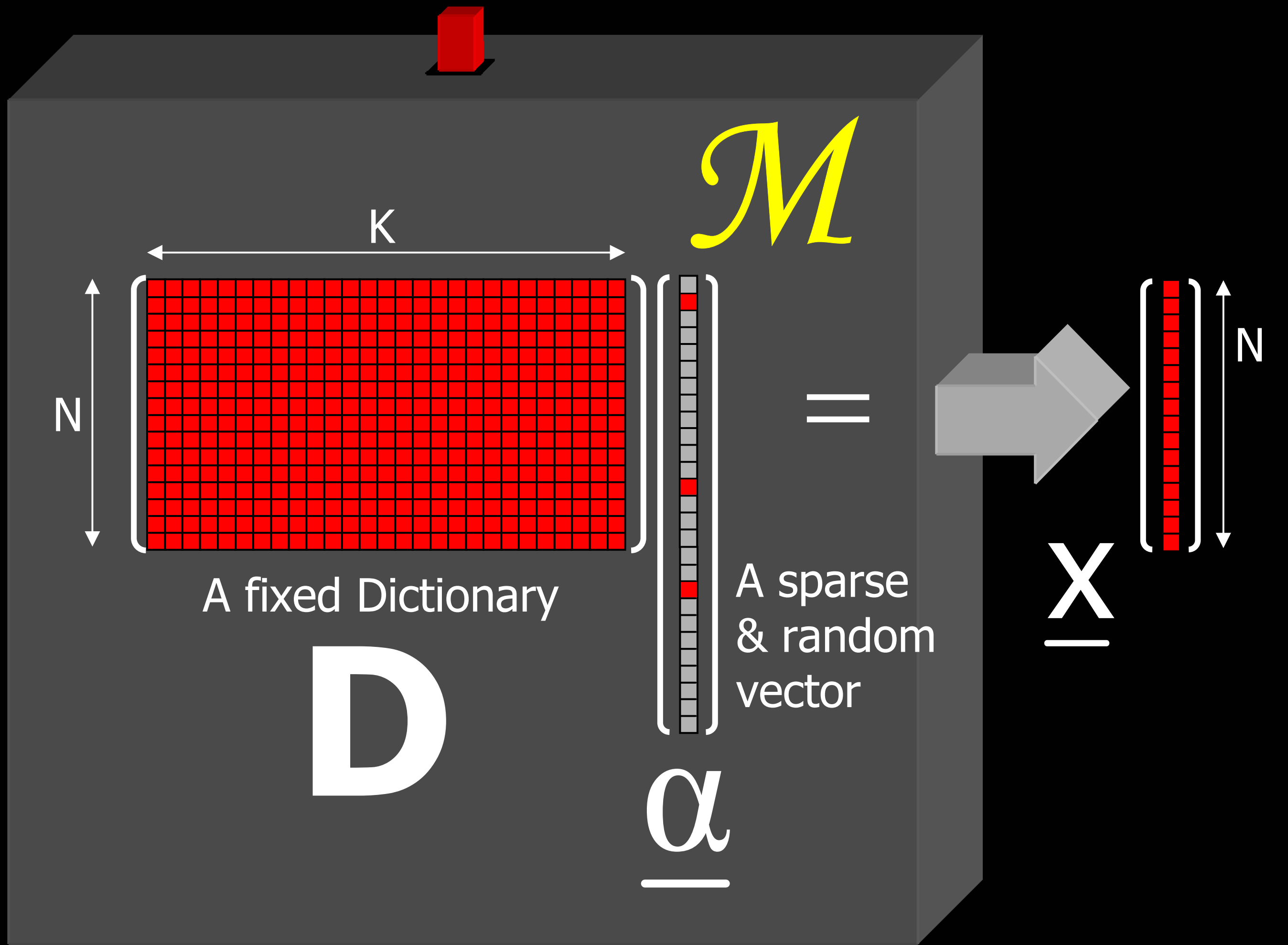
3. The Quest for a Dictionary – Practice

