Machine Learning for Signal Processing
Sparse and Overcomplete Representations

Bhiksha Raj
(slides from Sourish Chaudhuri and Abelino Jimenez)

1

So far

Can we use linear composition to identify **basic units** that compose the signal?

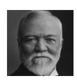
2

So far

Weights

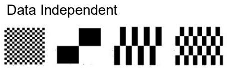
$$D \cdot \alpha = X$$
$$\approx$$

Data



Basis

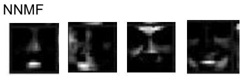
Data Independent



ICA

PCA

NNMF



basis <= dim X

3

Just in case you missed it..

- Remember, #(Basis Vectors)= #unknowns

$$D \cdot \alpha = X$$

Basis Vectors

Weights

Input data

Standard representations: number of bases <= dimension of data

4

A limitation we saw earlier

- Mathematical restrictions on the number of bases have no connection to reality
  - Universe does not respect your mathematical representations of the data
  - In reality: number of building blocks that compose any kind of data is unlimited
- One solution we saw earlier: picking *one* “closest” building block to represent any input
- Today: Learning linear compositional representations without restrictions on the number of basic units

5

Poll 1

6

Key Topics in this Lecture

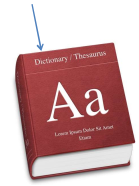
- Basics – Component-based representations
  - *Overcomplete* and Sparse Representations,
  - *Dictionaries*
- Pursuit Algorithms
- How to learn a dictionary
- Why is an overcomplete representation powerful?

Sparse and Overcomplete Representations 7

7

Representing Data

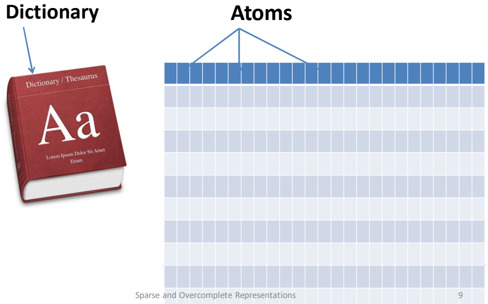
Dictionary (codebook)



