Bayesian Reconstruction of Hyperspectral Images by Using CS Measurements and a Local Structured Prior

ICASSP - New Orleans

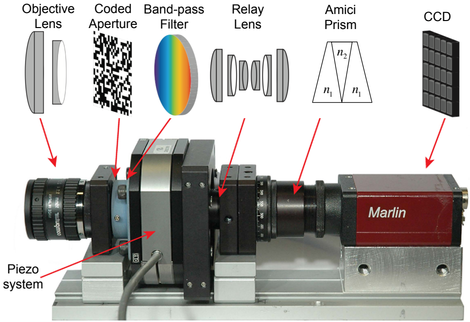
Facundo Costa, Jean-Yves Tourneret, Hadj Batatia [(1)],
Yuri Mejía, Henry Arguello [(2)]

(1) University of Toulouse, ENSEEIHT-IRIT-TéSA, France
(2) Universidad Industrial de Santander, Colombia
jyt@n7.fr

March 2017

Outline

Compressive Spectral Imaging (CSI)



Coded Aperture Snapshot Spectral Imaging (CASSI)

Compressive Spectral Imaging (CSI): recovering the full spatial and spectral information of a scene from undersampled random projections acquired by a compressive spectral imager such as CASSI.

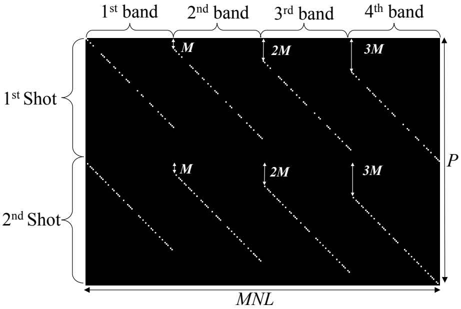
Compressive Sensing Measurements

Sensing matrix

$$\boldsymbol{y} = \boldsymbol{\Phi}\boldsymbol{x} + \boldsymbol{e}$$

where $\boldsymbol{\Phi}$ is fixed and $\boldsymbol{e}$ is an additive Gaussian noise, i.e., $\boldsymbol{e} \sim \mathcal{N}(0, \sigma_n^2)$

- Diagonal patterns related to the coded aperture
- Shifted patterns due to the prism effect
- Possible acquisition of multiple snapshots



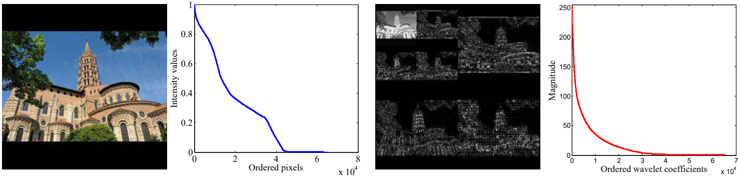
Compressive Sensing Measurements

Sparse representation of the image

$$\boldsymbol{x} = \boldsymbol{\Psi}\boldsymbol{\theta}$$

where $\boldsymbol{\Psi}$ is constructed from predefined atoms

e.g., using the wavelet transform



Problem: how to estimate the unknown image $\boldsymbol{x}$ from compressed measurements $\boldsymbol{y} = \boldsymbol{\Phi}\boldsymbol{x} + \boldsymbol{e}$?

Fusion as an Inverse Problem

Data fidelity term

$$\frac{1}{2}\|\boldsymbol{y}-\boldsymbol{\Phi}\boldsymbol{x}\|_2^2=\frac{1}{2}\|\boldsymbol{y}-\boldsymbol{H}\boldsymbol{\theta}\|_2^2$$

Sparse regularization

$$\varphi_1(\boldsymbol{\theta})=\|\boldsymbol{\theta}\|_1$$

Spatial regularization

$$\varphi_2(\boldsymbol{\theta})=\|(\boldsymbol{B}-\boldsymbol{I})\boldsymbol{\Psi}\boldsymbol{\theta}\|_2^2$$

where $\boldsymbol{B}$ is an appropriate weighting matrix (low-pass filter)