   Introducing the K-SVD

4. Results

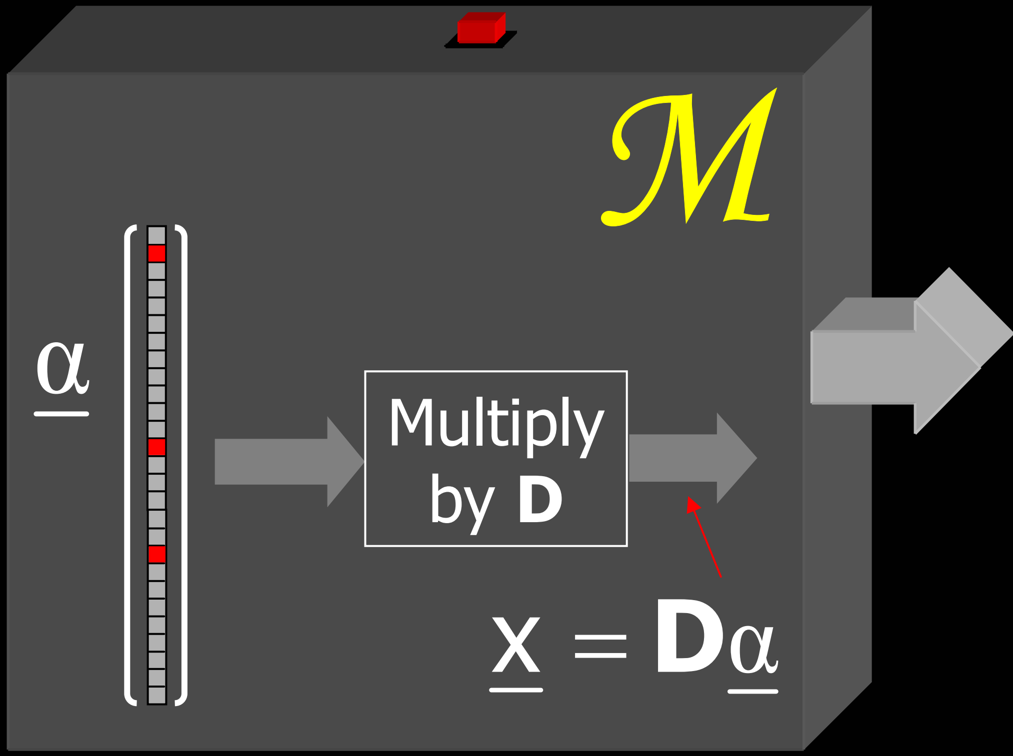
   Preliminary results and applications

Generating Signals in *Sparseland*



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

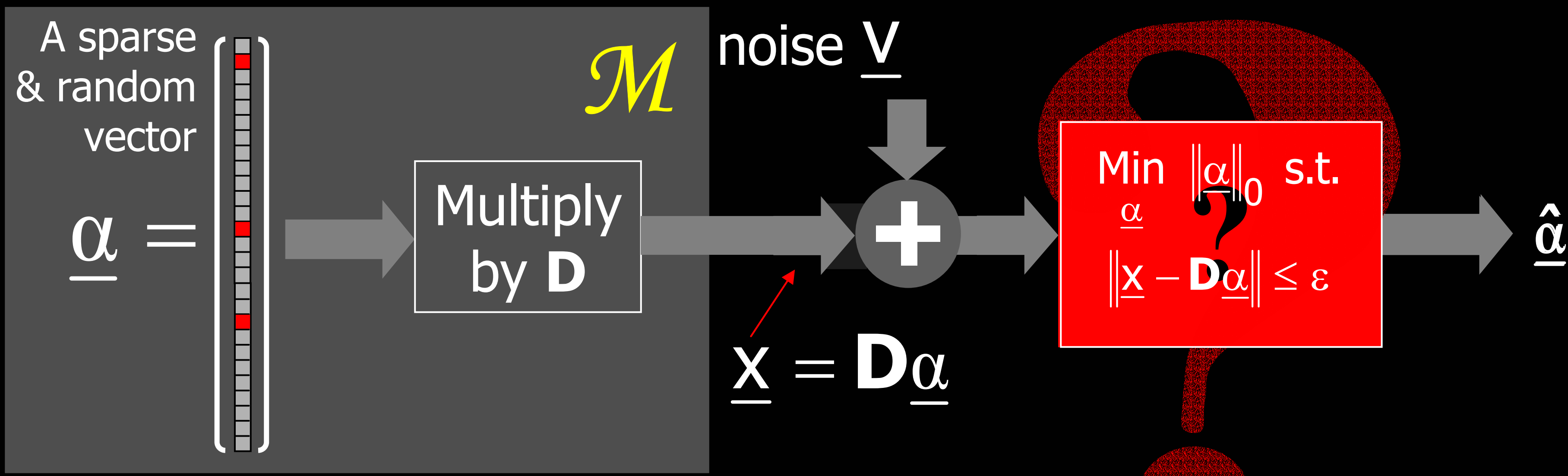
Sparseland Signals Are Interesting



- **Simple:** Every generated signal is built as a linear combination of **few** atoms from our dictionary **D**
- **Rich:** A general model: the obtained signals are a special type mixture-of-Gaussians (or Laplacians).
- **Popular:** Recent work on signal and image processing adopt this model and successfully deploys it to applications.

→ ***Sparseland* is here !?**

Signal Processing in *Sparseland*



4 Major Questions

- Is $\hat{\underline{\alpha}} = \underline{\alpha}$ or even close?
- Practical ways to get $\hat{\underline{\alpha}}$?
- How effective?

Recent results give optimistic answers to these questions

- How do we get **D**? **OUR FOCUS TODAY!!**

Agenda

1. A Visit to *Sparseland*

   Motivating redundancy & Sparsity

2. **The Quest for a Dictionary – Fundamentals**

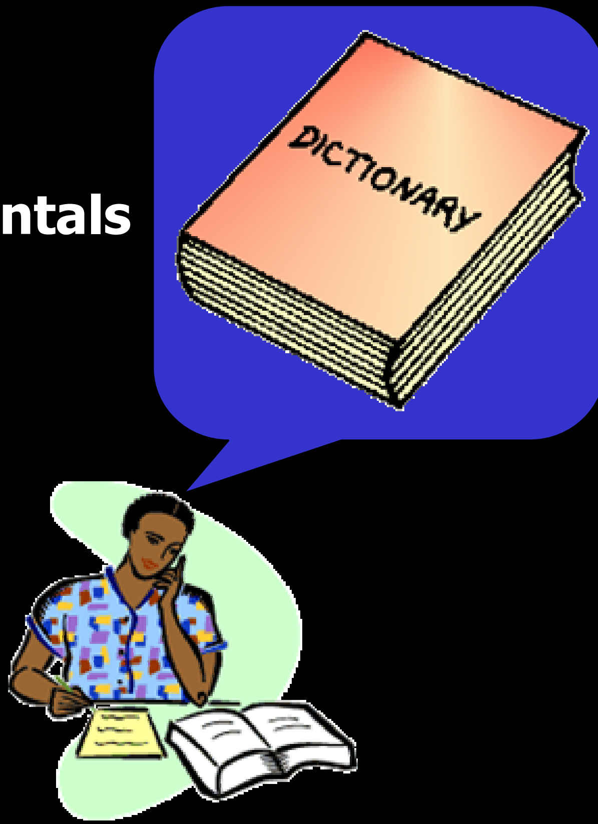
   Common approaches?

