Dictionary Learning

ICCV 2009

Bach, Mairal, Ponce, Sapiro

Restoration by Energy Minimization

Restoration/representation algorithms are often related to the minimization of an energy function of the form

$$f(\underline{x}) = \frac{1}{2}\|\underline{x} - \underline{y}\|_2^2 + \Pr(\underline{x})$$

Relation to measurements

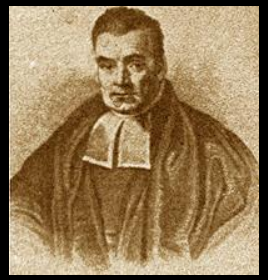
Prior or regularization

$\underline{y}$: Given measurements

$\underline{x}$: Unknown to be recovered

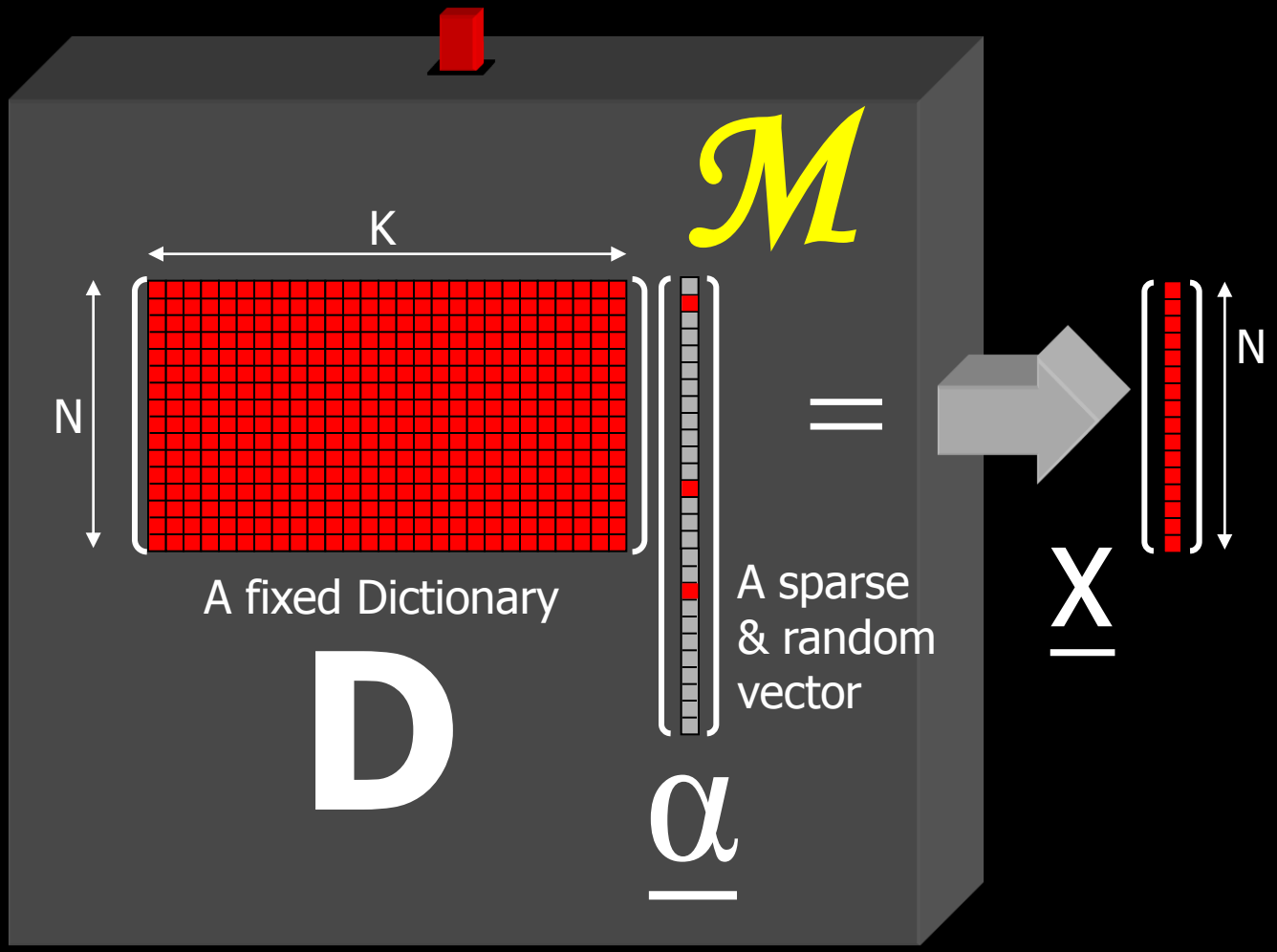
- Bayesian type of approach

- What is the prior? What is the image model?

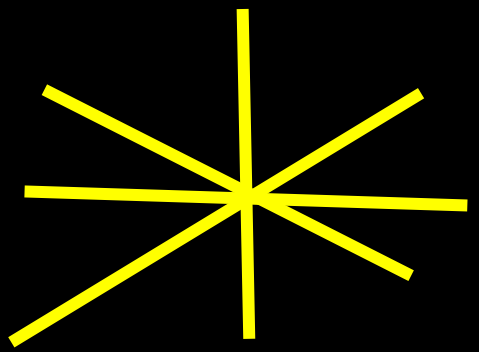


Thomas Bayes
1702 - 1761

The *Sparseland* Model for Images



- Every column in **D** (dictionary) is a prototype signal (Atom).
- The vector $\underline{\alpha}$ contains very few (say L) non-zeros.



