DYNAMIC FILTERING OF SPARSE SIGNALS USING REWEIGHTED ℓ_1

Adam S. Charles, Student Member, IEEE and Christopher J. Rozell, Senior Member, IEEE

School of Electrical and Computer Engineering
Georgia Institute of Technology
Atlanta, GA
30332-0250 USA
acharles6,crozell@gatech.edu

ABSTRACT

Accurate estimation of undersampled time-varying signals improves as stronger signal models provide more information to aid the estimator. In class Kalman filter-type algorithms, dynamic models of signal evolution are highly leveraged but there is little exploitation of structure within a signal at a given time. In contrast, standard sparse approximation schemes (e.g., L1 minimization) utilize strong structural models for a single signal, but do not admit obvious ways to incorporate dynamic models for data streams. In this work we introduce a causal estimation algorithm to estimate time-varying sparse signals. This algorithm is based on a hierarchical probabilistic model that uses re-weighted L1 minimization as its core computation, and propagates second order statistics through time similar to classic Kalman filtering. The resulting algorithm achieves very good performance, and appears to be particularly robust to errors in the dynamic signal model.

1. INTRODUCTION

Many real-time applications, such as real-time compressive video acquisition or real-time network tomography, require us to causally estimate a time-varying signal from a sequence of measurements. Such causal signal estimation is sometimes referred to as dynamic filtering. While batch methods that seek to estimate a series of correlated signals have flexibility in terms of how to leverage correlations between the signals, dynamic filtering is a more restrictive problem in at least three ways. First, no future data is available, limiting the information that can be used at each time step. Second, since dynamic filtering typically needs to be performed over long time spans, approaches require a concise and efficient way to store the previous information needed for estimation. Lastly, the actual estimation procedure at each time step needs to be computationally efficient to be effectively applied to the data stream.

In this paper, we focus on time-varying signals where we have an *a-priori* model for how the signal evolves in time,

$$\boldsymbol{x}_n = f_n\left(\boldsymbol{x}_{n-1}\right) + \boldsymbol{\nu}_n$$

where $\boldsymbol{x}_n \in \mathbb{R}^N$ is our signal of interest, $f_n(\cdot) : \mathbb{R}^N \to \mathbb{R}^N$ is our dynamic evolution function (which is potentially different at each time step) and $\boldsymbol{\nu}_n \in \mathbb{R}^N$ is the innovations. The innovations represents the error in our assumed dynamic model $f_n(\cdot)$. In this work we will be particularly interested in the case where the signal $\boldsymbol{x}_n$ is *sparse* in some representation (i.e $\boldsymbol{x}_n = \boldsymbol{\Psi}\boldsymbol{a}_n$ where $\boldsymbol{a}_n$ is mostly composed of zeros). We denote a signal to be s-sparse if only at most s non-zero coefficients are present. The signals themselves are sensed via a linear sensing matrix

$$\boldsymbol{y}_n = \boldsymbol{G}_n\boldsymbol{x}_n + \boldsymbol{\epsilon}_n$$

where $\boldsymbol{y}_n \in \mathbb{R}^M$ are the measurements taken at each iteration, $\boldsymbol{G}_n \in \mathbb{R}^{M\times N}$ is the sensing matrix and $\boldsymbol{\epsilon}_n$ is the measurement error. While in some applications it may be convenient to have $\boldsymbol{G}_n$ be the same at each iteration, here we treat the general case where the measurements are different for all n.

Our goal is to use only information available at time n to infer $\boldsymbol{x}_n$. This means that we have access to all $\boldsymbol{y}_k$ for $k \leq n$. While generally we could design an estimation procedure which uses all $\boldsymbol{y}_k$ directly to infer $\boldsymbol{x}_n$, such an estimation procedure would be computationally impractical as n becomes large. Instead we focus on methods similar to Kalman filtering [1] which use local information efficiently by retaining a set of parameters to use in the estimation procedure. In standard streaming estimation procedures such as Kalman filtering, the parameters used are the covariance matrix and the previous state estimate. The use of the covariance matrix, however, relies on Gaussian and linear assumptions that are not present in applications where the signals and innovations do not follow Gaussian statistics. In this work we stray from these assumptions since sparse signals follow a different set of statistics, requiring us to utilize different parameters native to sparse signals.

