# The Invention of Compressive Sensing and Recent Results:

## From Spectrum-Blind Sampling and Image Compression on the Fly to New Solutions with Realistic Performance Guarantees

Yoram Bresler [1]

ECE and CSL, Univ. of Illinois at Urbana-Champaign

IEEE Statistical Signal Processing Workshop
August 8, 2012

ILLINOIS
UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN

[1]Supported in part by NSF grants No. CCF 06-35234 and CCF 10-18660.

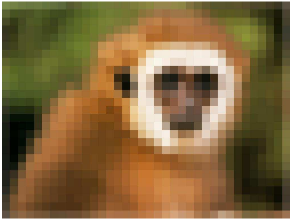
# Part I

# The Invention of Compressive Sensing

Since 1699, when French explorers
landed at the great bend of the
Mississippi River and celebrated
the first Mardi Gras in North
New Orleans has brewed a
melange of cultures. It was French
then Spanish, then French, again,
sold to the United States. Through
these years, and even into the
others arrived from everywhere:
Acadians (Cajuns), Africans,
Scene
Photodiode
A/D
Bitstream
Reconstruction
Image
DMD
Array
RNG

# Sampling

- Interface between analog and digital world

- Fundamental question: When can we reconstruct a signal from its samples?
- Whittaker-Nyquist-Kotelnikov-Raabe-Gabor-Shannon-Someya sampling of BL signals

# What is Compressed Sensing/Compressed Sampling?

Convex relaxation?

Fundamental limits?

Theoretical guarantee?

Gaussian i.i.d sensing matrix?

Greedy algorithm?

Sparsity?

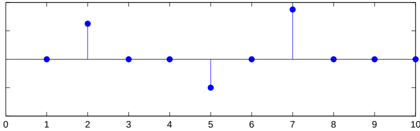
Random sampling?

Minimum sampling rate?

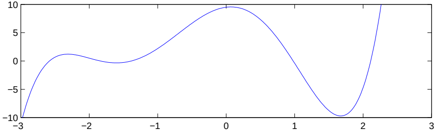
$\ell_1$ recovery?

# Sparse Signals

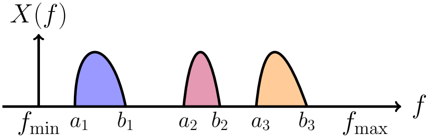
- Key notion: sparse (or sparsely representable) signals




- Sparsity level: $3$
- Sparsity rate: $3/10$




- $f(t) = c_0 + c_1 t + c_2 t_2 + c_3 t^3 + c_4 t^4 + c_5 t^5$
- Sparsity level: 6
- Sparsity rate: $6/\infty$

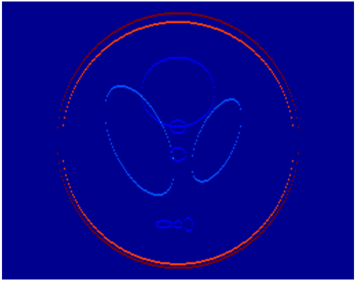



- Sparsity rate = occupied bandwidth/total bandwidth

# Sparse Signals

- Key notion: sparse (or sparsely representable) signals

Sparsity rate: 0.4946



Sparsity rate: 0.0324

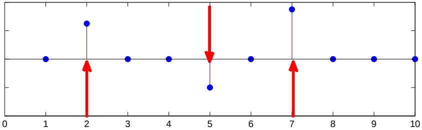
# Toward a Definition of Compressed Sampling

(i) **Sampling at the sparsity rate:** Signal is reconstructed from samples acquired at a rate essentially proportional to sparsity rate.

# Toward a Definition of Compressed Sampling

(i) **Sampling at the sparsity rate:** Signal is reconstructed from samples acquired at a rate essentially proportional to sparsity rate.

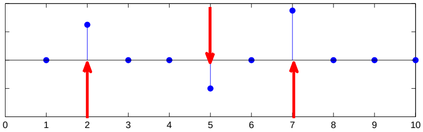
- Examples:



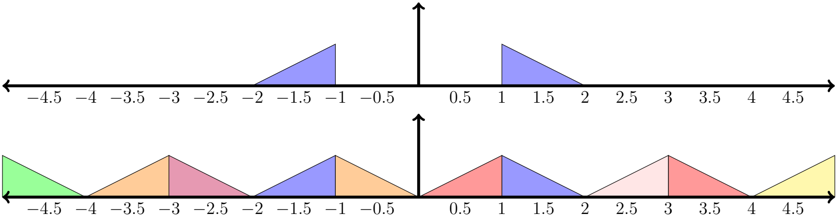
# Toward a Definition of Compressed Sampling

(i) **Sampling at the sparsity rate:** Signal is reconstructed from samples acquired at a rate essentially proportional to sparsity rate.

- Examples:

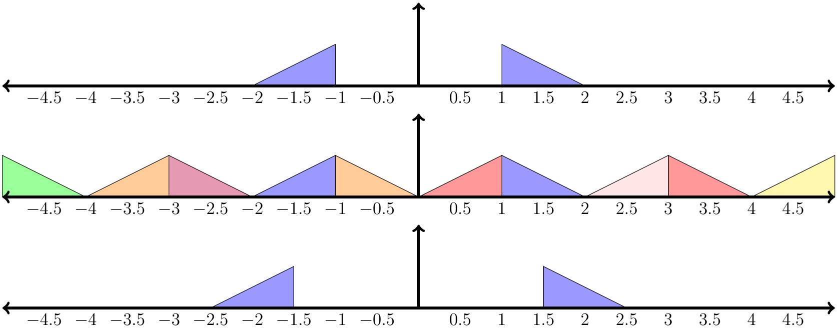



- Signal Bandwidth: $2$
- Sampling Frequency: $f_s = 2$



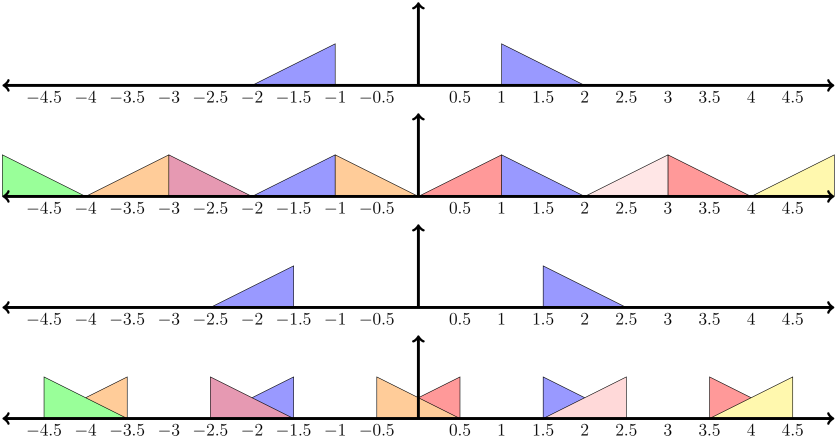
# Toward a Definition of Compressed Sampling

(i) **Sampling at the sparsity rate:** Signal is reconstructed from samples acquired at a rate essentially proportional to sparsity rate.



# Toward a Definition of Compressed Sampling

(i) **Sampling at the sparsity rate:** Signal is reconstructed from samples acquired at a rate essentially proportional to sparsity rate.



# Toward a Definition of Compressed Sampling

(i) *Sampling at the sparsity rate*

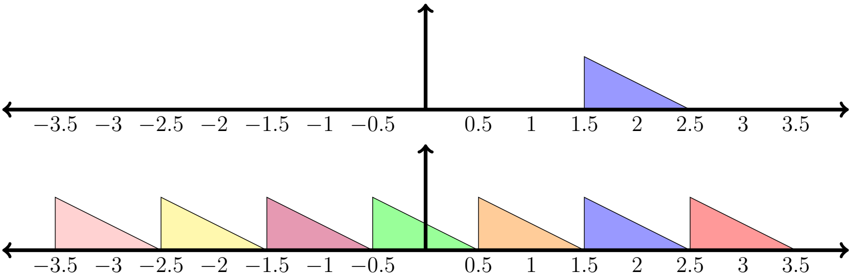
(ii) **Universality:** Sampling scheme (pattern) is *universal*- same for all signals in a given class.

# Toward a Definition of Compressed Sampling

(i) *Sampling at the sparsity rate*

(ii) **Universality:** Sampling scheme (pattern) is *universal*- same for all signals in a given class.

- Sampling Frequency: $f_s = B = 1$




- No aliasing, but can’t recover without additional information!

# Toward a Definition of Compressed Sampling

(i) *Sampling at the sparsity rate*

(ii) *Universality*

(iii) **Blind Recovery**: Signal can be (uniquely, stably) recovered without detailed prior knowledge or side information.

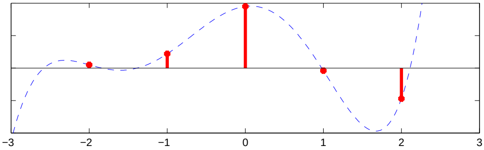
# Toward a Definition of Compressed Sampling

(i) *Sampling at the sparsity rate*

(ii) *Universality*

(iii) **Blind Recovery**: Signal can be (uniquely, stably) recovered without detailed prior knowledge or side information.

- Polynomial sampling satisfies all three, but isn't "compressed sensing"




$$L(x) = \sum_{k=0}^{n} y_k \prod_{\substack{m=0 \\ m\neq k}}^{k} \frac{x - x_m}{x_k - x_m}$$

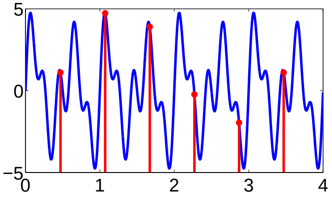
# Toward a Definition of Compressed Sampling

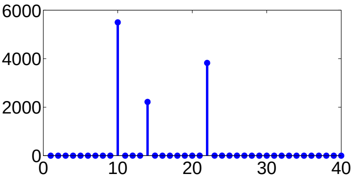
(i) *Sampling at the sparsity rate*
(ii) *Universality*
(iii) *Blind Recovery*
(iv) **Sampling rate independent of signal dimension**: Signal lives in high dimensional space, but # of samples is much smaller than dimension of space.

# Toward a Definition of Compressed Sampling

(i) *Sampling at the sparsity rate*
(ii) *Universality*
(iii) *Blind Recovery*
(iv) **Sampling rate independent of signal dimension**: Signal lives in high dimensional space, but # of samples is much smaller than dimension of space.

- Example: Spectral estimation of a sum of sinusoids with uniformly spaced sampling




$$x_c(t) = \sum_{i=1}^{d} s_i e^{j2\pi f_i T} \qquad x(m) = x_c(mT) \qquad \lambda_i = e^{j2\pi f_i T}$$

# Prony's Method (1795)

- $x(m)$ should be a solution to a difference equation:

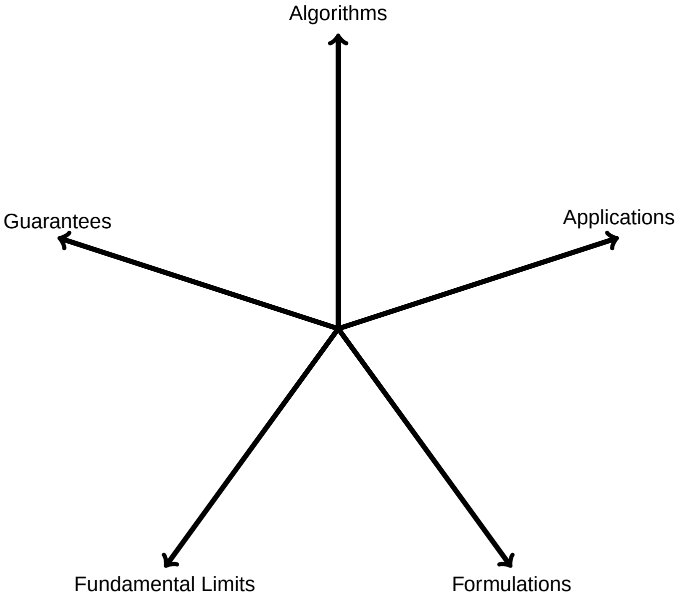
$$x(m) = -b_1 x(m-1) - b_2 x(m-2) \ldots - b_d x(m-d)$$
$$m = d, d+1, \ldots M+1$$
$$b(z) = 1 + b_1 z^{-1} + \ldots b_d z^{-d} = 0$$
$$z_i = \lambda_i$$

- If $M = 2d$, system matrix is $d \times d$ and nonsingular!
- $\Rightarrow$ unique solution for $\lambda_i, s_i$
- $\Rightarrow$ perfect recovery of $x(t)$ from $2d$ samples.
- Satisfies all four defining features of CS with an efficient, guaranteed algorithm.

