

# Dynamic Filtering of Time-Varying Sparse Signals via $\ell_1$ Minimization

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

## Abstract

Despite the importance of sparsity signal models and the increasing prevalence of high-dimensional streaming data, there are relatively few algorithms for dynamic filtering of time-varying sparse signals. Of the existing algorithms, fewer still provide strong performance guarantees. This paper examines two algorithms for dynamic filtering of sparse signals that are based on efficient  $\ell_1$  optimization methods. We first present an analysis for one simple algorithm (BPDN-DF) that works well when the system dynamics are known exactly. We then introduce a novel second algorithm (RWL1-DF) that is more computationally complex than BPDN-DF but performs better in practice, especially in the case where the system dynamics model is inaccurate. Robustness to model inaccuracy is achieved by using a hierarchical probabilistic data model and propagating higher-order statistics from the previous estimate (akin to Kalman filtering) in the sparse inference process. We demonstrate the properties of these algorithms on both simulated data as well as natural video sequences. Taken together, the algorithms presented in this paper represent the first strong performance analysis of dynamic filtering algorithms for time-varying sparse signals as well as state-of-the-art performance in this emerging application.

## I. INTRODUCTION

Many signal processing applications involve performing inference on high-dimensional time-varying signals. The past decade of work in signal and image processing has demonstrated that signal priors based on the notion of sparsity (e.g., compressibility in a basis) lead to state-of-the-art performance in many linear inverse problems (e.g., undersampling, inpainting, denoising, compressed sensing, etc. [1]–[3]) across several different application domains (e.g. natural images [1], audio [4], hyperspectral imagery [5], etc.). In this paradigm, a static signal  $\mathbf{x}$  is observed through a linear measurement process  $\mathbf{y} = \Phi\mathbf{x} + \epsilon$ , and the signal is estimated through performing penalized least-squares with a sparsity-inducing regularization function.

Manuscript received June 29, 2015. This work was supported in part by NSF grant CCF-1409422 and the James S. McDonnell Foundation.

A. Charles, A. Balavoine and C. Rozell are with the School of Electrical and Computer Engineering, Georgia Institute of Technology, Atlanta, GA, 30332-0250 USA (e-mail: acharles6,aurele.balavoine,crozell@gatech.edu). The authors would like to thank S. Asif and A. Kressner for valuable feedback, and C. Cadieu for the database of documentary videos.

With the recent increase in streaming data sources, it is critical to develop algorithms that can operate efficiently on these data streams. In *dynamic filtering*, a causal algorithm estimates the current system state recursively based on the previous state estimate (as opposed to batch methods which jointly estimate the time-varying states once all data is collected). Classic dynamic filtering approaches stem from Kalman filtering [6], which has restrictive assumptions (linear systems and Gaussian distributions) and involves large matrix inversion that is inappropriate for high-dimensional data. Given the historical success of dynamic filtering and recent achievements of sparsity models, there are many applications where one can reasonably expect significant gains from jointly leveraging sparsity and models of the signal dynamics. Unfortunately, compared to the individual literatures on sparse signal estimation and dynamic filtering, there is comparatively little work incorporating both types of information jointly. In particular, existing work (reviewed in detail in Section II-C) does not integrate the powerful  $\ell_1$  optimization methods from the sparsity literature with the successful Kalman philosophy of using higher-order statistics to propagate information about past states to the current estimate. Furthermore, existing reports lack strong convergence and performance guarantees.

In this paper we focus on the problem of estimating a time-varying state  $\mathbf{x}_n \in \mathbb{R}^N$  that evolves according to a Markov process

$$\mathbf{x}_n = f_n(\mathbf{x}_{n-1}) + \boldsymbol{\nu}_n, \quad (1)$$

where  $\mathbf{x}_n$  obeys a sparsity model at each time step and  $f_n(\cdot) : \mathbb{R}^N \rightarrow \mathbb{R}^N$  is the operator (assumed known) describing the system evolution with some error  $\boldsymbol{\nu}_n \in \mathbb{R}^N$  (called the *innovation*). The desired state is hidden and only observed through noisy linear measurements  $\mathbf{y} \in \mathbb{R}^M$ :

$$\mathbf{y}_n = \mathbf{G}_n \mathbf{x}_n + \boldsymbol{\epsilon}_n, \quad (2)$$

where  $\mathbf{G}_n \in \mathbb{R}^{M \times N}$  is the (known but potentially time-varying) measurement matrix<sup>1</sup> that has error  $\boldsymbol{\epsilon}_n \in \mathbb{R}^M$  and may have  $M \ll N$ .

In this paper we examine two methods for  $\ell_1$ -based dynamic filtering of sparse signals. We first consider the case where the system dynamics are known with high accuracy. For this case we present a simple and efficient algorithm (BPDN-DF) that combines dynamic penalization with  $\ell_1$  optimization and we present analytic accuracy guarantees on its performance. We then introduce a novel second algorithm (RWL1-DF) that performs better in practice at the cost of a marginally higher computational complexity than BPDN-DF. In particular, RWL1-DF continues to perform well when the system dynamics model is highly inaccurate. RWL1-DF achieves robustness to model inaccuracy via

<sup>1</sup>For static problems we will use  $\Phi$  as the observation matrix and for dynamic problems we will use  $\mathbf{G}$  in order to better differentiate the two cases.

a hierarchical probabilistic data model and propagation of higher-order statistics from the previous estimate in the sparse inference process, much as in traditional Kalman filtering. We demonstrate the properties of these algorithms both in extensive simulations as well as on natural video sequences. In total, we present here both the first strong performance analysis of causal dynamic filtering for temporally evolving sparse signals, as well as state-of-the-art performance in this emerging application.

## II. BACKGROUND AND RELATED WORK

### A. Sparse Signal Estimation

In the sparsity model, we assume the signal of interest can be written  $\mathbf{x} = \mathbf{W}\mathbf{z}$ , where the coefficient vector  $\mathbf{z} \in \mathbb{R}^{N'}$  is mostly composed of zeros and the dictionary  $\mathbf{W}$  may be orthonormal or overcomplete. In the Bayesian interpretation, this model assumes a prior distribution on  $\mathbf{z}$  which has *high kurtosis* to encourage coefficients to be zero (or close to zero). Most commonly, an independent, identically distributed (IID) Laplacian distribution is used as the coefficient prior  $p(\mathbf{z}) = \frac{\gamma}{2} e^{-\gamma \|\mathbf{z}\|_1}$ , where  $\gamma$  is related to the variance of the individual coefficients and  $\|\mathbf{z}\|_1 = \sum_i |z[i]|$  indicates the  $\ell_1$  norm. Under a Gaussian noise assumption on the measurements  $\mathbf{y} = \mathbf{\Phi}\mathbf{x} + \epsilon$ , the corresponding MAP estimate of the signal reduces to the Basis Pursuit De-Noising (BPDN) optimization program

$$\hat{\mathbf{z}} = \arg \min_{\mathbf{z}} \left[ \frac{1}{2\sigma_\epsilon^2} \|\mathbf{y} - \mathbf{\Phi}\mathbf{W}\mathbf{z}\|_2^2 + \gamma \|\mathbf{z}\|_1 \right], \quad (3)$$

where  $\sigma_\epsilon^2$  is the measurement noise variance and the signal estimate is  $\hat{\mathbf{x}} = \mathbf{W}\hat{\mathbf{z}}$ . BPDN has been a particularly popular approach due to its strong performance guarantees [7] and the development of many specialized optimization approaches [8]–[13]. As a guarantee of the measurement quality, we say that  $\mathbf{\Phi}$  satisfies the restricted isometry property with parameters  $2S$  and  $\delta$  (RIP( $2S, \delta$ )) with respect to  $\mathbf{W}$  if for every  $2S$ -sparse vector  $\mathbf{z}$  we have that

$$C(1 - \delta) \|\mathbf{z}\|_2^2 \leq \|\mathbf{\Phi}\mathbf{W}\mathbf{z}\|_2^2 \leq C(1 + \delta) \|\mathbf{z}\|_2^2. \quad (4)$$

If  $\mathbf{z}$  has only  $S$  non-zeros and  $\mathbf{\Phi}$  satisfies RIP( $2S, \delta$ ) for an appropriate value of  $\delta$ , then the BPDN estimation error can be shown to be bounded as

$$\|\mathbf{z} - \hat{\mathbf{z}}\|_2 \leq C_1 \|\epsilon\|_2 + \frac{C_2}{\sqrt{S}} \|\mathbf{z} - \tilde{\mathbf{z}}_S\|_1, \quad (5)$$

where  $C_1$  and  $C_2$  are constants and  $\tilde{\mathbf{z}}_S$  is the best  $S$ -term approximation to  $\mathbf{z}$  [14]–[16]. The RIP condition serves to quantify the quality of the measurements by bounding how similar measurements of different signals could be. The choice of basis  $\mathbf{W}$  is particularly important in ensuring the RIP holds. For orthonormal  $\mathbf{W}$ , simply ensuring that the dot products between the rows of  $\mathbf{\Phi}$  and the columns of  $\mathbf{W}$  are small (i.e. incoherent) is sufficient. In the case of overcomplete dictionaries, the inner products between the columns of  $\mathbf{W}$  must also be small, a property

that can still be found in tight frames.

In (3), the variance parameter  $\gamma$  is assumed known and identical for each coefficient. The RWL1 algorithm [17] improves performance by generalizing these assumptions through a hierarchical probabilistic model known as the Laplacian scale mixture (LSM) [18]. The LSM introduces a second layer of random variables  $\boldsymbol{\lambda} \in \mathbb{R}^{N'}$  representing the variance of the first level Laplacian coefficients  $\mathbf{z}$ , allowing a variable SNR on each coefficient that is also inferred from the data. The LSM uses the Gamma distribution (a conjugate prior for the Laplacian) as the hyperprior on  $\boldsymbol{\lambda}[i]$

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

where  $\Gamma(\cdot)$  is the Gamma function and  $\alpha$  and  $\theta$  are hyper-parameters with  $E[\boldsymbol{\lambda}[i]] = \alpha\theta$  and  $\text{Var}[\boldsymbol{\lambda}[i]] = \alpha\theta^2$ .

