Wavelets, Sparsity and their Applications
Session 8: Sparse Signal Representation

Pier Luigi Dragotti
Imperial College London
p.dragotti@imperial.ac.uk

Outline

- **Signal Representation Problem**
  - Bases and Frames
  - Analysis and Synthesis Models
  - Wavelet Representation Revisited
- **Sparsity in Union of Bases:**
  - l_0 and l_1 optimizations
  - Sparse Representation Key Bounds
  - OMP and its performance
- **Sparsity according to Prony**
- **Approximate sparsity and iterative shrinkage algorithms**
- **Applications**
- **Beyond Traditional Sparsity**

Imperial College London

Signal Representations

$$x(t)=\sum_{i=-\infty}^{\infty}\alpha_i\varphi_i(t)$$

Key Ingredients:

- a set of 'atoms': $\{\varphi_i\}$
- a inner product: $\langle x,\varphi_i\rangle=\int x(t)\varphi_i(t)dt$
- a synthesis formula: $x(t)=\sum_{i=-\infty}^{\infty}\alpha_i\varphi_i(t)$

Many choices of $\{\varphi_i\}$

- orthonormal bases (e.g., Fourier series, Haar wavelet, Daubechies wavelets)
- biorthogonal bases (e.g., splines)
- overcomplete expansions or frames

Signal Representation: Frames

Today's focus is on overcomplete representations and on finite dimensional signals (i.e., vectors)

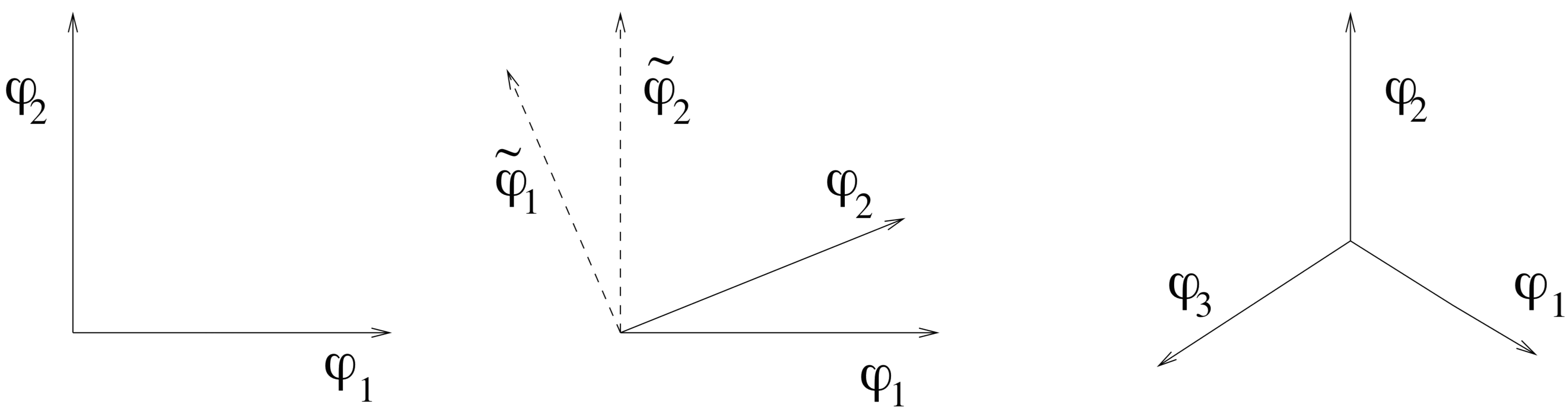
Definition(informal): The set of 'atoms' $\{\varphi_i\}_{i\in\mathcal{K}} \in V$ is called a *frame* of V when it is *overcomplete* meaning that for any signal *x* there exists a sequence α_i such that

$$x(t) = \sum_{i=-\infty}^{\infty} \alpha_i \varphi_i(t)$$

but the sequence is **not** unique.

Imperial College London

Bases and Frames: Geometric Interpretation



a) Orthogonal Basis b) Biorthogonal Basis c) Frame

Imperial College London

Bases and Frames: Matrix Interpretation

- Assume the 'atoms' $\{\varphi_i\}$ are finite dimensional column vectors of size *N*
- Stack them one next to the other to form the *synthesis* matrix *M*:

$$M = \begin{bmatrix} \uparrow & \cdots & \uparrow & \cdots \\ \varphi_1 & \cdots & \varphi_i & \cdots \\ \downarrow & \cdots & \downarrow & \cdots \end{bmatrix}$$

- In the case of frames, *M* is invertible but is *'fat'*

$$M^{-1} = M^H$$

Analysis and Synthesis Formulas: Frames Case

- In the frame case the synthesis matrix *M,* in the the synthesis formula $x = M\alpha$, is "fat".

- From now on we use the terminology 'dictionary' and denote the matrix with *D*.

- Orthogonal/biorthogonal bases are easier to handle but their construction is very constrained. In many practical contexts we prefer the flexibility provided by frames, but need to devise a strategy to find α_i since the solution is not unique.

Sparse Representation in a union of two bases

- Wavelets provide sparse representations of piecewise smooth images.
- In matrix/vector form ***y=Wα***
- Here the matrix ***W*** has size N × N and models the discrete-time wavelet transform of finite dimensional signals.

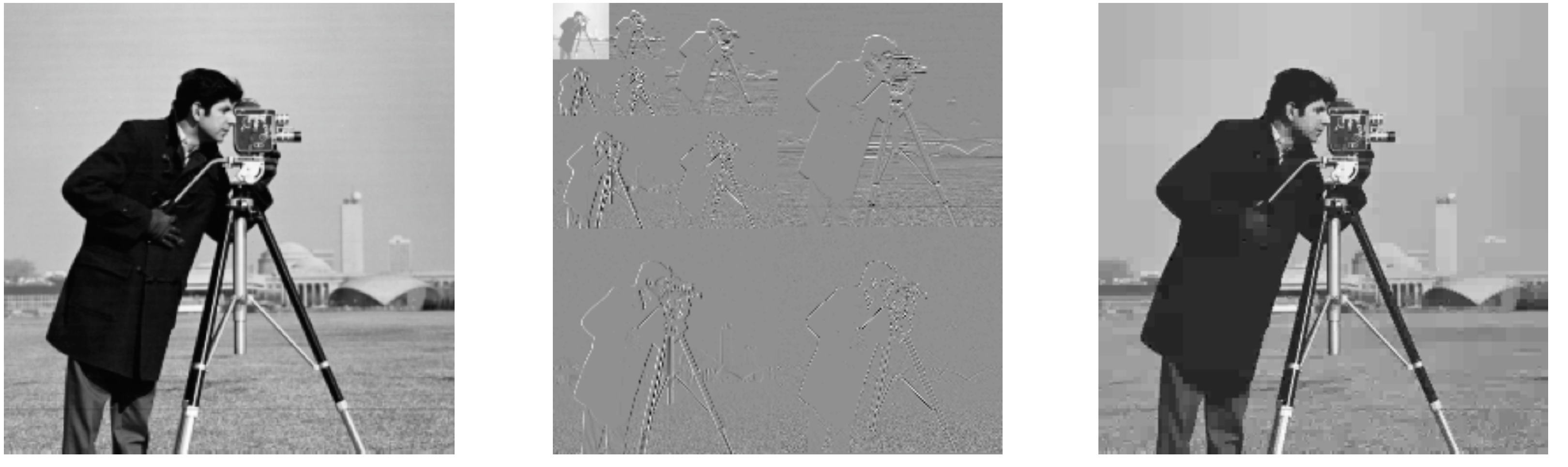
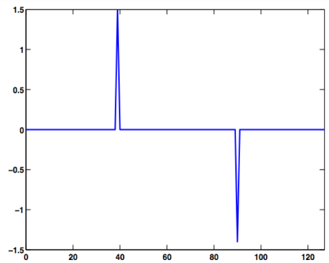