# Was CS invented in 1795?

- **Missing feature**: Broad applicability to important classes of practical signals.

# Aspects of Compressed Sensing

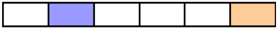


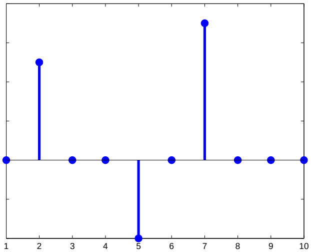
# Aspects of Compressed Sensing

- Formulations
  - Signals in Euclidean space (discrete-index, finite length)
  - 1-dimensional, multi-dimensional
  - Analog signals
  - Fourier-sparse signals
  - Sparsifiable signals
  - Signals sparse in a dictionary
  - Generalized sampling
- Fundamental Limits: Necessary and sufficent conditions
- Recovery algorithms
  - Computational efficency
  - Empirical performance
  - Theoretical analysis & guarantees

# Aspects of Compressed Sensing

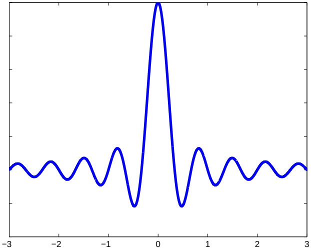
- Formulations
  - Signals in Euclidean space (discrete-index, finite length)
  - 1-dimensional, multi-dimensional
  - Analog signals
  - Fourier-sparse signals
  - Sparsifiable signals
  - Signals sparse in a dictionary
  - Generalized sampling
- Fundamental Limits: Necessary and sufficent conditions
- Recovery algorithms
  - Computational efficency
  - Empirical performance
  - Theoretical analysis & guarantees





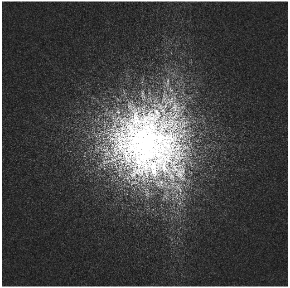
# Aspects of Compressed Sensing

- Formulations
  - Signals in Euclidean space (discrete-index, finite length)
  - 1-dimensional, multi-dimensional
  - Analog signals
  - Fourier-sparse signals
  - Sparsifiable signals
  - Signals sparse in a dictionary
  - Generalized sampling
- Fundamental Limits: Necessary and sufficent conditions
- Recovery algorithms
  - Computational efficency
  - Empirical performance
  - Theoretical analysis & guarantees



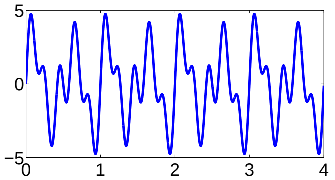

# Aspects of Compressed Sensing

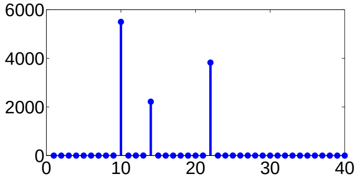
- Formulations
  - Signals in Euclidean space (discrete-index, finite length)
  - 1-dimensional, multi-dimensional
  - Analog signals
  - Fourier-sparse signals
  - Sparsifiable signals
  - Signals sparse in a dictionary
  - Generalized sampling
- Fundamental Limits: Necessary and sufficent conditions
- Recovery algorithms
  - Computational efficency
  - Empirical performance
  - Theoretical analysis & guarantees



# Aspects of Compressed Sensing

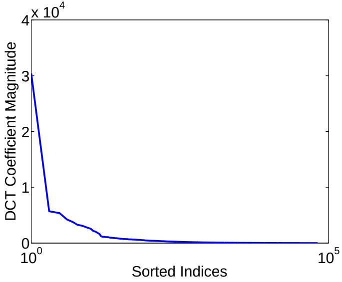
- Formulations
  - Signals in Euclidean space (discrete-index, finite length)
  - 1-dimensional, multi-dimensional
  - Analog signals
  - Fourier-sparse signals
  - Sparsifiable signals
  - Signals sparse in a dictionary
  - Generalized sampling
- Fundamental Limits: Necessary and sufficent conditions
- Recovery algorithms
  - Computational efficency
  - Empirical performance
  - Theoretical analysis & guarantees



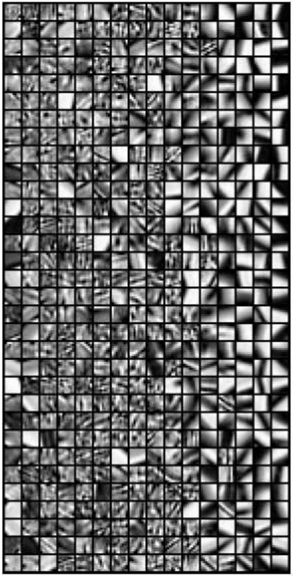
# Aspects of Compressed Sensing

- Formulations
  - Signals in Euclidean space (discrete-index, finite length)
  - 1-dimensional, multi-dimensional
  - Analog signals
  - Fourier-sparse signals
  - Sparsifiable signals
  - Signals sparse in a dictionary
  - Generalized sampling
- Fundamental Limits: Necessary and sufficent conditions
- Recovery algorithms
  - Computational efficency
  - Empirical performance
  - Theoretical analysis & guarantees



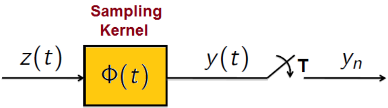
# Aspects of Compressed Sensing

- Formulations
  - Signals in Euclidean space (discrete-index, finite length)
  - 1-dimensional, multi-dimensional
  - Analog signals
  - Fourier-sparse signals
  - Sparsifiable signals
  - Signals sparse in a dictionary
  - Generalized sampling
- Fundamental Limits: Necessary and sufficent conditions
- Recovery algorithms
  - Computational efficency
  - Empirical performance
  - Theoretical analysis & guarantees



# Aspects of Compressed Sensing

- Formulations
  - Signals in Euclidean space (discrete-index, finite length)
  - 1-dimensional, multi-dimensional
  - Analog signals
  - Fourier-sparse signals
  - Sparsifiable signals
  - Signals sparse in a dictionary
  - Generalized sampling
- Fundamental Limits: Necessary and sufficent conditions
- Recovery algorithms
  - Computational efficency
  - Empirical performance
  - Theoretical analysis & guarantees



# Aspects of Compressed Sensing

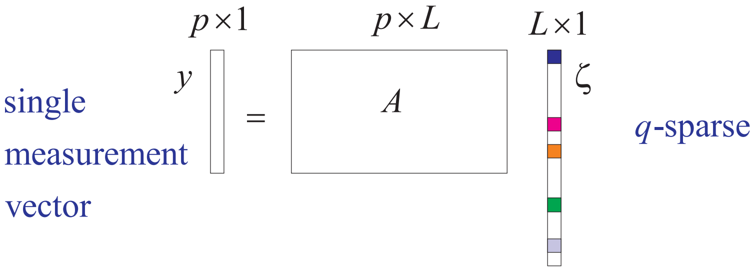
- Formulations
  - Signals in Euclidean space (discrete-index, finite length)
  - 1-dimensional, multi-dimensional
  - Analog signals
  - Fourier-sparse signals
  - Sparsifiable signals
  - Signals sparse in a dictionary
  - Generalized sampling
- Fundamental Limits: Necessary and sufficent conditions
- Recovery algorithms
  - Computational efficency
  - Empirical performance
  - Theoretical analysis & guarantees

# Aspects of Compressed Sensing

- Formulations
  - Signals in Euclidean space (discrete-index, finite length)
  - 1-dimensional, multi-dimensional
  - Analog signals
  - Fourier-sparse signals
  - Sparsifiable signals
  - Signals sparse in a dictionary
  - Generalized sampling
- Fundamental Limits: Necessary and sufficent conditions
- Recovery algorithms
  - Computational efficency
  - Empirical performance
  - Theoretical analysis & guarantees

# Compressive Sensing – Discrete, finite dimensional

- $y = A\zeta, \quad A \in \mathbb{C}^{p\times L},\ \zeta \in \mathbb{C}^L$
- $p =$ (# of measurements), $\|\zeta\|_0 =$ (sparsity level)
- Single measurement vector (SMV)



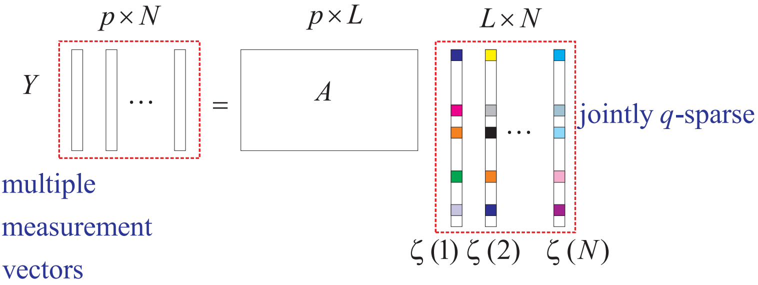

- $\mathrm{krank}(A) \geq s \implies$ Every $s+1$ columns of $A$ are linearly independent

P0: $\min_{\zeta} \|\zeta\|_0 \quad \text{subject to} \quad y = A\zeta.$

- Support recovery problem: Find $\{k_i\}_{i=1}^{q}$ s.t. $\zeta_{k_i} \neq 0$

# Compressive Sensing – Discrete, finite dimensional

- "Multiple Measurement Vector Problem" (MMV)
  - or "Jointly-sparse recovery"



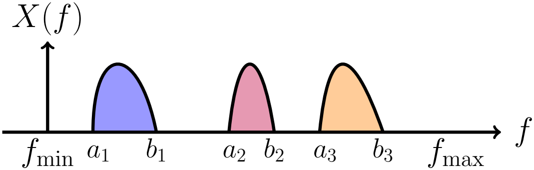

$$y(n) = A\zeta(n), \quad n = 1, 2, \cdots, n, \ N = \text{\# snapshots}$$
$$Y = A[\zeta(1),\ \zeta(2), \cdots, \zeta(N)]$$

- Joint support recovery problem

# Spectrum Blind Sampling (SBS) [Feng & Bresler 1996]

## Compressed Sensing of Analog Signals

- Multiband signal




- 1-D continuous-time signals

$$x(t) \leftrightarrow X(f) \quad t, f \in \mathbb{R}$$

- Spectrum-Sparse

$$X(f) = 0, f \notin \mathcal{F}, \quad \mathcal{F} = \bigcup_{i=1}^{n} [a_i, b_i) \subseteq [f_{\min}, f_{\max}]$$

$$\frac{\lambda(\mathcal{F})}{f_{\max} - f_{\min}} \leq \Omega < 1$$

# Questions

- Sampling rate requirements?
  - Known $\mathcal{F}$
  - Unknown $\mathcal{F}$
- Sampling at the minimum rate?
  - Design of sampling scheme
  - Reconstruction
- How achieve:
  - Universal (non-adaptive sampling)
  - Perfect (or robust) blind reconstruction
  - Computation linear in the data size

# Sampling Rate

- Throw in the towel (sufficient condition)

Nyquist sampling: $f_{\text{Nyq}} = f_{\max} - f_{\min}$

- Necessary condition, arbitrary pointwise sampling [Landau '67]

Landau lower bound: $D^- \geq \lambda(\mathcal{F})$

- Sufficient condition, known $\mathcal{F}$, packable or not [Kahn & Liu '65]

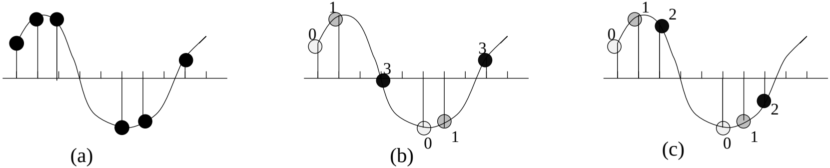
$$D^- \approx \lambda(\mathcal{F}) \leq \Omega f_{\text{Nyq}}$$

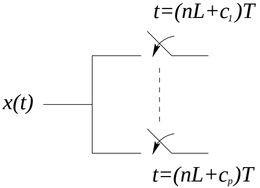
- Sufficient condition, unknown $\mathcal{F}$ [Feng & Bresler 1996]

$$D^- \approx \begin{cases} 2\Omega f_{\text{Nyq}} & \text{for all signals} \\ \Omega f_{\text{Nyq}} & \text{for almost all signals} \end{cases}$$

# Periodic Sampling

Non-uniform sampling (not reprentable by cosets) | Multicoset sampling Cosets = 0, 1, 3 | Bunched sampling Cosets = 0,1,2



