Sparsity in Astrophysics

Astrostatistics meets Astroinformatics

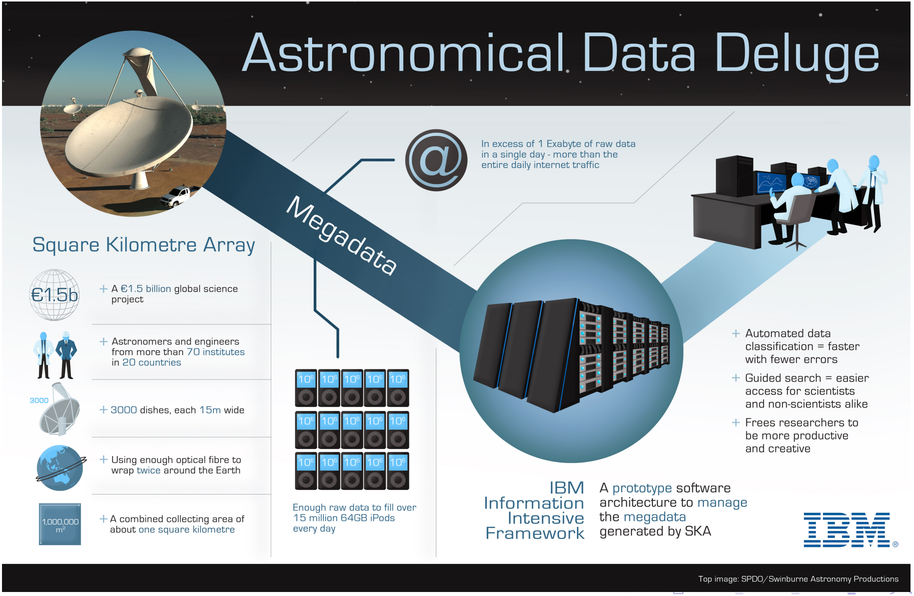
Jason McEwen
www.jasonmcewen.org
@jasonmcewen

Mullard Space Science Laboratory (MSSL)
University College London (UCL)

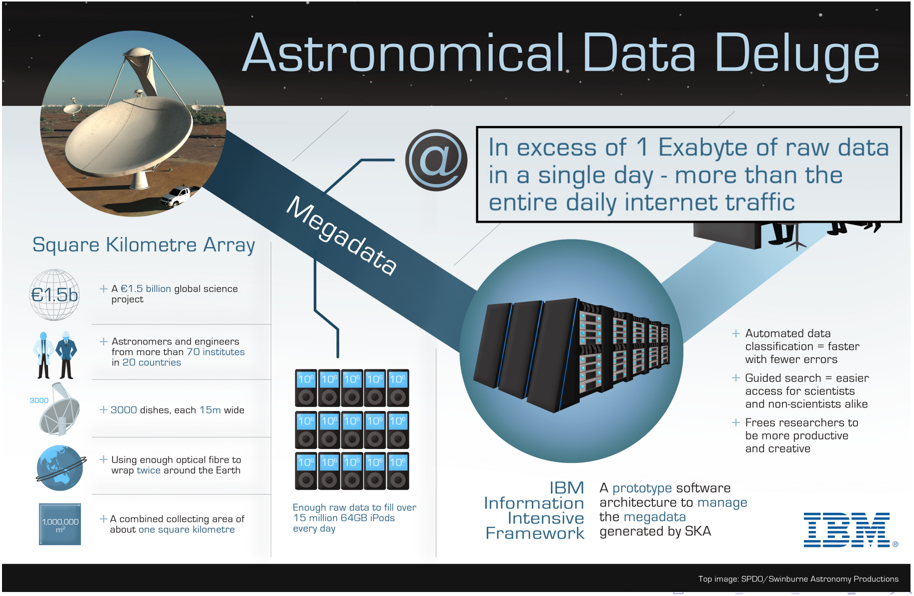
SuSTaIn EdgeCutter Workshop on Astrostatistics
Royal Statistical Society, London, December 2014

SKA movie

The SKA poses a considerable big-data challenge



The SKA poses a considerable big-data challenge



Outline

1. Sparsity
   - What is sparsity?
   - Why is sparsity useful?
   - How can we construct sparsifying transforms?

2. Compressive Sensing
   - Introduction
   - Analysis vs synthesis
   - Bayesian interpretations

3. Radio Interferometric Imaging
   - Interferometric imaging
   - Spread spectrum

Outline

1. Sparsity
   - What is sparsity?
   - Why is sparsity useful?
   - How can we construct sparsifying transforms?

2. Compressive Sensing
   - Introduction
   - Analysis vs synthesis
   - Bayesian interpretations

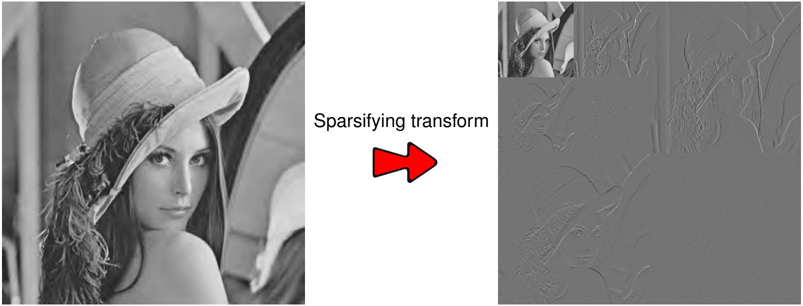
3. Radio Interferometric Imaging
   - Interferometric imaging
   - Spread spectrum

What is sparsity?

— representation of data in such a way that many data points are zero

What is sparsity?

What is sparsity?

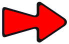


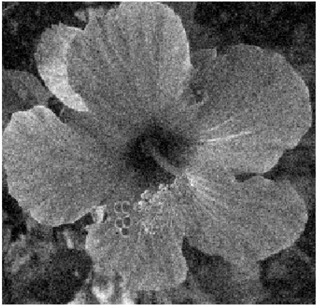
Why is sparsity useful?

— efficient characterisation of structure

Why is sparsity useful?

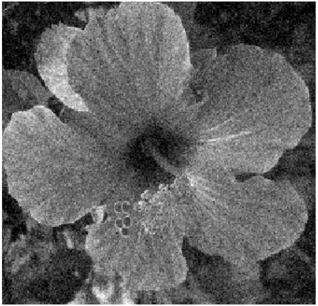
Add noise



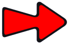


[Credit: http://www.ceremade.dauphine.fr/~peyre/numerical-tour/tours/denoisingwav_2_wavelet_2d/]

Why is sparsity useful?

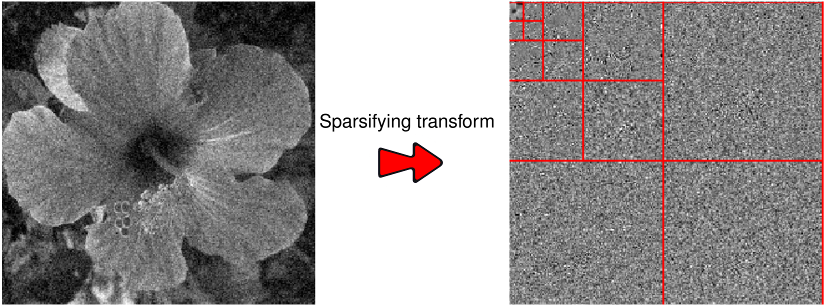


Sparsifying transform



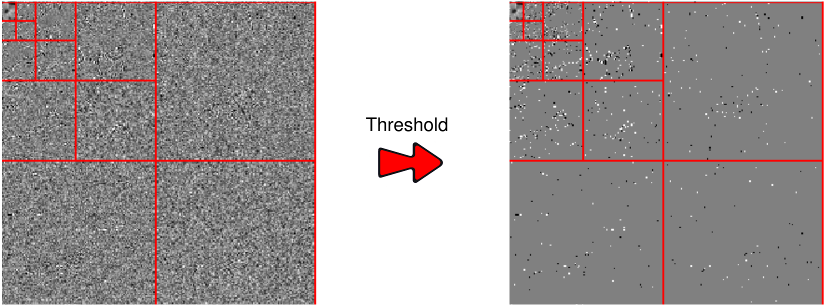
[Credit: http://www.ceremade.dauphine.fr/~peyre/numerical-tour/tours/denoisingwav_2_wavelet_2d/]

Why is sparsity useful?



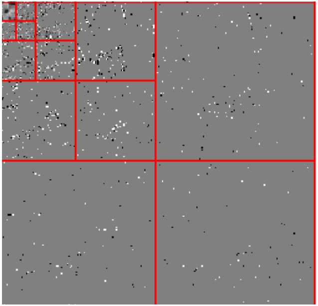
[Credit: `http://www.ceremade.dauphine.fr/~peyre/numerical-tour/tours/denoisingwav_2_wavelet_2d/`]

Why is sparsity useful?



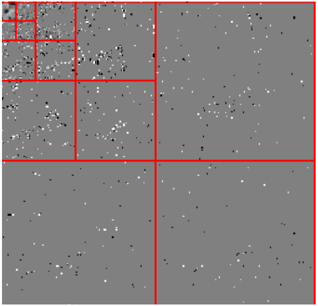
[Credit: http://www.ceremade.dauphine.fr/~peyre/numerical-tour/tours/denoisingwav_2_wavelet_2d/]

Why is sparsity useful?

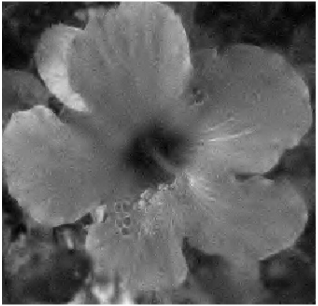


[Credit: http://www.ceremade.dauphine.fr/~peyre/numerical-tour/tours/denoisingwav_2_wavelet_2d/]

Why is sparsity useful?



Inverse transform



[Credit: http://www.ceremade.dauphine.fr/~peyre/numerical-tour/tours/denoisingwav_2_wavelet_2d/]

How can we construct sparsifying transforms?

— many signals in nature have spatially localised, scale-dependent features

How can we construct sparsifying transforms?