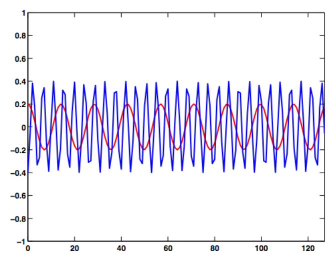
Figure: Cameraman is reconstructed using only 8% of the wavelet coefficients

- How about textures? The DCT is maybe better for textured regions
- **Key insight**: use an overcomplete dictionary (frame) ***D*** made of the union of two bases to obtain even sparser representations of images

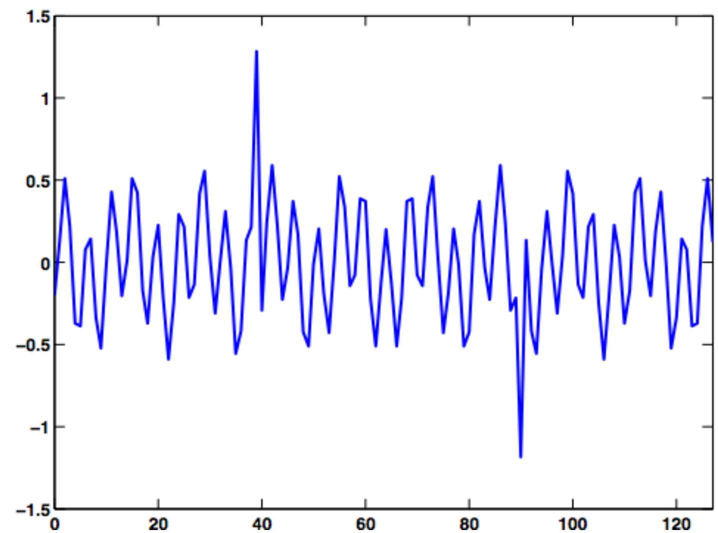
Sparse Representation in Fourier and Canonical Bases



two spikes



two complex exponentials



$\boldsymbol{y}$ *(real part plotted)*

The above signal, ***y***, is a combination of two spikes and two complex exponentials of different frequency (real part of ***y*** plotted). In matrix vector form:

$$\mathbf{y} = [\ \mathbf{I} \quad \mathbf{F}]\ \alpha = \mathbf{D}\alpha$$

where ***I*** is the N × N identity matrix and ***F*** is the N × N Fourier transform. The matrix ***D*** models the over-complete dictionary and has size N × 2N

- *Source separation: decompose signals into a smooth part and local innovations*

- *Prototype for the following problem:*

> Given two bases (or frames) $\boldsymbol{D} = [\boldsymbol{\Psi}, \boldsymbol{\Phi}]$. Represent an observed signal as a superposition of a few atoms from $\boldsymbol{\Psi}$ and a few atoms from $\boldsymbol{\Phi}$.

Example: *(Curvelets + DCT)*

images from [Elad, Starck, Querre, Donoho, 2005]

Sparse Representation in Fourier and Canonical Bases

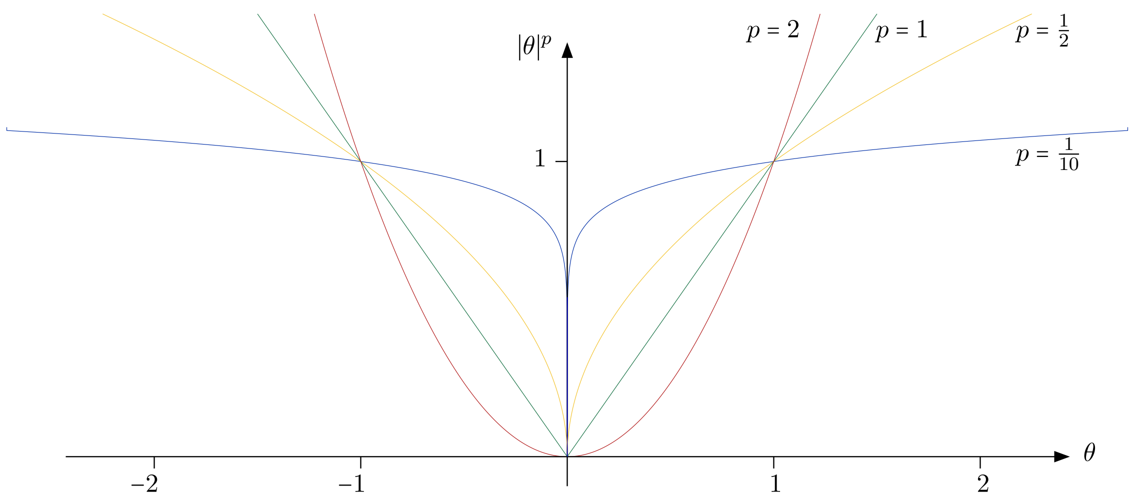
- Given ***y***, you want to find its sparse representation

- Ideally you want to solve

$$(P_0): \quad \min \|\alpha\|_0 \quad \text{s.t.} \quad \mathbf{y} = \mathbf{D}\alpha.$$

- Alternatively you may consider the following convex relaxation

$$(P_1): \quad \min \|\alpha\|_1 \quad \text{s.t.} \quad \mathbf{y} = \mathbf{D}\alpha.$$



Sparse Representation in Fourier and Canonical Bases

- Given ***y***, you want to find its sparse representation

- Ideally you want to solve

$$(P_0): \qquad \min \|\alpha\|_0 \qquad \text{s.t.} \qquad \mathbf{y} = \mathbf{D}\alpha.$$

- Alternatively you may consider the following convex relaxation

$$(P_1): \qquad \min \|\alpha\|_1 \qquad \text{s.t.} \qquad \mathbf{y} = \mathbf{D}\alpha.$$

- The problem (P_1) can be solved using convex optimization methods (i.e., linear programming) or greedy algorithms

Imperial College London

Sparse Representation in Fourier and Canonical Bases

- Given $\mathbf{y} = [\ \mathbf{I} \quad \mathbf{F}]\ \alpha = \mathbf{D}\alpha$ you want to find its K-sparse representation
- Ideally you want to solve

$$(P_0): \qquad \min \|\alpha\|_0 \qquad \text{s.t.} \qquad \mathbf{y} = \mathbf{D}\alpha.$$

- Alternatively you may consider the following convex relaxation

$$(P_1): \qquad \min \|\alpha\|_1 \qquad \text{s.t.} \qquad \mathbf{y} = \mathbf{D}\alpha.$$

- Key result due to Donoho-Huo [IEEE Trans. on Information Theory 2001]
  - (P_0) is unique when the sparsity K satisfies $K < \sqrt{N}$
  - (P_0) and (P_1) are equivalent when $K < \frac{1}{2}\sqrt{N}$