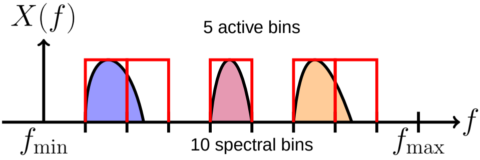




$$\frac{1}{T} \geq f_{\text{Nyq}}$$

- $(L, p)$ sampling pattern: $\mathcal{C} = \{c_i : 1 \leq i \leq p\}$
- Period = $L$

# Periodic Sampling

- Spectral support at resolution $L$




Active spectral cells:

$$\mathcal{F}_i \cap \mathcal{F} \neq \emptyset, i = 1, 2, \ldots, q$$

$$\Omega_L \triangleq \frac{q}{L} \approx \Omega$$

$$\mathcal{F}_0 = [0, \frac{1}{LT}], \quad \mathcal{F}_\ell = \mathcal{F}_0 + \ell/L, \quad \ell = 1, \ldots, L-1$$

- Vectorized signal spectrum

$$X_\ell(f) \triangleq X(f + \frac{\ell}{LT})\chi_{\mathcal{F}_0}(f)$$

$$\zeta(f) \triangleq [X_1(f), \ldots, X_L(f)]', f \in \mathcal{F}_0$$

- Vectorized sample spectrum

$$x_i(m) = x(Lm + c_i), m \in \mathbb{Z} \leftrightarrow X_i(f)$$

$$y(f) \triangleq LT[e^{-j2\pi c_1 fT} X_1(LTf), \ldots, e^{-j2\pi c_1 fT} X_p(LTf)]'$$

$y(f) = A\zeta(f), f \in \mathcal{F}_0, \quad A \in \mathbb{C}^{p \times L} =$ submatrix of DFT $L \times L$

# Compressive Sensing & Blind Reconstruction

$$y(f) = A\zeta(f), f \in \mathcal{F}_0, \quad A \in \mathbb{C}^{p \times L} = \text{submatrix of DFT } L \times L$$

$$\|\zeta(f)\|_0 \le \Omega_L L, \quad f \in \mathcal{F}_0$$

**P0:** for each $f \in \mathcal{F}_0$
$$\min_{\zeta(f)} \quad \|\zeta(f)\|_0 \quad \text{subject to} \quad y(f) = A\zeta(f)$$

- Common sparsity pattern for all
- MMV Problem
- But, uncountably infinite number of measurement vectors

# Spectral Support Recovery

- Active spectral cell indices:

$$\mathbf{k} = [k_1, k_2, \ldots, k_q]' \;:\; \mathcal{F}_{k_r} \cap \mathcal{F} \neq \emptyset$$
$$k_r \in \{1, 2, \ldots, L\} \quad r = 1, \ldots, q,$$

**P1:** $$\hat{\mathbf{k}} = \arg \min_{q, \|\mathbf{k}\|_0 = q} \int_{f \in \mathcal{F}_0} \| P^{\perp}_{\mathcal{R}(A_k)} y(f) \|_2^2 df$$

**P2:** $$\hat{\mathbf{k}} = \arg \min_{q, \|\mathbf{k}\|_0 = q} \operatorname{tr}\left( P^{\perp}_{\mathcal{R}(A_k)} R \right)$$

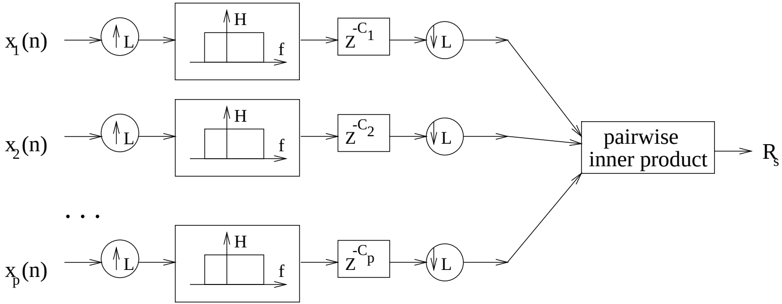
$$R \triangleq \int_{f \in \mathcal{F}_0} y(f) y^*(f) df$$

$$(R)_{k\ell} = \left\langle x_k(n - \frac{c_k}{L}), x_\ell(n - \frac{c_\ell}{L}) \right\rangle$$

- Finite dimensional problem!

# Finite-dimensional optimization

**P2:** $\hat{\mathbf{k}} = \arg \min_{q, \|\mathbf{k}\|_0 = q} \mathrm{tr}[P^{\perp}_{\mathcal{R}(A_{\mathbf{k}})} R]$



# More equivalent problems

**P3**: Given $\mathbb{R}^{p\times p}$ find the smallest integer $q$ and spectral cell index vector $\mathbf{k}$ of length $q$ such that $R = A_{\mathbf{k}} Z A_{\mathbf{k}}^*$ for some $Z \in \mathbb{C}^{q\times q}, Z \geq 0$

- Let $R = U_s \Lambda_s U_s^*, U \in \mathbb{C}^{p\times r}, r = \mathrm{rank}(R) = \mathrm{rank}(Z)$

**P4**: Given $U$ find the smallest integer $q$ and spectral cell index vector $\mathbf{k}$ of length $q$ such that $U = A_{\mathbf{k}} Q$ for some $Q \in \mathbb{C}^{q\times q}$

**Proposition**

Problems P0 – P4 are equivalent

- P4 is the (now) classical MMV problem!
- Mishali+Eldar (2007) proposed to solve it using MMV compressed sensing methods.

# Algebraic Bound: Fundamental Limit

[Wax and Ziskind '89], [Feng 1997], [Cotter et al. '05], [Chen and Huo '06]

- $\mathrm{krank}(A) \geq s \Leftrightarrow$ any $s$ columns of $A$ are linearly independent
  $\Rightarrow \mathrm{rank}(AX_0) = \mathrm{rank}(X_0) = r$

- Full krank (= ambiguity free): $\mathrm{krank}(A) = m$
  (always $\mathrm{krank}(A) \leq m + 1$)

- Algebraic bound for full krank case:
  support $J_0$ (indices of nonzero rows of $X_0$) can be uniquely determined iff the sparsity level $s$ satisfies

$$s < \frac{m + r}{2}$$

# MUSIC  Algorithm for Spectrum Blind Sampling

**MU**ltiple **SI**gnal **C**lassification

[Feng & Bresler 1996], [Feng 1997]

- $R = A_{\mathbf{k}} Z A_{\mathbf{k}}^*, \quad Z = \int_{f \in \mathcal{F}_0} \mathbf{z}(f)\mathbf{z}^*(f) df \geq 0$
- P3: Find smallest $q$ and $\mathbf{k} \in [L]^q$: $R = A_{\mathbf{k}} Z A_{\mathbf{k}}^*$
- MUSIC-like algorithm: (MUSIC [Schmidt '78], [Bienvenu & Kopp '78])

$$R = \left[U_s \ : \ U_S^L\right] \begin{bmatrix} \Lambda & 0 \\ 0 & 0 \end{bmatrix} \left[U_s \ : \ U_S^L\right]^*$$

$\hat{q} =$ # non-zero eigenvalues of $R$

$$\hat{\mathbf{k}} = \{k : U_S^* a_k = 0\}$$

- **Theorem**: If $\mathrm{rank}(Z) = q$ (full rank) and $\mathrm{krank}(A) \geq q + 1$, then $\mathbf{k} = \hat{\mathbf{k}}$, $q = \hat{q}$ (perfect recovery).

# Alternating Projections for Spectrum Blind Sampling

- Alternating Projection (AP) Algorithm [Wax & Ziskind '88]
- Application to the sampling problem [Feng & Bresler 1996], [Venkataramani & Bresler 1998]
- Solve

$$\text{P2}: \quad \min_{q, \|k\|_0 = q} \operatorname{tr}\left(P_{A_k}^{\perp} R\right)$$

or

$$\min_{q} \|q\|_0 \quad \text{subject to} \quad \operatorname{tr}\left(P_{A_k}^{\perp} R\right) \leq \epsilon$$

**Theorem (Feng & Bresler,1996, 1997; Venkataramani & Bresler, 1998)**

*Suppose* $\operatorname{krank}(A) = p$. *Then if* $p > \frac{q + rank(Z)}{2}$, *the minimizing* $k$ *is the correct support. Furthermore, for almost all* $Z$ *of any rank, the minimizing* $k$ *is the correct support if* $p > q + 1$.

# Alternating Projections for Spectrum Blind Sampling

- Solve:

$$\text{P2}: \quad \min_{q, \|k\|_0 = q} \operatorname{tr}\left(P_{A_k}^{\perp} R\right)$$

or

$$\min_{q} \|q\|_0 \quad \text{subject to} \quad \operatorname{tr}\left(P_{A_k}^{\perp} R\right) \leq \epsilon$$

- Greedy algorithm (normalized version of M-OMP, now known as Orthogonal Least Squares- OLS)
- In the $m$-th step, fix $\{k_1, \cdots, k_{m-1}\} \triangleq \mathbf{k}^{m-1}$, optimize over $\mathbf{k}$

$$\min_{k \in [L]} \operatorname{tr}\left(P_{[A_{\mathbf{k}^{m-1}},\ a_k]}^{\perp} R\right)$$

For Computational efficiency: $P_{[A,\ b]} = P_A + P_{P_A^{\perp} b}$

# Sampling Conditions

**Theorem (Feng & Bresler, 1996)**

*Let $\mathcal{X}$ consist of signals with spectral support of spectral occupancy at most $\Omega_L$ at resolution L. Then, all signals in $\mathcal{X}$ can be uniquely reconstructed from their samples on a periodic pattern with average rate $\approx 2\Omega_L f_{\max}$ and almost all signals can be reconstructed from samples at rate $\approx \Omega_L f_{\max}$.*

**Proof.**

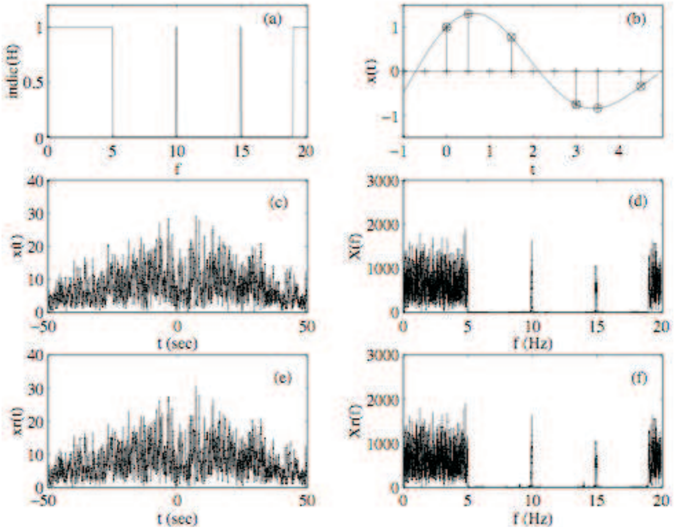
(a) For bunched pattern with $p \geq 2q + 1$

$$\mathrm{krank}(A) \geq 2q + 1$$

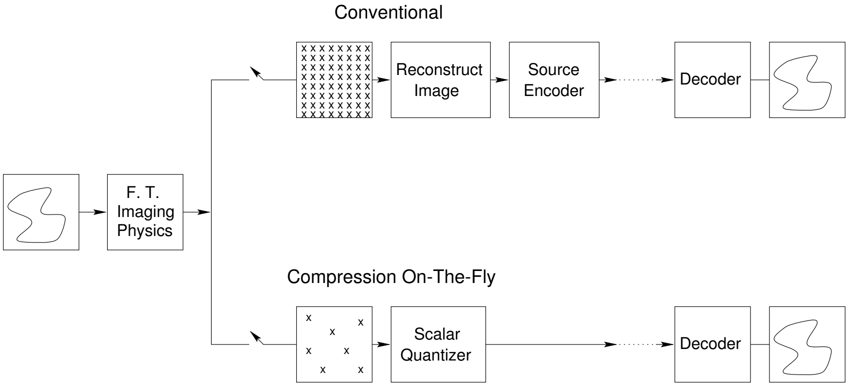
(b) Almost all $Z$ have full rank.

