Stochastic Filtering via Reweighted-ℓ_1

Adam Charles and Christopher J. Rozell
Georgia Institute of Technology

Introduction

Sparsity Models have proven indispensable in estimating single signals in measurement poor applications [1].

However, many applications require recovery of correlated sparse signals (e.g. [2]):

- Recover: a set of signals $\boldsymbol{x}_k \in \mathbb{R}^N$ where $k \in K \subset \mathbb{Z}^D$
- Given: measurements $\boldsymbol{y}_k \in \mathbb{R}^M$, $(M < N)$:

$$\boldsymbol{y}_k = \boldsymbol{\Phi}_k \boldsymbol{x}_k + \boldsymbol{\epsilon}_k$$

$\boldsymbol{\Phi}_k$: measurement matrix $\quad\boldsymbol{\epsilon}_k$: noise

Sparsity Model:

- $\boldsymbol{x}_k = \boldsymbol{\Psi}\boldsymbol{a}_k$, $\boldsymbol{a}_k$ is mostly zeros
- $p(\boldsymbol{a}_k)$ is highly kurtotic (e.g. Laplacian)
- Recover each $\boldsymbol{a}_k$ independently

E.g. MAP estimate with Laplacian priors:

$$\widehat{\boldsymbol{a}}_k = \arg\min_{\boldsymbol{a}} \|\boldsymbol{y}_k - \boldsymbol{\Phi}\boldsymbol{\Psi}\boldsymbol{a}\|_2^2 + \gamma\|\boldsymbol{a}\|_1$$
$$\widehat{\boldsymbol{x}}_k = \boldsymbol{\Psi}\widehat{\boldsymbol{a}}_k$$

Correlation Model:

- $\boldsymbol{x}_k$ depends probabilistically on $\boldsymbol{x}_{l\neq k}$

$$p\left(\boldsymbol{x}_k\right) = h\left(\boldsymbol{x}_{l\neq k}\right)$$

- Ideally, h is local (only needs l : $\|l-k\|_q < r$)
- MAP recovery requires joint or conditional distributions over K

We propose a hierarchical model to robustly recover correlated sparse signals

Reweighted-ℓ_1 Framework

We propose stochastic filtering using second order moments native to sparse signal estimation. Based on reweighted-l1 optimization [3], using weights to propagate information, e.g. [4].

Reweighted-l1 optimization can be viewed as a second order model for sparse signals.

- Gaussian Measurements:

$$p(\boldsymbol{y}_k|\boldsymbol{a}_k) \sim \mathcal{N}(\boldsymbol{\Phi}\boldsymbol{\Psi}\boldsymbol{a}_k, \sigma_\epsilon^2\boldsymbol{I})$$

- Laplacian Conditional:

$$p(\boldsymbol{a}_k[i]|\boldsymbol{\lambda}_k[i]) = \frac{\lambda_0\boldsymbol{\lambda}_k[i]}{2}e^{-\lambda_0\boldsymbol{\lambda}_k[i]|\boldsymbol{a}_k[i]|}$$

- Gamma Hyperprior on Variance:

$$p(\boldsymbol{\lambda}_k[i]) = \frac{\boldsymbol{\lambda}_k^{\alpha-1}[i]}{\boldsymbol{\theta}_k^{\alpha}[i]\Gamma(\alpha)}e^{-\boldsymbol{\lambda}_k[i]/\boldsymbol{\theta}_k[i]}$$

- Marginal Prior

$$p(\boldsymbol{a}_k[i]) = \frac{\alpha\lambda_0\boldsymbol{\theta}_k[i]}{2(\lambda_0\boldsymbol{\theta}_k[i]|\boldsymbol{a}_k[i]|+1)^{\alpha+1}}$$

- Variance parameters can be correlated via $\boldsymbol{\theta}_k$:

$$\boldsymbol{\theta}_k = \xi\left(g\left(\boldsymbol{a}_{l\in K}\right)+\eta\right)$$

- Expectation Maximization approach to optimization

Algorithm: iterate until convergence

M-step:

$$\widehat{\boldsymbol{a}}_k^t = \arg\min_{\boldsymbol{a}}\left[\|\boldsymbol{y}_k - \boldsymbol{\Phi}\boldsymbol{\Psi}\boldsymbol{a}\|_2^2 + 2\sigma_\epsilon^2\lambda_0\sum\boldsymbol{\lambda}_k^{t-1}[i]|\boldsymbol{a}[i]|\right]$$