Sparsity in Pairs of Orthogonal Bases

- Key extensions due to Elad-Bruckstein [IEEE Trans. on Information Theory 2002]
- Given an arbitrary pair of orthonormal bases $\mathbf{\Psi_N}$ and $\mathbf{\Phi_N}$ and the mutual coherence

$$\mu(\mathbf{D}) = \max_{1\leq k,j\leq N, k\neq j} \frac{|\mathbf{d}_k^T \mathbf{d}_j|}{\|\mathbf{d}_k\|\|\mathbf{d}_j\|},$$

- (P0) is unique when $K < 1/\mu(\mathbf{D})$
- (P0) and (P1) are equivalent when

$$2\mu(\mathbf{D})^2 K_p K_q + \mu(\mathbf{D}) \max\{K_p, K_q\} - 1 < 0, \qquad (\mathbf{Tight\ Bound})$$

- Alternatively (P0) and (P1) are equivalent when

$$K < (\sqrt{2} - 0.5)/\mu(\mathbf{D}) \sim 0.9/\mu(\mathbf{D}) \qquad (\mathbf{Weak\ Bound})$$

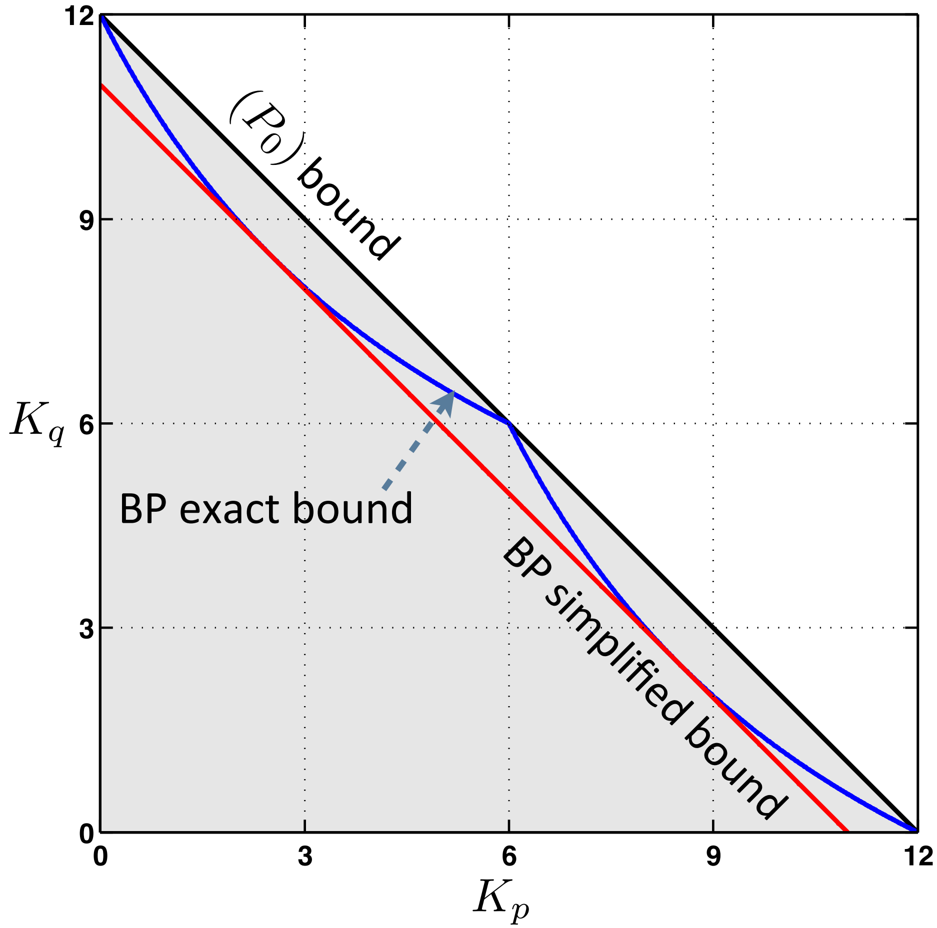
Sparsity in Pairs of Orthogonal Bases

- (P0) is unique when $$K < 1/\mu(\mathbf{D})$$

- (P0) and (P1) are equivalent when
$$2\mu(\mathbf{D})^2 K_p K_q + \mu(\mathbf{D}) \max\{K_p, K_q\} - 1 < 0, \qquad (\textbf{Tight Bound})$$
- Alternatively (P0) and (P1) are equivalent when
$$K < (\sqrt{2} - 0.5)/\mu(\mathbf{D}) \sim 0.9/\mu(\mathbf{D}) \qquad (\textbf{Weak Bound})$$

- Please note:
  - $K=K_p+K_q$
  - In Fourier and Canonical case $$\mu(\mathbf{D}) = \sqrt{N}$$

Sparsity Bounds in Pairs of Orthogonal Bases



Sparsity Bounds in Ovecomplete Dictionaries

Extensions [Tropp-04, GribonvalN:03, Elad-10]

- For a generic over-complete dictionary D, (P_1) is equivalent to (P_0) when[2]

$$K < \frac{1}{2}\left(1 + \frac{1}{\mu}\right).$$

- When D is a concatenation of J orthonormal dictionaries (P_1) is equivalent to (P_0) when

$$K < \left[\sqrt{2} - 1 + \frac{1}{2(J-1)}\right]\mu^{-1}$$

Sparse Representation via OMP

Algorithm 5 *OMP*—Orthogonal Matching Pursuit

Input: Dictionary $\boldsymbol{D} = [\boldsymbol{d}_1, \ldots, \boldsymbol{d}_L] \in \mathbb{C}^{N\times L}$, observation $\boldsymbol{y} \in \mathbb{C}^N$ and error threshold η [optional argument, maximum number of iterations $K_{\max}$].

Output: Sparse vector $\boldsymbol{x} \in \mathbb{C}^L$.

Initialise index $k = 0$.
Initialise solution $\boldsymbol{x}^{(0)} = \boldsymbol{0}$.
Initialise residual $\boldsymbol{r}^{(0)} = \boldsymbol{y} - \boldsymbol{D}\,\boldsymbol{x}^{(0)} = \boldsymbol{y}$.
Initialise support $\mathcal{S}^{(0)} = \emptyset$.
while $\left\|\boldsymbol{r}^{(k)}\right\|_2^2 > \eta$ [optional: $k < K_{\max}$] **do**
    $k \leftarrow k + 1$
    Compute $e[i] = \left\|z[i]\,\boldsymbol{d}_i - \boldsymbol{r}^{(k-1)}\right\|_2^2$ for $i \in \{1, \ldots, L\} \setminus \mathcal{S}^{(k-1)}$, where $z[i] = \frac{\boldsymbol{d}_i^H\,\boldsymbol{r}^{(k-1)}}{\|\boldsymbol{d}_i\|_2^2}$.
    Find index $i_0 = \arg\min_{i \in \{1,\ldots,L\}\setminus\mathcal{S}^{(k-1)}} \{e[i]\}$.
    Update support $\mathcal{S}^{(k)} = \mathcal{S}^{(k-1)} \cup \{i_0\}$.
    Compute solution $\boldsymbol{x}^{(k)} = \arg\min_{\tilde{\boldsymbol{x}} \in \mathbb{C}^L} \|\boldsymbol{D}\,\tilde{\boldsymbol{x}} - \boldsymbol{y}\|_2^2$ subject to $\operatorname{supp}\{\tilde{\boldsymbol{x}}\} = \mathcal{S}^{(k)}$.
    Update residual $\boldsymbol{r}^{(k)} = \boldsymbol{y} - \boldsymbol{D}\,\boldsymbol{x}^{(k)}$.
end while

