

HYPERSPECTRAL IMAGE COMPRESSED SENSING VIA LOW-RANK AND JOINT-SPARSE MATRIX RECOVERY

Mohammad Golbabaee and Pierre Vandergheynst

Signal Processing Institute, Ecole Polytechnique Fédérale de Lausanne (EPFL), Switzerland
E-mail: {mohammad.golbabaee, pierre.vandergheynst}@epfl.ch

ABSTRACT

We propose a novel approach to reconstruct Hyperspectral images from very few number of noisy compressive measurements. Our reconstruction approach is based on a convex minimization which penalizes both the nuclear norm and the $\ell_{2,1}$ mixed-norm of the data matrix. Thus, the solution tends to have a simultaneous low-rank and joint-sparse structure. We explain how these two assumptions fit Hyperspectral data, and by several simulations we show that our proposed reconstruction scheme significantly enhances the state-of-the-art tradeoffs between the reconstruction error and the required number of CS measurements.

Index Terms— Hyperspectral images, Compressed sensing, Joint sparse signals, Low rank matrix recovery, Nuclear norm

1. INTRODUCTION

Hyperspectral Images (HSI) are huge collection of images that have been acquired simultaneously from a scene in a few hundred narrow adjacent frequency bands. All substances have their own specific *spectral signature* or frequency absorption features and therefore once the frequency bands are sampled with highly enough resolution, Hyperspectral imagery will be a very powerful tool for characterizing the components of the observed scenes. As a result, these type of images are found in a wide variety of applications in remote sensing such as detection and identification of the ground surface as well as atmospheric composition, analysis of soil type, agriculture, mineral exploration and environmental monitoring (e.g., oil/gas leakage from pipelines or natural wells). The price to pay for such high spatio-spectral resolution is to handle extremely large data size. For example, each instance of the HSI acquired by the NASA's Airborne Visible InfraRed Imaging Spectrometer (AVIRIS) contains more than 140 MBytes of data. Such enormous amount of information brings serious challenges, particularly, to the embedded systems, such as spacecrafts, where the power consumption, memory storage, computational complexity and bandwidth are posing tight constraints on system implementation.

In this regard and over the last twenty years, many compression methods have been developed to reduce the size of HSI, prior to storage or transmission. The key observation among all those work is the fact that, despite the huge size of HSI, there exist massive correlations both spatially and spectrally, thus exploiting them properly enables one to design efficient compression algorithms. One of the most efficient approaches [1] consists in 2D wavelet coding for the spatial domain, as the natural images can be typically represented

by few sparse wavelet coefficients. In addition, a Karhunen-Loève transform (KLT) is applied to compress data into few principal components along the spectral dimension. The KLT-based approach is data-dependent and in practice costs heavy computations and transmission (to the decoder) of the correlation matrix, however its high efficiency reveals an important point about the data structure and that is, HSI typically have very few principal components and thus they are low-rank. We will discuss in further details in Section 2 that *hyperspectral images are low-rank, and in addition they have a joint-sparse spatial wavelet representations*.

An $n_1 \times n_2$ matrix is *joint-sparse* if only few $k \ll n_1$ number of its rows contain nonzero elements. Once the indices of those rows are known and the matrix has rank $r \ll \min(n_1, n_2)$, the whole data will have no more than $r(k + n_2 - r)$ degrees of freedom, that is much less than its actual size $n_1 n_2$. Regarding this, one major question would naturally arise; Why do we need to make an effort to acquire the whole HSI with such enormous redundancy? It is important to note that such wasteful acquisition brings tremendous constraints to the onboard implementation in terms of power consumption both at the sampling and compression steps, and it also requires a large number of expensive photodiodes sensitive to the invisible light range. Therefore, there is a strong motivation to move toward designing more advanced techniques of sampling that economize as much as possible the available resources.

