Sparse Representations and the Basis Pursuit Algorithm*

Michael Elad

The Computer Science Department –

Scientific Computing & Computational mathematics (SCCM) program

Stanford University

November 2002

* Joint work with: Alfred M. Bruckstein – CS, Technion

David L. Donoho – Statistics, Stanford

Peyman Milanfar – EE, UCSC

Collaborators

Freddy Bruckstein

Computer Science
Department – Technion

Dave Donoho

Statistics Department
Stanford

Peyman Milanfar

EE - University of
California Santa-Cruz

General

- Basis Pursuit algorithm [Chen, Donoho and Saunders, 1995]:
  - Effective for finding sparse over-complete representations,
  - Effective for non-linear filtering of signals.
- Our work (in progress) – better understanding BP and deploying it in signal/image processing and computer vision applications.
- We believe that over-completeness has an important role!
- Today we discuss:
  - Understanding the BP: why successful? conditions?
  - Deploying the BP: through its relation to Bayesian (PDE) filtering.

Agenda

1. Introduction
   Previous and current work
2. Two Ortho-Bases
   Uncertainty → Uniqueness → Equivalence
3. Arbitrary dictionary
   Uniqueness → Equivalence
4. Basis Pursuit for Inverse Problems
   Basis Pursuit Denoising → Bayesian (PDE) methods
5. Discussion

Understanding the BP (items 2–3)

Using the BP for denoising (item 4)

Transforms

- Define the forward and backward transforms by (assume one-to-one mapping)

$$\text{Forward}: \quad \underline{\alpha} = T\{\underline{s}\}$$

$$\text{Backward}: \quad \underline{s} = T^{-1}\{\underline{\alpha}\}$$

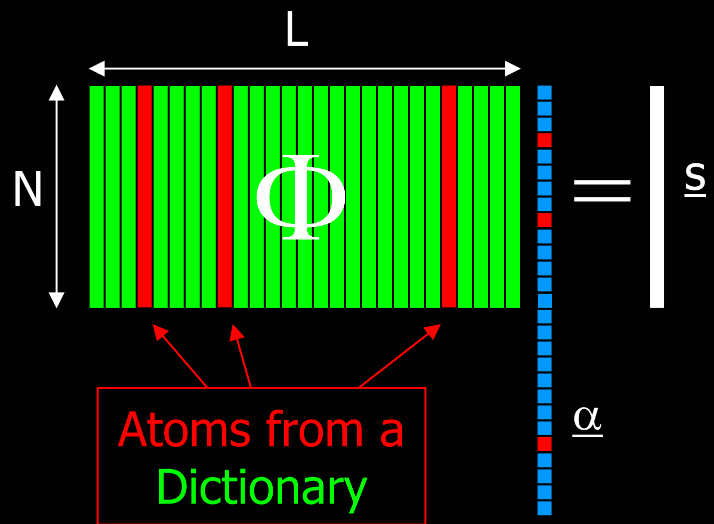
$\underline{s}$ – Signal (in the signal space C^N)

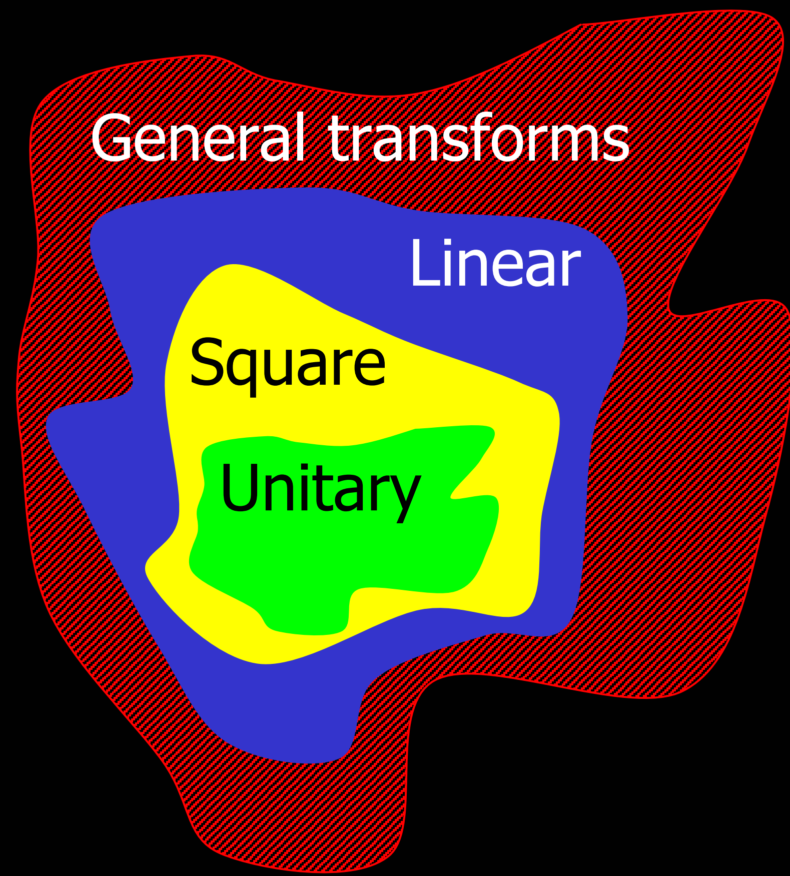
$\underline{\alpha}$ – Representation (in the transform domain C^L, $L \geq N$)

- Transforms T in signal and image processing used for coding, analysis, speed-up processing, feature extraction, filtering, …

The Linear Transforms

- Special interest - linear transforms (inverse) $\underline{s} = \Phi\underline{\alpha}$



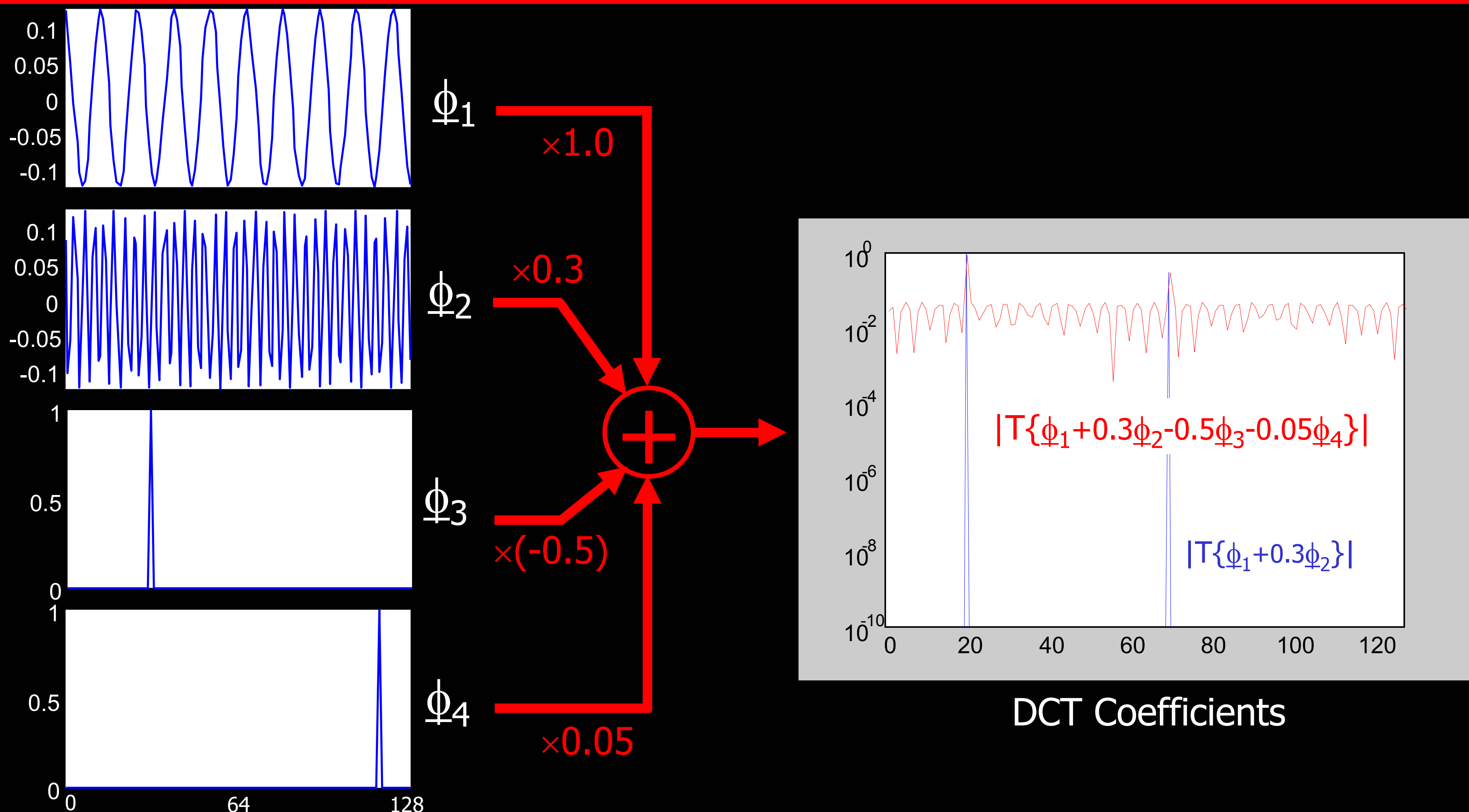


- In square linear transforms, Φ is an N-by-N & non-singular.

