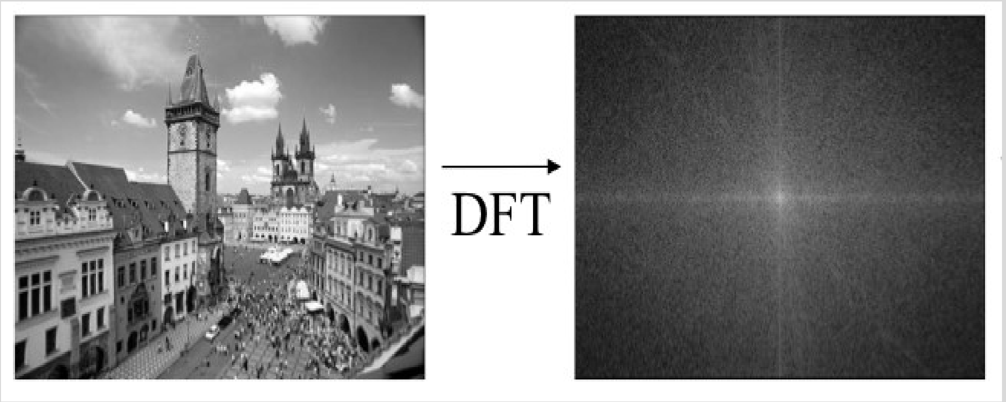
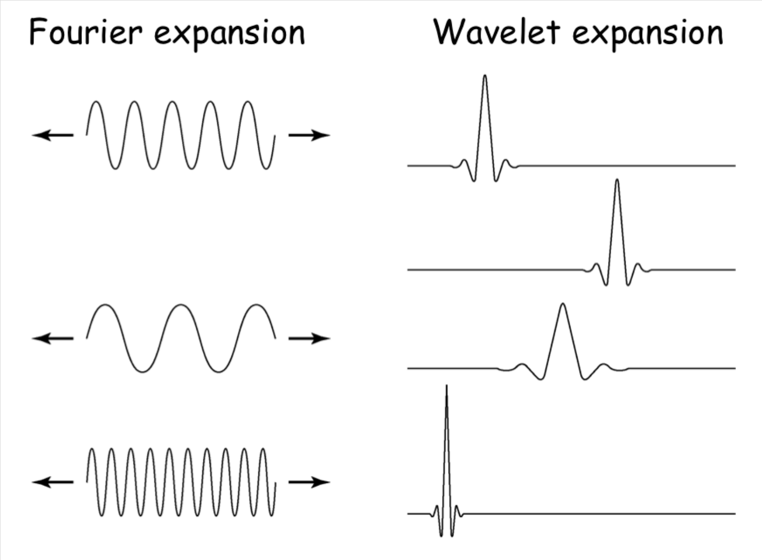
TA session 14 - 17.1.2016

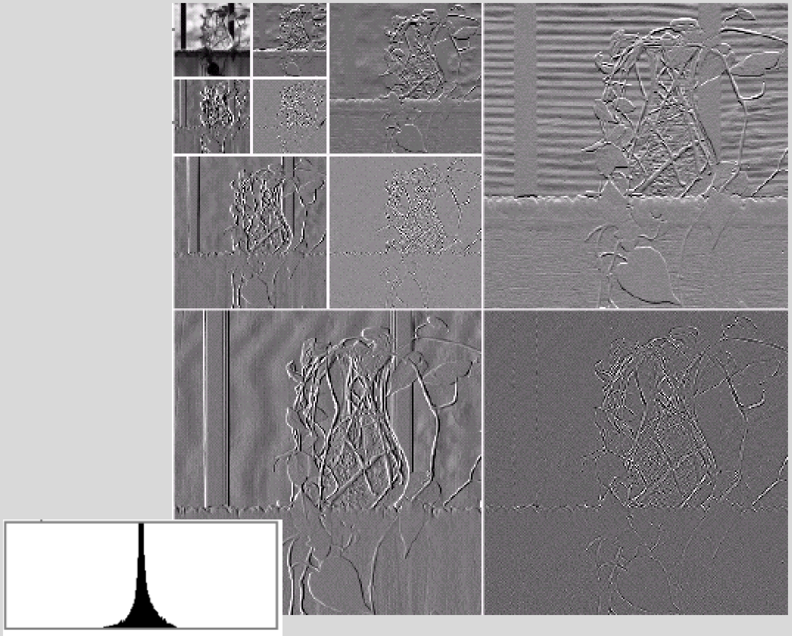
- Sparse representation
- Blending results ☺

Where we've already encountered sparsity in the course

e.g. in Fourier and Wavelet based compression



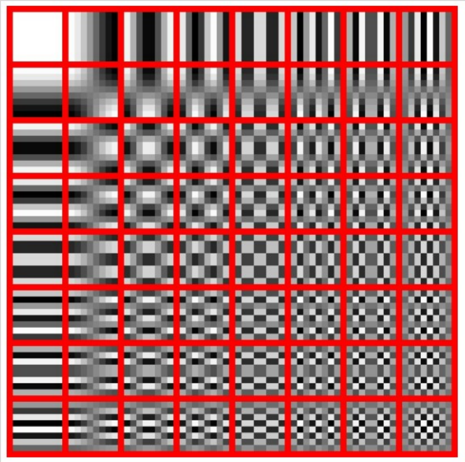




These were good analytical representations

Wavelets and the Fourier basis were 2 sets of complete orthonormal vectors we used to represent our images / image patches.