To date, a number of algorithms have been designed to address the problem of dynamically filtering undersampled, sparse, time-varying signals. Some of these methods seek to directly modify the equations that stem from the Kalman filter directly to account for sparsity [2], to use the Kalman filter equations on a restricted support [3], or to use sparse recovery optimization problems with additional norms to include temporal information [4]. Other methods work on specific time-varying models, such as characterizing the innovations to be sparser than the signal itself [5] or treating the signal's temporal evolution as a Gauss-Bernoulli signal with Markov transitions on the support [6]. We propose a methodology which uses precisely the same problem formulation as the Kalman filtering problem, but design a new probabilistic model to account for the non-Gaussian nature of the signal. We then derive an expectation-maximization (EM) algorithm that recovers the signals. Lastly, we test the resulting algorithm, reweighted ℓ_1 dynamic filtering (RWL1-DF), showing its effectiveness and robustness to the innovations' statistics.

This work was supported in part by NSF grants CCF-0830456 and CCF-0905346.

2. BACKGROUND

2.1. Kalman Filtering

Most standard dynamic filtering techniques are based on the Kalman filter [1]. In the Kalman filter framework, the signal at each time step is recovered using the estimate of the previous time step $\widehat{\boldsymbol{x}}_{n-1}$ and a calculated covariance for that estimate $\boldsymbol{P}_{k-1}$.

$$\widehat{\boldsymbol{x}}_n = \arg\min_{\boldsymbol{x}} \Big[\|\boldsymbol{y}_n - \boldsymbol{G}_n\boldsymbol{x}\|_{2,\boldsymbol{R}_n}^2 + \|\boldsymbol{x} - \boldsymbol{F}_n\widehat{\boldsymbol{x}}_{n-1}\|_{2,(\boldsymbol{Q}_n+\boldsymbol{F}\boldsymbol{P}_{n-1}\boldsymbol{F}^T)}^2 \Big]$$

where $\boldsymbol{R} \in \mathbb{R}^{M\times M}$ is the covariance matrix for the measurement error, $\boldsymbol{Q} \in \mathbb{R}^{N\times N}$ is the covariance matrix for the innovations, $\boldsymbol{F} \in \mathbb{R}^{N\times N}$ is the linear version of the dynamics function $f_n(\cdot)$, and the matrix weighted norm is defined as $\|\boldsymbol{z}\|_{2,\boldsymbol{R}}^2 = \boldsymbol{z}^T\boldsymbol{R}^{-1}\boldsymbol{z}$. The allure of the Kalman filter is that while the optimization in Equation (2.1) uses only local information, it solves a global optimization problem. This property, however, is due to the linearity of both the measurement and dynamics functions as well as the Gaussian nature of the signal, measurement error, and innovations.

In cases where the linearity and Gaussianity conditions are not met, alternate methods based on the Kalman filter have been proposed. The extended Kalman filter, for example, addresses the case of non-linear dynamics by linearizing around a point [7]. The EKF is limited, however, in that vary nonlinear functions are not well approximated by the linearization. Particle filtering techniques seek to bypass needing closed form solutions to the signal statistics by performing Monte-Carlo type sampling. The sampled points are used to approximate moments of the signal statistics to use in the estimation procedure. One of the more well known versions is the unscented Kalman filter (UKF) which uses a sampling scheme that targets minimal distortion in the second moment [8]. None of these various extensions, though, allow for straightforward incorporation of explicit signal structure such as sparsity.