E-step:

$$\boldsymbol{\lambda}_k^t[i] = \frac{2\tau}{\beta|\widehat{\boldsymbol{a}}_k^t[i]|+|g(\widehat{a}_{l\in K})[i]|+\eta}$$

Propagating information in the second order moments allows for more robustness to model mismatch

Special Cases: RWL1-DF and RWL1-SF

Reweighted-l1 Dynamic Filtering (RWL1-DF):

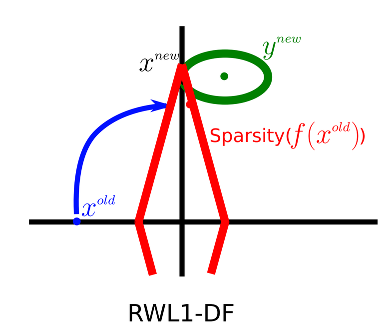
- Temporally located/correlated signals $(K \subset \mathbb{Z})$
- Example application: Compressive video recovery
- Temporal model: $\boldsymbol{x}_k = f(\boldsymbol{x}_{k-1}) + \boldsymbol{\nu}_k$
- Time Dependent Variance:

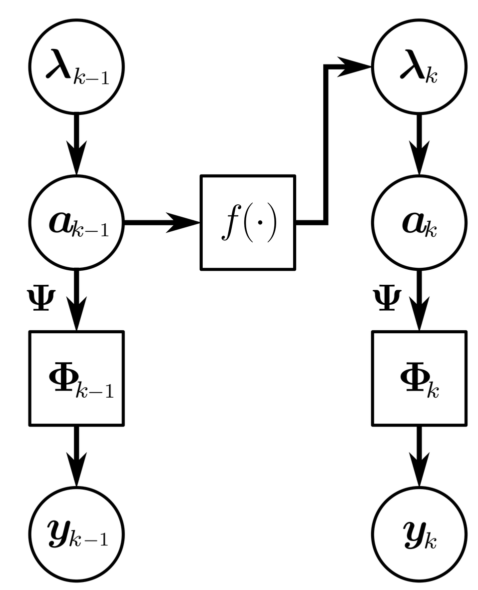
$$\boldsymbol{\theta}_k[i] = \frac{\xi}{|[\boldsymbol{\Psi}^T f(\boldsymbol{\Psi}\widehat{\boldsymbol{a}}_{k-1})][i]|+\eta}$$

- Exploit correlation for causal filtering (e.g. [5])

Algorithm: E-step becomes

$$\boldsymbol{\lambda}_k^t[i] = \frac{2\tau}{\beta|\widehat{\boldsymbol{a}}_k^t[i]|+|[\boldsymbol{\Psi}^T f(\boldsymbol{\Psi}\widehat{a}_{k-1})][i]|+\eta}$$





Reweighted-l1 Spatial Filtering (RWL1-SF):

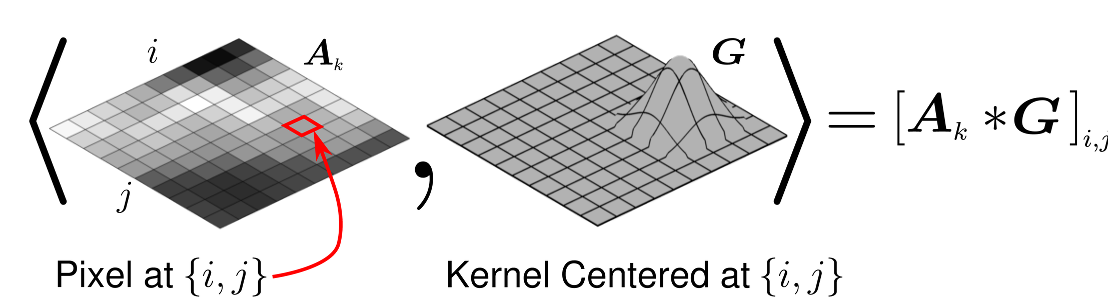
- Spatially located/correlated signals $(K \subset \mathbb{Z}^2)$
- Example application: Hyperspectral imagery (HSI)
- Spatial correlations defined by a kernel $\boldsymbol{G}$
- Spatially Dependent Variance:

$$\boldsymbol{\theta}_{i,j}[l] = \frac{\xi}{|\sum_m\sum_n \boldsymbol{G}_{i-n,j-m}\widehat{\boldsymbol{a}}_{i,j}[l]|+\eta}$$