E.g. the 2D DCT basis looked like this –



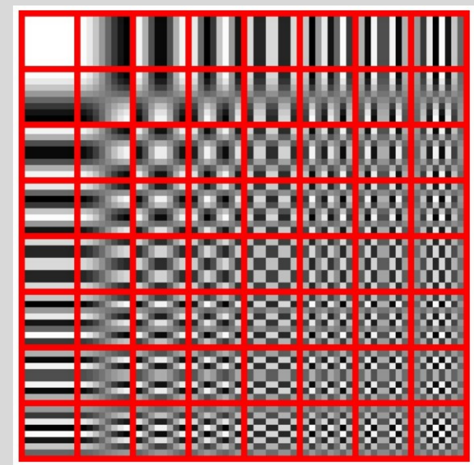
For some patch x, This was basically done as follows

$$\alpha[i] = \sum d_i \cdot x$$

These were good analytical representations

Wavelets and the Fourier basis were 2 sets of complete orthonormal vectors we used to represent our images / image patches.

E.g. the 2D DCT basis looked like this –



For some patch x, This was basically done as follows

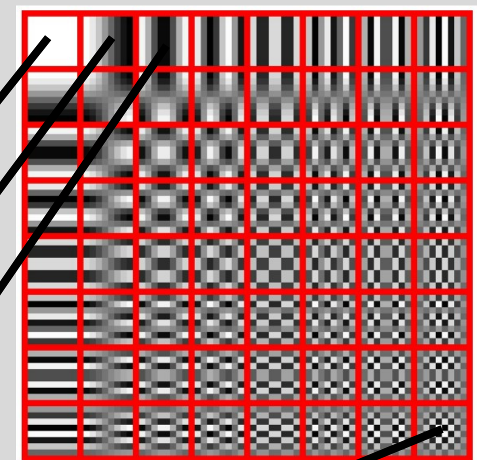
$$\alpha[i] = \sum d_i \cdot x$$

$$\underbrace{\begin{pmatrix} \mathbf{x} \end{pmatrix}}_{\mathbf{x}\in\mathbb{R}^m} \approx \underbrace{\left(\begin{array}{c|c|c|c} \mathbf{d}_1 & \mathbf{d}_2 & \cdots & \mathbf{d}_p \end{array}\right)}_{\mathbf{D}\in\mathbb{R}^{m\times p}} \underbrace{\begin{pmatrix} \alpha[1] \\ \alpha[2] \\ \vdots \\ \alpha[p] \end{pmatrix}}$$

These were good analytical representations

Wavelets and the Fourier basis were 2 sets of complete orthonormal vectors we used to represent our images / image patches.

E.g. the 2D DCT basis looked like this –



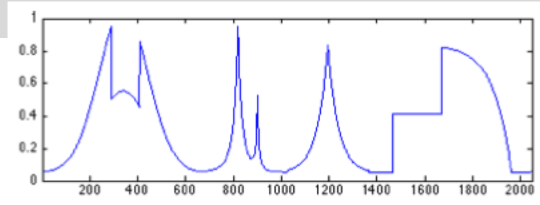
For some patch x, This was basically done as follows

$$\alpha[i] = \sum d_i \cdot x$$

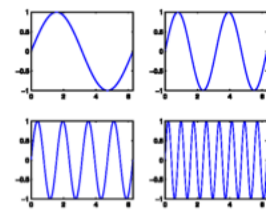
$$\underbrace{\begin{pmatrix} \mathbf{x} \end{pmatrix}}_{\mathbf{x} \in \mathbb{R}^m} \approx \underbrace{\begin{pmatrix} \mathbf{d}_1 & \vert & \mathbf{d}_2 & \vert & \cdots & \vert & \mathbf{d}_p \end{pmatrix}}_{\mathbf{D} \in \mathbb{R}^{m \times p}} \underbrace{\begin{pmatrix} \alpha[1] \\ \alpha[2] \\ \vdots \\ \alpha[p] \end{pmatrix}}$$

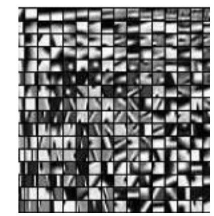
What we mean by sparsity in this context

Let $\boldsymbol{x}$ in $\mathbb{R}^m$ be a signal



Let $\boldsymbol{D} = [\boldsymbol{d}_1, \ldots, \boldsymbol{d}_p] \in \mathbb{R}^{m \times p}$ be a set of normalized “basis vector”. We call it **dictionary**





D is “adapted” to $\boldsymbol{x}$ if it can represent it with a few basis vectors: a **sparse vector** $\alpha \in \mathbb{R}^p$ such the $\boldsymbol{x} \approx \boldsymbol{D} \cdot \boldsymbol{\alpha}$. We call $\boldsymbol{\alpha}$ the sparse code.

