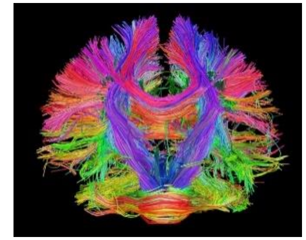
    

CS-570
Statistical Signal Processing

Lecture 6: Signal Representation (part b)

Spring Semester 2019

Grigorios Tsagkatakis

Today’s Objectives

Topics:

- Fourier and Wavelet Transform
- Sparsity and Dictionary Learning

Disclaimer: Material used:

Zhang, Zheng, et al. "A survey of sparse representation: algorithms and applications." *IEEE access* 3 (2015): 490-530.

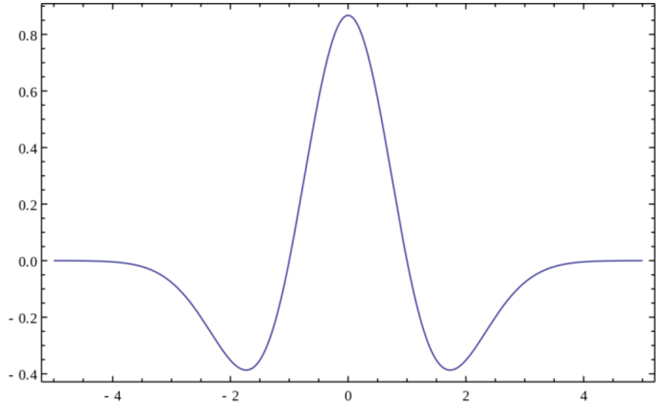
Wavelets

- Wavelet transform decomposes a signal into a set of basis functions.
- These basis functions are called *wavelets*
- Wavelets are obtained from a single prototype wavelet $\psi(t)$ called mother *wavelet* by *dilations* and *shifting:*

$$\psi_{a,b}(t) = \frac{1}{\sqrt{a}}\psi(\frac{t-b}{a})$$

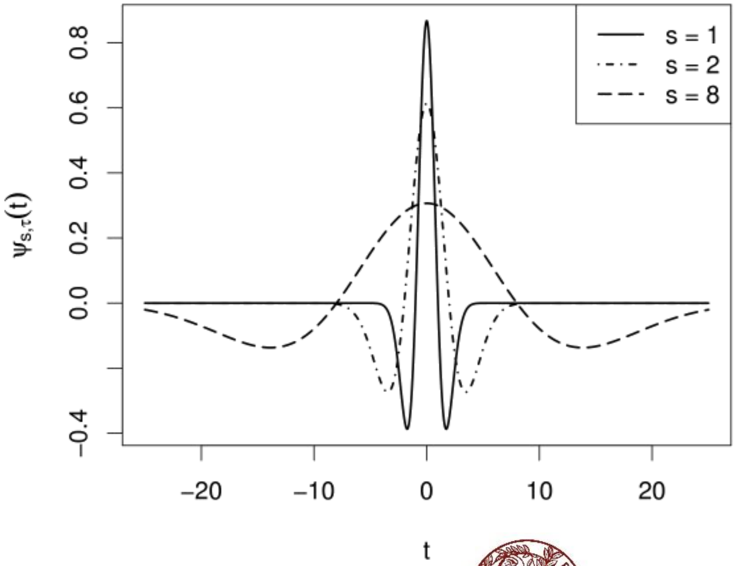
where a is the scaling parameter and b is the shifting parameter

Mother Wavelet examples



Mexican hat
negative normalized second derivative of a Gaussian function

$$\psi(t) = \frac{2}{\sqrt{3\sigma}\pi^{1/4}}\left(1 - \left(\frac{t}{\sigma}\right)^2\right)e^{-\frac{t^2}{2\sigma^2}}$$



DWT matrix

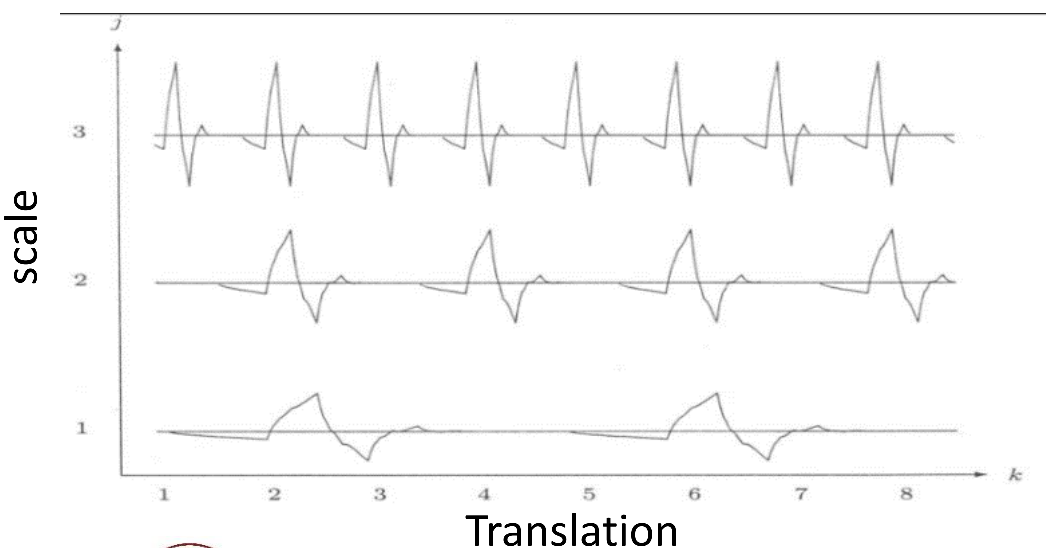
$$\begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & 0 & 0 \\ 0 & 0 & 1 & -1 \end{bmatrix}$$

Discrete Wavelet Transform (DWT)

- Forward $$a_{jk} = \sum_t f(t)\psi^*_{jk}(t)$$

- Inverse $$f(t) = \sum_k \sum_j a_{jk}\psi_{jk}(t)$$

Where $$\psi_{jk}(t) = 2^{j/2}\psi\left(2^j t - k\right)$$



Multiresolution Representation Using Wavelets

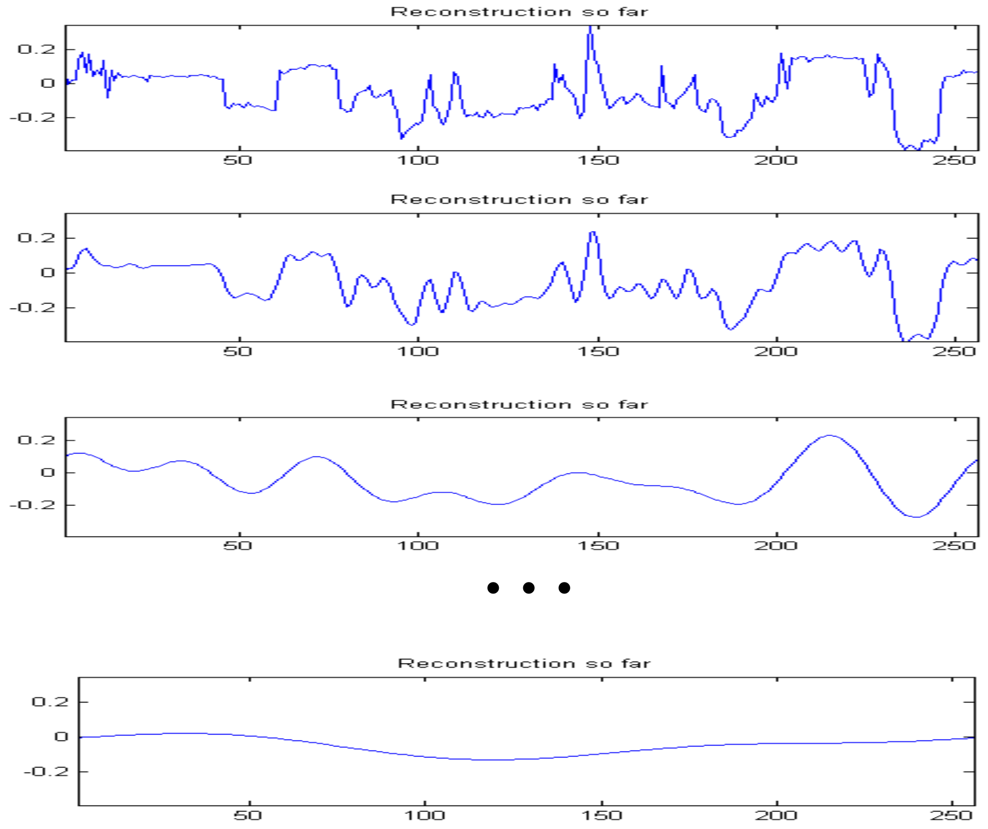
$f(t)$

$\hat{f}_1(t)$

$\hat{f}_2(t)$

...

$\hat{f}_s(t)$



high resolution

(more details)

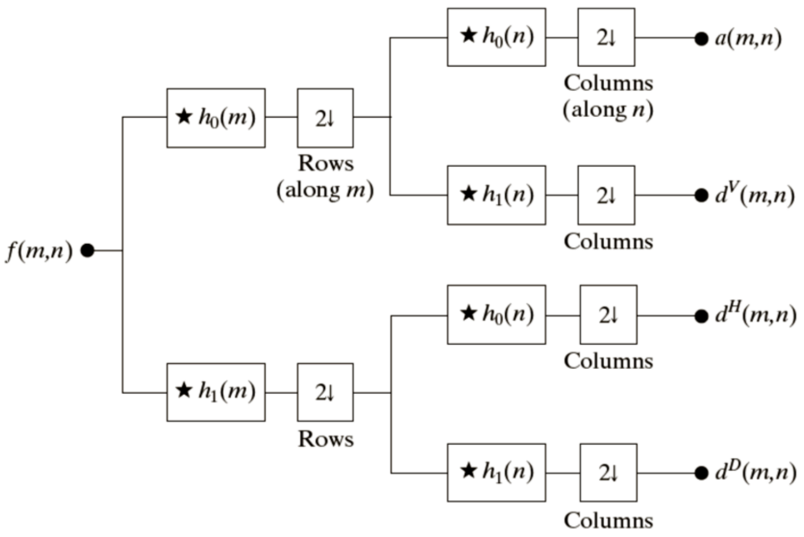
j

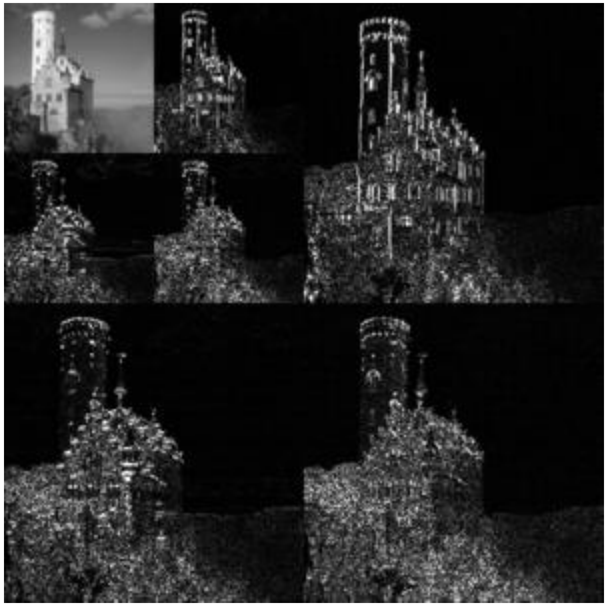
low resolution

(less details)