Sparse Representation via OMP

Orthogonal matching pursuit (OMP) finds the correct sparse representation when

$$K < \frac{1}{2}\left(1 + \frac{1}{\mu}\right). \tag{4}$$

Sketch of the Proof (Elad 2010, pages 65-67):
Assume the K non-zero entries are at the beginning of the vector in descending order with $y = Dx$. Thus

$$y = \sum_{l=1}^{K} x_l D_l \tag{5}$$

First iteration of OMP work properly if $|D_1^T y| > |D_i^T y|$ for any $i > K$. Using (5)

$$|\sum_{l=1}^{K} x_l D_1^T D_l| > |\sum_{l=1}^{K} x_l D_i^T D_l|$$

Sparse Representation via OMP

Sketch of the Proof (cont'd):
But

$$|\sum_{l=1}^{K} x_l D_1^T D_l| \geq |x_1| - \sum_{l=2}^{K} |x_l||D_1^T D_l| \geq |x_1| - \sum_{l=2}^{K} |x_l|\mu \geq |x_1|(1-\mu)(K-1).$$

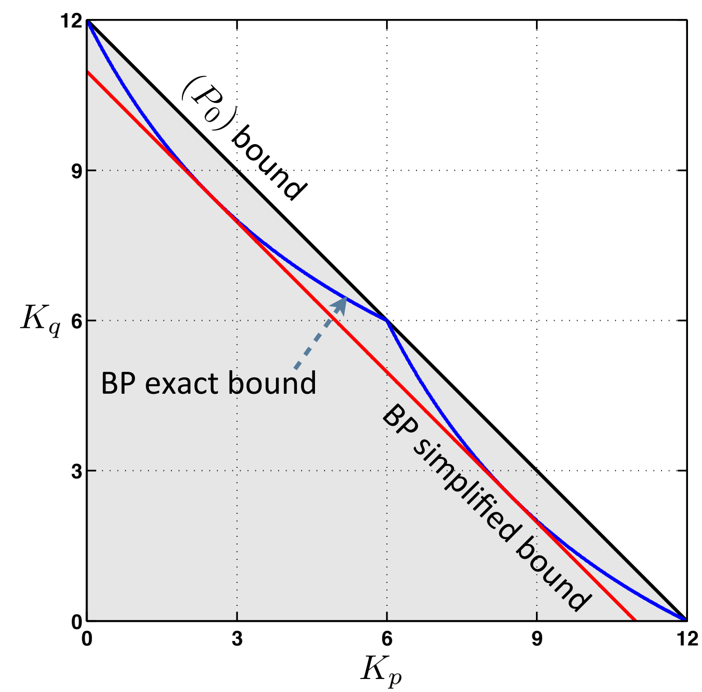
Moreover,

$$|\sum_{l=1}^{K} x_l D_i^T D_l| \leq \sum_{l=1}^{K} |x_l||D_i^T D_l| \leq \sum_{l=1}^{K} |x_l|\mu \leq |x_1|\mu K$$

Using these two bounds, we conclude that $|D_1^T y| > |D_i^T y|$ is satisfied when condition (4) is met. $\square$

The Tyranny of l_1: Is there life beyond BP?

- There is still a gap between l_0 and l_1 minimizations. Can we do better than Basis Pursuit?
- (P0) is NP-Hard for unrestricted dictionary. Can we say the same for structured dictionaries like Fourier and Identity?
- What happens when the unicity constraint is not met?



Sparsity according to Prony: Overview of Prony's Method

- Consider the case when the signal ***y*** is made only of *K* Fourier atoms, i.e., ***y=Fc*** for some K-sparse vector ***c***
- The sparse vector ***c*** can be reconstructed from only *2K consecutive* entries of ***y***

- *Sketch of the Proof:*

  - The *n*th entry of ***y*** is of the form

$$y_n = \frac{1}{\sqrt{N}} \sum_{k=0}^{K-1} c_{m_k} e^{j2\pi m_k n/N} = \sum_{k=0}^{K-1} \alpha_k u_k^n,$$

where m_k is the index of the *k*th nonzero element of ***c***

Overview of Prony's Method

- Consider the polynomial:

$$P(x) = \prod_{k=1}^{K}(x - u_k) = x^K + h_1 x^{K-1} + h_2 x^{K-2} + \ldots + h_{K-1}x + h_K.$$

- It is easy to verify that $h_n * y_n = 0$

- In matrix-vector form, we get

$$\begin{bmatrix} y_{l+K} & y_{l+K-1} & \cdots & y_l \\ y_{l+K+1} & y_{l+K} & \cdots & y_{l+1} \\ \vdots & \ddots & \ddots & \vdots \\ y_{l+2K-2} & \ddots & \ddots & \vdots \\ y_{l+2K-1} & y_{l+2K-2} & \cdots & y_{l+K-1} \end{bmatrix} \begin{bmatrix} 1 \\ h_1 \\ h_2 \\ \vdots \\ h_K \end{bmatrix} = \mathbf{T}_{K,l}\mathbf{h} = \mathbf{0}$$

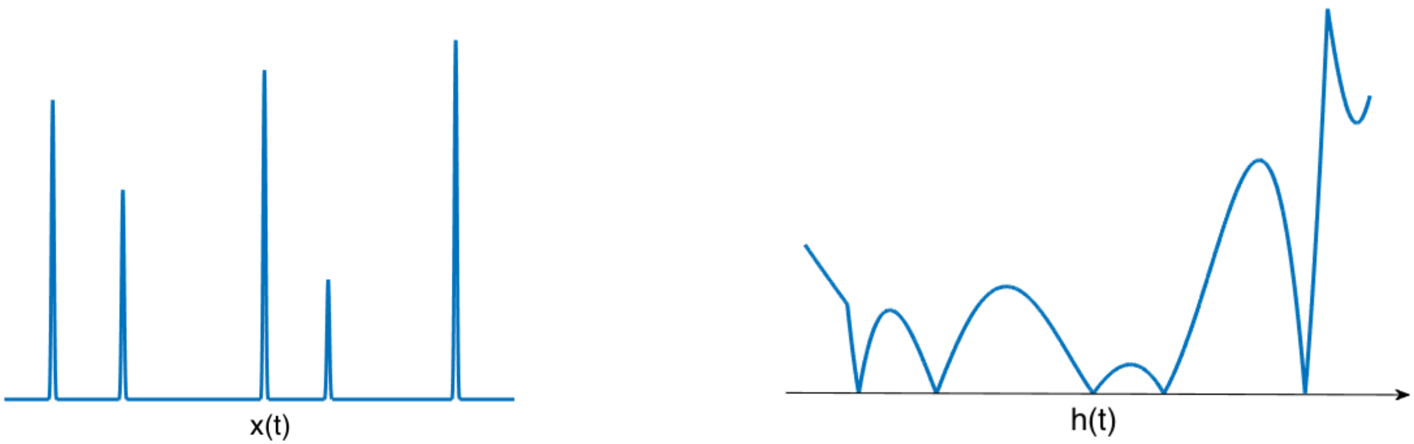
Overview of Prony's Method

- The vector of polynomial coefficients $\boldsymbol{h} = [1, h_1, \ldots, h_k]^T$ is in the null space of $\boldsymbol{T_{k,l}}$. Moreover, $\boldsymbol{T_{k,l}}$ has size *K x (K+1)* and has full rank, therefore, ***h*** is unique.

