Convergence of Basis Pursuit De-noising with Dynamic Filtering

Adam S. Charles and Christopher J. Rozell
Electrical and Computer Engineering
Georgia Institute of Technology
Atlanta, Georgia, USA
{acharles6,crozell}@gatech.edu

***Abstract*—Causal inference of dynamically changing signals is a vital task in many applications, including real-time image processing and channel estimation. Over the past few years, many algorithms have been proposed to accomplish this task, but extremely few algorithms have any theoretical guarantees on stability, convergence or performance. In this work we use results from the sparsity-based signal processing literature to demonstrate some basic bounds for one particular algorithm: basis pursuit de-noising with dynamic filtering (BPDN-DF). We show for what parameter ranges the algorithm remains stable for, and provide some guarantees on the steady-state approximation error.**

***Index Terms*—sparse signals, dynamic filtering, convergence**

I. Introduction

In the past few decades, signal processing tools have evolved to exploit the underlying structure present in many classes of signals. In particular, dynamic signal structures, (present in many temporally evolving signals such as video sequences and dynamic MRI), and sparsity signal structure (present in many high dimensional signals) have been explored extensively and have played a vital role in many state-of-the-art algorithms. The first of these signal structures deals with signals that evolve through time with a known dynamics update equation

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

where $\boldsymbol{x}_n \in \mathbb{R}^N$ is the dynamically evolving signal at time-step n, $f(\cdot) : \mathbb{R}^N \rightarrow \mathbb{R}^N$ is the dynamics function, and $\boldsymbol{\nu}_n \in \mathbb{R}^N$ is the innovations term that describes the imprecision in our knowledge of the dynamics function. In many applications we are only able to measure the dynamic signal indirectly via a measurement equation

$$\boldsymbol{y}_n = \boldsymbol{\Phi}\boldsymbol{x}_n + \boldsymbol{\epsilon}_n, \tag{2}$$

where $\boldsymbol{y}_n \in \mathbb{R}^M$ is a set of M linear measurements taken by the measurement matrix $\boldsymbol{\Phi} \in \mathbb{R}^{M\times N}$ and $\boldsymbol{\epsilon}_n \in \mathbb{R}^M$ is the measurement error vector. Since typically $M < N$, a major task is often to utilize the limited measurements in conjunction with the dynamics information to recover the underlying signal $\boldsymbol{x}_n$. Furthermore, many applications require real-time solutions, indicating that efficient and causal inference procedures are required. Many algorithms, including the widely used Kalman filtering algorithm and its many extensions and modifications have been highly successful when dynamics information is the only additional information available to supplement the measurements [1]–[3].

The second type of structure that has achieved widespread success is the sparsity signal structure. Many high dimensional signals can be considered to have parsimonious descriptions in some basis [4], [5], i.e. a signal $\boldsymbol{x}$ can be written as

$$\boldsymbol{x} = \boldsymbol{\Psi}\boldsymbol{a}, \tag{3}$$

where $\boldsymbol{\Psi} \in \mathbb{R}^{N\times N}$ is a known basis and $\boldsymbol{a} \in \mathbb{R}^N$ is the coefficient vector that only has $K \ll N$ non-zero entries. Such a signal is said to be K-sparse. This sparse structure can be used when undersampling signals as in the dynamics case to supplement the measurements and invert underdetermined linear systems. This ability has allowed sparsity to be leveraged to great effect in many areas of signal processing, including image processing, audio processing and remote sensing [4], [6], [7]. Specifically, both fast and efficient algorithms as well as theoretical guarantees for the resulting solution have been explored.