$$f(t) = \sum_k \sum_j a_{jk} \psi_{jk}(t)$$

DWT in images





The sparsity revolution

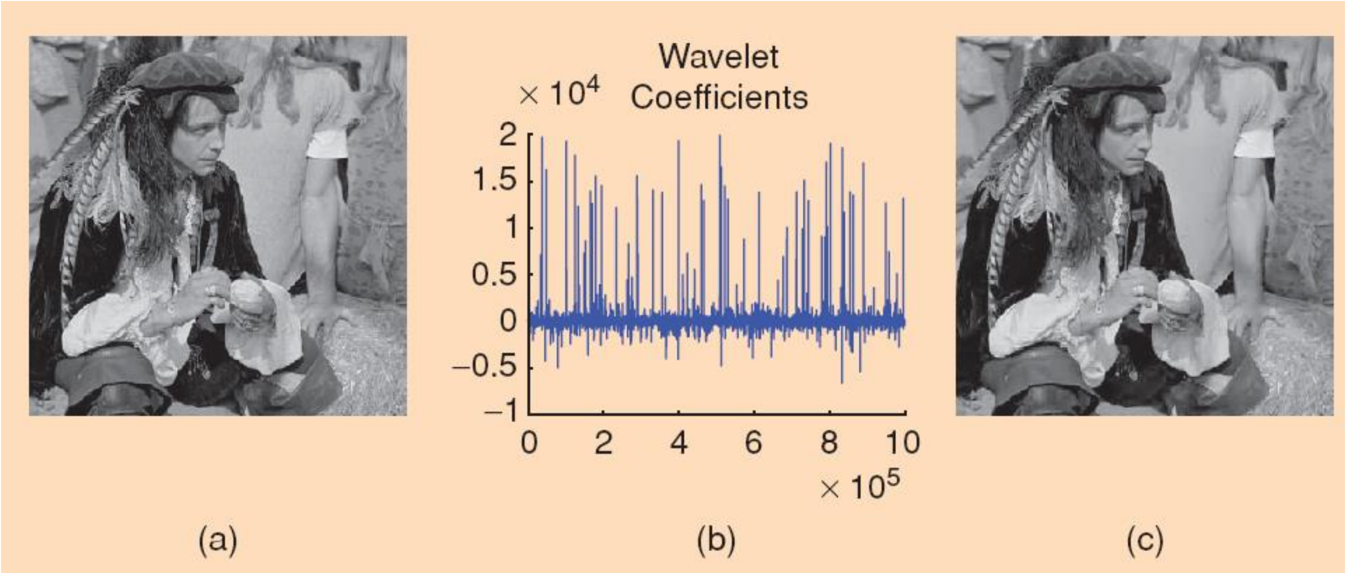
- Solve $y = Dx$ or $\min_x |y - Dx|_2$

Cases

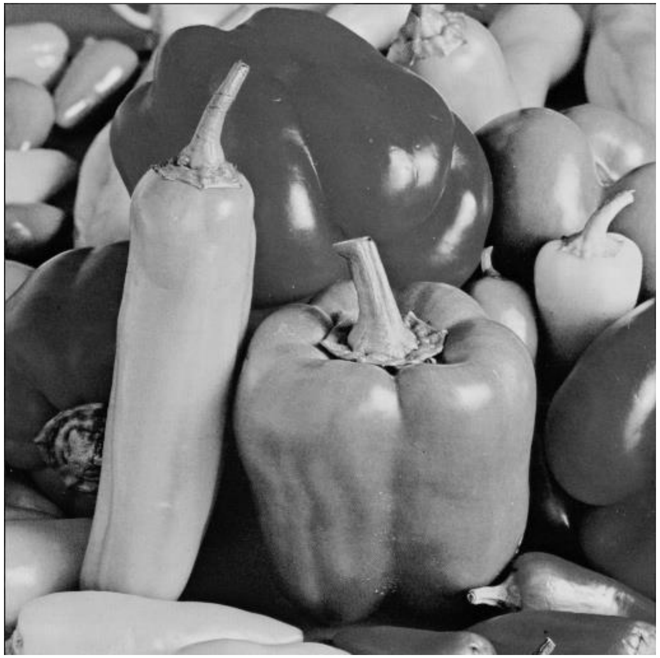
- D = DFT matrix
- D = DCT matrix
- D = DWT matrix
- D something else?

Sparsity

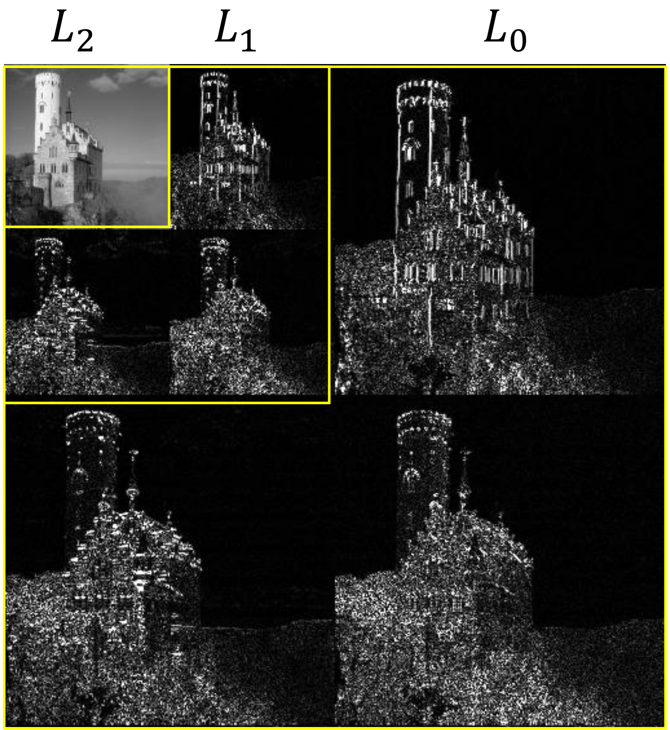
- The concept that most signals in our natural world are sparse



a. Original image
c. Image reconstructed by discarding the zero coefficients



Image



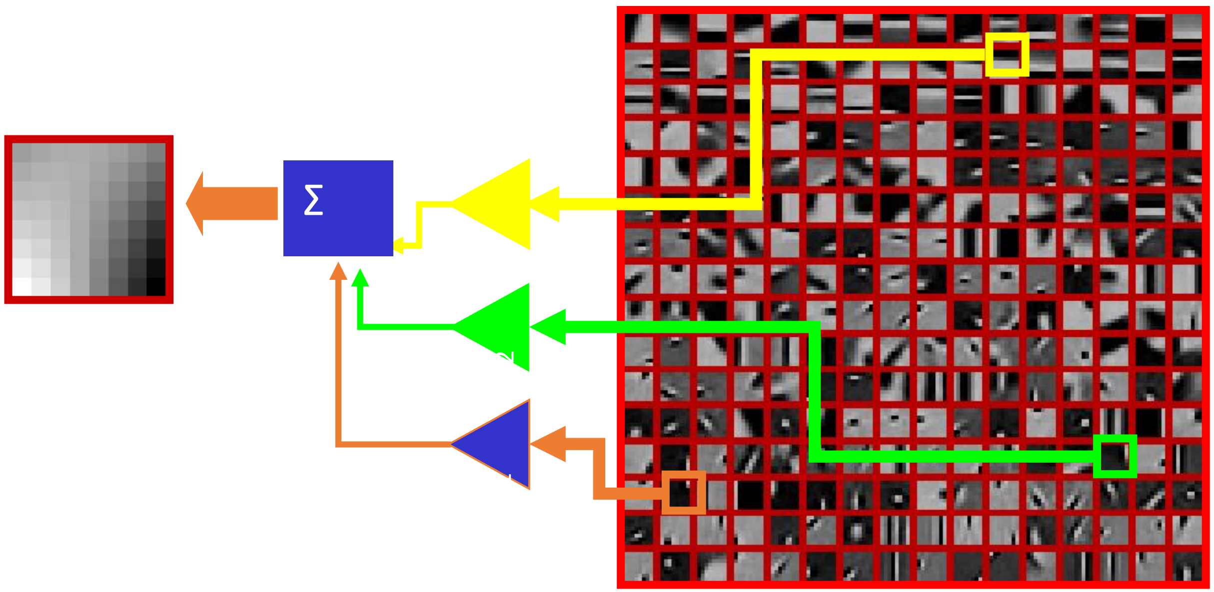
Wavelet transform

Sparse

Multiscale

Sparse representations framework

- Key idea $$\min \|\mathbf{y} - \mathbf{D}\mathbf{x}\|_2 \;\; \text{s.t.} \;\; \|\mathbf{x}\|_0 \leq K$$



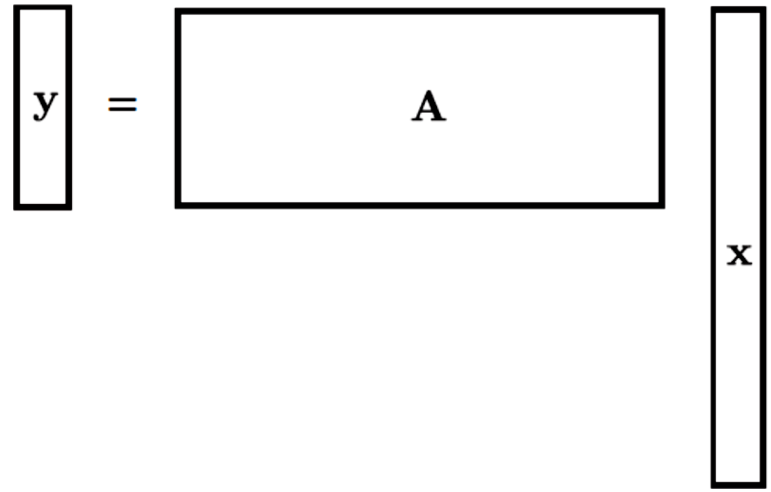
- *Sparseland* model: every signal can be described as a linear combination of few atoms

Sparsity

D is adapted to x if it can represent it with a few basis vectors(called atoms) - that is, there exists a sparse vector α in $\mathbb{R}^p$ such that $D\alpha \approx x$. We call α the sparse code.