What Should the Dictionary D Be?

$$\underline{\hat{\alpha}} = \arg\min_{\underline{\alpha}} \frac{1}{2} \left\| \mathbf{D}\underline{\alpha} - \underline{y} \right\|_2^2 \quad \text{s.t.} \quad \left\| \underline{\alpha} \right\|_0^0 \le L \quad \Rightarrow \quad \underline{\hat{x}} = \mathbf{D}\underline{\hat{\alpha}}$$

D should be chosen such that it sparsifies the representations

One approach to choose **D** is from a known set of transforms (Steerable wavelet, Curvelet, Contourlets, Bandlets, …)

Learn **D** :

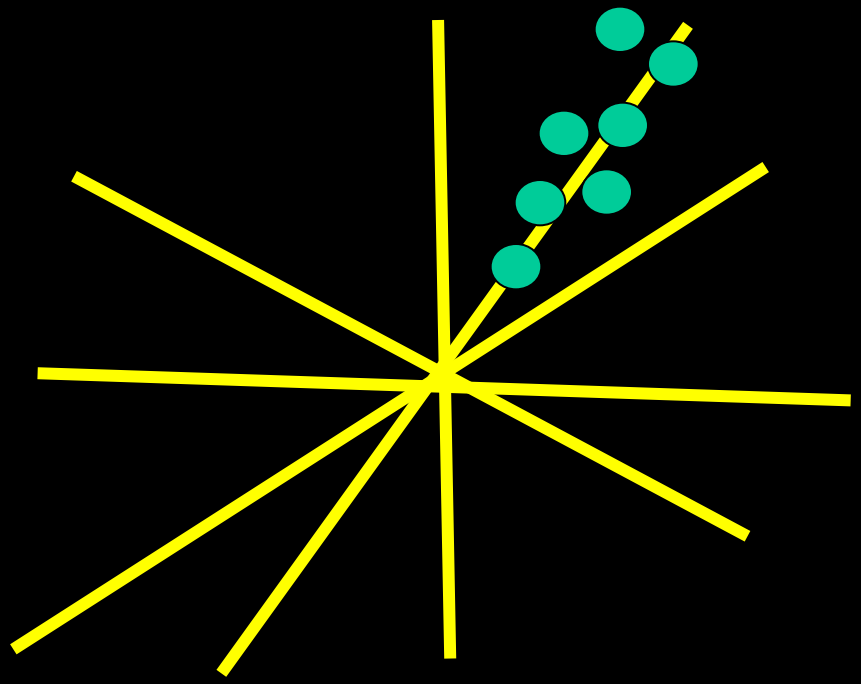
Multiscale Learning