1.1. Compressive Sampling and Low-Rank Matrix Recovery

Compressive sampling (CS) theory has been introduced as an alternative to the Shannon's sampling theorem, and it can be viewed as merging the two conventional sampling and compression blocks together in order to build acquisition systems achieving sampling rates much lower than the Nyquist rate [2] [3]. Instead of taking n periodic Nyquist samples to discretize data into the vector $x \in \mathbb{R}^n$, $m \ll n$ linear measurements are collected from data and without any further compression they are forwarded to the decoder. The set of measurement $y \in \mathbb{R}^m$ can be expressed in the matrix form $y = Ax + z$, where each measurement is the inner product of x and a row of the measurement matrix $A \in \mathbb{R}^{m \times n}$, and $z \in \mathbb{R}^m$ represents the noise vector due to the quantization, transmission, etc. It has been shown that, if the original signal has a *sparse* representation in an orthonormal basis (i.e., $x = \Phi\theta$, $\Phi \in \mathbb{R}^{n \times n}$, $\theta \in \mathbb{R}^n$), and if the measurement matrix is drawn at random from certain distributions (e.g., A being i.i.d subgaussian matrix), then a robust reconstruction is achievable by solving the following convex optimization, called the *Basis Pursuit Denoising* (BPDN):

$$\arg \min_{\theta \in \mathbb{R}^n} \|\theta\|_{\ell_1} \quad \text{subject to} \quad \|y - A\Phi\theta\|_{\ell_2} \leq \varepsilon. \quad (1)$$

In this case, the required number of measurements are $m \geq O(k \log(n/k))$, where $k \ll n$ denotes the number of the nonzero

This research was supported by the Swiss National Science Foundation through grant 200021-117884 and the EU Framework 7 FET-Open project FP7-ICT-225913-SMALL.

elements of θ . Lately, many other works have been focusing on enhancing the reconstruction by enforcing some additional a priori *structures* among coefficients e.g., replacing ℓ_1 by $\ell_{2,1}$ for joint-sparse data (will be defined in the next sections).

Similar results have been obtained regarding compressed sampling of low-rank matrices. Assume a linear mapping $\mathcal{A} : \mathbb{R}^{n_1 \times n_2} \rightarrow \mathbb{R}^m$ applies on a rank $r \ll \min(n_1, n_2)$ data matrix $X \in \mathbb{R}^{n_1 \times n_2}$ to collect $m \ll n_1 n_2$ linear measurements $y \in \mathbb{R}^m$ i.e.,

$$y = \mathcal{A}(X) + z. \quad (2)$$

Candés and Plan [4] show that, for an \mathcal{A} which is drawn at random (e.g. i.i.d. subgaussian ensemble), a robust recovery of X is achievable from the following convex *nuclear norm minimization*

$$\underset{X \in \mathbb{R}^{n_1 \times n_2}}{\operatorname{argmin}} \|X\|_* \quad \text{subject to} \quad \|y - \mathcal{A}(X)\|_{\ell_2} \leq \varepsilon, \quad (3)$$

and we require $m \geq \mathcal{O}(rn_1 + rn_2)$, that is the same order as the degrees of freedom of such low-rank matrix. Note that in (3), $\|\cdot\|_*$ stands for the nuclear norm of a matrix that is the sum of its singular values. The striking point here is that, unlike to the KLT-based compression, the sensor side does not need to know, compute or transmit (to the decoder) the observations correlation matrix. The sensor design remains universal for any data matrix (with any underlying correlation matrix), and as simple as acquiring few linear measurements, while at decoder side a more complex nonlinear approximation method applies to recover the underlying data.

1.2. Our Main Contributions

In the Hyperspectral imagery, the *asymmetric* complexity of the CS sampling-reconstruction suits perfectly the aforementioned practical requirements imposed by an onboard design. Recently, a few number of novel acquisition setups have been proposed based on the compressed sensing idea in order to acquire the HSI by very few number of measurements [5] [6]. For data reconstruction, the authors of [6] apply the TwIst algorithm [7] independently on each spectral band in order to find 2D images with few gradient variations. This approach clearly neglects the existing correlations in the spectral domain. In [8] the authors additionally take into account the piecewise smooth variation of HSI along the spectral domain and reconstruct a sparse representation of HSI in a 3D wavelet basis using the standard BPDN method i.e., Φ in (1) would be replaced by the *Kronecker* orthonormal basis composed of 2D spatial and 1D spectral wavelet basis. Other approaches have been recently developed in order to take advantage of both spectral and spatial correlations simultaneously, however none of them take advantage of one of the most influential priors for HSI compression, that is the low-rank structure of the data.