- Exploit correlation for spatial filtering

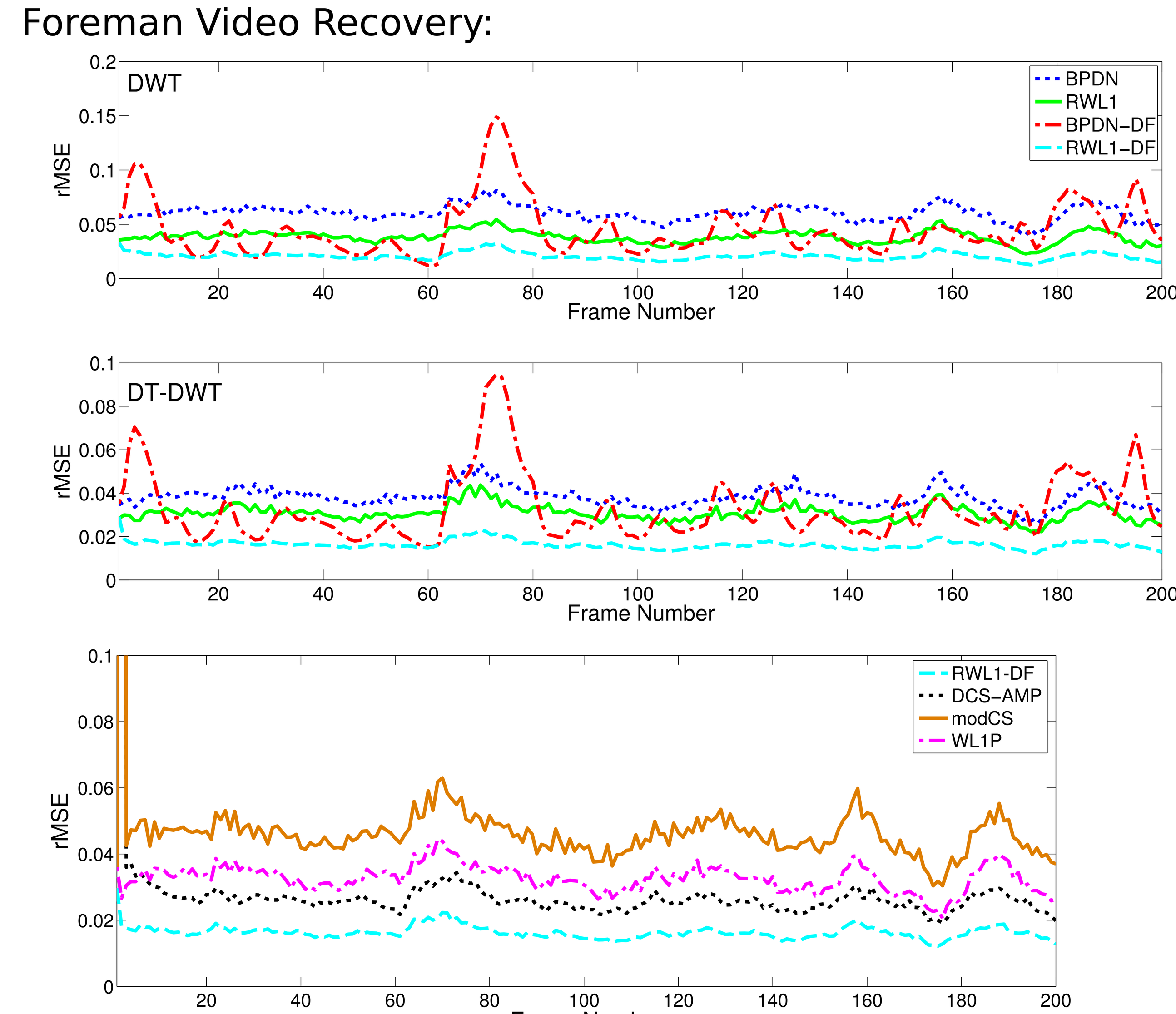
Algorithm: E-step becomes

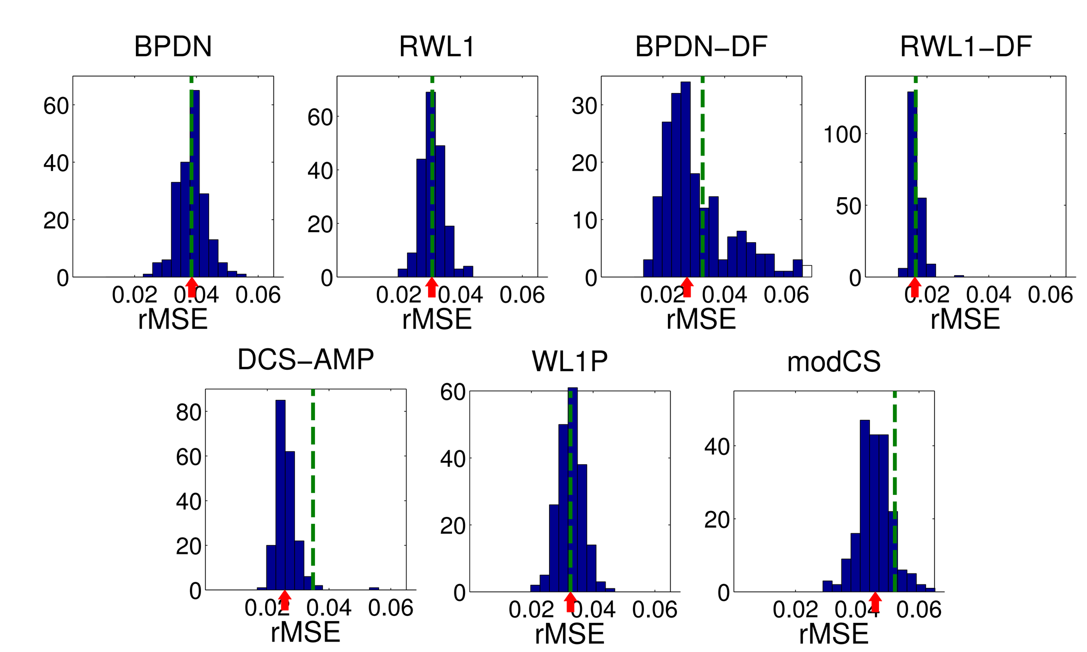
$$\boldsymbol{\lambda}_{i,j}^t[l] = \frac{2\tau}{|[\boldsymbol{G}*\widehat{\boldsymbol{A}}_l]_{i,j}|+\eta}$$