Fourier (1807)

Haar (1909)

Morlet and Grossman (1981)

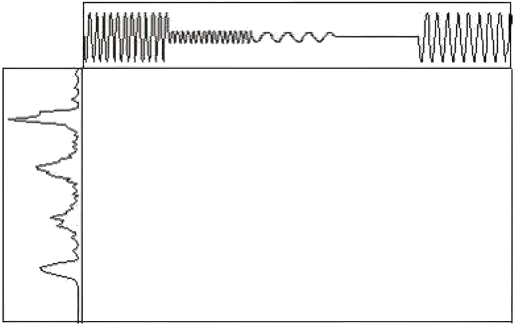


Figure: Fourier vs wavelet transform [Credit: http://www.wavelet.org/tutorial/]

How can we construct sparsifying transforms?

Fourier (1807)

Haar (1909)

Morlet and Grossman (1981)

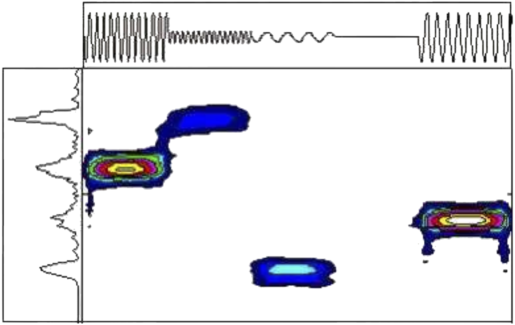


Figure: Fourier vs wavelet transform [Credit: `http://www.wavelet.org/tutorial/`]

How can we construct sparsifying transforms?

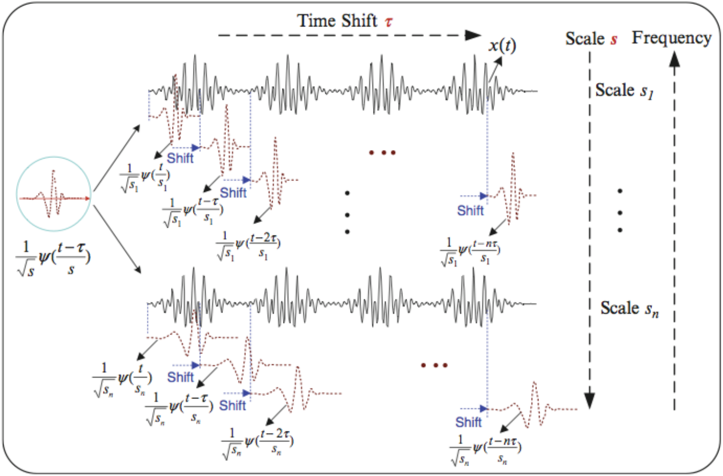


Figure: Wavelet scaling and shifting [Credit: Gao & Yan (2010)]

Outline

1. Sparsity
   - What is sparsity?
   - Why is sparsity useful?
   - How can we construct sparsifying transforms?

2. Compressive Sensing
   - Introduction
   - Analysis vs synthesis
   - Bayesian interpretations

3. Radio Interferometric Imaging
   - Interferometric imaging
   - Spread spectrum

Compressive sensing

"Nothing short of revolutionary."

– National Science Foundation

- Developed by Candes *et al.* 2006 and Donoho 2006 (and others).
- Although many underlying ideas around for a long time.

(a) Emmanuel Candes

(b) David Donoho

Compressive sensing

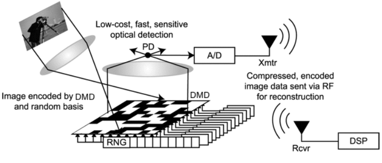
- Next evolution of wavelet analysis $\rightarrow$ wavelets are a key ingredient.
- Mystery of JPEG compression (discrete cosine transform; wavelet transform).
- Move compression to the acquisition stage $\rightarrow$ compressive sensing.
- Acquisition versus imaging.

Compressive sensing

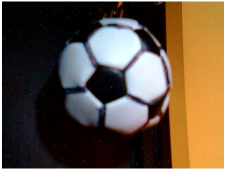
- Next evolution of wavelet analysis $\rightarrow$ wavelets are a key ingredient.
- Mystery of JPEG compression (discrete cosine transform; wavelet transform).
- Move compression to the acquisition stage $\rightarrow$ compressive sensing.
- Acquisition versus imaging.

Compressive sensing

- Next evolution of wavelet analysis $\rightarrow$ wavelets are a key ingredient.
- Mystery of JPEG compression (discrete cosine transform; wavelet transform).
- Move compression to the acquisition stage $\rightarrow$ compressive sensing.
- Acquisition versus imaging.



(a) Architecture



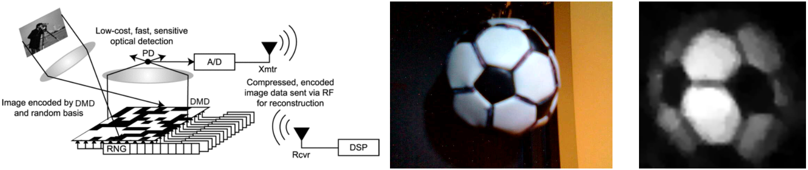
(b) Scene

(c) Recov. (20% meas.)

Figure: Single pixel camera

Compressive sensing

- Next evolution of wavelet analysis $\rightarrow$ wavelets are a key ingredient.
- Mystery of JPEG compression (discrete cosine transform; wavelet transform).
- Move compression to the acquisition stage $\rightarrow$ compressive sensing.
- Acquisition versus imaging.



(a) Architecture (b) Scene (c) Recov. (20% meas.)

Figure: Single pixel camera

An introduction to compressive sensing

Operator description

- Linear operator (linear algebra) representation of signal decomposition:

$$x(t) = \sum_i \alpha_i \Psi_i(t) \quad \rightarrow \quad \boldsymbol{x} = \sum_i \Psi_i \alpha_i = \begin{pmatrix} | \\ \Psi_0 \\ | \end{pmatrix} \alpha_0 + \begin{pmatrix} | \\ \Psi_1 \\ | \end{pmatrix} \alpha_1 + \cdots \quad \rightarrow \quad \boxed{\boldsymbol{x} = \Psi \boldsymbol{\alpha}}$$

- Linear operator (linear algebra) representation of measurement:

$$y_i = \langle x, \Phi_j \rangle \quad \rightarrow \quad \boldsymbol{y} = \begin{pmatrix} - \Phi_0 - \\ - \Phi_1 - \\ \vdots \end{pmatrix} \boldsymbol{x} \quad \rightarrow \quad \boxed{\boldsymbol{y} = \Phi \boldsymbol{x}}$$

- Putting it together: $\boxed{\boldsymbol{y} = \Phi \boldsymbol{x} = \Phi \Psi \boldsymbol{\alpha}}$

An introduction to compressive sensing

Operator description

- Linear operator (linear algebra) representation of signal decomposition:

$$x(t) = \sum_i \alpha_i \Psi_i(t) \quad \rightarrow \quad \boldsymbol{x} = \sum_i \Psi_i \alpha_i = \begin{pmatrix} | \\ \Psi_0 \\ | \end{pmatrix} \alpha_0 + \begin{pmatrix} | \\ \Psi_1 \\ | \end{pmatrix} \alpha_1 + \cdots \quad \rightarrow \quad \boxed{\boldsymbol{x} = \Psi \boldsymbol{\alpha}}$$

- Linear operator (linear algebra) representation of measurement:

$$y_i = \langle x, \Phi_j \rangle \quad \rightarrow \quad \boldsymbol{y} = \begin{pmatrix} - \; \Phi_0 \; - \\ - \; \Phi_1 \; - \\ \vdots \end{pmatrix} \boldsymbol{x} \quad \rightarrow \quad \boxed{\boldsymbol{y} = \Phi \boldsymbol{x}}$$

- Putting it together: $\boxed{\boldsymbol{y} = \Phi \boldsymbol{x} = \Phi \Psi \boldsymbol{\alpha}}$

An introduction to compressive sensing

Operator description

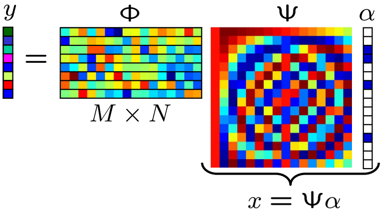
- Linear operator (linear algebra) representation of signal decomposition:

$$x(t) = \sum_i \alpha_i \Psi_i(t) \quad \rightarrow \quad \boldsymbol{x} = \sum_i \Psi_i \alpha_i = \begin{pmatrix} | \\ \Psi_0 \\ | \end{pmatrix} \alpha_0 + \begin{pmatrix} | \\ \Psi_1 \\ | \end{pmatrix} \alpha_1 + \cdots \quad \rightarrow \quad \boxed{\boldsymbol{x} = \Psi \boldsymbol{\alpha}}$$

- Linear operator (linear algebra) representation of measurement:

$$y_i = \langle x, \Phi_j \rangle \quad \rightarrow \quad \boldsymbol{y} = \begin{pmatrix} - \ \Phi_0 \ - \\ - \ \Phi_1 \ - \\ \vdots \end{pmatrix} \boldsymbol{x} \quad \rightarrow \quad \boxed{\boldsymbol{y} = \Phi \boldsymbol{x}}$$

- Putting it together: $\boxed{\boldsymbol{y} = \Phi \boldsymbol{x} = \Phi \Psi \boldsymbol{\alpha}}$



An introduction to compressive sensing

Promoting sparsity via ℓ_1 minimisation

- Ill-posed inverse problem:

$$\boldsymbol{y} = \Phi \boldsymbol{x} + \boldsymbol{n} = \Phi \Psi \boldsymbol{\alpha} + \boldsymbol{n} .$$

- Recall norms given by:

$$\|\alpha\|_0 = \text{no. non-zero elements} \qquad \|\alpha\|_1 = \sum_i |\alpha_i| \qquad \|\alpha\|_2 = \Big(\sum_i |\alpha_i|^2\Big)^{1/2}$$

- Solve by imposing a regularising prior that the signal to be recovered is sparse in Ψ, *i.e.* solve the following ℓ_0 optimisation problem:

$$\boldsymbol{\alpha}^\star = \arg\min_{\boldsymbol{\alpha}} \|\alpha\|_0 \text{ such that } \|\boldsymbol{y} - \Phi\Psi\boldsymbol{\alpha}\|_2 \leq \epsilon ,$$

where the signal is synthesising by $\boldsymbol{x}^\star = \Psi \boldsymbol{\alpha}^\star$.

