

Exploiting Spatial Sparsity for Multi-spectral Image Reconstruction in Optical Interferometry

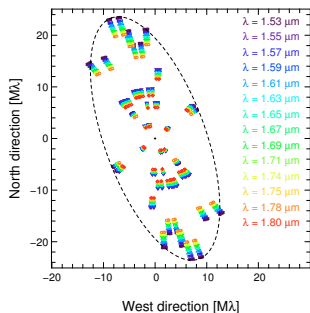
Éric Thiébaud & Ferréol Soulez

Centre de Recherche Astronomique de Lyon
Université Claude Bernard Lyon I
Ecole Normale Supérieure de Lyon
eric.thiebaut@univ-lyon1.fr

SPIE Conference in Amsterdam, 2012

Image Reconstruction in Interferometry

Interferometric Measurements



◀ (u, v) coverage for Arcturus interferometer; (IOTA/IONIC Lacour et al. 2008).

- interferometers sample the Fourier transform of the object brightness distribution:

$$\underbrace{y_{m,\ell}}_{\text{data}} = \underbrace{\widehat{I}_{\lambda_\ell}(\boldsymbol{\nu}_{m,\ell})}_{\text{complex vis.}} + \underbrace{e_{m,\ell}}_{\text{errors}} \quad \text{with} \quad \underbrace{\boldsymbol{\nu}_{m,\ell}}_{\text{spatial freq.}} = \underbrace{B_m/\lambda_\ell}_{\text{baseline/wavelength}}$$

- missing data (voids in (u, v) coverage, partial Fourier phase information, etc.)
- ▶ no unique solution
- ▶ image reconstruction must take into account the data and **priors**;

Image Reconstruction in a Nutshell

- **objective:** seek for the image that best match our priors while being compatible with the data
- in math:

$$\mathbf{x}^+ = \arg \min_{\mathbf{x} \in \mathbb{X}} \underbrace{f_{\text{prior}}(\mathbf{x})}_{\text{regularization}} \quad \text{s.t.} \quad \underbrace{f_{\text{data}}(\mathbf{x}|\mathbf{y})}_{\text{likelihood}} \leq \eta$$

with:

- $f_{\text{prior}}(\mathbf{x})$ is the *distance* to the priors;
- $f_{\text{data}}(\mathbf{x}|\mathbf{y})$ is the *distance* to the data, e.g. χ^2 ;
- η is a tolerance level;
- \mathbb{X} is the feasible set, e.g. $\mathbb{X} = \{\mathbf{x} \in \mathbb{R}_+^N\}$

the Lagrangian of this constrained problem writes:

$$\mathcal{L}(\mathbf{x}; \alpha) = f_{\text{prior}}(\mathbf{x}) + \alpha f_{\text{data}}(\mathbf{x}|\mathbf{y})$$

with $\alpha > 0$ the Lagrange multiplier tuned so that:

$$f_{\text{data}}(\mathbf{x}_\alpha|\mathbf{y}) = \eta \quad \text{with:} \quad \mathbf{x}_\alpha = \arg \min_{\mathbf{x} \in \mathbb{X}} \{f_{\text{prior}}(\mathbf{x}) + \alpha f_{\text{data}}(\mathbf{x}|\mathbf{y})\}$$

Direct Model and Likelihood

- **image parameters:**

$$\boxed{x_{n,\ell} = I_{\lambda_\ell}(\boldsymbol{\theta}_n)} \quad \text{with:} \quad \begin{cases} \lambda_\ell & \text{effective wavelength in } \ell\text{th spectral channel} \\ \boldsymbol{\theta}_n & \text{angular direction for } n\text{th spatial pixel} \\ \mathbf{B}_m & \text{projected position of } m\text{th observed baseline} \end{cases}$$

- **direct model** of the data:

$$y_{p,m,\ell} = \sum_n H_{p,m,n,\ell} x_{n,\ell} + e_{p,m,\ell} \quad \text{in short:} \quad \boxed{\mathbf{y} = \mathbf{H} \cdot \mathbf{x} + \mathbf{e}}$$

with:

$$H_{p,m,n,\ell} = \begin{cases} +\cos(\boldsymbol{\theta}_n^\top \cdot \mathbf{B}_m / \lambda_\ell) & \text{for } p = 1 \\ -\sin(\boldsymbol{\theta}_n^\top \cdot \mathbf{B}_m / \lambda_\ell) & \text{for } p = 2 \end{cases}$$

- **likelihood** (assuming Gaussian statistics for the errors):

$$\boxed{f_{\text{data}}(\mathbf{x}|\mathbf{y}) = \frac{1}{2} (\mathbf{H} \cdot \mathbf{x} - \mathbf{y})^\top \cdot \mathbf{W} \cdot (\mathbf{H} \cdot \mathbf{x} - \mathbf{y})} \quad \text{with:} \quad \mathbf{W} = \text{Cov}(\mathbf{e})^{-1}$$

- note: same wavelengths in the model than in the data (*i.e.* no spectral interpolation of the model)

Direct Model and Likelihood

- **image parameters:**

$$\boxed{x_{n,\ell} = I_{\lambda_\ell}(\boldsymbol{\theta}_n)} \quad \text{with:} \quad \begin{cases} \lambda_\ell & \text{effective wavelength in } \ell\text{th spectral channel} \\ \boldsymbol{\theta}_n & \text{angular direction for } n\text{th spatial pixel} \\ \mathbf{B}_m & \text{projected position of } m\text{th observed baseline} \end{cases}$$

- **direct model** of the data:

$$y_{p,m,\ell} = \sum_n H_{p,m,n,\ell} x_{n,\ell} + e_{p,m,\ell} \quad \text{in short:} \quad \boxed{\mathbf{y} = \mathbf{H} \cdot \mathbf{x} + \mathbf{e}}$$

with:

$$H_{p,m,n,\ell} = \begin{cases} +\cos(\boldsymbol{\theta}_n^\top \cdot \mathbf{B}_m / \lambda_\ell) & \text{for } p = 1 \\ -\sin(\boldsymbol{\theta}_n^\top \cdot \mathbf{B}_m / \lambda_\ell) & \text{for } p = 2 \end{cases}$$

- **likelihood** (assuming Gaussian statistics for the errors):

$$\boxed{f_{\text{data}}(\mathbf{x}|\mathbf{y}) = \frac{1}{2} (\mathbf{H} \cdot \mathbf{x} - \mathbf{y})^\top \cdot \mathbf{W} \cdot (\mathbf{H} \cdot \mathbf{x} - \mathbf{y})} \quad \text{with:} \quad \mathbf{W} = \text{Cov}(\mathbf{e})^{-1}$$

- note: same wavelengths in the model than in the data (*i.e.* no spectral interpolation of the model)

Direct Model and Likelihood

- **image parameters:**

$$\boxed{x_{n,\ell} = I_{\lambda_\ell}(\boldsymbol{\theta}_n)} \quad \text{with: } \begin{cases} \lambda_\ell & \text{effective wavelength in } \ell\text{th spectral channel} \\ \boldsymbol{\theta}_n & \text{angular direction for } n\text{th spatial pixel} \\ \mathbf{B}_m & \text{projected position of } m\text{th observed baseline} \end{cases}$$

- **direct model** of the data:

$$y_{p,m,\ell} = \sum_n H_{p,m,n,\ell} x_{n,\ell} + e_{p,m,\ell} \quad \text{in short: } \boxed{\mathbf{y} = \mathbf{H} \cdot \mathbf{x} + \mathbf{e}}$$

with:

$$H_{p,m,n,\ell} = \begin{cases} + \cos(\boldsymbol{\theta}_n^\top \cdot \mathbf{B}_m / \lambda_\ell) & \text{for } p = 1 \\ - \sin(\boldsymbol{\theta}_n^\top \cdot \mathbf{B}_m / \lambda_\ell) & \text{for } p = 2 \end{cases}$$