Very recently, algorithms have begun to be developed which leverage both dynamics and sparsity in applications where both structures are present (e.g. dynamic MRI). The goal of these algorithms is to obtain more accurate signal estimates at each time step than is achievable by leveraging either dynamics or sparsity alone. While some algorithms seek to modify well known algorithms, (e.g. Kalman filtering) to include sparsity assumptions [8], [9], other algorithms modify optimization programs used for sparse signal analysis to include dynamic information [10]–[13]. While many of these algorithms have been shown to work empirically, theoretical guarantees have largely only been proven for restricted signal models. In this work we focus on providing theoretical guarantees for one specific optimization-based approach called basis pursuit de-noising with dynamic filtering (BPDN-DF). In this procedure, a third norm which captures the dynamic signal structure is added to the basis-pursuit de-noising algorithm used extensively for sparse signal estimation. We establish parameter ranges that ensure stability of this algorithm, and provide extreme worst-case bounds for the steady-state error.

II. Background

A. Sparse approximation

Sparsity has been a highly successful signal model due to both the existence of efficient algorithms that can leverage sparsity and theoretical guarantees on the accuracy of such algorithms. One of the most widely used algorithms which can recover the sparse representation of a signal $\boldsymbol{a}$ from a small number of linear measurements is basis pursuit de-noising. In this algorithm, the optimization program used to recover the coefficients,

$$\widehat{\boldsymbol{a}} = \arg\min_{\boldsymbol{a}} \|\boldsymbol{y} - \boldsymbol{\Phi}\boldsymbol{\Psi}\boldsymbol{a}\|_2^2 + \gamma\|\boldsymbol{a}\|_1, \tag{4}$$

is a convex optimization program consisting of an ℓ_2 penalization on the measurement error (the measurement fidelity term) and an ℓ_1 term which encourages sparse solutions. The parameter γ determines the tradeoff between sparseness and measurement fidelity. The signal is then easily recovered via the multiplication $\widehat{\boldsymbol{x}} = \boldsymbol{\Psi}\widehat{\boldsymbol{a}}$. Since here we treat $\boldsymbol{\Psi}$ as a basis, estimation errors on $\boldsymbol{a}$ translate directly to estimation errors on $\boldsymbol{x}$, so we will focus on recovery results concerning $\boldsymbol{a}$.

Given certain conditions on the measurement matrix, it can be shown that the solution of BPDN can achieve a bounded accuracy. One of the most common properties of the measurement matrix which is used to determine these theoretical bounds is called the restricted isometry property (RIP). For a linear operator $\boldsymbol{\Phi} \in \mathbb{R}^{M\times N}$, we say that $\boldsymbol{\Phi}$ satisfies the RIP of order $2K$ with respect to a basis $\boldsymbol{\Psi}$ with parameters δ and C, if for every $2K$ sparse signal $\boldsymbol{a}$, the norm of $\|\boldsymbol{\Phi}\boldsymbol{\Psi}\boldsymbol{a}\|_2^2$ is bounded as,

$$C\left(1-\delta\right) \leq \|\boldsymbol{\Phi}\boldsymbol{\Psi}\boldsymbol{a}\|_2^2/\|\boldsymbol{a}\|_2^2 \leq C\left(1+\delta\right), \tag{5}$$

This condition essentially states that the difference between any two sparse signals is preserved to within a factor of $1\pm\delta$, indicating that different sparse signals are still distinguishable in the smaller measurement space. Using the RIP, it has been shown that the coefficients recovered via BPDN satisfy the bound

$$\|\boldsymbol{a} - \widehat{\boldsymbol{a}}\|_2 \leq C_1\|\boldsymbol{\epsilon}\|_2 + C_2\gamma\sqrt{q}, \tag{6}$$

where C_1 and C_2 are constants which depend on the RIP constant δ of the measurement matrix, and q is a constant multiple of the sparsity of $\boldsymbol{a}$.

B. Sparsity based dynamic filtering

A number of algorithms have recently combined the sparsity model with traditional dynamics models. While a number of these models perform batch-based optimization (solving for multiple, consecutive signals at once) [14]–[17], many more have focused on the traditional filtering problem where only the most recent signal $\boldsymbol{x}_n$ is recovered at each time n using only past measurements and estimated signals. An early attempt at leveraging both dynamic and sparse information modified the traditional Kalman filtering equations to provide sparse outputs [8], [9]. These methods retained the estimation and propagation of a covariance matrix, which can become computationally inefficient with high-dimensional signals. A more recent approach has modified the efficient optimization procedures, such as BPDN, to include dynamics information [10]–[13].