- Solving this problem is difficult (combinatorial).
- Instead, solve the ℓ_1 optimisation problem (convex):

$$\boldsymbol{\alpha}^\star = \arg\min_{\boldsymbol{\alpha}} \|\alpha\|_1 \text{ such that } \|\boldsymbol{y} - \Phi\Psi\boldsymbol{\alpha}\|_2 \leq \epsilon .$$

An introduction to compressive sensing

Promoting sparsity via ℓ_1 minimisation

- Ill-posed inverse problem:

$$\boldsymbol{y} = \Phi \boldsymbol{x} + \boldsymbol{n} = \Phi\Psi\boldsymbol{\alpha} + \boldsymbol{n} .$$

- Recall norms given by:

$$\|\alpha\|_0 = \text{no. non-zero elements} \qquad \|\alpha\|_1 = \sum_i |\alpha_i| \qquad \|\alpha\|_2 = \Big(\sum_i |\alpha_i|^2\Big)^{1/2}$$

- Solve by imposing a regularising prior that the signal to be recovered is sparse in Ψ, *i.e.* solve the following ℓ_0 optimisation problem:

$$\boldsymbol{\alpha}^\star = \arg\min_{\boldsymbol{\alpha}} \|\alpha\|_0 \text{ such that } \|\boldsymbol{y} - \Phi\Psi\boldsymbol{\alpha}\|_2 \leq \epsilon ,$$

where the signal is synthesising by $\boldsymbol{x}^\star = \Psi\boldsymbol{\alpha}^\star$.

- Solving this problem is difficult (combinatorial).

- Instead, solve the ℓ_1 optimisation problem (convex):

$$\boldsymbol{\alpha}^\star = \arg\min_{\boldsymbol{\alpha}} \|\alpha\|_1 \text{ such that } \|\boldsymbol{y} - \Phi\Psi\boldsymbol{\alpha}\|_2 \leq \epsilon .$$

An introduction to compressive sensing

Promoting sparsity via ℓ_1 minimisation

- Ill-posed inverse problem:

$$\boxed{\boldsymbol{y} = \Phi \boldsymbol{x} + \boldsymbol{n} = \Phi \Psi \boldsymbol{\alpha} + \boldsymbol{n}} .$$

- Recall norms given by:

$$\|\alpha\|_0 = \text{no. non-zero elements} \qquad \|\alpha\|_1 = \sum_i |\alpha_i| \qquad \|\alpha\|_2 = \Big(\sum_i |\alpha_i|^2\Big)^{1/2}$$

- Solve by imposing a regularising prior that the signal to be recovered is sparse in Ψ, *i.e.* solve the following ℓ_0 optimisation problem:

$$\boxed{\boldsymbol{\alpha}^\star = \underset{\boldsymbol{\alpha}}{\arg\min} \|\alpha\|_0 \ \text{ such that } \ \|\boldsymbol{y} - \Phi\Psi\boldsymbol{\alpha}\|_2 \leq \epsilon} ,$$

  where the signal is synthesising by $\boldsymbol{x}^\star = \Psi \boldsymbol{\alpha}^\star$.

- Solving this problem is difficult (combinatorial).
- Instead, solve the ℓ_1 optimisation problem (convex):

$$\boxed{\boldsymbol{\alpha}^\star = \underset{\boldsymbol{\alpha}}{\arg\min} \|\alpha\|_1 \ \text{ such that } \ \|\boldsymbol{y} - \Phi\Psi\boldsymbol{\alpha}\|_2 \leq \epsilon} .$$

An introduction to compressive sensing

Promoting sparsity via ℓ_1 minimisation

- Ill-posed inverse problem:

$$\boldsymbol{y} = \Phi \boldsymbol{x} + \boldsymbol{n} = \Phi \Psi \boldsymbol{\alpha} + \boldsymbol{n} .$$

- Recall norms given by:

$$\|\alpha\|_0 = \text{no. non-zero elements} \qquad \|\alpha\|_1 = \sum_i |\alpha_i| \qquad \|\alpha\|_2 = \Big(\sum_i |\alpha_i|^2\Big)^{1/2}$$

- Solve by imposing a regularising prior that the signal to be recovered is sparse in Ψ, *i.e.* solve the following ℓ_0 optimisation problem:

$$\boldsymbol{\alpha}^\star = \arg\min_{\boldsymbol{\alpha}} \|\alpha\|_0 \ \text{ such that } \ \|\boldsymbol{y} - \Phi\Psi\boldsymbol{\alpha}\|_2 \leq \epsilon ,$$

  where the signal is synthesising by $\boldsymbol{x}^\star = \Psi \boldsymbol{\alpha}^\star$.

- Solving this problem is difficult (combinatorial).

- Instead, solve the ℓ_1 optimisation problem (convex):

$$\boldsymbol{\alpha}^\star = \arg\min_{\boldsymbol{\alpha}} \|\alpha\|_1 \ \text{ such that } \ \|\boldsymbol{y} - \Phi\Psi\boldsymbol{\alpha}\|_2 \leq \epsilon .$$

An introduction to compressive sensing

Promoting sparsity via ℓ_1 minimisation

- Ill-posed inverse problem:

$$\boldsymbol{y} = \Phi \boldsymbol{x} + \boldsymbol{n} = \Phi \Psi \boldsymbol{\alpha} + \boldsymbol{n} .$$

- Recall norms given by:

$$\|\alpha\|_0 = \text{no. non-zero elements} \qquad \|\alpha\|_1 = \sum_i |\alpha_i| \qquad \|\alpha\|_2 = \Big(\sum_i |\alpha_i|^2\Big)^{1/2}$$

- Solve by imposing a regularising prior that the signal to be recovered is sparse in Ψ, *i.e.* solve the following ℓ_0 optimisation problem:

$$\boldsymbol{\alpha}^\star = \arg\min_{\boldsymbol{\alpha}} \|\alpha\|_0 \text{ such that } \|\boldsymbol{y} - \Phi\Psi\boldsymbol{\alpha}\|_2 \leq \epsilon ,$$

where the signal is synthesising by $\boldsymbol{x}^\star = \Psi \boldsymbol{\alpha}^\star$.

- Solving this problem is difficult (combinatorial).
- Instead, solve the ℓ_1 optimisation problem (convex):

$$\boldsymbol{\alpha}^\star = \arg\min_{\boldsymbol{\alpha}} \|\alpha\|_1 \text{ such that } \|\boldsymbol{y} - \Phi\Psi\boldsymbol{\alpha}\|_2 \leq \epsilon .$$

An introduction to compressive sensing
Union of subspaces

- Space of sparse vectors given by the union of subspaces aligned with the coordinate axes.

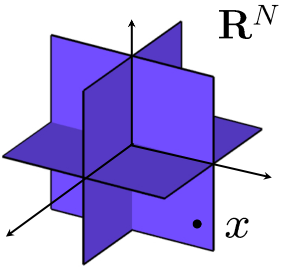


Figure: Space of the sparse vectors [Credit: Baraniuk]

An introduction to compressive sensing

Intuition

- Solutions of ℓ_0 and ℓ_1 problems often the same.
- Geometry of ℓ_0, ℓ_2 and ℓ_1 problems.

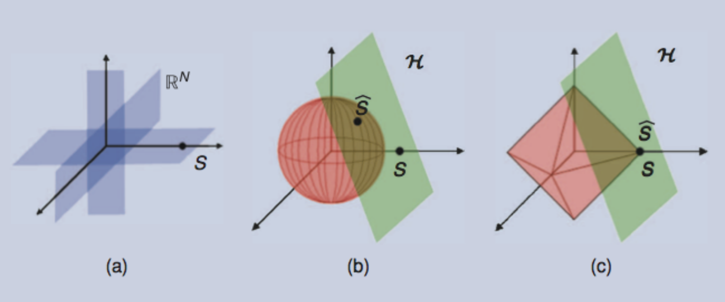


Figure: Geometry of (a) ℓ_0 (b) ℓ_2 and (c) ℓ_1 problems. [Credit: Baraniuk (2007)]

An introduction to compressive sensing

Coherence

- In the absence of noise, compressed sensing is exact!
- Number of measurements required to achieve exact reconstruction is given by

$$M \geq c\mu^2 K \log N ,$$

where K is the sparsity and N the dimensionality.

- The coherence between the measurement and sparsity basis is given by

$$\mu = \sqrt{N} \max_{i,j} |\langle \Psi_i, \Phi_j \rangle| .$$

- Robust to noise.

An introduction to compressive sensing

Coherence

- In the absence of noise, compressed sensing is exact!
- Number of measurements required to achieve exact reconstruction is given by

$$M \geq c\mu^2 K \log N ,$$

where K is the sparsity and N the dimensionality.

- The coherence between the measurement and sparsity basis is given by

$$\mu = \sqrt{N} \max_{i,j} |\langle \Psi_i, \Phi_j \rangle| .$$

- Robust to noise.

An introduction to compressive sensing

Coherence

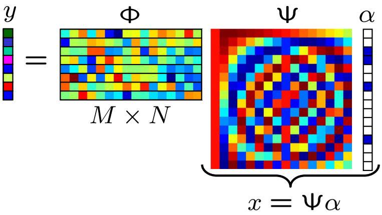
- In the absence of noise, compressed sensing is exact!
- Number of measurements required to achieve exact reconstruction is given by

$$M \geq c\mu^2 K \log N \,,$$

where K is the sparsity and N the dimensionality.

- The coherence between the measurement and sparsity basis is given by

$$\mu = \sqrt{N} \max_{i,j} |\langle \Psi_i, \Phi_j \rangle| \,.$$



- Robust to noise.

An introduction to compressive sensing

Coherence

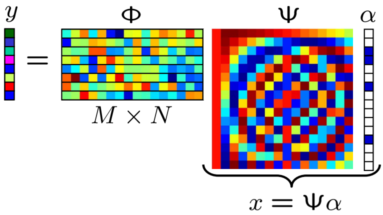
- In the absence of noise, compressed sensing is exact!
- Number of measurements required to achieve exact reconstruction is given by

$$M \geq c\mu^2 K \log N ,$$

where K is the sparsity and N the dimensionality.