Sparse and Overcomplete Representations 8

8

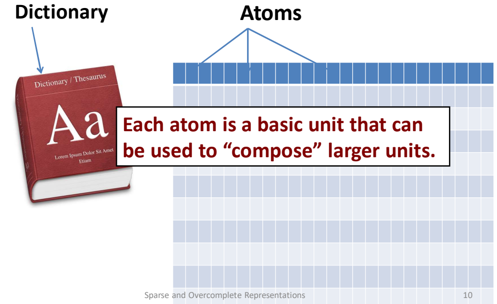
Representing Data



Sparse and Overcomplete Representations 9

9

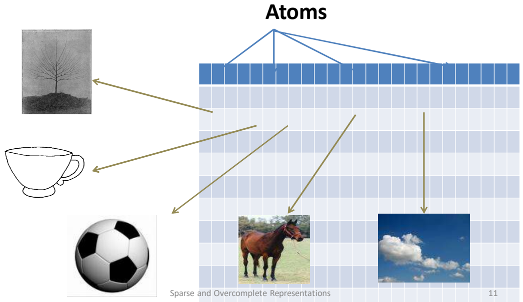
Representing Data



Sparse and Overcomplete Representations 10

10

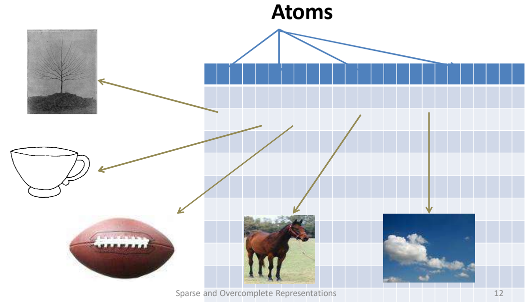
Representing Data



Sparse and Overcomplete Representations 11

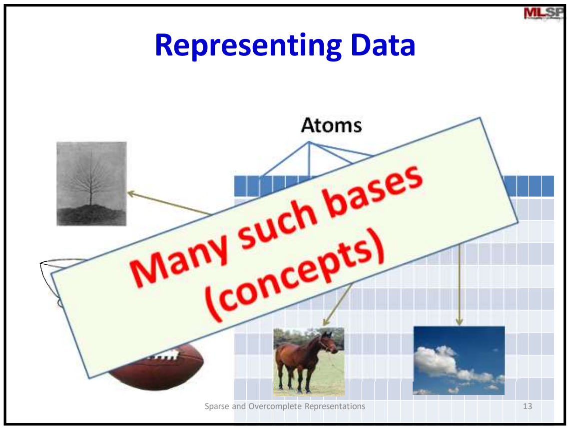
11

Representing Data

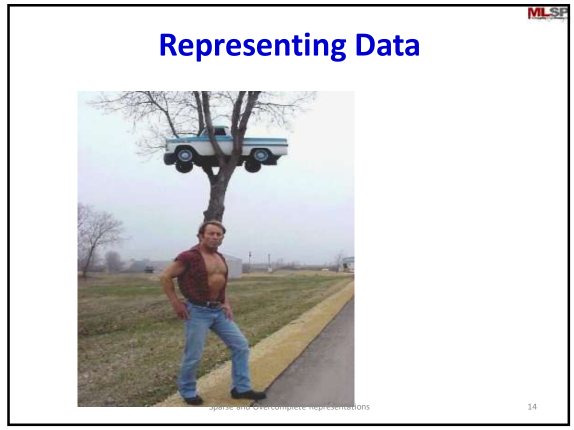


Sparse and Overcomplete Representations 12

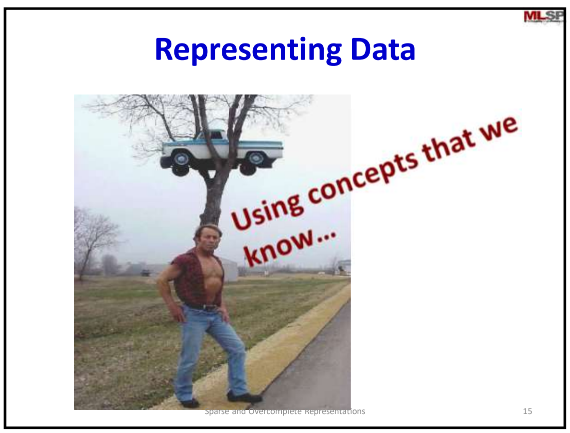
12



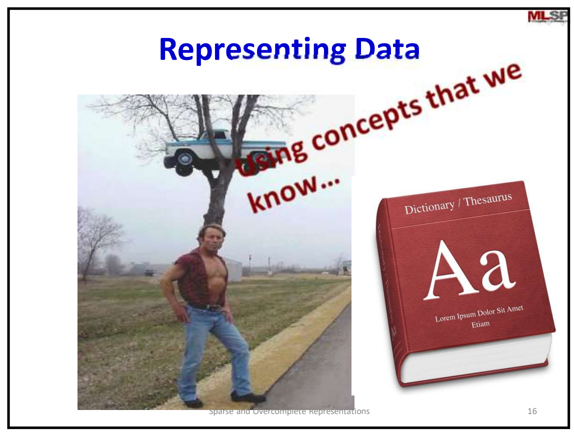
13



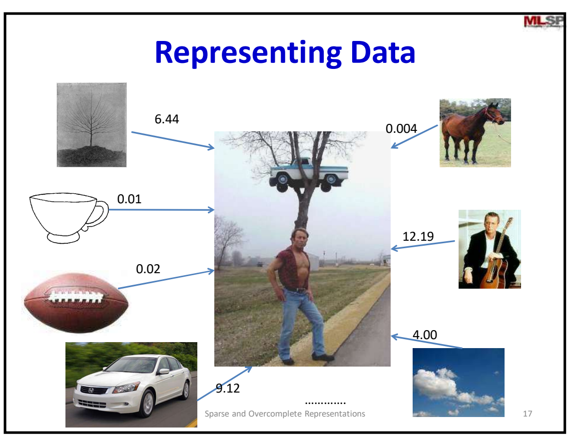
14



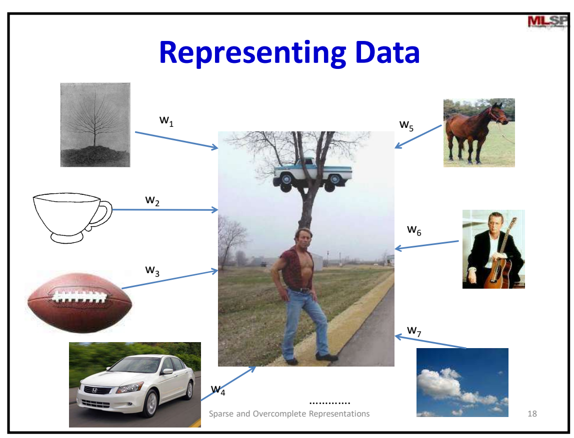
15



16

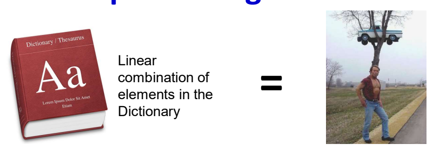


17



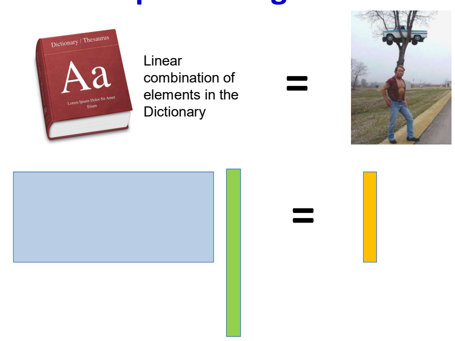
18

Representing Data



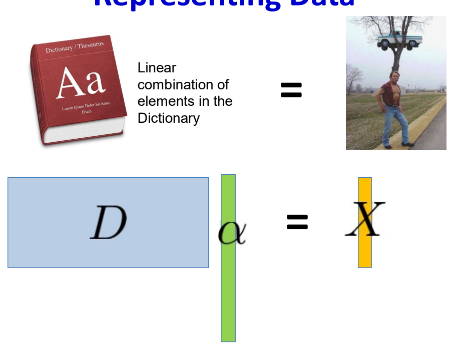
19

Representing Data



20

Representing Data



21

Quick Linear Algebra Refresher

- Remember, #(Basis Vectors)= #unknowns

$$D \cdot \alpha = X$$

Basis Vectors ***(from Dictionary)***

Weights

Input data

22

Overcomplete Representations

- What is the dimensionality of the input image? (say 64x64 image)
  - **4096**
- What is the dimensionality of the dictionary? (each image = 64x64 pixels)
  - **4096 x N**

23

Overcomplete Representations

- What is the dimensionality of the input image? (say 64x64 image)
  - **4096**
- What is the dimensionality of the dictionary? (each image = 64x64 pixels)
  - **4096 x N** ???

24

Overcomplete Representations

- What is the dimensionality of the input image? (say 64x64 image)
  - ➢ **4096**
- What is the dimensionality of the dictionary? (each image = 64x64 pixels)
  - ➢ **4096 x N**

VERY LARGE!!!

25

Overcomplete Representations

- What is the dimensionality of the input image? (say 64x64 image)

If N > 4096 (as it likely is)
we have an **overcomplete** representation

- What is the dimensionality of the dictionary? (each image = 64x64 pixels)
  - ➢ **4096 x N**

VERY LARGE!!!

26

Overcomplete Representations

- What is the dimensionality of the input image? (say 64x64 image)

More generally:

If #(dictionary units) > dimensions of input

we have an **overcomplete** representation

- ➢ **4096 x N**

VERY LARGE!!!

27

Quick Linear Algebra Refresher

- Remember, #(Basis Vectors)= #unknowns

$$D \cdot \alpha = X$$

Dictionary Units

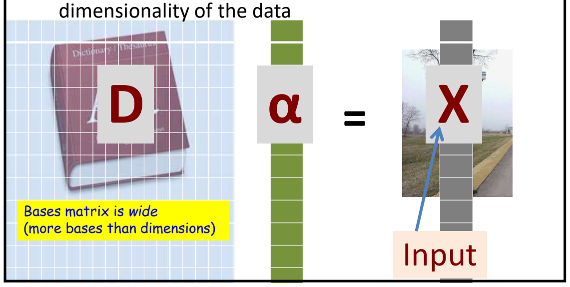
Weights

Input data

28

Dictionary based Representations

- Overcomplete "dictionary"-based representations are linear-composition-based representations with more "atomic building blocks" than the data dimensionality of the data



29

Why Dictionary-based Representations?

- Dictionary based representations are semantically more meaningful
- Enable content-based description
  - Bases can capture entire structures in data
  - E.g. notes in music
  - E.g. image structures (such as faces) in images
- Enable content-based processing
  - Reconstructing, separating, denoising, manipulating speech/music signals
  - Coding, compression, etc.
- Statistical reasons: We will get to that shortly..

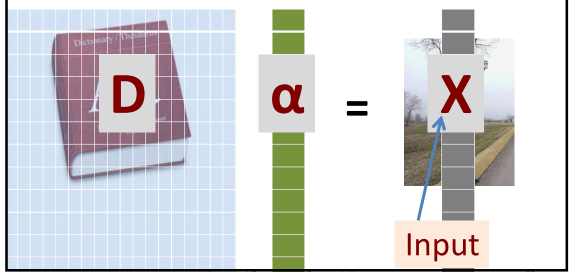
30

Poll 2

31

Problems

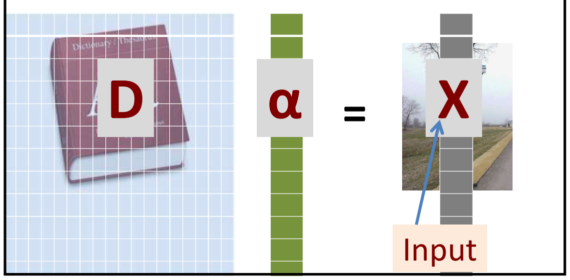
- How to obtain the dictionary
  - Which will give us meaningful representations
- How to compute the weights?



32

Problems

- How to obtain the dictionary
  - Which will give us meaningful representations
- How to compute the weights?



33

Quick Linear Algebra Refresher

- Remember, #(Basis Vectors)= #unknowns

$$D \cdot \alpha = X$$

Dictionary entries

Weights

Input data

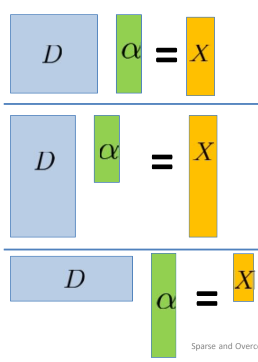
When can we solve for α?