While optimization-based methods have been empirically successful, few of the theoretical guarantees from the sparsity framework have carried through to the dynamical setting. Exceptions include the work in [18] and [19]. In [18] a bound is provided for the modCS algorithm under a model where the innovation is much sparser than the signal and the dynamics model has bounded changes on the support. In [19] the authors do not assume any dynamics model explicitly but still observe the tracking abilities of the iterative soft thresholding algorithm (ISTA).

C. Basis pursuit de-noising with dynamic filtering

To expand the theoretical guarantees currently available, we analyze here the BPDN-DF algorithm. In the BPDN-DF algorithm the optimization in Equation (4) is modified by adding a third norm as

$$\begin{aligned}\widehat{\boldsymbol{a}}_n \quad = \quad & \arg\min_{\boldsymbol{a}} \|\boldsymbol{y}_n - \boldsymbol{\Phi}\boldsymbol{a}\|_2^2 + \gamma\|\boldsymbol{a}\|_1 \\ & + \kappa\|\boldsymbol{\Psi}\boldsymbol{a} - f(\boldsymbol{\Psi}\widehat{\boldsymbol{a}}_{n-1})\|_2^2,\end{aligned} \tag{7}$$

where κ is a second tradeoff parameter. While in the previous literature, ℓ_p norms with $p \neq 2$ were considered for the third dynamics fidelity term, we focus here on the ℓ_2 case since fast solvers are readily available for this optimization program. With two tradeoff parameters to consider in BPDN-DF, setting these parameters can be difficult. With the dynamic nature of the algorithm to consider, poor performance can propagate through the algorithm and can potentially cause instabilities. Thus showing which ranges of parameters yield a stable algorithm could yield insight into the algorithm's performance. We dedicate the remainder of this paper to describing a first result along these lines which analyzes the behavior of this algorithm. Specifically we seek constraints on κ which ensure that the algorithm remains stable for large classes of dynamics functions $f(\cdot)$, and to then provide worst case scenario bounds for the error of the resulting estimate.

III. Convergence of BPDN-DF

Our main result is summarized in the following theorem:

Theorem III.1. *Suppose that at each time-step n, $\boldsymbol{\Phi} \in \mathbb{R}^{M\times N}$ satisfies RIP($2K$,δ), $\gamma > 0$ and $\kappa > 0$ are known constants. Additionally, suppose that the dynamics function $f(\cdot)$ satisfies $\|f(\boldsymbol{a}_1) - f(\boldsymbol{a}_2)\|_2 \leq f^*\|\boldsymbol{a}_1 - \boldsymbol{a}_2\|_2$ and that for all $n \geq 0$ the error and innovations satisfy $\|\boldsymbol{\epsilon}_n\|_2 \leq \overline{\epsilon}$ and $\|\boldsymbol{\nu}_n\|_2 \leq \overline{\nu}$. Under these conditions, the result of solving the optimization program of Equation (7) satisfies*

$$\|\widehat{\boldsymbol{a}}_n - \boldsymbol{a}_n\|_2 \leq \beta^n\left(\|\boldsymbol{e}_0\|_2 - \frac{\alpha}{1-\beta}\right) + \frac{\alpha}{1-\beta},$$

where the constant α is given by

$$\alpha = C_1\frac{1}{\sqrt{1+\kappa}}\overline{\epsilon} + C_1\sqrt{\frac{\kappa}{1+\kappa}}\overline{\nu} + C_2\frac{\gamma}{1+\kappa}\sqrt{q},$$

and the linear convergence rate is

$$\beta = C_1\sqrt{\frac{\kappa}{1+\kappa}}f^*,$$

and the constants C_1 and C_2 are the constants from the bounds on solving the static BPDN problem with sparsity K and a modified RIP parameter $\widetilde{\delta} = \boldsymbol{\delta}/(1+\kappa)$.