2.2. Sparse Signal Recovery

We seek to incorporate sparsity structure in to dynamic filtering due to its growing utility in a number of important applications (e.g. inverse problems in image processing [9] and hyperspectral imagery [10]). With sparsity knowledge of a signal, the signal can be recovered from many fewer measurements than would otherwise be required [11]. Typical measurement rates grow linearly with the sparsity and polylogarithmically with the ambient dimension [11]. In standard sparse recovery, the signal coefficients $\boldsymbol{a}$ may be recovered using the Basis Pursuit De-Noising (BPDN) optimization

$$\widehat{\boldsymbol{a}}_n = \arg\min_{\boldsymbol{x}} \left[\|\boldsymbol{y}_n - \boldsymbol{G}_n\boldsymbol{\Psi}\boldsymbol{a}\right]_2^2 + \lambda\|\boldsymbol{a}\|_1$$

where $\|\boldsymbol{z}\|_1 = \sum_i |\boldsymbol{z}[i]|$ is referred to as the ℓ_1 norm and λ trades off between the ℓ_2 data fidelity term and the ℓ_1 sparsity inducing norm. The signal is then reconstructed via $\widehat{\boldsymbol{x}} = \boldsymbol{\Psi}\widehat{\boldsymbol{a}}$.

The BPDN optimization assumes that the variable λ (representing the SNR for each coefficient) is known *a-priori* and the same for each coefficient, which may not be the most accurate signal model. One way to extend BPDN is to use a different value of λ for each coefficient and adapt these values depending on the data. While a complete optimization problem can be written in terms of minimizing a cost function for both the coefficients $\boldsymbol{a}$ and the λ parameters $\boldsymbol{\lambda}$ (where $\dim(\boldsymbol{\lambda}) = \dim(\boldsymbol{a})$), the desired (non-convex) program is typically solved via a variational algorithm where $\boldsymbol{\lambda}$ is updated between solving a series of weighted BPDN programs [12]. In particular, this reweighted ℓ_1 (RWL1) optimization program solves

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

with the weight update

$$\widehat{\boldsymbol{\lambda}}_n^t[i] = \frac{\tau}{|\widehat{\boldsymbol{a}}_n^t[i]| + \eta}$$

where λ_0, α and β are constants, t is the algorithmic iteration number (taken to increase until some convergence criterion has been met) and the signal estimate is $\widehat{\boldsymbol{x}}_n = \boldsymbol{\Psi}\widehat{\boldsymbol{a}}_n$.

As described in [13], the RWL1 approach described above can be viewed as a Bayesian inference problem for a hierarchical probabilistic model. In [13], the coefficients $\boldsymbol{a}_n$ and are treated as Laplacian random variables (conditioned on λ variables) to be inferred from the linear measurements under Gaussian measurement noise assumptions (Gaussian likelihood $p(\boldsymbol{y}|\boldsymbol{a})$). The λ parameters are treated as random variables with Gamma hyperpriors, and these values modulate the variances of the Laplacian coefficient priors. Once these distributions are defined, the reweighted optimization comes from applying an expectation-maximization (EM) approach to solving the complete MAP inference. The distribution on λ can be adjusted based on any information we have about the signal being estimated (i.e., encouraging or discouraging coefficients to be active a priori by making λ likely to be small or large). Therefore, this hierarchical probabilistic model provides a way to include additional regularization information from the temporal dynamics model into the second moments of the variables of interest, much like in the Kalman filtering framework.

3. RE-WEIGHTED ℓ_1 DYNAMIC FILTERING

In our approach, we wish to encourage the signal estimate to take on values predicted using a temporal model while not explicitly penalizing errors. In essence, we seek to influence the estimate by having the weights (in the RWL1 iterations) be affected both by the past signal estimate as well as the new measurements. Our probabilistic model accomplishes this task via the second order variables, embodied by $\boldsymbol{\lambda}$.