Lack Of Universality

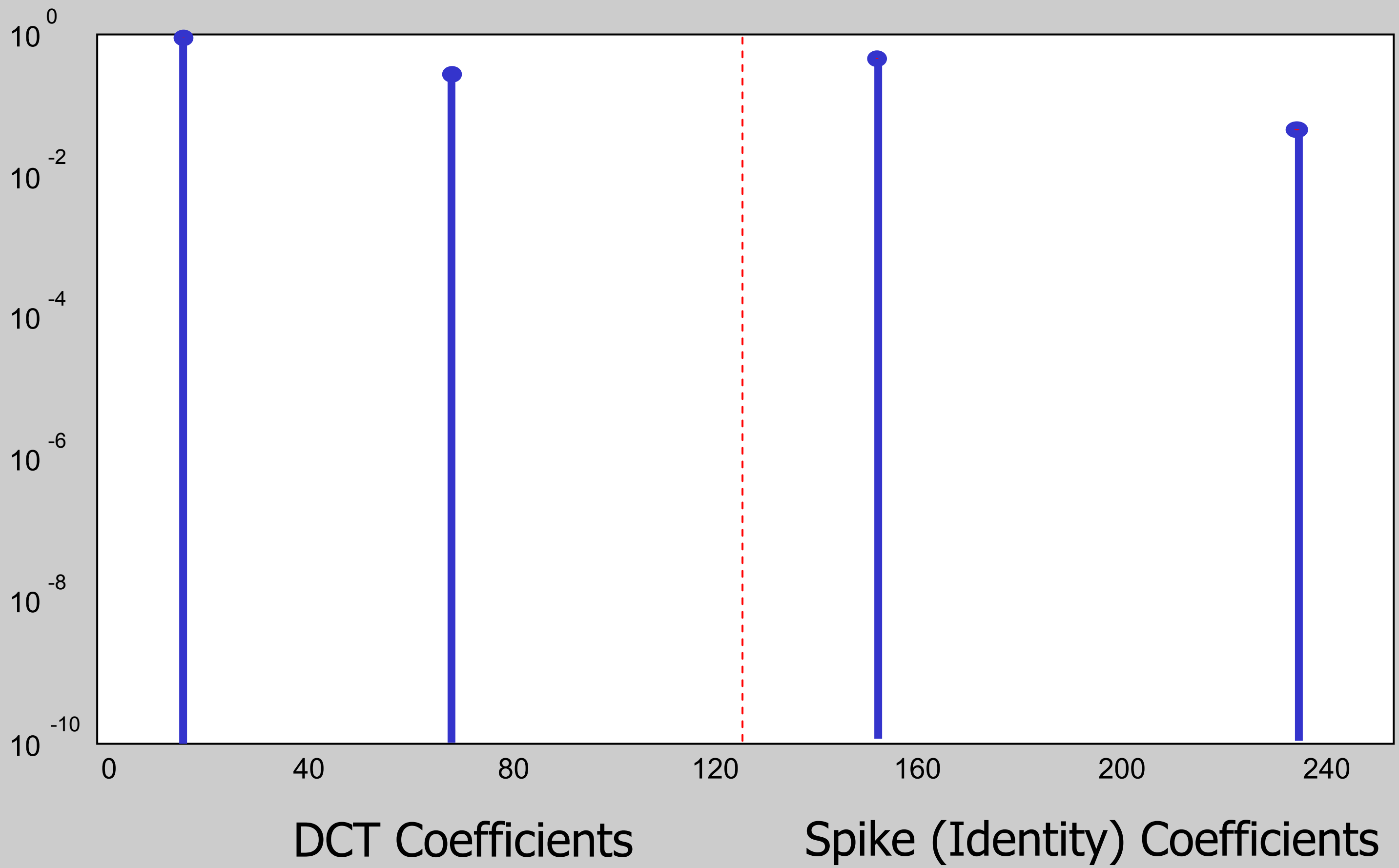
- Many available square linear transforms – sinusoids, wavelets, packets, ridgelets, curvelets, …
- Successful transform – one which leads to sparse representations.
- Observation: Lack of universality - Different bases good for different purposes.
  - Sound = harmonic music (Fourier) + click noise (Wavelet),
  - Image = lines (Ridgelets) + points (Wavelets).
- Proposed solution: Over-complete dictionaries, and possibly combination of bases.

Example – Composed Signal



DCT Coefficients

Example – Desired Decomposition



Matching Pursuit

- Given d unitary matrices $\{\Phi_k,\ 1 \le k \le d\}$, define a dictionary $\Phi = [\Phi_1, \Phi_2, \ldots \Phi_d]$ [Mallat & Zhang (1993)].
- Combined representation per a signal $\underline{s}$ by

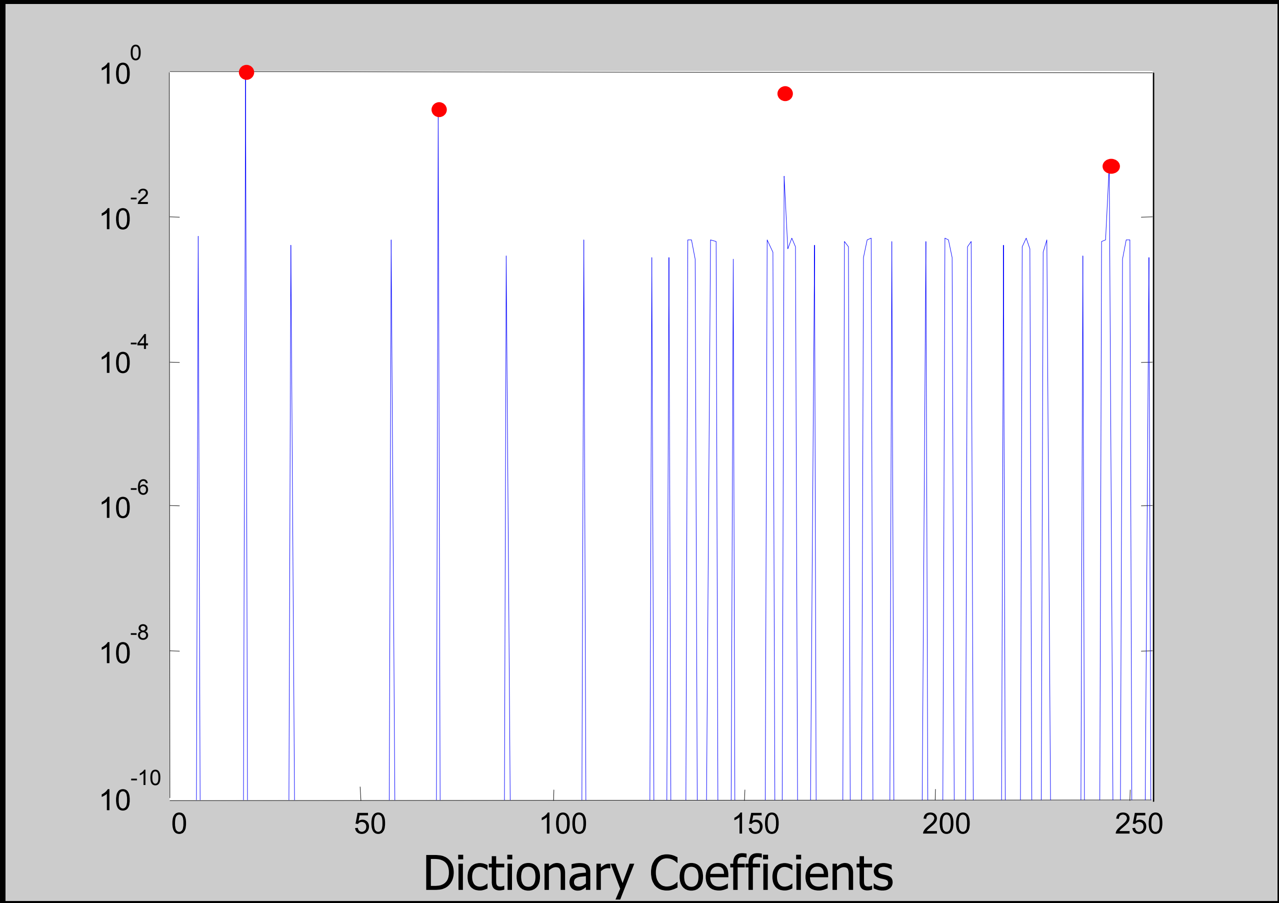
$$\underline{s} = \Phi \underline{\alpha}$$

- Non-unique solution $\underline{\alpha}$ - Solve for maximal sparsity

$$P_0: \quad \underset{\underline{\alpha}}{\text{Min}} \ \|\underline{\alpha}\|_0 \quad \text{s.t.} \quad \underline{s} = \Phi \underline{\alpha}$$

- Hard to solve – a sub-optimal greedy sequential solver: "Matching Pursuit algorithm".