34

Quick Linear Algebra Refresher

$$D \cdot \alpha = X$$

D: full rank



Unique solution

We may have no exact solution

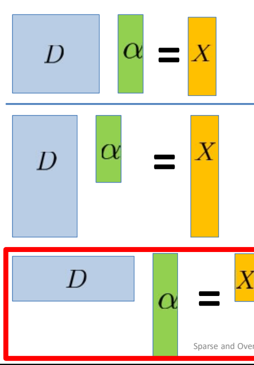
Infinite Solutions

35

Quick Linear Algebra Refresher

$$D \cdot \alpha = X$$

D: full rank



Unique solution

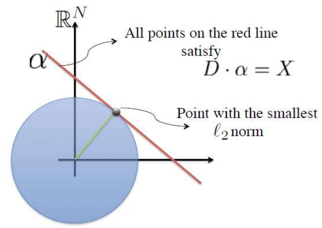
We may have no exact solution

Our Case

Infinite Solutions

36

Using Pseudo-Inverse?



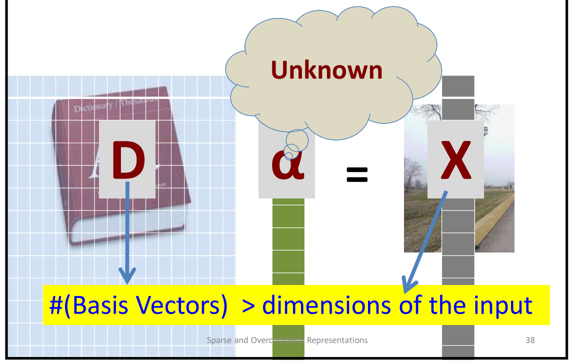
This is equivalent to

minimize $\|\alpha\|_2$ subject to $D\alpha = X$

α will generally be "dense"

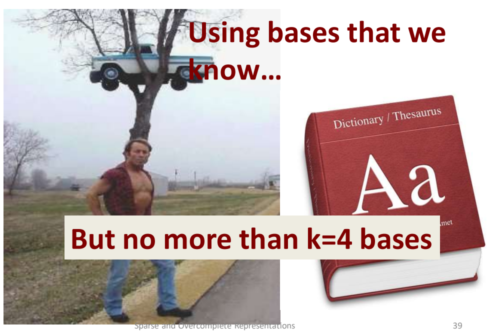
37

Overcomplete Representation



38

Representing Data

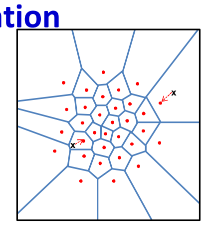


39

Alternate view: Recall quantization

$$V = \sum_i w_i d_i$$

$$V = \mathbf{D}\mathbf{w} \qquad |\mathbf{w}| = 1$$
$$|\mathbf{w}|_0 = 1$$



- d_i are the "representative" vectors of each cluster
- Restriction: only one of the w_i is 1, the rest are 0
  - $\sum_i w_i = 0$
  - $\mathbf{w}$ is unit length and one-sparse
- What if we let *more* than one entry of $\mathbf{w}$ to be non zero?

40

Overcompleteness and Sparsity

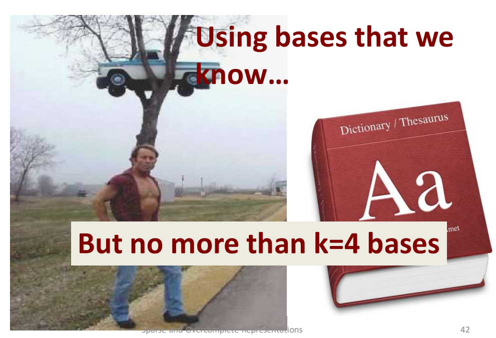
- To solve an overcomplete system of the type:

  D.α = X

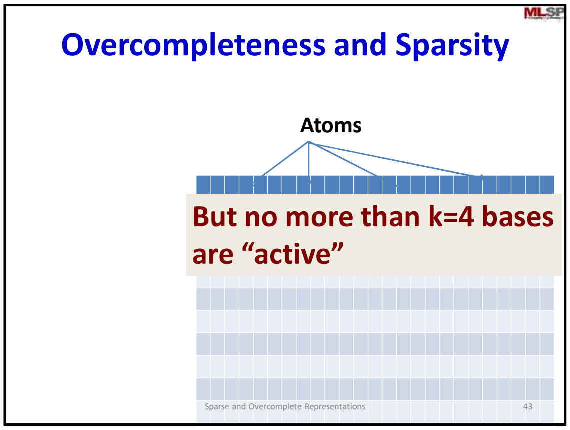
- Make assumptions about the data.
- Suppose, we say that **X** is composed of no more than a fixed number (**k**) of "bases" from **D** (k ≤ dim(**X**))
  - The term "bases" is an abuse of terminology..
- Now, we can find the set of **k** bases that best fit the data point, **X**.