□

# Numerical Experiments

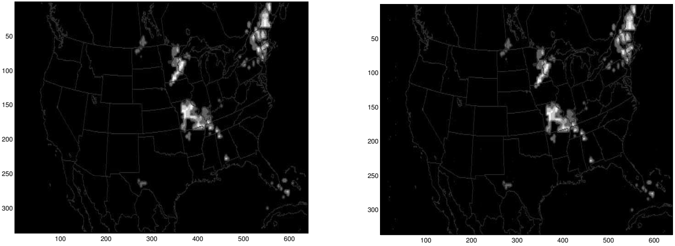


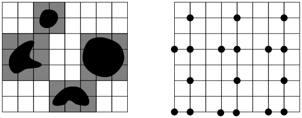
# Compression On-The-Fly




Bresler et al., 1999 - IEEE IT Workshop on Detection Estimation Classification and Imaging

# 2D Spectrum-Blind Sampling*



(a) (b)

*ICIP 1998

Venkataramani & Bresler 1998

# Image Compression On-The-Fly by Universal Sampling in Fourier Imaging Systems[1]

Yoram Bresler, Michael Gastpar, and Raman Venkataramani
University of Illinois
Coordinated Science Laboratory and ECE Department
1308 W. Main St. Urbana, IL 61801
Email: ybresler@uiuc.edu

*Abstract* — **An image compression system is proposed, where compression is trivially accomplished by minimally redundant acquisition, but decoding requires a nonlinear estimation step. The performance of this system is evaluated on examples and by studying the operational rate-distortion curves for a simple source model.**

## I. Introduction

Many important imaging systems, including radar imaging by real or synthetic aperture (SAR), and magnetic resonance imaging (MRI), acquire samples of the 2D Fourier transform of the image, rather than the image itself. The output of these so called Fourier imaging systems is often compressed, owing to channel or storage limitations. None the less, the traditional paradigm in these systems has been to acquire large quantities of data so as to allow the formation of high resolution images, and only then exploit the redundancy in the data to compress it. Because this redundancy usually takes the form of spatial correlations, it is often only apparent after formation of the image – a computationally expensive process.

This paper considers an alternative paradigm: directly acquire minimally redundant information, simplifying, or even eliminating the need for further compression. The approach relies on new results in sampling theory [1, 2, 3], and applies to a limited but fairly wide class of so-called sparse or LSI-sparsifiable images.

## II. Universal sampling and spectrum-blind reconstruction

is piece-wise constant on a number of regions is LSI sparsifiable by $h$ an appropriate differential operator and a $G$ supported on the frequency axes. We show that any image in this class can be perfectly recovered from a small fraction of the Nyquist-rate samples of $F$, provided that the selected samples are chosen on an appropriate *universal* sampling pattern.

## III. Compression by Sampling

For compression, the selected samples are scalar or vector quantized, followed by entropy coding. In general, the computational requirements of this scheme, which can avoid not only the image formation step, but also the acquisition and storage of the entire data set, will be far smaller than those of the conventional system. To analyze the effectiveness of this scheme, a simple doubly-stochastic Bernoulli-Gaussian model is considered for $f$ or for $H * f$. We compare the rate-distortion curves (computed numerically using the Blahut-Arimoto algorithm) for this source with the operational rate-distortion of the proposed scheme, and also illustrate the results on sample images.

## IV. Conclusion

Although in general suboptimal, the proposed scheme may be of interest in asymmetric compression applications requiring very low complexity compression (with possibly high complexity decoding), or when sparse acquisition is desirable for other reasons, such as sensor cost or physical constraints.

## References

# Sampling Pattern Design [Feng & Bresler 1996], [Feng 1997]

- Universality (maximal Kruskal Rank) is easy:
- Bunched pattern
  $\Leftrightarrow A$ Vandermond
  ($\Leftrightarrow$ uniform linear array w/ half wavelength spacing
  $\Leftrightarrow$ "Ambiguity free array manifold")
- Can we do better? Design Criterion:

$$\kappa_q \triangleq \max_{\|\mathbf{k}\|_0 = q} \frac{\sigma^2_{\max}(A_{\mathbf{k}})}{\sigma^2_{\min}(A_{\mathbf{k}})} \quad (q\text{-sparse condition \#})$$

- Select sampling pattern (row of DFT matrix) to minimize $\kappa_q$

$$Z(f) = A^L_{\mathbf{k}} y(f)$$

- Deterministic design (exhaustive search)
- Monte-Carlo design
  - Random sampling patterns
  - Random $\mathbf{k} \in [L]^q$ for testing

# Spectrum Blind Sampling and Image Compression on the Fly

[Bresler,Feng, Venkataramani, Gastpar, 1996-1999]

- Analog (infinite dimensional) compressed sensing
- Formulated as an MMV problem with infinitely many snapshots (locations of active spectral bins $\leftrightarrows$ support)
- Reduction to a finite dimensional MMV problem through eigenvalue decomposition of empirical correlation.
- Guaranteed MUSIC algorithm for support recover in full rank case.
- Conditions for $\ell_0$ recovery in arbitrary rank case.

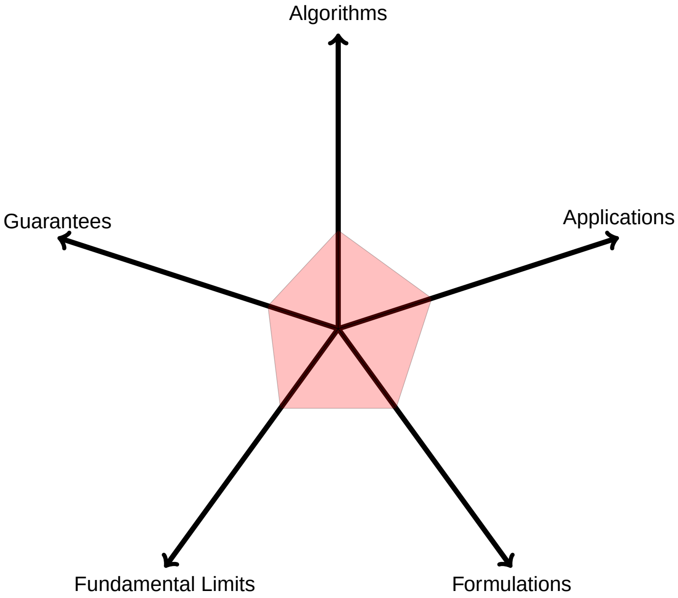
# Spectrum Blind Sampling and Image Compression on the Fly

[Bresler,Feng, Venkataramani, Gastpar, 1996-1999]

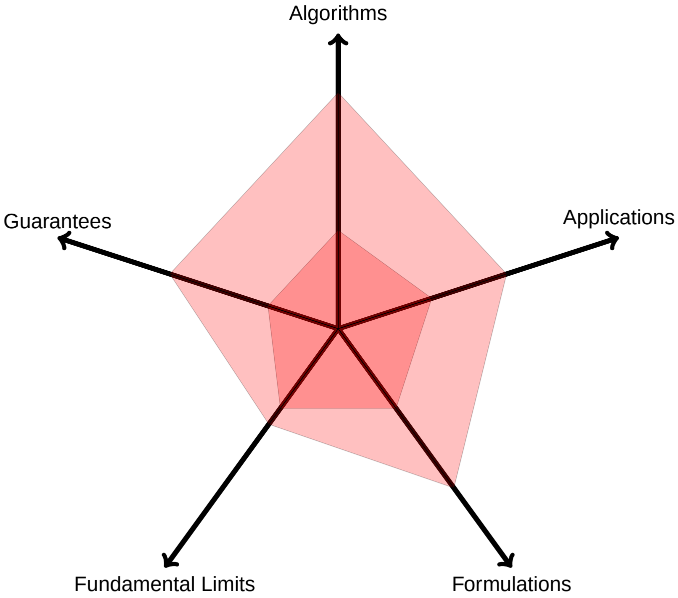
- M-OLS: Greedy algorithm for arbitrary rank case with good empirical performance.
- Both deterministic and randomized schemes for design of sampling scheme, with

  explicit RIP-like criterion
- Versions for 1D and 2D
- Versions for discrete-index and finite dimensions
- Applications to imaging
- Information-theoretic analysis of fundamental limitations

# Compressed Sensing in 1999



# Compressed Sensing Today



# Is Random Sampling a defining feature of Compressed Sampling?

- Even uniformly spaced sampling works fine in many cases (e.g, sinuoid retrieval, sensor array processing, etc)
- There are non-random, periodic, non-uniform universal sampling schemes [Feng & Bresler, 1996] [Venkataramani & Bresler, 1998]
- Expander graphs [Indyk, 2010]
- Other designs [Calderbank, 2010]
- $\Rightarrow$ Random sampling in CS is a **Marriage of Convenience:** convenient tool to generate sampling schemes with provably good properties, and to prove theoretical results about such schemes.
- Random sampling is neither essential for compressed sensing, nor is it optimal for finite size problems.

# Is $\ell_1$ Relaxation for Sparse Recovery a defining feature of Compressed Sensing?

- $\ell_1$ penalties used since the 1970's for sparse recovery with missing frequency information in seismic signal processing
  [Claerbout & Muir, 1973] [Levy & Fullager, 1981] [Santosa & Symes, 1986]
- Recovered sparse signal, but not at the sparsity rate! Only had a fraction of missing frequency data.
- $\ell_1$ sufficient condition from the 1980's:
  $2s(N-m) < N.M < N/2 \rightarrow s = 0$. To recover one spike, need at least half of the Fourier measurements!
- Similar condition from early Fourier uncertainty principle analysis
  [Donoho & Logan, 1992]

# Is $\ell_1$ Relaxation for Sparse Recovery a defining feature of Compressed Sensing?

- $\ell_1$ is just a method for sparse recovery that admits efficent algorithms.
- Early work (1970's – 2000) on sparse recovery using $\ell_1$ was not on Compressed Sensing

# Partial list of Contributions in Compressed Sensing

1991 Leahy & Jeffs : $\ell_p$ minimization ($0 < p < 1$) for sparse beamforming array

1992 Rao et al. : iterative reweighted least squares for Magnetoencephalography (MEG)

1995 Rao et al. : FOCUSS – $\ell_p$ relaxation ($0 < p < 1$) using reweighted least-squares with equality constraints

1996 J.J. Fuchs: Sensor array processing using $\ell_1$.

**1996** **Feng & Bresler** : spectrum blind sampling, guarantee of MUSIC for MMV, spark conditions

**1996** **Delaney & Bresler** : iterative tomographic reconstruction from sparse samples with non-convex relaxation and reweighting algorithms

**1997** **Harikumar & Bresler** : sparse solutions to linear inverse problems using convex and non-convex relaxations, and reweighting algorithms

**1997** **Feng** : algebraic conditions for MMV

**1998** **Venkataramani & Bresler** : algebraic conditions (spark) for uniqueness

**1999** **Bresler, Gastpar, & Venkataramani** : "Image Compression on the Fly"

1999 Fuch et al. : $\ell_1$ minimization

**2000** **Couvreur & Bresler** : guarantee for backward greedy

**2000** **Gastpar & Bresler** : information-theoretic analysis of compressive sensing with noise and with quantization

# Partial list of Contributions in Compressed Sensing

2001 Chen, Donoho, & Saunders : basis pursuit, $\ell_1$ minimization

2002 Vetterli et al.: sampling signals with finite rate of innovation

**2002 Ye, Bresler & Moulin** : non-linear image reconstruction from sparse Fourier samples, using sparsity of object edges

2003 Donoho & Elad : algebraic conditions (spark, coherence) for the uniqueness of the sparsest representation, $\ell_1$-$\ell_0$ equivalence

2004 Tropp : guarantee of OMP by cumulative coherence

2005 Candes & Tao : restricted isometry property, guarantee of $\ell_1$ minimization

2005 Rao et al.: algorithms for MMV (M-OMP / M-FOCUSS / M-ORMP), algebraic bound

2006 Donoho : guarantee of basis pursuit for random sensing matrix

2006 Candes Roberg Tao : guarantee of basis pursuit for random partial Fourier matrix

2006 Tropp : guarantee of convex relaxation for MMV

2006 Tropp et al : S-OMP + guarantee for MMV

2006 Chen & Huo : guarantee of convex relaxation for MMV for the noiseless case

2006 Baron et al.: distributed compressed sensing

2006 Indyk et al. : $\ell_1$-RIP of expander graph

2008 Baraniuk et al. : RIP of i.i.d. Gaussian matrix

2008 Baraniuk et al.: single pixel camera for compressed sensing

# Partial list of Contributions in Compressed Sensing

2008 Chartrand : RIP and guarantee of $\ell_p$ minimization ($0 < p < 1$)
2008 Lu & Do : theory of sampling signals from a union of subspaces
2008 Rudelson & Vershynin : RIP of random rows of DFT matrix
2008 Tropp : conditioning of random submatrices
2009 Dai & Milenkovic : subspace pursuit + guarantee
2009 Needell & Tropp : CoSaMP + guarantee
2009 Blumensath :iterative hard thresholding + guarantee
2009 Wainwright : information theoretic analysis
2009 Wainwright : guarantee of LASSO
2009 Reeves & Gastpar : information theoretic analysis
2009 Goyal et al. : information theoretic analysis
2009 Guo, Baron, & Shamai : information theoretic analysis
2009 Eldar & Mishali : repetition and variations on spectrum blind sampling
2010 Akcakaya & Tarokh : information theoretic analysis
2010 Baraniuk et al.: model based compressed sensing, algorithms and guarantees
2010 Calderbank et al. : statistical RIP of Deterministic Matrices
**2010** **Lee & Bresler** : SA-MUSIC + guarantee for MMV
2010 Kim, Lee, & Ye : compressive MUSIC + guarantee for MMV