In our probabilistic model construction, we retain key features of the RWL1 model that give it the sparsity inducing properties while inducing temporal correlation by tying together the distributions of the hyperpriors with the previous state. Our variable probability distributions are defined as

$$p(\boldsymbol{y}_n|\boldsymbol{a}_n) \propto e^{-\frac{1}{2\sigma^2}\|\boldsymbol{y}_n - \boldsymbol{G}_n\boldsymbol{\Psi}\boldsymbol{a}_n\|_2^2} \quad (1)$$

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

$$p(\boldsymbol{\lambda}_n[i]|\boldsymbol{\lambda}_{n-1}) = \frac{\boldsymbol{\lambda}_n^{\alpha-1}[i]}{\boldsymbol{\theta}_n^{\alpha}[i]\Gamma(\alpha)}e^{-\boldsymbol{\lambda}_n[i]/\boldsymbol{\theta}_n} \quad (3)$$

where α and λ_0 are constants, and $\boldsymbol{\theta}_n$ is the vector which controls the mean and variance of the Gamma distribution over $\boldsymbol{\lambda}_n$,

$$\boldsymbol{\theta}_n[i] = \xi\left(|\boldsymbol{\Psi}^{-1}f_n(\boldsymbol{\Psi}\boldsymbol{a}_{n-1})[i]| + \eta\right)^{-1}$$

for some model parameters ξ and η. The graphical representation of this probabilistic model (shown in Figure 1) shows that the variable dependencies causally feed forward in time, implying additional regularization which we can leverage in the signal estimation.

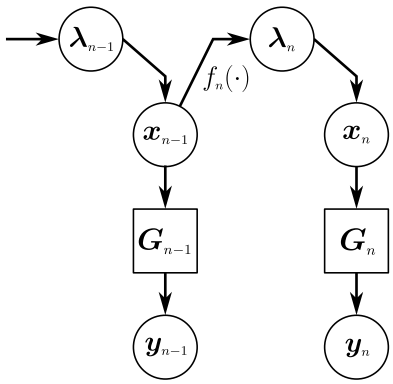

Fig. 1. The flow of information in the RWL1-DF model has temporal priors where the current variance values are dependent on a function of the past state.

We can use these distributions to write a MAP estimate for $\boldsymbol{a}_n$ and $\boldsymbol{\lambda}_n$ at time n as

$$\left\{\widehat{\boldsymbol{a}}, \widehat{\boldsymbol{\lambda}}\right\} = \arg\min_{\boldsymbol{a},\boldsymbol{\lambda}} -\log\left(p(\boldsymbol{y}_n|\boldsymbol{a}_n)p(\boldsymbol{a}_n|\boldsymbol{\lambda}_n)p(\boldsymbol{\lambda}|\widehat{\boldsymbol{\lambda}}_{n-1})\right)$$

By plugging in the distributions in Equations (1)- (3), we can use the EM approach to derive update steps for $\widehat{\boldsymbol{a}}_n$ and $\widehat{\boldsymbol{\lambda}}_n$. Specifically, we derive the update equations

$$\widehat{\boldsymbol{a}}_n^t = \arg\min_{\boldsymbol{a}} [\|\boldsymbol{y}_n - \boldsymbol{G}_n\boldsymbol{\Psi}\boldsymbol{a}]_2^2 + \lambda_0 \sum |\widehat{\boldsymbol{\lambda}}_n^{t-1}[i]\boldsymbol{a}[i]| \quad (4)$$

and

$$\widehat{\boldsymbol{\lambda}}_n^t[i] = \frac{(1+\alpha)\xi}{\lambda_0\xi|\widehat{\boldsymbol{a}}_n^t[i]| + |\boldsymbol{\Psi}^{-1}f_n\left(\boldsymbol{\Psi}\widehat{\boldsymbol{a}}_{n-1}\right)[i]| + \eta}. \quad (5)$$

In the standard RWL1 algorithm, values that are estimated to be high in the weighted ℓ_1 optimization result in lower weights via the weight update step, while off values result in higher weights. In the next weighted ℓ_1 optimization, these weights encourage the 'on' values to remain on and obtain higher values while encouraging 'off' values to remain low and tend to zero. Equation (5) for our modified RWL1-DF algorithm allows the weights to also be effected by past estimate, encouraging predicted values to be active while still allowing for information from the measurements to contradict the prediction and correct the estimate. One major advantage here is that since no explicit norm ties the prediction and the estimate together, the estimation procedure should be more robust to the statistics of the innovations, a behaviour we see empirically in our simulations.