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

Theoretical model



Find $\boldsymbol{x} \in \mathbb{C}^p$ s.t. $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{y}$

where $\boldsymbol{A} = [\boldsymbol{a}_1, \cdots, \boldsymbol{a}_p] \in \mathbb{C}^{n \times p}$ obeys

- underdetermined system: $n < p$
- full-rank: $\text{rank}(\boldsymbol{A}) = n$

$\boldsymbol{A}$: an *over-complete basis / dictionary*; $\boldsymbol{a}_i$: atom;
$\boldsymbol{x}$: representation in this basis / dictionary

Motivation

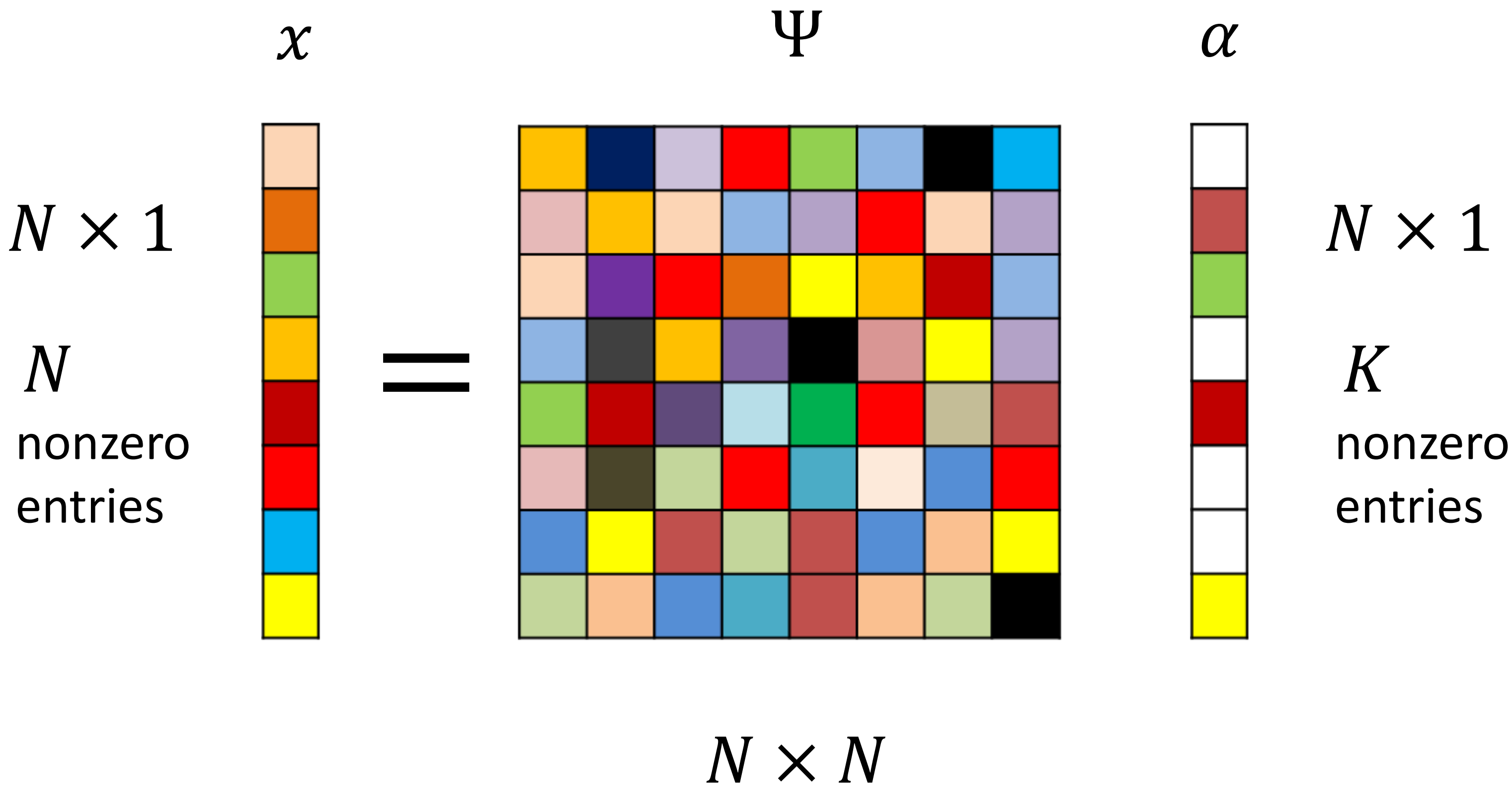
- Signal Transform: Given the signal, its sparsest (over-complete) representation x is its forward transform. Consider this for compression, feature extraction, analysis/synthesis of signals, …

- Signal Prior: in inverse problems seek a solution that has a sparse representation over a predetermined dictionary, and this way regularize the problem (just as TV, bilateral, Beltrami flow, wavelet, and other priors are used).

Sparse coding

General framework $\min_{\mathbf{x}} \|\mathbf{x}\|_0 \;\; \text{s.t.} \;\; \|\mathbf{y} - \mathbf{D}\mathbf{x}\|_2 \le \varepsilon$

- Challenge: finding $|x|_0$ is NP hard



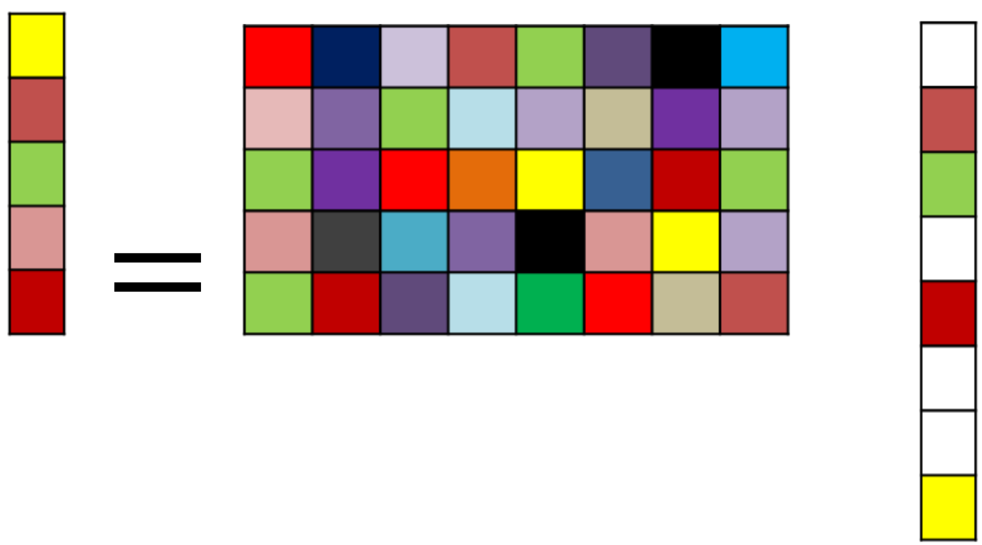
Sparse coding

General framework $\min_{\mathbf{x}} \|\mathbf{x}\|_0 \;\; \text{s.t.} \;\; \|\mathbf{y} - \mathbf{D}\mathbf{x}\|_2 \leq \varepsilon$

- Challenge: finding $|x|_0$ is NP hard

y $\quad$ Φ $\quad$ x

$M \times 1$ measurements



$N \times 1$ sparse signal

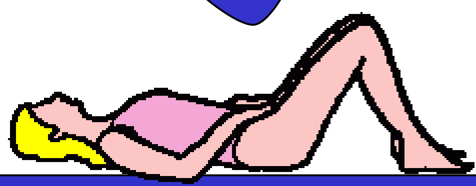
K nonzero entries

$K < M \ll N$

$M \times N$

Solution approaches

$$\min_{\underline{\alpha}} \|\underline{\alpha}\|_0^0 \quad \text{s.t.} \quad \|\mathbf{D}\underline{\alpha} - \underline{y}\|_2^2 \le \varepsilon^2$$