Example – Matching Pursuit



Basis Pursuit (BP)

- Facing the same problem, and the same optimization task [Chen, Donoho, Saunders (1995)]

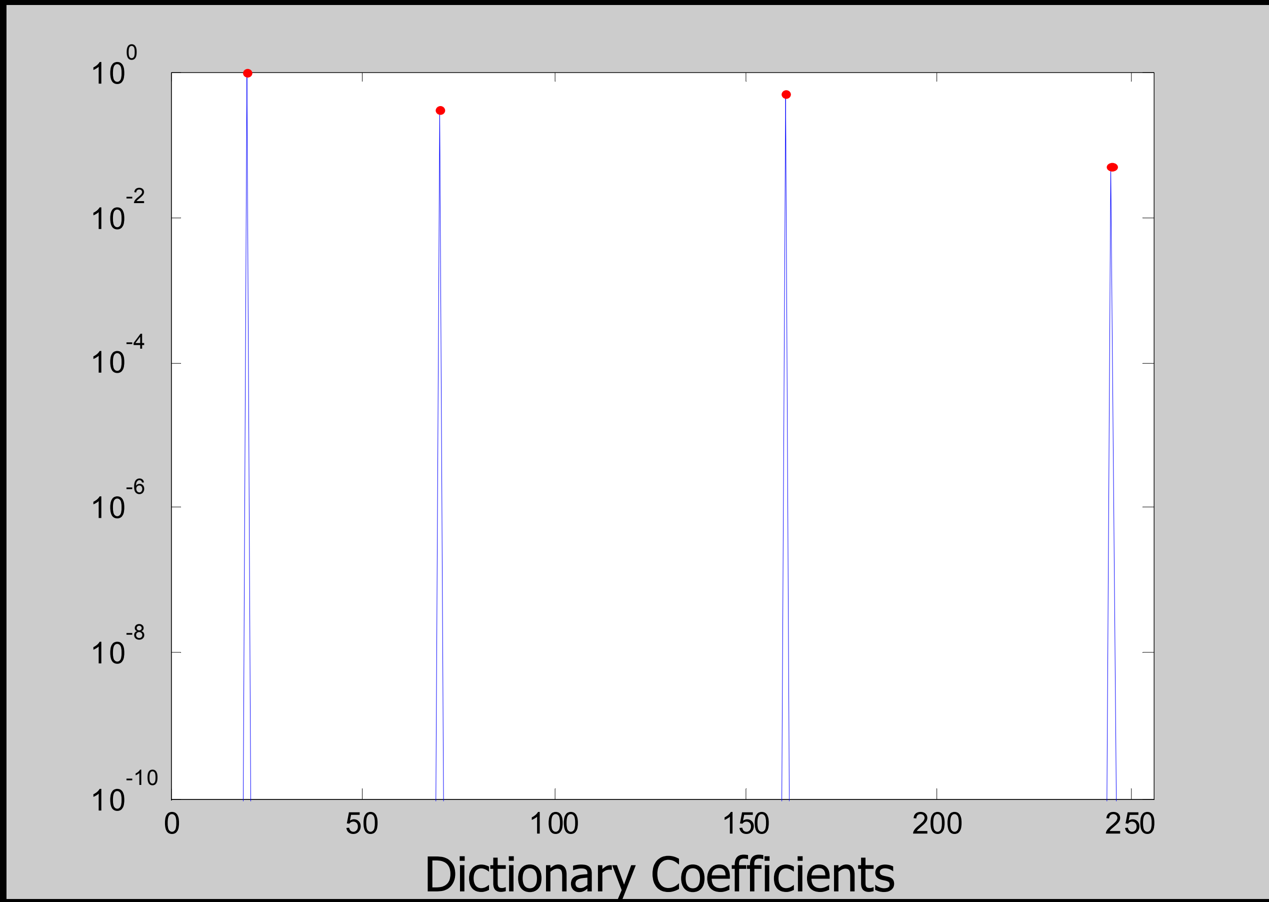
$$P_0: \quad \underset{\underline{\alpha}}{\text{Min}} \ \left\|\underline{\alpha}\right\|_0 \ \text{ s.t. } \ \underline{s} = \Phi\underline{\alpha}$$

- Hard to solve – replace the ℓ_0 norm by an ℓ_1: "Basis Pursuit algorithm"

$$P_1: \quad \underset{\underline{\alpha}}{\text{Min}} \ \left\|\underline{\alpha}\right\|_1 \ \text{ s.t. } \ \underline{s} = \Phi\underline{\alpha}$$

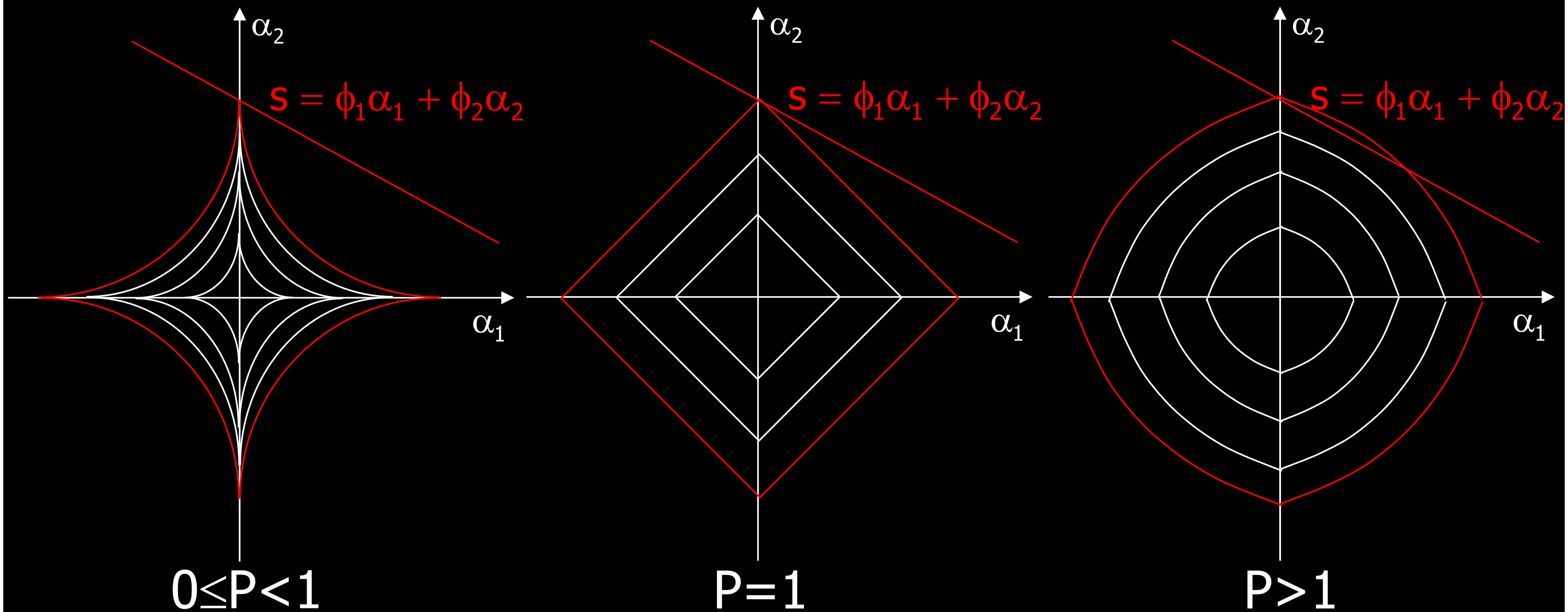
- **Interesting observation**: In many cases it successfully finds the sparsest representation.

Example – Basis Pursuit



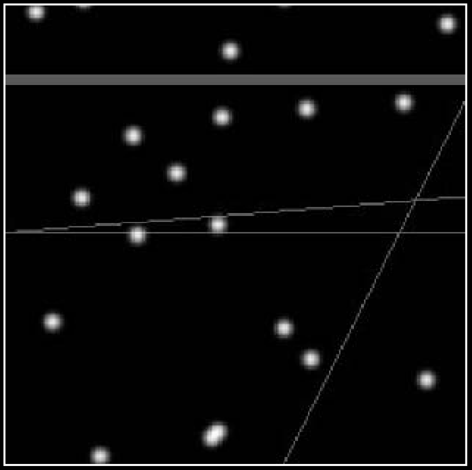
Why ℓ_1 ? 2D-Example

$$\underset{[\alpha_1,\alpha_2]}{\text{Min}} \ |\alpha_1|^p + |\alpha_2|^p \ \text{ s.t. } \ s = \phi_1\alpha_1 + \phi_2\alpha_2$$