- The coherence between the measurement and sparsity basis is given by

$$\mu = \sqrt{N} \max_{i,j} |\langle \Psi_i, \Phi_j \rangle| .$$



- Robust to noise.

Analysis vs synthesis

- Many new developments (*e.g.* analysis vs synthesis, structured sparsity).
- Typically sparsity assumption is justified by analysing example signals in terms of atoms of the dictionary.
- But this is different to synthesising signals from atoms.
- Suggests an analysis-based framework (Elad *et al.* 2007, Nam *et al.* 2012):

$$\boldsymbol{x}^\star = \arg\min_{\boldsymbol{x}} \|\Omega \boldsymbol{x}\|_1 \text{ such that } \|\boldsymbol{y} - \Phi \boldsymbol{x}\|_2 \leq \epsilon \, .$$

analysis

- Contrast with synthesis-based approach:

$$\boldsymbol{x}^\star = \Psi \cdot \arg\min_{\boldsymbol{\alpha}} \|\boldsymbol{\alpha}\|_1 \text{ such that } \|\boldsymbol{y} - \Phi \Psi \boldsymbol{\alpha}\|_2 \leq \epsilon \, .$$

synthesis

- For orthogonal bases $\Omega = \Psi^\dagger$ and the two approaches are identical.

Analysis vs synthesis

- Many new developments (*e.g.* analysis vs synthesis, structured sparsity).
- Typically sparsity assumption is justified by analysing example signals in terms of atoms of the dictionary.
- But this is different to synthesising signals from atoms.
- Suggests an analysis-based framework (Elad *et al.* 2007, Nam *et al.* 2012):

$$\boldsymbol{x}^\star = \arg\min_{\boldsymbol{x}} \|\Omega \boldsymbol{x}\|_1 \ \text{ such that } \ \|\boldsymbol{y} - \Phi \boldsymbol{x}\|_2 \leq \epsilon \, .$$

analysis

- Contrast with synthesis-based approach:

$$\boldsymbol{x}^\star = \Psi \cdot \arg\min_{\boldsymbol{\alpha}} \|\boldsymbol{\alpha}\|_1 \ \text{ such that } \ \|\boldsymbol{y} - \Phi \Psi \boldsymbol{\alpha}\|_2 \leq \epsilon \, .$$

synthesis

- For orthogonal bases $\Omega = \Psi^\dagger$ and the two approaches are identical.

Analysis vs synthesis

- Many new developments (*e.g.* analysis vs synthesis, structured sparsity).
- Typically sparsity assumption is justified by analysing example signals in terms of atoms of the dictionary.
- But this is different to synthesising signals from atoms.
- Suggests an analysis-based framework (Elad *et al.* 2007, Nam *et al.* 2012):

$$\boldsymbol{x}^\star = \arg\min_{\boldsymbol{x}} \|\Omega \boldsymbol{x}\|_1 \text{ such that } \|\boldsymbol{y} - \Phi \boldsymbol{x}\|_2 \leq \epsilon \,.$$

analysis

- Contrast with synthesis-based approach:

$$\boldsymbol{x}^\star = \Psi \cdot \arg\min_{\boldsymbol{\alpha}} \|\boldsymbol{\alpha}\|_1 \text{ such that } \|\boldsymbol{y} - \Phi \Psi \boldsymbol{\alpha}\|_2 \leq \epsilon \,.$$

synthesis

- For orthogonal bases $\Omega = \Psi^\dagger$ and the two approaches are identical.

Analysis vs synthesis

- Many new developments (*e.g.* analysis vs synthesis, structured sparsity).
- Typically sparsity assumption is justified by analysing example signals in terms of atoms of the dictionary.
- But this is different to synthesising signals from atoms.
- Suggests an analysis-based framework (Elad *et al.* 2007, Nam *et al.* 2012):

$$\boldsymbol{x}^\star = \underset{\boldsymbol{x}}{\arg\min} \, \|\Omega \boldsymbol{x}\|_1 \ \text{such that} \ \|\boldsymbol{y} - \Phi \boldsymbol{x}\|_2 \leq \epsilon \, .$$

analysis

- Contrast with synthesis-based approach:

$$\boldsymbol{x}^\star = \Psi \cdot \underset{\boldsymbol{\alpha}}{\arg\min} \, \|\boldsymbol{\alpha}\|_1 \ \text{such that} \ \|\boldsymbol{y} - \Phi \Psi \boldsymbol{\alpha}\|_2 \leq \epsilon \, .$$

synthesis

- For orthogonal bases $\Omega = \Psi^\dagger$ and the two approaches are identical.

Analysis vs synthesis

- Many new developments (*e.g.* analysis vs synthesis, structured sparsity).
- Typically sparsity assumption is justified by analysing example signals in terms of atoms of the dictionary.
- But this is different to synthesising signals from atoms.
- Suggests an analysis-based framework (Elad *et al.* 2007, Nam *et al.* 2012):

$$\boldsymbol{x}^\star = \arg\min_{\boldsymbol{x}} \|\Omega \boldsymbol{x}\|_1 \ \text{such that} \ \|\boldsymbol{y} - \Phi \boldsymbol{x}\|_2 \leq \epsilon \, .$$

analysis

- Contrast with synthesis-based approach:

$$\boldsymbol{x}^\star = \Psi \cdot \arg\min_{\boldsymbol{\alpha}} \|\boldsymbol{\alpha}\|_1 \ \text{such that} \ \|\boldsymbol{y} - \Phi \Psi \boldsymbol{\alpha}\|_2 \leq \epsilon \, .$$

synthesis

- For orthogonal bases $\Omega = \Psi^\dagger$ and the two approaches are identical.

Bayesian interpretations

One Bayesian interpretation of the synthesis-based approach

- Consider the inverse problem:

$$\boldsymbol{y} = \Phi \Psi \boldsymbol{\alpha} + \boldsymbol{n} .$$

- Assume Gaussian noise, yielding the likelihood:

$$\mathrm{P}(\boldsymbol{y} \,|\, \boldsymbol{\alpha}) \propto \exp\Bigl(\| \boldsymbol{y} - \Phi \Psi \boldsymbol{\alpha} \|_2^2 / (2\sigma^2) \Bigr) .$$

- Consider the Laplacian prior:

$$\mathrm{P}(\boldsymbol{\alpha}) \propto \exp\Bigl(-\beta \| \boldsymbol{\alpha} \|_1 \Bigr) .$$

- The maximum *a-posteriori* (MAP) estimate (with $\lambda = 2\beta\sigma^2$) is

$$\boldsymbol{x}^\star_{\text{MAP-Synthesis}} = \Psi \cdot \arg\max_{\boldsymbol{\alpha}} \mathrm{P}(\boldsymbol{\alpha} \,|\, \boldsymbol{y}) = \Psi \cdot \arg\min_{\boldsymbol{\alpha}} \| \boldsymbol{y} - \Phi \Psi \boldsymbol{\alpha} \|_2^2 + \lambda \| \boldsymbol{\alpha} \|_1 .$$

synthesis

- One possible Bayesian interpretation!
- Signal may be ℓ_0-sparse, then solving ℓ_1 problem finds the correct ℓ_0-sparse solution!

Bayesian interpretations

One Bayesian interpretation of the synthesis-based approach

- Consider the inverse problem:

$$\boldsymbol{y} = \Phi \Psi \boldsymbol{\alpha} + \boldsymbol{n} \, .$$

- Assume Gaussian noise, yielding the likelihood:

$$\mathrm{P}(\boldsymbol{y} \,|\, \boldsymbol{\alpha}) \propto \exp\Big(\| \boldsymbol{y} - \Phi \Psi \boldsymbol{\alpha} \|_2^2 / (2\sigma^2) \Big) \, .$$

- Consider the Laplacian prior:

$$\mathrm{P}(\boldsymbol{\alpha}) \propto \exp\Big(-\beta \| \boldsymbol{\alpha} \|_1 \Big) \, .$$

- The maximum *a-posteriori* (MAP) estimate (with $\lambda = 2\beta\sigma^2$) is

$$\boldsymbol{x}^\star_{\text{MAP-Synthesis}} = \Psi \cdot \underset{\boldsymbol{\alpha}}{\arg\max}\, \mathrm{P}(\boldsymbol{\alpha} \,|\, \boldsymbol{y}) = \Psi \cdot \underset{\boldsymbol{\alpha}}{\arg\min}\, \| \boldsymbol{y} - \Phi \Psi \boldsymbol{\alpha} \|_2^2 + \lambda \| \boldsymbol{\alpha} \|_1 \, .$$

synthesis

- One possible Bayesian interpretation!
- Signal may be ℓ_0-sparse, then solving ℓ_1 problem finds the correct ℓ_0-sparse solution!

Bayesian interpretations

One Bayesian interpretation of the synthesis-based approach

- Consider the inverse problem:

$$\boldsymbol{y} = \Phi \Psi \boldsymbol{\alpha} + \boldsymbol{n} \, .$$

- Assume Gaussian noise, yielding the likelihood:

$$\mathrm{P}(\boldsymbol{y} \,|\, \boldsymbol{\alpha}) \propto \exp\Bigl(\| \boldsymbol{y} - \Phi \Psi \boldsymbol{\alpha} \|_2^2 / (2\sigma^2) \Bigr) \, .$$

- Consider the Laplacian prior:

$$\mathrm{P}(\boldsymbol{\alpha}) \propto \exp\Bigl(-\beta \| \boldsymbol{\alpha} \|_1 \Bigr) \, .$$

- The maximum *a-posteriori* (MAP) estimate (with $\lambda = 2\beta\sigma^2$) is

$$\boldsymbol{x}^\star_{\text{MAP-Synthesis}} = \Psi \cdot \underset{\boldsymbol{\alpha}}{\arg\max}\, \mathrm{P}(\boldsymbol{\alpha} \,|\, \boldsymbol{y}) = \Psi \cdot \underset{\boldsymbol{\alpha}}{\arg\min}\, \| \boldsymbol{y} - \Phi \Psi \boldsymbol{\alpha} \|_2^2 + \lambda \| \boldsymbol{\alpha} \|_1 \, .$$

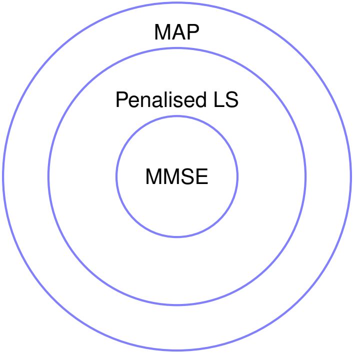
synthesis

- One possible Bayesian interpretation!
- Signal may be ℓ_0-sparse, then solving ℓ_1 problem finds the correct ℓ_0-sparse solution!