3. The Quest for a Dictionary – Practice

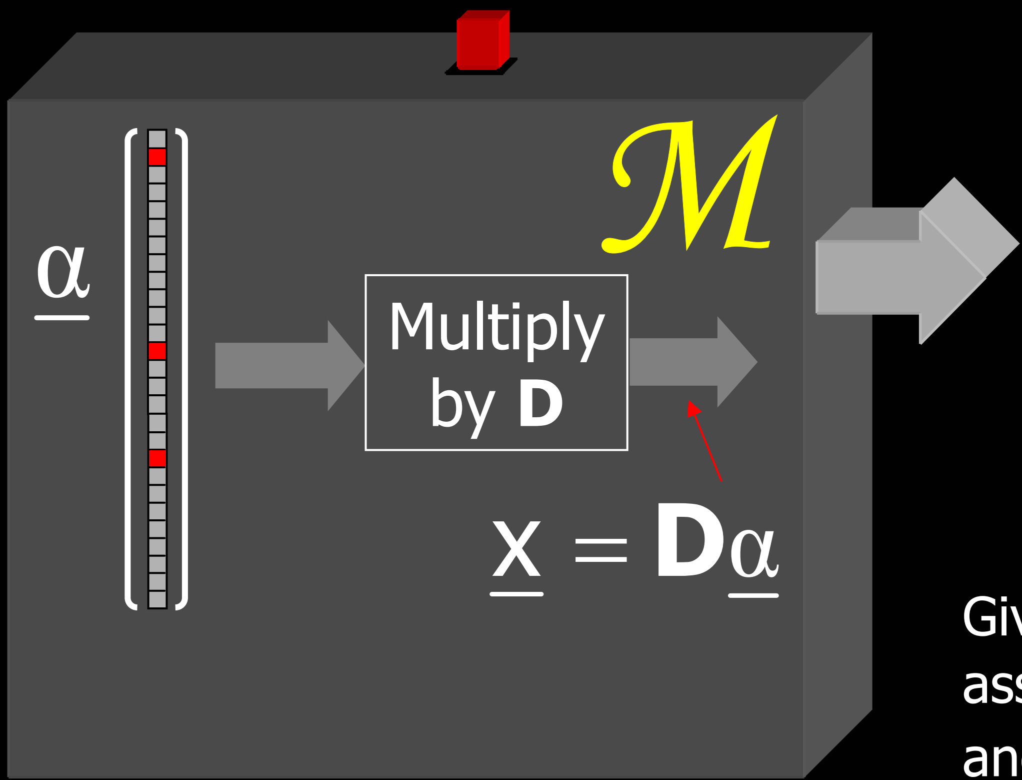
   Introducing the K-SVD

4. Results

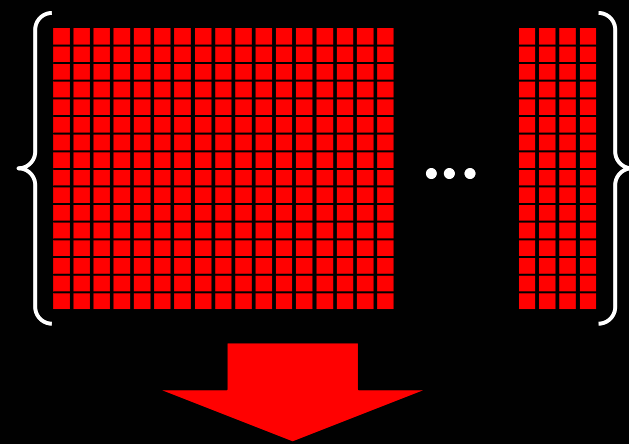
   Preliminary results and applications



Problem Setting



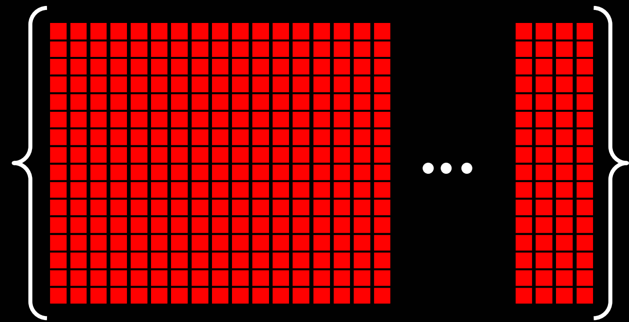
$$\left\{ \underline{X}_j \right\}_{j=1}^{P}$$



Given a set of data examples, we assume the *Sparseland* model and ask

$$\mathbf{D} = ?$$

Choose D – Modeling Approach



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

Replace the model with another, well-defined simple mathematical, model (e.g. images as piece-wise C^2 smooth regions with C^2 smooth edges) and fit a dictionary accordingly, based on existing methods.

Examples: Curvelet [Candes & Donoho]
Contourlet [Do & Vetterli]
Bandlet [Mallat & Le-Pennec]
and others ...

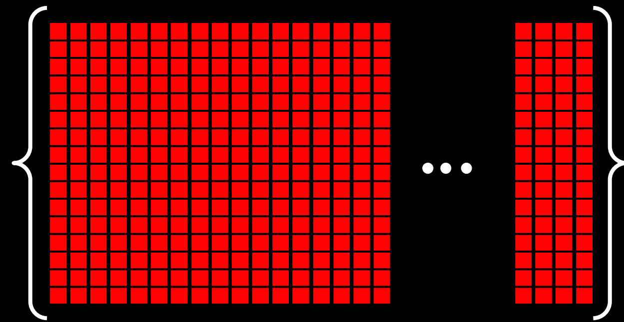
Pros:

- Build on existing methods,
- Fast transforms,
- Proven optimality for the model.

Cons:

- Relation to *Sparseland*? Linearity?
- How to adapt to other signals?
- Bad for "smaller" signal families.

Choose D – Training Approach



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

Train the dictionary directly based on the given examples, optimizing w.r.t. sparsity and other desired properties (normalized atoms, etc.).

Examples: ML (H
MAP [Le
MOD [Eng
ICA-like [Kreu l.]
and others ...

OUR APPROACH FOR TODAY

Pro

e true objectives,

dapted to small families.

No fast version,

- Slow training,
- Scalability issues?

Agenda

1. A Visit to *Sparseland*

   Motivating redundancy & Sparsity

2. The Quest for a Dictionary – Fundamentals

   Common Approaches?

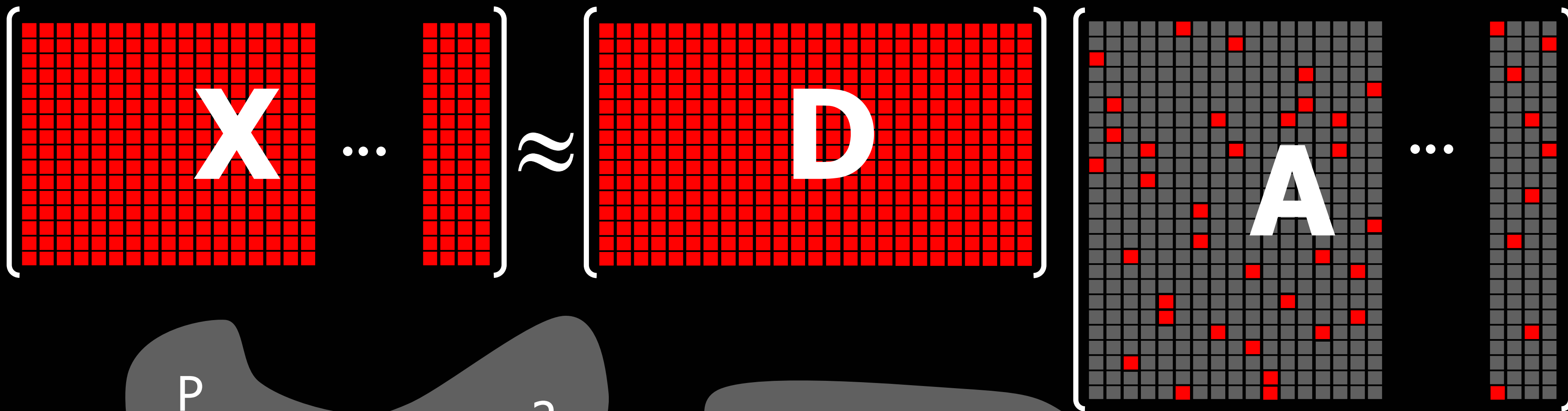
3. **The Quest for a Dictionary – Practice**