Assuming  $\alpha = 1$  and a Laplacian conditional distribution on the coefficients  $p(\mathbf{z}[i]|\boldsymbol{\lambda}[i]) = \frac{\boldsymbol{\lambda}[i]}{2} e^{-\boldsymbol{\lambda}[i]|\mathbf{z}[i]|}$ , the MAP estimate for the full model can be written as

$$\begin{aligned} \{\hat{\mathbf{z}}, \hat{\boldsymbol{\lambda}}\} = \arg \min_{\mathbf{z}, \boldsymbol{\lambda}} & \|\mathbf{y} - \boldsymbol{\Phi} \mathbf{W} \mathbf{z}\|_2^2 + \sum_i |\boldsymbol{\lambda}[i] \mathbf{z}[i]| \\ & - \log \left( \sum_i \boldsymbol{\lambda}[i] \right) + \theta^{-1} \sum_i \boldsymbol{\lambda}[i]. \end{aligned}$$

Though this optimization program is difficult to optimize directly, employing expectation-maximization (EM) [18] results in the RWL1 algorithm [17] where the following equations are iterated until a convergence criteria is met:

$$\begin{aligned} \hat{\mathbf{z}}^t &= \arg \min_{\mathbf{z}} \|\mathbf{y} - \boldsymbol{\Phi} \mathbf{W} \mathbf{z}\|_2^2 + \lambda_0 \sum_i |\boldsymbol{\lambda}^t[i] \mathbf{z}[i]| \\ \boldsymbol{\lambda}^{t+1}[k] &= \frac{\beta}{|\hat{\mathbf{z}}^t[k]| + \eta}, \end{aligned}$$

with  $t$  indicating the iteration number and  $\{\lambda_0, \beta, \eta\}$  constant hyperprior parameters. Several papers have followed up on this RWL1 approach, including a description of it as a compromise between  $\ell_1$  and  $\ell_0$  minimization [19], [20] and an analysis of performance guarantees [21]–[23]. Intuitively, coefficients with little current evidence will have their variances decreased so that future iterations are more likely to estimate them close to zero, and coefficients with strong current evidence will have their variances increased to make future iterations more likely to use those coefficients.

## B. Dynamic Signal Estimation

In dynamic filtering, knowledge of the system dynamics is used to set a prior probability distribution on the current estimate. In the classic Kalman filter setup, the linear and Gaussian assumptions (including the system dynamics  $f_k(\mathbf{x}) = \mathbf{F}_k \mathbf{x}$ ) translates to a Gaussian assumption on the innovations  $\boldsymbol{\nu}_k \sim \mathcal{N}(\mathbf{0}, \mathbf{Q})$  and on the distribution of the

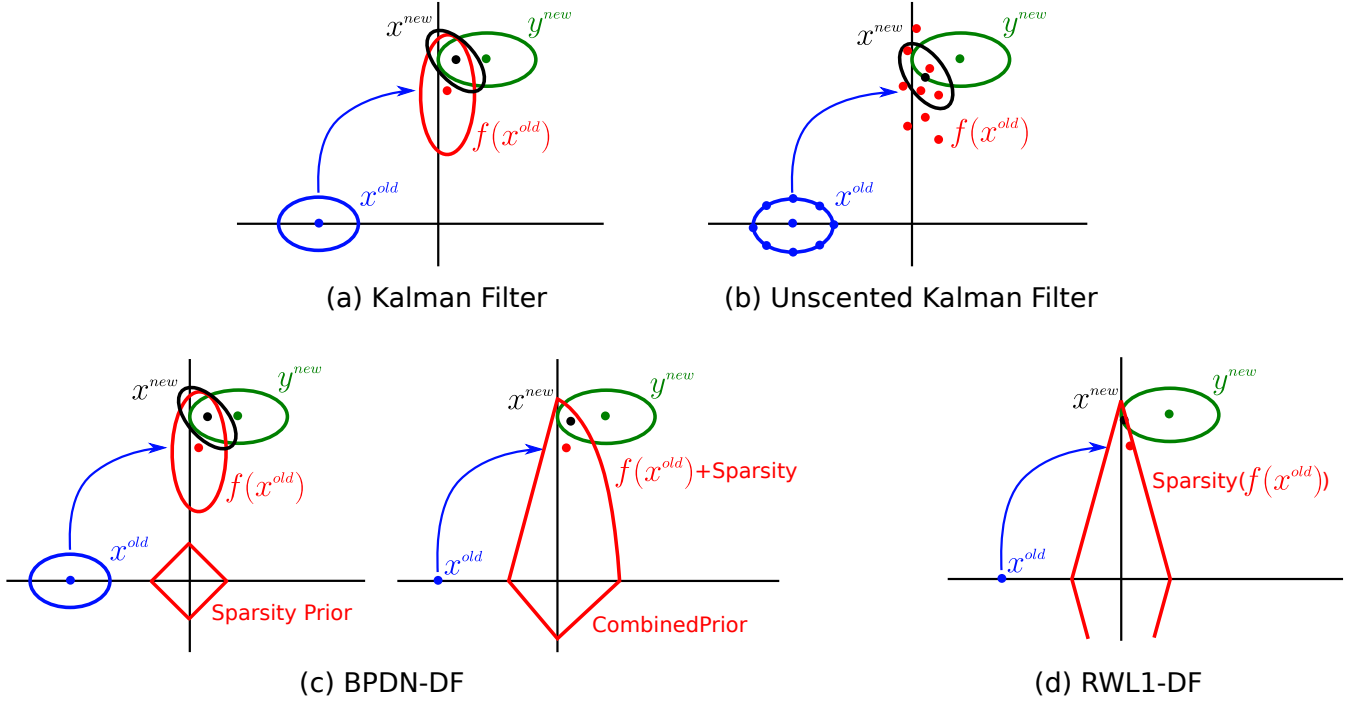


Fig. 1. Information propagation in dynamic filtering algorithms. (a) Classic Kalman filtering propagates the current estimated mean and covariance (blue) through the system dynamics to generate a prior distribution (red) for the next state estimate. This prior is used in conjunction with the likelihood function (green) from the measurements to form a posterior (black) used in estimation. (b) The unscented Kalman filter (and particle filtering generally) estimate the prior distribution via a sampling process that empirically approximates the distribution. (c) BPDN-DF adds a sparsity inducing term to the prior, resulting in a combined prior that is curved outward reminiscent of an  $\ell_2$  ball. (d) RWL1-DF uses the previous estimate to set the parameters of a prior that has the diamond-like shape known to promote sparse solutions.

current state conditioned on the previous state  $p(\mathbf{x}_k|\mathbf{x}_{k-1}) \sim \mathcal{N}(\mathbf{F}_k\mathbf{x}_{k-1}, \mathbf{Q})$ . Similarly, assuming  $\epsilon_k \sim \mathcal{N}(\mathbf{0}, \mathbf{R})$ , the likelihood on the measurements at time  $k$  conditioned on the current state is  $p(\mathbf{y}_k|\mathbf{x}_k) \sim \mathcal{N}(\mathbf{G}_k\mathbf{x}_k, \mathbf{R})$ . Using the Markov property of the state distributions we can write the current state's marginal distribution as  $p(\mathbf{x}_n) \sim \mathcal{N}(\hat{\mathbf{x}}_{n-1}, \mathbf{F}_n\mathbf{P}_{n-1}\mathbf{F}_n^T + \mathbf{Q})$ , where  $\mathbf{P}_n$  is the covariance matrix of the previous state estimate. The Kalman filter infers the current state via a MAP estimate

$$\hat{\mathbf{x}}_n = \arg \min_{\mathbf{x}_n} \left[ \|\mathbf{y}_n - \mathbf{G}_n\mathbf{x}_n\|_{\mathbf{R},2}^2 + \|\mathbf{x}_n - \mathbf{F}_n\hat{\mathbf{x}}_{n-1}\|_{\mathbf{F}_n\mathbf{P}_{n-1}\mathbf{F}_n^T + \mathbf{Q},2}^2 \right], \quad (6)$$

which has an analytic solution (known as the Kalman update equations [6]). The causal estimation procedure is depicted in Figure 1(a). Intuitively, the estimator propagates the previous state estimate (and its covariance matrix) through the system dynamics to get a prediction (i.e., a prior) on the current state. This prior is combined with the likelihood resulting from the current measurement to generate a posterior distribution, which is used to estimate the current state. Though the causal estimator is computationally simpler than joint estimation of the states, it still calculates the same estimate for  $\hat{\mathbf{x}}_n$  as if all of the previous data had been used due to the information propagation

at each iteration via the covariance matrices.

Unfortunately, the analytic simplicity and optimality guarantees of the Kalman filter are highly dependent on the linear and Gaussian model assumptions. Though not optimal estimators, many heuristic approaches have been introduced that follow the spirit of the Kalman filter while incorporating nonlinear system dynamics (e.g., Extended Kalman Filter [24]) or non-Gaussian structure. An alternate approach for highly non-linear functions or non-Gaussian statistics is particle filtering, which uses discrete points (particles) to approximate the relevant distributions and to propagate those distributions through nonlinear dynamics functions. The Unscented Kalman filter [25] (depicted in Figure 1(b)) is similar but uses deterministic (rather than Monte-Carlo) particles. Though particle filtering approaches are general, these methods can become intractable in high-dimensional state spaces and do not explicitly utilize information about sparsity structure.

### C. Sparsity in Dynamic Signals

Recent work has begun to address time-varying sparse signal estimation. One method to leverage both temporal correlations and sparsity is to take a batch processing approach (e.g. [26]–[32]). Batch processing is useful in a number of application, in particular when measurements can be amassed over a time period and analyzed simultaneously. While batch processing can achieve lower estimation errors and permit greater freedom in cost function design, these methods cannot be run causally, excluding them, for instance, from real-time applications. Additionally, operating on larger batches of measurements may require increased computational complexity (e.g. memory or computational time) than online methods. Thus, batch processing techniques are significantly different from our focus on causal (online) estimation. Recent work has also addressed sparse state estimation via efficient updates of previous solutions using specific optimization algorithms, such as homotopy methods [33], [34] or iterative thresholding [16]. While utilizing the dynamic information implicitly to reduce computational complexity, these approaches do not attempt to improve the estimator quality through modeling signal dynamics.

Another branch of previous work essentially modifies the Kalman filter equations to induce sparsity in the estimate. One approach in this vein employs a pseudo-norm in the Kalman update equations to encourage sparser solutions [35] and then enforces an  $\ell_1$  constraint on the state after the Kalman update. Another approach [36], [37] works in two-steps, first performing support estimation using an  $\ell_1$  cost function and then running the traditional Kalman equations on a restricted support set. While the work in [36], [37] assumes sparse innovations, the robustness to model mismatch has not been fully explored. As will be shown in Section V, these approaches appear to lack robustness to the innovations statistics, perhaps due to the support set estimation being particularly sensitive to estimation errors. From a computational perspective, storing and inverting full covariance matrices is also prohibitive for the high-dimensional data where sparsity models have been most successful.