Bayesian interpretations

Other Bayesian interpretations of the synthesis-based approach

- Other Bayesian interpretations are also possible (Gribonval 2011).
- Minimum mean square error (MMSE) estimators
  - $\subset$ synthesis-based estimators with appropriate penalty function, *i.e.* penalised least-squares (LS)
  - $\subset$ MAP estimators



Bayesian interpretations

One Bayesian interpretation of the analysis-based approach

- For the analysis-based approach, the MAP estimate is then

$$\boldsymbol{x}^\star_{\text{MAP-Analysis}} = \arg\max_{\boldsymbol{x}} \mathrm{P}(\boldsymbol{x} \mid \boldsymbol{y}) = \arg\min_{\boldsymbol{x}} \|\boldsymbol{y} - \Phi\boldsymbol{x}\|_2^2 + \lambda\|\Omega\boldsymbol{x}\|_1 \; .$$

analysis

- Identical to the synthesis-based approach if $\Omega = \Psi^\dagger$.
- But for redundant dictionaries, the analysis-based MAP estimate is

$$\boldsymbol{x}^\star_{\text{MAP-Analysis}} = \Omega^\dagger \cdot \arg\min_{\boldsymbol{\gamma}\in \text{column space } \Omega} \|\boldsymbol{y} - \Phi\Omega^\dagger\boldsymbol{\gamma}\|_2^2 + \lambda\|\boldsymbol{\gamma}\|_1 \; .$$

analysis

- Analysis-based approach more restrictive than synthesis-based.
- Similar ideas promoted by Maisinger & Hobson (2004) in a Bayesian framework for wavelet MEM (maximum entropy method).

Bayesian interpretations

One Bayesian interpretation of the analysis-based approach

- For the analysis-based approach, the MAP estimate is then

$$\boldsymbol{x}^\star_{\text{MAP-Analysis}} = \arg\max_{\boldsymbol{x}} \mathrm{P}(\boldsymbol{x} \,|\, \boldsymbol{y}) = \arg\min_{\boldsymbol{x}} \|\boldsymbol{y} - \Phi\boldsymbol{x}\|_2^2 + \lambda\|\Omega\boldsymbol{x}\|_1 \; .$$

analysis

- Identical to the synthesis-based approach if $\Omega = \Psi^\dagger$.

- But for redundant dictionaries, the analysis-based MAP estimate is

$$\boldsymbol{x}^\star_{\text{MAP-Analysis}} = \Omega^\dagger \cdot \arg\min_{\boldsymbol{\gamma} \in \text{column space } \Omega} \|\boldsymbol{y} - \Phi\Omega^\dagger\boldsymbol{\gamma}\|_2^2 + \lambda\|\boldsymbol{\gamma}\|_1 \; .$$

analysis

- Analysis-based approach more restrictive than synthesis-based.

- Similar ideas promoted by Maisinger & Hobson (2004) in a Bayesian framework for wavelet MEM (maximum entropy method).

Bayesian interpretations

One Bayesian interpretation of the analysis-based approach

- For the analysis-based approach, the MAP estimate is then

$$x^\star_{\text{MAP-Analysis}} = \arg\max_{x} \text{P}(x \mid y) = \arg\min_{x} \|y - \Phi x\|_2^2 + \lambda \|\Omega x\|_1 \; .$$

analysis

- Identical to the synthesis-based approach if $\Omega = \Psi^\dagger$.
- But for redundant dictionaries, the analysis-based MAP estimate is

$$x^\star_{\text{MAP-Analysis}} = \Omega^\dagger \cdot \arg\min_{\gamma \in \text{column space } \Omega} \|y - \Phi\Omega^\dagger \gamma\|_2^2 + \lambda \|\gamma\|_1 \; .$$

analysis

- Analysis-based approach more restrictive than synthesis-based.
- Similar ideas promoted by Maisinger & Hobson (2004) in a Bayesian framework for wavelet MEM (maximum entropy method).

Bayesian interpretations

One Bayesian interpretation of the analysis-based approach

- For the analysis-based approach, the MAP estimate is then

$$\boldsymbol{x}^{\star}_{\text{MAP-Analysis}} = \arg\max_{\boldsymbol{x}} \mathrm{P}(\boldsymbol{x} \mid \boldsymbol{y}) = \arg\min_{\boldsymbol{x}} \|\boldsymbol{y} - \Phi \boldsymbol{x}\|_2^2 + \lambda \|\Omega \boldsymbol{x}\|_1 \; .$$

analysis

- Identical to the synthesis-based approach if $\Omega = \Psi^\dagger$.
- But for redundant dictionaries, the analysis-based MAP estimate is

$$\boldsymbol{x}^{\star}_{\text{MAP-Analysis}} = \Omega^\dagger \cdot \arg\min_{\boldsymbol{\gamma} \in \text{column space } \Omega} \|\boldsymbol{y} - \Phi \Omega^\dagger \boldsymbol{\gamma}\|_2^2 + \lambda \|\boldsymbol{\gamma}\|_1 \; .$$

analysis

- Analysis-based approach more restrictive than synthesis-based.
- Similar ideas promoted by Maisinger & Hobson (2004) in a Bayesian framework for wavelet MEM (maximum entropy method).

Outline

1. Sparsity
   - What is sparsity?
   - Why is sparsity useful?
   - How can we construct sparsifying transforms?
2. Compressive Sensing
   - Introduction
   - Analysis vs synthesis
   - Bayesian interpretations
3. Radio Interferometric Imaging
   - Interferometric imaging
   - Spread spectrum

Radio interferometric imaging

Inverse problem

- Consider the ill-posed inverse problem of radio interferometric imaging:

$$\boxed{\boldsymbol{y} = \Phi \boldsymbol{x} + \boldsymbol{n}} \,,$$

where $\boldsymbol{y}$ are the measured visibilities, Φ is the linear measurement operator, $\boldsymbol{x}$ is the underlying image and $\boldsymbol{n}$ is instrumental noise.

- Measurement operator $\boxed{\Phi = \mathbf{M}\,\mathbf{F}\,\mathbf{C}\,\mathbf{A}}$ may incorporate:
  - primary beam $\mathbf{A}$ of the telescope;
  - w-modulation modulation $\mathbf{C}$;
  - Fourier transform $\mathbf{F}$;
  - masking $\mathbf{M}$ which encodes the incomplete measurements taken by the interferometer.

Radio interferometric imaging

Inverse problem

- Consider the ill-posed inverse problem of radio interferometric imaging:

$$\boldsymbol{y} = \Phi \boldsymbol{x} + \boldsymbol{n} ,$$

where $\boldsymbol{y}$ are the measured visibilities, Φ is the linear measurement operator, $\boldsymbol{x}$ is the underlying image and $\boldsymbol{n}$ is instrumental noise.

- Measurement operator $\Phi = \mathbf{M}\,\mathbf{F}\,\mathbf{C}\,\mathbf{A}$ may incorporate:
  - primary beam $\mathbf{A}$ of the telescope;
  - w-modulation modulation $\mathbf{C}$;
  - Fourier transform $\mathbf{F}$;
  - masking $\mathbf{M}$ which encodes the incomplete measurements taken by the interferometer.

Radio interferometric imaging

Inverse problem

- Consider the ill-posed inverse problem of radio interferometric imaging:

$$\boldsymbol{y} = \Phi \boldsymbol{x} + \boldsymbol{n} \,,$$

where $\boldsymbol{y}$ are the measured visibilities, Φ is the linear measurement operator, $\boldsymbol{x}$ is the underlying image and $\boldsymbol{n}$ is instrumental noise.

- Measurement operator $\Phi = \mathbf{M}\,\mathbf{F}\,\mathbf{C}\,\mathbf{A}$ may incorporate:
  - primary beam $\mathbf{A}$ of the telescope;
  - w-modulation modulation $\mathbf{C}$;
  - Fourier transform $\mathbf{F}$;
  - masking $\mathbf{M}$ which encodes the incomplete measurements taken by the interferometer.

Interferometric imaging: recover an image from noisy and incomplete Fourier measurements.

Radio interferometric imaging

Imaging

- Solve the interferometric imaging problem

$$\boldsymbol{y} = \Phi \boldsymbol{x} + \boldsymbol{n} \quad \text{with} \quad \Phi = \mathbf{M}\,\mathbf{F}\,\mathbf{C}\,\mathbf{A} ,$$

by applying a prior on sparsity of the signal in a sparsifying dictionary Ψ.

- Basis pursuit (BP) denoising problem

$$\boldsymbol{\alpha}^\star = \underset{\boldsymbol{\alpha}}{\arg\min} \|\boldsymbol{\alpha}\|_1 \;\text{ such that }\; \|\boldsymbol{y} - \Phi\Psi\boldsymbol{\alpha}\|_2 \leq \epsilon ,$$

synthesis

where the image is synthesised by $\boldsymbol{x}^\star = \Psi\boldsymbol{\alpha}^\star$.

- Application to simulations by Wiaux *et al.* 2009, McEwen & Wiaux 2011.
- Application of (unconstrained) problem to LOFAR by Garsden, Starck *et al.* 2014.

Radio interferometric imaging
Imaging

- Solve the interferometric imaging problem

$$\boxed{\boldsymbol{y} = \Phi \boldsymbol{x} + \boldsymbol{n} \quad \text{with} \quad \Phi = \mathbf{M}\,\mathbf{F}\,\mathbf{C}\,\mathbf{A}},$$

by applying a prior on sparsity of the signal in a sparsifying dictionary Ψ.