This theorem essentially states that BPDN-DF is guaranteed to converge at a linear rate β so long as $\beta < 1$. Solving for κ in this constraint gives us an upper bound on κ

$$\kappa < \frac{1}{(C_1 * f^*)^2 - 1} \qquad C_1 * f^* > 1$$
$$\kappa > \frac{1}{1 - (C_1 * f^*)^2} \qquad C_1 * f^* < 1$$

which guarantees that there will be a range of parameters for which the algorithm is stable. In the first condition, a larger f^* requires a smaller C_1 value to have the same rage of admissible κ values. This means that less smooth dynamics functions need a more accurate BPDN solver to stay stable. Likewise, a less accurate solver requires a smoother dynamics function to be stable for the same κ range. In the second of these conditions, κ must be greater than a negative number, which implies that all positive κ values result in a stable algorithm.

To prove Theorem III.1, we first show that the BPDN-DF optimization problem at each iteration is a BPDN problem where the sensing matrix satisfies the RIP with a better constant than the associated inference that does not include dynamic filtering. Theorem III.1 is then a direct consequence of using the theoretical guarantees from [19], [20] to obtain a per-iterate error bound, which can be related to the error at the last iteration, allowing for a recursive error bound to be determined. First, we assume that the matrix $\boldsymbol{\Phi}$ satisfies the RIP($2K$, δ) with respect to signals sparse in $\boldsymbol{\Psi}$. We then note that we can combine the first and third terms in the BPDN-DF optimization Equation (7) into an augmented BPDN optimization

$$\widehat{\boldsymbol{a}}_n = \arg\min_{\boldsymbol{a}} \left\| \begin{bmatrix} \frac{1}{\sqrt{1+\kappa}}\boldsymbol{y} \\ \sqrt{\frac{\kappa}{1+\kappa}} f\left(\Psi \boldsymbol{a}_{n-1}\right) \end{bmatrix} - \begin{bmatrix} \frac{1}{\sqrt{1+\kappa}}\boldsymbol{\Phi\Psi} \\ \sqrt{\frac{\kappa}{1+\kappa}}\boldsymbol{\Psi} \end{bmatrix} \boldsymbol{a} \right\|_2^2 + \frac{\gamma}{1+\kappa}\|\boldsymbol{a}\|_1, \quad (8)$$

Which is essentially trying to solve the BPDN problem with the augmented matrix $\widetilde{\boldsymbol{\Phi}} = [\boldsymbol{\Psi}^T\boldsymbol{\Phi}^T, \sqrt{\kappa}\boldsymbol{\Psi}^T]^T$, and the factor of $1/(1+\kappa)$ is introduced to normalize the columns of the augmented measurement matrix. Thus the first step is to show that $\widetilde{\boldsymbol{\Phi}}$ satisfies the RIP as well, and for more favorable constants. Since we assumed that $\boldsymbol{\Phi}$ had RIP($2K$,δ), we can find the RIP of $\widetilde{\boldsymbol{\Phi}}$ by observing the upper and lower bounds of the norm of $\|\widetilde{\boldsymbol{\Phi}}\boldsymbol{a}\|_2^2$ for any $2K$-sparse $\boldsymbol{a}$:

$$\begin{aligned} \|\widetilde{\boldsymbol{\Phi}}\boldsymbol{\Psi}\boldsymbol{a}\|_2^2 &= \frac{1}{1+\kappa}\|\boldsymbol{\Phi\Psi a}\|_2^2 + \frac{\kappa}{1+\kappa}\|\boldsymbol{a}\|_2^2 \\ &\leq \frac{C}{1+\kappa}(1+\delta)\|\boldsymbol{a}\|_2^2 + \frac{\kappa}{1+\kappa}\|\boldsymbol{a}\|_2^2 \\ &\leq \frac{C+\kappa+C\delta}{1+\kappa}\|\boldsymbol{a}\|_2^2 \\ &\leq \frac{C+\kappa}{1+\kappa}(1+\frac{C}{C+\kappa}\delta)\|\boldsymbol{a}\|_2^2, \end{aligned}$$