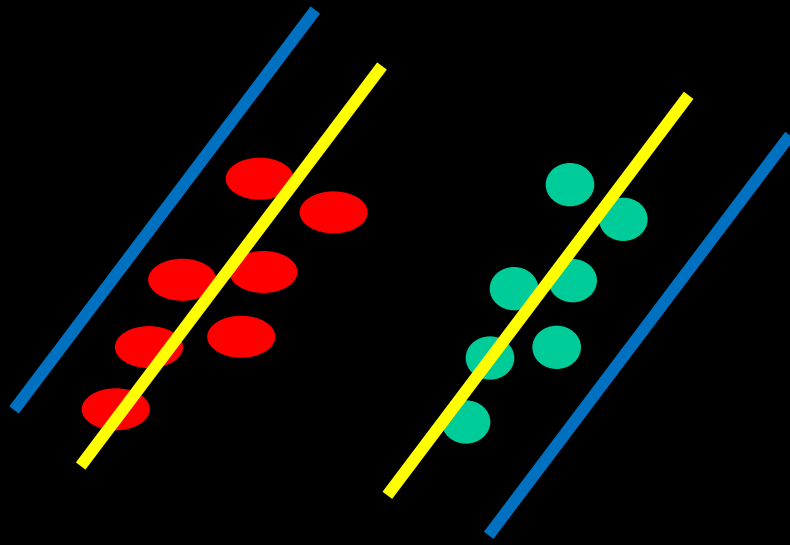
Color Image Examples

Task / sensing adapted

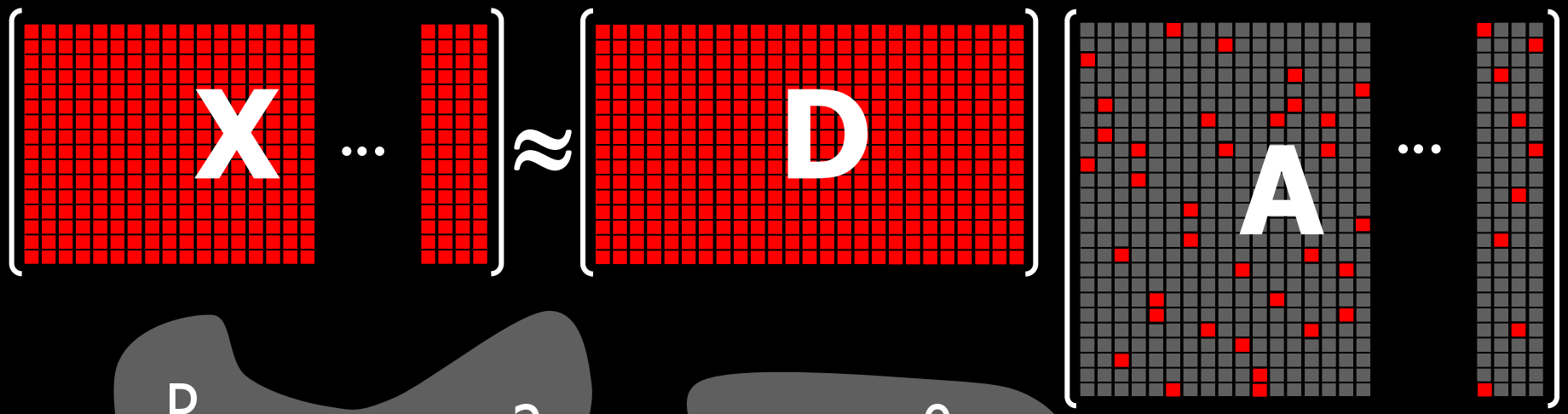
Internal structure

What is being learned?





Learning D to reconstruct



$$\min_{\mathbf{D},\mathbf{A}} \sum_{j=1}^{P} \left\| \mathbf{D}\underline{\alpha}_j - \underline{x}_j \right\|_2^2 \quad \text{s.t. } \forall j,\ \left\| \underline{\alpha}_j \right\|_0^0 \le L$$

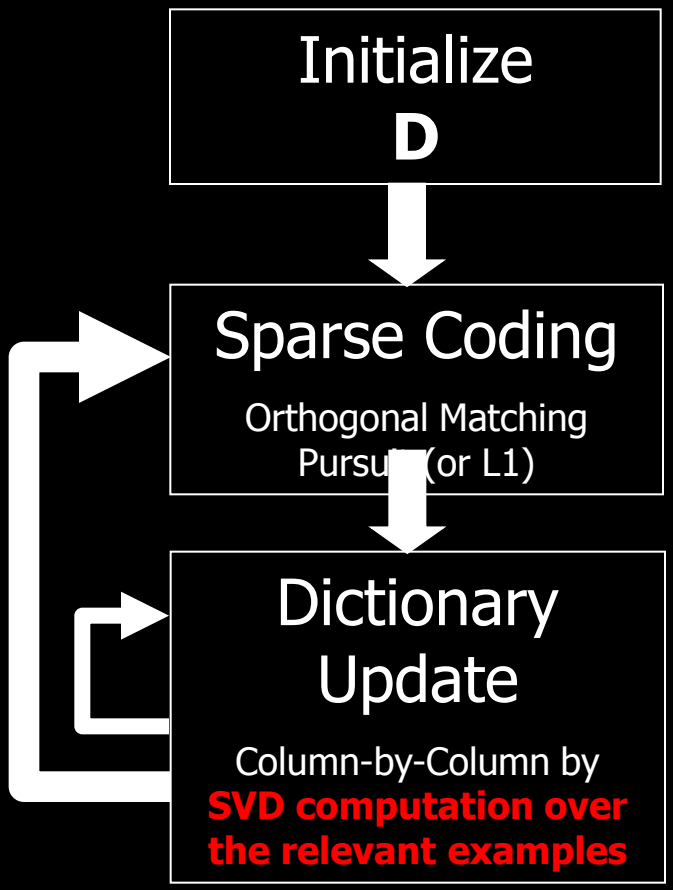
Each example is a linear combination of atoms from **D**

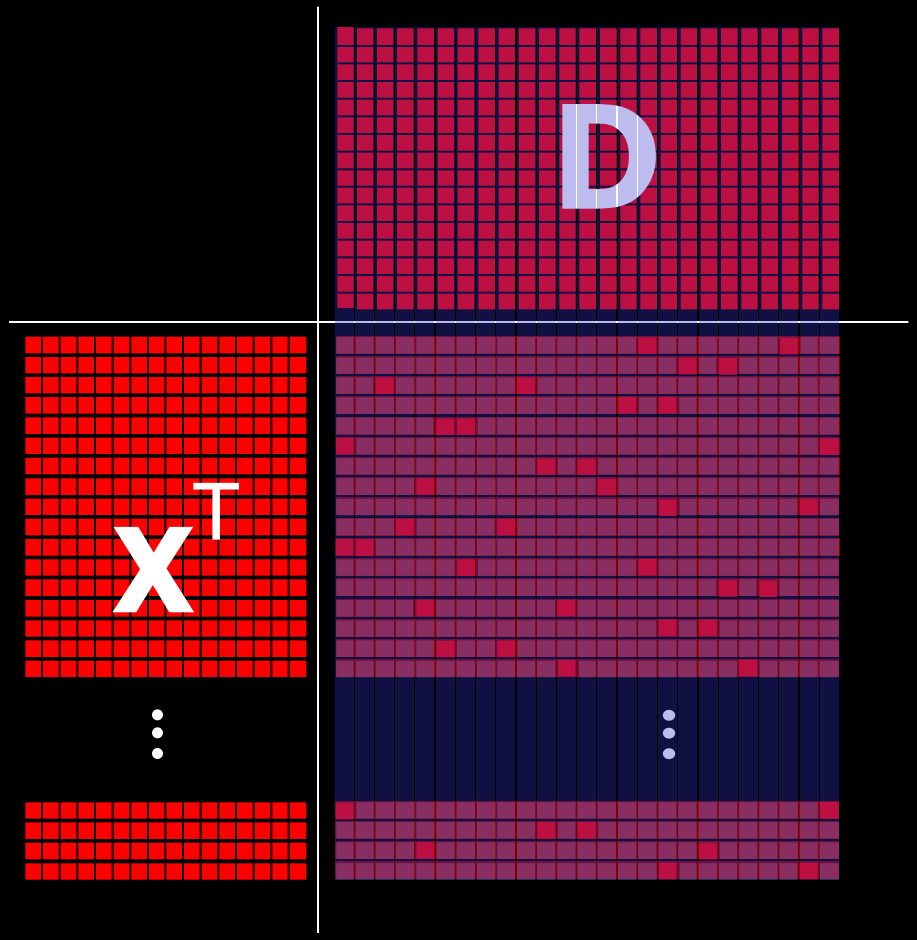
Each example has a sparse representation with no more than L atoms