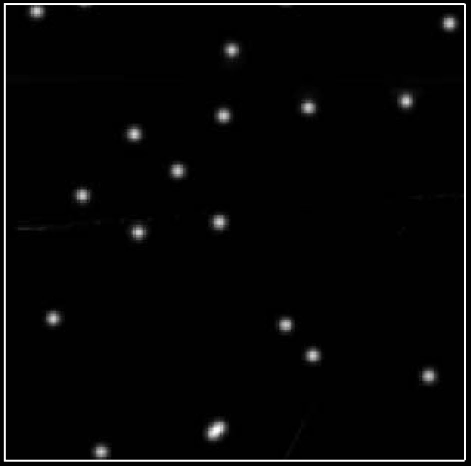


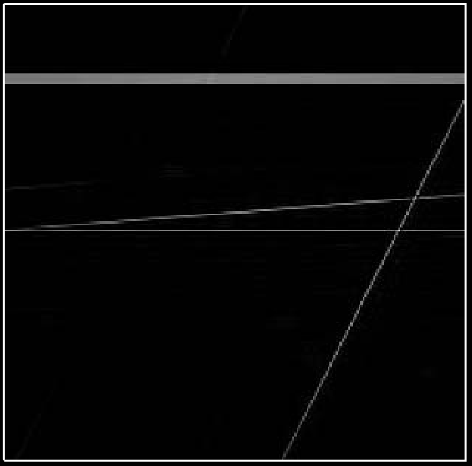
Example – Lines and Points*

Original image



Wavelet part of the noisy image

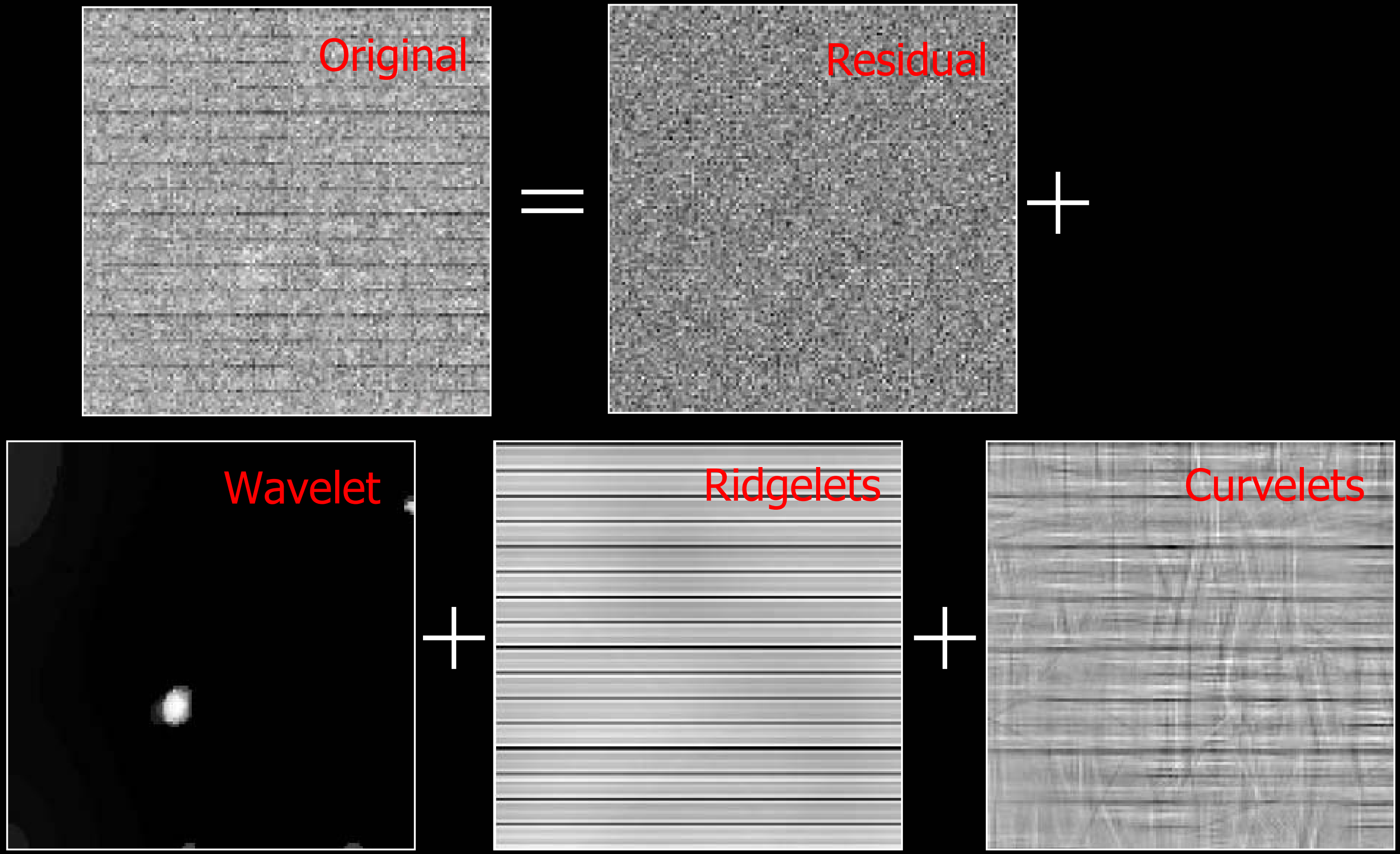




Ridgelets part of the image

* Experiments from Starck, Donoho, and Candes - Astronomy & Astrophysics 2002.

Example – Galaxy SBS 0335-052*



* Experiments from Starck, Donoho, and Candes - Astronomy & Astrophysics 2002.

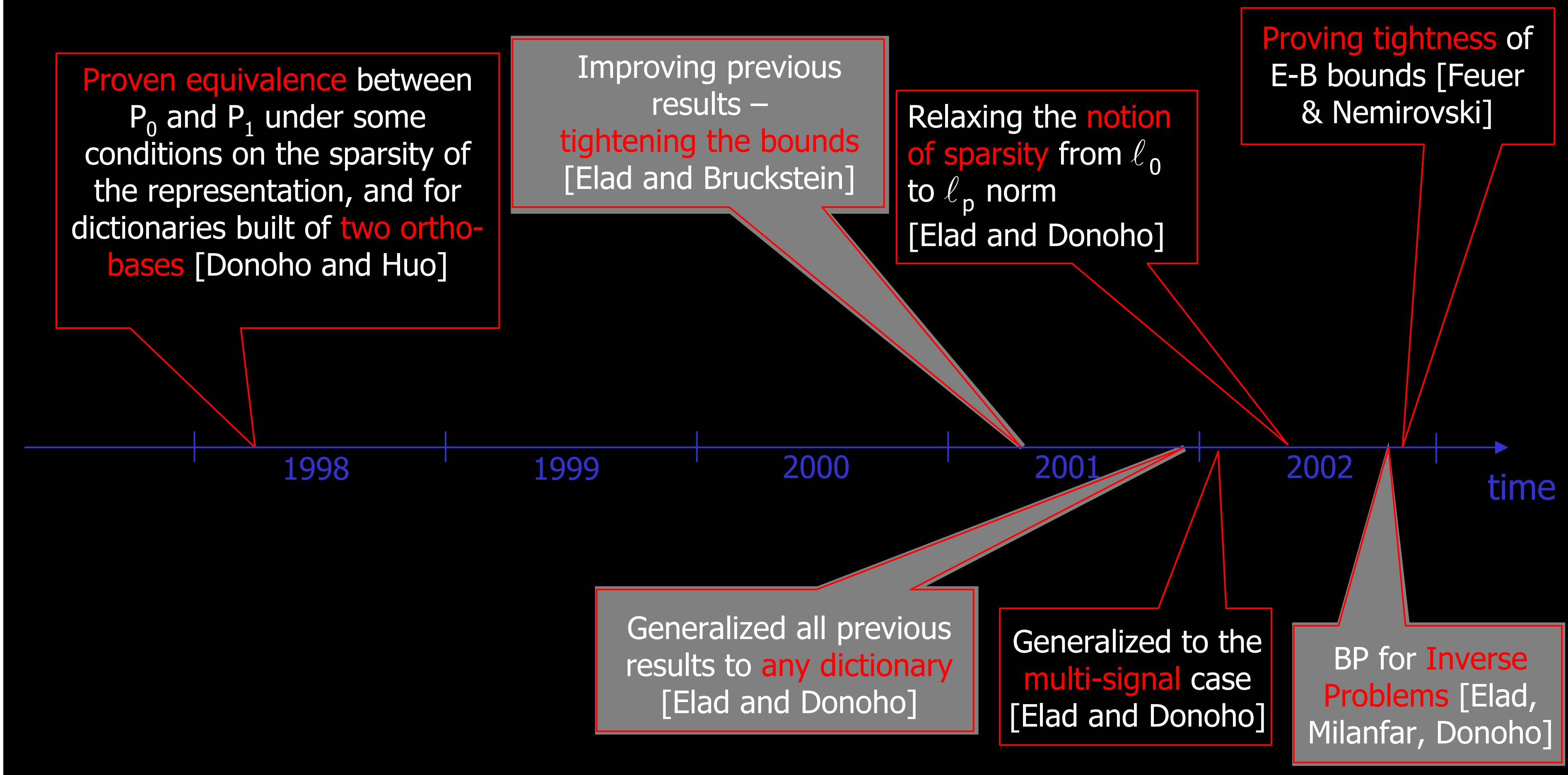
Non-Linear Filtering via BP

- Through the previous example – Basis Pursuit can be used for non-linear filtering.
- From Transforming to Filtering

$$\underset{\underline{\alpha}}{\text{Min}} \|\underline{\alpha}\|_1 \quad \text{s.t.} \quad \underline{s} = \Phi\underline{\alpha} \quad \Rightarrow \quad \underset{\underline{\alpha}}{\text{Min}} \|\underline{\alpha}\|_1 + \lambda \|\underline{s} - \Phi\underline{\alpha}\|_2^2$$

- What is the relation to alternative non-linear filtering methods, such as PDE based methods (TV, anisotropic diffusion …), Wavelet denoising?
- What is the role of over-completeness in inverse problems?

(Our) Recent Work



Before we dive …

- Given a dictionary ☞ and a signal $\underline{s}$, we want to find the sparse "atom decomposition" of the signal.
- Our goal is the solution of $\underset{\underline{\alpha}}{\text{Min}} \|\underline{\alpha}\|_0 \quad \text{s.t.} \quad \underline{s} = \Phi\underline{\alpha}$
- Basis Pursuit alternative is to solve instead

$$\underset{\underline{\alpha}}{\text{Min}} \|\underline{\alpha}\|_1 \quad \text{s.t.} \quad \underline{s} = \Phi\underline{\alpha}$$

- Our focus for now: Why should this work?