More recent approaches have considered direct coefficient transition modeling via Markov models [38], [39], either by utilizing message passing to propagate support information through time [39] or by using the previous estimate to influence coefficient selection through modified orthogonal matching pursuit [38]. These approaches do not rely on solving complex optimization problems and so are extremely computationally efficient. However, these approaches either restrict the dynamics model to specific applications [39] (i.e. the approach as specified and implemented restricts the dynamics function to  $f(\mathbf{x}) = \mathbf{x}$ ) or have limited robustness due to the fact that they strictly enforce a support set estimate [38]. Additionally, they currently lack strong accuracy guarantees.

In another direction, some proposed algorithms (e.g. [40]–[44]) essentially modify the BPDN objective function to include a term for the dynamic state prediction, including the BPDN-DF algorithm analyzed in the next section. Such approaches are simple and can work well, but lack robustness to model mismatch due to their explicit strong assumptions on the innovations statistics (see the discussion in [43]). Until now, these approaches have also lacked strong accuracy guarantees.

### III. BASIS PURSUIT DE-NOISING WITH DYNAMIC FILTERING

The first algorithm we explore is *basis-pursuit de-noising dynamic filtering* (BPDN-DF) [42], [43] (similar to [27], [40]), which is a simple and efficient method that can work well when the dynamics are known. BPDN-DF modifies the BPDN estimator to include a term from the Kalman filter optimization that enforces consistency with the model prediction from the previous state estimate

$$\begin{aligned} \hat{\mathbf{z}}_n = \arg \min_{\mathbf{z}} & \|\mathbf{y}_n - \Phi \mathbf{W} \mathbf{z}\|_2^2 + \gamma \|\mathbf{z}\|_1 \\ & + \kappa \|\mathbf{W} \mathbf{z} - f(\hat{\mathbf{x}}_{n-1})\|_2^2, \end{aligned} \quad (7)$$

where  $\gamma$  and  $\kappa$  are tradeoff parameters and the state is recovered via  $\hat{\mathbf{x}}_n = \mathbf{W} \hat{\mathbf{z}}_n$ . Thus BPDN-DF seeks a sparse vector that matches both the measurements and the causal prediction obtained via the known dynamics model. As depicted in Figure 1(c), BPDN-DF combines the penalty for deviations from the dynamic prediction  $f(\hat{\mathbf{x}}_{n-1})$  with the typical measurement likelihood and the sparsity penalties from BPDN to obtain an overall cost function for estimating the current state.

BPDN-DF has shown empirical performance improvements over standard Kalman filters and independent BPDN estimation at each time-step (i.e., using BPDN at each time step without temporal regularization) [43]. Additionally, the algorithm is simple and requires a similar computational complexity as standard BPDN, making it appropriate for high-dimensional data. However, while BPDN-DF does pass first-order information forward from the previous state to the current estimator, this procedure creates a combined regularizer with a “rounded” shape in some directions (illustrated in Figure 1(c)). This bulging of the combined prior starts to resemble the  $\ell_2$  ball, and is in conflict with

the geometric structures known to produce sparse solutions (e.g., the  $\ell_1$  ball) [3], [45]. This can cause robustness problems in BPDN-DF when there are inaccuracies in the dynamics model (demonstrated later).

Aware of these potential drawbacks, analytic performance guarantees would be valuable because BPDN-DF remains an effective and computationally attractive algorithm for some applications. Prior work in [44] provided guarantees on the statistical regret of an estimator similar to BPDN-DF called the dynamic mirror descent (DMD). Although the DMD is more general in terms of its formulation, the guarantees it provides (i.e., bounded regret) are in terms of a comparison with an optimal non-dynamical cost function and are therefore difficult to directly relate to the cost function in Equation 7. While our preliminary analysis of BPDN-DF using simple techniques provided some accuracy guarantees, these bounds were too loose to be an accurate reflection of the estimator performance [46]. In this work we adopt a proof technique similar to [16], [47] which provides much stronger accuracy guarantees for BPDN-DF, as summarized in the following theorem.

**Theorem III.1.** *Suppose that at each time  $n$ ,  $\mathbf{G}_n$  satisfies  $\text{RIP}(S + 2q, \delta)$  for some constant  $q \geq 0$ , and the error and innovations satisfy  $\|\boldsymbol{\epsilon}_n\|_2 \leq \bar{\epsilon}$  and  $\|\boldsymbol{\nu}_n\|_2 \leq \bar{\nu}$ . Furthermore, suppose that for all  $\mathbf{x}_1, \mathbf{x}_2$ , the dynamics satisfies  $\|f(\mathbf{x}_1) - f(\mathbf{x}_2)\|_2 \leq f^* \|\mathbf{x}_1 - \mathbf{x}_2\|_2$  for a universal constant  $f^*$ , and the coefficients at each  $n$  satisfy  $\|\mathbf{z}_n\|_2 \leq b$ . If  $\gamma > 0$ ,  $\kappa > 0$  are known constants and the following condition is met:*

$$\begin{aligned} \kappa((1 - f^*)b - (1 + \delta)\gamma\sqrt{q} - \bar{\nu}) \geq \\ (1 + \delta)\gamma\sqrt{q} + \sqrt{1 + \delta}\bar{\epsilon} - (1 - \delta)b, \end{aligned}$$

then the solution to Equation (7) satisfies

$$\|\hat{\mathbf{z}}_n - \mathbf{z}_n\|_2 \leq \beta^n \|\mathbf{e}_0\|_2 + (1 - \beta^n)(C_1\sqrt{q} + C_2\bar{\epsilon} + C_3\bar{\nu})$$

where the constants are defined as:

$$\begin{aligned} \beta &= \frac{\kappa f^*}{1 + \kappa - \delta} \\ C_1 &= \frac{(1 + \delta)(1 + \kappa)\gamma}{1 + \kappa - \delta - \kappa f^*} \\ C_2 &= \frac{\sqrt{1 + \delta}}{1 + \kappa - \delta - \kappa f^*} \\ C_3 &= \frac{\kappa}{1 + \kappa - \delta - \kappa f^*}. \end{aligned} \tag{8}$$

The full proof of Theorem III.1 is outlined in Appendix VII-A. Note that the error bound has two distinct components: a transient term depending on the initial uncertainty that decays if  $\beta^n \rightarrow 0$ , and a steady state term that emerges depending on the inherent difficulty of the problem (i.e., the noise sources and the sparsity of the



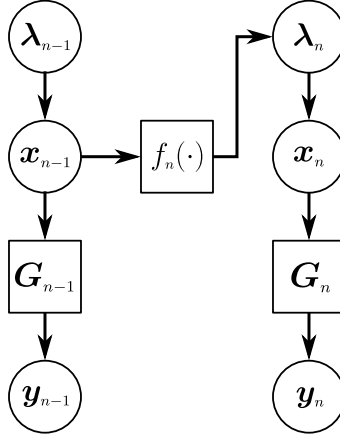


Fig. 2. The RWL1-DF algorithm inserts the dynamic information at the second layer of the LSM model for each time step. The graphical model depicts the model dependencies, where prior state estimates are used to set the hyperpriors for the second level variables controlling the variances (i.e., SNRs) of the state estimates at the next time step.

signal). If  $\kappa = 0$  (removing the dynamics in the estimation), we obtain exactly the results in [16] for solving BPDN with no dynamic filtering term. Alternatively, as  $\gamma \rightarrow 0$ , the result in Theorem III.1 ceases to hold as convergence is no longer guaranteed. Specifically,  $\gamma = 0$  violates a condition in Lemma VII.1, which is required to prove algorithmic convergence. We note that the same condition in Lemma VII.1 necessitates the condition  $\|z_n\|_2 \leq b$ , indicating that  $b \rightarrow \infty$  can also cause a loss of convergence, and thus accuracy, guarantees.

Theorem III.1 implies that BPDN-DF is only guaranteed to converge if  $\beta < 1$ . Since most parameters are system or signal dependent (and therefore not controllable by the user), we interpret this as a condition on the parameters  $\gamma$  and  $\kappa$ . In particular, since  $\gamma$  does not appear in the expression for  $\beta$ , we can consider this to be a bound on  $\kappa$ ,

$$\kappa < \frac{1 - \delta}{f^* - 1} \quad \text{if } f^* > 1. \quad (9)$$

This condition essentially compares the smoothness of the dynamics function with the RIP conditioning of the measurements. For example, as  $\delta$  approaches 1, the allowable range of  $\kappa$  shrinks indicating that the dynamics should be deemphasized in BPDN-DF. Alternatively, as  $f^*$  approaches 1, the range of  $\kappa$  increases, indicating that the dynamics can be emphasized in the optimization cost. Interestingly, if  $f^* \leq 1$ , the requirement that  $\beta < 1$  induces no additional restrictions on  $\kappa$  as the numerator of Equation (9) is positive. With a negative denominator in Equation (9), this condition is redundant with the prior requirement that  $\kappa$  is positive. The steady-state error bound depends on both the maximum measurement error energy  $\bar{\epsilon}$  and the maximum innovations energy  $\bar{\nu}$ . The parameter  $\kappa$  trades off between the two, where the trade-off takes into account the RIP conditioning  $\delta$  as well as the dynamics smoothness  $f^*$ .

#### IV. RE-WEIGHTED $\ell_1$ DYNAMIC FILTERING

Despite now having provable guarantees, performance of BPDN-DF degrades with increased inaccuracy in the dynamics model. In fact, Theorem III.1 requires that the innovations (a measure of the dynamics inaccuracy) satisfies certain inequalities with respect to the measurement noise and sparsity. In this section we present a novel algorithm called *re-weighted  $\ell_1$  dynamic filtering* (RWL1-DF) that uses a richer statistical model to improve performance, especially when the dynamics model is inaccurate. In contrast to existing algorithms (including BPDN-DF), RWL1-DF displays a combination of flexibility (allowing arbitrary dynamic models), computational efficiency (leveraging advances in  $\ell_1$  minimization), and robustness to dynamic model errors caused by sparse innovations. The main technical innovation we propose is to use the hyper-priors in the LSM model to incorporate a dynamic prediction into second-order statistics for the next estimate in a causal fashion (akin to Kalman filtering). The resulting estimation procedure is illustrated in Figure 1(d) and the graphical model is depicted in Figure 2. While some existing literature [41], [48]–[50] has proposed re-weighted estimation schemes to include prior information in a limited way (e.g., using only binary weights based on support set estimates), RWL1-DF modifies RWL1 specifically for dynamic filtering and uses a general probabilistic model that improves robustness by allowing arbitrary weightings.