similarly, for the lower bound, we obtain $\|\widetilde{\boldsymbol{\Phi}}\boldsymbol{\Psi}\boldsymbol{a}\|_2^2 \geq \frac{C+\kappa}{1+\kappa}(1-\frac{C}{C+\kappa}\delta)\|\boldsymbol{a}\|_2^2$. Thus the RIP constants for $\widetilde{\boldsymbol{\Phi}}$ are $\widetilde{C} = (C+\kappa)/(1+\kappa)$ and $\widetilde{\delta} = C\delta/(C+\kappa)$. Assuming $\boldsymbol{\Phi}$ is well normalized (i.e. $C = 1$), these expressions reduce to $\widetilde{C} = 1$ and $\widetilde{\delta} = \delta/(1+\kappa)$. Since κ is always positive, this implies that $\widetilde{\delta} < \delta$ and the conditioning on the augmented matrix is improved with respect to the original system. It remains, however, to show that the improved conditioning yields any tangible benefits given that new errors are introduced in the innovations term.

In the BPDN bounds we need to know the ℓ_2 error of the measurements σ_n, which in this case depends on both the actual measurement error as well as the dynamics error. The augmented system has to account for the errors not only in the dynamics model (the innovations term), but also in the previous estimate. We can thus bound the error by observing that

$$\begin{aligned} &\sqrt{\frac{\kappa}{1+\kappa}}(f(\boldsymbol{\Psi}\widehat{\boldsymbol{a}}_{n-1}) - \boldsymbol{\Psi}\boldsymbol{a}_n) \\ &\qquad = \sqrt{\frac{\kappa}{1+\kappa}}(f(\boldsymbol{\Psi}\widehat{\boldsymbol{a}}_{n-1}) - f(\boldsymbol{\Psi}\boldsymbol{a}_{n-1}) - \boldsymbol{\nu}_n), \end{aligned}$$

Using the smoothness assumption on $f(\cdot)$, we can see that

$$\begin{aligned} &\|\sqrt{\frac{\kappa}{1+\kappa}}(f(\boldsymbol{\Psi}\widehat{\boldsymbol{a}}_{n-1}) - \boldsymbol{\Psi}\boldsymbol{a}_n)\|_2 \\ &\qquad \leq \sqrt{\frac{\kappa}{1+\kappa}}(f^*\|\boldsymbol{e}+n-1\|_2 + \|\boldsymbol{\nu}_n\|_2), \end{aligned}$$

With this inequality, and the assumptions that $\|\boldsymbol{\epsilon}_n\|_2 \leq \overline{\epsilon}$ and $\|\boldsymbol{\nu}_n\|_2 \leq \overline{\nu}$ for all n, the effective measurement error on the augmented is then

$$\begin{aligned} \sigma_n &= \|\widetilde{\boldsymbol{y}}_n - \widetilde{\boldsymbol{\Phi}}\boldsymbol{a}_n\|_2 \\ &\leq \frac{\|\boldsymbol{\epsilon}_n\|_2}{\sqrt{1+\kappa}} + \sqrt{\frac{\kappa}{1+\kappa}}f^*\|\boldsymbol{e}_{n-1}\|_2 + \sqrt{\frac{\kappa}{1+\kappa}}\|\boldsymbol{\nu}_n\|_2 \\ &\leq \frac{1}{\sqrt{1+\kappa}}\overline{\epsilon} + \sqrt{\frac{\kappa}{1+\kappa}}f^*\|\boldsymbol{e}_{n-1}\|_2 + \sqrt{\frac{\kappa}{1+\kappa}}\overline{\nu} \quad (9) \end{aligned}$$

where f^* is the Lipshitz constant for the function f.