- **likelihood** (assuming Gaussian statistics for the errors):

$$\boxed{f_{\text{data}}(\mathbf{x}|\mathbf{y}) = \frac{1}{2} (\mathbf{H} \cdot \mathbf{x} - \mathbf{y})^\top \cdot \mathbf{W} \cdot (\mathbf{H} \cdot \mathbf{x} - \mathbf{y})} \quad \text{with: } \mathbf{W} = \text{Cov}(\mathbf{e})^{-1}$$

- note: same wavelengths in the model than in the data (*i.e.* no spectral interpolation of the model)

Direct Model and Likelihood

- **image parameters:**

$$\boxed{x_{n,\ell} = I_{\lambda_\ell}(\boldsymbol{\theta}_n)} \quad \text{with: } \begin{cases} \lambda_\ell & \text{effective wavelength in } \ell\text{th spectral channel} \\ \boldsymbol{\theta}_n & \text{angular direction for } n\text{th spatial pixel} \\ \mathbf{B}_m & \text{projected position of } m\text{th observed baseline} \end{cases}$$

- **direct model** of the data:

$$y_{p,m,\ell} = \sum_n H_{p,m,n,\ell} x_{n,\ell} + e_{p,m,\ell} \quad \text{in short: } \boxed{\mathbf{y} = \mathbf{H} \cdot \mathbf{x} + \mathbf{e}}$$

with:

$$H_{p,m,n,\ell} = \begin{cases} + \cos(\boldsymbol{\theta}_n^\top \cdot \mathbf{B}_m / \lambda_\ell) & \text{for } p = 1 \\ - \sin(\boldsymbol{\theta}_n^\top \cdot \mathbf{B}_m / \lambda_\ell) & \text{for } p = 2 \end{cases}$$

- **likelihood** (assuming Gaussian statistics for the errors):

$$\boxed{f_{\text{data}}(\mathbf{x}|\mathbf{y}) = \frac{1}{2} (\mathbf{H} \cdot \mathbf{x} - \mathbf{y})^\top \cdot \mathbf{W} \cdot (\mathbf{H} \cdot \mathbf{x} - \mathbf{y})} \quad \text{with: } \mathbf{W} = \text{Cov}(\mathbf{e})^{-1}$$

- note: same wavelengths in the model than in the data (*i.e.* no spectral interpolation of the model)

Regularization by Means of Sparsity Prior

Regularization by Sparsity Prior

- object of interest: point-like sources (e.g. stellar cluster, galactic center, ...);
 \implies our priors are that the object is **spatially sparse**
- in math, sparsity $\sim \ell_0$ norm:

$$f_{\text{prior}}(\mathbf{x}) = \|\mathbf{x}\|_0 = \mathbf{1}^\top \cdot \iota(\mathbf{x}) \quad \text{with:} \quad \iota(\mathbf{x})_{n,\ell} = \begin{cases} 1 & \text{if } x_{n,\ell} \neq 0; \\ 0 & \text{if } x_{n,\ell} = 0; \end{cases}$$

- however ℓ_0 norm yields non-convex penalty; thus sparsity prior is imposed via ℓ_1 norm (Donoho 2006):

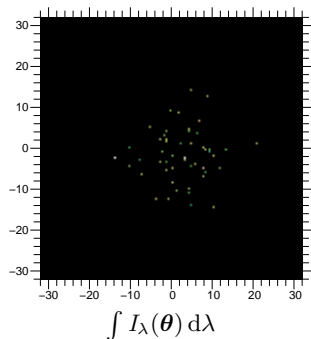
$$f_{\text{sparse}}(\mathbf{x}) = \|\mathbf{x}\|_1 = \sum_{n,\ell} |x_{n,\ell}| \underbrace{= \mathbf{1}^\top \cdot \mathbf{x}}_{\text{because } \mathbf{x} \geq 0}$$

Reconstruction of a Star Cluster with Full Sparsity

- simulation:

- 50 random stars with spectra randomly taken from the library provided by Jacoby et al. (1984)
- $\sim 10^6$ parameters: 128×128 pixels (0.5 mas/pixel) \times 100 spectral channels ($493 \text{ nm} \leq \lambda \leq 507 \text{ nm}$ and $\Delta\lambda = 0.14 \text{ nm}$)
- 100 baselines with $\|B\| \leq 180 \text{ m}$
- Gaussian noise with $\text{SNR} \leq 100$

- spectrally integrated images:

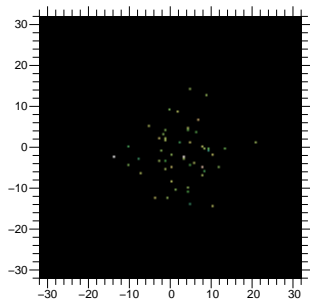


Reconstruction of a Star Cluster with Full Sparsity

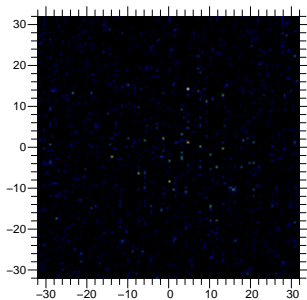
- simulation:

- 50 random stars with spectra randomly taken from the library provided by Jacoby et al. (1984)
- $\sim 10^6$ parameters: 128×128 pixels (0.5 mas/pixel) \times 100 spectral channels ($493 \text{ nm} \leq \lambda \leq 507 \text{ nm}$ and $\Delta\lambda = 0.14 \text{ nm}$)
- 100 baselines with $\|B\| \leq 180 \text{ m}$
- Gaussian noise with $\text{SNR} \leq 100$

- spectrally integrated images:



$$\int I_\lambda(\theta) d\lambda$$



reconst. with f_{sparse}

Trans-spectral Constraints

- with $f_{\text{sparse}}(\mathbf{x}) = \|\mathbf{x}\|_1$, the resulting criterion is separable along the spectral axis, thus image reconstruction makes use of the different spectral channel **independently**;
- we want **trans-spectral continuity** (i.e., in our case, a source is sparse in space but not in wavelength);
- following Fornasier and Rauhut (2008) and Soulez et al. (2011) we propose to use a structured norm:

$$f_{\text{group}}(\mathbf{x}) = \sum_n \left[\sum_{\ell} x_{n,\ell}^2 \right]^{\frac{1}{2}}$$

note that $\left[\sum_{\ell} x_{n,\ell}^2 \right]^{\frac{1}{2}}$ is ℓ_2 norm of spectrum at n th pixel

- example with 3 pixels \times 3 wavelengths:

\uparrow \leftarrow	<table border="1" style="border-collapse: collapse; text-align: center;"> <tr><td>1</td><td>0</td><td>0</td></tr> <tr><td>0</td><td>1</td><td>0</td></tr> <tr><td>0</td><td>0</td><td>1</td></tr> </table>	1	0	0	0	1	0	0	0	1	\uparrow \leftarrow	<table border="1" style="border-collapse: collapse; text-align: center;"> <tr><td>0</td><td>1</td><td>0</td></tr> <tr><td>0</td><td>1</td><td>0</td></tr> <tr><td>0</td><td>0</td><td>1</td></tr> </table>	0	1	0	0	1	0	0	0	1	\uparrow \leftarrow	<table border="1" style="border-collapse: collapse; text-align: center;"> <tr><td>0</td><td>1</td><td>0</td></tr> <tr><td>0</td><td>1</td><td>0</td></tr> <tr><td>0</td><td>1</td><td>0</td></tr> </table>	0	1	0	0	1	0	0	1	0
1	0	0																														
0	1	0																														
0	0	1																														
0	1	0																														
0	1	0																														
0	0	1																														
0	1	0																														
0	1	0																														
0	1	0																														
	$\theta \rightarrow$		$\theta \rightarrow$		$\theta \rightarrow$																											
	$f_{\text{group}}(\mathbf{x}) = 3$		$1 + \sqrt{2} \simeq 2.41$		$\sqrt{3} \simeq 1.73$																											

- the cost is minimal when the chromatic emission is grouped at the same position;
- this is the same behavior as in Total Variation (Rudin et al. 1992);