# Would a rose by any other name smell as sweet?

Spectrum-Blind Sampling

Compressed Sampling

Image Compression on the Fly

Compressed Sensing

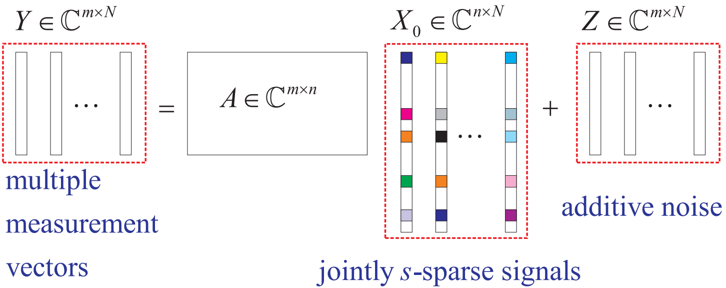
# Part II

## Subspace Methods for Joint Sparse Recovery

- Why joint sparse recovery (MMV)?
- Non-asymptotic guarantees for MUSIC with noise
- MUSIC Revitalized: overcoming the limitations of MMV algorithms

# Problem Statement

- Multiple Measurement Vectors (MMV)



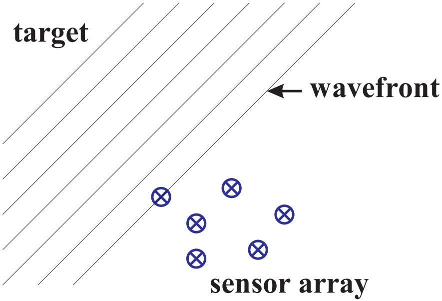

- Joint sparse recovery: given $Y$ and $A$, estimate

$$J_0 = \mathrm{supp}(X_0) \triangleq \{\text{indices of nonzero rows of } X_0\}$$

# Joint Sparse Recovery (JSR)?

- Discretized Direction of Arrival Estimation [Feng & Bresler, 1996, Malioutov et al. 2005]
- Compressed Sensing of Analog Signals
  - Spectrum-Blind Sampling of Multiband Signals [Feng & Bresler 1996, Mishali & Eldar 2009]
  - Modulated Wideband Converter [Mishali & Eldar, 2010]
- Compressive Fourier Imaging [Bresler & Feng, Bresler, Venkataramani, & Gastpar, 1998]
- Diffuse Optical Imaging [Lee, Kim, Bresler, Ye 2011]
- Multivariate Regression [Obozinski et al. 2011]

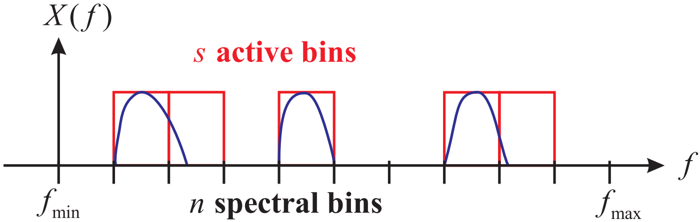
# Discretized Direction of Arrival (DOA) Estimation



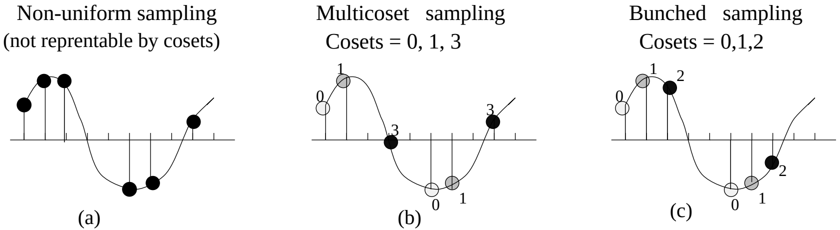

- Direction of arrival is discretized (the sine of the angle is uniformly discretized on the interval $[-1, 1)$, larger $n \to$ more accurate)
- $m$ sensors, $s$ sources, $N$ snapshots
- Discretized DOA estimation = joint sparse recovery with partial DFT matrix [Feng & Bresler 1996, Malioutov et al. 2005]]

# Spectrum Blind Sampling (SBS) [Feng & Bresler 1996]

- Multiband signal




- Multi coset sampling at sub-Nyquist rate

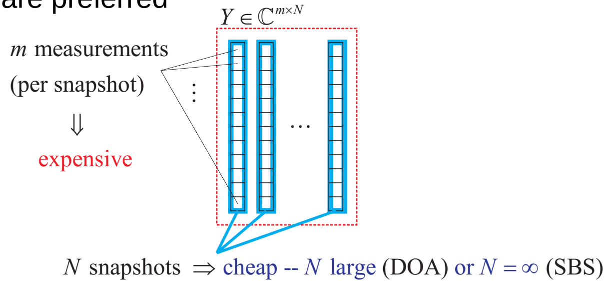


# Spectrum Blind Sampling

- Analog (infinite dimensional) compressed sensing
- Formulated as an MMV problem with $\infty$ snapshots
- Reduced to a finite dimensional MMV problem through eigenvalue decomposition of empirical correlation matrix [Feng & Bresler 1996]
- Wish to recover sparse support = locations of active spectral bins
- $m$: # cosets in spectrum blind sampling (multi-coset sampling)
- $m/n$: subsampling factor relative to Nyquist-rate
  $\Rightarrow$ Make $m$ small!

# Special MMV Problems (DOA, SBS)

- We focus on special MMV problems (DOA, SBS) where subspace methods are preferred




- Support recovery is explicitly required.
- $A$ is a random partial DFT matrix.
  (Columns of $A$ are not statistically independent)

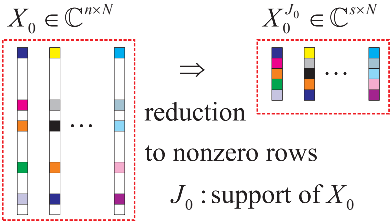
# Playing MUSIC with Noise and Finite Data

- Assumption: universal sampling pattern
- Ideal Scenario: zero noise, or white noise with $\infty$ snapshots

  $\Rightarrow$ perfect signal subspace estimate

  $\Rightarrow$ With "full-rank" signal, MUSIC is guaranteed for $m > s$ [Feng & Bresler, 1996]
- Non ideal Scenario
  - Q: What happens with finite data and noise?
  - A: Expect graceful degradation.
  - Non-asymptotic guarantees?

# MUSIC for JSR: Signal Subspace Estimation

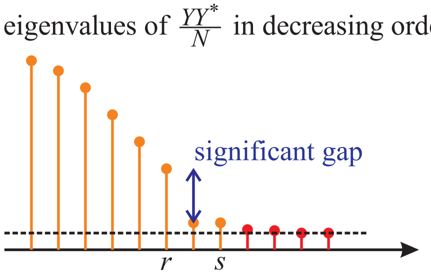



- $\mathcal{S} \triangleq \mathcal{R}(A_{J_0} X_0^{J_0})$: subspace spanned by signal component
- Goal: estimate a signal subspace (an $r$-dim subspace of $\mathcal{S}$) from finitely many noisy measurement vectors

$$Y = A_{J_0} X_0^{J_0} + W$$

# MUSIC for JSR: Signal Subspace Estimation

- Rank-revealing eigenvalue decomposition of sample correlation matrix




- $U$ (= $r$-principal eigenvectors of $\frac{YY^*}{N}$)
  $\Rightarrow \widehat{\mathcal{S}} = \mathcal{R}(U)$ estimated signal subspace
- $r$ can be smaller than $\mathrm{rank}(X_0^{J_0})$

# MUSIC for JSR: Support Recovery

- Algebraic subspace criterion:

$$P_{\widehat{\mathcal{S}}}^{\perp} a_k = 0 \quad \text{iff} \quad k \in \widehat{J}$$

- An analytic subspace criterion (thresholding with normalization): find $s$ indices $k$'s that maximize

$$\frac{\|P_{\widehat{\mathcal{S}}} a_k\|_2}{\|a_k\|_2} = \|P_{\widehat{\mathcal{S}}} P_{\mathcal{R}(a_k)}\|_{\ell_2^n \to \ell_2^n}$$

# Algebraic Analysis of MUSIC in an Ideal Case

**Definition (Kruskal rank)**

$\mathrm{krank}(A)$ is maximum number $q$ such that every $q$ columns of $A$ are linearly independent.

- Ideal scenario
  - $\mathcal{S} \triangleq \mathcal{R}(AX_0) = R(A_{J_0})$: full row rank condition of $X_0^{J_0}$
  - $\widehat{\mathcal{S}} = \mathcal{S}$: perfect subspace estimation
- MUSIC is guaranteed (in this scenario) if $\mathrm{krank}(A) > s$
- $A$ is a partial DFT matrix with a universal sampling pattern ($\mathrm{krank}(A) = m$) (e.g. consecutive rows)
  $\Rightarrow$ MUSIC is guaranteed (in this scenario) if $m \geq s + 1$
- $m \geq s + 1$ is also a necessary condition for ANY method

# Restricted Isometry Properties

- *$s$-restricted isometry property (RIP)* [Candes & Tao 2005]:
  - Any $s$ columns of $A$ are uniformly well conditioned
  - Provides a uniform guarantee for all sparse signals
  - Holds for partial DFT if $m \geq Cs \ln^4 n$ where $C$ is an unspecified constant
- Weak-1 RIP [Eldar & Rauhut 2009]:
  - For fixed $s$ columns & any one additional column
  - Provides an instance guarantee for a single sparse signal
  - Holds for partial DFT if $m \geq Cs \ln n$ where $C$ is a fully specified constant
- *Weak-1 restricted isometry constant (weak-1 RIC)*:

$$\delta_{s+1}^{\text{weak}}(A; J) = \max_{k \notin J} \|A_{J\cup\{k\}}^* A_{J\cup\{k\}} - I_{s+1}\|.$$

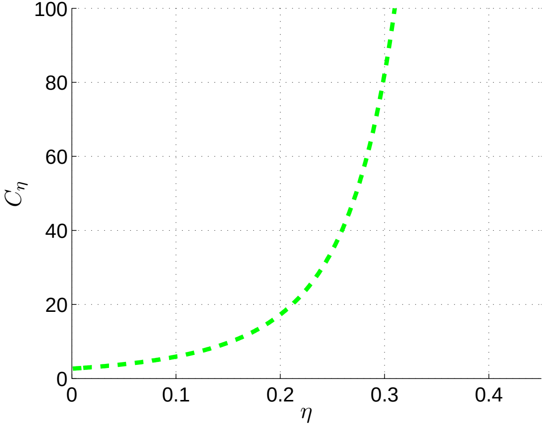
# MUSIC with Imperfect Subspace Estimate

- Scenario
  - $\mathcal{S} \triangleq \mathcal{R}(AX_0) = R(A_{J_0})$: full row rank condition of $X_0^{J_0}$
  - $\|P_{\widehat{\mathcal{S}}} - P_{\mathcal{S}}\| \leq \eta$: good subspace estimation
- MUSIC is guaranteed if $\delta_{s+1}^{\text{weak}}(A; J_0) < (1-2\eta)^2$
- Suppose $A$ is a random partial DFT matrix
  $\Rightarrow$ MUSIC is guaranteed with probability $1-\epsilon$ if

$$m \geq \underbrace{\left(\frac{8(1-\eta+\eta^2)}{3(1-2\eta)^4}\right)}_{=C_\eta} \underbrace{\left\{\ln\left(\frac{2(n-s)}{\epsilon}\right) + \ln(s+1)\right\}(s+1)}_{\approx s\ln n}.$$

# MUSIC with Imperfect Subspace Estimate

- Constant Factor Vs. Perturbation in Subspace Estimate



# Comparison of Instance Performance Guarantees

- Guarantees hold for a single signal instance with high probability when $m \geq Cs \ln n$ for random partial DFT matrix

| | Eldar & Rauhut | Candes & Plan | Lee, Bresler, Junge |
|---|---|---|---|
| SMV/MMV | MMV | SMV | MMV |
| Algorithm | M-BP | Lasso | MUSIC |
| Constant $C$ | specified (e.g. < 100) | unspecified (e.g. < 17,000) | specified (e.g. < 100) |
| Signal Model | exactly sparse multichannel model | arbitrary | exactly sparse full row rank |
| Noise Assumption | noiseless | i.i.d. Gaussian | good subspace est |
| Noise Amplification in Reconstruction | N/A | poly $\log n$ | specified constant |

# Conclusion?

- MUSIC proposed for JSR in the 1990s is still a good algorithm (in special scenarios)
  - Good empirical performance at low computational cost
  - Guaranteed by minimal requirements on the number of measurements
- But, when $\dim(\mathcal{S}) < s$ (rank defective case), MUSIC loses both advantages.
- A problem in practice! (Signal correlation, multipath effects, etc)
- Goal: improved JSR for the rank-defective case using the subspace idea behind MUSIC
  - Keep the order of # measurements in the guarantee of MUSIC at the cost of increased constant
  - Good empirical performance at low computational cost

# Subspace-Augmented MUSIC (SA-MUSIC)

- Q: can we make $\widehat{\mathcal{S}}$ into a good estimate of $\mathcal{R}(A_{J_0})$?
  $\Longrightarrow$ Given a partial support with at least $s-r$ elements, we can augment $\widehat{\mathcal{S}}$ to an $s$-dim subspace $\widetilde{\mathcal{S}}$ = estimate of $\mathcal{R}(A_{J_0})$.