Field & Olshausen ('96)
Engan et. al. ('99)
Lewicki & Sejnowski ('00)
Cotter et. al. ('03)
Gribonval et. al. ('04)
Aharon, Elad, & Bruckstein ('04)
Aharon, Elad, & Bruckstein ('05)
Ng et al. ('07)
Mairal, Sapiro, Elad ('08)

The K–SVD Algorithm – General

Aharon, Elad, & Bruckstein (`04)



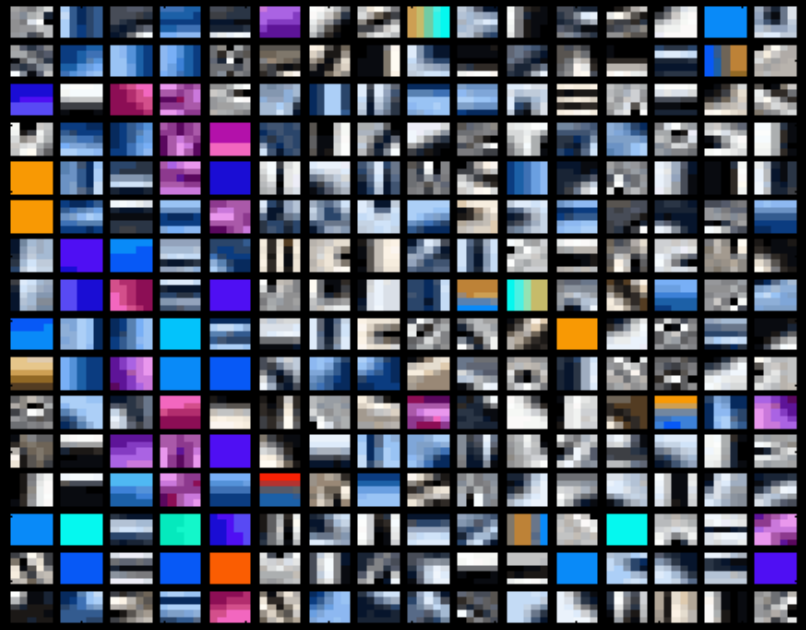


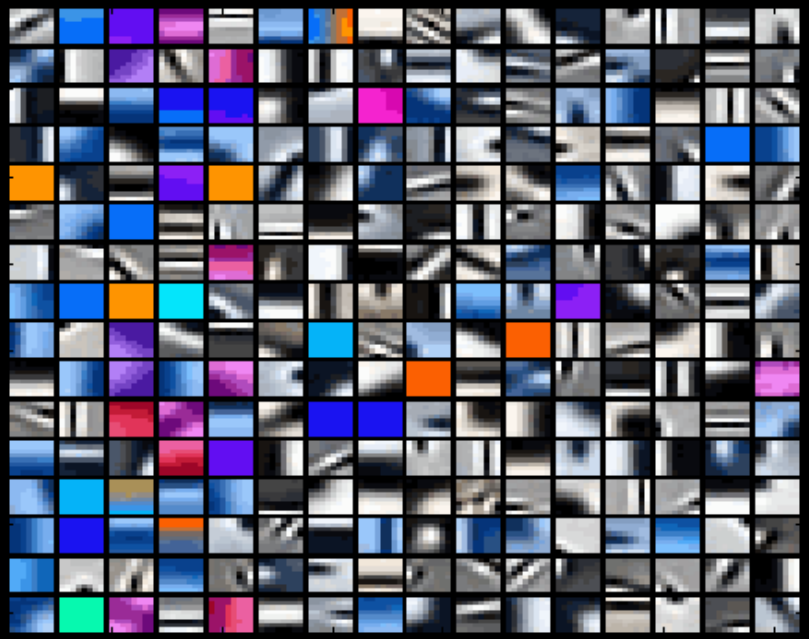
Non-uniform noise

$$\begin{aligned}\{\widehat{\alpha_{ij}}, \hat{\mathbf{D}}, \hat{\mathbf{x}}\} &= \arg \min_{\mathbf{D},\alpha_{ij},\mathbf{x}} \lambda||\beta \otimes (\mathbf{x}-\mathbf{y})||_2^2 \\ &+ \sum_{i,j} \mu_{ij}||\alpha_{ij}||_0 \\ &+ \sum_{ij} ||(\mathbf{R}_{ij}\beta) \otimes (\mathbf{D}\alpha_{ij} - \mathbf{R}_{ij}\mathbf{x})||_2^2.\end{aligned}$$

Show me the pictures

Change the Metric in the OMP





$$< \mathbf{y}, \mathbf{x} >_\gamma = \mathbf{y}^T\mathbf{x} + \frac{\gamma}{n^2}\mathbf{y}^T\mathbf{K}^T\mathbf{K}\mathbf{x} = \mathbf{y}^T(\mathbf{I} + \frac{\gamma}{n}\mathbf{K})\mathbf{x},$$

$$\mathbf{K} = \begin{pmatrix} \mathbf{J}_n & 0 & 0 \\ 0 & \mathbf{J}_n & 0 \\ 0 & 0 & \mathbf{J}_n \end{pmatrix}.$$