Trans-spectral Constraints

- with $f_{\text{sparse}}(\mathbf{x}) = \|\mathbf{x}\|_1$, the resulting criterion is separable along the spectral axis, thus image reconstruction makes use of the different spectral channel **independently**;
- we want **trans-spectral continuity** (*i.e.*, in our case, a source is sparse in space but not in wavelength);
- following Fornasier and Rauhut (2008) and Soulez et al. (2011) we propose to use a structured norm:

$$f_{\text{group}}(\mathbf{x}) = \sum_n \left[\sum_{\ell} x_{n,\ell}^2 \right]^{\frac{1}{2}}$$

note that $\left[\sum_{\ell} x_{n,\ell}^2 \right]^{\frac{1}{2}}$ is ℓ_2 norm of spectrum at n th pixel

- example with 3 pixels \times 3 wavelengths:

$$\begin{array}{ccc} \uparrow & & \\ \leftarrow & \begin{array}{|c|c|c|} \hline 1 & 0 & 0 \\ \hline 0 & 1 & 0 \\ \hline 0 & 0 & 1 \\ \hline \end{array} & \\ \theta \rightarrow & & \end{array}$$

$$f_{\text{group}}(\mathbf{x}) = 3$$

$$\begin{array}{ccc} \uparrow & & \\ \leftarrow & \begin{array}{|c|c|c|} \hline 0 & 1 & 0 \\ \hline 0 & 1 & 0 \\ \hline 0 & 0 & 1 \\ \hline \end{array} & \\ \theta \rightarrow & & \end{array}$$

$$1 + \sqrt{2} \simeq 2.41$$

$$\begin{array}{ccc} \uparrow & & \\ \leftarrow & \begin{array}{|c|c|c|} \hline 0 & 1 & 0 \\ \hline 0 & 1 & 0 \\ \hline 0 & 1 & 0 \\ \hline \end{array} & \\ \theta \rightarrow & & \end{array}$$

$$\sqrt{3} \simeq 1.73$$

- the cost is minimal when the chromatic emission is grouped at the same position;
- this is the same behavior as in Total Variation (Rudin et al. 1992);

Trans-spectral Constraints

- with $f_{\text{sparse}}(\mathbf{x}) = \|\mathbf{x}\|_1$, the resulting criterion is separable along the spectral axis, thus image reconstruction makes use of the different spectral channel **independently**;
- we want **trans-spectral continuity** (*i.e.*, in our case, a source is sparse in space but not in wavelength);
- following Fornasier and Rauhut (2008) and Soulez et al. (2011) we propose to use a structured norm:

$$f_{\text{group}}(\mathbf{x}) = \sum_n \left[\sum_{\ell} x_{n,\ell}^2 \right]^{\frac{1}{2}}$$

note that $\left[\sum_{\ell} x_{n,\ell}^2 \right]^{\frac{1}{2}}$ is ℓ_2 norm of spectrum at n th pixel

- example with 3 pixels \times 3 wavelengths:

$$f_{\text{group}}(\mathbf{x}) = 3 \quad 1 + \sqrt{2} \simeq 2.41 \quad \sqrt{3} \simeq 1.73$$

- the cost is minimal when the chromatic emission is grouped at the same position;
- this is the same behavior as in Total Variation (Rudin et al. 1992);

Trans-spectral Constraints

- with $f_{\text{sparse}}(\mathbf{x}) = \|\mathbf{x}\|_1$, the resulting criterion is separable along the spectral axis, thus image reconstruction makes use of the different spectral channel **independently**;
- we want **trans-spectral continuity** (*i.e.*, in our case, a source is sparse in space but not in wavelength);
- following Fornasier and Rauhut (2008) and Soulez et al. (2011) we propose to use a structured norm:

$$f_{\text{group}}(\mathbf{x}) = \sum_n \left[\sum_{\ell} x_{n,\ell}^2 \right]^{\frac{1}{2}}$$

note that $\left[\sum_{\ell} x_{n,\ell}^2 \right]^{\frac{1}{2}}$ is ℓ_2 norm of spectrum at n th pixel

- example with 3 pixels \times 3 wavelengths:

$$\begin{array}{c} \uparrow \\ \leftarrow \\ \begin{array}{|c|c|c|} \hline 1 & 0 & 0 \\ \hline 0 & 1 & 0 \\ \hline 0 & 0 & 1 \\ \hline \end{array} \\ \theta \rightarrow \end{array}$$

$$f_{\text{group}}(\mathbf{x}) = 3$$

$$\begin{array}{c} \uparrow \\ \leftarrow \\ \begin{array}{|c|c|c|} \hline 0 & 1 & 0 \\ \hline 0 & 1 & 0 \\ \hline 0 & 0 & 1 \\ \hline \end{array} \\ \theta \rightarrow \end{array}$$

$$1 + \sqrt{2} \simeq 2.41$$

$$\begin{array}{c} \uparrow \\ \leftarrow \\ \begin{array}{|c|c|c|} \hline 0 & 1 & 0 \\ \hline 0 & 1 & 0 \\ \hline 0 & 1 & 0 \\ \hline \end{array} \\ \theta \rightarrow \end{array}$$

$$\sqrt{3} \simeq 1.73$$

- the cost is minimal when the chromatic emission is grouped at the same position;
- this is the same behavior as in Total Variation (Rudin et al. 1992);

Trans-spectral Constraints

- with $f_{\text{sparse}}(\mathbf{x}) = \|\mathbf{x}\|_1$, the resulting criterion is separable along the spectral axis, thus image reconstruction makes use of the different spectral channel **independently**;
- we want **trans-spectral continuity** (*i.e.*, in our case, a source is sparse in space but not in wavelength);
- following Fornasier and Rauhut (2008) and Soulez et al. (2011) we propose to use a structured norm:

$$f_{\text{group}}(\mathbf{x}) = \sum_n \left[\sum_{\ell} x_{n,\ell}^2 \right]^{\frac{1}{2}}$$

note that $\left[\sum_{\ell} x_{n,\ell}^2 \right]^{\frac{1}{2}}$ is ℓ_2 norm of spectrum at n th pixel

- example with 3 pixels \times 3 wavelengths:

$$\begin{array}{c} \uparrow \\ \swarrow \\ \begin{array}{|c|c|c|} \hline 1 & 0 & 0 \\ \hline 0 & 1 & 0 \\ \hline 0 & 0 & 1 \\ \hline \end{array} \\ \theta \longrightarrow \end{array}$$

$$f_{\text{group}}(\mathbf{x}) = 3$$

$$\begin{array}{c} \uparrow \\ \swarrow \\ \begin{array}{|c|c|c|} \hline 0 & 1 & 0 \\ \hline 0 & 1 & 0 \\ \hline 0 & 0 & 1 \\ \hline \end{array} \\ \theta \longrightarrow \end{array}$$

$$1 + \sqrt{2} \simeq 2.41$$

$$\begin{array}{c} \uparrow \\ \swarrow \\ \begin{array}{|c|c|c|} \hline 0 & 1 & 0 \\ \hline 0 & 1 & 0 \\ \hline 0 & 1 & 0 \\ \hline \end{array} \\ \theta \longrightarrow \end{array}$$

$$\sqrt{3} \simeq 1.73$$

- the cost is minimal when the chromatic emission is grouped at the same position;
- this is the same behavior as in Total Variation (Rudin et al. 1992);

Trans-spectral Constraints

- with $f_{\text{sparse}}(\mathbf{x}) = \|\mathbf{x}\|_1$, the resulting criterion is separable along the spectral axis, thus image reconstruction makes use of the different spectral channel **independently**;
- we want **trans-spectral continuity** (*i.e.*, in our case, a source is sparse in space but not in wavelength);
- following Fornasier and Rauhut (2008) and Soulez et al. (2011) we propose to use a structured norm:

$$f_{\text{group}}(\mathbf{x}) = \sum_n \left[\sum_{\ell} x_{n,\ell}^2 \right]^{\frac{1}{2}}$$

note that $\left[\sum_{\ell} x_{n,\ell}^2 \right]^{\frac{1}{2}}$ is ℓ_2 norm of spectrum at n th pixel