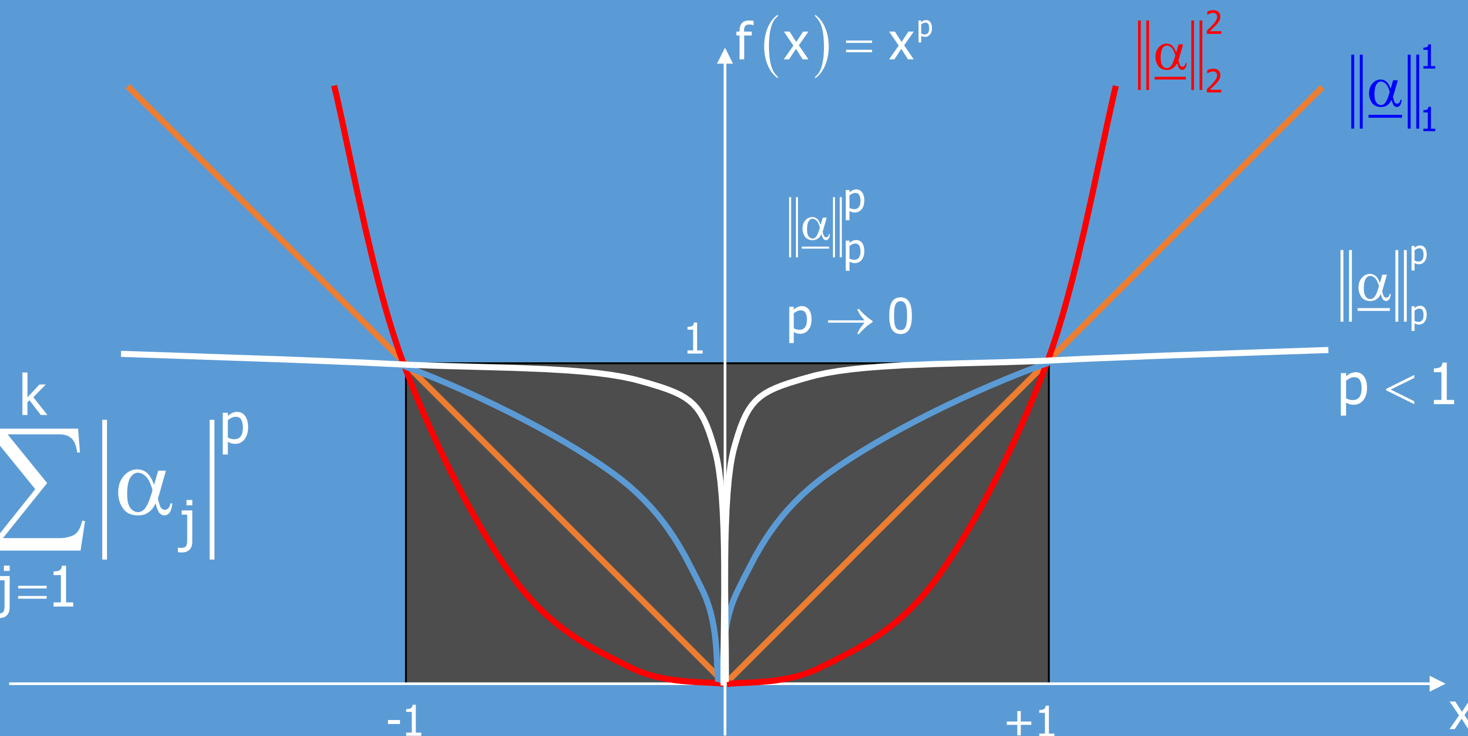


Relaxation methods

Greedy methods

Relaxation

$$\|\underline{\alpha}\|_p^p = \sum_{j=1}^{k} |\alpha_j|^p$$



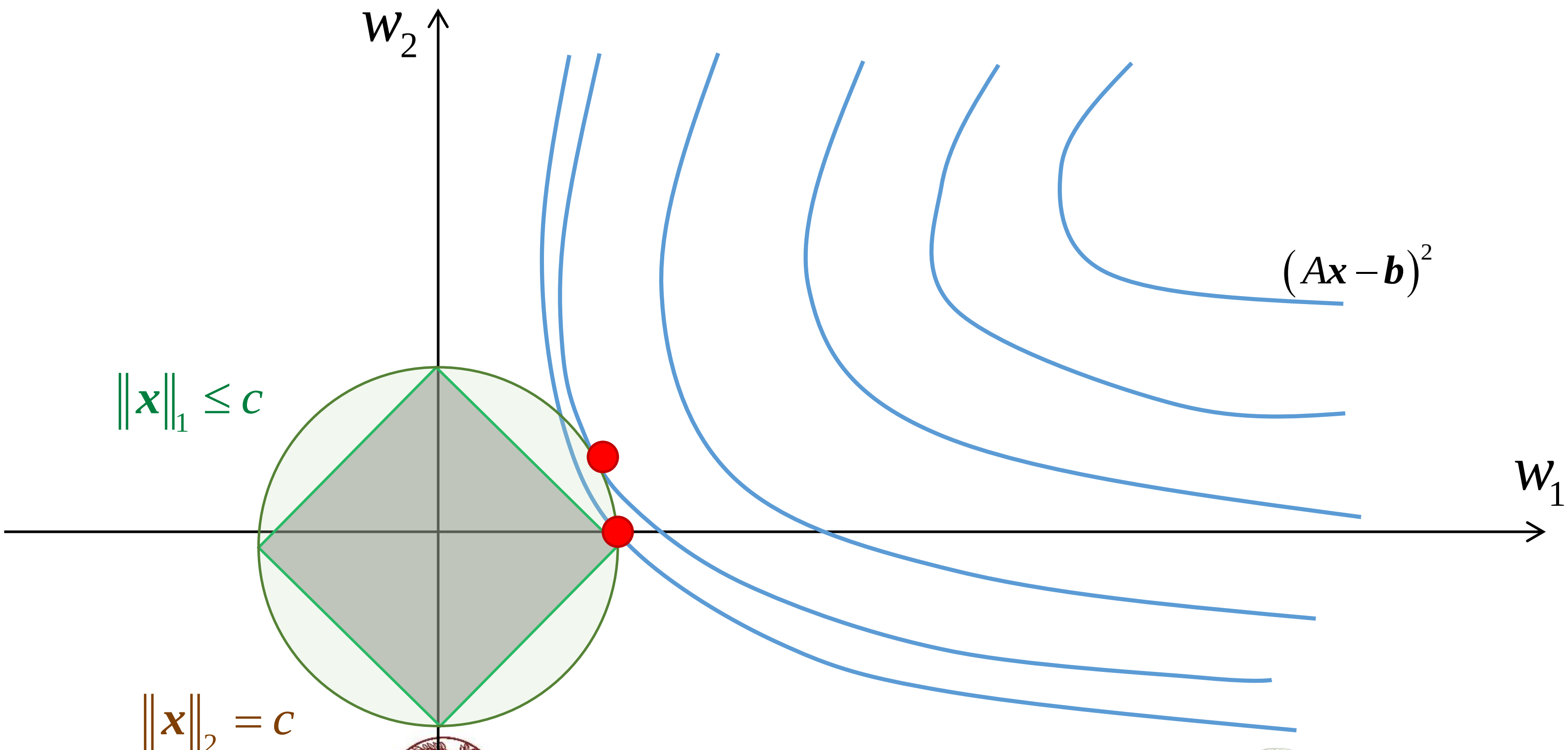
Relaxation

- Replace l_0 with l_1

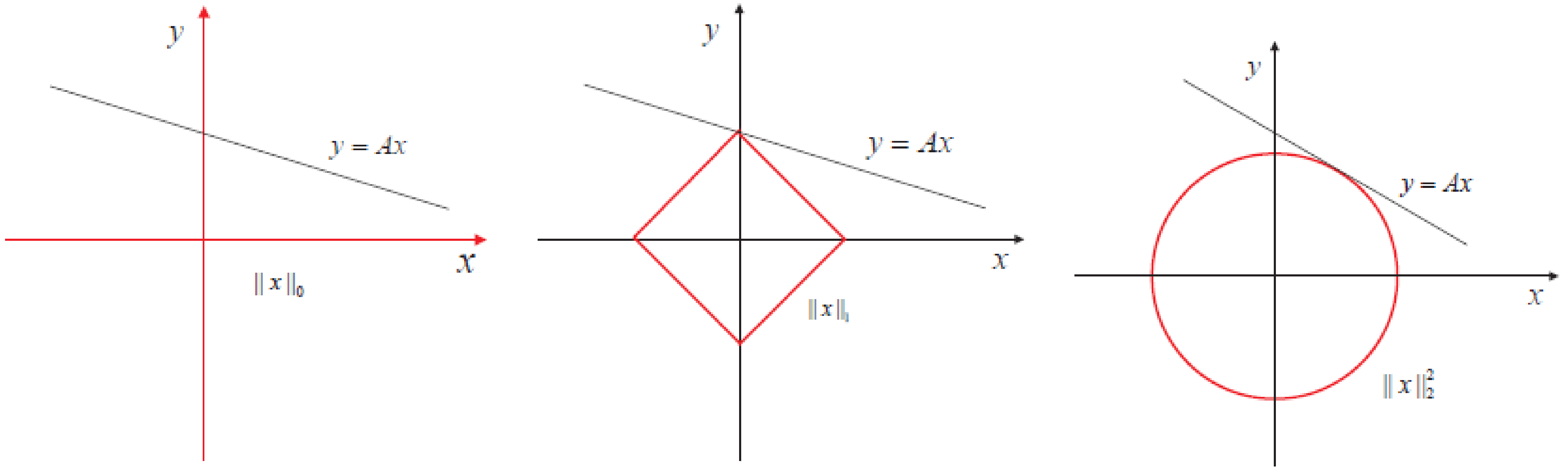
$$\boldsymbol{x}^{***} = \arg\min_{\boldsymbol{x}} \|\boldsymbol{x}\|_1 = \sum_{n=1}^{N} \left|x^{(n)}\right|, \text{ s.t. } \left(A\boldsymbol{x}-\boldsymbol{b}\right)^2 = 0$$

$$\boldsymbol{x}^{***} = \boldsymbol{x}^{*}$$

$$\arg\min_{\boldsymbol{x}} \|\boldsymbol{x}\|_1, \text{ s.t. } \left(A\boldsymbol{x}-\boldsymbol{b}\right)^2 = 0 \;\rightarrow\; \arg\min_{\boldsymbol{x}} \left(A\boldsymbol{x}-\boldsymbol{b}\right)^2, \text{ s.t. } \|\boldsymbol{x}\|_1 \le c$$



The geometry of the solutions of different norm regularization in 2-D space



L_1 minimization formulations

<u>Regularize with approximation error</u>

- Noise free case $$\hat{\boldsymbol{\alpha}} = \arg\min_{\boldsymbol{\alpha}} \|\boldsymbol{\alpha}\|_1 \quad s.t. \quad \boldsymbol{y} = X\boldsymbol{\alpha}$$

- Noisy case $$\hat{\boldsymbol{\alpha}} = \arg\min_{\boldsymbol{\alpha}} \|\boldsymbol{\alpha}\|_1 \quad s.t. \quad \|\boldsymbol{y} - X\boldsymbol{\alpha}\|_2^2 \leq \varepsilon$$

<u>Regularize with sparsity</u>

- Noise free case $$\hat{\boldsymbol{\alpha}} = \arg\min_{\boldsymbol{\alpha}} \|\boldsymbol{y} - X\boldsymbol{\alpha}\|_2^2 \quad s.t. \quad \|\boldsymbol{\alpha}\|_1 \leq \tau$$

<u>Lagrangian formulation</u>

$$\hat{\boldsymbol{\alpha}} = L(\boldsymbol{\alpha}, \lambda) = \arg\min_{\boldsymbol{\alpha}} \frac{1}{2}\|y - X\boldsymbol{\alpha}\|_2^2 + \lambda\|\boldsymbol{\alpha}\|_1$$

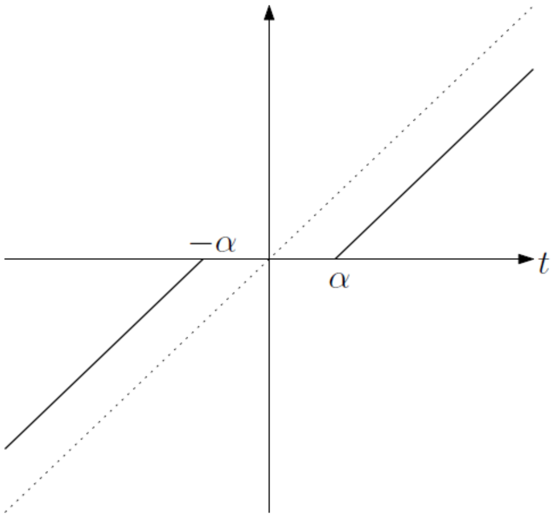
Motivation

- Nature of H
  - Convex
  - Differentiable
  - $\nabla H(x') = \mathrm{D}^T(Dx' - y)$
- Basic Intuition
  - Take an arbitrary x'
  - Calculate $x' - \tau\nabla H(x')$
  - Use the *shrinkage* operator
  - Make corrections and iterate

Shrinkage operator

We define the shrinkage operator as follows

$$shrink(x, \alpha) = \begin{cases} x - \alpha & \text{if } \alpha < x \\ 0 & \text{if } -\alpha \le x \le \alpha \\ x + \alpha & \text{if } x < -\alpha \end{cases}$$