RWL1-DF: Compressive Video Recovery

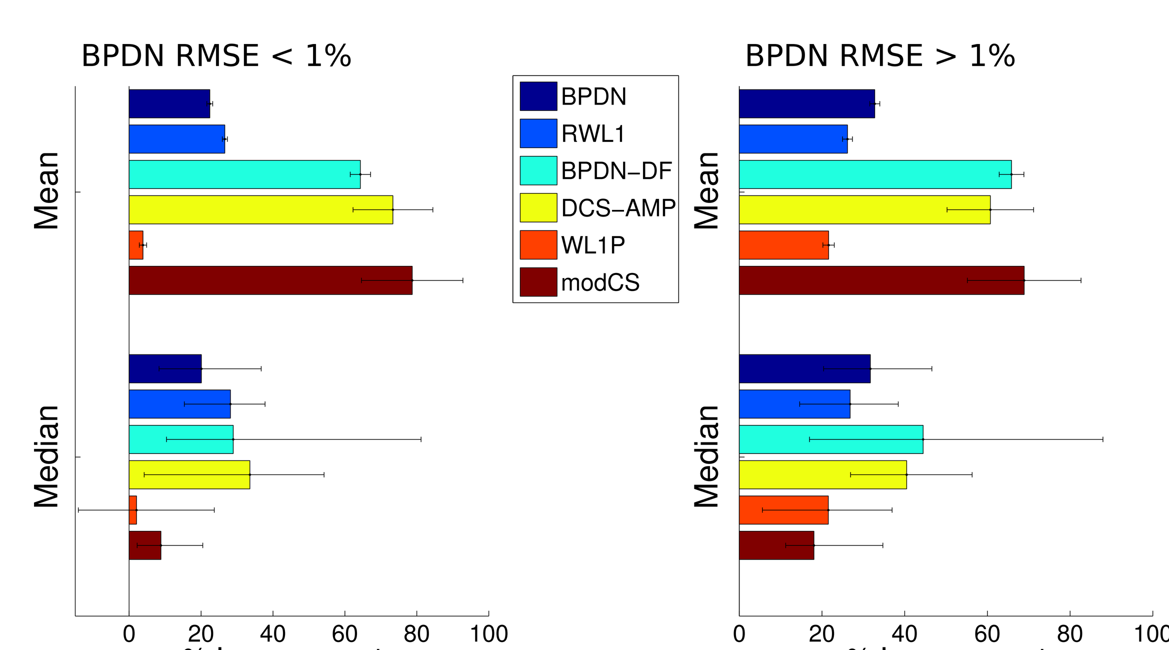
Foreman Video Recovery:





The Foreman video sequence recovered from randomly selected noiselet measurements (M = 0.25N) [7]. Comparisons shown to BPDN, BPDN-DF [6], RWL1 [3], DCS-AMP [8], and modCS [9]. .