41

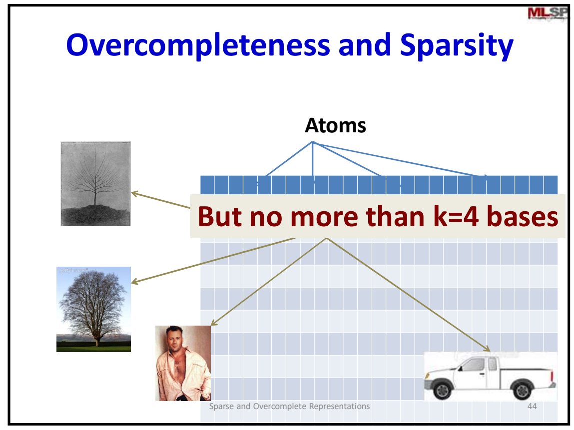
Representing Data



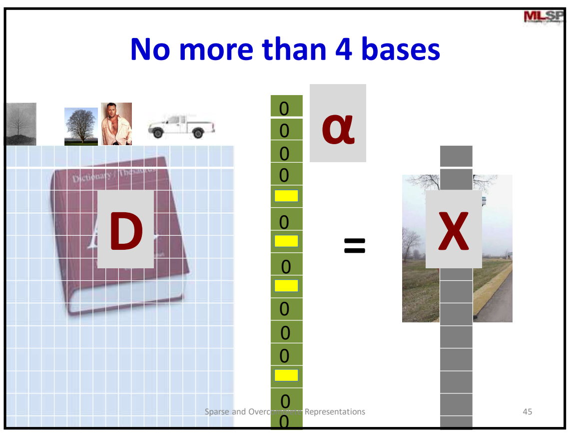
42



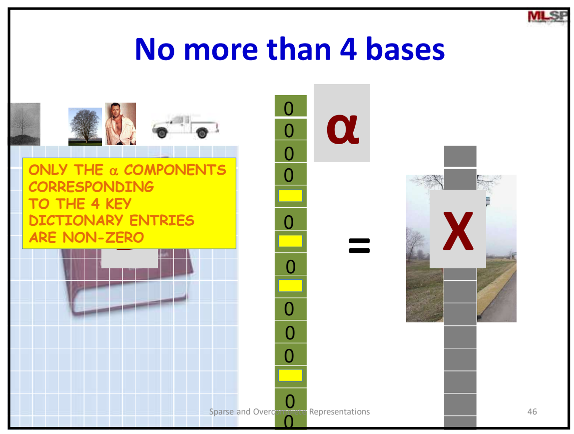
43



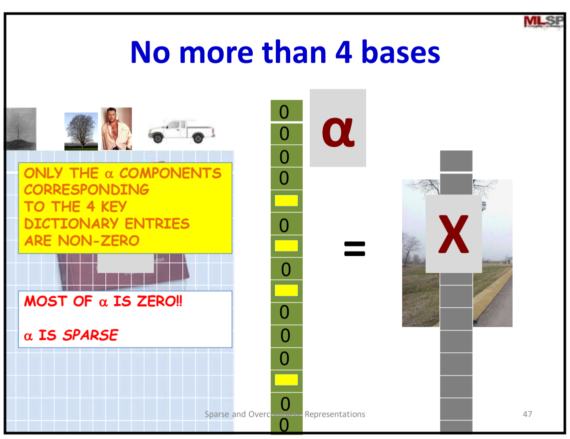
44



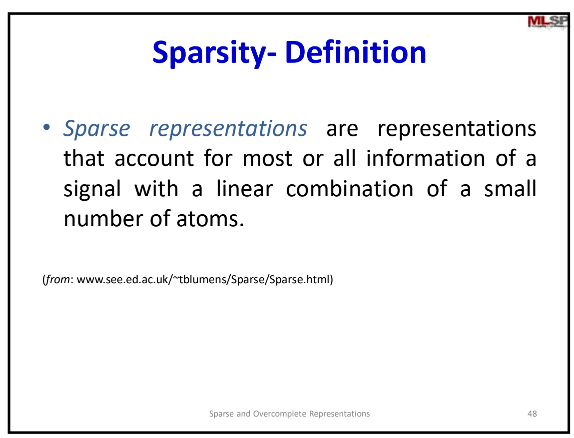
45



46



47



48

The Sparsity Problem

- We don't really know **k**
- You are given a signal **X**
- Assuming **X** was generated using the dictionary, can we find α that generated it?

49

The Sparsity Problem

- We want to use as few dictionary entries as possible to do this.

$$\boxed{\begin{aligned} &\underset{\underline{\alpha}}{Min}\ \|\underline{\alpha}\|_0 \\ &s.t.\ \ \underline{X} = \mathbf{D}\underline{\alpha} \end{aligned}}$$

50

The Sparsity Problem

- We want to use as few dictionary entries as possible to do this.

$$\boxed{\begin{aligned} &\underset{\underline{\alpha}}{Min}\ \|\underline{\alpha}\|_0 \\ &s.t.\ \ \underline{X} = \mathbf{D}\underline{\alpha} \end{aligned}}$$

Counts the number of non-zero elements in α

51

The Sparsity Problem

- We want to use **as few dictionary entries** as possible to do this
  - Ockham's razor: Choose the simplest explanation invoking the fewest variables

$$\boxed{\begin{aligned} &\underset{\underline{\alpha}}{Min}\ \|\underline{\alpha}\|_0 \\ &s.t.\ \ \underline{X} = \mathbf{D}\underline{\alpha} \end{aligned}}$$

52

Poll 3

53

The Sparsity Problem

- We want to use as few dictionary entries as possible to do this.

$$\boxed{\begin{aligned} &\underset{\underline{\alpha}}{Min}\ \|\underline{\alpha}\|_0 \\ &s.t.\ \ \underline{X} = \mathbf{D}\underline{\alpha} \end{aligned}}$$

How can we solve the above?

54

Obtaining Sparse Solutions

- We will look at 2 algorithms:
  - Matching Pursuit (MP)
  - Basis Pursuit (BP)

55

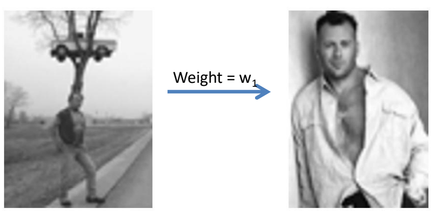
Matching Pursuit (MP)

- Greedy algorithm
- Finds an atom in the dictionary that best matches the input signal
- Remove the weighted value of this atom from the signal
- Again, find an atom in the dictionary that best matches the remaining signal.
- Continue till a defined stop condition is satisfied.

56

Matching Pursuit

- Find the dictionary atom that best matches the given signal.



57

Matching Pursuit

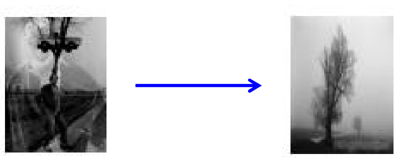
- Remove weighted image to obtain updated signal

Find best match for this signal from the dictionary

58

Matching Pursuit

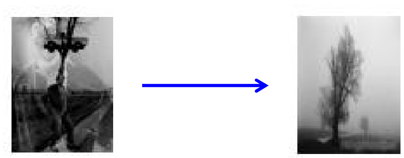
- Find best match for updated signal



59

Matching Pursuit

- Find best match for updated signal



Iterate till you reach a stopping condition,
norm(ResidualInputSignal) < threshold

60

Matching Pursuit

```
Algorithm Matching Pursuit
 Input: Signal: f(t).
 Output: List of coefficients: (a_n, g_γn).
 Initialization:
   Rf_1 ← f(t);
 Repeat
   find g_γn ∈ D with maximum inner product < Rf_n, g_γn >;
   a_n ← < Rf_n, g_γn >;
   Rf_{n+1} ← Rf_n − a_n g_γn;
   n ← n + 1;
 Until stop condition (for example: || Rf_n || < threshold)
```

From http://en.wikipedia.org/wiki/Matching_pursuit

61

Matching Pursuit

- Problems ???

62

Matching Pursuit

- Main Problem
  - Computational complexity
  - The entire dictionary has to be searched at every iteration

63

Comparing MP and BP

Matching Pursuit	Basis Pursuit
Hard thresholding (remember the equations)	
Greedy optimization at each step	
Weights obtained using greedy rules	

64

Basis Pursuit (BP)

- Remember,

$$\boxed{\begin{array}{l} \underset{\underline{\alpha}}{Min} \ \|\underline{\alpha}\|_0 \\ s.t. \ \ \underline{X} = \mathbf{D}\underline{\alpha} \end{array}}$$

65

Basis Pursuit

- Remember,

$$\boxed{\begin{array}{l} \underset{\underline{\alpha}}{Min} \ \|\underline{\alpha}\|_0 \\ s.t. \ \ \underline{X} = \mathbf{D}\underline{\alpha} \end{array}}$$

In the general case, this is intractable

66

Basis Pursuit

- Remember,

$$\min_{\alpha} \|\alpha\|_0 \quad s.t. \; X = \mathbf{D}\alpha$$

In the general case, this is intractable

Requires combinatorial optimization

67

Basis Pursuit

- Replace the intractable expression by an expression that is solvable

$$\min_{\alpha} \|\alpha\|_1 \quad s.t. \; X = \mathbf{D}\alpha$$

68

Basis Pursuit

- Replace the intractable expression by an expression that is solvable

$$\min_{\alpha} \|\alpha\|_1 \quad s.t. \; X = \mathbf{D}\alpha$$

This will provide identical solutions when $\mathbf{D}$ obeys the ***Restricted Isometry Property***.