Example: Non-uniform noise

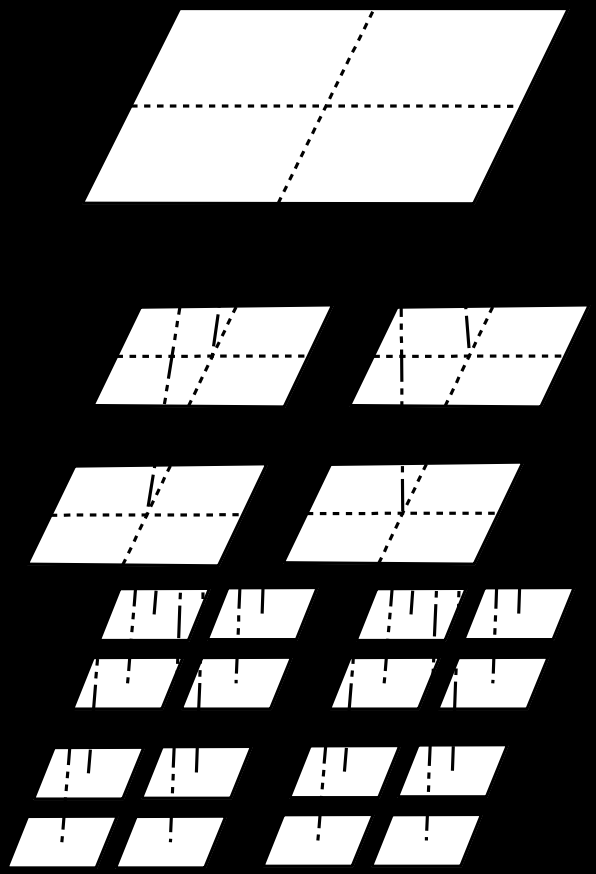
Example: Inpainting

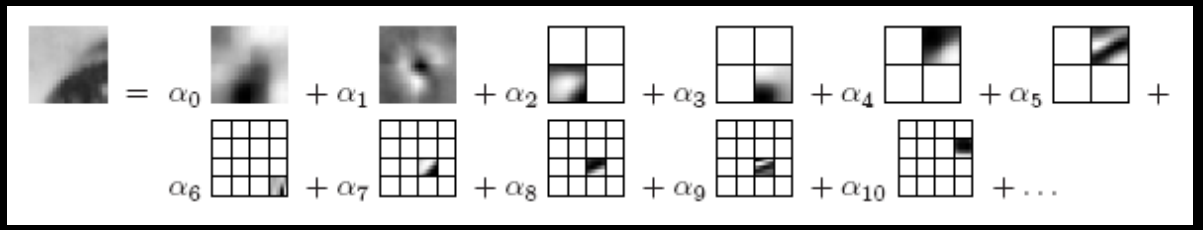
Example: Demoisaic

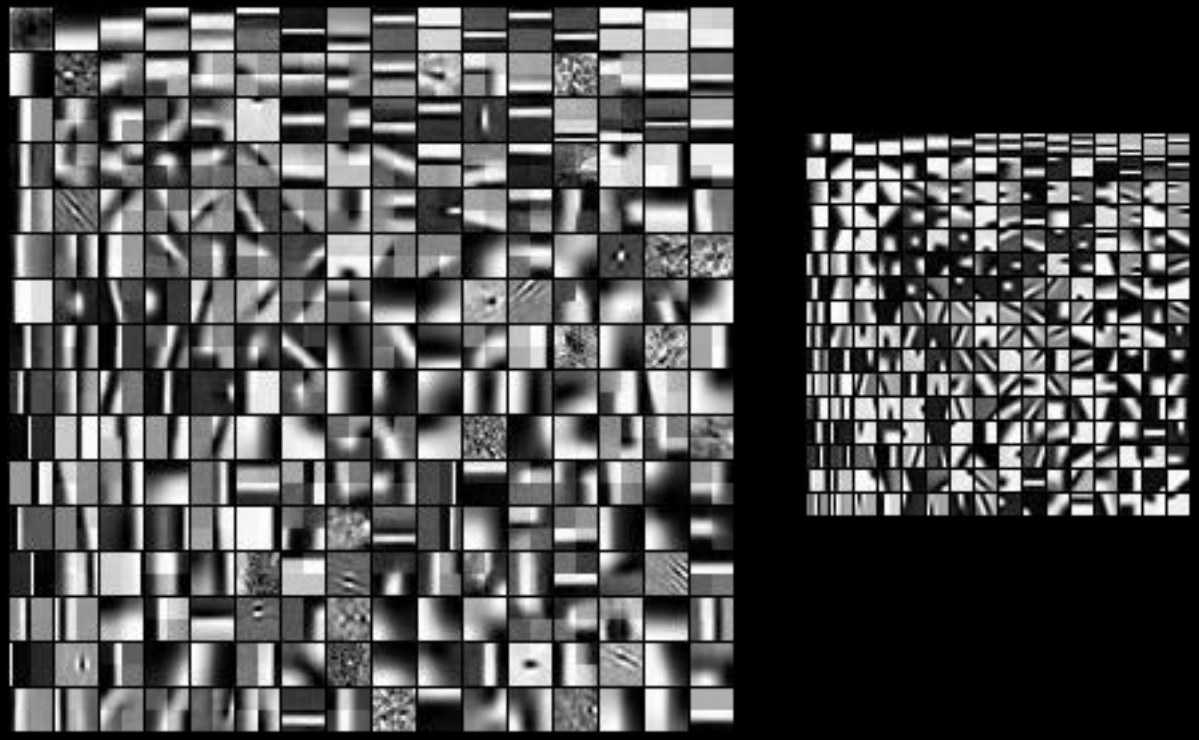
Example: Inpainting

Multiscale Dictionaries

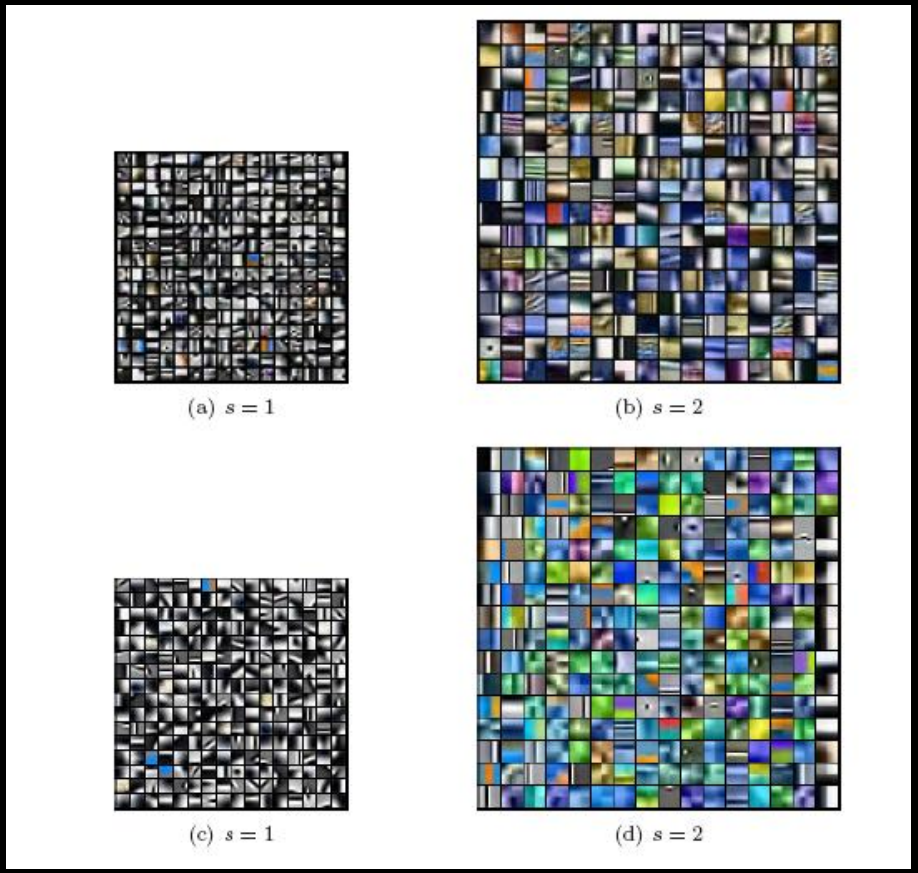
Learned multiscale dictionary







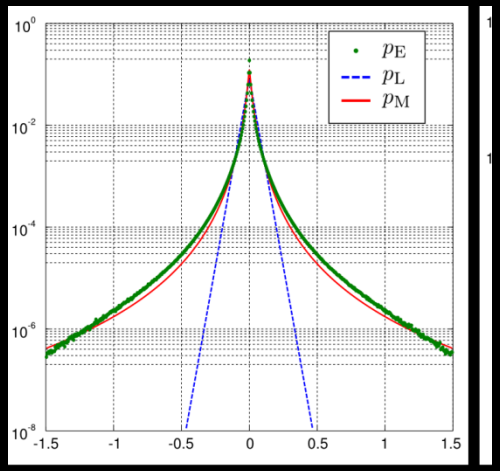
Color multiscale dictionaries