Technically, similar to the LSM model, our proposed model describes the conditional distribution on the sparse coefficients  $z$  at time  $n$  as a zero-mean Laplacian with different variances:

$$p(z_n[i] | \lambda_n) = \frac{\lambda_0 \lambda_n[i]}{2} e^{-\lambda_0 \lambda_n[i] |z_n[i]|}.$$

We also set the scale variables controlling the coefficient variance to be Gamma distributed:

$$p(\lambda_n[i]) = \frac{\lambda_n^{\alpha-1}[i]}{\theta_n^\alpha[i] \Gamma(\alpha)} e^{-\lambda_n[i]/\theta_n[i]},$$

and allow each of these variables to have a different expected value, ( $\alpha \theta_n[i] = E[\lambda_n[i]]$ ) by modifying  $\theta_n[i]$ . To introduce dynamic information, the expected value of each scale variable is set based on the previous state estimate. In particular, if the model prediction of  $z_n[i]$  based on the previous state is large (or small), the variance of that coefficient at the current estimate is made large (or small) by making  $\lambda_n[i]$  small (or large). Large variances allow the model flexibility to choose from a wide-range of non-zero values for coefficients that are likely to be active and small variances (with a mean of zero) encourage the model to make such coefficients inactive. By “encouraging” the model through the use of second order statistics (instead of forcing the model to use an estimated coefficient subset) the model remains robust and flexible. While there are many possible choices for the mapping of predictions into the model at the next time step, in this paper we choose

$$\theta_n[i] = \xi (|\mathbf{W}^{-1} f_n(\mathbf{W} \hat{\mathbf{z}}_{n-1})[i]| + \eta)^{-1},$$

where  $\mathbf{W}^{-1}f_n(\mathbf{W}\hat{\mathbf{z}}_{n-1})[i]$  is the  $i^{\text{th}}$  coefficient of the previous signal propagated through the dynamics,  $\eta$  is a linear offset and  $\xi$  is a multiplicative constant. Note that any general model for the dynamics is allowable and we are not restricted to linear dynamical systems in the prediction. Note also that the absolute values are necessary because  $\lambda$  determines a variance, which must be strictly positive. The parameter  $\eta$  determines the distribution of the variance when the coefficient is predicted to be zero, resulting in  $\mathbf{W}^{-1}f_n(\mathbf{W}\hat{\mathbf{z}}_{n-1})[i] = 0$ . This parameter reflects the magnitude of the innovations in the erroneous model predictions.

Under this model, the joint MAP estimate of all model parameters becomes

$$\begin{aligned} [\hat{\mathbf{z}}_n, \hat{\boldsymbol{\lambda}}_n] &= \arg \min_{[\mathbf{z}, \boldsymbol{\lambda}]} \|\mathbf{y}_n - \mathbf{G}_n \mathbf{W} \mathbf{z}\|_2^2 + \beta \sum_i |\boldsymbol{\lambda}[i] \mathbf{z}[i]| \\ &\quad - \alpha \sum_i \log(\boldsymbol{\lambda}[i]) \\ &\quad + \frac{\lambda_0}{\xi} \sum_i \boldsymbol{\lambda}[i] (|\mathbf{W}^{-1}f_n(\mathbf{W}\hat{\mathbf{z}}_{n-1})[i]| + \eta). \end{aligned} \quad (10)$$

As in the LSM case, the optimization in (10) is not easily solved for both  $\mathbf{z}_n$  and  $\boldsymbol{\lambda}_n$  jointly, but the model yields a simple form when using an EM approach. The precise steps of the iteration in this case are

$$\begin{aligned} \text{E step:} \quad & \hat{\boldsymbol{\lambda}}_n^t = E_{p(\boldsymbol{\lambda}|\hat{\mathbf{z}}_n^t)}[\boldsymbol{\lambda}] \\ \text{M step:} \quad & \hat{\mathbf{z}}_n^t = \arg \min_{\mathbf{z}_n} -\log \left[ p(\mathbf{z}_n | \hat{\boldsymbol{\lambda}}) \right] \end{aligned}$$

where  $t$  denotes the EM iteration number and  $E_{p(\boldsymbol{\lambda}|\hat{\mathbf{z}}_n^t)}[\cdot]$  denotes the expectation with respect to the conditional distribution  $p(\boldsymbol{\lambda}|\hat{\mathbf{z}}_n^t)$ . We can write the maximization step as

$$\hat{\mathbf{z}}_n^t = \arg \min_{\mathbf{z}} \|\mathbf{y}_n - \mathbf{G} \mathbf{W} \mathbf{z}\|_2^2 + \lambda_0 \sum_i \left| \hat{\boldsymbol{\lambda}}^t[i] \mathbf{z}[i] \right|, \quad (11)$$

since the MAP optimization conditioned on the  $\lambda$  parameters reduces to a weighted  $\ell_1$  optimization.

The conjugacy of the Gamma and Laplacian distributions admit a simple closed-form solution for the expectation step in this model that is similar to the original LSM model [18],

$$\boldsymbol{\lambda}_n^{t+1}[i] = \frac{2\tau}{\beta |\hat{\mathbf{z}}_n^t[i]| + |\mathbf{W}^{-1}f_n(\mathbf{W}\hat{\mathbf{z}}_{n-1})[i]| + \eta}, \quad (12)$$

where  $\tau = (\alpha + 1)\xi$  is a constant scaling value,  $\beta = \lambda_0\xi$  can be interpreted as a tradeoff between the measurement and the prediction, and the signal of interest can again be recovered via  $\hat{\mathbf{x}}_n = \mathbf{W}\hat{\mathbf{z}}_n$ . The resulting procedure that iterates between Equations (11) and (12) looks nearly identical to the static RWL1 algorithm except that the denominator in the  $\boldsymbol{\lambda}$  update contains a term depending on the previous state. This term encourages smaller  $\boldsymbol{\lambda}$  values (i.e., higher variances) in the elements that are predicted to be highly active according to  $f_n(\hat{\mathbf{x}}_{n-1})$ . This graduated encouragement of coefficients selected by the prediction allows the algorithm to perform especially well when the

states and innovations are sparse while retaining good performance when the innovations are denser. Furthermore, the simple form means that the recent work in  $\ell_1$  optimization methods can be leveraged for computationally efficient recursive updates, requiring only a small number of  $\ell_1$  optimizations at each time step. In particular, since no covariance matrix inversion is required and many modern  $\ell_1$  estimation methods require only matrix multiplication (and no inversion), this approach is also amenable to high-dimensional data analysis.

Despite being highly nonlinear, RWL1-DF has demonstrable stability and convergence properties. First, the RWL1-DF update equations in (11) and (12) directly guarantee that the coefficient and variance estimates will be bounded. To see this for an arbitrary time step  $n$ , note that the variance estimates are within the range  $\lambda_n^{t+1}[i] \in (0, 2\tau/\eta]$  for each EM iteration. With bounded variances, the weighted BPDN optimization in (12) will yield a solution where the output coefficients are also bounded. Second, existing guarantees for the EM algorithm guarantee that the EM iterations in the proposed algorithm (i.e., an estimate at a single time step) have coefficient differences that asymptotically converge to zero,  $\lim_{t \rightarrow \infty} \|z_n^t - z_n^{t+1}\| = 0$  [51]. Existing stronger convergence guarantees (from EM or from the very limited analysis on RWL1 algorithms [21]) do not apply in this case because of the technical details of RWL1-DF. Despite the lack of stronger convergence guarantees, the numerical results in section V demonstrate that the algorithm converges in practice with just a few EM iterations.

## V. SIMULATION RESULTS

While the dynamic filtering approaches discussed here are general inference tools with many potential applications, we focus our evaluation on compressed sensing (CS) recovery as an example test case. In all simulations we compare the performance with existing algorithms discussed in Section II-C when possible, noting in each particular case where algorithms were unable to be evaluated because they are incompatible or computationally prohibitive. Standard Kalman filtering is not shown because it performs very poorly in these type of simulations (i.e., it doesn't converge to a stable estimate with the sparse statistics of the applications we use) [43]. The performance of independent BPDN (Equation (3)) and independent RWL1 (Equations (6)), each applied independently at each time step with no temporal information, are also shown to highlight the benefit of including dynamic information.

### A. Stylized tracking scenario

To explore the performance and robustness of the algorithms in detail, we first perform inference on synthetic data that simulates a stylized tracking scenario. The use of synthetic data provides us with “ground truth” so we can make controlled variations of the data characteristics. In this setup, we generate an image with  $S$  non-zero moving pixels of various intensities that move with time and represent targets that must be tracked. The movement of these non-zero pixels  $F_k$  is specified to be constant motion, and the simulated dynamics includes a sparse innovations term (i.e., dynamic model error) that causes target motion to change in each time step for some percentage of the

pixels  $p$ . In other words, at every time step there is a probability  $p$  of each target abruptly changing directions to violate the dynamics model assumed by the inference algorithms. This process simulates an innovation that is approximately  $2Sp$ -sparse at every iteration, allowing us to evaluate the algorithm’s robustness to a type of model mismatch (i.e., shot noise) that is particularly challenging for Kalman filter techniques.