- Prony's method summary:
  - Given the input y_n, build the Toeplitz matrix $\boldsymbol{T_{k,l}}$ and solve for ***h***.
  - Find the roots $P(x) = 1 + \sum_{n=1}^{K} h_k x^{K-k}$. These roots are exactly the exponentials $\{u_k\}_{k=0}^{K-1}$. They give you the locations of the non-zero entries of your vector.
  - Given the locations of the non-zero entries find their amplitudes (this is a linear problem)

Imperial College London

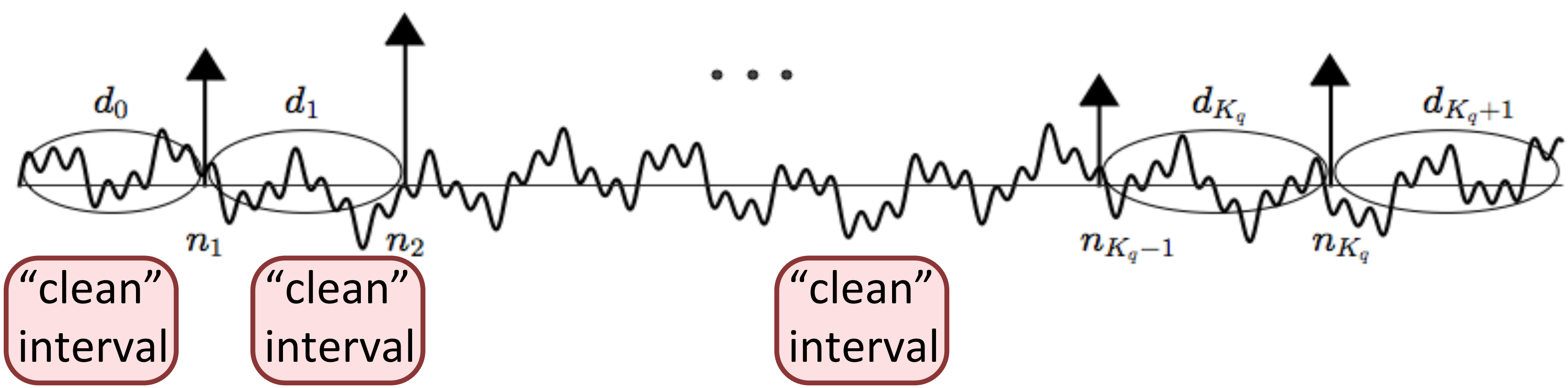
Prony's vs OMP



- OMP looks sequentially for the largest coefficients of ***x***
- Prony's find the non-zero entry of ***x*** by setting them to zero. That is, it builds a mask ***h*** with K zeros such that $x(t)h(t)=0$

ProSparse – Prony's based Sparsity

- Given $\boldsymbol{y} = [\boldsymbol{F}, \boldsymbol{I}] = \boldsymbol{D}\boldsymbol{x}$ where $\boldsymbol{x}$ is (K_p,K_q)-sparse



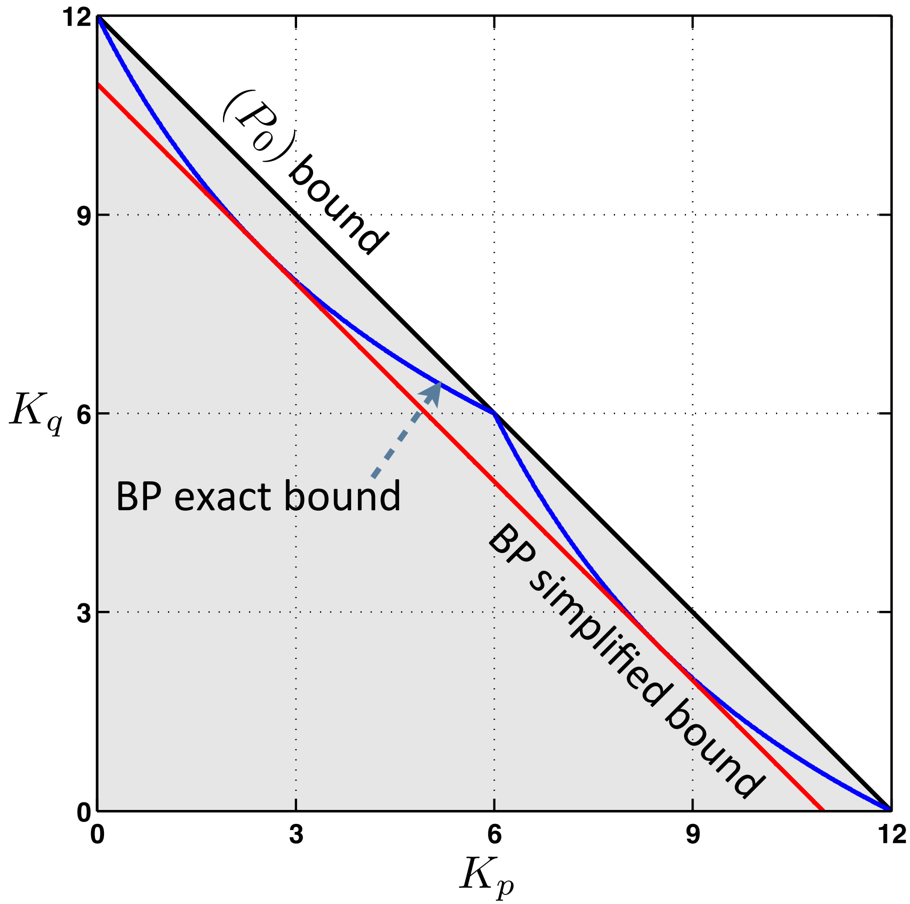
- K_p Fourier atoms —> need a "clean" interval of length $2K_p$
- For sufficiently sparse signals, such intervals always exist
- Sequential search and test: polynomial complexity