Basis pursuit

Input: Matrix Φ, Signal measurement y, parameter sequence μ_n

Output: Signal estimate $\hat{x}$

Initialization: $\hat{x}_0=0, \quad r=y, \quad k=0$

Algorithm 1 Signal estimate $\hat{x}$

---

while Halting Criterion is false **do**

  $k \longleftarrow k+1$

  $x \longleftarrow \hat{x} - \tau\Phi^T r$

  $\hat{x} \longleftarrow \text{shrink}(x, \mu_k \tau)$

  $r \longleftarrow y - \Phi\hat{x}$

end while

return $\hat{x}$

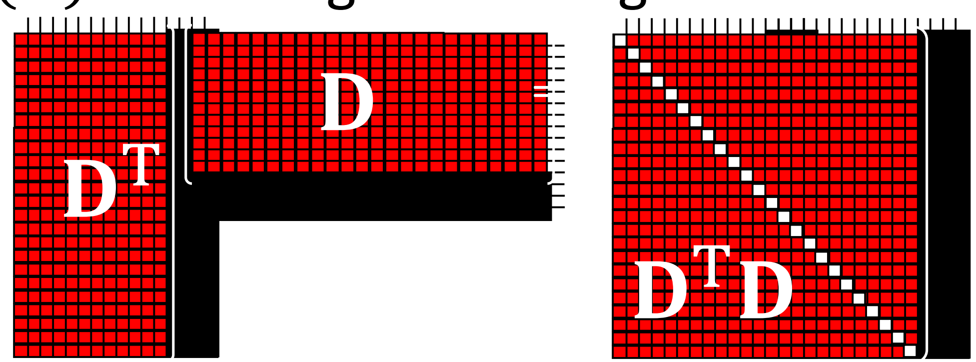
Basis-Pursuit Success

Theorem: Given a noisy signal $y = \mathbf{D}\alpha + v$ where $\|v\|_2 \leq \varepsilon$ and α is sufficiently sparse,

then Basis-Pursuit: $\min_\alpha \ \|\alpha\|_1 \ \ \text{s.t.} \ \ \|\mathbf{D}\alpha - y\|_2 \leq \varepsilon$

leads to a stable result: $\|\hat{\alpha} - \alpha\|_2^2 \leq \frac{4\varepsilon^2}{1-\mu(4\|\alpha\|_0 - 1)}$

The Mutual Coherence $\mu(\mathbf{D})$ is the largest off-diagonal entry in absolute value



Alternating direction method of multipliers (ADMM)

Original problem $\hat{\boldsymbol{\alpha}} = L(\boldsymbol{\alpha}, \lambda) = \arg\min_{\boldsymbol{\alpha}} \frac{1}{2}\|y - X\boldsymbol{\alpha}\|_2^2 + \tau\|\boldsymbol{\alpha}\|_1$

- Introduce auxiliary variable s

$$\arg\min_{\boldsymbol{\alpha}, s} \frac{1}{2\tau}\|\boldsymbol{s}\|_2 + \|\boldsymbol{\alpha}\|_1 \quad s.t. \quad \boldsymbol{s} = \boldsymbol{y} - X\boldsymbol{\alpha}$$

- Form the Augmented Lagrangian function

$$\arg\min_{\boldsymbol{\alpha}, \boldsymbol{s}, \lambda} L(\boldsymbol{\alpha}, \boldsymbol{s}, \boldsymbol{\lambda}) = \frac{1}{2\tau}\|\boldsymbol{s}\|_2 + \|\boldsymbol{\alpha}\|_1 - \lambda^T(\boldsymbol{s} + X\boldsymbol{\alpha} - y) + \frac{\mu}{2}\|\boldsymbol{s} + X\boldsymbol{\alpha} - \boldsymbol{y}\|_2^2$$

Where λ is the Lagrange multiplier vector and μ is a penalty paramter

Alternating direction method of multipliers (ADMM)

- General ADMM framework

Iterate

$$\begin{cases} \boldsymbol{s}^{t+1} = \arg\min L(\boldsymbol{s}, \boldsymbol{\alpha}^t, \boldsymbol{\lambda}^t) \\ \boldsymbol{\alpha}^{t+1} = \arg\min L(\boldsymbol{s}^{t+1}, \boldsymbol{\alpha}, \boldsymbol{\lambda}^t) \\ \boldsymbol{\lambda}^{t+1} = \boldsymbol{\lambda}^t - \mu(\boldsymbol{s}^{t+1} + X\boldsymbol{\alpha}^{t+1} - \boldsymbol{y}) \end{cases}$$

ADMM step 1

$$\arg\min L_{\boldsymbol{s}}(\boldsymbol{s}, \boldsymbol{\alpha}^t, \boldsymbol{\lambda}^t) = \frac{1}{2\tau}\|\boldsymbol{s}\|_2 + \|\boldsymbol{\alpha}^t\|_1 - (\boldsymbol{\lambda}^t)^T(\boldsymbol{s} + X\boldsymbol{\alpha}^t - \boldsymbol{y}) + \frac{\mu}{2}\|\boldsymbol{s} + X\boldsymbol{\alpha}^t - \boldsymbol{y}\|_2^2$$

$$= \frac{1}{2\tau}\|\boldsymbol{s}\|_2 - (\boldsymbol{\lambda}^t)^T\boldsymbol{s} + \frac{\mu}{2}\|\boldsymbol{s} + X\boldsymbol{\alpha}^t - \boldsymbol{y}\|_2^2 + \|\boldsymbol{\alpha}^t\|_1 - (\boldsymbol{\lambda}^t)^T(X\boldsymbol{\alpha}^t - \boldsymbol{y})$$

$$\Longrightarrow \quad \boldsymbol{s}^{t+1} = \frac{\tau}{1+\mu\tau}(\boldsymbol{\lambda}^t - \mu(\boldsymbol{y} - X\boldsymbol{\alpha}^t))$$

ADMM step 2

$$\arg\min L_{\boldsymbol{\alpha}}(\boldsymbol{s}^{t+1}, \boldsymbol{\alpha}, \lambda^t) = \frac{1}{2\tau}\|\boldsymbol{s}^{t+1}\|_2 + \|\boldsymbol{\alpha}\|_1 - (\boldsymbol{\lambda})^T(\boldsymbol{s}^{t+1} + X\boldsymbol{\alpha} - \boldsymbol{y}) + \frac{\mu}{2}\|\boldsymbol{s}^{t+1} + X\boldsymbol{\alpha} - \boldsymbol{y}\|_2^2$$

$$\Rightarrow \quad \boldsymbol{\alpha}^{t+1} = soft\{\boldsymbol{\alpha}^t - \tau X^T(\boldsymbol{s}^{t+1} + X\boldsymbol{\alpha}^t - \boldsymbol{y} - \boldsymbol{\lambda}^t/\mu), \frac{\tau}{\mu}\}$$

$$soft(\sigma, \eta) = sign(\sigma)\max\{|\sigma| - \eta, 0\}$$

ADMM step 3

- Update Lagrangian parameter $\boldsymbol{\lambda}^{t+1} = \boldsymbol{\lambda}^{t} - \mu(\boldsymbol{s}^{t+1} + X\boldsymbol{\alpha}^{t+1} - \boldsymbol{y})$

Algorithm 4. Alternating direction method (ADM) based sparse representation strategy

Task: To address the unconstraint problem:

$$\hat{\alpha} = \arg\min_{\alpha} \frac{1}{2}\|y - X\alpha\|_2^2 + \tau\|\alpha\|_1$$

Input: Probe sample y, the measurement matrix X, small constant λ

Initialization: $t = 0$, $s^0 = 0$, $\alpha^0 = 0$, $\lambda^0 = 0$, $\tau = 1.01$, μ is a small constant.

Step 1: Construct the constraint optimization problem of problem III.12 by introducing the auxiliary parameter and its augmented Lagrangian function, i.e. problem (V.22) and (V.23).

While not converged do

Step 2: Update the value of the s^{t+1} by using Eq. (V.25).

Step 2: Update the value of the α^{t+1} by using Eq. (V.29).

Step 3: Update the value of the λ^{t+1} by using Eq. (V.24(c)).

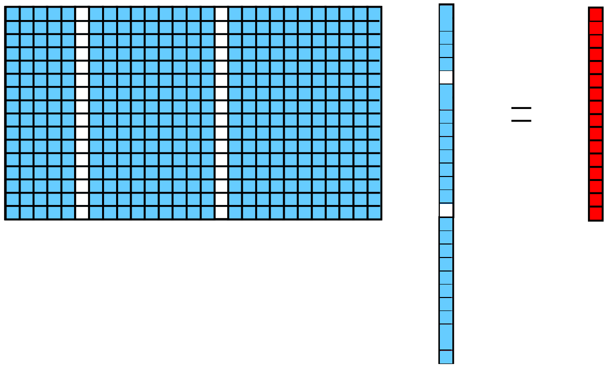
Step 4: $\mu^{t+1} = \tau\mu^t$ and $t = t + 1$.

End While

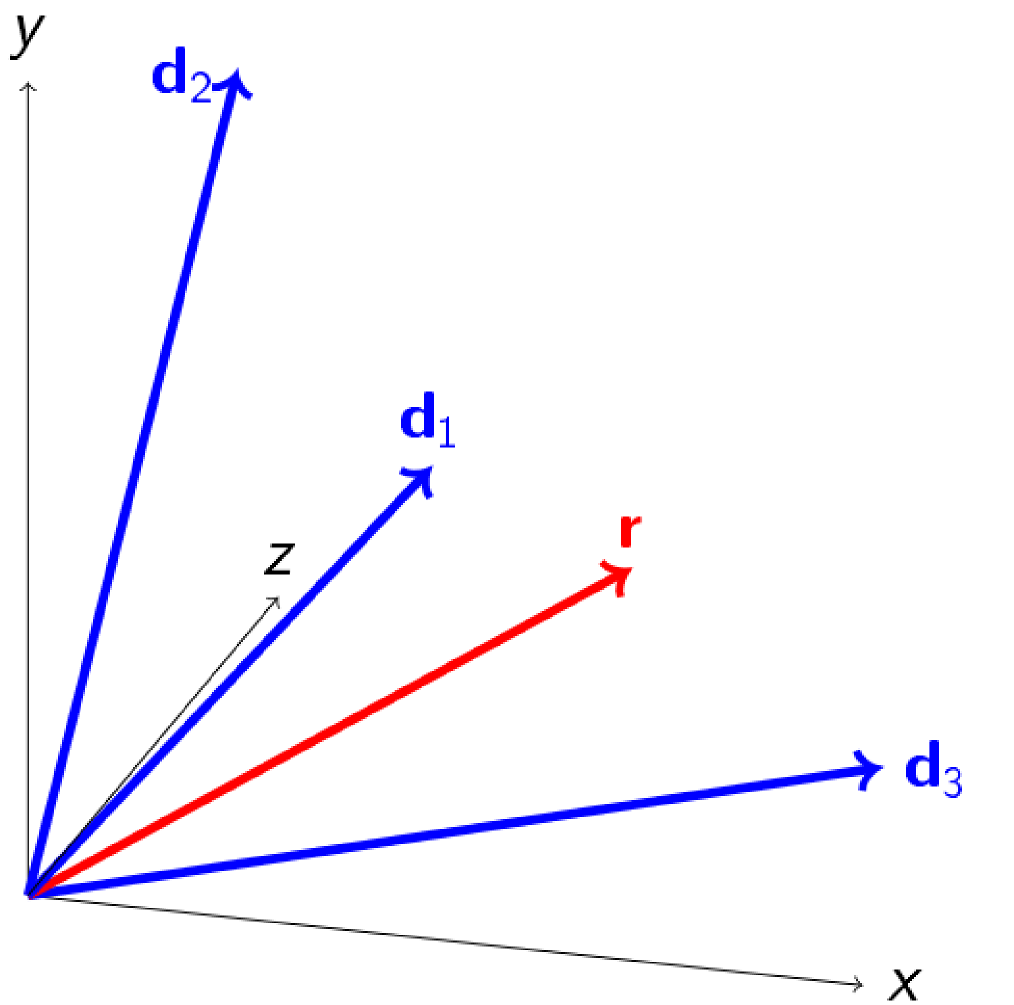
Output: α^{t+1}

Matching Pursuit Algorithms

- 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 …
- Repeat step 1 until convergence.