Conclusion

$$\arg\min_{\boldsymbol{\theta}}\left[\frac{1}{2}\|\boldsymbol{y}-\boldsymbol{H}\boldsymbol{\theta}\|_2^2+\tau\varphi_1(\boldsymbol{\theta})+\lambda\varphi_2(\boldsymbol{\theta})\right]$$

Outline

Bayesian LASSO[1]

Observation model

$$\boldsymbol{y} = \boldsymbol{H}\boldsymbol{\theta} + \boldsymbol{n}$$

where $\boldsymbol{\theta}$ is sparse and $n \sim \mathcal{N}(0, \sigma_n^2)$

Optimization problem

$$\arg\min_{\boldsymbol{\theta}} \left[\frac{1}{2\sigma_n^2} ||\boldsymbol{y} - \boldsymbol{H}\boldsymbol{\theta}||^2 + \tau ||\boldsymbol{\theta}||_1 \right]$$

Problem: how to adjust the regularization parameter τ?

Equivalent problem

$$\arg\max_{\boldsymbol{\theta}} \left[\exp\left(-\frac{1}{2\sigma_n^2} ||\boldsymbol{y} - \boldsymbol{H}\boldsymbol{\theta}||^2 \right) \exp(-\tau ||\boldsymbol{\theta}||_1) \right]$$

[1] T. Park and G. Casella, "The Bayesian Lasso," Journal of the American Statistical Association, vol. 103, no. 482, pp. 681-686, 2008.

Bayesian LASSO

Observation model

$$\boldsymbol{y} = \boldsymbol{H}\boldsymbol{\theta} + \boldsymbol{n}, \quad \text{sparse } \boldsymbol{\theta}, n \sim \mathcal{N}(0, \sigma_n^2 \boldsymbol{I}_N)$$

Bayesian formulation

- Gaussian likelihood

$$f(\boldsymbol{y}|\boldsymbol{\theta}) = \mathcal{N}(\boldsymbol{H}\boldsymbol{\theta}, \sigma_n^2 \boldsymbol{I}_N) \propto \exp\left(-\frac{1}{2\sigma_n^2}||\boldsymbol{y} - \boldsymbol{H}\boldsymbol{\theta}||^2\right)$$

- Independent Laplacian priors

$$f(\boldsymbol{\theta}|\tau) = \prod_{i=1}^{p} \exp(-\tau|\theta_i|) = \exp(-\tau||\boldsymbol{\theta}||_1)$$

- Posterior

$$f(\boldsymbol{\theta}|\boldsymbol{y}) \propto \exp\left(-\frac{1}{2\sigma_n^2}||\boldsymbol{y} - \boldsymbol{H}\boldsymbol{\theta}||^2\right) \exp(-\tau||\boldsymbol{\theta}||_1)$$

Hierarchical Bayesian Model

- Gaussian likelihood

$$f(\boldsymbol{y}|\boldsymbol{\theta}, \sigma_n^2) = \mathcal{N}(\boldsymbol{H}\boldsymbol{\theta}, \sigma_n^2 \boldsymbol{I}) \propto \exp\left(-\frac{1}{2\sigma_n^2}||\boldsymbol{y} - \boldsymbol{H}\boldsymbol{\theta}||^2\right)$$

- Independent Laplacian priors

$$f(\boldsymbol{\theta}|\tau) = \prod_{i=1}^{p} \exp(-\tau|\theta_i|) = \exp(-\tau||\boldsymbol{\theta}||_1)$$

- Joint noise variance and hyperparameter prior

$$\pi(\tau, \sigma_n^2)$$

- Posterior

$$f(\boldsymbol{\theta}, \sigma_n^2, \tau|\boldsymbol{y}) \propto \exp\left(-\frac{1}{2\sigma_n^2}||\boldsymbol{y} - \boldsymbol{H}\boldsymbol{\theta}||^2\right) \exp(-\tau||\boldsymbol{\theta}||_1)\pi(\tau, \sigma_n^2)$$

How can we estimate $\boldsymbol{\theta}, \sigma_n^2, \tau$?