   Introducing the K-SVD

4. Results

   Preliminary results and applications

Practical Approach – 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

(n,K,L are assumed known)

Field & Olshausen (96')
Engan et. al. (99')
Lewicki & Sejnowski (00')
Cotter et. al. (03')
Gribonval et. al. (04')

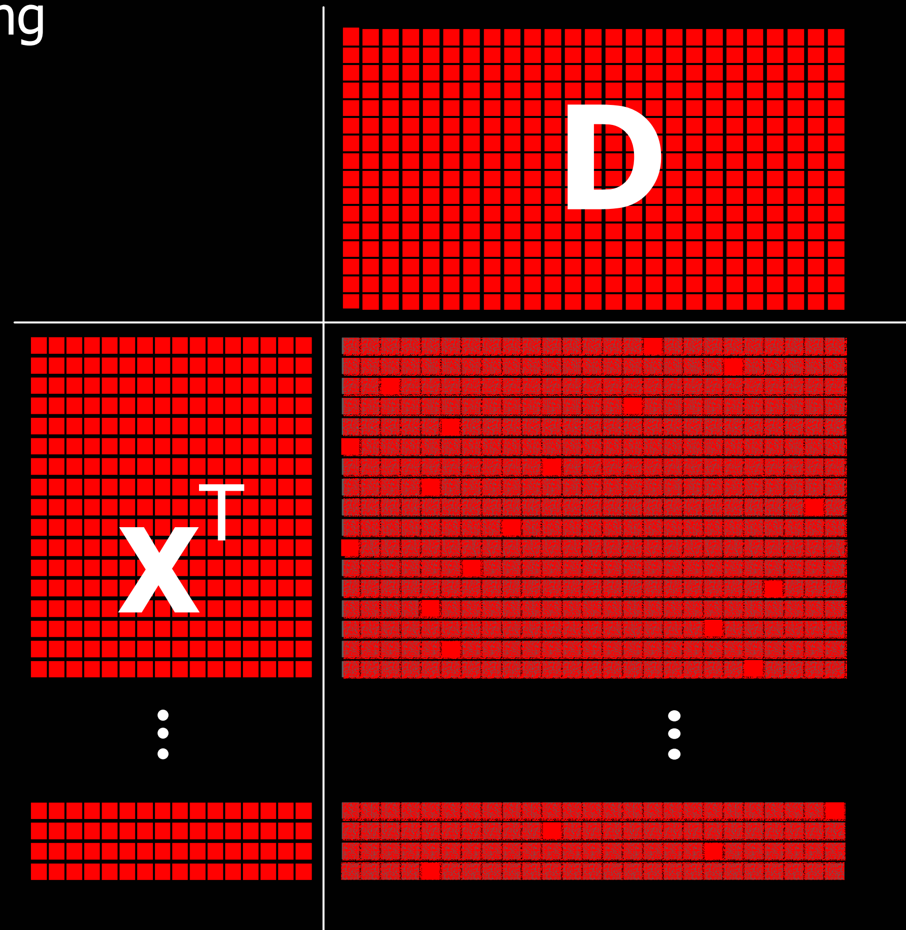
K–Means For Clustering

Clustering: An extreme sparse coding

Initialize
D

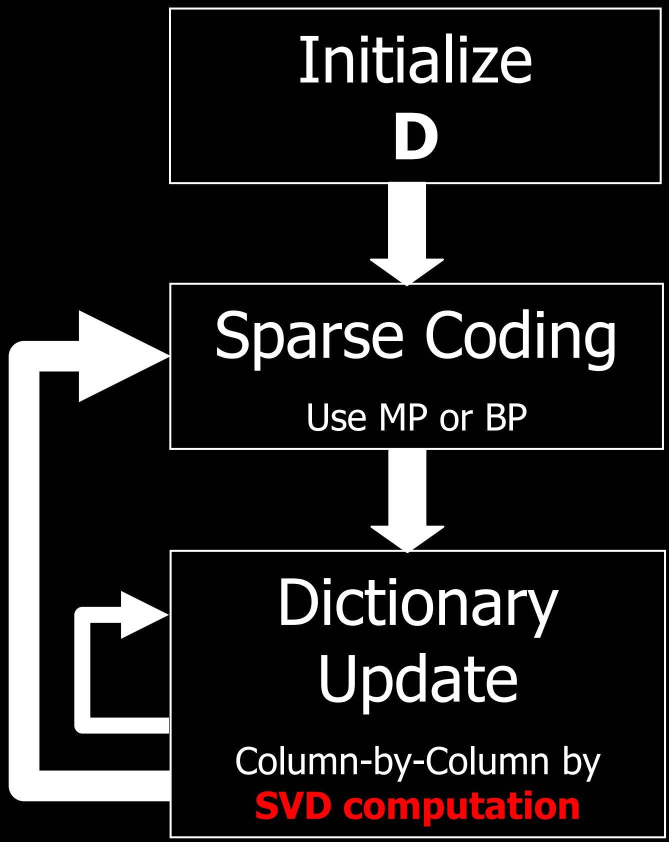
Sparse Coding
Nearest Neighbor

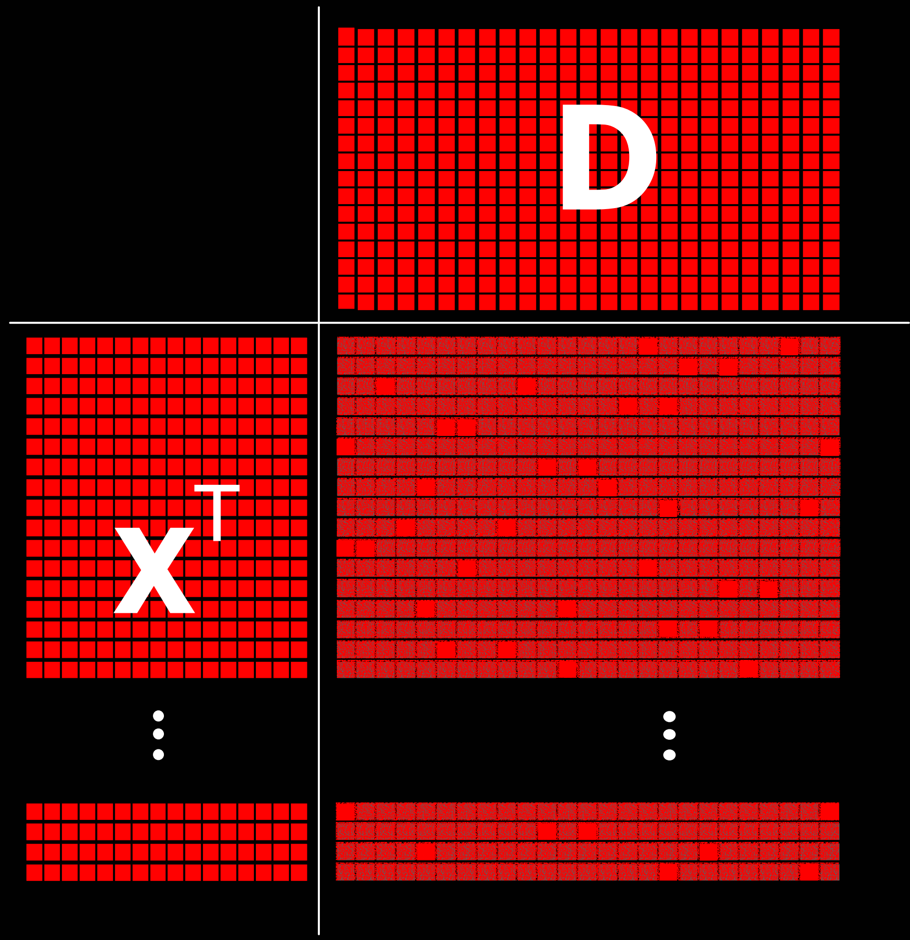
Dictionary Update
Column-by-Column by
Mean computation over the relevant examples