$$\underbrace{\begin{pmatrix} \mathbf{x} \end{pmatrix}}_{\mathbf{x}\in\mathbb{R}^m} \approx \underbrace{\begin{pmatrix} \mathbf{d}_1 & \mathbf{d}_2 & \cdots & \mathbf{d}_p \end{pmatrix}}_{\mathbf{D}\in\mathbb{R}^{m\times p}} \underbrace{\begin{pmatrix} \alpha[1] \\ \alpha[2] \\ \vdots \\ \alpha[p] \end{pmatrix}}_{\alpha\in\mathbb{R}^p,\,\textbf{sparse}}$$

Where do we expect to encounter sparsity

- Example: Patches taken from a large page of text (same font).
- But this is apparently a very common feature of various types of signals and classes of images + biological motivation

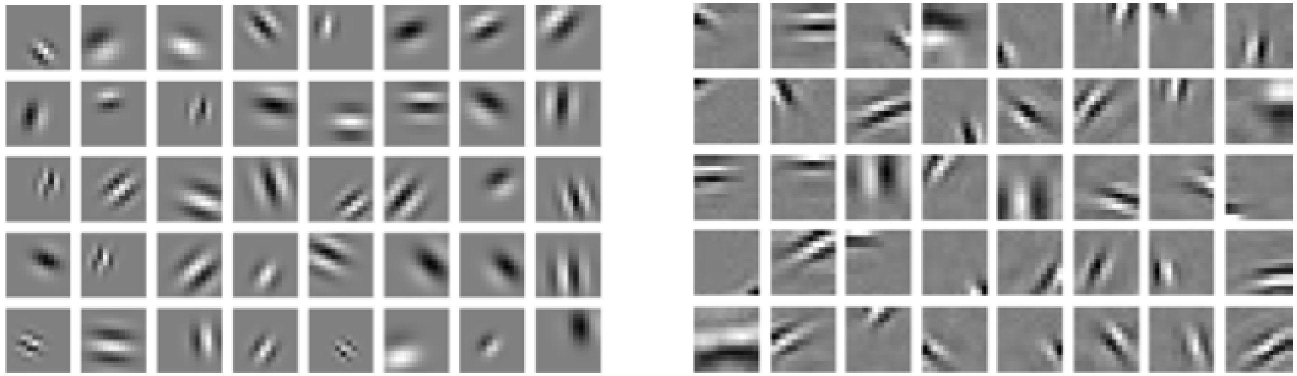
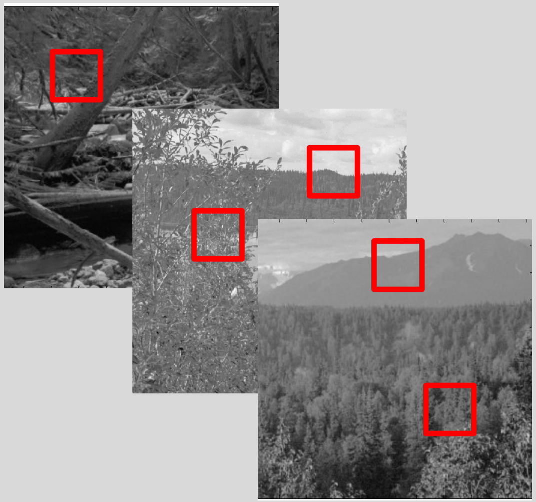


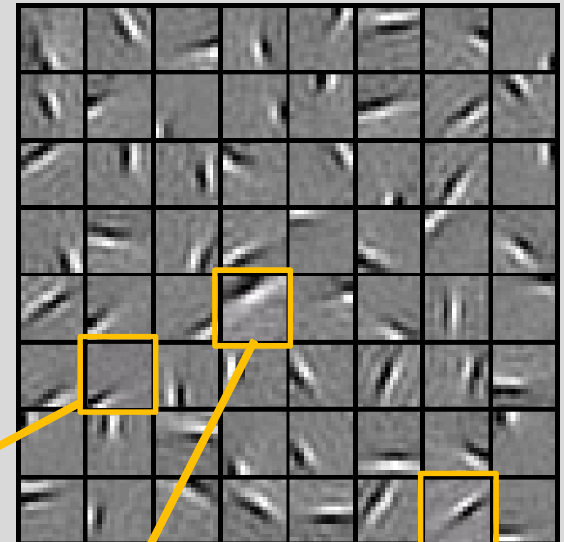
Fig. 2. Left: A few 12×12 Gabor atoms at different scales and orientations. Right: A few atoms trained by Olshausen and Field (extracted from [34]).