The main contribution of our work is developing a reconstruction method that simultaneously brings the two important priors of HSI into the consideration; (i) the sparse wavelet representation, and (ii) the low-rank structure due to the high correlations. These two assumptions significantly reduce the degrees of freedom of data, and neglecting one would demand much higher number of measurements for a decent CS recovery. Our proposed scheme is based on convex optimization and it recovers a low-rank *and* joint-sparse representation of data that fits well the CS measurements. We will show by number of simulations that the global minima of the proposed approach significantly enhances the HSI reconstruction, comparing to the prior arts. Notably and in contrast with HSI compression technique in [1], the sensor side does not need to compute or transmit any correlation matrix and its complexity reduces to taking few number of linear measurements and forward them to the decoder.

2. HYPERSPECTRAL DATA CORRELATION MODEL

In order to represent the Hyperspectral images we define a matrix $X \in \mathbb{R}^{n_1 \times n_2}$ where each column, say X_j , corresponds to a 2D spatial image (reshaped in a vector) in the corresponding spectral band j . We denote by n_2 the number of spectral bands and by n_1 the resolution of the spatial images per each band. Moreover, let us denote by $\Theta \in \mathbb{R}^{n_1 \times n_2}$ the matrix whose columns contains the spatial wavelet coefficients of X for all spectral bands i.e., $X_j = \Phi_{2D}\Theta_j$ and thus $X = \Phi_{2D}\Theta$, where Φ_{2D} is the corresponding 2D spatial wavelet basis. Typically there are high correlations among the spectral bands (columns of X), because the whole region of observation can be decomposed into few number of subregions (*source images*) that each contains a certain material with distinct spectral signature. Assume the observation region is composed of $\rho \ll \min(n_1, n_2)$ different elements. Let $H \in \mathbb{R}_+^{n_2 \times \rho}$ be the matrix whose columns contains spectral signatures of those materials along n_2 frequency bands, and $S \in \mathbb{R}_+^{n_1 \times \rho}$ be the matrix containing the corresponding source images i.e., S_i^n indicates the percentage of the material i in pixel n . Then the whole HSI matrix can be factorized into $X = SH^T$, that is known as the *linear mixture model*.

This simple observation justifies the key assumptions we made about the correlation model of HSI. Hyperspectral images are low-rank and spatially they have joint-sparse wavelet representations. Indeed, once a matrix can be factorized as above the rank would be bounded i.e., $\mathbf{Rank}(X) = r \leq \rho$. In addition, there are regularities in the source images so that all S_j 's appear to be spatially piecewise smooth and thus, they have sparse spatial wavelet representations. Therefore due to the linear mixture model, and once the sources are not many, the wavelet sparsity would propagate to the columns of Θ , and in addition all the columns would share a common *sparsity support* i.e., nonzero coefficients would share the same indices along the columns.

3. COMPRESSIVE HYPERSPECTRAL IMAGERS

Sensors are collecting $m \ll n_1 n_2$ linear measurements from the HSI matrix X in a vector $y \in \mathbb{R}^m$ the same as in (2). Note that, \mathcal{A} can be explicitly expressed by a matrix $A \in \mathbb{R}^{m \times n_1 n_2}$ through the equivalent expression $y = AX_{vec} + z$, wherein the columns of X are stacked into the vector X_{vec} . Several camera designs have been so far proposed for the single-channel image compressive acquisition. A common point among those is the use of a *random pattern* to modulate the light prior to the measurement collection. As an example, the random convolution measurement scheme have been proposed by [9], convolves the image light-field with a random pattern using few optical blocks, and finally a few number of random pixels have been acquired from the resulting modulation. All those setups can be easily extended to the hyperspectral imaging, by repeating the same acquisition scheme for all spectral bands, however by using an *independent random pattern per channel*. In this case the corresponding measurement matrix A , would be a block diagonal matrix of the form

$$A = \begin{bmatrix} A_1 & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & A_2 & \dots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \dots & A_J \end{bmatrix}, \quad (4)$$

wherein $A_j \in \mathbb{R}^{\hat{m} \times N}$ is the random measurement matrix applies on channel j independently from the other spectral bands, and \hat{m} denotes the number of measurements collected per channel i.e., $m =$

$\hat{m}J$. In contrast with the Single-pixel hyperspectral imager in [5] which uses a unique random pattern for all spectral bands (i.e., $A_1 = A_2 = \dots = A_J$), using independent blocks as in (4) leads the measurements to benefit more efficiently the existing information *diversity* across multiple spectral channels.

For more real-time acquisitions, another camera design have been implemented by Wagadarikar *et al.* The Coded Aperture Snapshot Spectral Imager (CASSI) captures a few thousands of CS measurements in snapshot. There, by using optical light modulators the whole spectral information are encoded into a single 2D spatial image, and thus the sampling matrix has a different form than in (4) (for more details see [6]).