- Basis pursuit (BP) denoising problem

$$\boxed{\boldsymbol{\alpha}^\star = \underset{\boldsymbol{\alpha}}{\arg\min} \|\alpha\|_1 \ \text{ such that } \ \|\boldsymbol{y} - \Phi\Psi\boldsymbol{\alpha}\|_2 \leq \epsilon\,,}$$

synthesis

where the image is synthesised by $\boldsymbol{x}^\star = \Psi \boldsymbol{\alpha}^\star$.

- Application to simulations by Wiaux *et al.* 2009, McEwen & Wiaux 2011.
- Application of (unconstrained) problem to LOFAR by Garsden, Starck *et al.* 2014.

Radio interferometric imaging

Imaging

- Solve the interferometric imaging problem

$$\boldsymbol{y} = \Phi \boldsymbol{x} + \boldsymbol{n} \quad \text{with} \quad \Phi = \mathbf{M}\,\mathbf{F}\,\mathbf{C}\,\mathbf{A} \,,$$

by applying a prior on sparsity of the signal in a sparsifying dictionary Ψ.

- Basis pursuit (BP) denoising problem

$$\boldsymbol{\alpha}^\star = \arg\min_{\boldsymbol{\alpha}} \|\alpha\|_1 \;\; \text{such that} \;\; \|\boldsymbol{y} - \Phi\Psi\boldsymbol{\alpha}\|_2 \leq \epsilon \,,$$

synthesis

where the image is synthesised by $\boldsymbol{x}^\star = \Psi\boldsymbol{\alpha}^\star$.

- Application to simulations by Wiaux *et al.* 2009, McEwen & Wiaux 2011.
- Application of (unconstrained) problem to LOFAR by Garsden, Starck *et al.* 2014.

SARA for radio interferometric imaging

Algorithm

- Sparsity averaging reweighted analysis (SARA) for RI imaging (Carrillo, McEwen & Wiaux 2013, 2014)
- Consider a dictionary composed of a concatenation of orthonormal bases, i.e.

$$\Psi = \frac{1}{\sqrt{q}}[\Psi_1, \Psi_2, \ldots, \Psi_q],$$

  thus $\Psi \in \mathbb{R}^{N \times D}$ with $D = qN$.

- We consider the following bases: Dirac (*i.e.* pixel basis); Haar wavelets (promotes gradient sparsity); Daubechies wavelet bases two to eight. $\Rightarrow$ concatenation of 9 bases.
- Promote average sparsity by solving the reweighted ℓ_1 analysis problem:

$$\min_{\bar{\boldsymbol{x}} \in \mathbb{R}^N} \|W \Psi^T \bar{\boldsymbol{x}}\|_1 \quad \text{subject to} \quad \|\boldsymbol{y} - \Phi \bar{\boldsymbol{x}}\|_2 \leq \epsilon \quad \text{and} \quad \bar{\boldsymbol{x}} \geq 0 \,, \tag{SARA}$$

  where $W \in \mathbb{R}^{D \times D}$ is a diagonal matrix with positive weights.

- Solve a sequence of reweighted ℓ_1 problems using the solution of the previous problem as the inverse weights $\rightarrow$ approximate the ℓ_0 problem.

SARA for radio interferometric imaging

Algorithm

- Sparsity averaging reweighted analysis (SARA) for RI imaging (Carrillo, McEwen & Wiaux 2013, 2014)

- Consider a dictionary composed of a concatenation of orthonormal bases, i.e.

$$\Psi = \frac{1}{\sqrt{q}}[\Psi_1, \Psi_2, \ldots, \Psi_q],$$

thus $\Psi \in \mathbb{R}^{N \times D}$ with $D = qN$.

- We consider the following bases: Dirac (*i.e.* pixel basis); Haar wavelets (promotes gradient sparsity); Daubechies wavelet bases two to eight. $\Rightarrow$ concatenation of 9 bases.

- Promote average sparsity by solving the reweighted ℓ_1 analysis problem:

$$\min_{\bar{\boldsymbol{x}} \in \mathbb{R}^N} \|W \Psi^T \bar{\boldsymbol{x}}\|_1 \quad \text{subject to} \quad \|\boldsymbol{y} - \Phi \bar{\boldsymbol{x}}\|_2 \leq \epsilon \quad \text{and} \quad \bar{\boldsymbol{x}} \geq 0\,,$$

SARA

where $W \in \mathbb{R}^{D \times D}$ is a diagonal matrix with positive weights.

- Solve a sequence of reweighted ℓ_1 problems using the solution of the previous problem as the inverse weights $\rightarrow$ approximate the ℓ_0 problem.

SARA for radio interferometric imaging

Algorithm

- Sparsity averaging reweighted analysis (SARA) for RI imaging (Carrillo, McEwen & Wiaux 2013, 2014)

- Consider a dictionary composed of a concatenation of orthonormal bases, i.e.

$$\Psi = \frac{1}{\sqrt{q}}[\Psi_1, \Psi_2, \dots, \Psi_q],$$

  thus $\Psi \in \mathbb{R}^{N \times D}$ with $D = qN$.

- We consider the following bases: Dirac (*i.e.* pixel basis); Haar wavelets (promotes gradient sparsity); Daubechies wavelet bases two to eight. $\Rightarrow$ concatenation of 9 bases.

- Promote average sparsity by solving the reweighted ℓ_1 analysis problem:

$$\min_{\bar{\boldsymbol{x}} \in \mathbb{R}^N} \|W \Psi^T \bar{\boldsymbol{x}}\|_1 \quad \text{subject to} \quad \|\boldsymbol{y} - \Phi \bar{\boldsymbol{x}}\|_2 \leq \epsilon \quad \text{and} \quad \bar{\boldsymbol{x}} \geq 0 \,,$$

SARA

  where $W \in \mathbb{R}^{D \times D}$ is a diagonal matrix with positive weights.

- Solve a sequence of reweighted ℓ_1 problems using the solution of the previous problem as the inverse weights $\rightarrow$ approximate the ℓ_0 problem.

SARA for radio interferometric imaging

Algorithm

- Sparsity averaging reweighted analysis (SARA) for RI imaging (Carrillo, McEwen & Wiaux 2013, 2014)

- Consider a dictionary composed of a concatenation of orthonormal bases, i.e.

$$\Psi = \frac{1}{\sqrt{q}}[\Psi_1, \Psi_2, \ldots, \Psi_q],$$

  thus $\Psi \in \mathbb{R}^{N \times D}$ with $D = qN$.

- We consider the following bases: Dirac (*i.e.* pixel basis); Haar wavelets (promotes gradient sparsity); Daubechies wavelet bases two to eight. $\Rightarrow$ concatenation of 9 bases.

- Promote average sparsity by solving the reweighted ℓ_1 analysis problem:

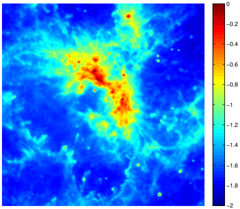
$$\min_{\bar{\boldsymbol{x}} \in \mathbb{R}^N} \|W \Psi^T \bar{\boldsymbol{x}}\|_1 \quad \text{subject to} \quad \|\boldsymbol{y} - \Phi \bar{\boldsymbol{x}}\|_2 \leq \epsilon \quad \text{and} \quad \bar{\boldsymbol{x}} \geq 0\,, \qquad \text{SARA}$$

  where $W \in \mathbb{R}^{D \times D}$ is a diagonal matrix with positive weights.

- Solve a sequence of reweighted ℓ_1 problems using the solution of the previous problem as the inverse weights $\rightarrow$ approximate the ℓ_0 problem.

SARA for radio interferometric imaging

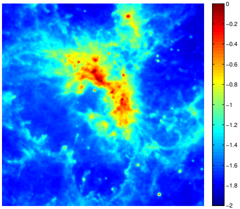
Results on simulations



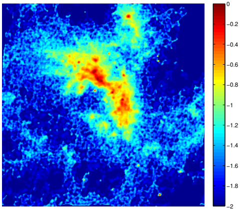
(a) Original

SARA for radio interferometric imaging

Results on simulations



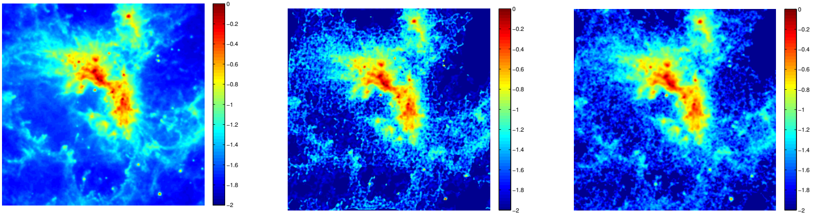
(a) Original



(b) “CLEAN” (SNR=16.67 dB)

SARA for radio interferometric imaging

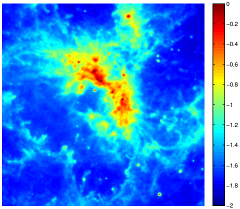
Results on simulations



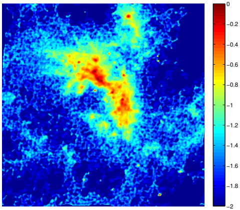
(a) Original (b) “CLEAN” (SNR=16.67 dB) (c) “MS-CLEAN” (SNR=17.87 dB)

SARA for radio interferometric imaging

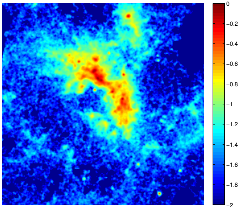
Results on simulations



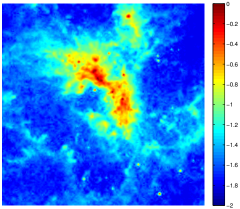
(a) Original



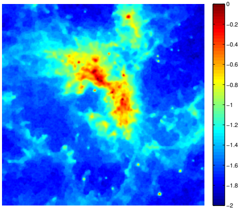
(b) “CLEAN” (SNR=16.67 dB)



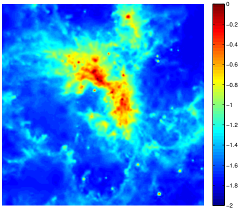
(c) “MS-CLEAN” (SNR=17.87 dB)



(d) BPDb8 (SNR=24.53 dB)



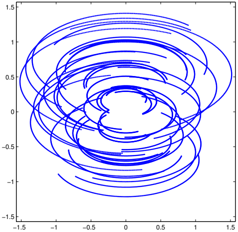
(e) TV (SNR=26.47 dB)



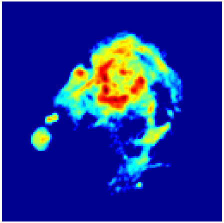
(f) SARA (SNR=29.08 dB)

SARA for radio interferometric imaging

Results on simulations for continuous visiblities



(a) Coverage

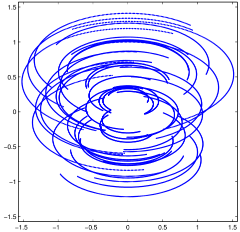


(b) M31 (ground truth)

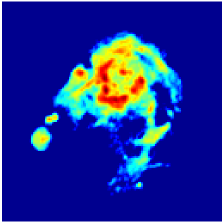
Figure: Reconstructed images from continuous visibilities.