Outline

The Bayesian LASSO

Posterior

$$f(\boldsymbol{\theta}, \sigma_n^2, \tau | \boldsymbol{y}) \propto \exp\left(-\frac{1}{2\sigma_n^2}||\boldsymbol{y} - \boldsymbol{H}\boldsymbol{\theta}||^2\right)\exp(-\tau||\boldsymbol{\theta}||_1)\pi(\tau, \sigma_n^2)$$

Completion

- Scale mixture of a Gaussians distributions

$$\frac{\tau}{2}e^{-\tau|\theta_i|} = \int_0^\infty \frac{1}{\sqrt{2\pi s}}e^{-\frac{\theta_i^2}{2s}}\frac{\tau^2}{2}e^{-\frac{\tau^2 s}{2}}ds$$

- Hierarchical representation

$$\begin{aligned} \boldsymbol{y} &\sim \mathcal{N}(\boldsymbol{y}; \boldsymbol{H}\boldsymbol{\theta}, \sigma_n^2\boldsymbol{I}_N) \\ \boldsymbol{\theta}|\sigma_n^2, s_1^2, ..., s_p^2 &\sim \mathcal{N}(\boldsymbol{\theta}; \mathbf{0}_p, \sigma_n^2\boldsymbol{D}_p),\ \boldsymbol{D}_p = \text{diag}(s_1^2, ..., s_p^2) \\ s_1^2, ..., s_p^2|\tau &\sim \prod_{j=1}^p\left(\frac{\tau^2}{2}e^{-\frac{\tau^2 s_j^2}{2}}\right), \quad \pi(\tau) \sim 1/\tau \\ \pi(\sigma_n^2) &\sim 1/\sigma_n^2 \quad \text{(Jeffreys prior)} \end{aligned}$$

Generalized Inverse Gaussian Distribution

$$\pi(x) = \left(\frac{a}{b}\right)^{1/4} K_{1/2}^{-1}\left(\sqrt{ab}\right) \frac{1}{\sqrt{x}} \exp\left[-\frac{1}{2}\left(\frac{b}{x} + ax\right)\right] I_{\mathbb{R}+}(x)$$

where $K_{1/2}$ is a modified Bessel function, hence

$$\begin{aligned} \int_0^\infty \frac{1}{\sqrt{x}} \exp\left[-\frac{1}{2}\left(\frac{b}{x} + ax\right)\right] dx &= \left(\frac{b}{a}\right)^{1/4} K_{1/2}\left(\sqrt{ab}\right) \\ &= \sqrt{\frac{\pi}{2}} \frac{1}{\sqrt{a}} \exp\left(-\sqrt{ab}\right). \end{aligned}$$

can be used to demonstrate that the Laplace distribution is a scale mixture of Gaussian distributions.

Gibbs Sampler

Algorithm 1 Gibbs sampler

```
Initialize τ and σ_n^2
Sample θ from its prior distribution
repeat
    for i = 1 to p do
        Sample s_i^2 from f(s_i^2 | θ_i, σ_n^2, τ)
    end for
    Sample θ from f(θ | y, σ_n^2, s^2)
    Sample τ from f(λ | θ)
    Sample a from f(a | δ^2)
    Sample σ_n^2 from f(σ_n^2 | y, θ, δ^2)
until convergence
```

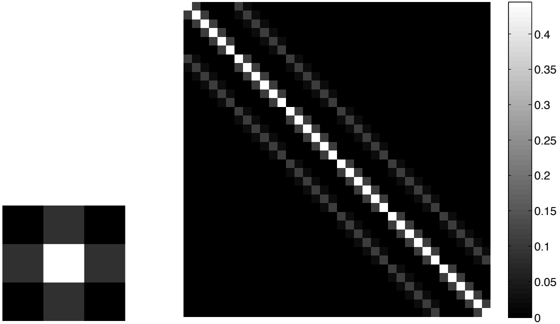
Outline

- 1: Compressive Sensing Measurements
- 2: Bayesian Formulation
- 3: An MCMC Method
- 4: Spatial Regularization
- 5: Simulation Results