In detail, we create  $24 \times 24$  pixel videos ( $N = 24^2$ ) with 20 moving particles ( $S = 20$ ). The vectors are observed with Gaussian measurement matrices (with normalized columns) that are independently drawn at each iteration, and we add Gaussian measurement noise with variance  $\sigma_\epsilon^2 = 0.001$ . We vary the number of measurements to observe the reconstruction capability of the algorithm in highly undersampled regimes, but the number of measurements per time step is always constant within a trial. All simulations average the results of 40 independent runs and display reconstruction results as the relative mean-squared error (rMSE) for each frame, calculated as:

$$\frac{\|\mathbf{x}_k - \hat{\mathbf{x}}_k\|_2^2}{\|\mathbf{x}\|_2^2}. \quad (13)$$

For independent BPDN, at each iteration we use the value  $\lambda = 0.55\sigma_\epsilon^2$ . For independent RWL1 we use  $\lambda_0 = 0.0011$ ,  $\tau = 1$  and  $\nu = 0.01$ . For BPDN-DF we use  $\gamma = 0.5\sigma_\epsilon^2$  and  $\kappa = 0.0007/(p+1)$ . For RWL1-DF we use  $\lambda_0 = 0.0011$ ,  $\tau = 1$  and  $\nu = 1 - 2p/S$ . These parameters were optimized using a manual parameter sweep. Furthermore, for comparison we also show the performance of an optimal oracle least-squares solution, where the support at each iteration is known ( $\mathbf{G}_k$  becomes overdetermined). Note that this scenario is particularly challenging for many existing algorithms because of the arbitrary model dynamics (e.g.,  $\mathbf{F}_k \neq I$ ) that may vary with time. In particular, comparisons with the algorithms described in [40], [41], [48], [49] (or modifications of them to accommodate arbitrary dynamics) were attempted and are not shown here because they still performed significantly worse than static estimation (e.g., BPDN) even after extensive searching for good parameter settings.

The iterative algorithms based on RWL1 are stopped when the relative norm-squared difference between coefficients at consecutive iterations  $\|z_n^t - z_n^{t+1}\|_2^2 / \|z_n^t\|_2^2$  falls below a specified threshold (we use 0.1%), which typically requires just a few (typically 5-7) iterations. Figure 3 shows the relative coefficient change over EM iteration, demonstrating that RWL1 (i.e. without dynamics) convergence occurs by 10 iterations and RWL1-DF actually converges faster due to the improved performance from incorporating dynamic information. Figure 4(a) shows a single trial with  $M = 80$  measurements and  $2Sp = 5$  innovation errors at each time step. Note that RWL1-DF reaches substantially lower steady-state recovery error than BPDN-DF, illustrating the net improvements gained by using second-order statistics in the estimation. Figure 4(b) displays the results of varying the number of measurements while holding  $2Sp = 5$ . While performance for all algorithms becomes comparable for large numbers of measurements, it is clear that exploiting temporal information can most improve performance in the

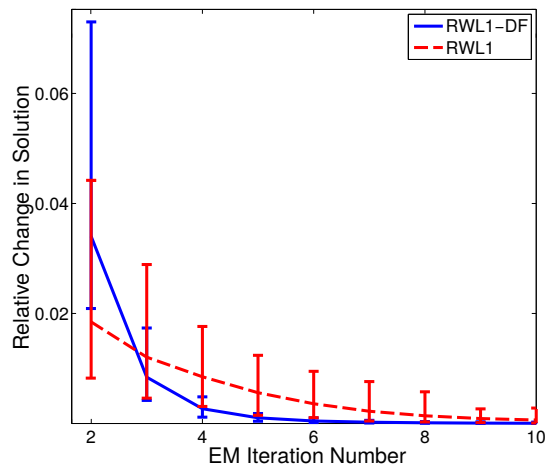


Fig. 3. Convergence of RWL1 and RWL1-DF. The mean relative change (over 200 frames) in the coefficients is plotted, with the error bars indicating the maximum and minimum values. The relative norm difference of the coefficients in the RWL1 and RWL1-DF algorithms falls quickly over the first 10 iterations. The dynamic information helps the RWL1-DF algorithm converge faster, requiring approximately 5-7 iterations to converge.

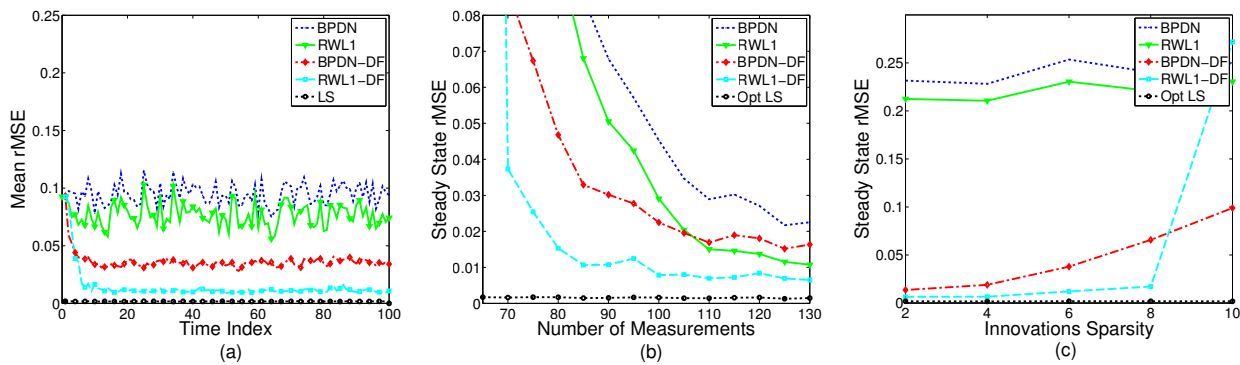


Fig. 4. Behavior of the RWL1-DF algorithm on synthetic data. (a) RWL1-DF converges to a lower mean rMSE than static sparse estimation or BPDN-DF. Shown for  $M = 80$ ,  $N = 576$ ,  $S = 20$  and  $p = 0.25$ . (b) When sweeping the number of measurements  $M$  for  $N = 576$ ,  $S = 20$ , and  $p = 0.25$ , we observe that the performance improvement for RWL1-DF is especially distinct in the highly undersampled regime. Each point is the average steady state rMSE over 40 independent trials. (c) RWL1-DF is also more robust to model mismatch in the innovation statistics. Shown here for different innovations sparsity ( $2Sp$ ) for  $M = 70$ ,  $N = 576$ , and  $S = 20$ . Each data point is the result of averaging the steady-state rMSE over 40 independent trials. Note that when BPDN-DF starts to perform better ( $2Sp = 10$ ), the innovations are actually half of the total support set and a Gaussian innovations model may be more accurate than a sparse innovation model.

highly undersampled regime. In particular, RWL1-DF is able to sustain virtually the same steady-state rMSE down to much more aggressive levels of undersampling than BPDN-DF. Finally, we explore the robustness of each algorithm to model errors by fixing the number of measurements ( $M = 70$ ) and varying the sparsity of the innovations  $2Sp$ . Figure 4(c) shows the results, illustrating that RWL1-DF uses the second-order statistics to sustain better performance than BPDN-DF when the innovations are sparse (i.e., shot noise). We note that when  $2Sp > 8$ , the total number of model errors is 50% of the signal sparsity, indicating both that the innovation is dense, and that the innovations has as much, or more, energy than the signal. The combination of the innovations statistics mismatch (sparse vs. dense) and reduced predictive power of the dynamics model produces a phase transition where RWL1-DF is no longer able to correct large innovation values.

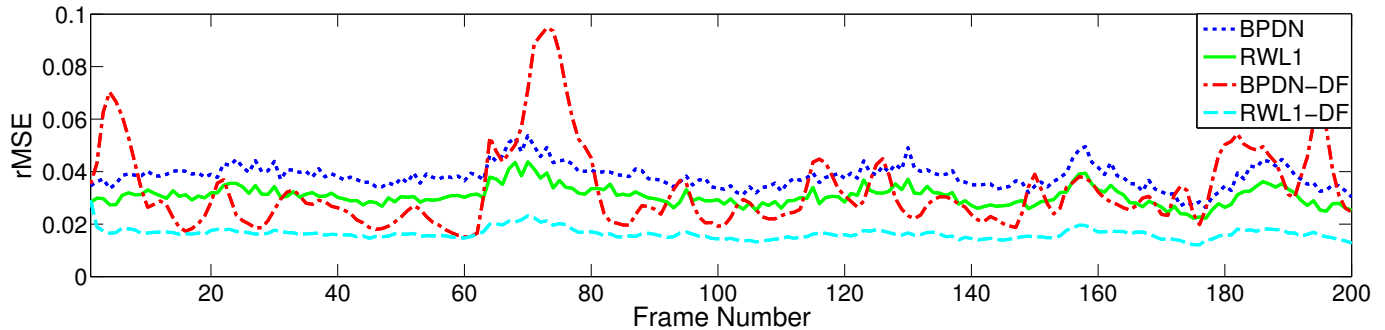


Fig. 5. CS recovery of the full Foreman video sequence. Each curve represents the rMSE for recovery from subsampled noiselets ( $M/N = 0.25$ ) using the DT-DWT as the sparsifying basis. The independent BPDN recovery (dotted blue curve) and the independent re-weighted BPDN (solid green curve) retain a steady rMSE over time. RWL1-DF (dashed cyan curve) converges on a lower rMSE than either time-independent estimation and remains at approximately steady-state for the remainder of the video sequence. BPDN-DF (the dot-dash red curve) can converge to low rMSE values, but is highly unstable and can yield very poor results when the model is not accurate due to motion in the scene.

### B. CS recovery of natural video sequences

To test the utility of RWL1-DF on natural signals, we explore its performance on a simulation of compressively sampled natural video sequences. These results will report in-depth comparison of a single challenging video sequence (the Foreman sequence<sup>2</sup>) as well as aggregate statistics from a batch of video from a BBC nature documentary (as used in [52]). The documentary footage is valuable as broad comparison because it contains many different types of motion, including static frames with localized changes and highly dynamic frames with moving subjects across large portions of the visual field.

In our simulation of CS video recovery, we take the time-varying hidden state  $\mathbf{x}_n$  to be the wavelet (synthesis) coefficients at each frame of the video. While the true frame-by-frame dynamics of natural video are likely to be complex and non-linear, for this simulation we use a simple first-order model predicting that the coefficients will remain the same from one frame to the next:  $f_k(\mathbf{x}) = \mathbf{x}$  for all  $k$ . Improvements to this simple model would improve any algorithm that can incorporate more advanced models, including RWL1-DF which allows arbitrary dynamics models. A benefit to assuming stationary dynamics is that it allows us to compare recovery performance with a number of existing algorithms that do not currently have arbitrary dynamics as part of the approach, including DCS-AMP [53] and modCS [41]. To demonstrate the advantage of using the coefficient values instead of only support information, we also compare to a modified version of the binary-weighted BPDN algorithm described in [48] where we have added support set prediction based on model dynamics. We call this approach weighted  $\ell_1$  with prior information (WL1P).