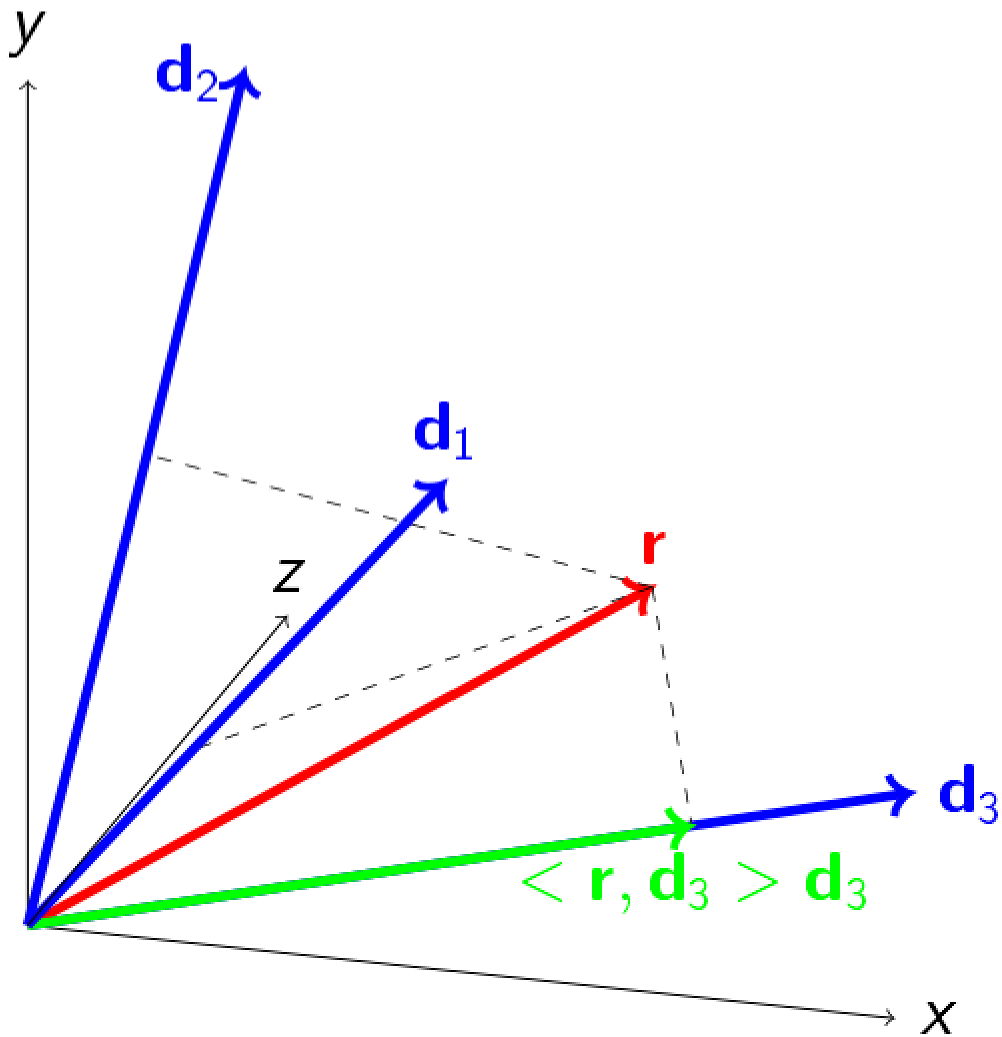


Matching Pursuit



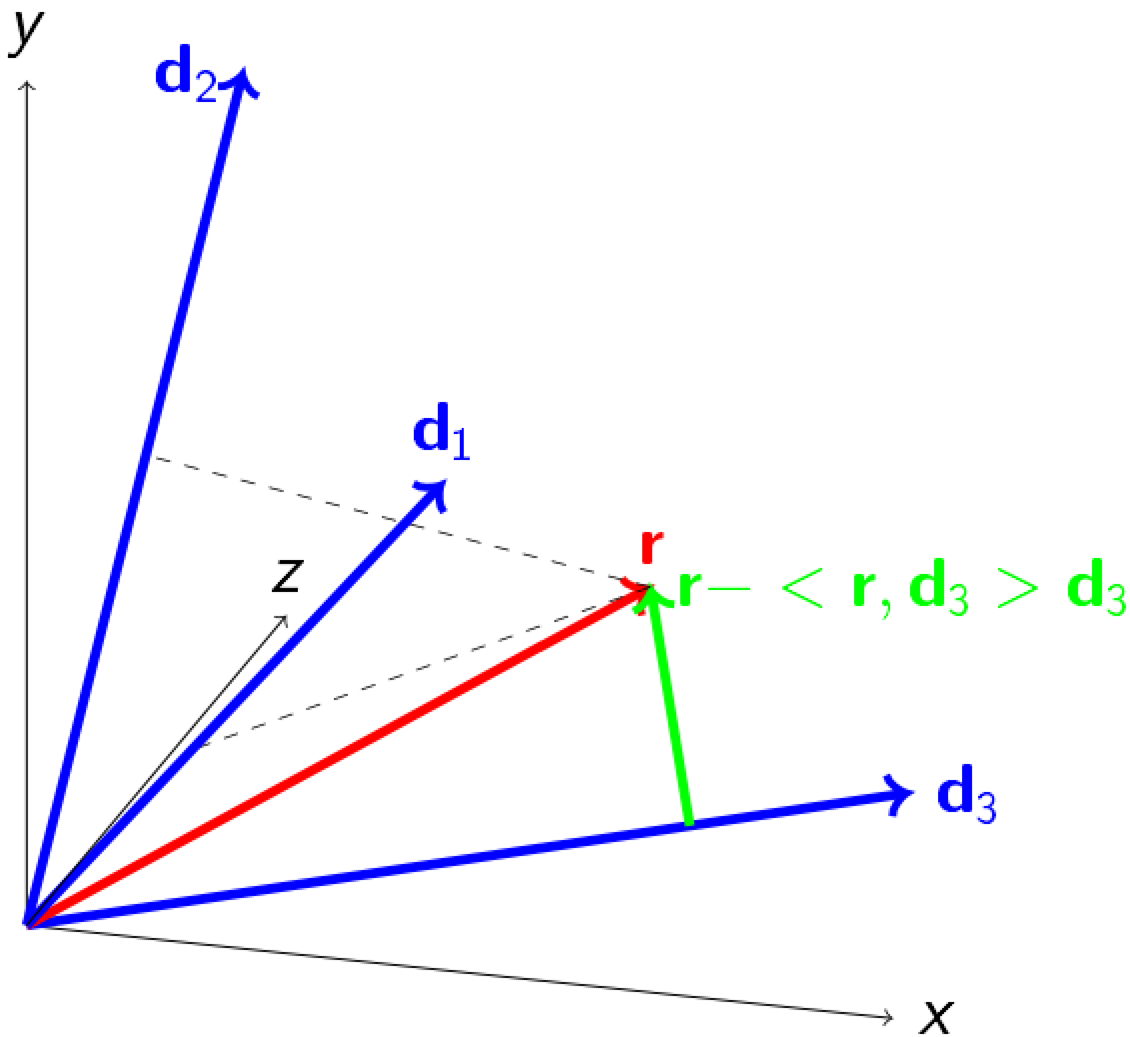
$\alpha=(0,0,0)$

Matching Pursuit



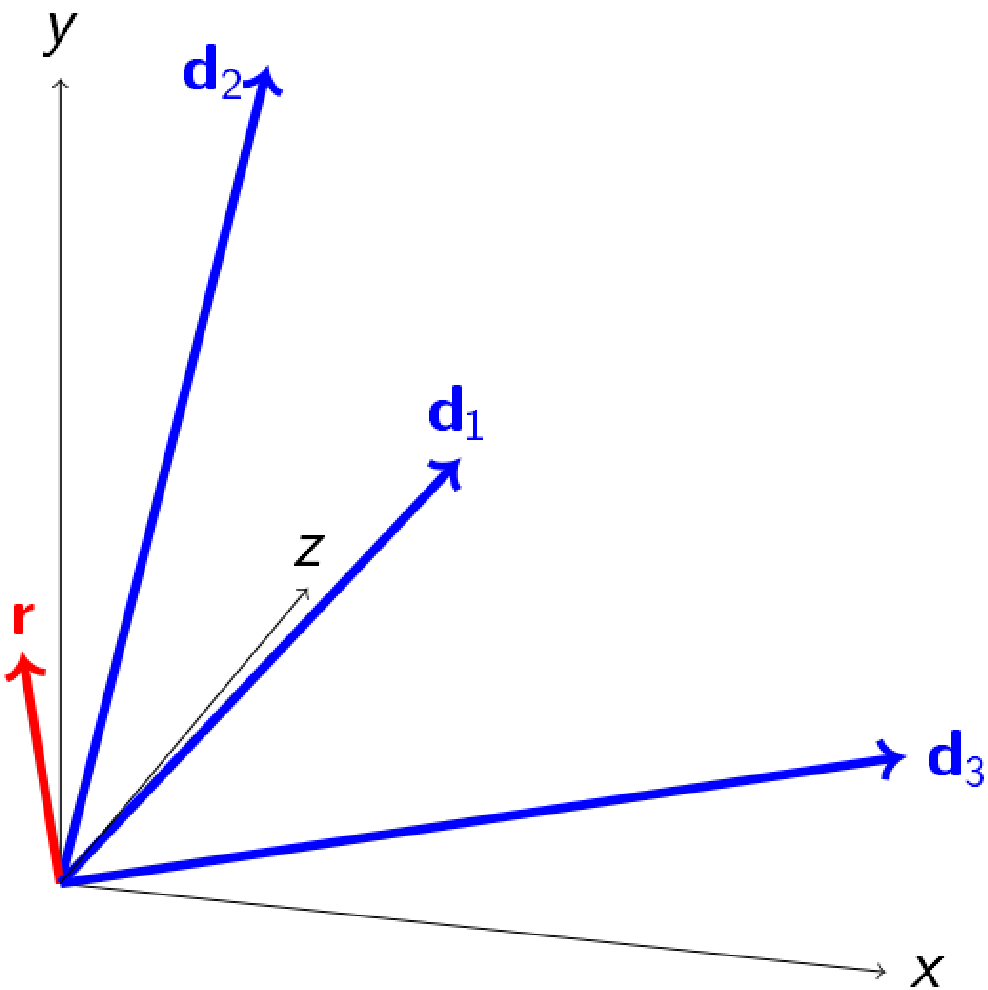
$\alpha=(0,0,0)$

Matching Pursuit