Include Spatial Regularization into Bayesian LASSO

Optimization problem

$$\arg\min_{\boldsymbol{\theta}} \left[\frac{1}{2\sigma_n^2} ||\boldsymbol{y} - \boldsymbol{H}\boldsymbol{\theta}||^2 + \tau ||\boldsymbol{\theta}||_1 + \lambda ||(\boldsymbol{B} - \boldsymbol{I})\boldsymbol{\Psi}\boldsymbol{\theta}||^2 \right]$$



(a) (b)

(a) Zero mean Gaussian filter of size 3×3 with $\sigma = 0.6$, (b) matrix B created by using the Gaussian filter of (a).

Include Spatial Regularization into the Bayesian LASSO

Bayesian formulation

- Equivalent problem

$$\arg\max_{\boldsymbol{\theta}} \left[\exp\left(-\frac{1}{2\sigma_n^2}||\boldsymbol{y}-\boldsymbol{H}\boldsymbol{\theta}||^2 - \tau||\boldsymbol{\theta}||_1\right)\exp(-\lambda||(\boldsymbol{B}-\boldsymbol{I})\boldsymbol{\Psi}\boldsymbol{\theta}||^2)\right]$$

- Our proposal

Gibbs Sampler

Algorithm 2 Gibbs sampler

Initialize a, σ_n^2 and λ
Sample $\boldsymbol{\theta}$ from its prior distribution
repeat
  for $i = 1$ to N **do**
    Sample δ_i^2 from $f(\delta_i^2|\theta_i, \sigma_n^2, a)$
  end for
  Sample $\boldsymbol{\theta}$ from $f(\boldsymbol{\theta}|\boldsymbol{y}, \sigma_n^2, \boldsymbol{\delta}^2, \lambda)$
  Sample λ from $f(\lambda|\boldsymbol{\theta})$
  Sample a from $f(a|\boldsymbol{\delta}^2)$
  Sample σ_n^2 from $f(\sigma_n^2|\boldsymbol{y}, \boldsymbol{\theta}, \boldsymbol{\delta}^2)$
until convergence

Conditional Distributions of $f(\sigma_n^2, \boldsymbol{\theta}, a, \lambda, \delta_i^2|\boldsymbol{y})$

Full conditionals $f(\delta_i^2|\theta_i, \sigma_n^2, a)$, $f(\boldsymbol{\theta}|\boldsymbol{y}, \sigma_n^2, \boldsymbol{\delta}^2, \lambda)$, $f(\lambda|\boldsymbol{\theta})$, $f(a|\boldsymbol{\delta}^2)$ and $f(\sigma_n^2|\boldsymbol{y}, \boldsymbol{\theta}, \boldsymbol{\delta^2})$ associated with the posterior distribution of interest.

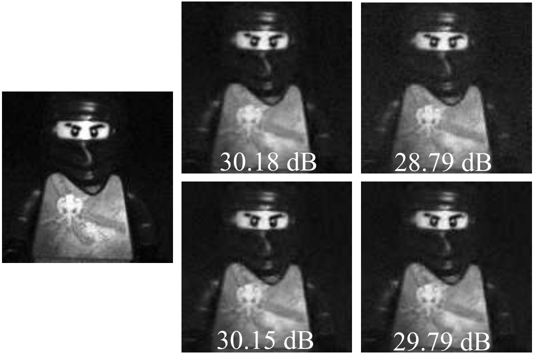
δ_i^2	$\mathcal{GIG}\left(\frac{1}{2}, a, \frac{\theta_i^2}{\sigma_n^2}\right)$
$\boldsymbol{\theta}$	$\mathcal{N}\left(\frac{\boldsymbol{\Sigma H}^T\boldsymbol{y}}{\sigma_n^2}, \boldsymbol{\Sigma}\right), \boldsymbol{\Sigma}^{-1} = \frac{1}{\sigma_n^2}(\boldsymbol{H}^T\boldsymbol{H} + \boldsymbol{\Delta}^{-1}) + \lambda\boldsymbol{C}^{-1}$
λ	$\mathcal{G}\left(\frac{NML}{2} + \alpha_\lambda, \frac{\|\|(\boldsymbol{B}-\boldsymbol{I})\boldsymbol{\Psi\theta}\|\|^2}{2} + \beta_\lambda\right)$
a	$\mathcal{G}\left(NML, \frac{\|\|\boldsymbol{\delta}\|\|^2}{2}\right)$
σ_n^2	$\mathcal{IG}\left(\frac{NML+P}{2}, \frac{1}{2}\left[\|\|\boldsymbol{y} - \boldsymbol{H\theta}\|\|^2 + \sum \frac{\theta_i^2}{\delta_i^2}\right]\right)$