We use a four-times overcomplete dual-tree discrete wavelet transform (DT-DWT) [54] as the sparsity basis and we simulate CS measurements by applying a subsampled noiselet transform [55] to each frame. We take  $M =$

<sup>2</sup>The Foreman sequence is available at: <http://www.hlevkin.com/TestVideo/foreman.yuv>

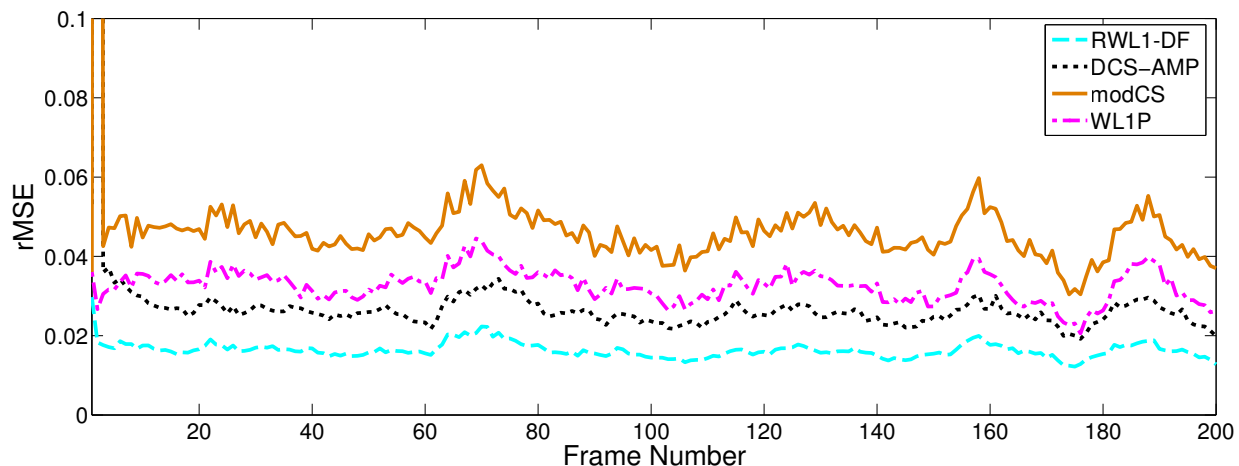


Fig. 6. A comparison of RWL1-DF with existing recovery algorithms (DCS-AMP, modCS and WL1P) for the Foreman video sequence. Each curve represents the rMSE for recovery from subsampled noiselets ( $M/N = 0.25$ ) using the DT-DWT as the sparsifying basis.

$0.25N$  measurements per frame (where  $N = 128^2$ ) with a measurement noise variance of  $10^{-4}$ .<sup>3</sup> We solve all optimization programs using the TFOCS package [57] due its stability during RWL1 optimization and the ability to use fast implicit operators for matrix multiplications. In the Foreman video sequence, we simulate CS measurements on a  $128 \times 128$  pixel portion of the video, using parameters  $\lambda = 0.001$  for BPDN,  $\lambda_0 = 0.11$ ,  $\tau = 0.2$  and  $\nu = 0.1$  for RWL1,  $\gamma = 0.01$  and  $\kappa = 0.2$  for BPDN-DF, and  $\lambda_0 = 0.003$ ,  $\tau = 4$ ,  $\beta = 1$  and  $\nu = 1$  for RWL1-DF. These parameters were found using a manual parameter sweep to optimize performance for each algorithm and number of CS measurements. While the number of measurements was fixed in this simulation for computational tractability, recovery performance could be altered for all algorithms by adjusting the number of measurements.

Figure 5 shows the recovery of 200 consecutive frames of video in the Foreman sequence. We see that RWL1-DF converges to the lowest steady-state rMSE and is able to largely sustain that performance over the sequence. In contrast, BPDN-DF cycles through periods of good performance and poor performance, sometimes performing worse than not using temporal information at all. In essence, BPDN-DF is not robust to model errors, and each time there is motion in the scene (violating the simple dynamics model) the algorithm has to re-converge. The RWL1-DF approach does not exhibit this performance oscillation because the use of second-order statistics to propagate temporal information is less rigid, allowing for more robustness during model errors. Figure 6 shows a comparison of recovery for the Foreman video sequence across several existing algorithms, including DCS-AMP, modCS and WL1P. Again we optimize algorithms parameters manually to achieve the best aggregate performance over the video sequence. To summarize the performance over the entire video sequence, Figure 7 shows histogram plots of the rMSE values over the 200 frames for each recovery. The mean and median for each histogram are

<sup>3</sup>We use the noiselet transform because it can be computed with an efficient implicit transform and has enough similarity to a random measurement that it works well in CS [56].



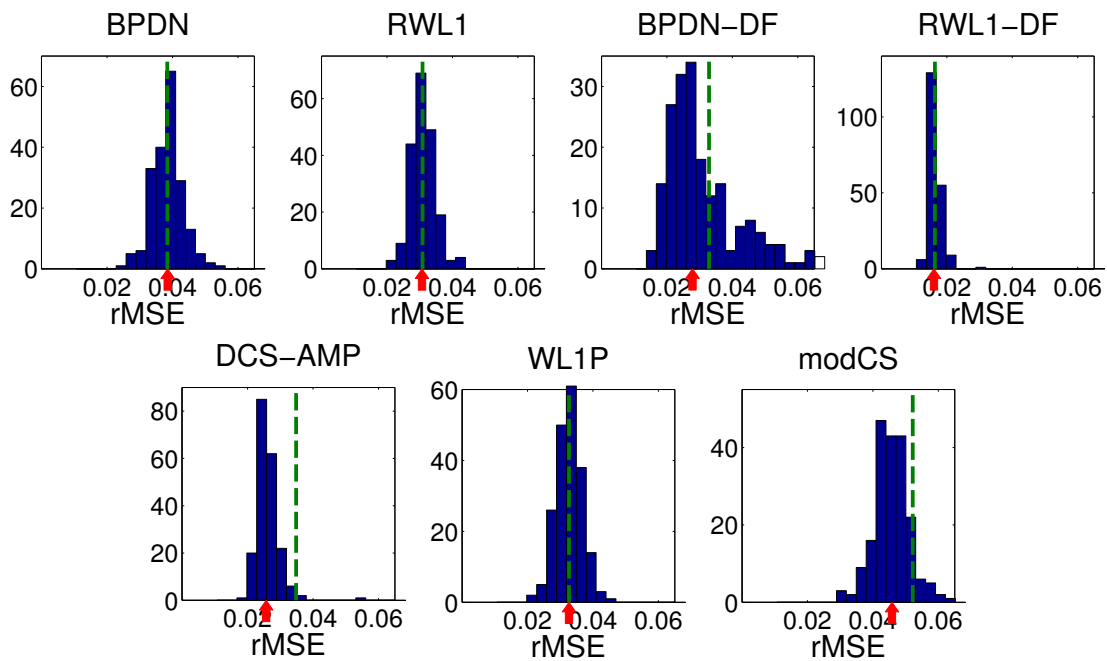


Fig. 7. Histogram of the rMSE for the compared algorithms when recovering the Foreman video sequence with the DT-DWT as the sparsifying basis. RWL1-DF achieves a lower mean (indicated by the dashed green lines) and median (indicated by the red arrows), with a tightly concentrated error distribution due to the robustness to model mismatch (producing few outliers). Specific mean and median values are shown in Table I.

	BPDN	RWL1	BPDN-DF	RWL1-DF	DCS-AMP	WL1P	modCS
Mean rMSE	3.84%	3.09%	3.29%	<b>1.63%</b>	3.48%	3.27%	5.22%
Median rMSE	3.85%	3.07%	2.78%	<b>1.61%</b>	2.57%	3.27%	4.58%

TABLE I  
MEAN AND MEDIAN VALUES FOR COMPRESSIVE RECOVERY OF THE FOREMAN VIDEO SEQUENCE.

represented by the green dashed line and the red arrow respectively, and are listed in Table I. The recovery errors for BPDN-DF are significantly spread out, achieving nearly the same median error as the independent algorithms and having a much higher mean error due to the large excursions during model mismatch. BPDN-DF actually may be a worse choice than not using any temporal information at all when model errors are present. In contrast, RWL1-DF shows a much tighter distribution of errors, having a mean and median significantly lower than alternate approaches.

In addition to the in-depth comparison on the single Foreman video sequence, we also perform the same CS recovery task on a database of video sequences from a nature BBC documentary to investigate the performance across a wider range of video characteristics (i.e., including video clips with localized motion and global motion in the scene). We simulated CS measurements for 24 sequences (48-frames each) in the same manner as the Foreman sequence and recovered the frames using the same methodology described above (including parameters

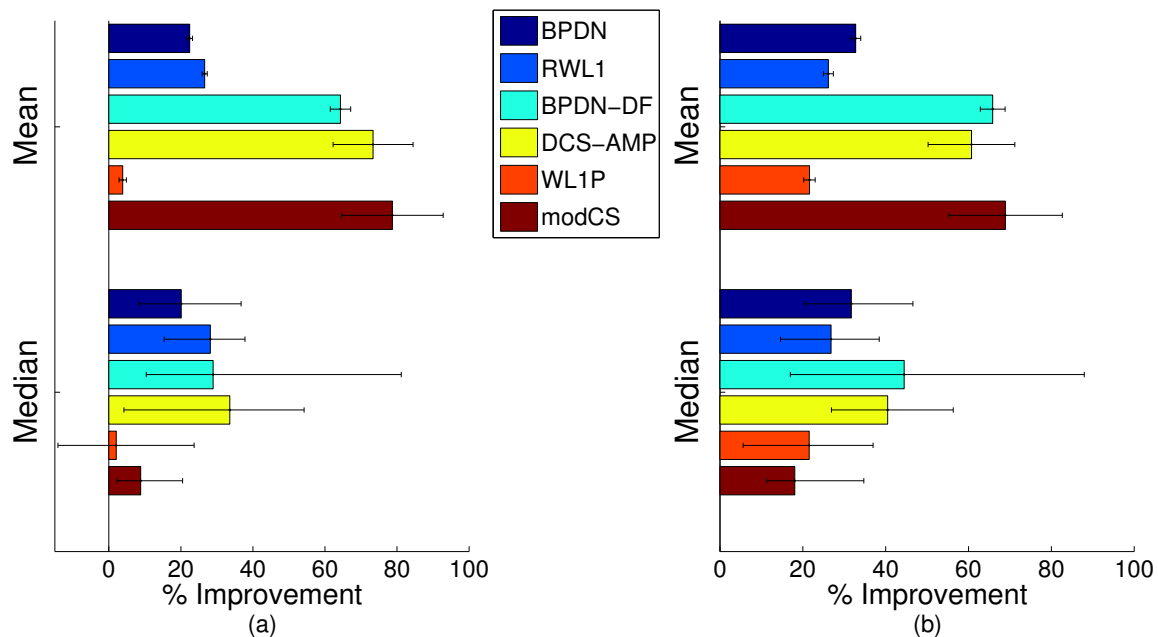


Fig. 8. Percent improvement of RWL1-DF over other algorithms for compressive recovery of video sequences (calculated as described in the text). Displayed is both the mean improvement (error bars indicating the normalized standard deviation) and the median improvement (error bars indicating the 25th and 75th percentile) for each algorithm. (a) Results for the full database of 24 sequences from a BBC nature documentary. (b) Results for the 13 sequences that were especially challenging for CS recovery, illustrating the benefits of RWL1-DF in this particularly difficult regime.

optimized on the Foreman sequence). Figure 8 shows the mean and median *improvement* of RWL1-DF relative to the other algorithms being evaluated. Specifically, the plotted mean improvement is the average of the rMSE difference between RWL1-DF and the comparison algorithm at each frame normalized by the average rMSE for the comparison algorithm across the whole video clip. The median improvement is calculated in the same manner.