4. LOW-RANK AND JOINT-SPARSE CS RECOVERY

We propose the following convex optimization for recovery of low-rank and joint-sparse matrices and thus, to reconstruct the wavelet coefficient matrix Θ of the HSI (equivalently X) from their CS measurements:

$$\begin{aligned} \arg \min_{\Theta} \quad & \|\Theta\|_* + \lambda \|\Theta\|_{\ell_{2,1}} \\ \text{subject to} \quad & \|y - \mathcal{A}(\Phi_{2D}\Theta)\|_{\ell_2} \leq \varepsilon. \end{aligned} \quad (5)$$

The $\ell_{2,1}$ mixed-norm is defined for a given matrix Θ as the sum of the ℓ_2 norm of its rows i.e., $\|\Theta\|_{\ell_{2,1}} = \sum_i \|\Theta^i\|_{\ell_2}$, where Θ^i is the i th row of Θ . As mentioned above, penalizing the Nuclear norm or the $\ell_{2,1}$ mixed-norm have been widely used in the literature to impose the reconstructed data to have a low-rank or joint-sparse structure, respectively. However, by penalizing both terms with a proper regularization factor λ in (5), we tend to impose the solutions to satisfy *simultaneously* both properties i.e., our principal priors for the HSI reconstruction. Therefore, applying this approach can efficiently take advantage of the limited degrees of freedom of such data structure in order to economize the number of measurements required for the HSI reconstruction. In order to solve (5), let us rewrite it as

$$\arg \min_{\Theta} f_1(\Theta) + f_2(\Theta) + f_3(\Theta), \quad (6)$$

where $f_1(\Theta) = \|\Theta\|_*$, $f_2(\Theta) = \lambda \|\Theta\|_{\ell_{2,1}}$, and $f_3(\Theta) = i_{\mathcal{B}_{\ell_2}}(\Theta)$. $\mathcal{B}_{\ell_2} \subset \mathbb{R}^{n_1 \times n_2}$ is the convex set corresponding to the matrices that satisfy the fidelity constraint $\|y - \mathcal{A}(\Phi_{2D}\Theta)\|_{\ell_2} \leq \varepsilon$, and $i_{\mathcal{B}_{\ell_2}}$ is the indicator function of this set defined as:

$$i_{\mathcal{B}_{\ell_2}}(\Theta) = \begin{cases} 0 & \text{if } \Theta \in \mathcal{B}_{\ell_2} \\ +\infty & \text{otherwise} \end{cases}$$

Minimizing summation of lower semi-continuous convex functions in (6) can be effectively done using the Parallel Proximal Algorithm (PPXA) proposed in [10]. PPXA is an iterative method that at each iteration computes the *proximity operators* of the functions f_i in the summation, averages their results and updates the solution until the convergence point. The proximity operator of a function $f_i(\Theta)$ is defined as $prox_{f_i} : \mathbb{R}^{n_1 \times n_2} \rightarrow \mathbb{R}^{n_1 \times n_2}$:

$$\operatorname{argmin}_{\tilde{\Theta} \in \mathbb{R}^{n_1 \times n_2}} f_i(\tilde{\Theta}) + \frac{1}{2} \|\Theta - \tilde{\Theta}\|_F^2, \quad (7)$$

wherein, $\|\cdot\|_F$ denotes the matrix Frobenius norm. Regarding this definition, the proximal operator of the indicator function f_3 is obviously the orthogonal projection onto the ℓ_2 ball \mathcal{B}_{ℓ_2} , which can be

Algorithm 1: The Parallel Proximal Algorithm to solve (5)

Input: $(y, \mathcal{A}, \Phi_{2D})$, the regularization triple $(\varepsilon, \lambda, \gamma > 0)$.