The K–SVD Algorithm – General

Aharon, Elad, & Bruckstein (`04)





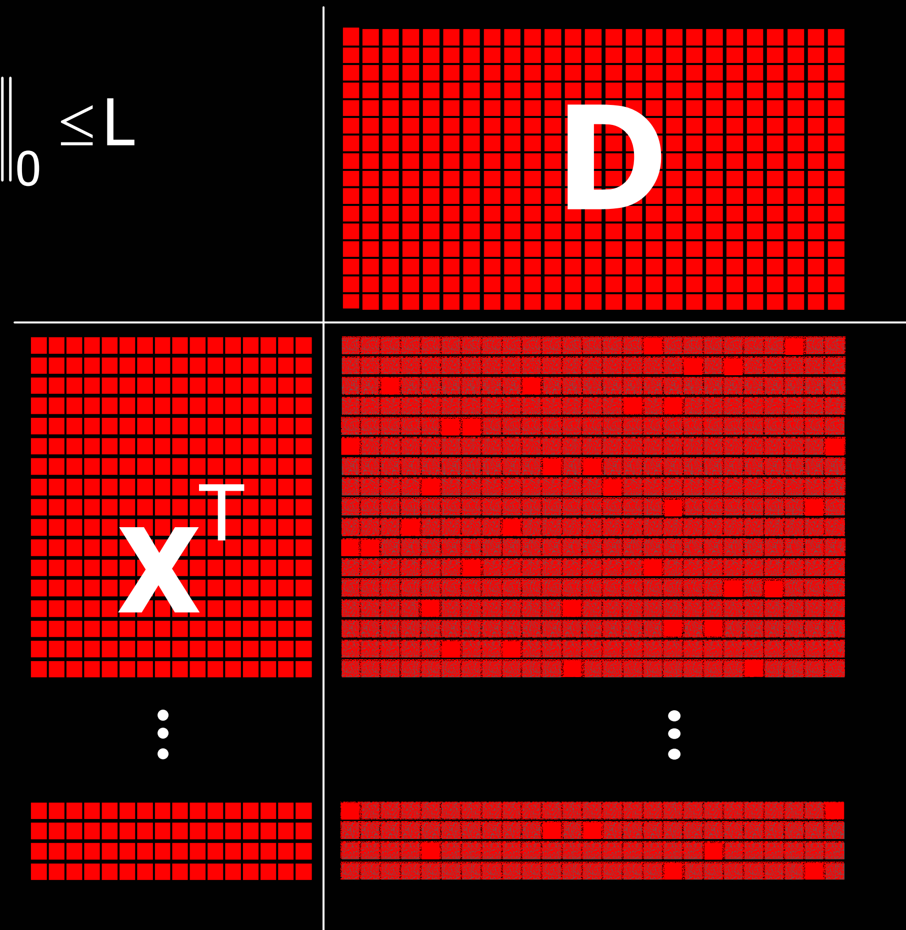
K–SVD: Sparse Coding Stage

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

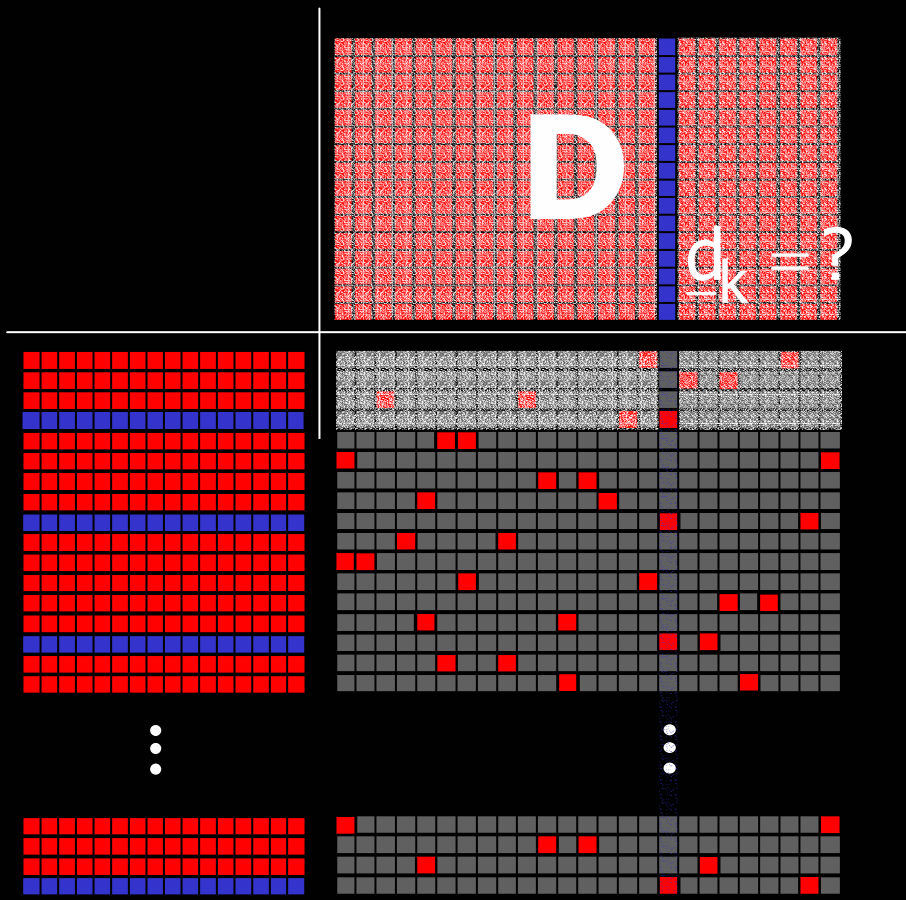
D is known!
For the j[th] item
we solve

$$\underset{\underline{\alpha}}{\text{Min}} \left\| \mathbf{D}\underline{\alpha} - \underline{x}_j \right\|_2^2 \quad \text{s.t.} \quad \left\| \underline{\alpha} \right\|_0 \le L$$

Pursuit Problem !!!