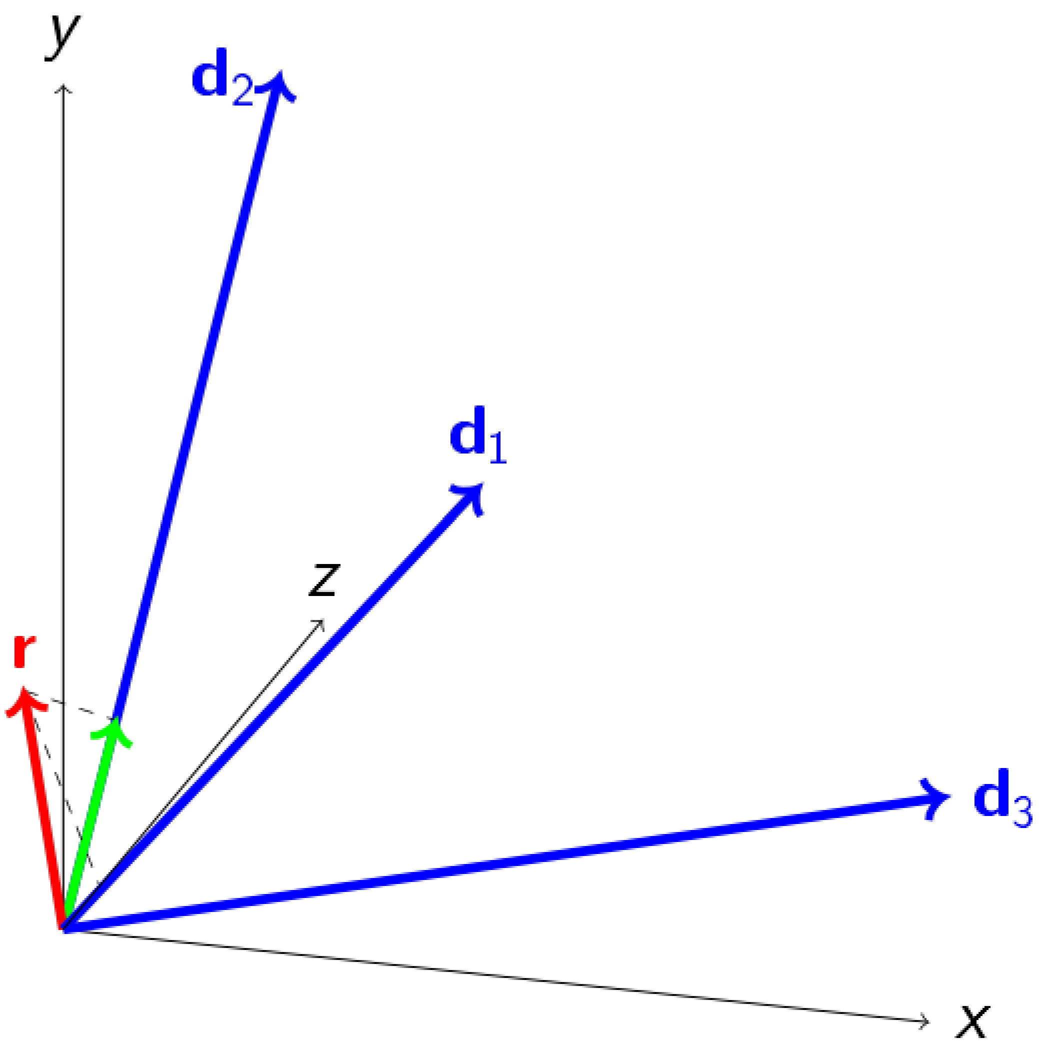
$\alpha=(0,0,0)$

Matching Pursuit



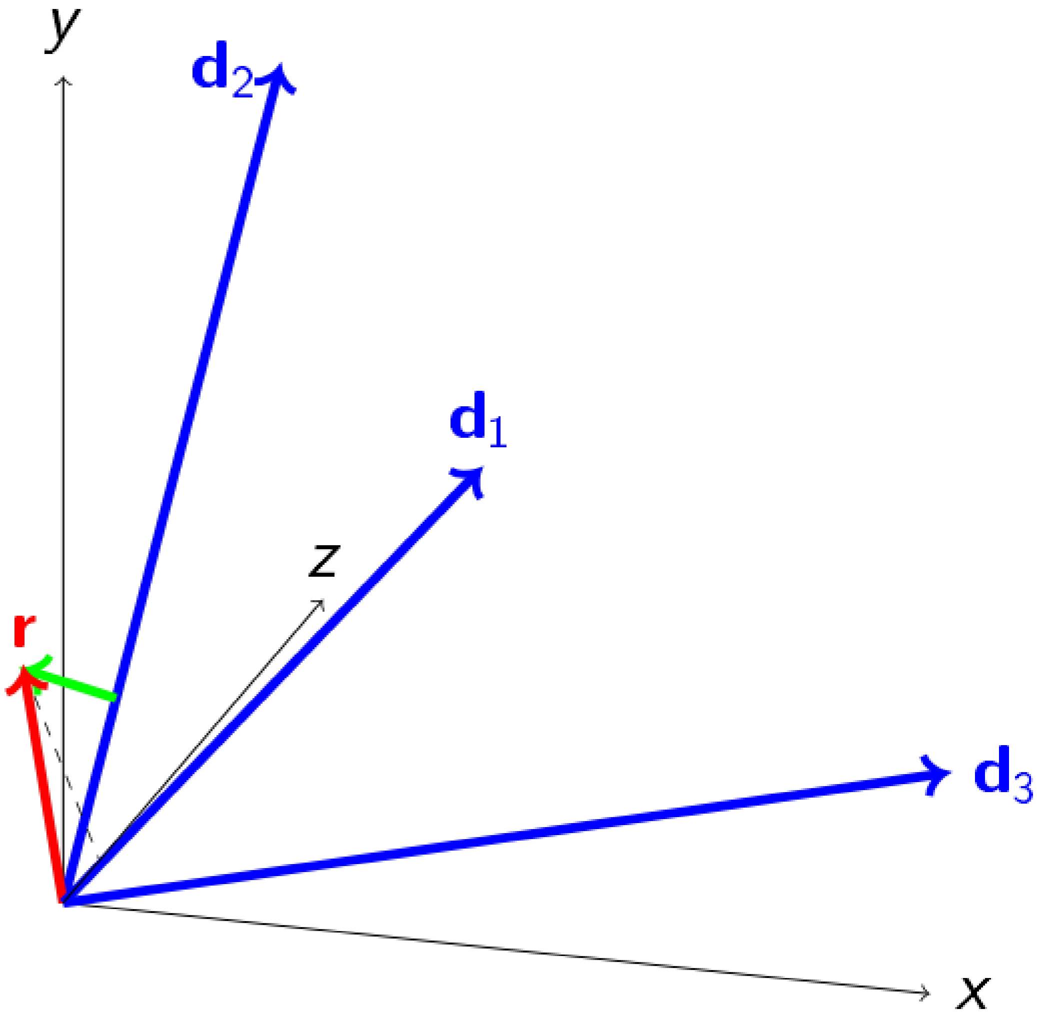
α=(0,0,0.75)

Matching Pursuit



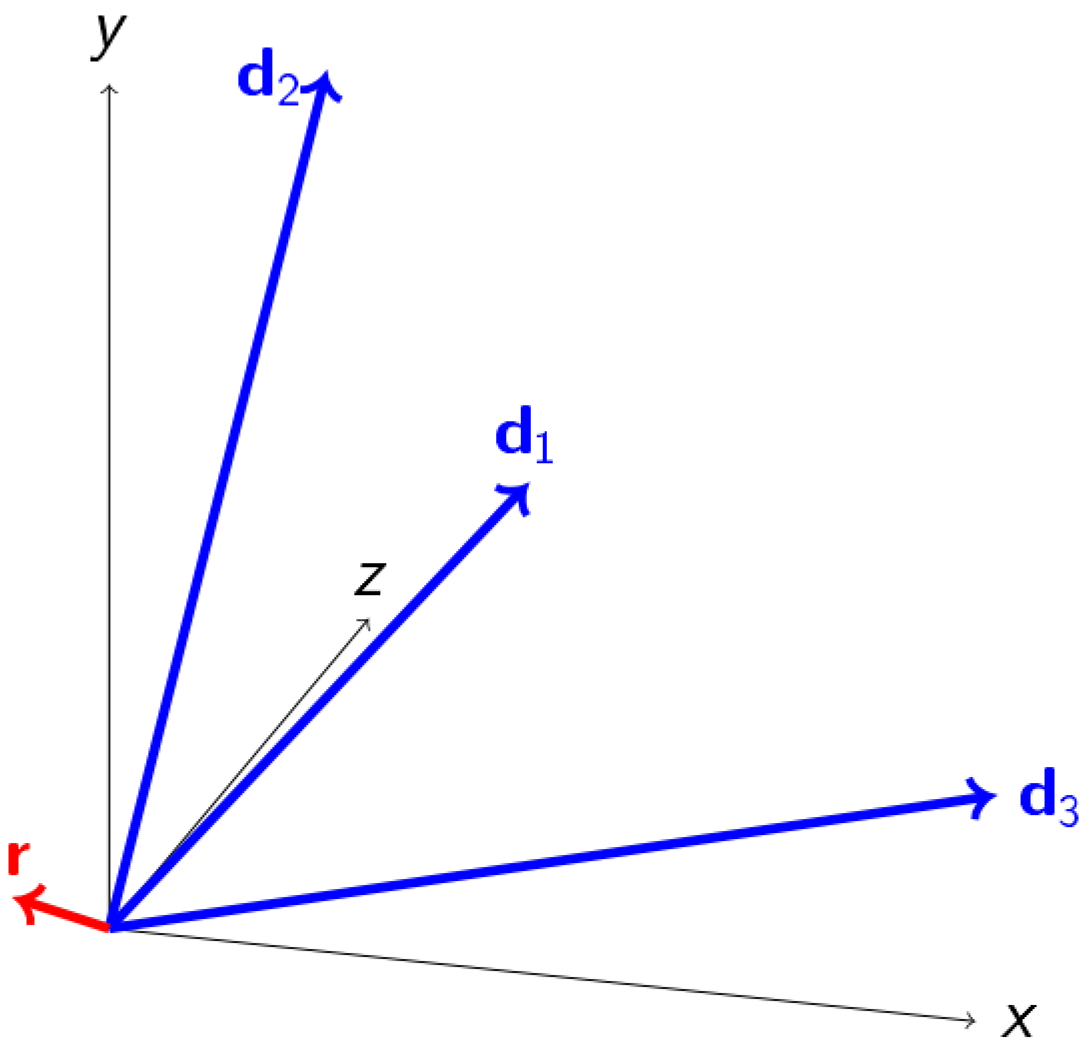
α=(0,0,0.75)