Extending the Models

Universal Coding and Incoherent Dictionaries

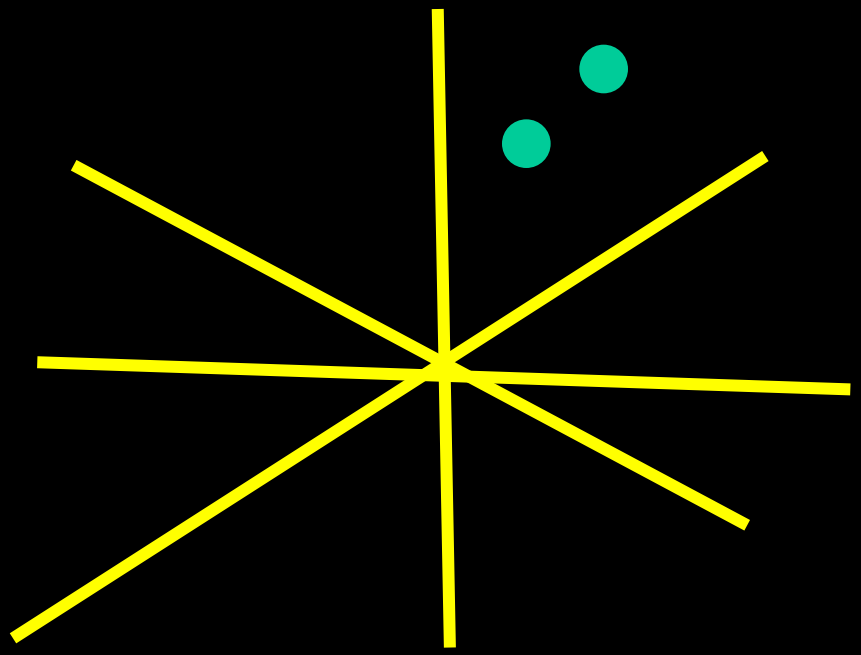
$$f(\mathbf{X},\mathbf{D},\mathbf{A}) = \|\mathbf{X}-\mathbf{D}\mathbf{A}\|_F^2 + \lambda\sum_{j=1}^{N}\sum_{i=1}^{K}\log\left(|\alpha_{ij}|+\beta\right) + \zeta\left\|\mathbf{D}^T\mathbf{D}-\mathbf{I}_K\right\|_F^2 + \eta\sum_{k=1}^{K}(\|\mathbf{D}_k\|_2^2-1)^2.$$