- example with 3 pixels \times 3 wavelengths:

$$\begin{array}{c} \uparrow \\ \swarrow \\ \begin{array}{|c|c|c|} \hline 1 & 0 & 0 \\ \hline 0 & 1 & 0 \\ \hline 0 & 0 & 1 \\ \hline \end{array} \\ \theta \longrightarrow \end{array}$$

$$f_{\text{group}}(\mathbf{x}) = 3$$

$$\begin{array}{c} \uparrow \\ \swarrow \\ \begin{array}{|c|c|c|} \hline 0 & 1 & 0 \\ \hline 0 & 1 & 0 \\ \hline 0 & 0 & 1 \\ \hline \end{array} \\ \theta \longrightarrow \end{array}$$

$$1 + \sqrt{2} \simeq 2.41$$

$$\begin{array}{c} \uparrow \\ \swarrow \\ \begin{array}{|c|c|c|} \hline 0 & 1 & 0 \\ \hline 0 & 1 & 0 \\ \hline 0 & 1 & 0 \\ \hline \end{array} \\ \theta \longrightarrow \end{array}$$

$$\sqrt{3} \simeq 1.73$$

- the cost is minimal when the chromatic emission is grouped at the same position;
- this is the same behavior as in Total Variation (Rudin et al. 1992);

Trans-spectral Constraints

- with $f_{\text{sparse}}(\mathbf{x}) = \|\mathbf{x}\|_1$, the resulting criterion is separable along the spectral axis, thus image reconstruction makes use of the different spectral channel **independently**;
- we want **trans-spectral continuity** (*i.e.*, in our case, a source is sparse in space but not in wavelength);
- following Fornasier and Rauhut (2008) and Soulez et al. (2011) we propose to use a structured norm:

$$f_{\text{group}}(\mathbf{x}) = \sum_n \left[\sum_{\ell} x_{n,\ell}^2 \right]^{\frac{1}{2}}$$

note that $\left[\sum_{\ell} x_{n,\ell}^2 \right]^{\frac{1}{2}}$ is ℓ_2 norm of spectrum at n th pixel

- example with 3 pixels \times 3 wavelengths:

$$\begin{array}{c} \uparrow \\ \swarrow \\ \begin{array}{|c|c|c|} \hline 1 & 0 & 0 \\ \hline 0 & 1 & 0 \\ \hline 0 & 0 & 1 \\ \hline \end{array} \\ \theta \longrightarrow \end{array}$$

$$f_{\text{group}}(\mathbf{x}) = 3$$

$$\begin{array}{c} \uparrow \\ \swarrow \\ \begin{array}{|c|c|c|} \hline 0 & 1 & 0 \\ \hline 0 & 1 & 0 \\ \hline 0 & 0 & 1 \\ \hline \end{array} \\ \theta \longrightarrow \end{array}$$

$$1 + \sqrt{2} \simeq 2.41$$

$$\begin{array}{c} \uparrow \\ \swarrow \\ \begin{array}{|c|c|c|} \hline 0 & 1 & 0 \\ \hline 0 & 1 & 0 \\ \hline 0 & 1 & 0 \\ \hline \end{array} \\ \theta \longrightarrow \end{array}$$

$$\sqrt{3} \simeq 1.73$$

- the cost is minimal when the chromatic emission is grouped at the same position;
- this is the same behavior as in Total Variation (Rudin et al. 1992);

Trans-spectral Constraints

- with $f_{\text{sparse}}(\mathbf{x}) = \|\mathbf{x}\|_1$, the resulting criterion is separable along the spectral axis, thus image reconstruction makes use of the different spectral channel **independently**;
- we want **trans-spectral continuity** (*i.e.*, in our case, a source is sparse in space but not in wavelength);
- following Fornasier and Rauhut (2008) and Soulez et al. (2011) we propose to use a structured norm:

$$f_{\text{group}}(\mathbf{x}) = \sum_n \left[\sum_{\ell} x_{n,\ell}^2 \right]^{\frac{1}{2}}$$

note that $\left[\sum_{\ell} x_{n,\ell}^2 \right]^{\frac{1}{2}}$ is ℓ_2 norm of spectrum at n th pixel

- example with 3 pixels \times 3 wavelengths:

$$\begin{array}{c} \uparrow \\ \swarrow \\ \begin{array}{|c|c|c|} \hline 1 & 0 & 0 \\ \hline 0 & 1 & 0 \\ \hline 0 & 0 & 1 \\ \hline \end{array} \\ \theta \longrightarrow \end{array}$$

$$f_{\text{group}}(\mathbf{x}) = 3$$

$$\begin{array}{c} \uparrow \\ \swarrow \\ \begin{array}{|c|c|c|} \hline 0 & 1 & 0 \\ \hline 0 & 1 & 0 \\ \hline 0 & 0 & 1 \\ \hline \end{array} \\ \theta \longrightarrow \end{array}$$

$$1 + \sqrt{2} \simeq 2.41$$

$$\begin{array}{c} \uparrow \\ \swarrow \\ \begin{array}{|c|c|c|} \hline 0 & 1 & 0 \\ \hline 0 & 1 & 0 \\ \hline 0 & 1 & 0 \\ \hline \end{array} \\ \theta \longrightarrow \end{array}$$

$$\sqrt{3} \simeq 1.73$$

- the cost is minimal when the chromatic emission is grouped at the same position;
- this is the same behavior as in Total Variation (Rudin et al. 1992);

Trans-spectral Constraints

- with $f_{\text{sparse}}(\mathbf{x}) = \|\mathbf{x}\|_1$, the resulting criterion is separable along the spectral axis, thus image reconstruction makes use of the different spectral channel **independently**;
- we want **trans-spectral continuity** (*i.e.*, in our case, a source is sparse in space but not in wavelength);
- following Fornasier and Rauhut (2008) and Soulez et al. (2011) we propose to use a structured norm:

$$f_{\text{group}}(\mathbf{x}) = \sum_n \left[\sum_{\ell} x_{n,\ell}^2 \right]^{\frac{1}{2}}$$

note that $\left[\sum_{\ell} x_{n,\ell}^2 \right]^{\frac{1}{2}}$ is ℓ_2 norm of spectrum at n th pixel

- example with 3 pixels \times 3 wavelengths:

$$\begin{array}{c} \uparrow \\ \swarrow \\ \begin{array}{|c|c|c|} \hline 1 & 0 & 0 \\ \hline 0 & 1 & 0 \\ \hline 0 & 0 & 1 \\ \hline \end{array} \\ \theta \longrightarrow \end{array}$$

$$f_{\text{group}}(\mathbf{x}) = 3$$

$$\begin{array}{c} \uparrow \\ \swarrow \\ \begin{array}{|c|c|c|} \hline 0 & 1 & 0 \\ \hline 0 & 1 & 0 \\ \hline 0 & 0 & 1 \\ \hline \end{array} \\ \theta \longrightarrow \end{array}$$

$$1 + \sqrt{2} \simeq 2.41$$

$$\begin{array}{c} \uparrow \\ \swarrow \\ \begin{array}{|c|c|c|} \hline 0 & 1 & 0 \\ \hline 0 & 1 & 0 \\ \hline 0 & 1 & 0 \\ \hline \end{array} \\ \theta \longrightarrow \end{array}$$

$$\sqrt{3} \simeq 1.73$$

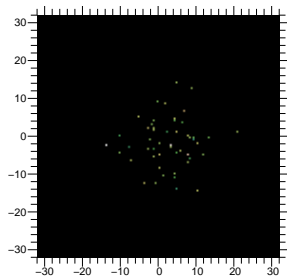
- the cost is minimal when the chromatic emission is grouped at the same position;
- this is the same behavior as in Total Variation (Rudin et al. 1992);