The recovery results for this video database show consistent performance improvements for RWL1-DF when evaluated over all video sequences in this database. Additionally, we note that some video sequences were significantly richer in texture and motion than others, resulting in a more challenging recovery task. We identified 13 such video clips that were especially challenging (i.e, those where the average rMSE reconstruction for BPDN is over 1%). For these clips we plot the mean and median percent improvement in Figure 8. RWL1-DF shows very significant improvements within these video sequences, indicating that RWL1-DF is especially beneficial in challenging recovery scenarios. Additionally, an adaptation of a similar RWL1 method to a spatial filtering setting has shown similar performance improvements in the most challenging data examples [58].

## VI. CONCLUSIONS

In this work we explore two causal dynamic filtering schemes for time-varying sparse signals that are based on  $\ell_1$  minimization approaches. We examine both a simple method (BPDN-DF) that admits a theoretical analysis as well present a novel algorithm (RWL1-DF) that achieves state-of-the-art performance by leveraging second-order

statistics in a similar manner as Kalman filtering. We conclude from these results that  $\ell_1$  minimization methods have significant potential to be extended from static sparse signal estimation to time-varying data streaming settings. While not discussed in detail, the approaches in this paper can be solved efficiently through simple approaches such as an Iterative Shrinkage and Thresholding Algorithm (ISTA). Any future advances in numerical algorithms for  $\ell_1$  optimization will allow for faster and larger scale implementations of these approaches.

There are a number of potential modifications and improvements to the described approaches that will be pursued in future work. First, while classic dynamic filtering approaches assume known dynamics, it would be natural to estimate or learn complex dynamics as part of the proposed model structure. Second, this paper follows the convention of classic dynamic filtering work by assuming that the dynamics model operates in the state space (i.e., predicting the variances from the previous state estimates). In scenarios where the dynamics operations are not derived from physical principles and may be learned from data, RWL1-DF may be more effective when the dynamics model predicts the variances from the previous variance estimates.

## VII. APPENDIX

### A. Temporal bound derivation

This appendix proves the result in Theorem III.1. While the results do not depend on the numerical algorithm used to solve the optimization, the proof technique will leverage ISTA in the analysis [16], [47], [59]. This proof will follow by using the ISTA update equation at each time-step  $n$  and algorithmic iteration  $l$ , along with the RIP condition on  $\Phi$  and Lemma VII.1 (presented below), to obtain a recurrence relation for the absolute error between the current signal estimate. We then solve the recurrence relation for the analytical upper bound for each  $l, n$  independent of the previous ISTA iteration  $l + 1$  (but still dependent on the previous time-step  $n - 1$ ), allowing us to take  $l \rightarrow \infty$  and obtaining a recursive steady state bound in terms of only  $n$ . We then solve the new recursion equation in terms of  $n$ , yielding an upper bound for all  $n$  depending only on the initial condition and the properties of the dynamics and measurement functions.

We start by noting that ISTA for BPDN-DF can be written as

$$\begin{aligned} \mathbf{u}_n^{l+1} &= \hat{\mathbf{z}}_n^l + \frac{\eta}{1 + \kappa} \mathbf{W}^T \left( \mathbf{G}_n^T \left( \mathbf{y}_n - \mathbf{G}_n \mathbf{W} \hat{\mathbf{z}}_n^l \right) \right. \\ &\quad \left. + \kappa \left( f(\mathbf{W} \hat{\mathbf{z}}_{n-1}) - \mathbf{W} \hat{\mathbf{z}}_n^l \right) \right) \\ \hat{\mathbf{z}}_n^{l+1} &= T_\gamma(\mathbf{u}_n^{l+1}), \end{aligned} \tag{14}$$

where  $\mathbf{u}$  is an un-thresholded state,  $T_\gamma(\cdot)$  is a soft-thresholding function with threshold parameter  $\gamma$ , and  $\eta$  is the algorithm's step size. The proof continues by demonstrating that error between the estimate and the true signal contracts at each algorithmic step. Using the fact that the converged solution of ISTA minimizes the corresponding

cost function [47], [59], the error bound on the ISTA solution is then also a bound for the overall BPDN-DF minimizer. For ease of notation, we omit temporal subscripts, assuming that all variables, unless otherwise stated, have temporal subscript  $n$ . Additionally, we define the previous (steady state) estimate as  $\widehat{\mathbf{z}}_n = \widetilde{\mathbf{z}}$  and refer to  $\mathbf{z}_n^\dagger$  as the true coefficients at time  $n$ . We define two subsets of the sparse vector  $\mathbf{z}$ :  $J$  and  $J'$ . The set  $J = J[l+1]$  is the union of the current set of active coefficients in the ISTA algorithm  $\Gamma[l]$ , the  $q$  largest elements of the vector  $\mathbf{u}$   $\Delta[l]$ , and the true active set  $\Gamma_\dagger$ . In [16] it is shown that  $|J| = |\Delta[l+1] \cup \Gamma[l] \cup \Gamma_\dagger| \leq S + 2q$ . Similarly we define  $J' = J[l+2] = \Delta[l+1] \cup \Gamma[l+1] \cup \Gamma_\dagger$ . To start, we require the following lemma which bounds the norm of  $\mathbf{u}$  at each algorithmic step and at each  $n$ :

**Lemma VII.1.** *Suppose that the same conditions as in Theorem III.1 hold. Additionally, assume that  $\|\mathbf{z}_n^\dagger\|_2 \leq b$  for all  $n$ . The vector  $\mathbf{u}_J^l$  (the ISTA variables  $\mathbf{u}^l$  restricted to the support subset  $J$ ) at each algorithmic iteration  $l$  obtained via ISTA (iterating Equation (14)) with a step size of  $\mu$  satisfies*

$$\begin{aligned} \|\mathbf{u}_J^l\|_2 \leq & \left( |\eta - 1| + \frac{\eta\delta}{1 + \kappa} \right) \frac{\gamma}{1 + \kappa} \sqrt{q} + \left( \eta + \eta \frac{\kappa f^* + \delta}{\kappa + 1} \right) b \\ & + \eta \frac{\sqrt{1 + \delta}}{1 + \kappa} \bar{\epsilon} + \frac{\eta\kappa}{1 + \kappa} \bar{\nu}, \end{aligned}$$

And with the restriction

$$\begin{aligned} & \tilde{\eta}(\kappa + \kappa f^* + 1 + \delta)b + \tilde{\eta}\sqrt{1 + \delta}\bar{\epsilon} + \tilde{\eta}\kappa\bar{\nu} \\ & \leq \left( 1 - \frac{|\tilde{\eta}(1 + \kappa) - 1| + \tilde{\eta}\delta}{1 + \kappa} \right) \gamma\sqrt{q}, \end{aligned}$$

then  $\|\mathbf{u}_J^l\|_2$  is simply bounded by  $\|\mathbf{u}_J^l\|_2 \leq \gamma\sqrt{q}$

We start by writing the norm using the definition of  $\mathbf{u}$  in the ISTA algorithm:

$$\begin{aligned} \|\mathbf{u}_J^l\|_2 = & \left\| \mathbf{z}_J^l + \frac{\eta}{1 + \kappa} (\mathbf{G}\mathbf{W})_J^T (\mathbf{y} - \mathbf{G}\mathbf{W}\mathbf{z}^l) \right. \\ & \left. + \frac{\eta\kappa}{1 + \kappa} \mathbf{W}_J^T (f(\mathbf{W}\widetilde{\mathbf{z}}) - \mathbf{W}\mathbf{z}^l) \right\|_2. \end{aligned}$$

Using the fact that  $\mathbf{y} = \mathbf{G}\mathbf{W}\mathbf{z}^\dagger + \boldsymbol{\epsilon}$ ,

$$\begin{aligned} \|\mathbf{u}_J^l\|_2 = & \left\| \mathbf{z}_J^l + \frac{\eta}{1 + \kappa} (\mathbf{G}\mathbf{W})_J^T \mathbf{G}\mathbf{W} (\mathbf{z}^\dagger - \mathbf{z}^l) \right. \\ & \left. + \frac{\eta}{1 + \kappa} \mathbf{W}^T \mathbf{G}_J^T \boldsymbol{\epsilon} + \frac{\eta\kappa}{1 + \kappa} \mathbf{W}_J^T (f(\mathbf{W}\widetilde{\mathbf{z}}) - \mathbf{W}\mathbf{z}^l) \right\|_2. \end{aligned} \quad (15)$$

To properly reduce the portion of this expression depending on the dynamics  $f(\cdot)$ , we note that since the dynamics

satisfies  $\mathbf{W}z_n^\dagger = f(\mathbf{W}z_{n-1}^\dagger) + \boldsymbol{\nu}_n$ ,

$$\begin{aligned} f(\mathbf{W}\tilde{\mathbf{z}}) - \mathbf{W}z^l &= f(\mathbf{W}\tilde{\mathbf{z}}) - \mathbf{W}z^\dagger + \mathbf{W}z^\dagger - \mathbf{W}z^l \\ &= f(\mathbf{W}\tilde{\mathbf{z}}) - f(\mathbf{W}z_{n-1}^\dagger) - \boldsymbol{\nu} \\ &\quad + \mathbf{W}(z^\dagger - z^l). \end{aligned}$$