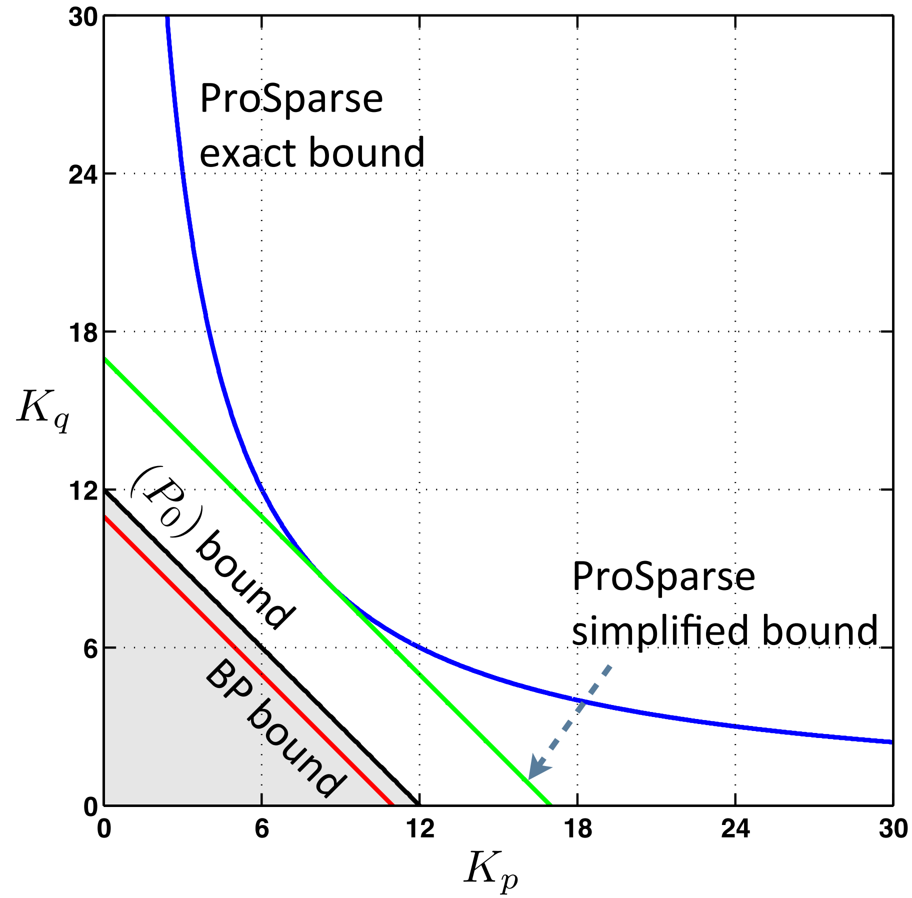
ProSparse Properties

***Theorem* [Dragotti & Lu, IEEE IT. 2014]** *Let* $\boldsymbol{D} = [\boldsymbol{F}, \boldsymbol{I}]$ *and* $\boldsymbol{y} \in \mathbb{C}^N$ *an arbitrary signal. There exists an algorithm, with a worst-case complexity of* $\mathcal{O}(N^3)$ *, that finds **all** (K_p,K_q)-sparse signals* $\boldsymbol{x}$ *such that*

$$\boldsymbol{y} = \boldsymbol{D}\boldsymbol{x} \quad \text{and} \quad K_p K_q < N/2$$

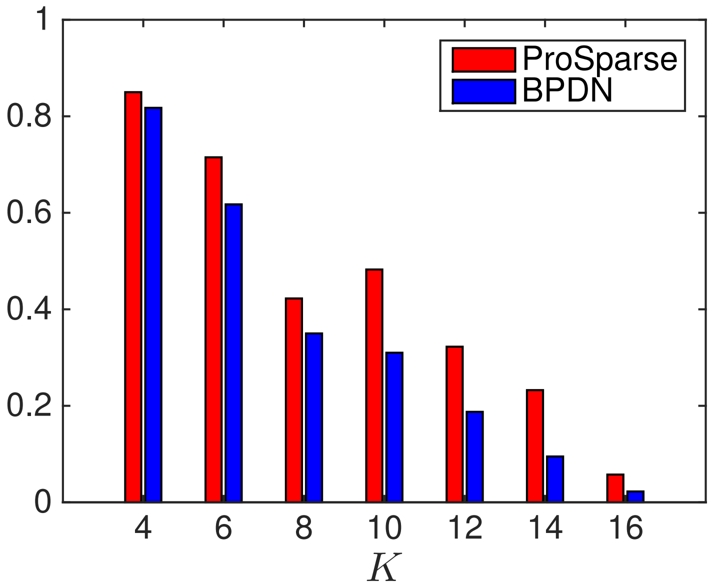
ProSparse Bounds vs BP



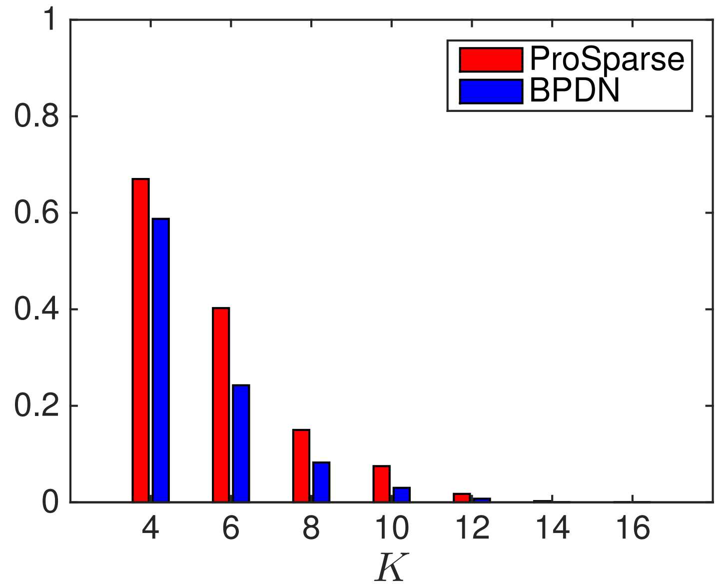


ProSparse vs BP (noisy case)

- Support Recovery



(a) SNR = 10 dB



(b) SNR = 5 dB

Recovery of Approximately Sparse Signals

- Assume ***y*** is not exactly sparse. You may also observe a corrupted (e.g., noisy version) of ***y***
- Rather than solving (P_1)

$$(P_1): \quad \min \|\alpha\|_1 \quad \text{s.t.} \quad \mathbf{y} = \mathbf{D}\alpha.$$

- Consider the following relaxed version:

$$\min_{\alpha} \|\mathbf{y} - \mathbf{D}\alpha\|_2^2 + \lambda\|\alpha\|_1$$

- This basic formulation applies to denoising, deconvolution, inpainting (see for example Elad et al. SPIE 2007)

Recovery of Approximately Sparse Signals

- Consider the following relaxed version:

$$\min_{\alpha} \|\mathbf{y} - \mathbf{D}\alpha\|_2^2 + \lambda\|\alpha\|_1$$

- This basic formulation applies to denoising, deconvolution, inpainting (see for example Elad et al. SPIE 2007)

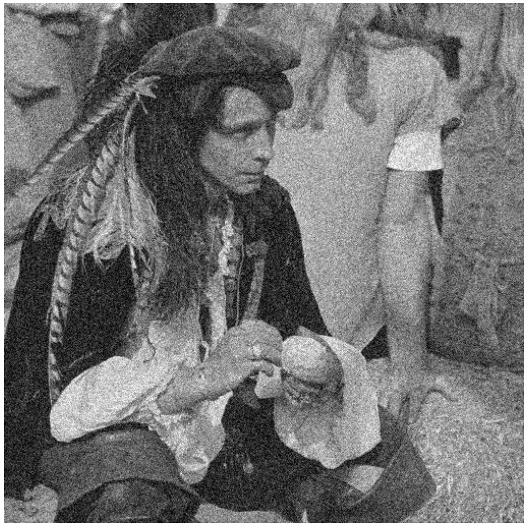
- Can be solved iteratively as follows

$$\alpha_{i+1} = \mathcal{S}_{\lambda/c}\left(\frac{1}{c}\mathbf{D}^T(\mathbf{y} - \mathbf{D}\alpha_i) + \alpha_i\right)$$

- Here $\mathcal{S}$ is a shrinkage operator (e.g., soft-threshold)