Agenda

1. Introduction

   Previous and current work

2. Two Ortho-Bases

   Uncertainty → Uniqueness → Equivalence

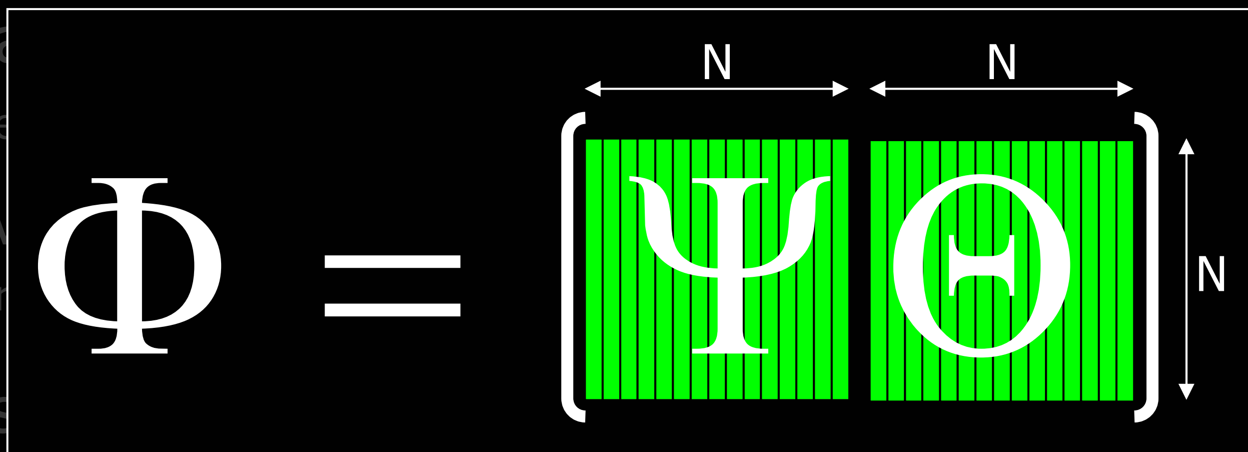
3. Arbitr

   Uniquene

4. BP Inv

   Basis Pur

5. Discus



Our Objective

Given a signal $\underline{s}$, and its two representations using Ψ and Θ, what is the lower bound on the sparsity of both?

Our Objective is

$$\underline{s} = \Psi\underline{\alpha} \qquad \underline{s} = \Theta\underline{\beta} \quad \Rightarrow \quad \|\underline{\alpha}\|_0 + \|\underline{\beta}\|_0 \geq \mathrm{Thr}(\Psi, \Theta)$$

We will show that such rule immediately leads to a practical result regarding the solution of the P_0 problem.

Mutual Incoherence

$$\text{Define} \quad M = \max_{1 \le k,j \le N} \left(\left| \underline{\psi}_k^H \underline{\theta}_j \right| \right)$$

- M – mutual incoherence between Ψ and Θ.
- M plays an important role in the desired uncertainty rule.
- Properties
    - Generally, $1/\sqrt{N} \le M \le 1$.
    - For Fourier+Trivial (identity) matrices $M = 1/\sqrt{N}$.
    - For random pairs of ortho-matrices $M \approx 2\sqrt{\log_e N}/\sqrt{N}$.

Uncertainty Rule

Theorem 1

$$\|\underline{\alpha}\|_0 + \|\underline{\beta}\|_0 \geq 2\sqrt{\|\underline{\alpha}\|_0 \cdot \|\underline{\beta}\|_0} \geq \frac{2}{M}$$ *

Examples:

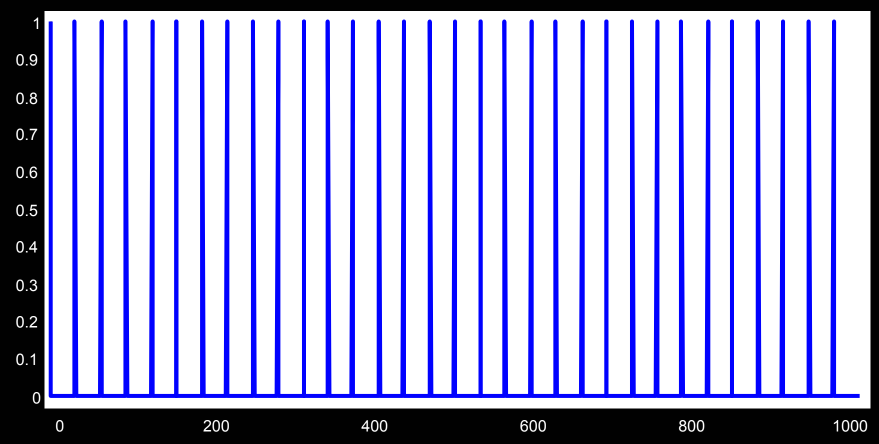
- $\Psi=\Theta$: M=1, leading to $\|\underline{\alpha}\|_0 + \|\underline{\beta}\|_0 \geq 2$.
- $\Psi=I$, $\Theta=F_N$ (DFT): $M = 1/\sqrt{N}$, leading to $\|\underline{\alpha}\|_0 + \|\underline{\beta}\|_0 \geq 2\sqrt{N}$.

* Donoho & Huo obtained a weaker bound $\|\underline{\alpha}\|_0 + \|\underline{\beta}\|_0 \geq \left(1 + M^{-1}\right)$

Example

$$\Psi = I,\ \Theta = F_N\ \text{(DFT)} \Rightarrow M = 1/\sqrt{N} \Rightarrow \|\underline{\alpha}\|_0 + \|\underline{\beta}\|_0 \ge 2\sqrt{N}$$

- For N=1024, $\|\underline{s}\|_0 + \|F \cdot \underline{s}\|_0 \ge 64$.
- The signal satisfying this bound: Picket-fence



Towards Uniqueness

- Given a unit norm signal $\underline{s}$, assume we hold two different representations for it using Φ

$$\underline{s} = \Phi\underline{\gamma}_1 = \Phi\underline{\gamma}_2$$

- Thus $\underline{0} = \Phi\underbrace{\left(\underline{\gamma}_1 - \underline{\gamma}_2\right)}_{\underline{x}} = [\Psi, \Theta]\begin{bmatrix}\underline{x}_1 \\ \underline{x}_2\end{bmatrix} \Rightarrow \boxed{\Psi\underline{x}_1 = -\Theta\underline{x}_2 = \underline{q}}$

- Based on the uncertainty theorem we just got:

$$\frac{2}{M} \le \left\|\underline{x}_1\right\|_0 + \left\|\underline{x}_2\right\|_0 = \left\|\underline{\gamma}_1 - \underline{\gamma}_2\right\|_0 \le \left\|\underline{\gamma}_1\right\|_0 + \left\|\underline{\gamma}_2\right\|_0$$

Uniqueness Rule

$$\frac{2}{M} \leq \left\| \underline{\gamma}_1 \right\|_0 + \left\| \underline{\gamma}_2 \right\|_0$$

In words: Any two different representations of the same signal CANNOT BE JOINTLY TOO SPARSE.

Theorem 2

If we found a representation that satisfy *

$$\frac{1}{M} > \left\| \underline{\gamma} \right\|_0$$

Then necessarily it is unique (the sparsest).

* Donoho & Huo obtained a weaker bound $\left\| \underline{\gamma} \right\|_0 < 0.5\left(1 + M^{-1}\right)$

Uniqueness Implication

- We are interested in solving

$$P_0: \quad \underset{\underline{\alpha}}{\text{Min}} \ \|\underline{\gamma}\|_0 \ \text{ s.t. } \ \underline{s} = [\Psi, \Theta]\underline{\gamma}.$$

- Somehow we obtain a candidate solution $\hat{\underline{\gamma}}$.
- The uniqueness theorem tells us that a simple test on $\hat{\underline{\gamma}}$ ($M \cdot \|\hat{\underline{\gamma}}\|_0 < 1$) could tell us if it is the solution of P_0.
- However:
  - If the test is negative, it says nothing.
  - This does not help in solving P_0.
  - This does not explain why P_1 may be a good replacement.

Equivalence - Goal

- We are going to solve the following problem

$$P_1 : \quad \underset{\underline{\alpha}}{\text{Min}} \; \left\|\underline{\gamma}\right\|_1 \; \text{ s.t. } \; \underline{s} = [\Psi, \Theta]\underline{\gamma} .$$

- The questions we ask are:
  - Will the P_1 solution coincide with the P_0 one?
  - What are the conditions for such success?

- We show that if indeed the P_0 solution is sparse **enough**, then P_1 solver finds it exactly.

Equivalence - Result

Given a signal $\underline{s}$ with a representation $\underline{s} = [\Psi, \Theta]\underline{\gamma}$,

Assuming a sparsity on $\underline{\gamma}$ such that (assume $k_1 < k_2$)

$$\underline{\gamma} = [\underbrace{\gamma_1 \quad \gamma_2 \quad \cdots \quad \gamma_N}_{k_1 \text{ non-zeros}}, \underbrace{\gamma_{N+1} \quad \gamma_{N+2} \quad \cdots \quad \gamma_{2N}}_{k_2 \text{ non-zeros}}]$$

Theorem 3

If k_1 and k_2 satisfy $2M^2 k_1 k_2 + M k_2 - 1 < 0$ then P_1 will find the correct solution.