Images of a Star Cluster with Structured Sparsity

- fully separable sparsity: $f_{\text{sparse}}(\mathbf{x}) = \|\mathbf{x}\|_1 = \sum_{n,\ell} |x_{n,\ell}| \underbrace{= \mathbf{1}^\top \cdot \mathbf{x}}_{\text{because } x \geq 0}$

- structured sparsity: $f_{\text{group}}(\mathbf{x}) = \sum_n \left[\sum_\ell x_{n,\ell}^2 \right]^{\frac{1}{2}}$

- spectrally integrated images:



$$\int I_\lambda(\boldsymbol{\theta}) d\lambda$$

 reconst. with f_{sparse}

Images of a Star Cluster with Structured Sparsity

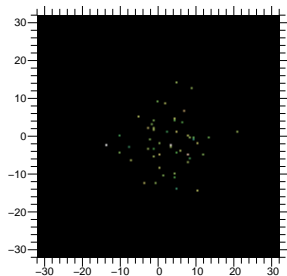
- fully separable sparsity:

$$f_{\text{sparse}}(\mathbf{x}) = \|\mathbf{x}\|_1 = \sum_{n,\ell} |x_{n,\ell}| \underbrace{= \mathbf{1}^\top \cdot \mathbf{x}}_{\text{because } \mathbf{x} \geq 0}$$

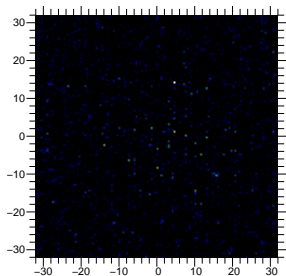
- structured sparsity:

$$f_{\text{group}}(\mathbf{x}) = \sum_n \left[\sum_\ell x_{n,\ell}^2 \right]^{\frac{1}{2}}$$

- spectrally integrated images:



$$\int I_\lambda(\boldsymbol{\theta}) d\lambda$$



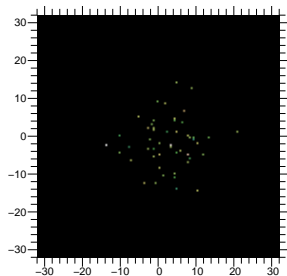
reconst. with f_{sparse}

Images of a Star Cluster with Structured Sparsity

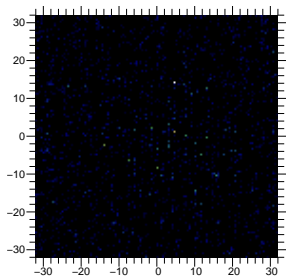
- fully separable sparsity: $f_{\text{sparse}}(\mathbf{x}) = \|\mathbf{x}\|_1 = \sum_{n,\ell} |x_{n,\ell}| \stackrel{\text{because } \mathbf{x} \geq 0}{=} \mathbf{1}^\top \cdot \mathbf{x}$

- structured sparsity: $f_{\text{group}}(\mathbf{x}) = \sum_n \left[\sum_\ell x_{n,\ell}^2 \right]^{\frac{1}{2}}$

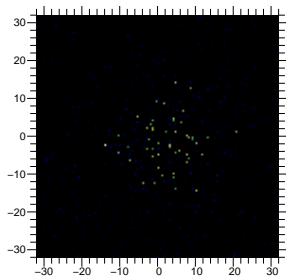
- spectrally integrated images:



$$\int I_\lambda(\boldsymbol{\theta}) d\lambda$$



reconst. with f_{sparse}



reconst. with f_{group}

Pushing Spectral Continuity to the Extreme: Assuming a Gray Object

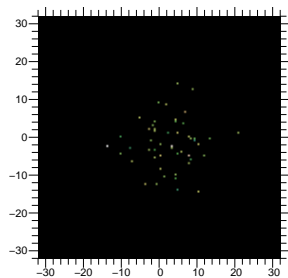
- an extreme case to impose spectral continuity is to assume a gray object:

$$\boxed{x_{n,\ell} = g_n} \quad \text{with} \quad x_{n,\ell} = I_{\lambda_\ell}(\boldsymbol{\theta}_n)$$

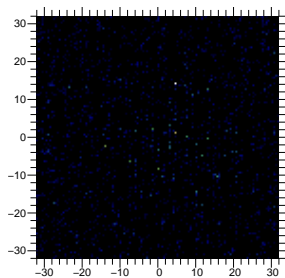
and impose spatial sparsity by:

$$\boxed{f_{\text{prior}}(\mathbf{g}) = \|\mathbf{g}\|_1}$$

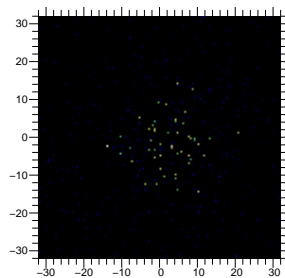
- spectrally integrated images:



$$\int I_{\lambda}(\boldsymbol{\theta}) d\lambda$$



reconst. with f_{sparse}



reconst. with f_{group}

Pushing Spectral Continuity to the Extreme: Assuming a Gray Object

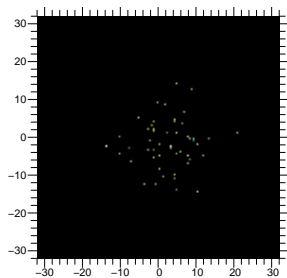
- an extreme case to impose spectral continuity is to assume a gray object:

$$\boxed{x_{n,\ell} = g_n} \quad \text{with} \quad x_{n,\ell} = I_{\lambda_\ell}(\boldsymbol{\theta}_n)$$

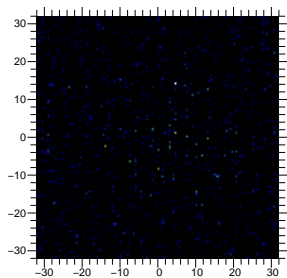
and impose spatial sparsity by:

$$\boxed{f_{\text{prior}}(\mathbf{g}) = \|\mathbf{g}\|_1}$$

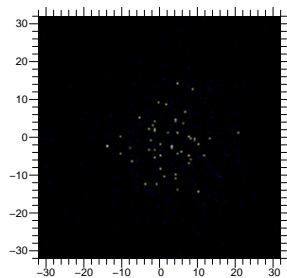
- spectrally integrated images:



$$\int I_{\lambda}(\boldsymbol{\theta}) d\lambda$$

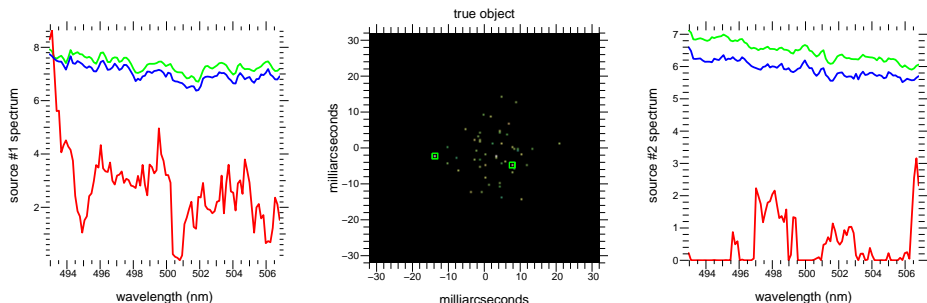


reconst. with f_{sparse}



reconst. as gray object

Spectra Extracted from the Restored 3-D Images

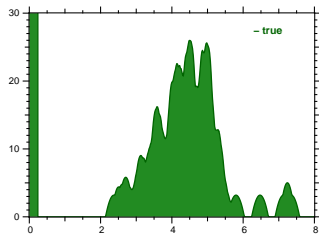


- we consider the spectra of two particular stars
 - in green: true spectrum
 - in red: spectrum in 3-D reconstruction with f_{sparse}
 - in blue: spectrum in 3-D reconstruction with f_{group}
- again transverse constraints yield superior results
- however there is a slight **bias** \implies make a reconstruction with support constraint to unbiased