Imperial College London

Application: Image Denoising

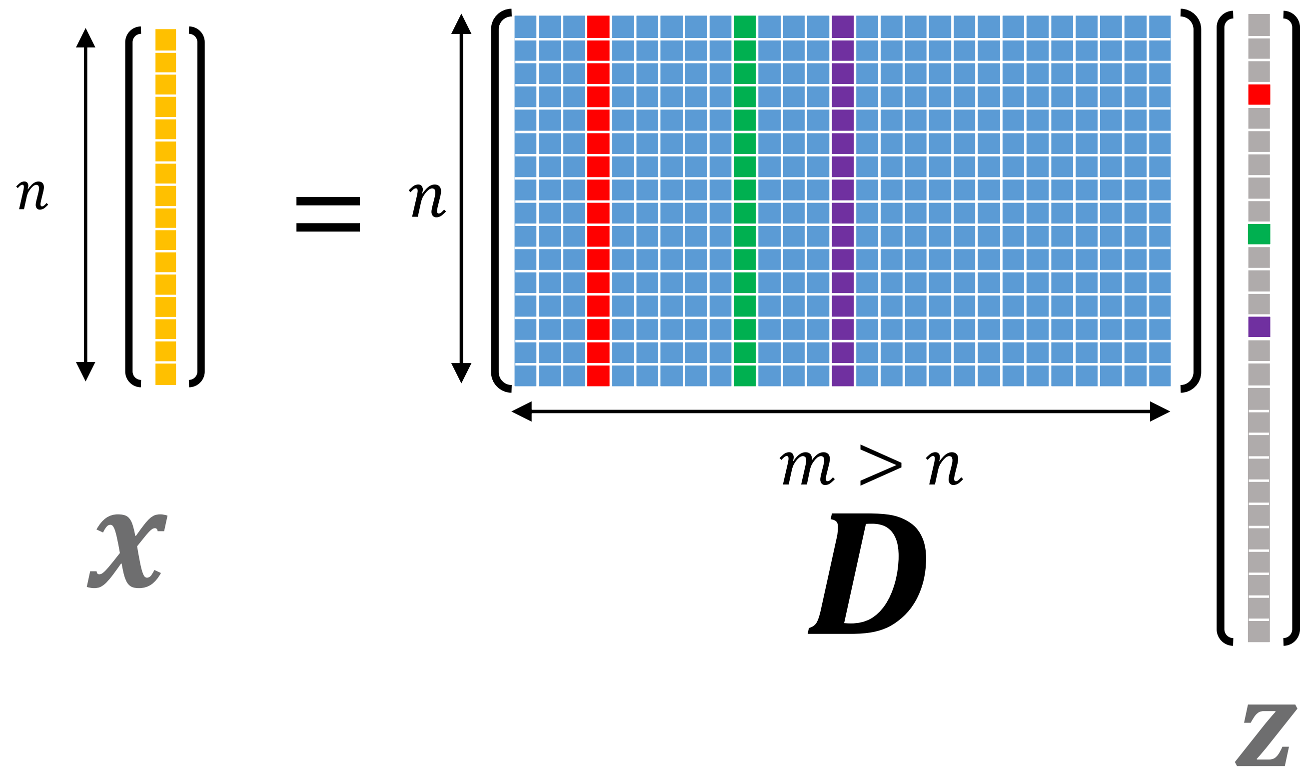


Original *Degraded (PSNR=20dB)* *State-of-the-Art* [BM3D-SAPCA]
(PSNR = 29.81 dB)

Note: BM3D uses *non-local* sparsity models

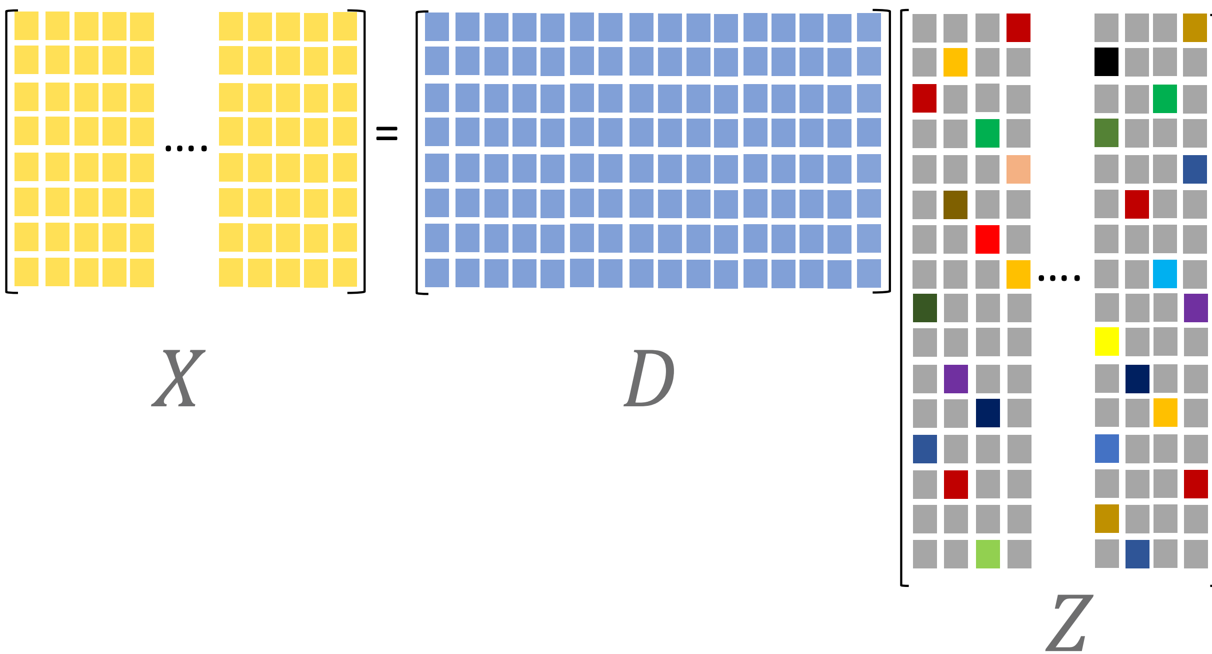
Dictionary Learning

- The **key insight** in the data-driven approach is that images or patches of images have a sparse representation in a redundant dictionary
- The dictionary is usually learned



Dictionary Learning

- Learn dictionaries by alternating between
  - Learning the sparse representations given the dictionaries (sparse coding step)
  - Update the dictionary given the sparse representations (dictionary update step)