Sparse coding illustration

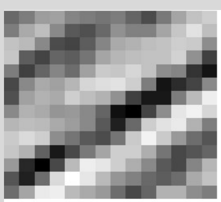
Natural Images



$\boldsymbol{D} = [\boldsymbol{d}_1, \dots, \boldsymbol{d}_{64}]$



Test example

 $\approx 0.8 \cdot$  $+\ 0.3 \cdot$  $+\ 0.5 \cdot$ 

$$x \approx 0.8 \cdot d_{36} + 0.3 \cdot d_{42} + 0.5 \cdot d_{36}$$

α = [0, 0, …, 0, **0.8**, 0, …, 0, **0.3**, 0, …, 0, **0.5**, …]

= [a_1, …, a_{64}] (feature representation)

Given a dictionary how can we decompose signal sparsely

$$\underbrace{\begin{pmatrix} \mathbf{x} \end{pmatrix}}_{\mathbf{x}\in\mathbb{R}^m} \approx \underbrace{\left(\begin{array}{c|c|c|c} \mathbf{d}_1 & \mathbf{d}_2 & \cdots & \mathbf{d}_p \end{array}\right)}_{\mathbf{D}\in\mathbb{R}^{m\times p}} \underbrace{\begin{pmatrix} \boldsymbol{\alpha}[1] \\ \boldsymbol{\alpha}[2] \\ \vdots \\ \boldsymbol{\alpha}[p] \end{pmatrix}}$$

$$\min_{\alpha\in\mathbb{R}^p} \underbrace{\frac{1}{2}||x - D\alpha||_2^2}_{\text{data fitting term}} + \underbrace{\lambda\psi(\alpha)}_{\text{sparsity-inducing regularization}}$$

- The term ψ induces sparsity
  - the ℓ_0 "pseudo-norm": $||\alpha||_0 \triangleq \#\{i \ \text{s.t.} \ \alpha_i \neq 0\}$ (NP-hard)
  - the ℓ_1 norm: $||\alpha||_1 \triangleq \sum_{i=1}^{p} |\alpha_i|$ (convex)

Greedy algorithm to find representation given D

Matching Pursuit

$$\min_{\boldsymbol{\alpha} \in \mathbb{R}^p} ||\underbrace{\mathbf{x} - \mathbf{D}\boldsymbol{\alpha}}_{\mathbf{r}}||_2^2 \text{ s.t. } ||\boldsymbol{\alpha}||_0 \leq L$$

1: $\boldsymbol{\alpha} \leftarrow 0$
2: $\mathbf{r} \leftarrow \mathbf{x}$ (residual).
3: **while** $||\boldsymbol{\alpha}||_0 < L$ **do**
4: Select the atom with maximum correlation with the residual

$$\hat{\imath} \leftarrow \arg\max_{i=1,\ldots,p} |\mathbf{d}_i^T \mathbf{r}|$$

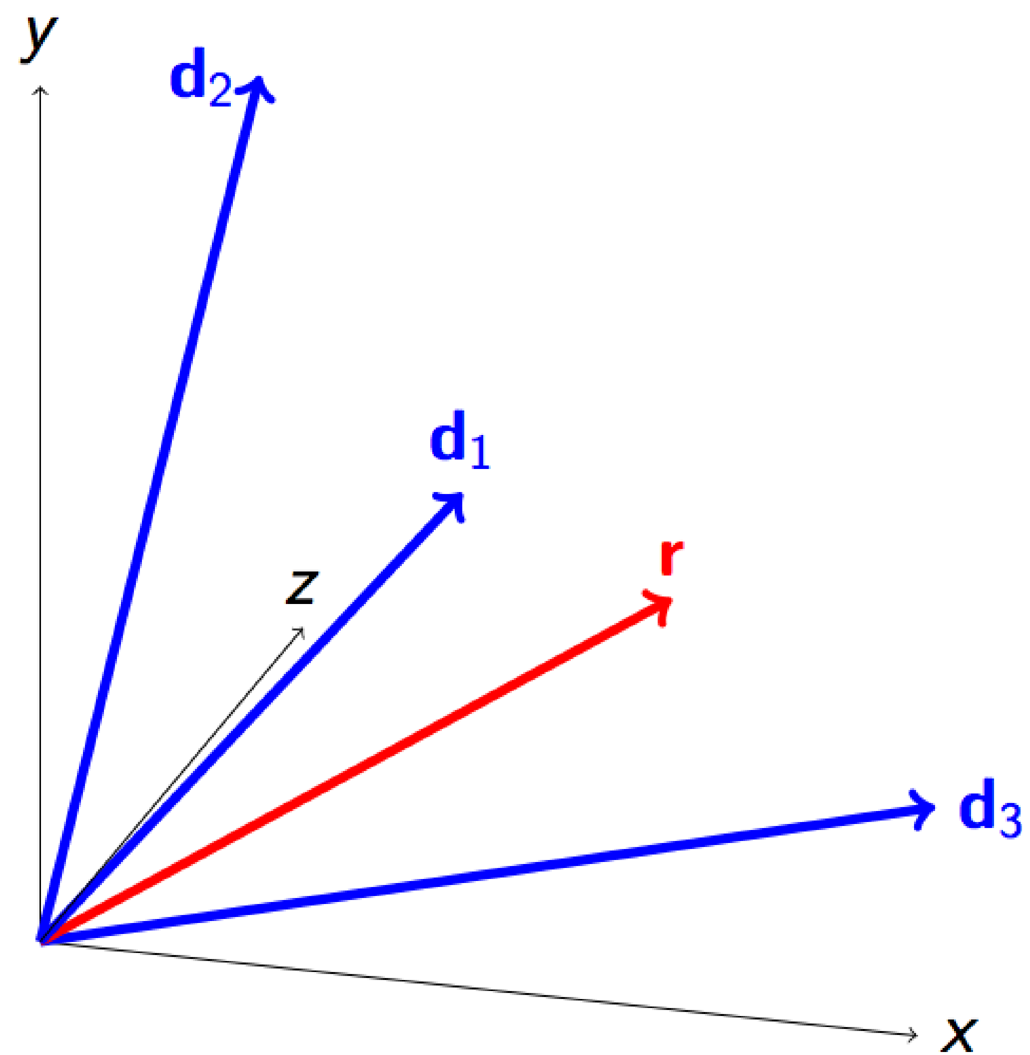
5: Update the residual and the coefficients