Distribution of Luminosity

- distribution of average pixel intensity:

$$\frac{1}{\Delta\lambda} \int I_\lambda(\boldsymbol{\theta}) d\lambda$$



- suggest thresholding to detect sources in reconstructed 3-D image:

reconst. with f_{group}

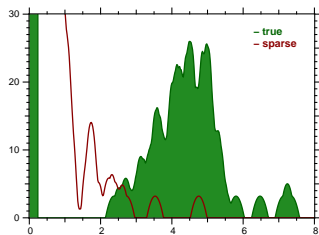
reconst. as gray object

- can be used to debias spectra

Distribution of Luminosity

- distribution of average pixel intensity:

$$\frac{1}{\Delta\lambda} \int I_\lambda(\boldsymbol{\theta}) d\lambda$$



- suggest thresholding to detect sources in reconstructed 3-D image:

reconst. with f_{group}

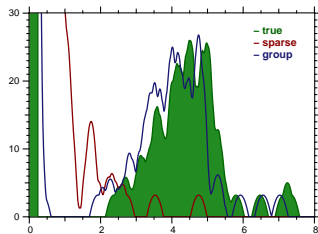
reconst. as gray object

- can be used to debias spectra

Distribution of Luminosity

- distribution of average pixel intensity:

$$\frac{1}{\Delta\lambda} \int I_\lambda(\boldsymbol{\theta}) d\lambda$$



- suggest thresholding to detect sources in reconstructed 3-D image:

reconst. with f_{group}

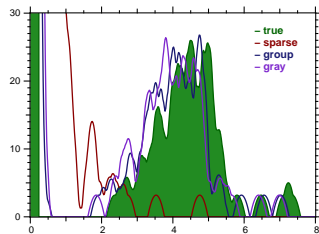
reconst. as gray object

- can be used to debias spectra

Distribution of Luminosity

- distribution of average pixel intensity:

$$\frac{1}{\Delta\lambda} \int I_\lambda(\boldsymbol{\theta}) d\lambda$$



- suggest thresholding to detect sources in reconstructed 3-D image:

reconst. with f_{group}

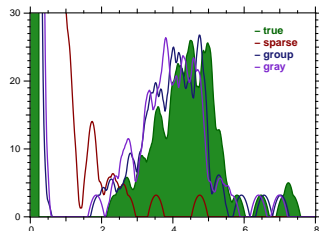
reconst. as gray object

- can be used to debias spectra

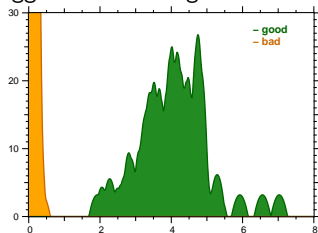
Distribution of Luminosity

- distribution of average pixel intensity:

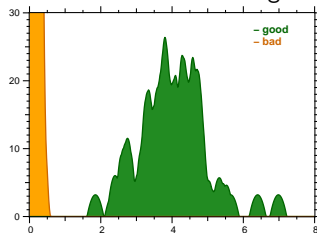
$$\frac{1}{\Delta\lambda} \int I_\lambda(\boldsymbol{\theta}) d\lambda$$



- suggest thresholding to detect sources in reconstructed 3-D image:



reconst. with f_{group}



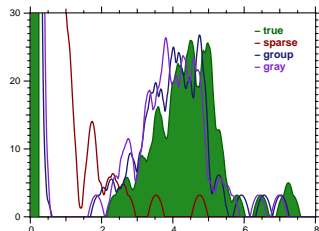
reconst. as gray object

- can be used to debias spectra

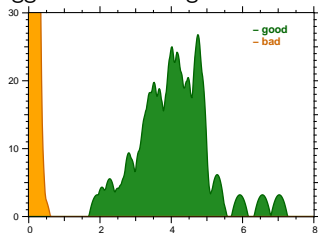
Distribution of Luminosity

- distribution of average pixel intensity:

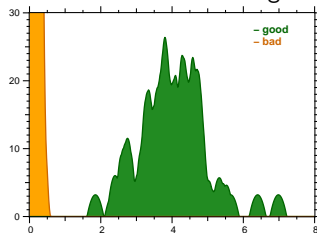
$$\frac{1}{\Delta\lambda} \int I_\lambda(\boldsymbol{\theta}) d\lambda$$



- suggest thresholding to detect sources in reconstructed 3-D image:



reconst. with f_{group}



reconst. as gray object

- can be used to debias spectra

Algorithm

Optimization of non-smooth criterion

- image reconstruction \sim solving problem:

$$\min_{\mathbf{x} \in \mathbb{X}} \{f_{\text{prior}}(\mathbf{x}) + \alpha f_{\text{data}}(\mathbf{x})\}$$

but $f_{\text{prior}}(\mathbf{x})$ is non-smooth (not \mathcal{C}^2)

\Rightarrow rules out quasi-Newton methods (such as VMLM-B in MiRA or WISARD)

- equivalent problem (variables splitting):

$$\min_{\mathbf{x} \in \mathbb{X}, \mathbf{z}} \{f_{\text{prior}}(\mathbf{x}) + \alpha f_{\text{data}}(\mathbf{z})\} \quad \text{s.t.:} \quad \mathbf{x} = \mathbf{z}$$

- augmented Lagrangian:

$$\mathcal{L}_\rho(\mathbf{x}, \mathbf{z}, \mathbf{u}) = f_{\text{prior}}(\mathbf{x}) + \alpha f_{\text{data}}(\mathbf{z}) + \mathbf{u}^\top \cdot (\mathbf{x} - \mathbf{z}) + \frac{\rho}{2} \|\mathbf{x} - \mathbf{z}\|_2^2$$

with $\rho > 0$ and \mathbf{u} the quadratic weight and Lagrange multipliers of the constraints

- solved by *alternating direction method of multipliers* (ADMM) Boyd et al. (2010)

Alternating Direction Method of Multipliers

0. Choose initial variables $\mathbf{z}^{(0)}$ and Lagrange multipliers $\mathbf{u}^{(0)}$. Then repeat, for $t = 1, 2, \dots$ until convergence, the following steps:
1. Choose $\rho^{(t)} > 0$ and update variables \mathbf{x} :

$$\mathbf{x}^{(t)} = \arg \min_{\mathbf{x} \in \mathbb{X}} \mathcal{L}_{\rho^{(t)}}(\mathbf{x}, \mathbf{z}^{(t-1)}, \mathbf{u}^{(t-1)}) = \arg \min_{\mathbf{x} \in \mathbb{X}} \left\{ f_{\text{prior}}(\mathbf{x}) + \frac{\rho^{(t)}}{2} \|\mathbf{x} - \tilde{\mathbf{x}}^{(t)}\|_2^2 \right\}$$

with: $\tilde{\mathbf{x}}^{(t)} = \mathbf{z}^{(t-1)} - \mathbf{u}^{(t-1)} / \rho^{(t)}$.

2. Update auxiliary variable \mathbf{z} :

$$\mathbf{z}^{(t)} = \arg \min_{\mathbf{z}} \mathcal{L}_{\rho^{(t)}}(\mathbf{x}^{(t)}, \mathbf{z}, \mathbf{u}^{(t-1)}) = \arg \min_{\mathbf{z}} \left\{ \alpha f_{\text{data}}(\mathbf{z}) + \frac{\rho^{(t)}}{2} \|\mathbf{z} - \tilde{\mathbf{z}}^{(t)}\|_2^2 \right\}$$

with: $\tilde{\mathbf{z}}^{(t)} = \mathbf{x}^{(t)} + \mathbf{u}^{(t-1)} / \rho^{(t)}$.