Imperial College London

Sparse Coding Step

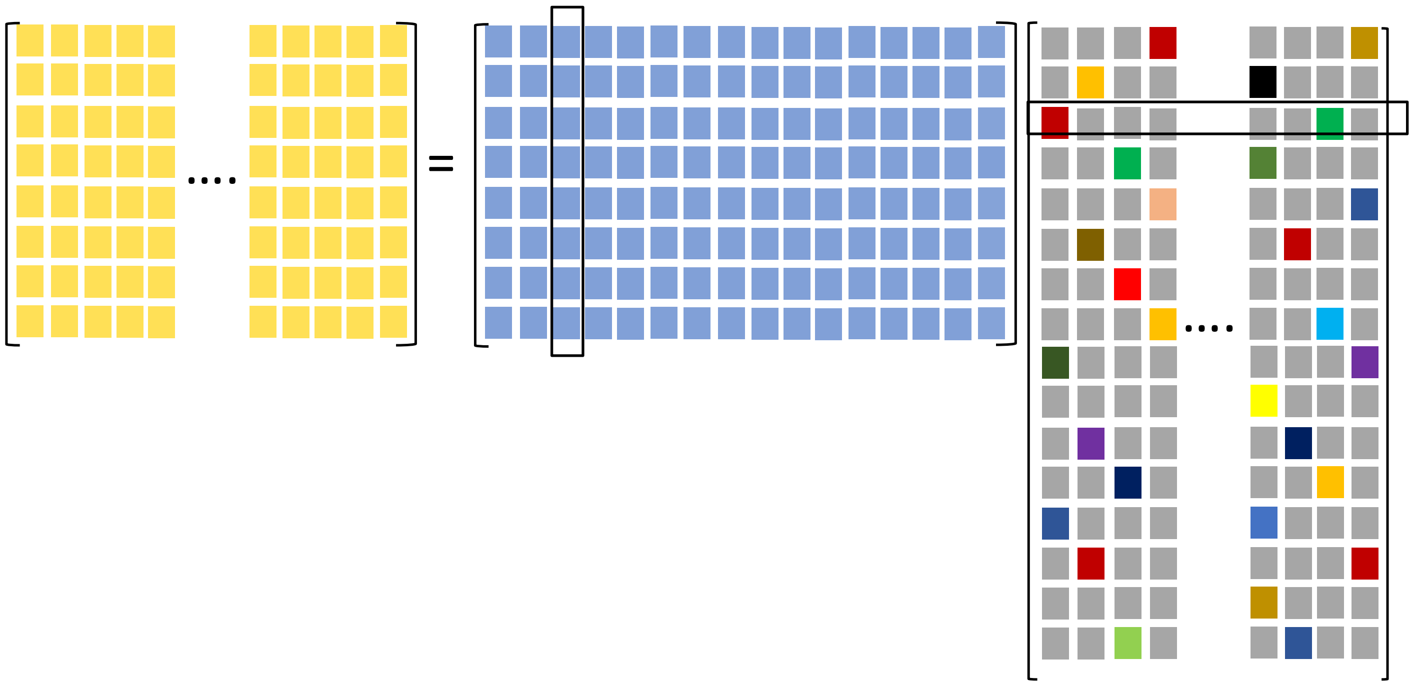
- Given $\boldsymbol{D}$, learn the sparse representations $\boldsymbol{z_i}$

- *Sparse Representation Algorithms:*
  - *Greedy algorithms:*
    - *Matching Pursuit (MP)*
    - *Orthogonal Matching Pursuit (O*
    - *...*
  - *Convex Relaxation Algorithms:*
    - *Basis Pursuit (BP)*
    - *....*

Imperial College London

Dictionary Update Step

- Given the sparse representations update the dictionary.
  - Many possible approaches, k-SVD (Aharon-Elad:06) is the most used

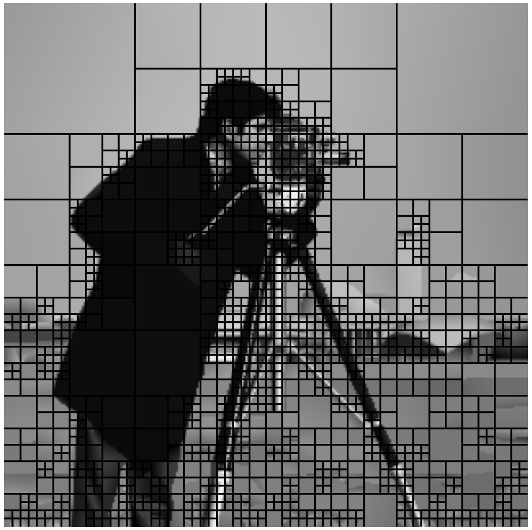


- *Find* $\mathbf{d}_i$ *and* $\mathbf{z}_i^T$ *that minimize* $\|\mathbf{E}_i - \mathbf{d}_i \mathbf{z}_i^T\|$
- *This is achieved by taking the SDV of* $\mathbf{E}_i$
- *(with a small caveat to keep* $\mathbf{z}_i^T$ *sparse)*

Beyond Traditional Sparsity Models

- Traditional sparsity models are essentially linear and apply essentially only to 1-D signals

- **Possible 2-D extension:** decompose the image with tiles of different size each being made of two smooth regions separated by a straight edge (semi-parametric model)



Beyond Traditional Sparsity Models (cont'd)

- The better the sparsity, the better the results ;-) Scholefield-Dragotti. IEEE Trans. Image Processing 2014

Degraded
(PSNR=10.6dB)

State-of-the-Art
(PSNR = 26.8 dB)

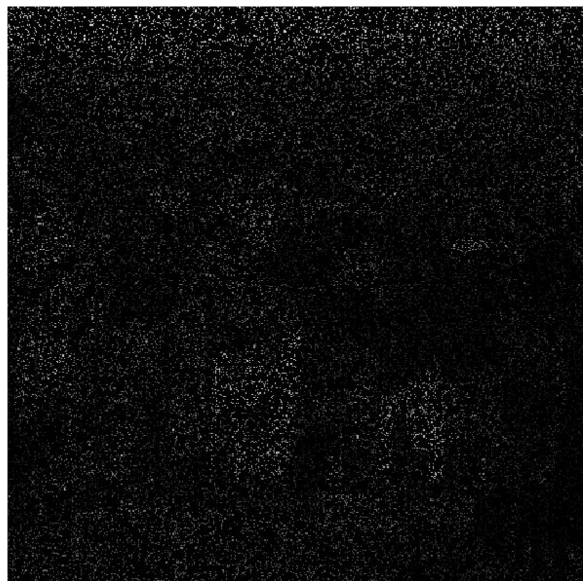
New Sparsity Model
(PSNR = 27.1 dB)

Imperial College London

Beyond Traditional Sparsity Models (cont'd)

- Inpainting: Scholefield-Dragotti. IEEE Trans. Image Processing 2014

Original



90% missing pixels

Inpainted using new Sparsity Model

Imperial College London

Summary

- Traditional sparsity is based around two pillars:
  - An expansion-based sparsity model
  - Reconstruction based on convex programming (e.g., BP)

- Room for more ***creative solutions*** both in terms of sparsity and reconstruction methods
  - E.g. Reconstruction using *ProSparse* outperforms Convex Programming in specific settings
  - E.g. Semi-Parametric sparsity models or non-local models for images outperforms state-of-the-art image processing algorithms

Imperial College London

References

Key papers and books on Sparse Signal Representation

- M. Elad,'Sparse and Redundant Representations', Springer, 2010
- D. L. Donoho and P.B. Starck, 'Uncertainty principles and signal recovery', SIAM J. Appl. Math., 1989, pp. 906-931.
- D.L. Donoho and X. Huo, 'Uncertainty principles and ideal atomic decomposition', IEEE Trans. on Info.. Theory, vol.47(7, pp.2845-62, November 2001.
- M. Elad et. al, 'A wide-angle view at iterated shrinkage algorithms', SPIE 2007

ProSparse:

- P. L. Dragotti and Y. M. Lu, "On Sparse Representation in Fourier and Local Bases," IEEE Transactions on Information Theory, vol. 60, no. 12, pp. 7888-7899, 2014.

BM3D:

- K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, "Image denoising by sparse 3D transform-domain collaborative filtering," IEEE Trans. Image Process., vol. 16, no. 8, pp. 2080-2095, August 2007.