$$\begin{aligned} \boldsymbol{\alpha}[\hat{\imath}] &\leftarrow \boldsymbol{\alpha}[\hat{\imath}] + \mathbf{d}_{\hat{\imath}}^T \mathbf{r} \\ \mathbf{r} &\leftarrow \mathbf{r} - (\mathbf{d}_{\hat{\imath}}^T \mathbf{r})\mathbf{d}_{\hat{\imath}} \end{aligned}$$

6: **end while**

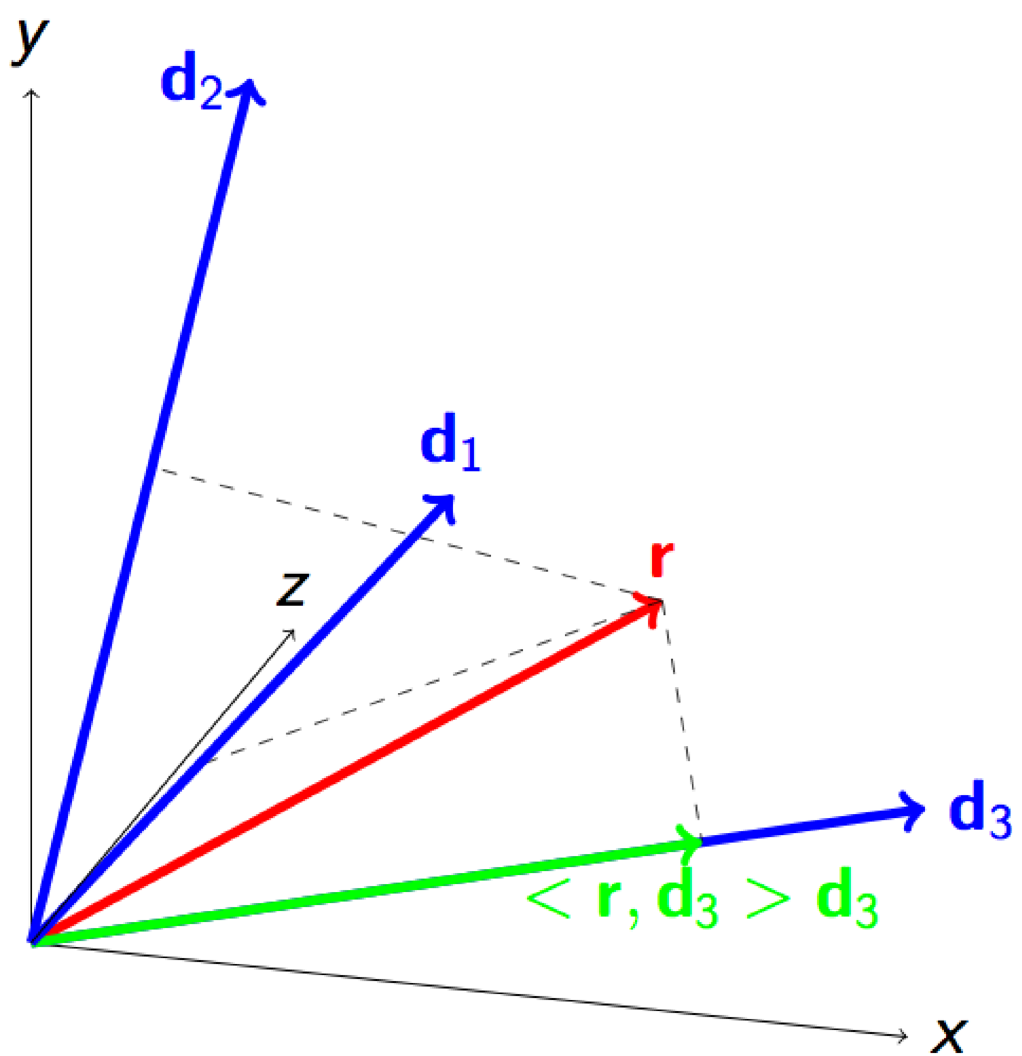
Matching Pursuit

$$\boldsymbol{\alpha} = (0, 0, 0)$$



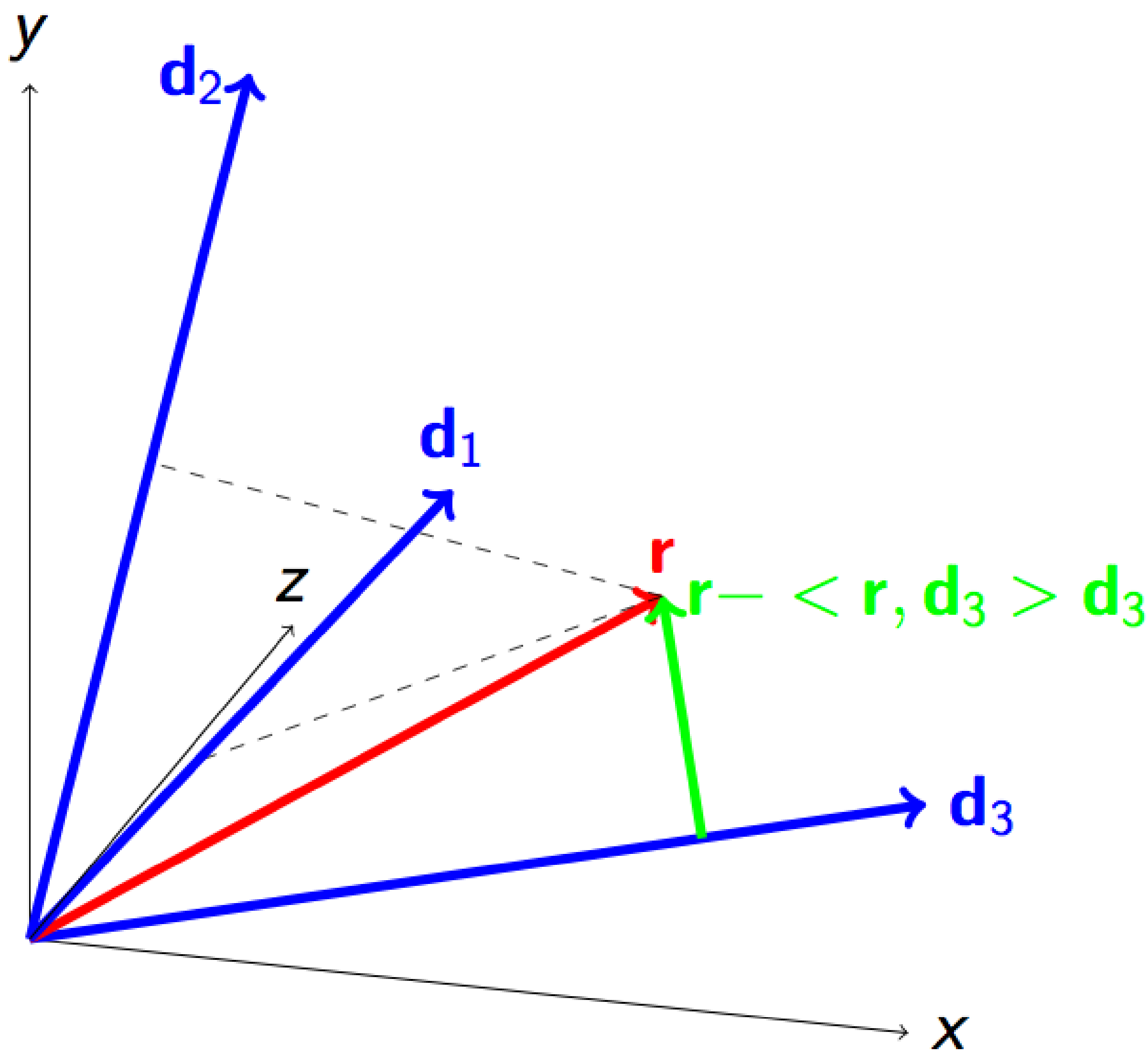
Matching Pursuit

$$\boldsymbol{\alpha} = (0, 0, 0)$$



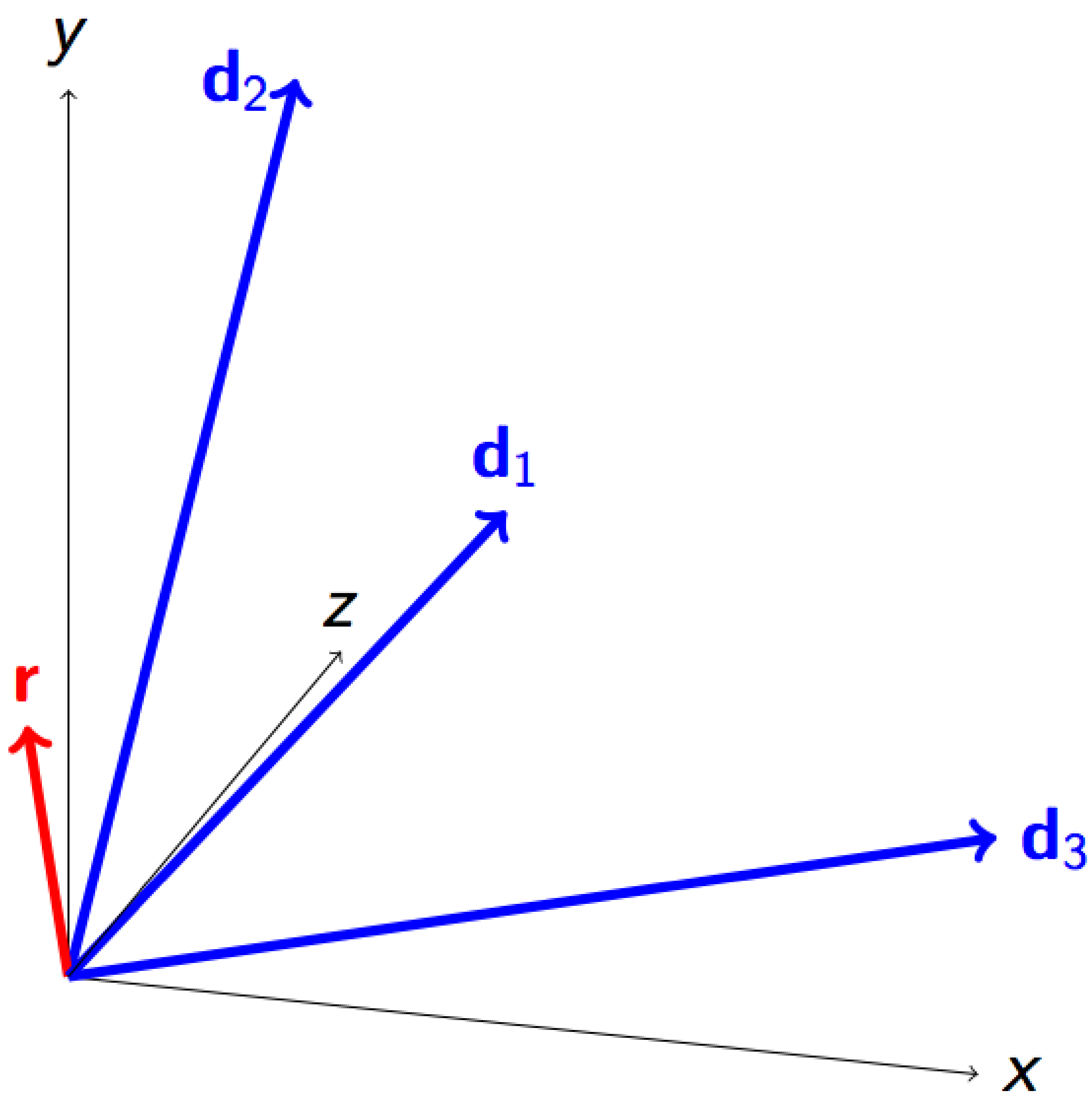
Matching Pursuit