Initializations:

$n = 0, \Theta_0 = \Gamma_{1,0} = \Gamma_{2,0} = \Gamma_{3,0} \in \mathbb{R}^{n_1 \times n_2}$

repeat

for $(i = 1 : 3)$ **do**

$P_{i,n} = prox_{\gamma f_i}(\Gamma_{i,n})$

end

$\Theta_{n+1} = (P_{1,n} + P_{2,n} + P_{3,n})/3$

for $(i = 1 : 3)$ **do**

$\Gamma_{i,n+1} = \Gamma_{i,n} + 2\Theta_{n+1} - \Theta_n - P_{i,n}$

end

until *convergence* ;

computed iteratively as proposed in [11] (and within a single iteration, once \mathcal{A} is a tight frame). Moreover, by standard calculation we have $\forall i \in \{1, \dots, n_1\}$

$$[prox_{f_2}(\Theta)]^i = \frac{\max(\|\Theta^i\|_{\ell_2} - \lambda, 0)}{\|\Theta^i\|_{\ell_2}} \Theta^i,$$

that is the *soft thresholding* operator applied on the rows of Θ to shrink their ℓ_2 norms. Finally, assume $\Theta = U\Sigma V^T$ is the 'economized' singular value decomposition of Θ with $\Sigma = \text{diag}(\sigma_1, \dots, \sigma_r)$, then the proximity of the nuclear norm would be the singular value soft thresholding i.e., $prox_{\omega f_1}(\Theta) = U\bar{\Sigma}V^T$ where $\bar{\Sigma} = \text{diag}(\bar{\sigma}_1, \dots, \bar{\sigma}_r)$ and $\bar{\sigma}_i = \max(\sigma_i - \omega, 0)$.

With these definitions, the full algorithm to solve (5) can be summarized in Algorithm 1. For the regularization parameters in (5), we set $\varepsilon = \|z\|_{\ell_2}$ that is the estimated noise power, and we suggest $\lambda \sim \sqrt{r/k}$ which requires a rough estimation about the rank and sparsity-level of data. Finally, $\gamma > 0$ is an optional parameter and it influences the speed of convergence.

5. SIMULATION RESULTS

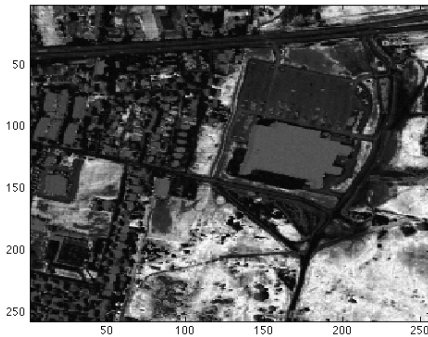
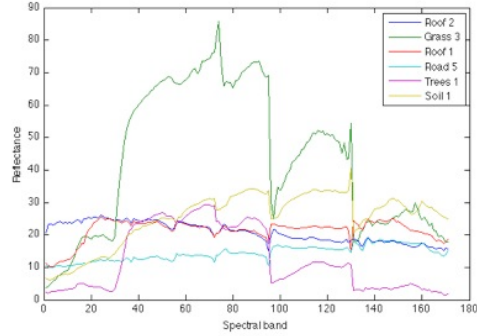
We evaluate the performance of our approach on the standard URBAN hyperspectral dataset. We crop the corresponding HSI to have the spatial resolution $n_1 = 256 \times 256$, and in addition we sub-sampled $n_2 = 170$ spectral bands. The dataset consists in mainly six source images corresponding to the spectral signature of certain materials (see Figure 1(a)). Therefore, we expect the HSI to be approximated by no more than six principal components i.e., $r \sim 6$.

For compressive acquisition, a block diagonal random matrix similar to (4) has been applied to collect m linear random measurements from data as in (2). An independent random convolution measurement matrix [9] has been used for each spectral band. In this case, the resulting sampling operator/matrix is a tight frame. Note that, in our experiments the CS measurements are corrupted by the additive white Gaussian noise whose power corresponds to the 'sampling SNR'.

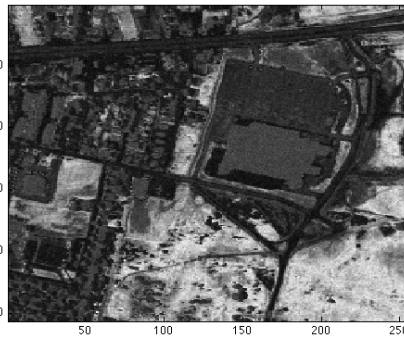
We use our recovery approach (5) to reconstruct a low-rank and joint-sparse representation of the spatial wavelet coefficients Θ . The reconstruction SNR has been evaluated in Table 5 for a fixed compression rate $\delta = m/n_1 n_2 = 1/16$, and under various noise powers. For such a low compression rate ($\sim 6\%$ of the whole HSI size) our results indicate a robust recovery against different noise regimes. We repeat this experiment for the noiseless case while the compression rate being halved $\sim 3\%$, and we achieve the reconstruction SNR of 18.7dB. Figures 1(b)-1(d) are illustrated as an evidence of the