4. RESULTS

We test the RWL1-DF algorithm both on simulated data and compressive recovery of a video sequence. We compare performance against the BPDN and RWL1 algorithms applied independently at each time step and a time dependent version of BPDN (BPDN Dynamic Filtering: BPDN-DF) that includes an additional history-dependent regularization norm as

$$\widehat{\boldsymbol{a}}_n = \arg\min_{\boldsymbol{a}} \|\boldsymbol{y}_n - \boldsymbol{G}_n\boldsymbol{\Psi}\boldsymbol{a}\|_2^2 + \lambda\|\boldsymbol{a}\|_1 + \gamma\|\boldsymbol{a} - f_n\left(\widehat{\boldsymbol{a}}_{n-1}\right)\|_p^p$$

where λ and γ are constants that trade off between historical information, sparsity, and measurement fidelity, p dictates the assumed innovations statistics, and the signal estimate is $\widehat{\boldsymbol{x}}_n = \boldsymbol{\Psi}\widehat{\boldsymbol{a}}_n$ [4, 14]. In simulated data where ground truth is known, we also compare against an optimal oracle (i.e., support is known *a-priori*) least-squares solution. To evaluate the recovery, we use the relative mean-squared error (rMSE),

$$e_{rMSE} = \frac{\|\widehat{\boldsymbol{x}} - \boldsymbol{x}\|_2^2}{\|\boldsymbol{x}\|_2^2}.$$

In the case of simulated data, we generate 100 time-step sequences of 500-length vectors that are 20-sparse. The dynamics function is a randomly drawn permutation matrix with a random scaling (the dynamics are different at each iteration). The dynamics are assumed to be known up to a few misplaced support, implying a sparse innovations where the sparsity is twice the number of misplaced values, and the measurements matrices are random Gaussian matrices. For the simulated data we use the recovery parameters: $\lambda = 5.5 * 10^{-4}$ for BPDN, $\lambda_0 = 0.0011$, $\tau = 2$ and $\eta = 0.01$ for RWL1, $\lambda = 5 * 10^{-4}$, $\gamma = 2.5 * 10^{-4}$ and $p = 1$ for BPDN-DF, and $\lambda_0 = 0.0011$, $\eta = 0.01$, $\lambda_0\xi = 1$ and $(1+\alpha)\xi = 2$ for RWL1-DF. First we fix the mean innovations sparsity to 6 (the support mismatch is Poisson with mean 3), and sweep the number of measurements from 55 to 110. The steady state rMSE was averaged over 40 trials and the results are plotted in Figure 2. The steady state rMSE for RWL1-DF stays at less than 1% rMSE down to 65 measurements, at which point BPDN-DF (with $p = 1$ due to the sparse innovations) has approximately 7% rMSE and both BPDN and RWL1 are over 20% rMSE. We then fix the number of measurements to $M = 70$ and sweep the mean innovations sparsity from 2 to 10 (again the mismatch is a Poisson). While for small innovations sparsity BPDN-DF and RWL1-DF both perform close to the optimal least-squares performance, RWL1-DF is much more robust to the change in innovations sparsity, retaining an rMSE error of less than 1.6% rMSE steady state error for up to 8-sparse innovations. BPDN-DF, meanwhile, increases to 8.6% rMSE.