K–SVD: Dictionary Update Stage



G_k : The examples in $\{\underline{x}_j\}_{j=1}^{P}$ that use the column $\underline{d}_k$.

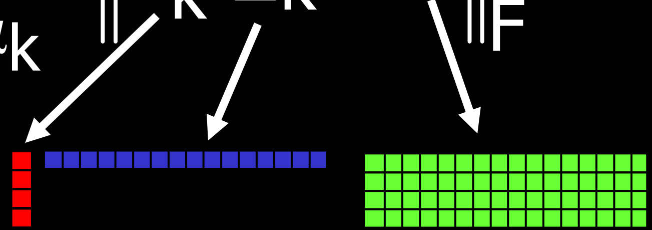
The content of $\underline{d}_k$ influences only the examples in G_k.

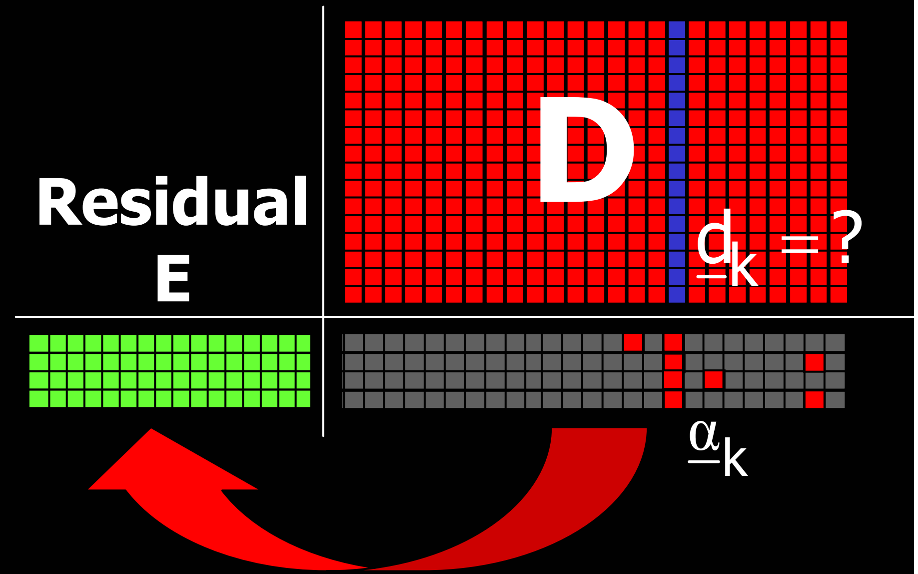
Let us fix all **A** and **D** apart from the k^{th} column and seek both $\underline{d}_k$ and the k^{th} column in **A** to better fit the **residual**!

K–SVD: Dictionary Update Stage

We should solve:

$$\min_{\underline{d}_k,\alpha_k} \left\| \alpha_k^T \underline{d}_k - \mathbf{E} \right\|_F^2$$





$\underline{d}_k$ is obtained by **SVD** on the examples' residual in G_k.

Agenda

1. A Visit to *Sparseland*

   Motivating redundancy & Sparsity

2. The Quest for a Dictionary – Fundamentals

   Common Approaches?

3. The Quest for a Dictionary – Practice

   Introducing the K-SVD

4. **Results**

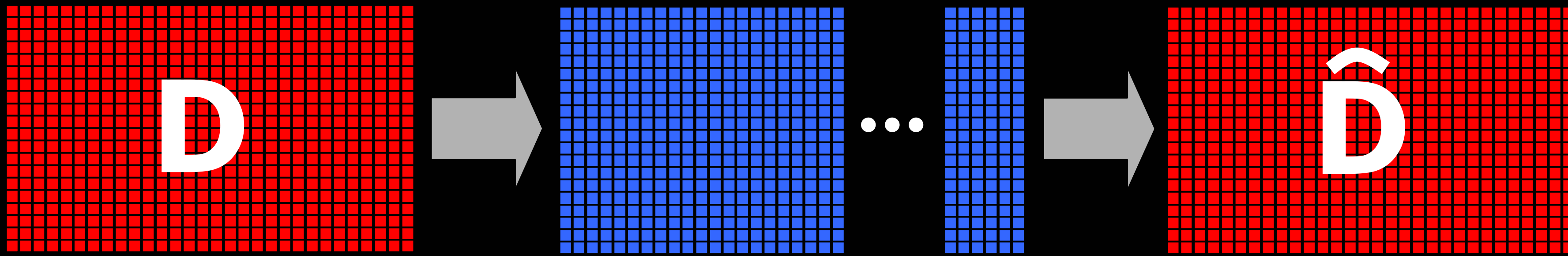
   Preliminary results and applications

K–SVD: A Synthetic Experiment

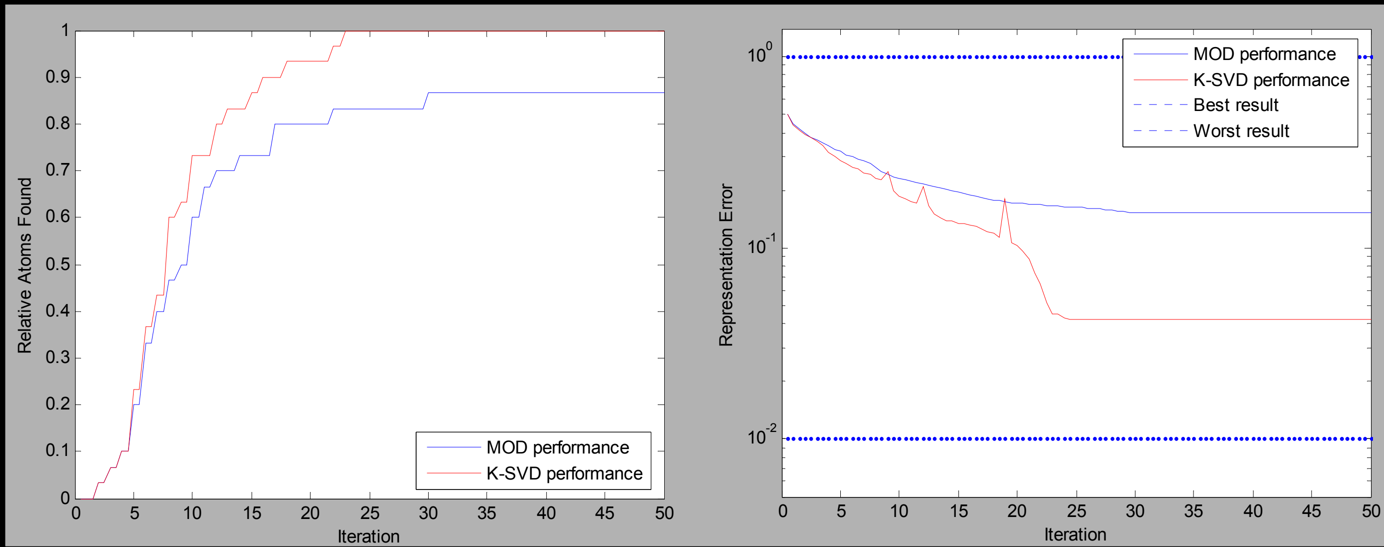
Create A 20×30 random dictionary with normalized columns