69

Basis Pursuit

- Replace the intractable expression by an expression that is solvable

$$\min_{\alpha} \|\alpha\|_1 \quad s.t. \; X = \mathbf{D}\alpha$$

Objective

Constraint

70

Basis Pursuit

- We can formulate the optimization term as:

$$\min_{\alpha} \{\|X - \mathbf{D}\alpha\|^2 + \lambda\|\alpha\|_1\}$$

Constraint Objective

71

Basis Pursuit

- We can formulate the optimization term as:

$$\min_{\alpha} \{\|X - \mathbf{D}\alpha\|^2 + \lambda\|\alpha\|_1\}$$

λ is a penalty term on the non-zero elements and promotes sparsity

72

Basis Pursuit

Equivalent to *LASSO*; for more details, see [this paper by Tibshirani](http://www-stat.stanford.edu/~tibs/ftp/lasso.ps)
http://www-stat.stanford.edu/~tibs/ftp/lasso.ps

$$\underset{\alpha}{Min}\ \{\|X - \mathbf{D}\alpha\|^2 + \lambda\|\alpha\|_1\}$$

λ is a penalty term on the non-zero elements and promotes sparsity

73

Basis Pursuit

$$\underset{\alpha}{Min}\{\|X - \mathbf{D}\alpha\|^2 + \lambda\|\alpha\|_1\}$$

$$\frac{\partial\|\alpha\|_1}{\partial\alpha_j} = \begin{cases} +1 & \text{at } \alpha_j > 0 \\ [-1,1] & \text{at } \alpha_j = 0 \\ -1 & \text{at } \alpha_j < 0 \end{cases}$$

- $\|\alpha\|_1$ is not differentiable at $\alpha_j = 0$
- Gradient of $\|\alpha\|_1$ for gradient descent update
- At optimum, following conditions hold

$$\nabla_j\|X - \mathbf{D}\alpha\|^2 + \lambda sign(\alpha_j) = 0, \quad \text{if } |\alpha_j| > 0$$
$$\nabla_j\|X - \mathbf{D}\alpha\|^2 \le \lambda, \quad \text{if } \alpha_j = 0$$

74

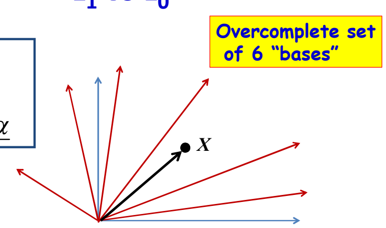
Basis Pursuit

- There are efficient ways to solve the LASSO formulation.
  - http://web.stanford.edu/~hastie/glmnet_matlab/
- Simplest solution: Coordinate descent algorithms
  - On webpage..

75

L_1 vs L_0

$$\underset{\alpha}{Min}\|\alpha\|_0$$
$$s.t.\ X = \mathbf{D}\alpha$$

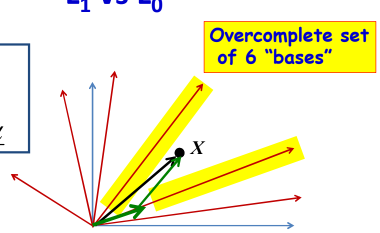


- **L_0 minimization**
  - Two-sparse solution
  - ANY pair of bases can explain X with 0 error

76

L_1 vs L_0

$$\underset{\alpha}{Min}\|\alpha\|_1$$
$$s.t.\ X = \mathbf{D}\alpha$$



- **L_1 minimization**
  - Two-sparse solution
  - All else being equal, the two closest bases are chosen

77

Comparing MP and BP

Matching Pursuit	Basis Pursuit
Hard thresholding	Soft thresholding
(remember the equations)	
Greedy optimization at each step	Global optimization
Weights obtained using greedy rules	Can force N-sparsity with appropriately chosen weights

78

General Formalisms

- L_0 minimization $\begin{array}{l} \underset{\alpha}{Min} \|\alpha\|_0 \\ s.t.\ X = \mathbf{D}\alpha \end{array}$
- L_0 constrained optimization $\begin{array}{l} \underset{\alpha}{Min} \|X - \mathbf{D}\alpha\|_2^2 \\ s.t. \|\alpha\|_0 < C \end{array}$
- L_1 minimization $\begin{array}{l} \underset{\alpha}{Min} \|\alpha\|_1 \\ s.t.\ X = \mathbf{D}\alpha \end{array}$
- L_1 constrained optimization $\begin{array}{l} \underset{\alpha}{Min} \|X - \mathbf{D}\alpha\|_2^2 \\ s.t. \|\alpha\|_1 < C \end{array}$

79

Many Other Methods..

- Iterative Hard Thresholding (IHT)
- CoSAMP
- OMP
- …

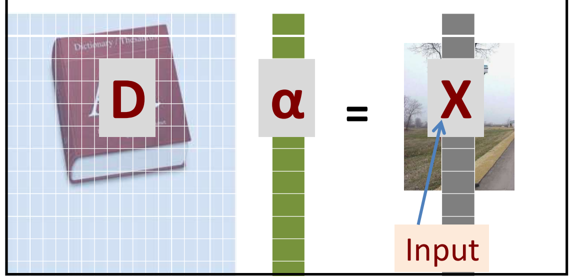
80

Poll 4

81

Problems

- How to obtain the dictionary
  - Which will give us meaningful representations
- How to compute the weights?



82

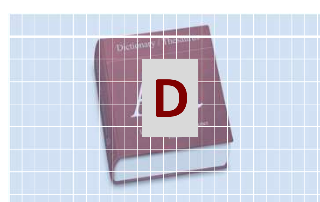
Trivial Solution

- D = Training data
- Impractical in most situations
  - Popular approach: sample random vectors from training data

83

Dictionaries: Compressive Sensing

- Just random vectors!



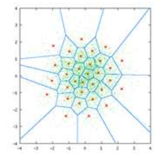
84

More Structured ways of Constructing Dictionaries

- Dictionary entries must be structurally "meaningful"
  – Represent true compositional units of data
- Have already encountered two ways of building dictionaries
  – NMF for non-negative data
  – K-means ..

85

K-Means for Composing Dictionaries

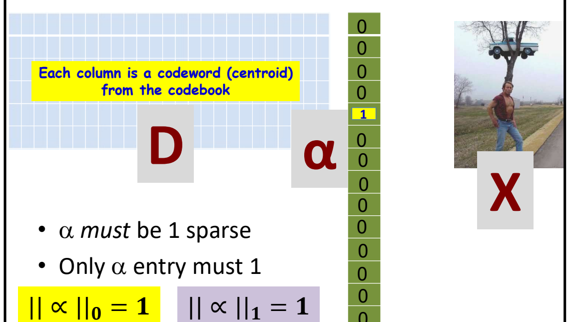


Train the codebook from training data using K-means

- Every vector is approximated by the centroid of the cluster it falls into
- Cluster means are "codebook" entries
  – Dictionary entries
  – Also compositional units the compose the data

86

K-Means for Dictionaries

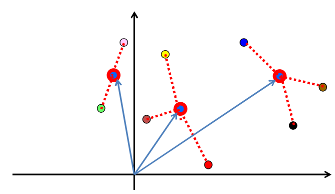


- α *must* be 1 sparse
- Only α entry must 1

$||\propto||_0 = 1$ $||\propto||_1 = 1$

87

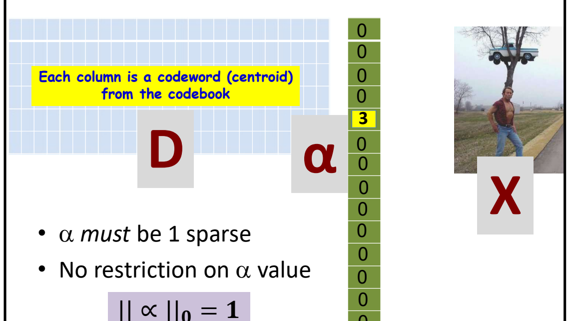
K-Means



- Learn Codewords to minimize the total squared length of the training vectors from the closest codeword