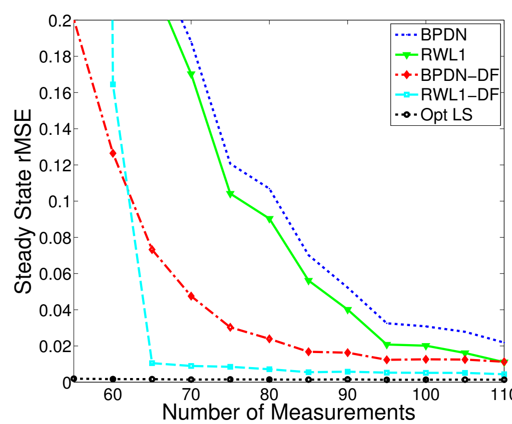

Fig. 2. We recover 500-length, 20-sparse signals from M measurements where we sweep M from 55 to 110. For smaller numbers of measurements, RWL1-DF achieves a lower steady-state rMSE than BPDN-DF, and the time independent BPDN and RWL1.

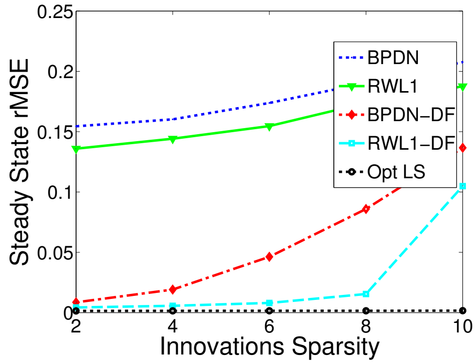

Fig. 3. We sweep the average innovations sparsity from 2 to 10, keeping the same model parameters as in Figure 2 and fixing the number of measurements to $M = 70$. While both BPDN-DF and RWL1-DF perform comparably at low innovations sparsity, the error from RWL1-DF remains low for higher innovations sparsity (higher model mismatch).

For the compressive video recovery, we take $M = 0.25N$ randomly subsampled noiselets from 200 consecutive frames of the foreman video sequence[1]. We recover the video sequence using the dual-tree wavelet transform as the sparsity basis [cite] and the following parameters: $\lambda = 0.01$ for BPDN, $\lambda_0 = 0.001$, $\tau = 0.05$ and $\eta = 0.1$ for RWL1, $\lambda = 0.01$, $\gamma = 0.3$ and $p = 2$ for BPDN-DF, and $\lambda_0 = 0.001$, $\eta = 0.2$, $\lambda_0\xi = 1$ and $(1 + \alpha)\xi = 0.4$ for RWL1-DF. The recovery rMSE of each frame (shown in Figure 4) shows that while BPDN-DF can use dynamical information to enhance the recovery from time-independent recovery, RWL1-DF achieves even lower errors, typically staying below 2% rMSE.

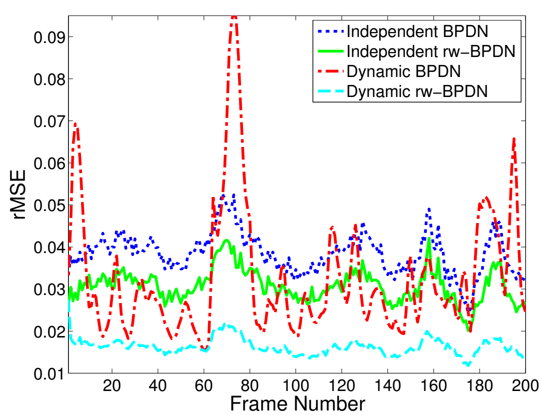

Fig. 4. Recovery of the foreman video sequence from compressive samples.

[1] The foreman video sequence can be found at http://www.hlevkin.com/TestVideo/foreman.yuv .

5. CONCLUSIONS

In this work we show the benefits of utilizing the weight parameters in the RWL1 algorithm for dynamic filtering of sparse signals. The recovery performance of our reweighted ℓ_1 dynamic filtering algorithm, in both simulated experiments and compressive recovery of video sequences, demonstrates improvement over time-independent recovery and simple norm-regularized time dependent recovery. In particular our simulations highlight a robustness of RWL1-DF to the innovations statistics, which can be particularly important in sparse signal regimes where the innovations may be difficult to quantify.