3. Update Lagrange multipliers \mathbf{u} :

$$\mathbf{u}^{(t)} = \mathbf{u}^{(t-1)} + \rho^{(t)} (\mathbf{x}^{(t)} - \mathbf{z}^{(t)}) .$$

Alternating Direction Method of Multipliers

0. Choose initial variables $\mathbf{z}^{(0)}$ and Lagrange multipliers $\mathbf{u}^{(0)}$. Then repeat, for $t = 1, 2, \dots$ until convergence, the following steps:
1. Choose $\rho^{(t)} > 0$ and update variables \mathbf{x} :

$$\mathbf{x}^{(t)} = \arg \min_{\mathbf{x} \in \mathbb{X}} \mathcal{L}_{\rho^{(t)}}(\mathbf{x}, \mathbf{z}^{(t-1)}, \mathbf{u}^{(t-1)}) = \arg \min_{\mathbf{x} \in \mathbb{X}} \left\{ f_{\text{prior}}(\mathbf{x}) + \frac{\rho^{(t)}}{2} \|\mathbf{x} - \tilde{\mathbf{x}}^{(t)}\|_2^2 \right\}$$

with: $\tilde{\mathbf{x}}^{(t)} = \mathbf{z}^{(t-1)} - \mathbf{u}^{(t-1)} / \rho^{(t)}$.

2. Update auxiliary variable \mathbf{z} :

$$\mathbf{z}^{(t)} = \arg \min_{\mathbf{z}} \mathcal{L}_{\rho^{(t)}}(\mathbf{x}^{(t)}, \mathbf{z}, \mathbf{u}^{(t-1)}) = \arg \min_{\mathbf{z}} \left\{ \alpha f_{\text{data}}(\mathbf{z}) + \frac{\rho^{(t)}}{2} \|\mathbf{z} - \tilde{\mathbf{z}}^{(t)}\|_2^2 \right\}$$

with: $\tilde{\mathbf{z}}^{(t)} = \mathbf{x}^{(t)} + \mathbf{u}^{(t-1)} / \rho^{(t)}$.

3. Update Lagrange multipliers \mathbf{u} :

$$\mathbf{u}^{(t)} = \mathbf{u}^{(t-1)} + \rho^{(t)} (\mathbf{x}^{(t)} - \mathbf{z}^{(t)}) .$$

Alternating Direction Method of Multipliers

0. Choose initial variables $\mathbf{z}^{(0)}$ and Lagrange multipliers $\mathbf{u}^{(0)}$. Then repeat, for $t = 1, 2, \dots$ until convergence, the following steps:
1. Choose $\rho^{(t)} > 0$ and update variables \mathbf{x} :

$$\mathbf{x}^{(t)} = \arg \min_{\mathbf{x} \in \mathbb{X}} \mathcal{L}_{\rho^{(t)}}(\mathbf{x}, \mathbf{z}^{(t-1)}, \mathbf{u}^{(t-1)}) = \arg \min_{\mathbf{x} \in \mathbb{X}} \left\{ f_{\text{prior}}(\mathbf{x}) + \frac{\rho^{(t)}}{2} \|\mathbf{x} - \tilde{\mathbf{x}}^{(t)}\|_2^2 \right\}$$

with: $\tilde{\mathbf{x}}^{(t)} = \mathbf{z}^{(t-1)} - \mathbf{u}^{(t-1)} / \rho^{(t)}$.

2. Update auxiliary variable \mathbf{z} :

$$\mathbf{z}^{(t)} = \arg \min_{\mathbf{z}} \mathcal{L}_{\rho^{(t)}}(\mathbf{x}^{(t)}, \mathbf{z}, \mathbf{u}^{(t-1)}) = \arg \min_{\mathbf{z}} \left\{ \alpha f_{\text{data}}(\mathbf{z}) + \frac{\rho^{(t)}}{2} \|\mathbf{z} - \tilde{\mathbf{z}}^{(t)}\|_2^2 \right\}$$

with: $\tilde{\mathbf{z}}^{(t)} = \mathbf{x}^{(t)} + \mathbf{u}^{(t-1)} / \rho^{(t)}$.

3. Update Lagrange multipliers \mathbf{u} :

$$\mathbf{u}^{(t)} = \mathbf{u}^{(t-1)} + \rho^{(t)} (\mathbf{x}^{(t)} - \mathbf{z}^{(t)}).$$

Alternating Direction Method of Multipliers

0. Choose initial variables $\mathbf{z}^{(0)}$ and Lagrange multipliers $\mathbf{u}^{(0)}$. Then repeat, for $t = 1, 2, \dots$ until convergence, the following steps:
1. Choose $\rho^{(t)} > 0$ and update variables \mathbf{x} :

$$\mathbf{x}^{(t)} = \arg \min_{\mathbf{x} \in \mathbb{X}} \mathcal{L}_{\rho^{(t)}}(\mathbf{x}, \mathbf{z}^{(t-1)}, \mathbf{u}^{(t-1)}) = \arg \min_{\mathbf{x} \in \mathbb{X}} \left\{ f_{\text{prior}}(\mathbf{x}) + \frac{\rho^{(t)}}{2} \|\mathbf{x} - \tilde{\mathbf{x}}^{(t)}\|_2^2 \right\}$$

with: $\tilde{\mathbf{x}}^{(t)} = \mathbf{z}^{(t-1)} - \mathbf{u}^{(t-1)} / \rho^{(t)}$.

2. Update auxiliary variable \mathbf{z} :

$$\mathbf{z}^{(t)} = \arg \min_{\mathbf{z}} \mathcal{L}_{\rho^{(t)}}(\mathbf{x}^{(t)}, \mathbf{z}, \mathbf{u}^{(t-1)}) = \arg \min_{\mathbf{z}} \left\{ \alpha f_{\text{data}}(\mathbf{z}) + \frac{\rho^{(t)}}{2} \|\mathbf{z} - \tilde{\mathbf{z}}^{(t)}\|_2^2 \right\}$$

with: $\tilde{\mathbf{z}}^{(t)} = \mathbf{x}^{(t)} + \mathbf{u}^{(t-1)} / \rho^{(t)}$.

3. Update Lagrange multipliers \mathbf{u} :

$$\mathbf{u}^{(t)} = \mathbf{u}^{(t-1)} + \rho^{(t)} (\mathbf{x}^{(t)} - \mathbf{z}^{(t)}) .$$

Updating the Auxiliary Variables

- Updating the auxiliary variables z amounts to solving:

$$z^{(t)} = \arg \min_z \left\{ \underbrace{\frac{\alpha}{2} (\mathbf{H} \cdot \mathbf{x} - \mathbf{y})^\top \cdot \mathbf{W} \cdot (\mathbf{H} \cdot \mathbf{x} - \mathbf{y})}_{\alpha f_{\text{data}}(z)} + \frac{\rho^{(t)}}{2} \|z - \tilde{z}^{(t)}\|_2^2 \right\}$$

with: $\tilde{z}^{(t)} = \mathbf{x}^{(t)} + \mathbf{u}^{(t-1)} / \rho^{(t)}$.

- This has an analytical solution:

$$z^{(t)} = \left(\alpha \mathbf{H}^\top \cdot \mathbf{W} \cdot \mathbf{H} + \rho^{(t)} \mathbf{I} \right)^{-1} \cdot \left(\alpha \mathbf{H}^\top \cdot \mathbf{W} \cdot \mathbf{y} + \rho^{(t)} \tilde{z}^{(t)} \right)$$

which can be computed by means of **linear conjugate gradient method** (Hestenes and Stiefel 1952)

- Computations are wavelength-separable \implies trivial to parallelize:

$$\text{solve, } \forall \ell \quad \mathbf{A}_\ell^{(t)} \cdot z_\ell^{(t)} = b_\ell^{(t)} \quad \text{with: } \begin{cases} \mathbf{A}_\ell^{(t)} = \alpha \mathbf{H}_\ell^\top \cdot \mathbf{W}_\ell \cdot \mathbf{H}_\ell + \rho^{(t)} \mathbf{I} \\ b_\ell^{(t)} = \alpha \mathbf{H}_\ell^\top \cdot \mathbf{W}_\ell \cdot \mathbf{y}_\ell + \rho^{(t)} \tilde{z}_\ell^{(t)} \end{cases}$$

Updating the Auxiliary Variables

- Updating the auxiliary variables z amounts to solving:

$$z^{(t)} = \arg \min_z \left\{ \underbrace{\frac{\alpha}{2} (\mathbf{H} \cdot \mathbf{x} - \mathbf{y})^\top \cdot \mathbf{W} \cdot (\mathbf{H} \cdot \mathbf{x} - \mathbf{y})}_{\alpha f_{\text{data}}(z)} + \frac{\rho^{(t)}}{2} \|z - \tilde{z}^{(t)}\|_2^2 \right\}$$

with: $\tilde{z}^{(t)} = \mathbf{x}^{(t)} + \mathbf{u}^{(t-1)} / \rho^{(t)}$.