The general form of the BPDN solution satisfies

$$\|\boldsymbol{a}_n - \widehat{\boldsymbol{a}}_n\|_2 \leq C_1\sigma_n + C_2\gamma\sqrt{q},$$

where C_1 and C_2 are constants, which can vary depending on the techniques used [19], [20]. We can use this bound with the per-time-step σ_n from Equation (9) to find the time-dependent bound

$$\begin{aligned} \|\boldsymbol{e}_n\|_2 &\leq \frac{C_1\left(\bar{\epsilon} + \sqrt{\kappa} f^* \|\boldsymbol{e}_{n-1}\|_2 + \sqrt{\kappa}\bar{\nu}\right)}{\sqrt{1+\kappa}} + \frac{C_2 \gamma}{1+\kappa}\sqrt{q} \\ &= \left(C_1 \sqrt{\frac{\kappa}{1+\kappa}} f^*\right) \|\boldsymbol{e}_{n-1}\|_2 \\ &\quad + \left(\frac{C_1 \bar{\epsilon}}{\sqrt{1+\kappa}} + C_1 \sqrt{\frac{\kappa}{1+\kappa}} \bar{\nu} + \frac{C_2 \gamma}{1+\kappa}\sqrt{q}\right) \\ &= \beta \|\boldsymbol{e}_{n-1}\|_2 + \alpha \end{aligned} \qquad (10)$$

where

$$\beta = C_1 \sqrt{\frac{\kappa}{1+\kappa}} f^* \qquad (11)$$

and

$$\alpha = C_1 \frac{1}{\sqrt{1+\kappa}} \bar{\epsilon} + C_1 \sqrt{\frac{\kappa}{1+\kappa}} \bar{\nu} + C_2 \frac{\gamma}{1+\kappa} \sqrt{q} \qquad (12)$$

This relationship is essentially a simple linear difference equation and is easily solved for the error at each time step:

$$\|\boldsymbol{e}_n\|_2 \leq \beta^n \left(\|\boldsymbol{e}_0\|_2 - \frac{\alpha}{1-\beta}\right) + \frac{\alpha}{1-\beta} \qquad (13)$$

indicating that this algorithm converges linearly with rate β when $\beta < 1$ and the steady state error as $n \to \infty$ is $\|\boldsymbol{e}\infty\|_2 \leq \alpha/(1-\beta)$.

IV. Simulations

We validate our bound by comparing to the empirical behavior of BPDN-DF. We run BPDN-DF on sequences of 100 $K = 15$-sparse signals of size $N = 576$. At each time step we take $M = 68$ measurements. We recover the sequence of signals using BPDN-DF with $\gamma = 5.5 \times 10^{-4}$ and sweep κ over 30 possible values. We average all our results over 50 trials. We fit our theoretical bounds by selecting C_1 and C_2 such that they fall above the empirical curves. Figure 1 shows That the convergence time increases as predicted by the theory ($n$convergence $\propto \log^{-1}\beta$). The worst-case-scenario nature of the bound, however, creates a gap in the predicted steady-state error. The theoretical curve for the error does not predict the dip that occurs for the optimal κ value, and instead has a monotonically increasing value from $\kappa = 0$, the point that corresponds to simply running BPDN independently at each iteration.

V. Conclusions

In this work we have derived worst-case bounds for the BPDN-DF algorithm based on results from the sparse approximation literature. Our results guarantee ranges of the parameter κ for which the BPDN-DF algorithm converges to a steady-state error and bounds the time to convergence. In exploring the upper bound for κ we see that there appears to be a tradeoff between the accuracy of the BPDN solver (in terms of C_1) and the smoothness of the dynamics function (in terms of f^*.