88

Length-unconstrained K-Means for Dictionaries

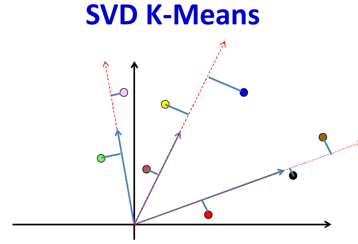


- α *must* be 1 sparse
- No restriction on α value

$||\propto||_0 = 1$

89

SVD K-Means

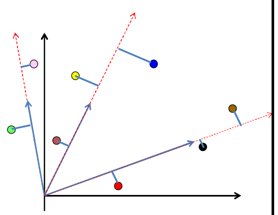


- Learn Codewords to minimize the total squared *projection error* of the training vectors from the closest codeword

90

SVD K–means

1. Initialize a set of centroids randomly
2. For each data point x, find the projection from the centroid for each cluster
   - $p_{cluster} = \left| x^T m_{cluster} \right|$
3. Put data point in the cluster of the closest centroid
   - Cluster for which $p_{cluster}$ is *maximum*
4. When all data points are clustered, recompute centroids

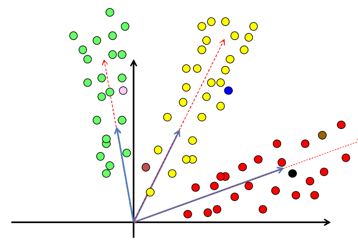


$$m_{cluster} = \text{Principal Eigenvector}(\{x \mid x \in cluster\})$$

91

91

Problem

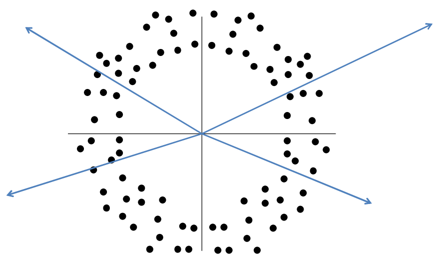


- Only represents *Radial* patterns

Sparse and Overcomplete Representations 92

92

What about this pattern?

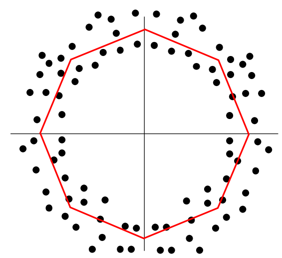


- Dictionary entries that represent radial patterns will not capture this structure
  - 1-sparse representations will not do

93

93

What about this pattern?

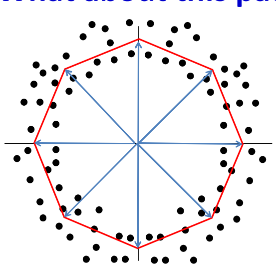


- We need ***AFFINE*** *patterns*

Sparse and Overcomplete Representations 94

94

What about this pattern?

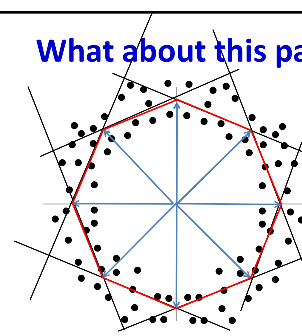


- We need ***AFFINE*** *patterns*
- *Each vector is modeled by a linear combination of K (here 2) bases*

95

95

What about this pattern?



Every line is a (constrained) combination of two bases

2-sparse

Constraint:
Line = a.b_1 + (1-a)b_2

- We need ***AFFINE*** *patterns*
- *Each vector is modeled by a linear combination of K (here 2) bases*

96

96

Codebooks for K sparsity?

Each column is a codeword (centroid) from the codebook

D

α

X

$$\alpha = [0, 0, 0, 0, 3, 0, 1, 0, 0, 4, 0, 0, 0, 0]^T$$

- α *must* be *k* sparse
- No restriction on α value

$|| \propto ||_0 = k$

97

Formalizing

Given training data

$$\{X_1, X_2, ..., X_T\}$$

We want to find a dictionary D, such that

$$D\alpha_i = X_i$$

With α_i sparse

98

Formalizing

Two objectives:

- Approximation $\|D\alpha_i - X_i\|$
- Sparsity in coefficients $\|\alpha_i\|_1$

$$\min_{D,\alpha_i} \sum_{i=1}^{T} \|X_i - D\alpha_i\|^2 + \|\alpha_i\|_1$$

NON-Convex!!!

99

An iterative method

- Given D, estimate α_i to get sparse solution
  - We can use any method

$$\min_{\alpha_i} \sum_{i=1}^{T} \|X_i - D\alpha_i\|^2 + \|\alpha_i\|_1$$

- Given α_i, estimate D

$$\min_{D} \sum_{i=1}^{T} \|X_i - D\alpha_i\|^2$$

Difficult!

100

K SVD

- Initialize Codebook

D =

1. For every vector, compute K-sparse alphas
   - Using any pursuit algorithm

α =

0	0	2	0	0
1	0	0	0	0
0	0	0	1	0
0	7	0	0	0
0	0	0	1	0
3	0	0	0	0
0	5	0	0	1
0	0	1	0	2

101

K-SVD

D = (D_j j != 1)

2. For each codeword (k):
- For each vector x that used k
  - Subtract the contribution of all other codewords to obtain $e_k(x)$
    - Codeword-specific residual
- Compute the principal Eigen vector of $\{e_k(x)\}$

3. Return to step 1

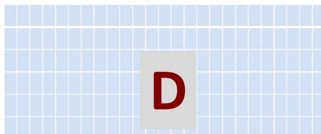
α = ($\alpha_j(x)$ j != 1)

0	0	2	0	0
1	0	0	0	0
0	0	0	1	0
0	7	0	0	0
0	0	0	1	0
3	0	0	0	0
0	5	0	0	1
0	0	1	0	2

$$e_k(x) = x - \sum_{j \neq k} \propto_j D_j$$

102

K-SVD



- Termination of each iteration: Updated dictionary
- Conclusion: A dictionary where any data vector can be composed of at most K dictionary entries
  - More generally, sparse composition

103

K-SVD algorithm (skip)

Initialization : Set the random normalized dictionary matrix $\mathbf{D}^{(0)} \in \mathbb{R}^{n\times K}$. Set $J = 1$.
Repeat until convergence,
Sparse Coding Stage: Use any pursuit algorithm to compute $\mathbf{x}_i$ for $i = 1, 2, \ldots, N$

$$\min_{\mathbf{x}} \{\|\mathbf{y}_i - \mathbf{D}\mathbf{x}\|_2^2\} \quad \text{subject to} \quad \|\mathbf{x}\|_0 \le T_0.$$

Codebook Update Stage: For $k = 1, 2, \ldots, K$

- Define the group of examples that use $\mathbf{d}_k$, $\omega_k = \{i|\ 1 \le i \le N,\ \mathbf{x}_i(k) \ne 0\}$.
- Compute

$$\mathbf{E}_k = \mathbf{Y} - \sum_{j\ne k} \mathbf{d}_j \mathbf{x}^j,$$

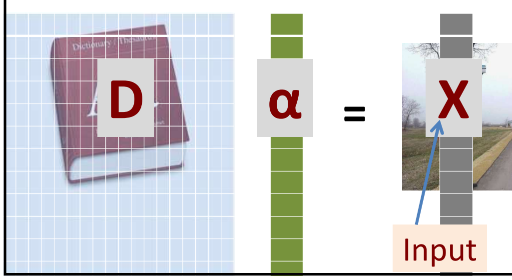
- Restrict $\mathbf{E}_k$ by choosing only the columns corresponding to those elements that initially used $\mathbf{d}_k$ in their representation, and obtain $\mathbf{E}_k^R$.
- Apply SVD decomposition $\mathbf{E}_k^R = \mathbf{U}\mathbf{\Delta}\mathbf{V}^T$. Update: $\mathbf{d}_k = \mathbf{u}_1, \mathbf{x}_R^k = \mathbf{\Delta}(1,1)\cdot \mathbf{v}_1$

Set $J = J + 1$.