As a note on the complexity of RWL1-DF, although we have added significant *a-priori* information in RWL1-DF as opposed to standard RWL1, the computational cost is essentially unchanged. Even the number of reweighting steps seems unchanged between the two algorithms. Recent advances in homotopy methods for RWL1 algorithms could significantly improve algorithmic speed [15]. Additionally, reweighted algorithms can be written as continuous time systems [16], enabling the possibility of analog implementations in dedicated hardware [17].

6. REFERENCES

[1] 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.

[2] A. Carmi, P. Gurfil, and D. Kanevsky, “Methods for sparse signal recovery using Kalman filtering pseudo-measurementnt norms and quasi-norms,” *IEEE Transactions on Signal Processing*, vol. 58, no. 4, pp. 2405–2409, Apr 2010.

[3] N. Vaswani, “Kalman filtered compressed sensing,” *Proc of ICIP 2008*, pp. 893–896, 2008.

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

[5] N. Vaswani, “Ls-cs-residual (ls-cs): Compressive sensing on the lease squares residual,” *IEEE Trans on Signal Processing*, vol. 58, no. 8, pp. 4108–4120, Aug 2010.

[6] J. Ziniel and P. Schniter, “Dynamic compressive sensing of time-varying signals via approximate message passing,” *Arxiv preprint arXiv:1205.4080*, 2012.

[7] S. Haykin, “Kalman filters,” in *Kalman Filtering and Neural Networks*, S. Haykin, Ed. John Wiley & Sons, Inc., 2001, pp. 1–22.

[8] E. A. Wan and R. van der Merwe, “The unscented Kalman filter,” in *Kalman Filtering and Neural Networks*, S. Haykin, Ed. John Wiley & Sons, Inc., 2001, pp. 221–282.

[9] M. Elad, M. Figueiredo, and Y. Ma, “On the role of sparse and redundant representations in image processing,” *IEEE Proc. - Spec, Issue on App. of Compressive Sensing & Sparse Rep.*, Oct 2008.

[10] A. S. Charles, B. A. Olshausen, and C. J. Rozell, “Learning sparse codes for hyperspectral imagery,” *IEEE Journal of Selected Topics in Signal Processing*, vol. 5, no. 5, pp. 963–978, 2011.

[11] E. Candès and J. Romberg, “Sparsity and incoherence in compressive sampling,” *Inverse problems*, vol. 23, p. 969, 2007.

[12] E. Candes, M. B. Wakin, and S. P. Boyd, “Enhancing sparsity by reweighted ℓ_1 minimization,” *Journal of Fourier Analysis and Applications*, vol. 14, no. 5, pp. 877–905, Dec 2008, special Issue on Sparsity.

[13] P. J. Garrigues and B. A. Olshausen, “Learning horizontal connections in a sparse coding model of natural images,” *Advances in Neural Information Processing Systems*, pp. 505–512, 2008.

[14] D. Sejdinovic, C. Andrieu, and R. Piechocki, “Bayesian sequential compressed sensing in sparse dynamical systems,” in *48th Allerton Conference*. IEEE, 2010, pp. 1730–1736.

[15] S. Asif and J. Romberg, “Fast and accurate algorithms for reweighted ℓ_1-norm minimization,” 2012, submitted.

[16] A. S. Charles, P. Garrigues, and C. J. Rozell, “A common network architecture efficiently implements a variety of sparsity-based inference problems,” *Neural Computation*, 2012, in press.

[17] S. Shapero, A. S. Charles, C. J. Rozell, and P. Hasler, “Low power sparse approximation on reconfigurable analog hardware,” *IEEE Journal on Emerging and Selected Topics in Circuits and Systems*, 2012, in Press.