Sampling $\boldsymbol{\theta}$ using a perturbation-optimization algorithm[2]

[2] F. Orieux, O. Feron and J. F. Giovannelli, "Sampling High-Dimensional Gaussian Distributions for General Linear Inverse Problems," IEEE Signal Processing Letters, vol. 19, no. 5, pp. 251-254, May 2012.

Outline

Qualitative Results



Seventh spectral band of the image: (Left) Ground truth. Reconstruction results for: (top center) the proposed method, (bottom center) SpaRSA[3] Smooth, (top right) Bayesian LASSO and (bottom right) SpaRSA LASSO.

[3] S. J. Wright, R. D. Nowak, and M. A. T. Figueiredo, "Sparse reconstruction by separable approximation," IEEE Transactions on Signal Processing, vol. 57, no. 7, pp. 2479—2493, July 2009.

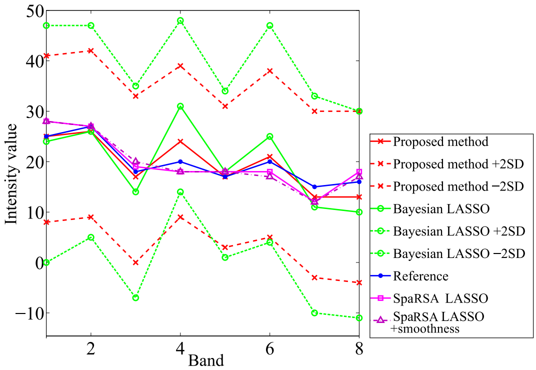


Figure: Spectral signature for pixel $\#(20, 33)$.

- The estimates obtained using the smoothing term are closer to the ground truth.
- Bayesian methods provide confidence measures for the estimates

Conclusions and Future Work

Conclusions

- Hierarchical Bayesian model solving the compressive spectral imaging problem by promoting the image to be sparse in a given basis and smooth in the spatial domain.
- A Gibbs sampler sampling the full image in a single step using a perturbation optimization algorithm
- Including a spatial smoothing term can improve the PSNR of the recovered image up to 2dB.

Prospects

- Other regularization terms: Total Variation? l_p regularization?
- Analyze the effects of the sensing matrix on the reconstruction performance and design an optimal sensing matrix

Thanks

Basis Representation

Basis representation

$$\mathbf{\Psi} = \mathbf{\Psi}_1 \otimes \mathbf{\Psi}_2 \otimes \mathbf{\Psi}_3$$

- $\mathbf{\Psi}_1 \otimes \mathbf{\Psi}_2$: 2D-Wavelet Symmlet 8 basis
- $\mathbf{\Psi}_3$: cosine basis.

PSNRs for Different Reconstruction Algorithms

Compression ratio	13%	26%	40%	53%	66%
Proposed method	24.4	27.1	28.6	29.6	30.4
Bayesian LASSO	22.9	26.0	27.5	28.4	28.4
SpaRSA smooth	25.2	27.1	28.8	29.7	30.6
SpaRSA LASSO	23.5	26.8	28.5	29.4	30.4

Table: PSNRs [dB] obtained by the different algorithms.

Computational Cost

Computational cost	Iterations	Seconds
Proposed method	500	20×10^3
Bayesian LASSO	750	18×10^3
SpaRSA smooth	300	316
SpaRSA LASSO	300	42

Table: Computational costs for a 53% compression ratio.