- This has an analytical solution:

$$z^{(t)} = (\alpha \mathbf{H}^\top \cdot \mathbf{W} \cdot \mathbf{H} + \rho^{(t)} \mathbf{I})^{-1} \cdot (\alpha \mathbf{H}^\top \cdot \mathbf{W} \cdot \mathbf{y} + \rho^{(t)} \tilde{z}^{(t)})$$

which can be computed by means of **linear conjugate gradient method** (Hestenes and Stiefel 1952)

- Computations are wavelength-separable \implies trivial to parallelize:

$$\text{solve, } \forall \ell \quad \mathbf{A}_\ell^{(t)} \cdot z_\ell^{(t)} = b_\ell^{(t)} \quad \text{with: } \begin{cases} \mathbf{A}_\ell^{(t)} = \alpha \mathbf{H}_\ell^\top \cdot \mathbf{W}_\ell \cdot \mathbf{H}_\ell + \rho^{(t)} \mathbf{I} \\ b_\ell^{(t)} = \alpha \mathbf{H}_\ell^\top \cdot \mathbf{W}_\ell \cdot \mathbf{y}_\ell + \rho^{(t)} \tilde{z}_\ell^{(t)} \end{cases}$$

Updating the Auxiliary Variables

- Updating the auxiliary variables z amounts to solving:

$$z^{(t)} = \arg \min_z \left\{ \underbrace{\frac{\alpha}{2} (\mathbf{H} \cdot \mathbf{x} - \mathbf{y})^\top \cdot \mathbf{W} \cdot (\mathbf{H} \cdot \mathbf{x} - \mathbf{y})}_{\alpha f_{\text{data}}(z)} + \frac{\rho^{(t)}}{2} \|z - \tilde{z}^{(t)}\|_2^2 \right\}$$

with: $\tilde{z}^{(t)} = \mathbf{x}^{(t)} + \mathbf{u}^{(t-1)} / \rho^{(t)}$.

- This has an analytical solution:

$$z^{(t)} = \left(\alpha \mathbf{H}^\top \cdot \mathbf{W} \cdot \mathbf{H} + \rho^{(t)} \mathbf{I} \right)^{-1} \cdot \left(\alpha \mathbf{H}^\top \cdot \mathbf{W} \cdot \mathbf{y} + \rho^{(t)} \tilde{z}^{(t)} \right)$$

which can be computed by means of **linear conjugate gradient method** (Hestenes and Stiefel 1952)

- Computations are wavelength-separable \implies trivial to parallelize:

$$\text{solve, } \forall \ell \quad \mathbf{A}_\ell^{(t)} \cdot z_\ell^{(t)} = b_\ell^{(t)} \quad \text{with: } \begin{cases} \mathbf{A}_\ell^{(t)} = \alpha \mathbf{H}_\ell^\top \cdot \mathbf{W}_\ell \cdot \mathbf{H}_\ell + \rho^{(t)} \mathbf{I} \\ b_\ell^{(t)} = \alpha \mathbf{H}_\ell^\top \cdot \mathbf{W}_\ell \cdot \mathbf{y}_\ell + \rho^{(t)} \tilde{z}_\ell^{(t)} \end{cases}$$

Updating the Variables

- objective:

$$\mathbf{x}^{(t)} = \arg \min_{\mathbf{x} \in \mathbb{X}} \left\{ f_{\text{prior}}(\mathbf{x}) + \frac{\rho^{(t)}}{2} \|\mathbf{x} - \tilde{\mathbf{x}}^{(t)}\|_2^2 \right\}$$

with: $\tilde{\mathbf{x}}^{(t)} = \mathbf{z}^{(t-1)} - \mathbf{u}^{(t-1)} / \rho^{(t)}$.

- for many different non-differentiable $f_{\text{prior}}(\mathbf{x})$ analytical expressions of the solution exist for $\mathbb{X} = \mathbb{R}^N$ (Combettes and Pesquet 2011);
- easy to adapt for $\mathbb{X} = \mathbb{R}_+^N$:
 - when $f_{\text{prior}}(\mathbf{x}) = f_{\text{sparse}}(\mathbf{x}) = \|\mathbf{x}\|_1$:

$$x_{n,\ell}^{(t)} = \max(0, \tilde{x}_{n,\ell}^{(t)} - 1/\rho^{(t)})$$

- when $f_{\text{prior}}(\mathbf{x}) = f_{\text{group}}(\mathbf{x}) = \sum_n \sqrt{\sum_\ell x_{n,\ell}^2}$:

$$x_{n,\ell}^{(t)} = \begin{cases} (1 - 1/\beta_n^{(t)}) \max(0, \tilde{x}_{n,\ell}^{(t)}) & \text{if } \beta_n^{(t)} > 1 \\ 0 & \text{else} \end{cases}$$

with:

$$\beta_n^{(t)} = \rho^{(t)} \sqrt{\sum_\ell \max(0, \tilde{x}_{n,\ell}^{(t)})}$$

Conclusions

Conclusions and Perspectives

Current Work:

- exploiting multi-variate data is powerful but require to account for *transverse* constraints (e.g. spectral **and** spatial regularization);
- non-smooth regularization requires specialized algorithms (e.g. a new algorithm based on ADMM strategy dedicated to multi-spectral observations of point-like sources);
- 3-D image reconstruction with a few 10^6 parameters is not an issue;

Future Work:

- consider more realistic data (non-linear model) and noise statistics;
- debiasing of the result;
- compare with multi-spectral matching-pursuit method (a kind of multi-spectral CLEAN);
- implement other regularizations (remember that means *new algorithms*):
 - explicit spectral continuity ($\ell_2 - \ell_1$ smoothness);
 - other spatial regularizations suitable for non-sparse sources (e.g. Total Variation);
- implement spectral interpolation;

References

- Boyd, S., Parikh, N., Chu, E., Peleato, B., and Eckstein, J.: 2010, *Foundations and Trends in Machine Learning* **3**, 1
- Combettes, P. L. and Pesquet, J.-C.: 2011, *Proximal splitting methods in signal processing*, Chapt. Fixed-Point Algorithms for Inverse Problems in Science and Engineering, pp 185–212, Springer, New York
- Donoho, D.: 2006, *Communications on Pure and Applied Mathematics* **59(7)**, 907
- Fornasier, M. and Rauhut, H.: 2008, *SIAM Journal on Numerical Analysis* **46(2)**, 577
- Hestenes, M. R. and Stiefel, E.: 1952, *Journal of Research of the National Bureau of Standards* **49(6)**, 409
- Jacoby, G. H., Hunter, D. A., and Christian, C. A.: 1984, *Astrophys. J. Suppl.* **56**, 257
- Lacour, S., Meimon, S., Thiébaud, E., Perrin, G., Verhoelst, T., Pedretti, E., Schuller, P. A., Mugnier, L., Monnier, J., Berger, J., Haubois, X., Poncelet, A., le Besnerais, G., Eriksson, K., Millan-Gabet, R., Lacasse, M., and Traub, W.: 2008, *Astron. Astrophys.* **485**, 561
- Rudin, L., Osher, S., and Fatemi, E.: 1992, *Physica D* **60**, 259–
- Soulez, F., Thiébaud, , Bongard, S., and Bacon, R.: 2011, in *3rd IEEE WHISPERS*, Lisbon, Portugal