- Consistent
- Improved generalization properties
- Improved active set computation
- Improved reconstruction
- Improved coding speed

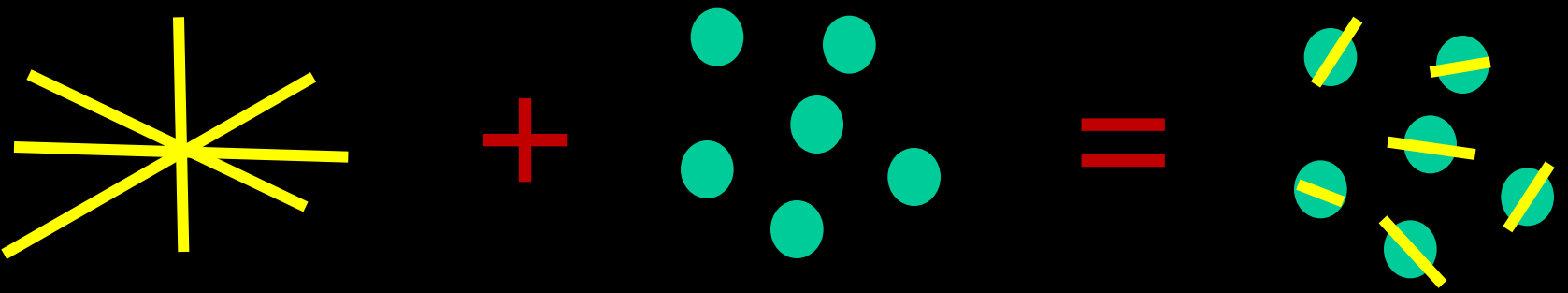
ℓ_0	n_ε	SC	$\mathcal{H}(n_\varepsilon)$	OLS PSNR
3	0	ℓ_1 MOL	35.6 **71.1**	37.4 **42.6**
5	1	ℓ_1 MOL	10.6 **43.2**	36.9 **42.2**
8	2	ℓ_1 MOL	7.6 **30.8**	37.6 **42.3**

Group Sparsity



Sparsity + Self-similarity=Group Sparsity

- Combine the two of the most successful models for images



Sparsity + Self-similarity=Group Sparsity

Sparsity + Self-similarity=Group Sparsity

Adobe Camera Raw

Proposed Method

Learning to Sense Sparse Images

Motivation

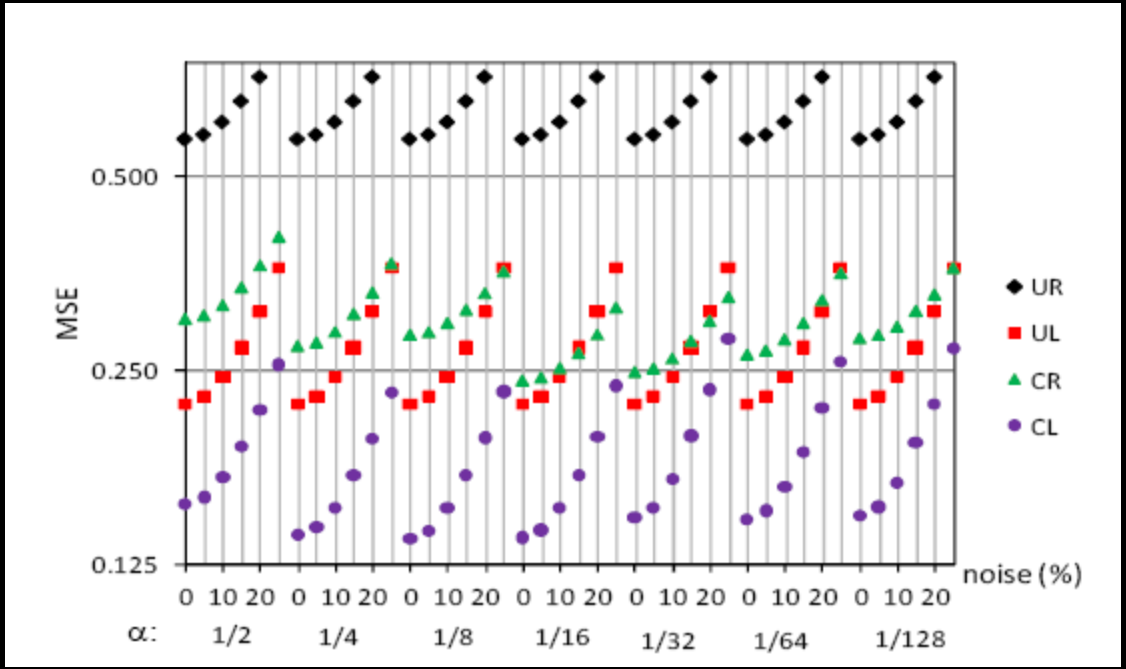
- Compressed sensing (Candes &Tao, Donoho, et al.)
  - Sparsity 
  - Random sampling
    - Universality
    - Stability
- Shall the sensing be adapted to the data type?
  - Yes! (Elad, Peyre, Weiss et al., Applebaum et al, this talk).
- Shall the sensing and dictionary be learned simultaneously?