(a) URBAN hyperspectral dataset: six sources and their corresponding spectral signatures



(b) $\delta = 1/16$ and sampling SNR= 40dB



(c) $\delta = 1/16$ and sampling SNR= 20dB



(d) $\delta = 1/32$ and sampling SNR= ∞

Fig. 1: Reconstruction of the URBAN HSI using (5) for different compression ratios and sampling noise powers, illustrated for the spectral channel $j = 33$

Sampling SNR (dB)	∞	40	20	10	0
Reconstruction SNR (dB)	42.7	34.5	21.1	14.1	6.6

Table 1: Reconstruction SNR using (5) for $\delta = 1/16$ and under different sampling noise regimes.

reconstruction quality that is suggested by (5) for such severe undersampling regimes (results are demonstrated for the spectral band $j = 33$). Note that, none of the standard recovery methods that use either the wavelet sparsity or the low-rank priors can achieve such performance. For a fair comparison, BPDN (1) has been applied for joint-recovery of the 3D spatio-spectral sparse wavelet coefficients (as proposed by [8]). The resulting reconstruction SNR for the same CS measurements and with compression rate $1/16$ is about 10dB, so that we find the recovered images not desirable for demonstration.

6. CONCLUSIONS

In this paper we have proposed a novel convex optimization formulation for recovering the hyperspectral images from very few CS measurements. Our approach penalizes both the nuclear norm and the $\ell_{2,1}$ mixed-norm of the data matrix in order to reconstruct simultaneously the low-rank and joint-sparse structure of the data. An algorithm to solve this convex optimization has been proposed based on the proximal splitting methods. By number of simulations we have shown that, our approach is robust against the noise and the number of measurements required for the HSI reconstruction is significantly reduced comparing to the other state-of-the-art methods.

For the future work, we investigate the possible enhancements that can be obtained by replacing the TV-norm instead of $\ell_{2,1}$ penalty on the wavelet coefficients, because in practice penalizing the TV-norm appears to better handle the image recovery comparing to the wavelet-based approaches.

7. REFERENCES

- [1] Q. Du and J. E. Fowler, "Hyperspectral image compression using jpeg2000 and principal component analysis," *IEEE Geosci. Remote Sens. Lett.*, vol. 4, no. 4, pp. 201–205, April 2007.
- [2] D.L. Donoho, "Compressed sensing," *IEEE Trans. on Inf. Theory*, vol. 52, no. 4, pp. 1289–1306, 2006.
- [3] E. J. Candes, J. Romberg, and T. Tao, "Stable signal recovery from incomplete and inaccurate measurements," *Pure Appl. Math.*, vol. 59, pp. 1207–1223, 2005.
- [4] E. J. Candes and Y. Plan, "Tight oracle bounds for low-rank matrix recovery from a minimal number of random measurements," *IEEE Trans. on Inf. Theory*, 2009.
- [5] T. Sun and K.F. Kelly, "Compressive sensing hyperspectral imager," *Comp. Optical Sensing and Imaging (COSI), San Jose, CA, Oct. 2009*.
- [6] A. Wagadarikar, R. John, R. Willett, and D. Brady, "Single disperser design for coded aperture snapshot spectral imaging," *Applied Optics*, vol. 47, pp. B44–B51, 2008.
- [7] J.M. Bioucas-Dias and M.A.T. Figueiredo, "A new twist: Two-step iterative shrinkage/thresholding algorithms for image restoration," *Image Processing, IEEE Trans. on*, vol. 16, no. 12, pp. 2992–3004, dec. 2007.
- [8] M.F. Duarte and R.G. Baraniuk, "Kronecker Compressive Sensing," *to appear in the IEEE Trans. on Image Processing*, 2009.
- [9] J. Romberg, "Compressive sensing by random convolution," *SIAM J. Imaging Sciences*, 2009.
- [10] P. L. Combettes and J. C. Pesquet, "Proximal splitting methods in signal processing," in: *Fixed-Point Algorithms for Inverse Problems in Science and Engineering*, Springer-Verlag, vol. 49, pp. 185–212, 2011.
- [11] Mohamed-Jalal Fadili and Jean-Luc Starck, "Monotone operator splitting for optimization problems in sparse recovery," in *ICIP*, 2009, pp. 1461–1464.