104

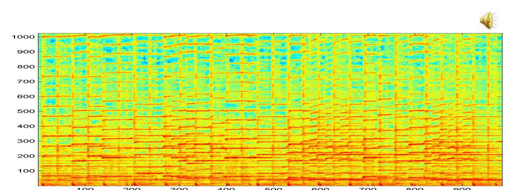
Problems

- How to obtain the dictionary
  - Which will give us meaningful representations
- How to compute the weights?



105

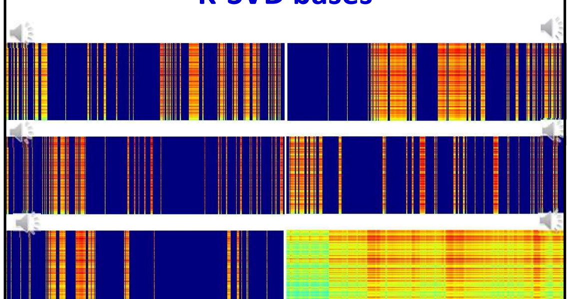
So how does that work



- In case you forgot this music…
- 975 vectors (1025 dimensions)
- N=12, K=5

106

K-SVD bases



107

Applications of Sparse Representations

- Many many applications
  - Signal representation
  - Statistical modelling
  - ..
  - We've seen one: Compressive sensing
- Another popular use
  - **Denoising**

108

Denoising

- As the name suggests, remove noise!

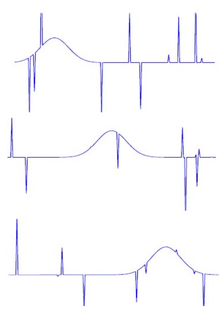
109

Denoising

- As the name suggests, remove noise!
- We will look at image denoising as an example

110

A toy example



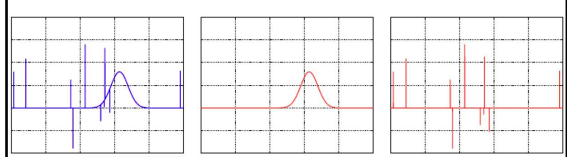
111

A toy example

$$D = [I\ G]$$

I Identity matrix

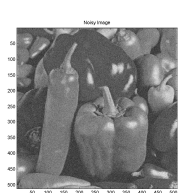
G Translation of a Gaussian pulse



112

Image Denoising

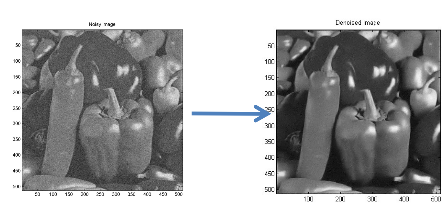
- Here's what we want



113

Image Denoising

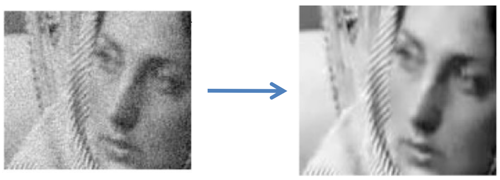
- Here's what we want



114

Image Denoising

- Here's what we want



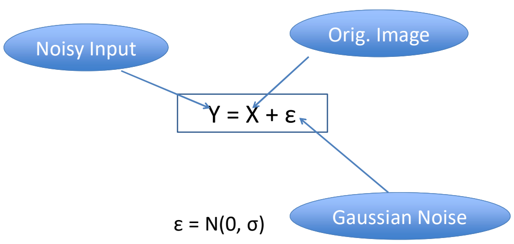
115

The Image Denoising Problem

- Given an image
- Remove Gaussian additive noise from it

116

Image Denoising



$\varepsilon = N(0, \sigma)$

117

Image Denoising

- Remove the noise from **Y**, to obtain **X** as best as possible.

118

Image Denoising

- Remove the noise from **Y**, to obtain **X** as best as possible
- Using sparse representations over learned dictionaries

119

Image Denoising

- Remove the noise from **Y**, to obtain **X** as best as possible
- Using sparse representations over learned dictionaries
- We will *learn* the dictionaries

120

Image Denoising

- Remove the noise from **Y**, to obtain **X** as best as possible
- Using sparse representations over learned dictionaries
- We will *learn* the dictionaries
- What data will we use? *The corrupted image itself!*

121

Image Denoising

- We use the data to be denoised to learn the dictionary.
- Training and denoising become an iterated process.
- We use image patches of size $\sqrt{n} \times \sqrt{n}$ pixels (i.e. if the image is 64x64, patches are 8x8)

122

Image Denoising

- The data dictionary D
  - Size = n x k (k > n)
  - This is known and fixed, to start with
  - Every image patch can be sparsely represented using D

123

Image Denoising

- Recall our equations from before.
- We want to find α so as to minimize the value of the equation below:

$$\underset{\underline{\alpha}}{Min}\ \{\|\underline{X} - \mathbf{D}\underline{\alpha}\|^2 + \lambda\|\underline{\alpha}\|_0\}$$

$$\underset{\underline{\alpha}}{Min}\{\|\underline{X} - \mathbf{D}\underline{\alpha}\|^2 + \lambda\|\underline{\alpha}\|_1\}$$

124

Image Denoising

$$\underset{\underline{\alpha}}{Min}\ \{\|\underline{X} - \mathbf{D}\underline{\alpha}\|^2 + \lambda\|\underline{\alpha}\|_1\}$$

- In the above, X is a patch.

125

Image Denoising

$$\underset{\underline{\alpha}}{Min}\ \{\|\underline{X} - \mathbf{D}\underline{\alpha}\|^2 + \lambda\|\underline{\alpha}\|_1\}$$

- In the above, X is a patch.
- If the larger image is fully expressed by the every patch in it, how can we go from patches to the image?

126

Image Denoising

$$\underset{\alpha_{ij},X}{Min}\ \{\mu\|X-Y\|_2^2+\sum_{ij}\|R_{ij}X-\mathbf{D}\alpha_{ij}\|_2^2+\sum_{ij}\lambda_{ij}\|\alpha_{ij}\|_0\}$$

Sparse and Overcomplete Representations 127

127

Image Denoising

$$\underset{\alpha_{ij},X}{Min}\ \{\mu\|X-Y\|_2^2+\sum_{ij}\|R_{ij}X-\mathbf{D}\alpha_{ij}\|_2^2+\sum_{ij}\lambda_{ij}\|\alpha_{ij}\|_0\}$$

(X – Y) is the error between the input and denoised image. μ is a penalty on the error.