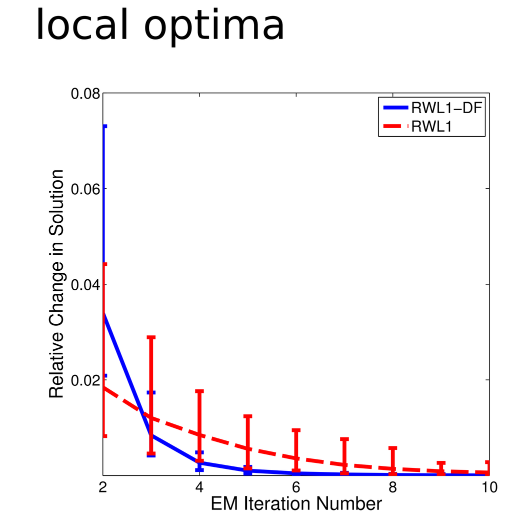
BBC "Natural" Video Recovery:



RWL1-DF recovers natural movies (segments of a BBC documentary) more accurately than competing algorithms Comparisons shown to DCS-AMP [8], modCS [9], RWL1 [3], BPDN-DF [4].

Convergence Notes:

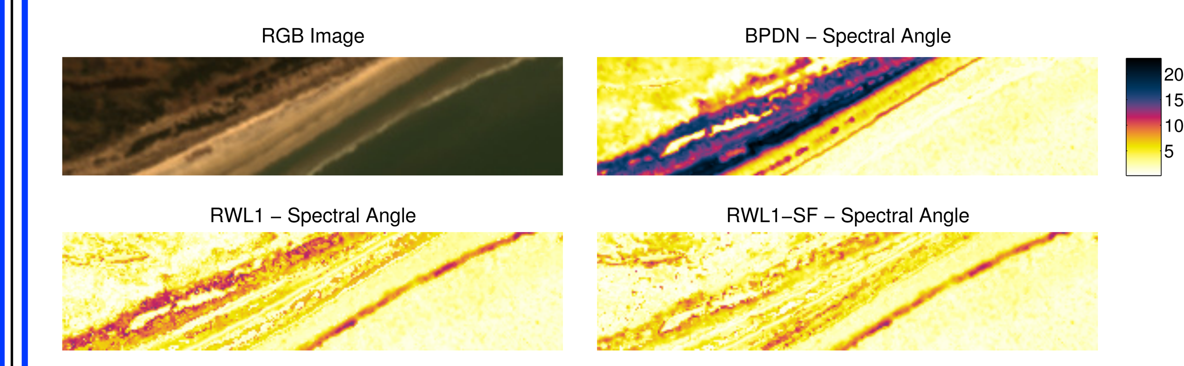
Expectation maximization ensures convergence to a local optima

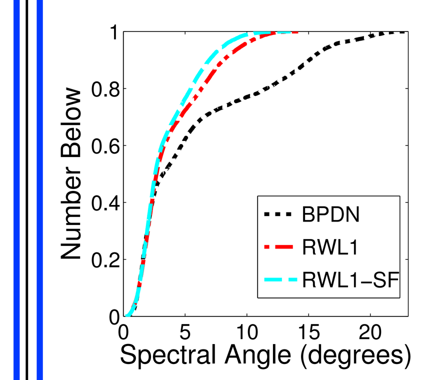


Empirically RWL1-DF converges faster than RWL1.