Some formulas….

$$min_{\mathbf{\Psi},\Phi,\mathbf{\Theta}} \{\alpha\|\mathbf{X}-\mathbf{\Psi\Theta}\|_F^2 + \|\mathbf{Y}-\mathbf{\Phi\Psi\Theta}\|_F^2\} \quad s.t.\ \forall i, \|\boldsymbol{\theta}_i\|_{\ell_0} \leq S$$

+ "RIP (Identity Gramm Matrix)"

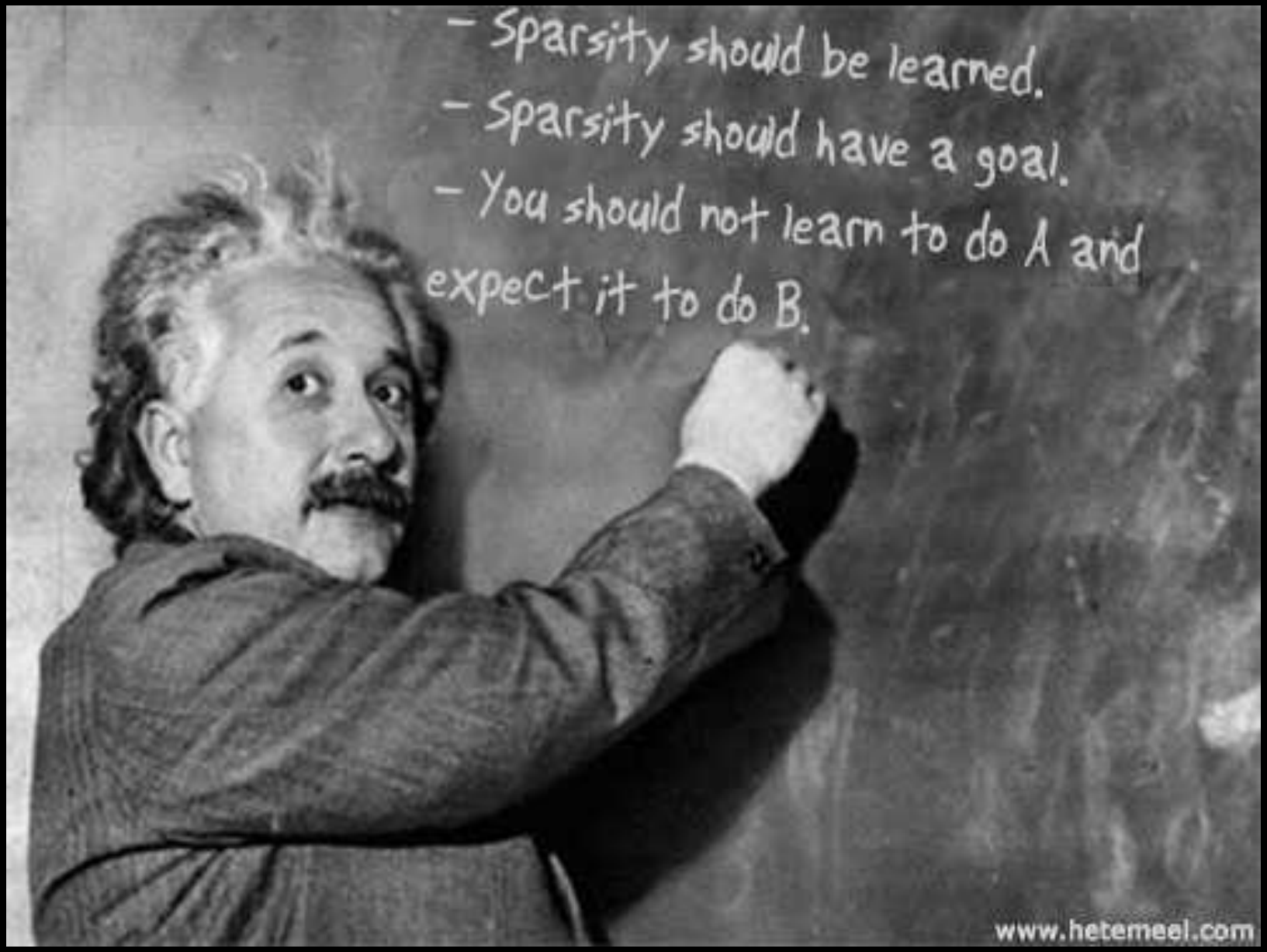
Design the dictionary and sensing together



Just Believe the Pictures

Just Believe the Pictures

Just Believe the Pictures


- Sparsity should be learned.
- Sparsity should have a goal.
- You should not learn to do A and expect it to do B.
www.hetemeel.com