Matching Pursuit



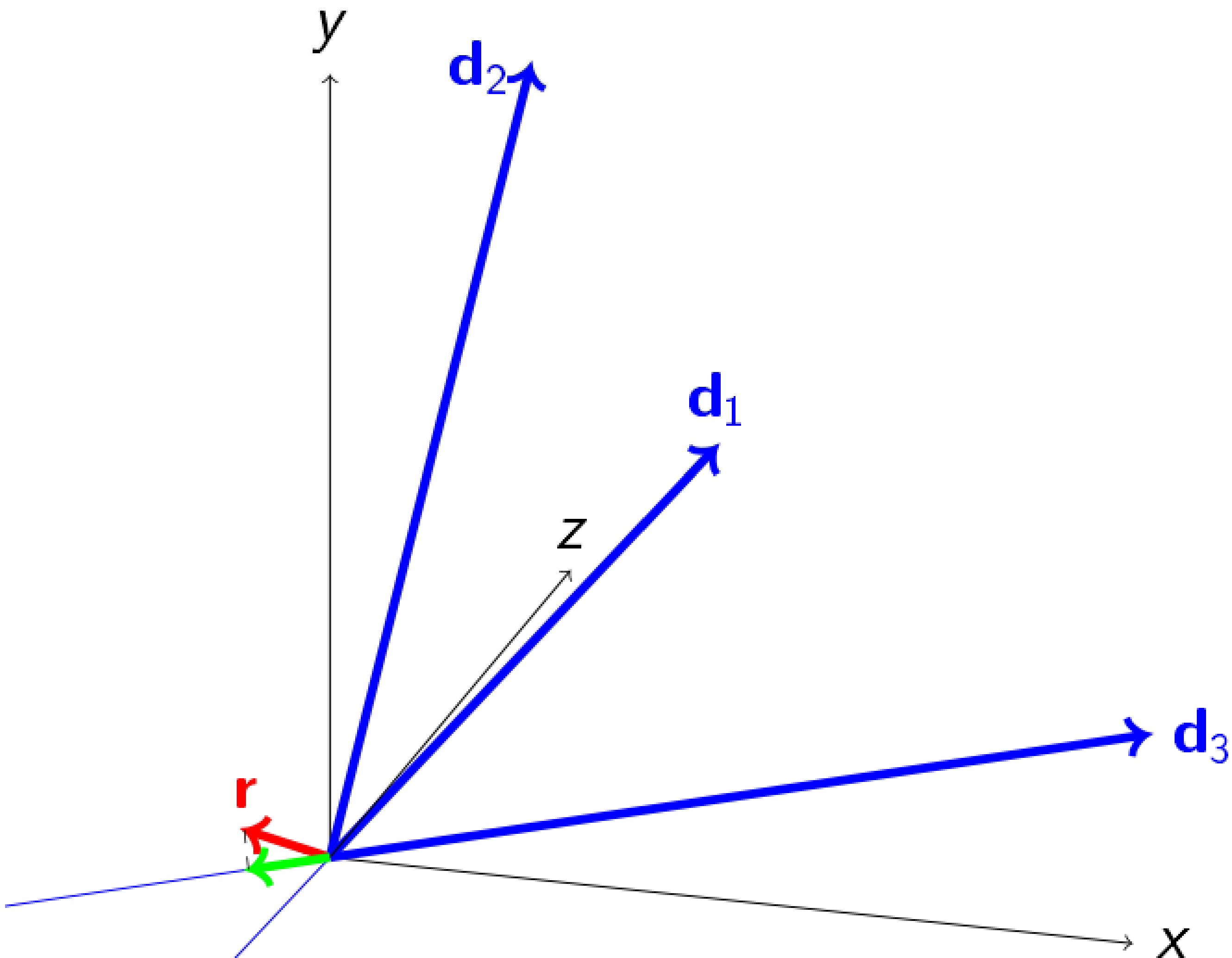
α=(0,0,0.75)

Matching Pursuit



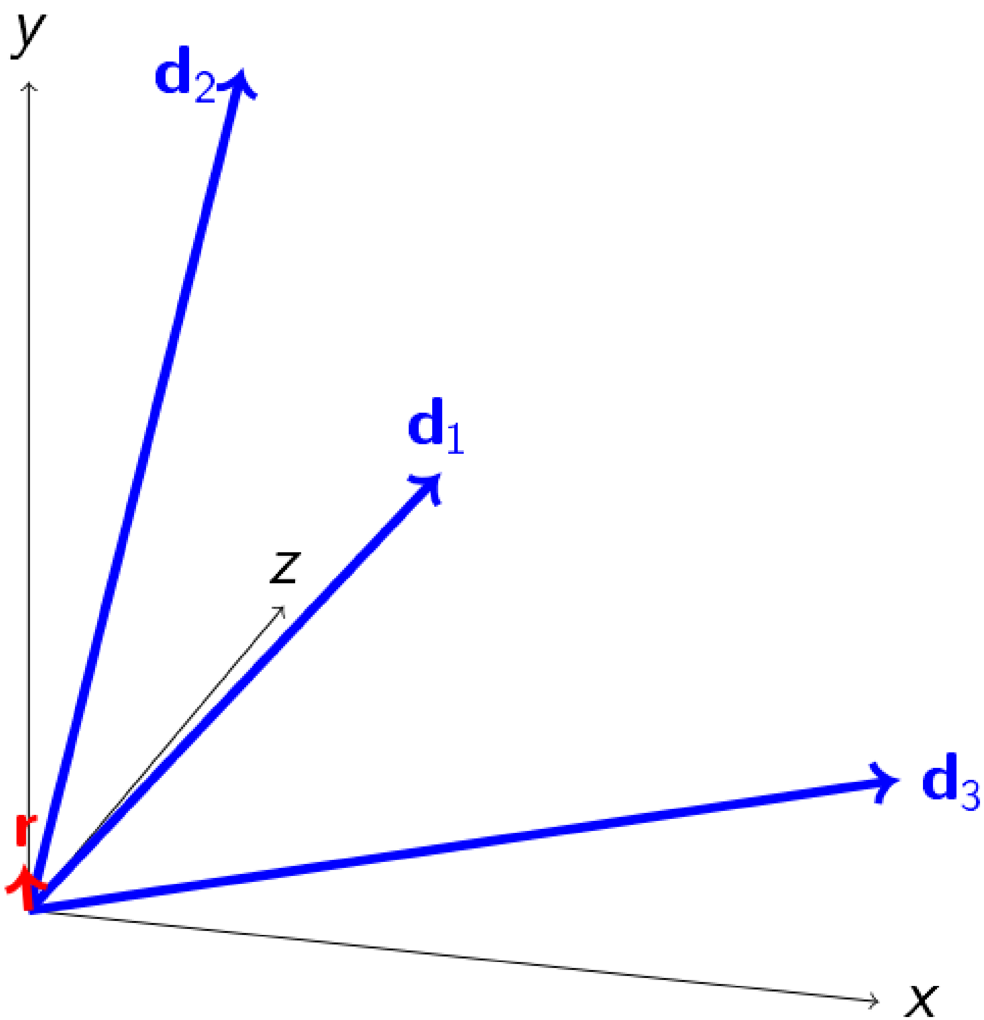
$\alpha=(0,0,0.75)$

Matching Pursuit



α=(0,0.24,0.75)

Matching Pursuit



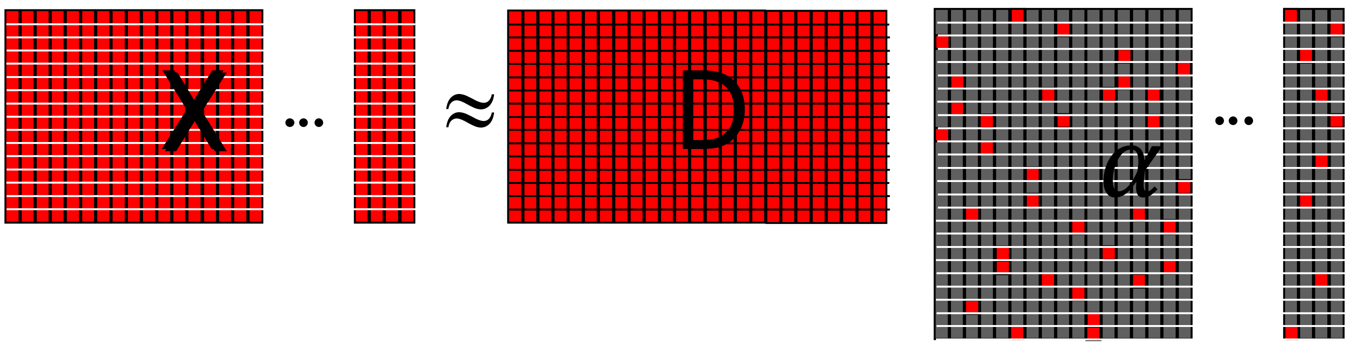
α=(0,0.24,0.65)

Matching Pursuit Algorithms

Algorithm 1 Greedy Matching Pursuit Algorithm

Input: Dictionary D, input signal x
Output: Sparse presentation vector α
Initialization:
$R \leftarrow x$
$n \leftarrow 1$
while $R_{n+1} > threshold$ **do**
  Find atom g_j with maximum inner product $| < R_n, g_j > |$
  $\alpha_i \leftarrow < R_n, g_j > / ||g_j||^2$
  $R_{n+1} \leftarrow R_n - \alpha_i g_j$
  $n \leftarrow n + 1$
end while
end

Matching Pursuit Algorithms



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

Each example is a linear combination of atoms from D

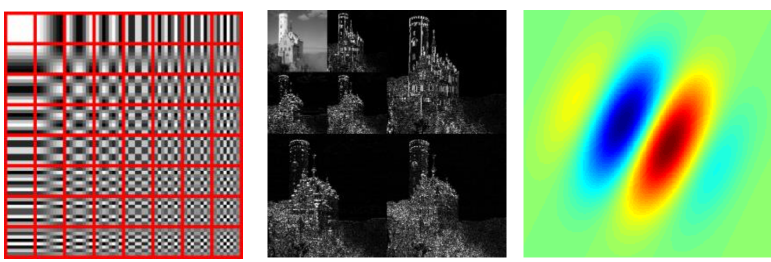
Each example has a sparse representation with no more than L atoms

Sparse Signal Modeling

Key idea $\min \|\mathbf{y} - \mathbf{D}\mathbf{s}\|_2 \ \text{ s.t. } \ \|\mathbf{s}\|_0 \leq K$

Greedy $\quad \|\mathbf{s}\|_1$

Dictionary learning



Parametric

$$\min \|\mathbf{Y} - \mathbf{D}\mathbf{S}\|_F$$
$$\text{s.t. } \ \|\mathbf{S}_i\|_1 \leq K, \|\mathbf{D}_i\|_2 \leq 1$$

K-SVD

Signal Processing

- Super resolution
- Denoising
- Demosaicing
- Enhancing

Computer Vision

- Image classification
- Object detection
- Tracking
- Change detection