SARA for radio interferometric imaging

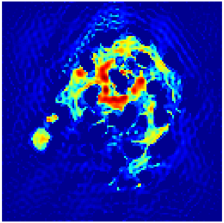
Results on simulations for continuous visiblities



(a) Coverage



(b) M31 (ground truth)

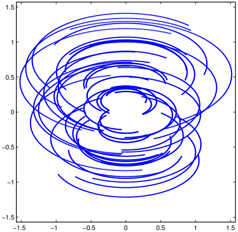


(c) “CLEAN” $\rightarrow$ SNR$= 8.2$dB

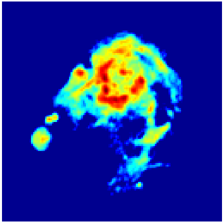
Figure: Reconstructed images from continuous visibilities.

SARA for radio interferometric imaging

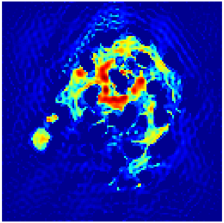
Results on simulations for continuous visiblities



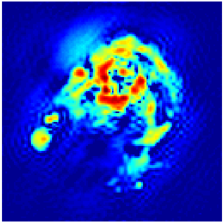
(a) Coverage



(b) M31 (ground truth)



(c) “CLEAN” $\rightarrow$ SNR= 8.2dB

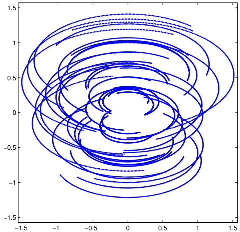


(d) “MS-CLEAN” $\rightarrow$ SNR= 11.1dB

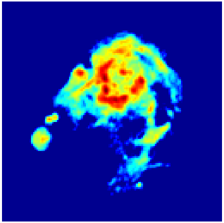
Figure: Reconstructed images from continuous visibilities.

SARA for radio interferometric imaging

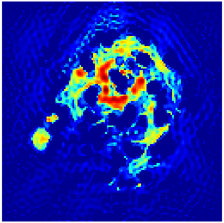
Results on simulations for continuous visiblities



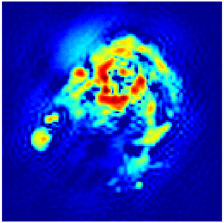
(a) Coverage



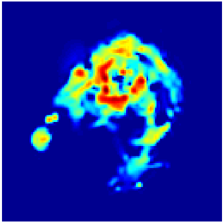
(b) M31 (ground truth)



(c) "CLEAN" $\rightarrow$ SNR$=8.2$dB



(d) "MS-CLEAN" $\rightarrow$ SNR$=11.1$dB



(e) SARA $\rightarrow$ SNR$=13.4$dB

Figure: Reconstructed images from continuous visibilities.

Spread spectrum effect

Review

- Use theory of compressive sensing to optimise telescope configurations.
- Non-coplanar baselines and wide fields $\rightarrow$ w-modulation $\rightarrow$ spread spectrum effect $\rightarrow$ improves reconstruction quality (first considered by Wiaux *et al.* 2009b).
- The w-modulation operator $\mathbf{C}$ has elements defined by

$$C(l,m) \equiv \mathrm{e}^{\mathrm{i}2\pi w\left(1-\sqrt{1-l^2-m^2}\right)} \simeq \mathrm{e}^{\mathrm{i}\pi w\|\boldsymbol{l}\|^2} \quad \text{for} \quad \|\boldsymbol{l}\|^4\, w \ll 1 ,$$

giving rise to to a linear chirp.

Spread spectrum effect

Review

- Use theory of compressive sensing to optimise telescope configurations.
- Non-coplanar baselines and wide fields $\rightarrow$ w-modulation $\rightarrow$ spread spectrum effect $\rightarrow$ improves reconstruction quality (first considered by Wiaux *et al.* 2009b).
- The w-modulation operator $\mathbf{C}$ has elements defined by

$$C(l,m) \equiv \mathrm{e}^{\mathrm{i}2\pi w\left(1-\sqrt{1-l^2-m^2}\right)} \simeq \mathrm{e}^{\mathrm{i}\pi w\|\boldsymbol{l}\|^2} \quad \text{for} \quad \|\boldsymbol{l}\|^4\, w \ll 1 \, ,$$

giving rise to to a linear chirp.

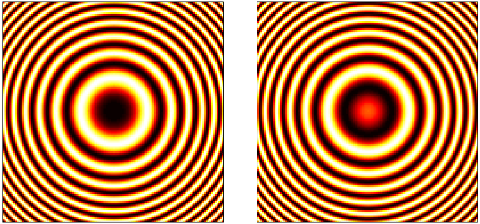
Spread spectrum effect

Review

- Use theory of compressive sensing to optimise telescope configurations.
- Non-coplanar baselines and wide fields $\rightarrow$ w-modulation $\rightarrow$ spread spectrum effect $\rightarrow$ improves reconstruction quality (first considered by Wiaux *et al.* 2009b).
- The w-modulation operator $\mathbf{C}$ has elements defined by

$$C(l,m) \equiv \mathrm{e}^{\mathrm{i}2\pi w\left(1-\sqrt{1-l^2-m^2}\right)} \simeq \mathrm{e}^{\mathrm{i}\pi w \|\boldsymbol{l}\|^2} \quad \text{for} \quad \|\boldsymbol{l}\|^4\, w \ll 1 ,$$

giving rise to to a linear chirp.



(a) Real part (b) Imaginary part

Figure: Chirp modulation.

Spread spectrum effect

Review

Spread spectrum effect in a nutshell

1. Radio interferometers take (essentially) Fourier measurements.
2. Recall, the coherence is the maximum inner product between measurement vectors and sparsifying atoms.
3. Thus, coherence is (essentially) the maximum of the Fourier coefficients of the atoms of the sparsifying dictionary.
4. w-modulation spreads the spectrum of the atoms of the sparsifying dictionary, reducing the maximum Fourier coefficient.
5. Spreading the spectrum reduces coherence, thus improving reconstruction fidelity.

- Consistent with findings of Carozzi et al. (2013) from information theoretic approach.
- Studied for constant w (for simplicity) by Wiaux *et al.* (2009b).
- Studied for varying w (with realistic images and various sparse representations) by Wolz *et al.* (2013) by developing fast sparse w-projection algorithm.

Spread spectrum effect

Review

Spread spectrum effect in a nutshell

1. Radio interferometers take (essentially) Fourier measurements.
2. Recall, the coherence is the maximum inner product between measurement vectors and sparsifying atoms.
3. Thus, coherence is (essentially) the maximum of the Fourier coefficients of the atoms of the sparsifying dictionary.
4. w-modulation spreads the spectrum of the atoms of the sparsifying dictionary, reducing the maximum Fourier coefficient.
5. Spreading the spectrum reduces coherence, thus improving reconstruction fidelity.

- Consistent with findings of Carozzi et al. (2013) from information theoretic approach.
- Studied for constant w (for simplicity) by Wiaux *et al.* (2009b).
- Studied for varying w (with realistic images and various sparse representations) by Wolz *et al.* (2013) by developing fast sparse w-projection algorithm.

Spread spectrum effect

Review

Spread spectrum effect in a nutshell

1. Radio interferometers take (essentially) Fourier measurements.
2. Recall, the coherence is the maximum inner product between measurement vectors and sparsifying atoms.
3. Thus, coherence is (essentially) the maximum of the Fourier coefficients of the atoms of the sparsifying dictionary.
4. w-modulation spreads the spectrum of the atoms of the sparsifying dictionary, reducing the maximum Fourier coefficient.
5. Spreading the spectrum reduces coherence, thus improving reconstruction fidelity.

- Consistent with findings of Carozzi et al. (2013) from information theoretic approach.
- Studied for constant w (for simplicity) by Wiaux *et al.* (2009b).
- Studied for varying w (with realistic images and various sparse representations) by Wolz *et al.* (2013) by developing fast sparse w-projection algorithm.

Spread spectrum effect
Review

Spread spectrum effect in a nutshell

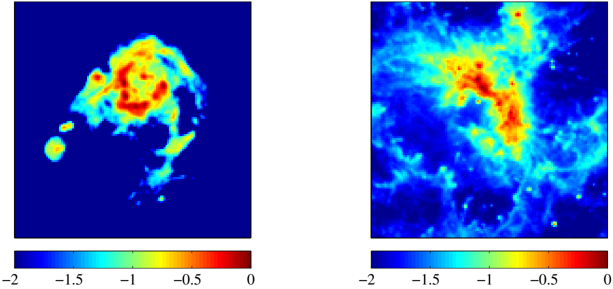
1. Radio interferometers take (essentially) Fourier measurements.
2. Recall, the coherence is the maximum inner product between measurement vectors and sparsifying atoms.
3. Thus, coherence is (essentially) the maximum of the Fourier coefficients of the atoms of the sparsifying dictionary.
4. w-modulation spreads the spectrum of the atoms of the sparsifying dictionary, reducing the maximum Fourier coefficient.
5. Spreading the spectrum reduces coherence, thus improving reconstruction fidelity.

- Consistent with findings of Carozzi et al. (2013) from information theoretic approach.
- Studied for constant w (for simplicity) by Wiaux *et al.* (2009b).
- Studied for varying w (with realistic images and various sparse representations) by Wolz *et al.* (2013) by developing fast sparse w-projection algorithm.