$\boldsymbol{\alpha} = (0, 0, 0)$

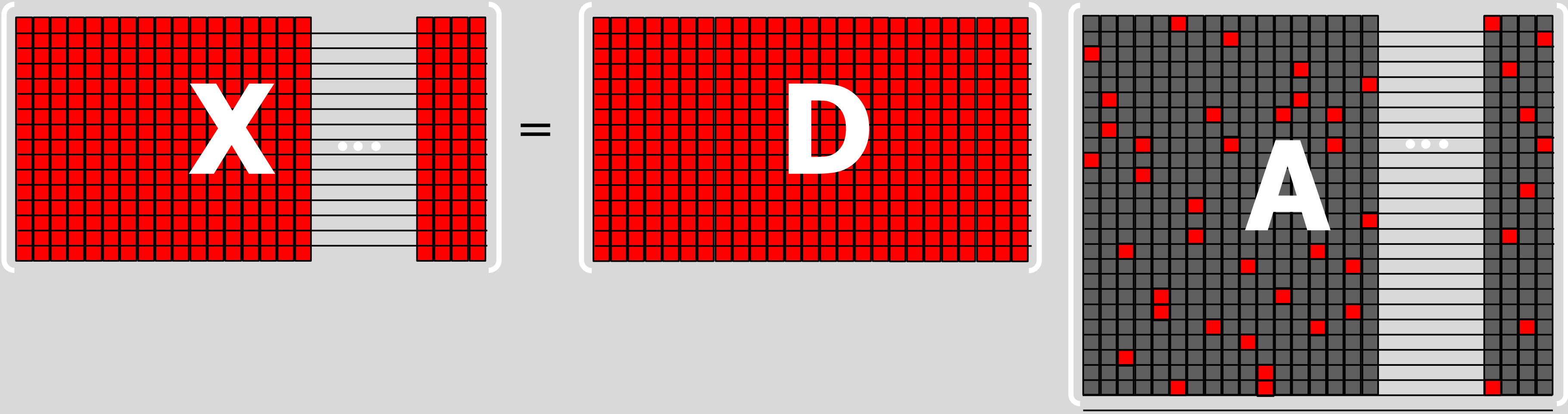


Matching Pursuit

$$\boldsymbol{\alpha} = (0, 0, 0.75)$$



Learning the dictionary



$$\min_{D,A} \|DA - X\|_2^2 \quad s.t. \quad \forall j, \left\|\vec{a}_j\right\|_0 \leq L$$