**Definition**

Matrix $X$ is *row-nondegenerate* if $\mathrm{krank}(X^*) = \mathrm{rank}(X)$.

- Rows of $X$ are in general position $\Rightarrow$ $X$ is row-nondegenrate

**Proposition (Subspace augmentation)**

*Suppose $X_0^{J_0}$ is row-nondegenerate. Let $\bar{\mathcal{S}}$ be an $r$-dim subspace of $\mathcal{S}$. Let $J_1 \subset J_0$ and $|J_1| \geq s-r$. If $A_{J_0}$ has full column rank, then*

$$\bar{\mathcal{S}} + \mathcal{R}(A_{J_1}) = \mathcal{R}(A_{J_0}).$$

# Subspace-Augmented MUSIC (SA-MUSIC)

- $\widehat{\mathcal{S}}$: estimated signal subspace of dim $r$
- $\bar{\mathcal{S}}$: an $r$-dim subspace of $\mathcal{S}$ that minimizes $\|P_{\widehat{\mathcal{S}}} - P_{\bar{\mathcal{S}}}\|$
- Given a partial support $J_1$ ($|J_1| \geq s - r$), we construct an augmented subspace $\widetilde{\mathcal{S}}$ by $\widetilde{\mathcal{S}} = \bar{\mathcal{S}} + \mathcal{R}(A_{J_1})$.
- SA-MUSIC: MUSIC applied to $\widetilde{\mathcal{S}}$.

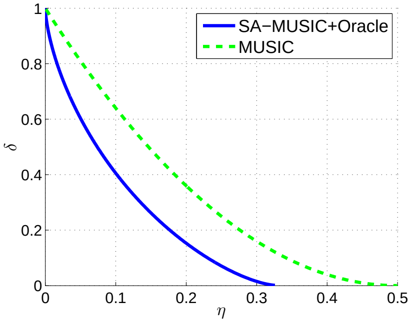
**Proposition (Perturbation in augmented subspace)**

*Suppose $X_0^{J_0}$ is row-nondegenerate and $\|P_{\widehat{\mathcal{S}}} - P_{\bar{\mathcal{S}}}\| \leq \eta$. If $\kappa(A_{J_0}) < \frac{1}{\eta}$, then*

$$\|P_{\widetilde{\mathcal{S}}} - P_{\mathcal{R}(A_{J_0})}\| \leq \frac{\kappa(A_{J_0})}{1/\eta - \kappa(A_{J_0})}.$$

# SA-MUSIC + "Oracle"

- Assumptions
  - Given arbitrary partial support $J_1$ of size $s - r$ (oracle)
  - $\|P_{\widehat{\mathcal{S}}} - P_{\bar{\mathcal{S}}}\| \leq \eta$: good subspace estimation
- SA-MUSIC is guaranteed by $\delta_{s+1}^{\text{weak}}(A; J_0) \leq \delta$



# Partial Support Recovery

- Q: how to obtain a correct partial support?
- A: *any MMV algorithm that can recover a partial support reliably.*
- Early stopping of forward greedy algorithms for partial support recovery saves computation and also avoids selection outside the true support
  $\Rightarrow$ forward greedy algorithms are good candidates!
- RA-ORMP [Davies and Eldar 2012] shares the same guarantee of MUSIC in an ideal scenario (full row rank, noiseless, exact sparsity) but is sensitive to noise

  $\Rightarrow$ Modification of RA-ORMP provides guarantees not restricted to the ideal scenario [Lee, Bresler, Junge 2012]

# Rank-Aware Order-Recursive Matching Pursuit (RA-ORMP)[Davis & Eldar 2012]

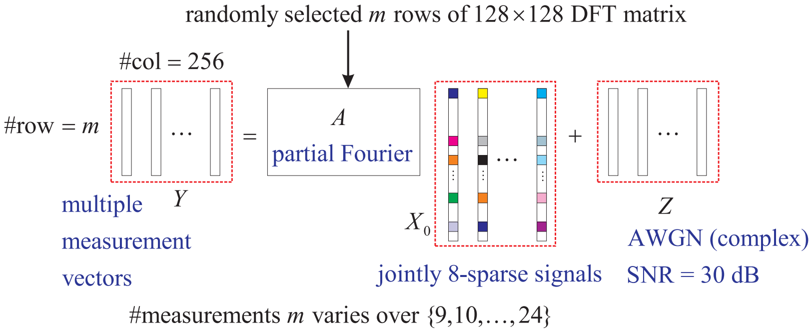
- A forward greedy algorithm
- Update rule: add the index for the subspace $\mathcal{R}(a_k)$ minimizing the distance between two subspaces $\mathcal{R}(P^{\perp}_{\mathcal{R}(A_J)} Y)$ and $P^{\perp}_{\mathcal{R}(A_J)} \mathcal{R}(a_k)$ in the "angle function" (a subspace metric)
- Rank-aware?: yes, in the noiseless case, RA-ORMP works with the subspaces whose dimensions are determined by the rank of $X_0^{J_0}$.
- In the noisy case, $Y$ may have full rank due to additive noise.
  $\Longrightarrow$ RA-ORMP loses its "rank-aware" ability.
- Q: what's the fix?

# Orthogonal Subspace Matching Pursuit (OSMP)

- Like MUSIC, OSMP performs rank-revealing EVD once to estimates $\widehat{\mathcal{S}}$.
- Update rule: add the index for the subspace $\mathcal{R}(a_k)$ that minimizing the distance between $P^{\perp}_{\mathcal{R}(A_J)}\widehat{\mathcal{S}}$ and $P^{\perp}_{\mathcal{R}(A_J)}\mathcal{R}(a_k)$
instead of the distance between $\mathcal{R}(P^{\perp}_{\mathcal{R}(A_J)}Y)$ and $P^{\perp}_{\mathcal{R}(A_J)}\mathcal{R}(a_k)$
- Slight but important change from RA-ORMP
- OSMP avoids the repetition of RR-EVD
- We propose to run OSMP up to $s - r$ iterations to get a partial support for SA-MUSIC ($\because$ empirically performed better in this way than running full $s$ steps).

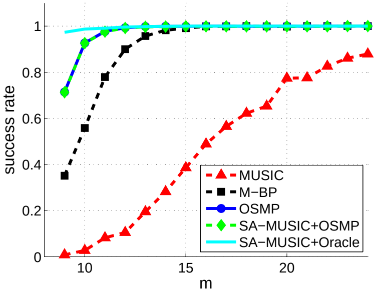
# Comparison of Empirical Performance

- Simulation Setup (relevant to DOA and SBS)

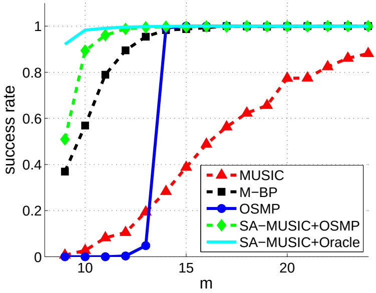



- All parameters except $m$ are fixed
  $\rightarrow$ Full support recovery with minimal $m$ is desired

# Empirical Performance of Subspace Methods



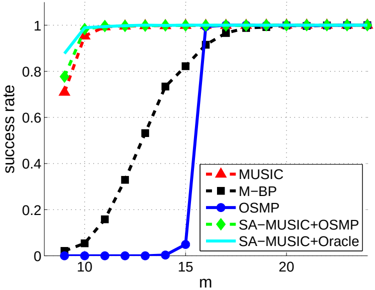

Noiseless



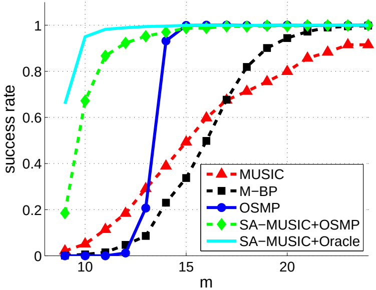

SNR = 30 dB

- $A \in \mathbb{C}^{m \times n}$: partial Fourier ($n = 128$). $X_0$: exactly 8 row-sparse. $N = 256$ snapshots. Rank-defective: $\mathrm{rank}(X_0) = 6$ and $\kappa(X_0) = 1$.
- MUSIC fails in the rank defective case.
- Full run of OSMP is vulnerable to noise.
- SA-MUSIC + OSMP is good!

# Empirical Performance of Subspace Methods



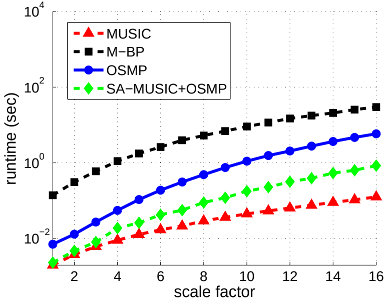

$$\kappa(X_0^{J_0}) = 10$$




$$\kappa(X_0^{J_0}) = 50$$

- Rank defective ($\dim(\widehat{\mathcal{S}}) < s$) when $X_0^{J_0}$ is ill conditioned.
- SA-MUSIC + OSMP is robust against this rank defect.

# Empirical Performance of Subspace Methods




- $n = (\text{scale factor}) \times 64$
  Other parameters also increase proportionally
- SA-MUSIC + OSMP is faster than M-BP!

# Analysis of OSMP

- Full Row Rank + Noiseless: OSMP = RA-ORMP $\Rightarrow$ RA-ORMP is guaranteed (in this scenario) if $\mathrm{krank}(A) > s$ [Davies & Eldar 2012]

- Full Row Rank + Imperfect Subspace Estimate: OSMP $\neq$ RA-ORMP $\Rightarrow$ MUSIC is guaranteed (in this scenario) if $\delta_{s+1}^{\text{weak}}(A; J_0) < (1 - 2\eta)^2$ [Lee, Bresler, Junge 2012]

- Rank-defective: OSMP is guaranteed if $\delta_{s+1}^{\text{weak}}(A; J_0) \leq \delta$ such that

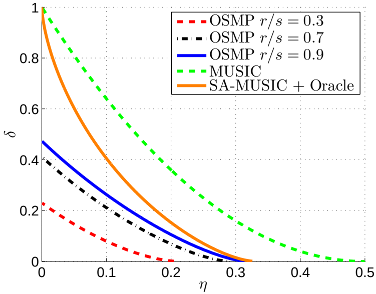
$$\eta \leq \sqrt{\frac{1-\delta}{1+\delta}} \cdot \frac{\sqrt{1-\delta} - \sqrt{\delta}}{2 + \sqrt{1-\delta} - \sqrt{\delta}}$$

$\Rightarrow$ Of course, also guarantees the first $s - r$ steps

[Lee, Bresler, Junge 2012]

# Analysis of OSMP in Rank Defective Case

- Weak-1 RIC Vs. Perturbation

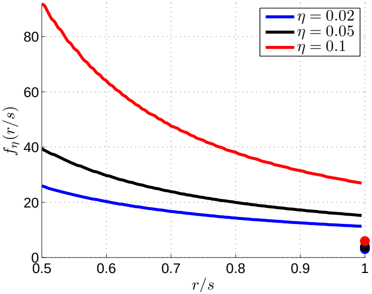



$\Rightarrow$ SA-MUSIC step is easier to guarantee than partial support recovery by OSMP
$\Rightarrow$ SA-MUSIC + OSMP is guaranteed by the same weak-1 RIP condition

# Analysis of SA-MUSIC+OSMP for Fourier

- SA-MUSIC + OSMP is guaranteed with probability $1 - \epsilon$ if

$$m \geq \underbrace{\left(\frac{8(1-\eta+\eta^2)}{3(1-2\eta)^4}\right)}_{=f_\eta(r/s)} \underbrace{\left\{\ln\left(\frac{2(n-s)}{\epsilon}\right) + \ln(s+1)\right\}(s+1)}_{\approx s \ln n}$$