A weaker requirement is given by $k_1 + k_2 < \frac{\sqrt{2} - 0.5}{M}$ *

* Donoho & Huo obtained a weaker bound $\|\underline{\gamma}\|_0 < 0.5(1 + M^{-1})$

The Various Bounds

Signal dimension: N=1024,

Dictionary: $\Psi=I$, $\Theta=F_N$,

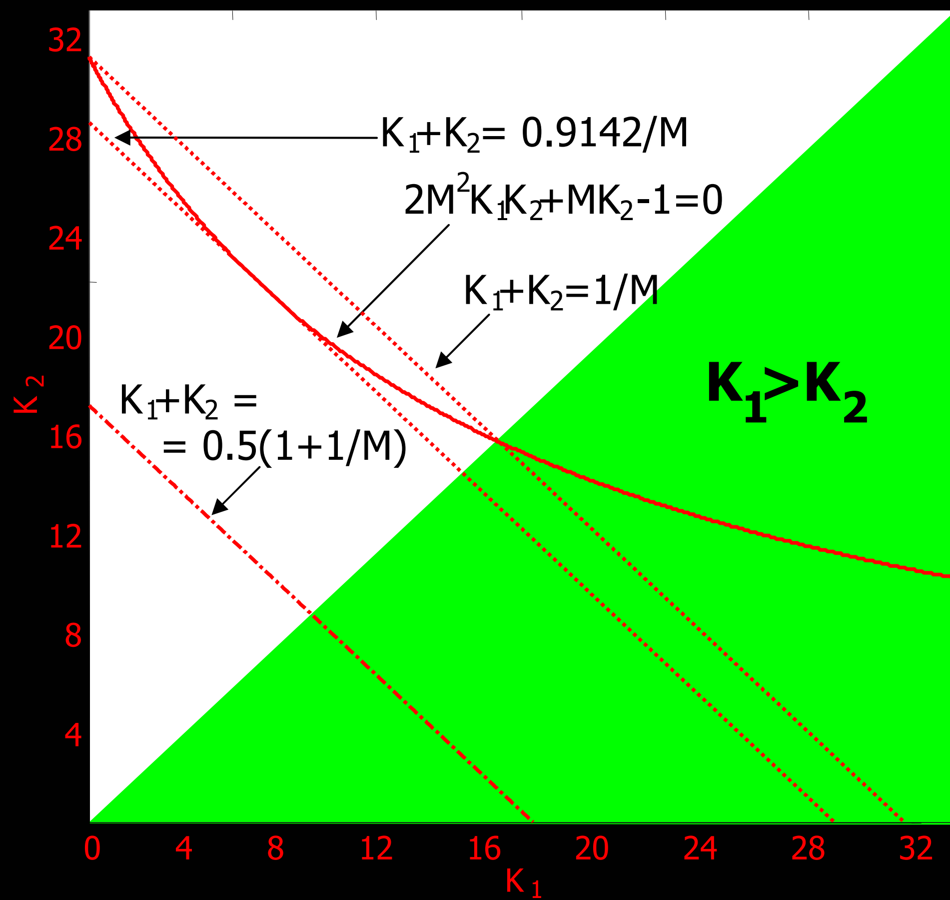
Mutual incoherence M=1/32.

<u>Results</u>

Uniqueness: 32 entries and below,

Equivalence:

- 16 entries and below (D-H),
- 29 entries and below (E-B).



Equivalence – Uniqueness Gap

- For uniqueness we got the requirement $\|\underline{\gamma}\|_0 < \frac{1}{M}$
- For equivalence we got the requirement $\|\underline{\gamma}\|_0 < \frac{\sqrt{2}-0.5}{M}$
- Is this gap due to careless bounding?
- Answer [by Feuer and Nemirovski, to appear in IEEE Transactions On Information Theory]: No, both bounds are indeed tight.

Agenda

1. Introduction

   Previous and current work

2. Two Ortho-Bases

   Uncertainty → Uniqueness → E

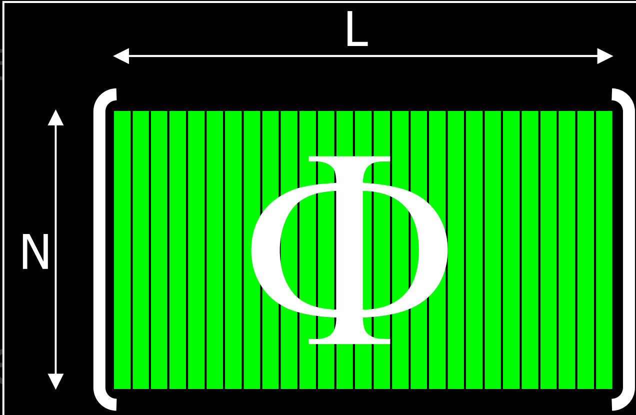
3. Arbitrary dictionary

   Uniqueness → Equivalence

4. Basis Pursuit for Inve

   Basis Pursuit Denoising → Bayesian (PDE) methods

5. Discussion



Every column is normalized to have an l_2 unit norm

Why General Dictionaries?

- Because in many situations
  - We would like to use more than just two ortho-bases (e.g. Wavelet, Fourier, and ridgelets);
  - We would like to use non-ortho bases (pseudo-polar FFT, Gabor transform, …),
  - In many situations we would like to use non-square transforms as our building blocks (Laplacian pyramid, shift-invariant Wavelet, …).
- In the following analysis we assume ARBITRARY DICTIONARY (frame). We show that BP is successful over such dictionaries as well.

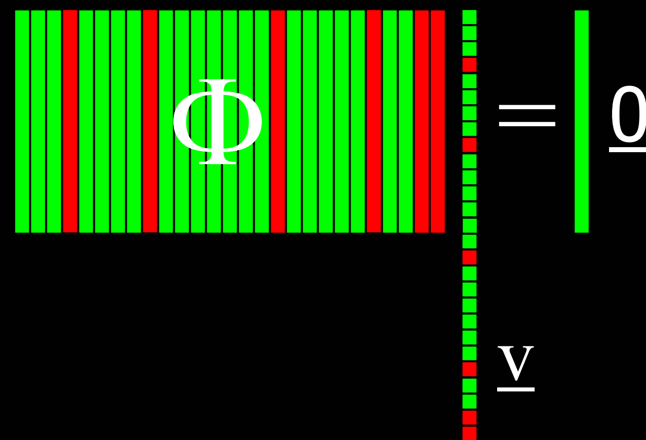
Uniqueness - Basics

- Given a unit norm signal $\underline{s}$, assume we hold two different representations for it using Φ

$$\underline{s} = \Phi\underline{\gamma}_1 = \Phi\underline{\gamma}_2 \quad \Rightarrow \quad \Phi\left(\underline{\gamma}_1 - \underline{\gamma}_2\right) = \underline{0}$$

- In the two-ortho case - simple splitting and use of the uncertainty rule – here there is no such splitting !!

- The equation $\Phi\underline{v} = \underline{0}$ implies a linear combination of columns from Φ that are linearly dependent. What is the smallest such group?



Uniqueness – Matrix "Spark"

Definition: Given a matrix Φ, define σ=Spark{Φ} as the smallest integer such that there exists at least one group of σ columns from Φ that is linearly dependent. The group realizing σ is defined as the "Critical Group".

Examples:

$$\text{Spark}\left\{\left[\begin{array}{cccc:c} 1 & 0 & \cdots & 0 & 1 \\ 0 & 1 & \cdots & 0 & 1 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & 1 \end{array}\right]\right\}=N+1; \quad \text{Spark}\left\{\left[\begin{array}{cccc:c} 1 & 0 & \cdots & 0 & 1 \\ 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & 0 \end{array}\right]\right\}=2$$

"Spark" versus "Rank"

The notion of spark is confusing – here is an attempt to compare it to the notion of rank

Rank	Spark
Definition: Maximal # of columns that are linearly independent	Definition: Minimal # of columns that are linearly dependent
Computation: Sequential - Take the first column, and add one column at a time, performing Gram-Schmidt orthogonalization. After L steps, count the number of non-zero vectors – This is the rank.	Computation: Combinatorial - sweep through 2^L combinations of columns to check linear dependence - the smallest group of linearly dependent vectors is the Spark.

Generally: $2 \le \sigma = \text{Spark}\{\Phi\} \le \text{Rank}\{\Phi\}+1$.

Uniqueness – Using the "Spark"

- Assume that we know the spark of Φ, denoted by σ.
- For any pair of representations of $\underline{s}$ we have

$$\underline{s} = \Phi\underline{\gamma}_1 = \Phi\underline{\gamma}_2 \quad \Rightarrow \quad \Phi\left(\underline{\gamma}_1 - \underline{\gamma}_2\right) = \underline{0}$$

- By the definition of the spark we know that if $\Phi\underline{v}=0$ then $\left\|\underline{v}\right\|_0 \geq \sigma$. Thus

$$\left\|\underline{\gamma}_1 - \underline{\gamma}_2\right\|_0 \geq \sigma$$

- From here we obtain the relationship

$$\sigma \leq \left\|\underline{\gamma}_1 - \underline{\gamma}_2\right\|_0 \leq \left\|\underline{\gamma}_1\right\|_0 + \left\|\underline{\gamma}_2\right\|_0$$

Uniqueness Rule – 1

$$\sigma \le \left\|\underline{\gamma}_1\right\|_0 + \left\|\underline{\gamma}_2\right\|_0$$

Any two different representations of the same signal using an **arbitrary dictionary** cannot be jointly sparse.

Theorem 4

If we found a representation that satisfy

$$\frac{\sigma}{2} > \left\|\underline{\gamma}\right\|_0$$

Then necessarily it is unique (the sparsest).

Lower bound on the "Spark"

- Define $$0(?) < M = \mathop{\mathrm{Max}}_{\substack{1\le k,j\le L \\ k\neq j}} \left\{ \left| \underline{\phi}_k^H \underline{\phi}_j \right| \right\} \le 1$$

  (notice the resemblance to the previous definition of M).

- We can show (based on Gerśgorin disks theorem) that a lower-bound on the spark is obtained by

$$\sigma \ge 1 + \frac{1}{M}.$$

- Since the Gerśgorin theorem is un-tight, this lower bound on the Spark is too pessimistic.

Uniqueness Rule – 2

$$1+\frac{1}{M}\le \sigma \le \|\underline{\gamma}_1\|_0+\|\underline{\gamma}_2\|_0$$

Any two different representations of the same signal using an **arbitrary dictionary** cannot be jointly sparse.

Theorem 5

If we found a representation that satisfy

$$\frac{\sigma}{2}\ge \frac{1}{2}\left(1+\frac{1}{M}\right)>\|\underline{\gamma}\|_0$$

Then necessarily it is unique (the sparsest). *

* This is the same as Donoho and Huo's bound! Have we lost tightness?