Generate 2000 signal examples with 3 atoms per each and add noise

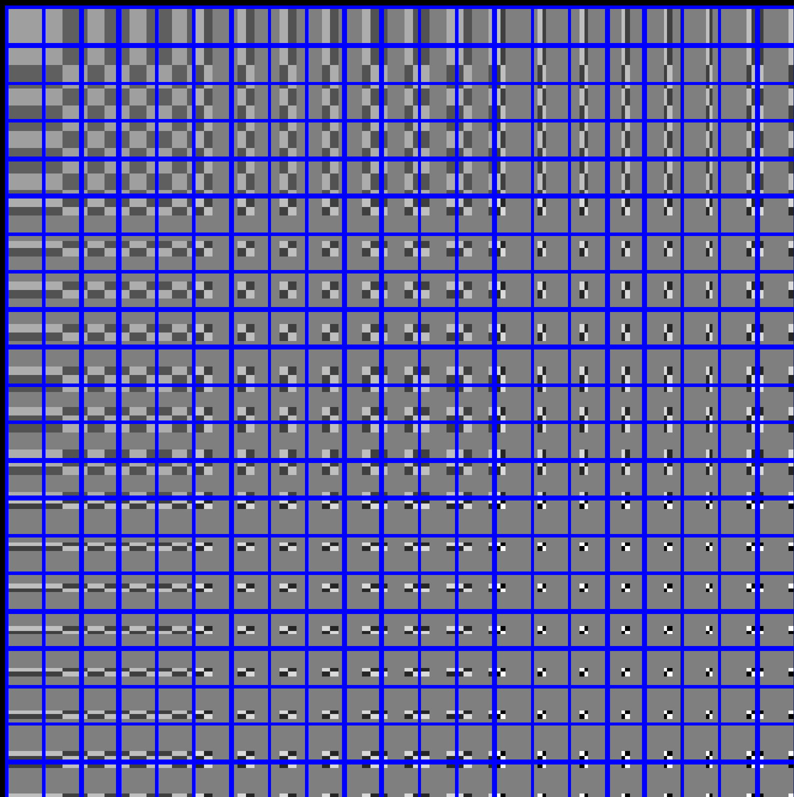
Train a dictionary using the KSVD and MOD and compare



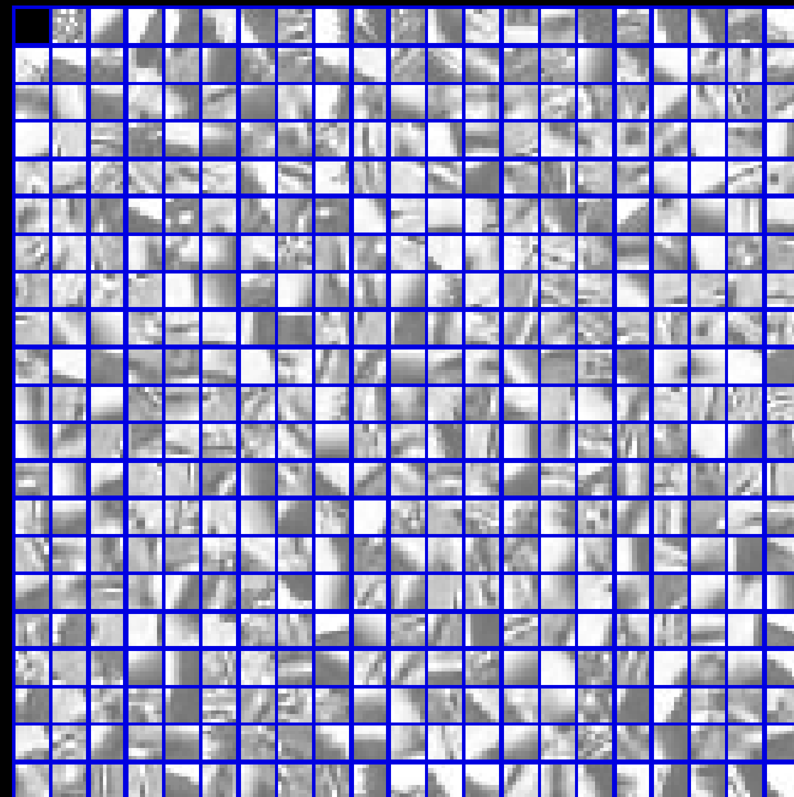
Results



K–SVD on Images



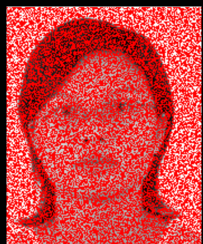
Overcomplete Haar



K-SVD: 10,000 sample 8-by-8 images.
441 dictionary elements.
Approximation method: OMP

Filling–In Missing Pixels

Given an image with missing values



Apply pursuit (per each block of size(8×8) using a decimated dictionary with rows removed