$\Rightarrow$ benefits from larger $r/s$

# Related Works

- Compressive MUSIC (CS-MUSIC): a similar independently developed idea [Kim, Lee, Ye 2012]
- Like SA-MUSIC, CS-MUSIC provides a complete JSR algorithm when combined with another algorithm for partial support recovery.
- OSMP provides better empirical performance and stronger guarantees than 2-thresholding and Subspace S-OMP proposed to be used with CS-MUSIC.
- CS-MUSIC analyzed for i.i.d. Gaussian $A$ in the asymptotic case $\Rightarrow$ No implication for problems of finite size or with non-Gaussian matrix (not relevant to compressed sensing)
- Obozinski et al 2011: sharp phase transition of group Lasso for i.i.d. Gaussian $A$ in the non-asymptotic case (target application: multivariate regression, another MMV problem)

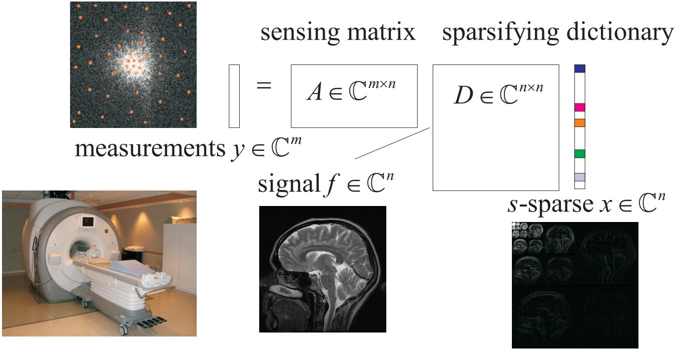
# Summary – Part 2

- New analysis shows MUSIC for JSR (from the 1990s) to be a guaranteed algorithm for the noisy, full-row-rank case, with best-in-class performance and guarantees.
- SA-MUSIC extends MUSIC to the rank-defective case.
- Overcomes the open problem in sensor array processing of signal coherence.
- Performance competitive with relaxation methods, with computation lower by orders of magnitude
- Performance guarantees
  - for the noisy case
  - with practical sensing matrices (random partial Fourier)
  - reveal the role of signal rank and advantage of MMV over SMV
  - modest, fully-specified empirical constant
  - non-asymptotic

# Part III

## Oblique Pursuits

# Compressed Sensing Fourier Imaging




- Sensing matrix $A$: dictated by the physics of modality
  $\Rightarrow$ i.i.d. Gaussian $A$ is not interesting
- Dictionary $D$: provides sparse approximation of signal of interest

# Problem Formulation

- $A$: random partial DFT matrix
  - $W$: $n \times n$ DFT matrix ($W^*W = nI_n$)
  - $\omega_1, \ldots, \omega_m$: $m$ i.i.d. copies of a random variable $\omega$

$$\omega \sim \mathbb{P}(\omega = j) = p_j \quad j = 1, \ldots, n$$

  - Construction of $A$: ($k$th row of $A$) $=$ ($\omega_k$th row of $W$) $\times \frac{1}{\sqrt{m}}$
- $D = [d_1, \ldots, d_n]$: sparsifying basis (invertible matrix).

*The algorithms and analysis are not restricted to this case.

# RIP-Guaranteed Recovery

- $s$-restricted isometry property (RIP) of $\Psi$:

$$\delta_s(\Psi) \triangleq \max_{|J|=s} \|\Psi_J^* \Psi - I_s\| \ll 1$$

☺ Various methods (BP, Lasso, CoSaMP, SP, IHT, HTP, etc) provide an approximation of $x^\star$ of guaranteed quality
if $\delta_{ks}(\Psi) < c$ for some $k \in \{2, 3, 4\}$ and $c \in (0, 1)$

# RIP of $\Psi = AD$

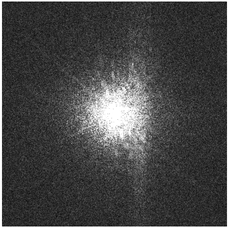
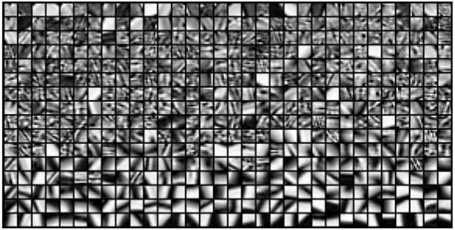
- Key assumptions:
  - $\mathbb{E}\Psi^*\Psi = I_n$: isotropy
  - $\max_{j,k} |(\Psi)_{j,k}| \le K$: incoherence

    Atoms (columns) of $D$ are incoherent to the rows of DFT
- Compressed sensing is possible:

$$m \ge C\delta^{-2} s \ln^4 n \Rightarrow \delta_s(\Psi) < \delta$$

- Q: When is “isotropy” satisfied?
  A: E.g. $D$ is an orthogonal matrix ($D^*D = I_n$) and $p_j = \frac{1}{n}, \forall j$
  (sampling with uniform distribution)

# Isotropy Breaks Down ...

- ☹ Energy of natural images is concentrated at lower frequencies
  $\Rightarrow$ Variable density recommended [Bresler et al. 1999, Lustig et al. 2007]

- ☹ Data-adaptive $D$: orthogonality constraint is too restrictive

$\Longrightarrow$ Motivates the study of the "anisotropic" case

# Q: Solution for the Anisotropic Case?

- Goal: close the gap between empirical success and theoretical guarantee of compressed sensing Fourier imaging
- Key idea: use properly designed $\widetilde{\Psi}^*$ instead of $\Psi^*$ for recovery by greedy algorithms
- Roadmap
  - extend RIP of $\Psi^*\Psi$ to restricted biorthogonality property (RBOP) of $\widetilde{\Psi}^*\Psi$
  - construction of $\widetilde{\Psi}$ that satisfies RBOP
  - modify RIP-guaranteed algorithms to RBOP-guaranteed algorithms (oblique pursuits)

# Q: Solution for the Anisotropic Case?

- Goal: close the gap between empirical success and theoretical guarantee of compressed sensing Fourier imaging
- Key idea: use properly designed $\widetilde{\Psi}^*$ instead of $\Psi^*$ for recovery by greedy algorithms
- Roadmap
  - extend RIP of $\Psi^*\Psi$ to restricted biorthogonality property (RBOP) of $\widetilde{\Psi}^*\Psi$
  - construction of $\widetilde{\Psi}$ that satisfies RBOP
  - modify RIP-guaranteed algorithms to RBOP-guaranteed algorithms (oblique pursuits)

# Q: Solution for the Anisotropic Case?

- Goal: close the gap between empirical success and theoretical guarantee of compressed sensing Fourier imaging

- Key idea: use properly designed $\widetilde{\Psi}^*$ instead of $\Psi^*$ for recovery by greedy algorithms

- Roadmap
  - extend RIP of $\Psi^*\Psi$ to restricted biorthogonality property (RBOP) of $\widetilde{\Psi}^*\Psi$
  - construction of $\widetilde{\Psi}$ that satisfies RBOP
  - modify RIP-guaranteed algorithms to RBOP-guaranteed algorithms (oblique pursuits)

# Restricted Biorthgonality Property

- Intuition: $\mathbb{E}\Psi^*\Psi = I_n$ too restrictive? How about $\mathbb{E}\Psi^*\Psi$ full rank? Much weaker condition!
- Restricted Biorthogonality Property (RBOP)

$$|\langle\widetilde{\Psi}y, \Psi x\rangle - \langle y, x\rangle| \leq \delta\|x\|_2\|y\|_2, \quad \forall \text{ jointly } s\text{-sparse } x, y$$

- Restricted biorthogonality Constant (RBOC)

$$\theta_s(\widetilde{\Psi}^*\Psi) \triangleq \max_{|J|=s} \|\widetilde{\Psi}_J^*\Psi_J - I_s\|$$

  $\theta_s(\widetilde{\Psi}^*\Psi)$ is the smallest $\delta$ satisfying the above inequality.

- RBOP of $(\Psi, \widetilde{\Psi}) \iff \forall J$ s.t. $|J| = s$, $(\Psi_J, \widetilde{\Psi}_J) \approx$ a biorthogonal basis

# Construction of $\widetilde{\Psi}$

- $\omega_1, \ldots, \omega_m$: $m$ i.i.d. copies of a random variable $\omega$

$$\omega \sim \mathbb{P}(\omega = j) = p_j \quad j = 1, \ldots, n$$

- Construction of $A$ and $\widetilde{A}$:

$$(k\text{th row of } A) = (\omega_k\text{th row of DFT}) \times \frac{1}{\sqrt{m}}$$

$$(k\text{th row of } \widetilde{A}) = (\omega_k\text{th row of DFT}) \times \frac{1}{\sqrt{m}} \times \frac{1}{np_{\omega_k}}$$

$\Rightarrow \mathbb{E}\widetilde{A}^* A = I_n$

- Construction of $\widetilde{D}$: $\widetilde{D} = D^\dagger = D(D^* D)^{-1}$
- Construction of $\widetilde{\Psi}$: $\widetilde{\Psi} = \widetilde{A}\widetilde{D}$
$\Rightarrow \mathbb{E}\widetilde{\Psi}^* \Psi = \mathbb{E}\widetilde{D}^* \widetilde{A}^* A D = I_n$

# RIP of $\Psi$: Anisotropic Case

- Incoherence: $\max_{j,k} |(AD)_{j,k}| \leq K$
- $\delta_s(\Psi) < \delta + \rho_0$ holds with probability $1 - \eta$ provided that

$$m \geq C_1(1+\rho_0)^2 K^2 \delta^{-2} s (\ln s)^2 \ln n \ln m$$
$$m \geq C_2 K^2 \delta^{-2} s \ln(\eta^{-1})$$

  for universal constants $C_1$ and $C_2$

- $\rho_0 = 0$ in the isotropic case ($D^* D = I_n$ and $p_j = \frac{1}{n}, \forall j$ uniform distr.)
- Penalty for anisotropy: constant $\rho_0$ increases as $(p_j)_{j=1}^n$ and $D$ deviate from the ideal case (uniform distr. & orthogonal matrix)
- $\rho_0$ does not vanish by increasing $m$
  Larger $\rho_0 \Rightarrow$ conservative upper bound on RIC (no guarantee) ☹

# RBOP of $(\Psi, \widetilde{\Psi})$

- $\theta_s(\widetilde{\Psi}^*\Psi) < \delta$ hold with probability $1 - \eta$ provided that

$$m \geq C_1(1+\rho_1)^2(\rho_2 K)^2\delta^{-2}s(\ln s)^2 \ln n \ln m,$$
$$m \geq C_2(\rho_2 K)^2\delta^{-2}s \ln(\eta^{-1})$$

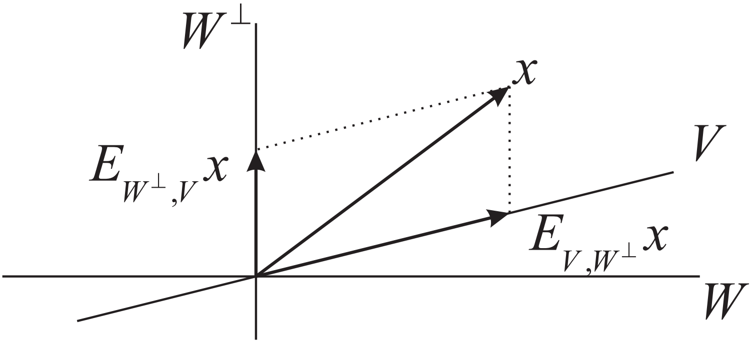
  for universal constants $C_1$ and $C_2$

- Penalty for anisotropy: constants $\rho_1$ and $\rho_2$ increase as $(p_j)_{j=1}^n$ and $D$ deviate from the ideal case (uniform distribution & orthogonal matrix)

- $p_j > 0, \forall j$ satisfied $\leftrightarrow \mathbb{E}\Psi^*\Psi$ has full rank (mild condition)

☺ Still provide good upper bound (of just $\delta$) on RBOC without the "isotropy" property

# Oblique Projection



# Building Blocks for Greedy Algorithms

- Greedy pursuit algorithms consist of common building blocks:
  1) Thresholding of correlation
  2) Least-squares
  3) Orthogonal matching
- RIP guarantees that these blocks work well so that the algorithms converge to the desired estimate
- We modify these blocks using $\widetilde{\Psi}$ so that they are guaranteed by RBOP of $(\Psi, \widetilde{\Psi})$