"Spark" Upper bound

- The Spark can be found by solving

$$\left\{ S_k : \ \underset{\underline{\gamma}}{\text{Min}} \left\|\underline{\gamma}\right\|_0 \quad \text{s.t.} \quad \Phi\underline{\gamma} = \underline{0} \ \ \& \ \ \gamma_k = 1 \right\}_{k=1}^{L} \Rightarrow \left\{ \underline{\gamma}_k^S \right\}_{k=1}^{L}$$

$$\Rightarrow \quad \sigma = \underset{1 \le k \le L}{\text{Min}} \left\| \underline{\gamma}_k^S \right\|_0$$

- Use Basis Pursuit

$$\left\{ Q_k : \ \underset{\underline{\gamma}}{\text{Min}} \left\|\underline{\gamma}\right\|_1 \quad \text{s.t.} \quad \Phi\underline{\gamma} = \underline{0} \ \ \& \ \ \gamma_k = 1 \right\}_{k=1}^{L} \Rightarrow \left\{ \underline{\gamma}_k^Q \right\}_{k=1}^{L}$$

- Clearly $\left\| \underline{\gamma}_k^Q \right\|_0 \ge \left\| \underline{\gamma}_k^S \right\|_0$. Thus $\sigma = \underset{1 \le k \le L}{\text{Min}} \left\| \underline{\gamma}_k^S \right\|_0 \le \underset{1 \le k \le L}{\text{Min}} \left\| \underline{\gamma}_k^Q \right\|_0$.

Equivalence – The Result

Following the same path as shown before for the equivalence theorem in the two-ortho case, and adopting the new definition of M we obtain the following result:

Theorem 6

Given a signal $\underline{s}$ with a representation $\underline{s} = \Phi\underline{\gamma}$,
Assuming that $\|\underline{\gamma}\|_0 < 0.5(1 + 1/M)$, P_1 (BP) is
Guaranteed to find the sparsest solution. *

* This is the same as Donoho and Huo's bound! Is it non-tight?

To Summarize so far …

Over-complete linear transforms – great for sparse representations

forward transform?

Basis Pursuit Algorithm

Why works so well?

We give explanations (uniqueness and equivalence) true for any dictionary

Practical Implications?

(a) Design of dictionaries,

(b) Test of solution for optimality,

(c) Applications of BP for scrambling, signal separation, inverse problems, …

Agenda

1. Introduction

   Previous and current work

2. Two Ortho-Bas

   Uncertainty → Unique

3. Arbitrary dictio

   Uniqueness → Equivalence

$$\underline{y} = \underline{x} + \underline{n}$$

4. Basis Pursuit for Inverse Problems

   Basis Pursuit Denoising → Bayesian (PDE) methods

5. Discussion

From Exact to Approximate BP

$$\text{A.}\quad \underset{\underline{\alpha}}{\text{Min}} \left\|\underline{\alpha}\right\|_1 \quad \text{s.t.} \quad \underline{y} = \Phi\underline{\alpha}$$

$$\text{B. } \underset{\underline{\alpha}}{\text{Min}} \left\|\underline{\alpha}\right\|_1 \quad \text{s.t.} \quad \left\|\underline{y} - \Phi\underline{\alpha}\right\|_2^2 \leq \delta^2$$

$$\text{C. } \underset{\underline{\alpha}}{\text{Min}} \quad \left\|\underline{\alpha}\right\|_1 + \lambda \left\|\underline{y} - \Phi\underline{\alpha}\right\|_2^2$$

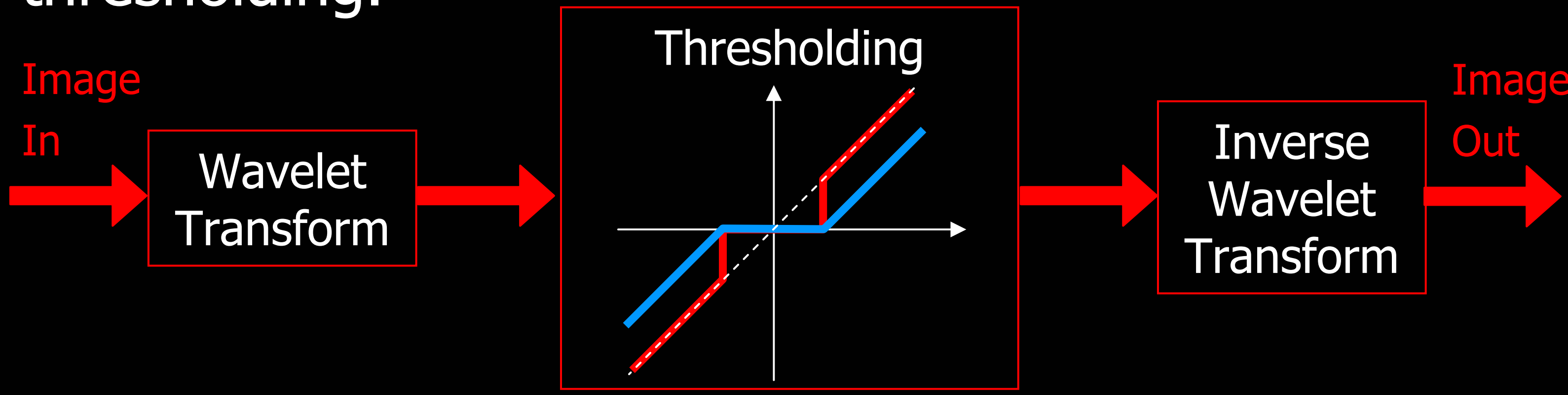
Wavelet Denoising

- Wavelet denoising by Donoho and Johnston (1994) –

$$\underset{\underline{x}}{\text{Min}} \left\|\underline{x}-\underline{y}\right\|_2^2 + \lambda\left\|W\underline{x}\right\|_p = \underset{\underline{\alpha}=W\underline{x}}{\text{Min}} \left\|W^T\underline{\alpha}-\underline{y}\right\|_2^2 + \lambda\left\|\underline{\alpha}\right\|_p$$

where W is an orthonormal matrix, and p=0 or 1.

- The result is very simple - hard (p=0) or soft (p=1) thresholding.



Shift Invariance Wavelet Denoising

- Major problem with Wavelet denoising – A shifted signal results with a different output - "shift-dependence".
- Proposed solution (Donoho and Coifman, 1995): Apply the Wavelet denoising for all shifted version of the W matrix and average – results very promising.
- In our language $\underset{\underline{\alpha}}{\text{Min}}\ \lambda\|\underline{\alpha}\|_1 + \left\|\left[W, DW, \cdots, D^{N-1}W\right]\underline{\alpha} - \underline{y}\right\|_2^2$.

$$\Rightarrow \left[W, DW, \cdots, D^{N-1}W\right]^{\#} = W^T\left[I, D, \cdots, D^{N-1}\right]^T$$

- Can be applied in the Bayesian approach – variant of the Bilateral filter.

Basis Pursuit Denoising

- A denoising algorithm is proposed for non-square dictionaries [Chen, Donoho & Saunders 1995]

$$\underset{\underline{\alpha}}{\text{Min}} \ \left\| \Phi \underline{\alpha} - \underline{y} \right\|_2^2 + \lambda \left\| \underline{\alpha} \right\|_1$$

- The solution now is not as simple as in the ortho-case, but the results are far better due to over-completeness!

- Interesting questions:
  - Which dictionary to choose?
  - Relation to other classic non-linear denoising algorithms?

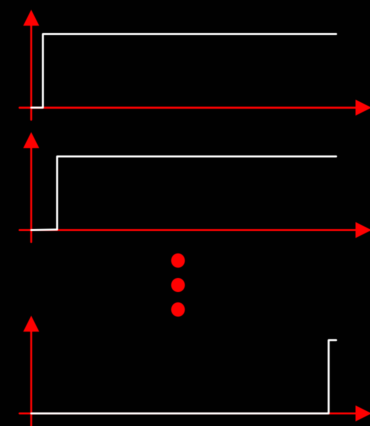
BP Denoising & Total Variation

- Relation between BP and the Total-Variation denoising algorithm [Rudin, Osher & Fatemi, 1992]? Answer is given by [Chen, Donoho & Saunders 1995]:

$$\text{TV}: \quad \underset{\underline{x}}{\text{Min}} \ \left\| \underline{x} - \underline{y} \right\|_2^2 + \lambda \text{TV}\{\underline{x}\}$$

- We have that $\text{TV}\{\underline{x}\} = \|\underline{\alpha}\|_1$ for $\underline{x} = \text{H}\underline{\alpha}$

  H is the *Heaviside* basis vectors.



$$\Rightarrow \quad \underset{\underline{\alpha}}{\text{Min}} \ \left\| \text{H}\underline{\alpha} - \underline{y} \right\|_2^2 + \lambda \|\underline{\alpha}\|_1$$

A General Bayesian Approach

- Our distributions are

$$P_{\underline{Y}/\underline{X}}\left(\underline{y}/\underline{x}\right)=C_1\cdot\exp\left\{\frac{1}{2\sigma_n^2}\left\|\underline{x}-\underline{y}\right\|_2^2\right\},\quad P_{\underline{X}}\left(\underline{x}\right)=C_2\cdot\exp\left\{\frac{-1}{2\sigma_x^2}\left\|\Omega^T\underline{x}\right\|_p\right\}$$

- Using the Maximum A-Posteriori Probability (MAP) we get

$$\hat{\underline{x}}_{MAP}=\underset{\underline{x}}{\text{ArgMax}}\ P_{\underline{X}/\underline{Y}}\left(\underline{x}/\underline{y}\right)=\underset{\underline{x}}{\text{ArgMax}}\ \frac{P_{\underline{Y}/\underline{X}}\left(\underline{y}/\underline{x}\right)P_{\underline{X}}\left(\underline{x}\right)}{P_{\underline{Y}}\left(\underline{y}\right)}$$

$$=\underset{\underline{x}}{\text{ArgMin}}\ \left\|\underline{x}-\underline{y}\right\|_2^2+\lambda\left\|\Omega^T\underline{x}\right\|_p$$

Generalized Result

- Bayesian denoising formulation $\underset{\underline{x}}{\text{Min}} \; \left\| \underline{x} - \underline{y} \right\|_2^2 + \lambda \left\| \Omega^T \underline{x} \right\|_p$

- Using $\Omega^T \underline{x} = \underline{\alpha} \;\Rightarrow\; \Omega\Omega^T \underline{x} = \Omega \underline{\alpha}$ and thus* $\Phi = \left(\Omega\Omega^T \right)^{-1} \Omega$

  we obtain $$\underset{\underline{\alpha}}{\text{Min}} \; \lambda \left\| \underline{\alpha} \right\|_p + \left\| \Phi \underline{\alpha} - \underline{y} \right\|_2^2$$

- Thus, we have a general relationship between Ω (Bayesian Prior operator) and Φ (dictionary).