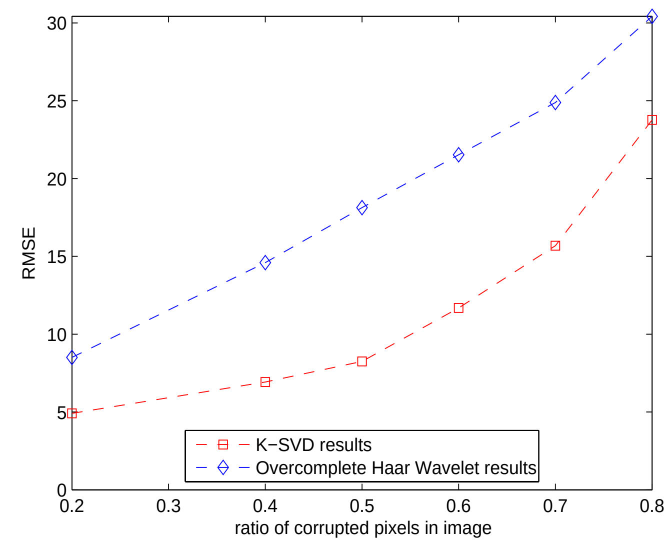
Multiply the found representation by the complete dictionary

Get the recovered image

60% missing pixels

K-SVD Results
Average # coefficients 4.08
RMSE: 11.68

Haar Results
Average # coefficients 4.42
RMSE: 21.52



Summary

Today we discussed:

1. A Visit to *Sparseland*
   Motivating redundancy & Sparsity
2. The Quest for a Dictionary – Fundamentals
   Common Approaches?
3. The Quest for a Dictionary – Practice
   Introducing the K-SVD
4. Results
   Preliminary results and applications

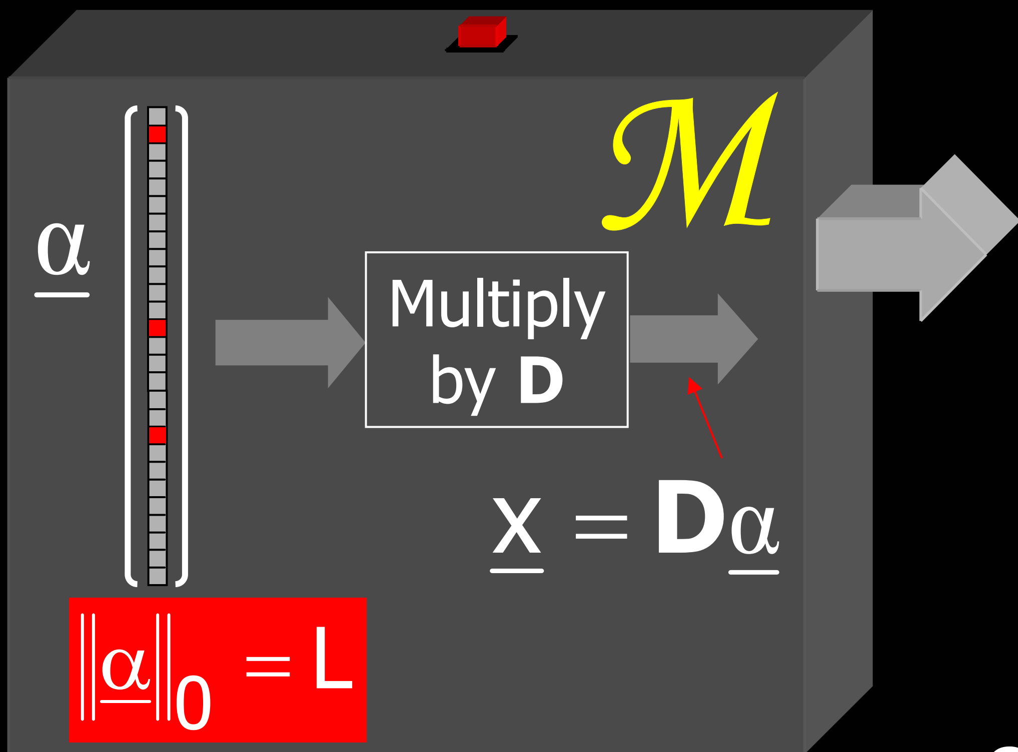
Open Questions:

1. Scalability – treatment of bigger blocks and large images.
2. Uniqueness? Influence of noise?
3. Equivalence? A guarantee to get the perfect dictionary?
4. Choosing K? What forces govern the redundancy?
5. Other applications? …

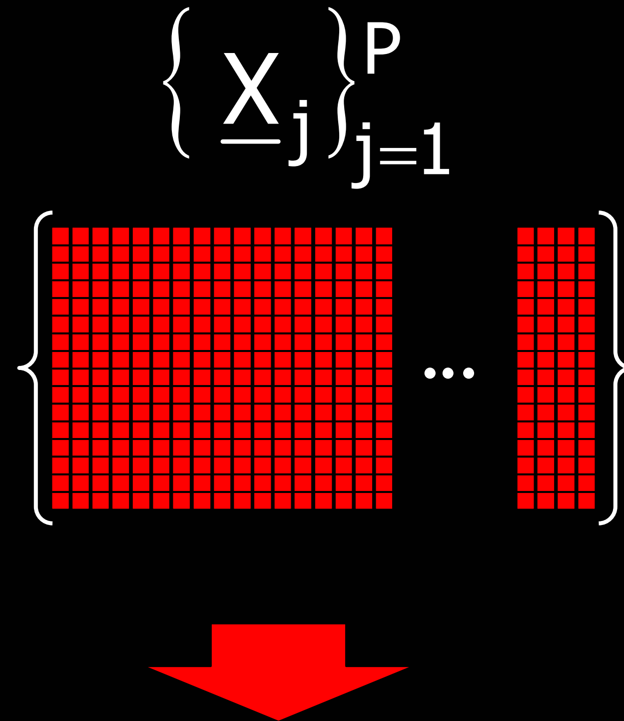
http://www.cs.technion.ac.il/~elad

Supplement Slides

Uniqueness?



(n,K,L are assumed known)



Is D unique in explaining the origin of the signals?

Uniqueness? YES !!

Uniqueness

Aharon, Elad, & Bruckstein (`05)

If $\left\{\underline{x}_j\right\}_{j=1}^{P}$ is rich enough* and if

$$L < \frac{\text{Spark}\left\{\mathbf{D}\right\}}{2}$$

then **D** is unique.

Comments:

- "Rich Enough": The signals from $\mathcal{M}$ could be clustered to $\binom{K}{L}$ groups that share the same support. At least L+1 examples per each are needed.
- This result is proved constructively, but the number of examples needed to pull this off is huge – we will show a far better method next.
- A parallel result that takes into account noise could be constructed similarly.

Naïve Compression

$$\text{BPP} = \frac{C\left(10 + \log K\right)}{N}$$

$$\text{BPP} = \frac{C \cdot 10}{N}$$

K-SVD dictionary

Haar dictionary

DCT dictionary

OMP with error bound

Naïve Compression

DCT
BPP = 1.01
RMSE = 8.14

K-SVD
BPP = 0.502
RMSE = 7.67

Haar
BPP = 0.784
RMSE = 8.41