# Building Blocks for Greedy Algorithms 1

- Conventional: matching to $\Psi$ (thresholding of correlation)
  - input: $\Psi x$ for $s$-sparse $x$
  - output: estimated support of $x$
  - find $s$ indices $k$'s that maximize $|(\Psi^* \Psi x)_k|$
  - RIP of $\Psi \implies \Psi^* \Psi x \approx x \implies$ good estimate

- Modified: matching to $\widetilde{\Psi}$ (thresholding of oblique correlation)
  - find $s$ indices $k$'s that maximize $|(\widetilde{\Psi}^* \Psi x)_k|$
  - RBOP of $(\Psi, \widetilde{\Psi}) \implies \widetilde{\Psi}^* \Psi x \approx x \implies$ good estimate

# Building Blocks for Greedy Algorithms 2

- Conventional: approx. by least-squares (pseudo inverse)
  - input: $J$, $\Psi x$ such that $|\mathrm{supp}(x) \cup J| \leq 2s$
  - output: estimate the components of $x$ on $J$
  - LS solution $\widehat{z} = \arg\min_z \|\Psi x - \Psi_{\widehat{J}} z\|_2^2$ satisfies

$$\widehat{z} = x|_J + (\Psi_J^* \Psi_J)^{-1} \Psi_J^* \Psi_{\mathrm{supp}(x)\setminus J} x|_{\mathrm{supp}(x)\setminus J}$$

  - RIP of $\Psi \implies \Psi_J^* \Psi_{\mathrm{supp}(x)\setminus J} \approx 0 \implies$ small error

- Modified: approx. by weighted LS (oblique inverse)
  - WLS solution $\widehat{z} = \arg\min_z \|\widetilde{\Psi}_J^* (\Psi x - \Psi_J z)\|_2^2$ satisfies

$$\widehat{z} = x|_J + (\widetilde{\Psi}_J^* \Psi_J)^{-1} \widetilde{\Psi}_J^* \Psi_{\mathrm{supp}(x)\setminus J} x|_{\mathrm{supp}(x)\setminus J}$$

  - RBOP of $(\Psi, \widetilde{\Psi}) \implies \widetilde{\Psi}_J^* \Psi_{\mathrm{supp}(x)\setminus J} \approx 0 \implies$ small error

# Building Blocks for Greedy Algorithms 3

- Conventional: orthogonal matching
  - input: $J$, $\Psi x$ such that $|\mathrm{supp}(x) \cup J| \leq 2s$
  - output: estimate support elements of $x$ outside $J$
  - find $s$ indices $k$'s that maximize $|(\Psi^* P_{\mathcal{R}(\Psi_J)^\perp} \Psi x)_k|$
  - RIP of $\Psi \implies \Psi^* P_{\mathcal{R}(\Psi_J)^\perp} \Psi x \approx x \Rightarrow$ good estimate

- Modified: oblique matching
  - find $s$ indices $k$'s that maximize $|(\widetilde{\Psi}^* E_{\mathcal{R}(\widetilde{\Psi}_J)^\perp, \mathcal{R}(\Psi_J)^\perp} \Psi x)_k|$
  - RBOP of $(\Psi, \widetilde{\Psi}) \implies \widetilde{\Psi}^* E_{\mathcal{R}(\widetilde{\Psi}_J)^\perp, \mathcal{R}(\Psi_J)^\perp} \Psi x \approx x \Rightarrow$ good estimate

# Oblique Pursuits

- Each building block has been modified.
- It remains to assemble the blocks.
- Subspace Pursuit (SP) $\rightarrow$ Oblique Subspace Pursuit (ObSP)
- CoSaMP $\rightarrow$ ObCoSaMP
- IHT $\rightarrow$ ObIHT
- HTP $\rightarrow$ ObHTP
- OMP $\rightarrow$ Oblique Matching Pursuit (ObMP)

$$\vdots$$

# RBOP-based Guarantees

- Oblique pursuits extend the RIP-based guarantees of the original counterparts to RBOP-based guarantees.
- $Dx^\star$: best $s$-sparse approximation of $f$ over $D$.
- Let ALG $\in \{\text{ObSP}, \text{ObCoSaMP}, \text{ObIHT}, \text{ObHTP}\}$. If $(AD, \widetilde{A}\widetilde{D})$ satisfies the RBOP, then ALG provides $\widehat{x}$ such that

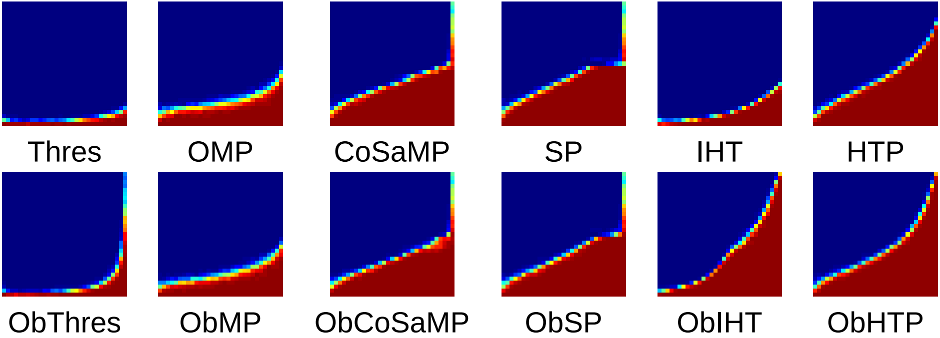
$$\|f - D\widehat{x}\|_2 \leq C \left( \epsilon + \frac{\|f - Dx^\star\|_1}{\sqrt{s}} \right)$$

| ObSP | ObCoSaMP | ObIHT | ObHTP |
|---|---|---|---|
| $\theta_{3s} < 0.325$ | $\theta_{4s} < 0.38$ | $\theta_{3s} < 0.5$ | $\theta_{3s} < 0.58$ |

- With exact arithmetic, ObCoSaMP, ObSP, and ObHTP converge in finite iterations (at most $O(s)$, possibly $O(\ln s)$).

# 1D CS with Biorthogonal Basis

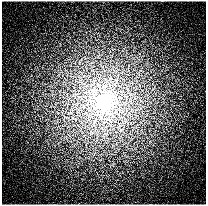
- $A \in \mathbb{C}^{m \times n}$: generic, $\kappa(\mathbb{E}A^*A) = 2$ (anisotropic)
- $\widetilde{A} \in \mathbb{C}^{m \times n}$ satisfies $\mathbb{E}\widetilde{A}^*A = I_n$
- $f$ is sparse in the std. basis ($D = I_n$), nonzeros are random $\pm 1$
- Support recovery rate was observed ($x$-axis: $m/n$, $y$-axis: $s/n$)



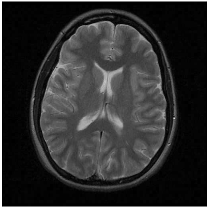

- Oblique pursuits perform competitively with, or sometime better than the conventional counterparts.

# 2D CS Fourier Imaging

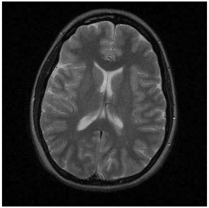
- Sensing matrix $A$: partial Fourier with nonuniform sampling
- Basis $D$: learned dictionary that applies to non-overlapping $8 \times 8$ patches (image size: $512 \times 512$)
- Isotropy property is not satisfied (variable density sampling)
- $D$ is incoherent to DFT with $K = 7.6$
- Input image: exactly $s$-sparse phantom $\frac{s}{n} = \frac{1}{8}$
- Downsample by 3, error measured by PSNR



Sampling Mask

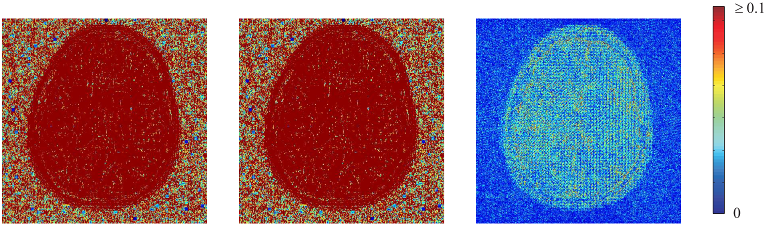


Input

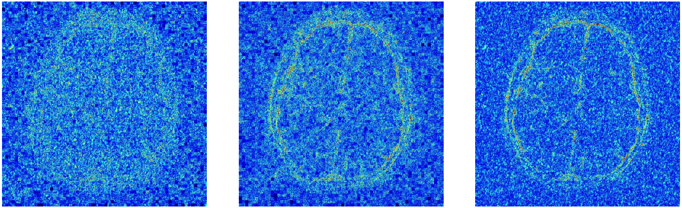


ObSP (36.26 dB)

# Error Images

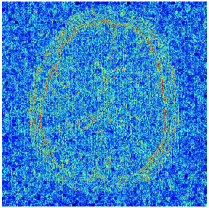



Thres 9.34 dB | IHT 9.34 dB | $\ell_1$ Analysis 29.75 dB

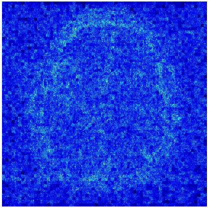


ObThres 31.01 dB | ObIHT 30.95 dB | Zero Filling 31.46 dB

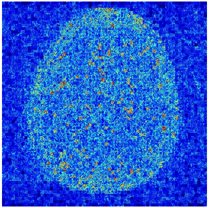
# Error Images



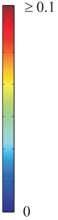
CoSaMP
29.38 dB

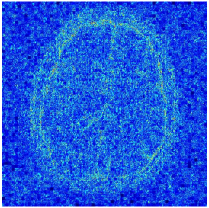


SP
34.58 dB

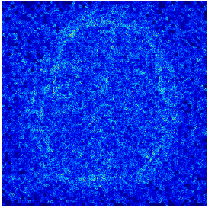


HTP
31.02 dB

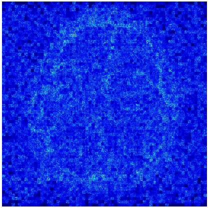


ObCoSaMP
32.27 dB



ObSP
36.26 dB



ObHTP
36.40 dB

# General Framework

- $(\phi_\omega)_{\omega\in\Omega}$: (continuous, not necessarily tight) frame for $\mathbb{C}^n$
- $(\widetilde{\phi}_\omega)_{\omega\in\Omega}$: oblique dual frame
- $\omega_1, \ldots, \omega_m$: $m$ i.i.d. copies of a random variable $\omega$ following a distribution $\nu$ on $\Omega$
- Construction of $A$ and $\widetilde{A}$:

$$(k\text{th row of } A) = \phi_{\omega_k}^T \times \frac{1}{\sqrt{m}}$$

$$(k\text{th row of } \widetilde{A}) = \widetilde{\phi}_{\omega_k}^T \times \frac{1}{\sqrt{m}} \times \frac{1}{np_{\omega_k}}$$

- $(D, \widetilde{D})$ satisfies RBOP (e.g. $\widetilde{D} = D$ if $D$ satisfies RIP)
- $\widetilde{\Psi} = \widetilde{A}\widetilde{D}$

# Discussion

- Schnass & Vandergheynst 2008
  - Modified thresholding and (O)MP using $\widetilde{\Psi}^*$ instead of $\Psi^*$
  - Heuristic design of $\widetilde{\Psi}$
  - Design criterion: guarantee by Babel function (more conservative than RIP)
- "RIPless" analysis of basis pursuit [Kueng & Gross 2012]: noiseless case
- Iterative greedy algorithms Vs. $\ell_1$
  - Guarantees of $\ell_1$ formulations are based on optimality conditions (KKT or dual certificate) and are not explicitly related to any specific algorithm.
  - Guarantees of greedy algorithms are based on the success of iterations of algorithms (Modification of algorithms were necessary).

# Summary – Part 3

- The isotropy property of $\Psi = AD$ is often violated in practice $\implies$ RIP-based guarantees for CS algorithms break down ☹

- New *oblique pursuits* replace "orthogonal-type" building blocks in existing greedy CS recovery algorithms by "oblique-type" or "bio-orthogonal-type" blocks.

- Oblique pursuits are guaranteed by a new Restricted Biorthogonality Property (RBOP), replacing RIP.

  $\Rightarrow$ Guarantees apply to the anisotropic case. ☺

- Oblique pursuit also improve empirical performance over conventional pursuits. ☺

# Conclusion

- Spectrum Blind Sampling and Image Compression on the Fly (Bresler, Gastpar, Feng, Venkataramani) - compressed sensing in the 1990’s.
- The ideas first proposed for SBS and Image Compression on the Fly lead to new, best in class guaranteed algorithms for jointly-sparse recovery.
- A new set of theoretical and algorithmic tools- oblique pursuits - overcome limitations of current methods for compressive sensing in physical imaging systems.