Sparse and Overcomplete Representations 128

128

Image Denoising

$$\underset{\alpha_{ij},X}{Min}\ \{\mu\|X-Y\|_2^2+\sum_{ij}\|R_{ij}X-\mathbf{D}\alpha_{ij}\|_2^2+\sum_{ij}\lambda_{ij}\|\alpha_{ij}\|_0\}$$

Error bounding in each patch
- R_{ij} selects the $(ij)^{th}$ patch
- Terms in summation = no. of patches

Sparse and Overcomplete Representations 129

129

Image Denoising

$$\underset{\alpha_{ij},X}{Min}\ \{\mu\|X-Y\|_2^2+\sum_{ij}\|R_{ij}X-\mathbf{D}\alpha_{ij}\|_2^2+\sum_{ij}\lambda_{ij}\|\alpha_{ij}\|_0\}$$

λ forces sparsity

Sparse and Overcomplete Representations 130

130

Image Denoising

- But, we don't "***know***" our dictionary D.
- We want to estimate D as well.

Sparse and Overcomplete Representations 131

131

Image Denoising

- But, we don't "***know***" our dictionary D.
- We want to estimate D as well.

$$\underset{D,\alpha_{ij},X}{Min}\ \{\mu\|X-Y\|_2^2+\sum_{ij}\|R_{ij}X-\mathbf{D}\alpha_{ij}\|_2^2+\sum_{ij}\lambda_{ij}\|\alpha_{ij}\|_0\}$$

We can use the previous equation itself!!!

Sparse and Overcomplete Representations 132

132

Image Denoising

$$\underset{D,\alpha_{ij},X}{Min}\ \{\mu\|\underline{X}-Y\|_2^2+\sum_{ij}\|\underline{R_{ij}X}-\mathbf{D}\underline{\alpha}_{ij}\|_2^2$$
$$+\sum_{ij}\lambda_{ij}\|\underline{\alpha_{ij}}\|_0\}$$

How do we estimate all 3 at once?

133

Image Denoising

$$\underset{D,\alpha_{ij},X}{Min}\ \{\mu\|\underline{X}-Y\|_2^2+\sum_{ij}\|\underline{R_{ij}X}-\mathbf{D}\underline{\alpha}_{ij}\|_2^2$$
$$+\sum_{ij}\lambda_{ij}\|\underline{\alpha_{ij}}\|_0\}$$

How do we estimate all 3 at once?

We cannot estimate them at the same time!

134

Image Denoising

$$\underset{D,\alpha_{ij},X}{Min}\ \{\mu\|\underline{X}-Y\|_2^2+\sum_{ij}\|\underline{R_{ij}X}-\mathbf{D}\underline{\alpha}_{ij}\|_2^2$$
$$+\sum_{ij}\lambda_{ij}\|\underline{\alpha_{ij}}\|_0\}$$

How do we estimate all 3 at once?

Fix 2, and find the optimal 3rd.

135

Image Denoising

$$\underset{D,\alpha_{ij},X}{Min}\ \{\mu\|\underline{X}-Y\|_2^2+\sum_{ij}\|\underline{R_{ij}X}-\mathbf{D}\underline{\alpha}_{ij}\|_2^2$$
$$+\sum_{ij}\lambda_{ij}\|\underline{\alpha_{ij}}\|_0\}$$

Initialize X = Y

136

Image Denoising

$$\underset{\alpha_{ij}}{Min}\ \{\underbrace{\mu\|\underline{X}-Y\|_2^2}_{0}+\sum_{ij}\|\underline{R_{ij}X}-\mathbf{D}\underline{\alpha}_{ij}\|_2^2$$
$$+\sum_{ij}\lambda_{ij}\|\underline{\alpha_{ij}}\|_0\}$$

Initialize X = Y, initialize D

You know how to solve the remaining portion for α – MP, BP!

137

Image Denoising

- Now, update the dictionary D.
- Update D one column at a time, following the [K-SVD algorithm]
- K-SVD maintains the sparsity structure

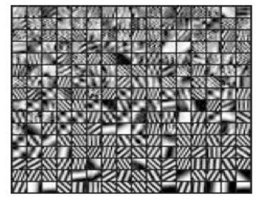
138

Image Denoising

- Now, update the dictionary D.
- Update D one column at a time, following the K-SVD algorithm
- K-SVD maintains the sparsity structure
- Iteratively update α and D

139

Image Denoising



Learned Dictionary for Face Image denoising

From: M. Elad and M. Aharon, *Image denoising via learned dictionaries and sparse representation*, CVPR, 2006.

140

Image Denoising

$$\underset{X}{Min}\ \{\mu\|X - Y\|_2^2 + \sum_{ij}\|R_{ij}X - \mathbf{D}\alpha_{ij}\|_2^2 + \sum_{ij}\lambda_{ij}\|\alpha_{ij}\|_0\}$$

Const. wrt X

We know D and α

The quadratic term above has a closed-form solution

141

Image Denoising

$$\underset{X}{Min}\ \{\mu\|X - Y\|_2^2 + \sum_{ij}\|R_{ij}X - \mathbf{D}\alpha_{ij}\|_2^2 + \sum_{ij}\lambda_{ij}\|\alpha_{ij}\|_0\}$$

Const. wrt X

We know D and α

$$X = (\mu I + \sum_{ij} R_{ij}^T R_{ij})^{-1}(\mu Y + \sum_{ij} R_{ij}^T D\alpha_{ij})$$

142

Image Denoising

- Summarizing… We wanted to obtain 3 things

143

Image Denoising

- Summarizing… We wanted to obtain 3 things

- Weights **α**
- Dictionary **D**
- Denoised Image **X**

144

Image Denoising

- Summarizing… We wanted to obtain 3 things

➢ Weights **α** – Your favorite pursuit algorithm
➢ Dictionary **D** – Using K-SVD
➢ Denoised Image **X**

145

Image Denoising

- Summarizing… We wanted to obtain 3 things

➢ Weights **α** – Your favorite pursuit algorithm
➢ Dictionary **D** – Using K-SVD
➢ Denoised Image **X**

Iterating

146

Image Denoising

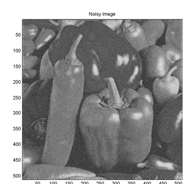
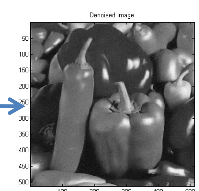
- Summarizing… We wanted to obtain 3 things

➢ Weights **α**
➢ Dictionary **D**
➢ Denoised Image **X**- Closed form solution

147

Image Denoising

- Here's what we want

148

Image Denoising

- Here's what we want


149

Comparing to Other Techniques

Non-Gaussian data

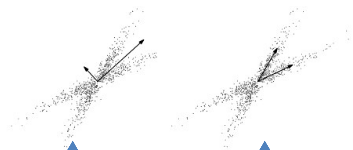

PCA of ICA Which is which?

Images from Lewicki and Sejnowski, *Learning Overcomplete Representations*, 2000.

150

Comparing to Other Techniques

Non-Gaussian data



PCA ICA

Images from Lewicki and Sejnowski, *Learning Overcomplete Representations*, 2000.

Sparse and Overcomplete Representations 151

151

Comparing to Other Techniques

Non-Gaussian data

PCA ICA

Images from Lewicki and Sejnowski, *Learning Overcomplete Representations*, 2000.

Sparse and Overcomplete Representations 152

152

Comparing to Other Techniques

Data still in 2-D space

ICA Overcomplete

Doesn't capture the underlying representation, which Overcomplete representations can do…

Sparse and Overcomplete Representations 153

153

Summary

- Overcomplete representations can be more powerful than component analysis techniques.
- Dictionary can be learned from data.
- Relative advantages and disadvantages of the pursuit algorithms.

Sparse and Overcomplete Representations 154

154