Spread spectrum effect for varying w

Results on simulations

- Perform simulations to assess the effectiveness of the spread spectrum effect in the presence of varying w.
- Consider idealised simulations with uniformly random visibility sampling.

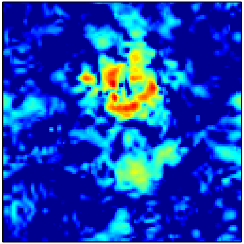


(a) M31 (b) 30Dor

Figure: Ground truth images in logarithmic scale.

Spread spectrum effect for varying w

Results on simulations

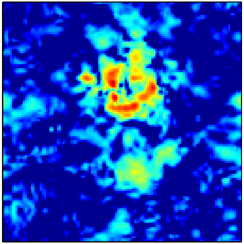


(a) $w_{\rm d} = 0 \rightarrow$ SNR$=5$dB

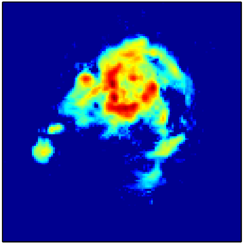
Figure: Reconstructed images of M31 for 10% coverage.

Spread spectrum effect for varying w

Results on simulations



(a) $w_{\mathrm{d}} = 0 \rightarrow$ SNR$= 5$dB

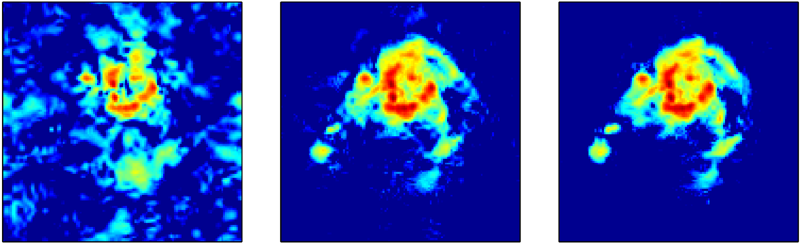


(c) $w_{\mathrm{d}} = 1 \rightarrow$ SNR$= 19$dB

Figure: Reconstructed images of M31 for 10% coverage.

Spread spectrum effect for varying w

Results on simulations

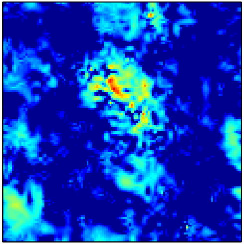


(a) $w_{\mathrm{d}} = 0 \rightarrow$ SNR$=$ 5dB (b) $w_{\mathrm{d}} \sim \mathcal{U}(0,1) \rightarrow$ SNR$=$ 16dB (c) $w_{\mathrm{d}} = 1 \rightarrow$ SNR$=$ 19dB

Figure: Reconstructed images of M31 for 10% coverage.

Spread spectrum effect for varying w

Results on simulations

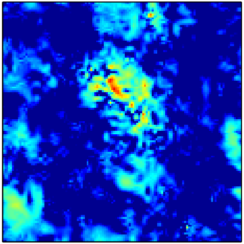


(a) $w_{\rm d} = 0 \rightarrow$ SNR$=$ 2dB

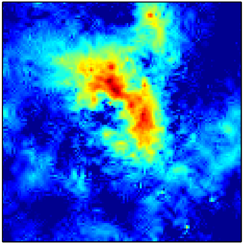
Figure: Reconstructed images of 30Dor for 10% coverage.

Spread spectrum effect for varying w

Results on simulations



(a) $w_{\rm d} = 0 \rightarrow$ SNR$= 2$dB

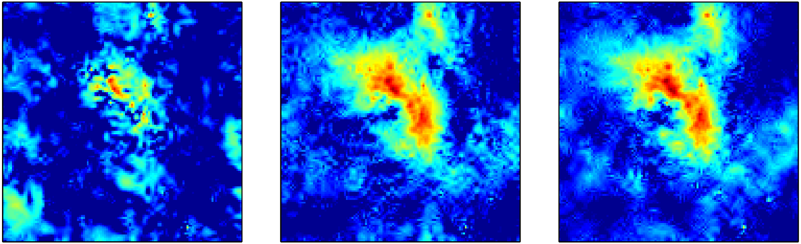


(c) $w_{\rm d} = 1 \rightarrow$ SNR$= 15$dB

Figure: Reconstructed images of 30Dor for 10% coverage.

Spread spectrum effect for varying w

Results on simulations

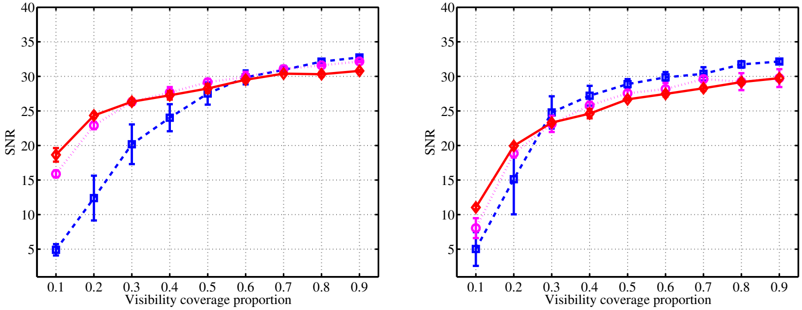


(a) $w_{\rm d} = 0 \rightarrow$ SNR$=$ 2dB (b) $w_{\rm d} \sim \mathcal{U}(0,1) \rightarrow$ SNR$=$ 12dB (c) $w_{\rm d} = 1 \rightarrow$ SNR$=$ 15dB

Figure: Reconstructed images of 30Dor for 10% coverage.

Spread spectrum effect for varying w

Results on simulations



(a) Daubechies 8 (Db8) wavelets (b) Dirac basis

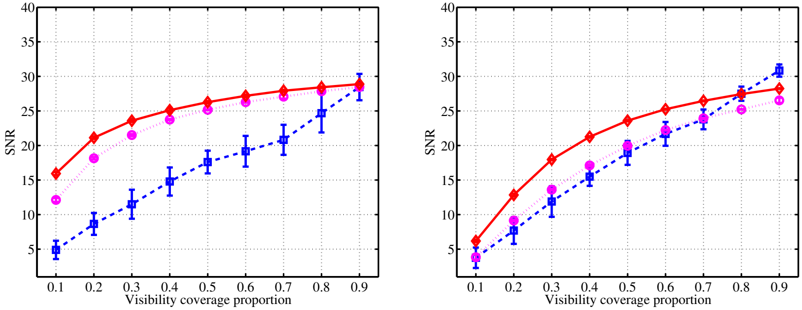
Figure: Reconstruction fidelity for M31.

Improvement in reconstruction fidelity due to the spread spectrum effect for varying w is almost as large as the case of constant maximum w!

- As expected, for the case where coherence is already optimal, there is little improvement.

Spread spectrum effect for varying w

Results on simulations



(a) Daubechies 8 (Db8) wavelets

(b) Dirac basis

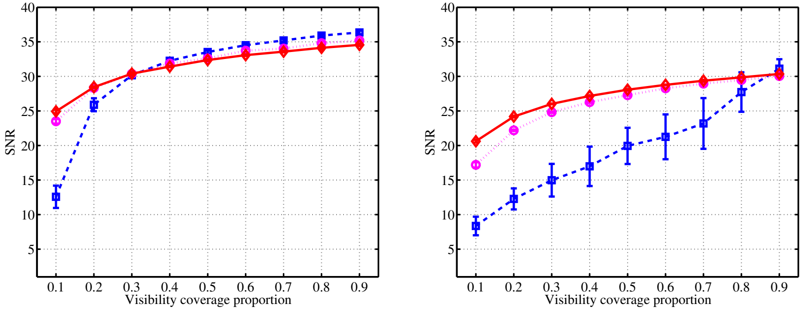
Figure: Reconstruction fidelity for 30Dor.

Improvement in reconstruction fidelity due to the spread spectrum effect for varying w is almost as large as the case of constant maximum w!

- As expected, for the case where coherence is already optimal, there is little improvement.

Spread spectrum effect for varying w

Results on simulations



(a) M31 (b) 30 Dor

Figure: Reconstruction fidelity using SARA.

Improvement in reconstruction fidelity due to the spread spectrum effect for varying w is almost as large as the case of constant maximum w!

- As expected, for the case where coherence is already optimal, there is little improvement.

Compressive sensing for radio interferometric imaging

Outlook

- Effectiveness of compressive sensing for radio interferometric imaging demonstrated.
- We have just released the PURIFY code to scale to real data.
- Includes state-of-the-art convex optimisation algorithms that support parallelisation.

Apply to observations made by real interferometric telescopes.

Compressive sensing for radio interferometric imaging
Outlook

- Effectiveness of compressive sensing for radio interferometric imaging demonstrated.
- We have just released the PURIFY code to scale to real data.
- Includes state-of-the-art convex optimisation algorithms that support parallelisation.

Apply to observations made by real interferometric telescopes.

PURIFY code http://basp-group.github.io/purify/

Next-generation radio interferometric imaging
Carrillo, McEwen, Wiaux

PURIFY is an open-source code that provides functionality to perform radio interferometric imaging, leveraging recent developments in the field of compressive sensing and convex optimisation.

Compressive sensing for radio interferometric imaging
Outlook

- Effectiveness of compressive sensing for radio interferometric imaging demonstrated.
- We have just released the PURIFY code to scale to real data.
- Includes state-of-the-art convex optimisation algorithms that support parallelisation.

Apply to observations made by real interferometric telescopes.

PURIFY code http://basp-group.github.io/purify/

Next-generation radio interferometric imaging
Carrillo, McEwen, Wiaux

PURIFY is an open-source code that provides functionality to perform radio interferometric imaging, leveraging recent developments in the field of compressive sensing and convex optimisation.

Summary

Astrostatistics is now a mature field.

Summary

Astrostatistics is now a mature field.

Informatics techniques (sparsity, wavelets, compressive sensing) are a complementary approach. . .

. . . leading to the emerging field of astroinformatics.

Summary

Astrostatistics is now a mature field.

Informatics techniques (sparsity, wavelets, compressive sensing) are a complementary approach. . .

. . . leading to the emerging field of astroinformatics.

Many codes for application to cosmological data (CMB, LSS) available from:

www.jasonmcewen.org

For postdoc opportunities see:

www.jasonmcewen.org