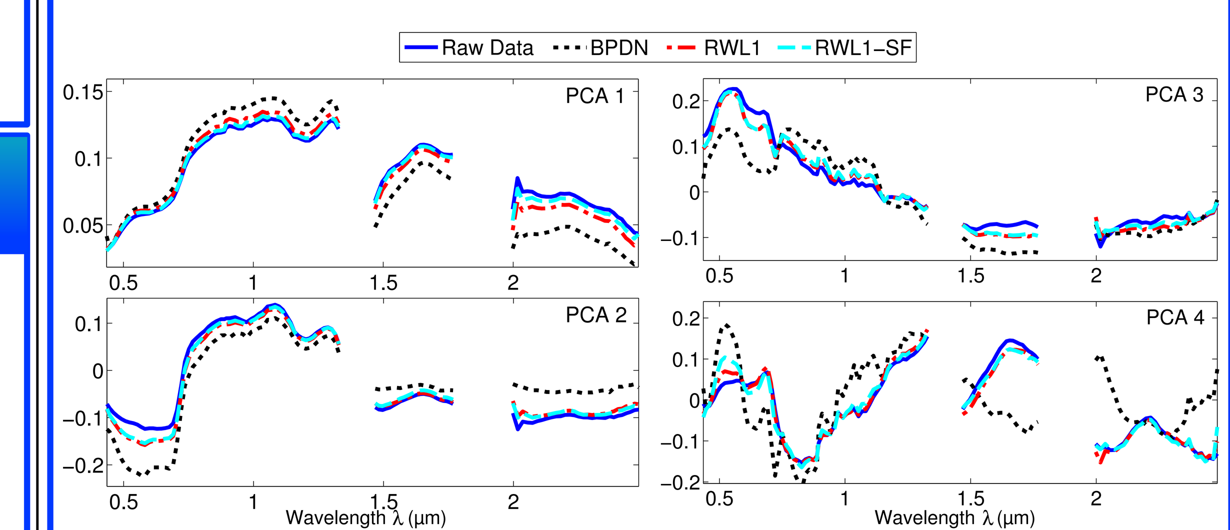
RWL1-SF: HSI Spectral SR

Task: super-resolution (SR) of multispectral imagery to hyperspectral imagery: inpainting+deblurring





A segment of the Smith Island HSI dataset (113 bands) recovered from 8 coarse measurements [10] using BPDN, RWL1 and RWL1-SF



Conclusions

Second order models allow for joint recovery of correlated, sparse signals.

- Improved estimation quality (RMSE)
- Robustness to innovations' statistics
- Computational complexity no more than a number of BPDN solutions

Future directions:

- Understand theoretical limits of the algorithm

References:

[1] M. Elad, M. Figueiredo, and Y. Ma, “On the role of sparse and redundant representations in image processing,” IEEE Proceedings - Special Issue on Applications of Compressive Sensing & Sparse Representation, Oct 2008..

[2] M. S. Asif, L. Hamilton, M. Brummer, and J. Romberg, “Motion-adaptive spatio-temporal regularization (master) for accelerated dynamic mri,” 2012, Submitted.

[3] E. Candes, M. B. Wakin, and S. P. Boyd, “Enhancing sparsity by reweighted l1 minimization,” Jour. of Fourier Anal. and App., vol. 14, no. 5, pp. 877–905, 2008.

[4] M. A. Khajehnejad, W. Xu, S. Avestimehr, and B. Hassibi, “Weighted l1 minimization for sparse recovery with prior information,” http://arxiv.org/abs/0901.2912v1, 2009.

[5] R. E. Kalman, “A new approach to linear filtering and prediction problems,” Transactions of the ASME-Journal of Basic Engineering, vol. 82, no. D, pp. 35–45, 1960.

[6] A. Charles, M. S. Asif, J. Romberg, and C. Rozell, “Sparsity penalties in dynamical system estimation,” Proc of the CISS, March 2011..

[7] R. Coifman, F. Geshwind, and Y. Meyer, “Noiselets,” Applied Computational Harmonic Analysis, vol. 10, no. 1, pp. 27–44, 2001.

[8] J. Ziniel, L. C. Potter, and P. Schniter, “Tracking and smoothing of time-varying sparse signals via approximate belief propagation,” Proc. of the Asilomar Conference on Signals, Systems and Computers, 2010.

[9] N. Vaswani and W. Lu, “Modified-cs: Modifying compressive sensing for problems with partially known support,” IEEE Trans. on Sig. Proc., vol. 58, no. 9, pp. 4595–4607, 2010.

[10] A. S. Charles, B. A. Olshausen and C. J. Rozell. Learning sparse codes for hyperspectral images. IEEE Journal of Selected Topics in Signal Processing, 5(5):963-978, 2011

Acknowledgements

This work was partially supported by NSF grant CCF-0830456

Contact

{acharles6,crozell}@gatech.edu