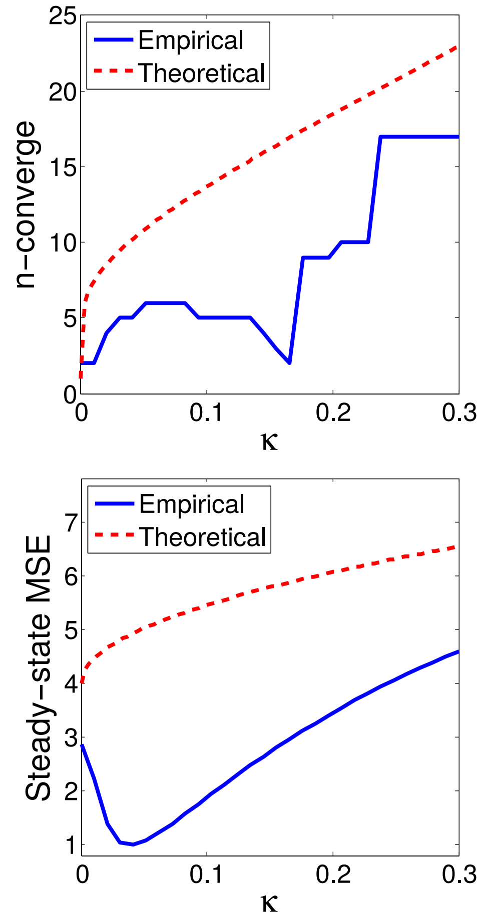

Fig. 1. The theoretical bound was fit to empirical curves of BPDN-DF's behavior as a function of κ. Top: The empirical number of iterations to convergence (solid blue curve) generally increases as a function of κ, as predicted by theory (dashed red curve). The dip in the empirical curve corresponds to the crossover point as the steady-state error increases from being below the initial error to being above the initial error. Bottom: The derived bound accounts for the worst possible recovery at each time-step, and thus yields an extreme upper bound in terms of the steady-state error.

Our empirical results show that while the bound does seem to capture the convergence properties, the nature of the worst-case bounds over-estimates the steady-state error, indicating that tighter bounds for the actual error can be pursued.

Acknowledgement

This work was partially supported by NSF grants CCF-0830456 and CCF-1409422.

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

[3] 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.

[4] 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.

[5] R. Baraniuk, "Compressive sensing," *IEEE Signal Processing Magazine*, vol. 24, no. 4, pp. 118–121, Jul 2007.

[6] M. Davies and L. Daudet, "Sparse audio representations using the mclt," *Signal processing*, vol. 86, no. 3, pp. 457–470, 2006.

[7] 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.

[8] N. Vaswani, "Kalman filtered compressed sensing," *Proc of ICIP 2008*, pp. 893–896, 2008.

[9] 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.

[10] 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.

[11] N. Vaswani and W. Lu, "Modified-cs: Modifying compressive sensing for problems with partially known support," *Signal Processing, IEEE Transactions on*, vol. 58, no. 9, pp. 4595–4607, 2010.

[12] D. Sejdinovic, C. Andrieu, and R. Piechocki, "Bayesian sequential compressed sensing in sparse dynamical systems," in *Communication, Control, and Computing (Allerton), 2010 48th Annual Allerton Conference on*. IEEE, 2010, pp. 1730–1736.

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

[14] S. Farahmand, G. Giannakis, and D. Angelosante, "Doubly robust smoothing of dynamical processes via outlier sparsity constraints," vol. 59, no. 10, pp. 4529–4543, Oct 2011.

[15] Z. Zhang and B. Rao, "Sparse signal recovery with temporally correlated source vectors using sparse bayesian learning," *Selected Topics in Signal Processing, IEEE Journal of*, no. 99, pp. 1–1, 2011.

[16] A. Sankaranarayanan, P. Turaga, R. Baraniuk, and R. Chellappa, "Compressive acquisition of dynamic scenes," *Computer Vision–ECCV 2010*, pp. 129–142, 2010.

[17] M. S. Asif, L. Hamilton, M. Brummer, and J. Romberg, "Motion-adaptive spatio-temporal regularization (master) for accelerated dynamic mri," 2012, submitted to Magnetic Resonance in Medicine.

[18] J. Zhan and N. Vaswani, "Time invariant error bounds for modified-cs based sparse signal sequence recovery," in *IEEE International Symposium on Information Theory*, July 2013, pp. 286–290.

[19] "Discrete and continuous-time soft-thresholding with dynamic inputs."

[20] T. Zhang, "Some sharp performance bounds for least squares regression with l1 regularization," *The Annals of Statistics*, vol. 37, no. 5A, pp. 2109–2144, 2009.