* The case of non-full-rank Ω can be dealt-with using sub-space projection as a pre-stage, and using Economy SVD for pseudo-inverse.

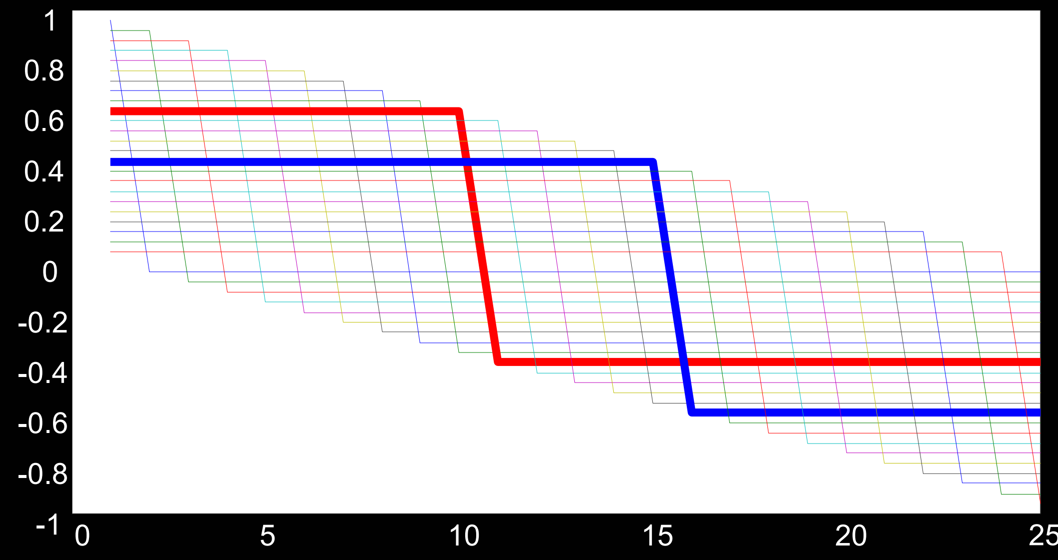
Example 1 – Total Variation

- Looking back at the TV approach we have (D – shift-right)

$$\underset{\underline{x}}{\text{Min}} \;\; \lambda\left\|\underline{x}-\underline{y}\right\|_2^2 + \left\|(I-D)\underline{x}\right\|_1$$

- Based on our result we have $(I-D)\underline{x}=\underline{\alpha} \;\Rightarrow\; \Phi=(I-D^T)^{\#}$

- Indeed we get a Heaviside basis. Moreover, finite support effects and singularity are taken into account properly.



Example 2 – Bilateral Filter

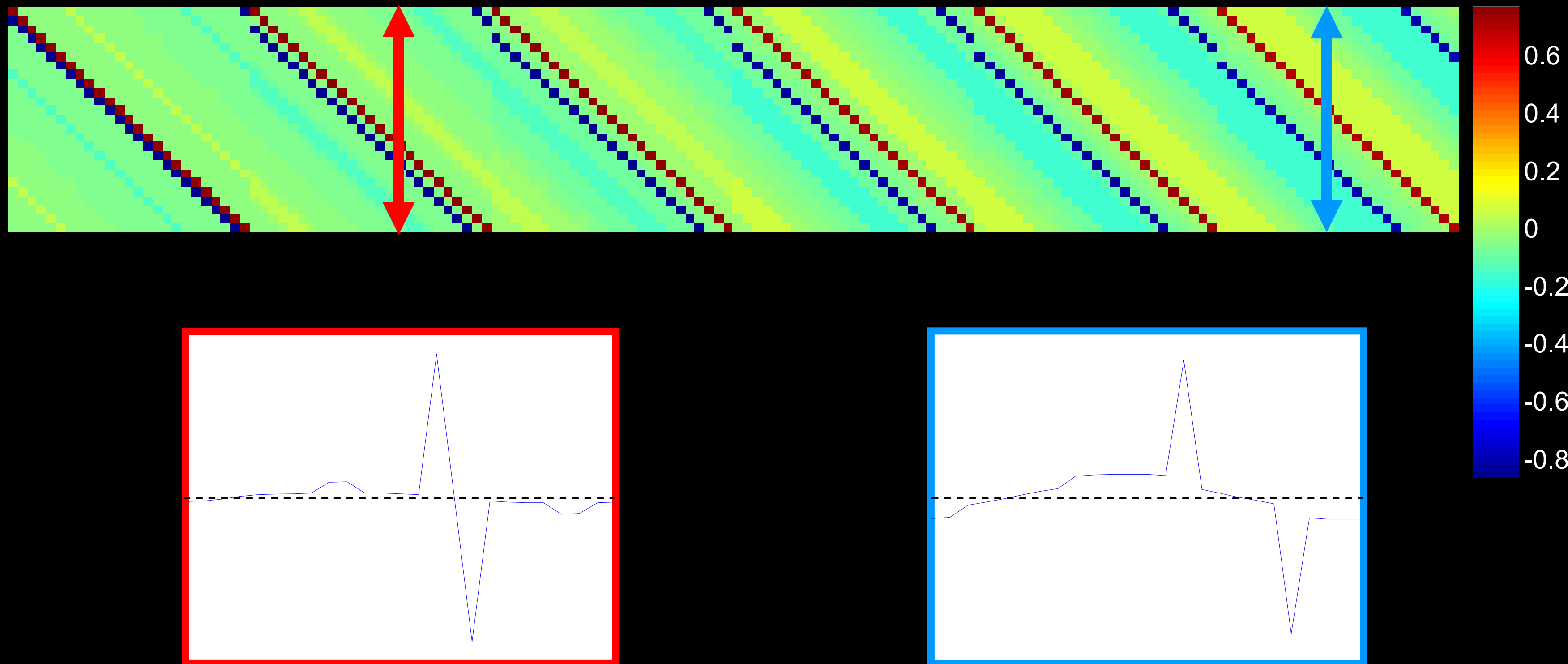
- ONE recent denoising algorithm of great impact:
  - Bilateral filter [Tomasi and Manduchi, 1998],
  - Digital TV [Chan, Osher and Shen, 2001],
  - Mean-Shift [Comaniciu and Meer, 2002].
- Recent work [Elad, 2001] show that these filters are essentially the same, being one Jacobi iteration minimizing

$$\underset{\underline{x}}{\text{Min}} \;\; \lambda \left\| \underline{x} - \underline{y} \right\|_2^2 + \left\| \begin{bmatrix} I - D^1 \\ \vdots \\ I - D^{k_0} \end{bmatrix} \underline{x} \right\|_p$$

- In [Elad, 2001] we give speed-up and other extensions for the above minimization – Implication: Speed-up the BP.

Example 2 – Bilateral Dictionary

The dictionary Φ has truncated (not all scales) multi-scaled and shift-invariant (all locations) 'derive-lets' :



Results

Original and noisy (σ^2=900) images

TV filtering:

10 iterations (MSE=146.3339) | 50 iterations (MSE=131.5013)

Wavelet Denoising (hard)

Using DB3
(MSE=154.1742)

Using DB5
(MSE=161.086)

Wavelet Denoising (soft)

Using DB3
(MSE=144.7436)

Using DB5
(MSE=150.7006)

Filtering via the Bilateral (BP equivalent):

2 iterations with 11×11 (MSE=89.2516)

Sub-gradient based 5×5 (MSE=93.4024)

Agenda

Part 5

Discussion

Summary

- Basis Pursuit is successful for
  - Forward transform – we shed light on this behavior.
  - Regularization scheme – we have shown relation to Bayesian non-linear filtering, and demonstrated the bilateral filter speed-up.
- The dream: the over-completeness idea is highly effective, and should replace existing methods in representation and inverse-problems.
- We would like to contribute to this change by
  - Supplying clear(er) explanations about the BP behavior,
  - Improve the involved numerical tools, and then
  - Deploy it to applications.

Future Work

- What dictionary to use? Relation to learning?
- BP beyond the bounds – Can we say more?
- Relaxed notion of sparsity? When zero is really zero?
- How to speed-up BP solver (both accurate and approximate)?
- Theory behind approximate BP?
- Applications – Demonstrating the concept for practical problems beyond denoising: Coding? Restoration? Signal separation? …