The examples are linear combinations of atoms from **D**

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

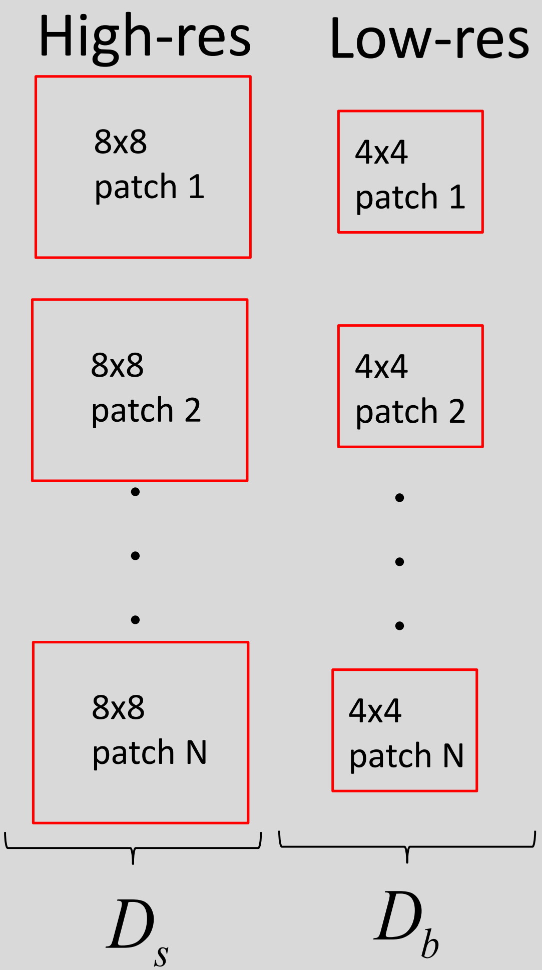
A partial list of applications

- Edge detection
- Single frame super-resolution (Digital zooming)
- Denoising
- Inpainting
- Inverse half-toning
- …

- There are also many computer vision applications which we won't list…

Single frame super-resolution (digital zooming)

Given a low resolution image: find a natural high resolution image that its downsampled version is similar to the input image.



Single frame super-resolution (digital zooming)

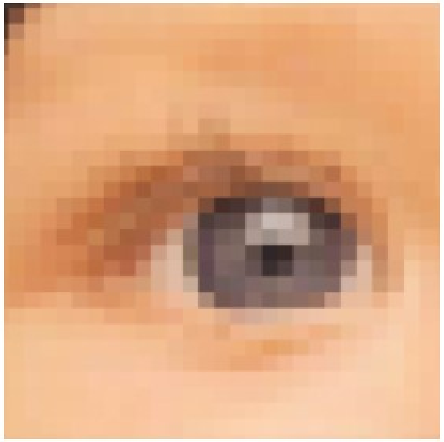
Given a low resolution image: find a natural high resolution image that its downsampled version is similar to the input image.

High-res	Low-res
8x8 patch 1	4x4 patch 1
8x8 patch 2	4x4 patch 2
⋮	⋮
8x8 patch N	4x4 patch N
D_s	D_b

For a new low-res image, find

$$\min_{\alpha} \|x - D_b \alpha\|_2^2 + \lambda \|\alpha\|_1$$

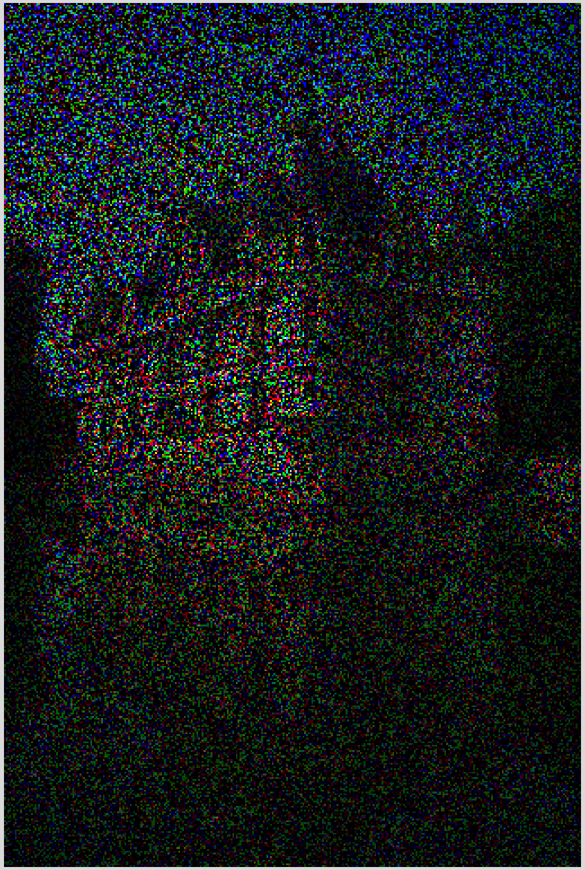
$$=> y = D_s \cdot \alpha$$



Example: denoising

*

Example: Inpainting



*

Example: Inpainting