Setting  $\tilde{\eta} = \eta/(1 + \kappa)$  and collecting similar terms,

$$\begin{aligned} \left\| \mathbf{u}_J^l \right\|_2 &= \left\| z_J^l + \tilde{\eta}((\mathbf{G}\mathbf{W})_J^T \mathbf{G}\mathbf{W} + \kappa \mathbf{I}_J) (z^\dagger - z^l) \right. \\ &\quad \left. + \tilde{\eta} \mathbf{W}^T (\mathbf{G}_J^T \boldsymbol{\epsilon} - \kappa \boldsymbol{\nu}) \right. \\ &\quad \left. + \tilde{\eta} \kappa \mathbf{W}_J^T (f(\mathbf{W}\tilde{\mathbf{z}}) - f(\mathbf{W}z_{n-1}^\dagger)) \right\|_2 \\ &\leq \left\| z_J^l + \tilde{\eta}((\mathbf{G}\mathbf{W})_J^T \mathbf{G}\mathbf{W} + \kappa \mathbf{I}_J) (z^\dagger - z^l) \right\|_2 \\ &\quad + \tilde{\eta} \sqrt{1 + \delta} \|\boldsymbol{\epsilon}\|_2 + \tilde{\eta} \kappa \|\boldsymbol{\nu}\|_2 \\ &\quad + \tilde{\eta} \kappa \left\| \mathbf{W}_J^T (f(\mathbf{W}\tilde{\mathbf{z}}) - f(\mathbf{W}z_{n-1}^\dagger)) \right\|_2 \\ &\leq \left\| z_J^l + \tilde{\eta}((\mathbf{G}\mathbf{W})_J^T \mathbf{G}\mathbf{W} + \kappa \mathbf{I}_J) (z^\dagger - z^l) \right\|_2 \\ &\quad + \tilde{\eta} \sqrt{1 + \delta} \|\boldsymbol{\epsilon}\|_2 + \tilde{\eta} \kappa \|\boldsymbol{\nu}\|_2 + \tilde{\eta} \kappa f^* \|\mathbf{e}_{n-1}\|_2 \\ &\leq \left\| (\tilde{\eta}(\mathbf{G}\mathbf{W})_J^T \mathbf{G}\mathbf{W} - (1 - \kappa \tilde{\eta}) \mathbf{I}_J) z_J^l \right\|_2 \\ &\quad + \left\| \tilde{\eta}(\mathbf{G}_J^T \mathbf{G} + \kappa \mathbf{I}_J) \mathbf{W} z^\dagger \right\|_2 + \tilde{\eta} \sqrt{1 + \delta} \|\boldsymbol{\epsilon}\|_2 \\ &\quad + \tilde{\eta} \kappa \|\boldsymbol{\nu}\|_2 + \tilde{\eta} \kappa f^* \|\mathbf{e}_{n-1}\|_2. \end{aligned} \tag{16}$$

where the first and third inequalities follow from the triangle inequality, the fact that  $\|\mathbf{W}\| \leq 1$  and the RIP of  $\mathbf{G}$ , and the second inequality follows from the smoothness of  $f(\cdot)$ .

We now use the Cauchy-Schwartz inequality with the fact that, since  $\mathbf{G}$  satisfies  $\text{RIP}(|J|, \delta)$  with respect to  $\mathbf{W}$ ,

$$\left\| \alpha(\mathbf{W}\mathbf{G})_J^T (\mathbf{G}\mathbf{W})_J + \beta \mathbf{I}_J \right\|_2 \leq |\alpha + \beta| + \alpha \delta, \tag{17}$$

for any constants  $\alpha, \beta$ , to bound the first two terms of Equation 16 as

$$\begin{aligned} \left\| \mathbf{u}_J^l \right\|_2 &\leq (|\tilde{\eta} + \kappa \tilde{\eta} - 1| + \tilde{\eta} \delta) \left\| z^l \right\|_2 + \tilde{\eta}(\kappa + 1 + \delta) \left\| z^\dagger \right\|_2 \\ &\quad + \tilde{\eta} \sqrt{1 + \delta} \|\boldsymbol{\epsilon}\|_2 + \tilde{\eta} \kappa \|\boldsymbol{\nu}\|_2 + \tilde{\eta} \kappa f^* \|\mathbf{e}_{n-1}\|_2, \end{aligned}$$

Simplifying, we obtain

$$\left\| \mathbf{u}_J^l \right\|_2 \leq (|\tilde{\eta}(1 + \kappa) - 1| + \tilde{\eta} \delta) \tilde{\gamma} \sqrt{q} + \tilde{\eta}(\kappa + 1 + \delta) b$$

$$+\tilde{\eta}\sqrt{1+\delta\bar{\epsilon}}+\tilde{\eta}\kappa\bar{\nu}+\tilde{\eta}\kappa f^* \|e_{n-1}\|_2,$$

where  $b$  is the maximum energy of  $z^\dagger$  (i.e.  $\|z^\dagger\|_2 \leq b$ ) and  $\tilde{\gamma} = \gamma/(1+\kappa)$ .

Ideally, this Lemma should be independent of the previous estimation error norm  $\|e_{n-1}\|_2$  for the Lemma to hold for all  $n$ . If we initialize the estimate with the zero vector, clearly  $\|e_0\|_2 = \|z_0^\dagger\|_2 \leq b$ . Thus, setting  $\|e_{n-1}\|_2 < b$  results in the Lemma statement and it remains to ensure that  $\|e_n\|_2 \leq b$  for all  $n$  through later parameter restrictions.

### B. First Recursion

With Lemma VII.1, we seek a recursive expression for the estimation error at algorithmic iteration  $l+1$ ,

$$\begin{aligned} \|e^{l+1}\|_2 &= \|z^{l+1} - z^\dagger\|_2 \\ &= \|z^{l+1} - u_{J'}^{l+1} + u_{J'}^{l+1} - z^\dagger\|_2 \\ &\leq \|z^{l+1} - u_{J'}^{l+1}\|_2 + \|u_{J'}^{l+1} - z^\dagger\|_2 \\ &\leq \|u_{J'}^{l+1}\|_2 + \|z_{J'} - z^\dagger \\ &\quad + \tilde{\eta}W^T \left( G_J^T G W (z^\dagger - z_{J'}^l) \right. \\ &\quad \left. + G_{J'}\epsilon + \kappa(f(W\tilde{z}) - Wz^l) \right)\|_2 \\ &= \|u_{J'}^{l+1}\|_2 + \|(\tilde{\eta}W^T G_J^T G W - I_{J'}) (z^\dagger - z_{J'}^l) \\ &\quad + \tilde{\eta}W^T G_{J'}\epsilon + \tilde{\eta}\kappa W^T (f(W\tilde{z}) - Wz^l)\|_2 \\ &\leq \gamma\sqrt{q} + \tilde{\eta}\kappa f^* \|e_{n-1}\|_2 + \tilde{\eta}\sqrt{1+\delta\bar{\epsilon}} + \tilde{\eta}\kappa\bar{\nu} \\ &\quad + (|\tilde{\eta} + \tilde{\eta}\kappa - 1| + \tilde{\eta}\delta)\|e^l\|_2, \end{aligned}$$

where the first inequality follows from the triangle inequality, the second inequality follows from the nature of the thresholding function and the definition of  $u$ , the third inequality follows from Lemma VII.1 for the first term and a similar set of steps as used to prove Lemma VII.1 used on the second term. This recursive formula can now be solved for  $\|e^l\|_2$ ,

$$\begin{aligned} \|e[l]\|_2 &\leq (|\eta - 1| + \tilde{\eta}\delta)^l \left( \|e^0\|_2 \right. \\ &\quad \left. - \frac{\gamma\sqrt{q} + \tilde{\eta}\kappa f^* \|e_{n-1}\|_2 + \tilde{\eta}\sqrt{1+\delta\bar{\epsilon}} + \tilde{\eta}\kappa\bar{\nu}}{1 - |\eta - 1| - \tilde{\eta}\delta} \right) \\ &\quad + \frac{\gamma\sqrt{q} + \tilde{\eta}\kappa f^* \|e_{n-1}\|_2 + \tilde{\eta}\sqrt{1+\delta\bar{\epsilon}} + \tilde{\eta}\kappa\bar{\nu}}{1 - |\eta - 1| - \tilde{\eta}\delta}. \end{aligned}$$

As  $l \rightarrow \infty$ , ISTA converges to a solution with a bounded error of

$$\|e_n\|_2 = \|e_n^\infty\|_2$$

$$\leq \frac{\gamma\sqrt{q} + \tilde{\eta}\kappa f^* \|e_{n-1}\|_2 + \tilde{\eta}\sqrt{1 + \delta\bar{\epsilon}} + \tilde{\eta}\kappa\bar{\nu}}{1 - |\eta - 1| - \tilde{\eta}\delta}.$$

### C. Second Recursion

The steady state error of ISTA can now be treated as a new recursive equation on  $\|e_n\|_2$  in terms of the signal time-step  $n$ . This expression can be solved for  $\|e_n\|_2$

$$\begin{aligned} \|e_n\|_2 \leq & \left( \frac{\tilde{\eta}\kappa f^*}{1 - |\eta - 1| - \tilde{\eta}\delta} \right)^n (\|e_0\|_2 \\ & - \frac{\gamma\sqrt{q} + \tilde{\eta}\sqrt{1 + \delta\bar{\epsilon}} + \tilde{\eta}\kappa\bar{\nu}}{1 - |\eta - 1| - \tilde{\eta}\delta - \tilde{\eta}\kappa f^*}) \\ & + \frac{\gamma\sqrt{q} + \tilde{\eta}\sqrt{1 + \delta\bar{\epsilon}} + \tilde{\eta}\kappa\bar{\nu}}{1 - |\eta - 1| - \tilde{\eta}\delta - \tilde{\eta}\kappa f^*}, \end{aligned}$$

which yields a bound for the error at every iteration  $n$ .

The only remaining task is to ensure  $\|e_n\|_2 \leq b$  for all  $n$  so that Lemma VII.1 holds. If  $b$  bounds  $\|e_n\|_2$  at each  $n$ , then

$$\beta^n (\|e_0\|_2 - \frac{\alpha}{1 - \beta}) + \frac{\alpha}{1 - \beta} \leq \beta^n (b - \frac{\alpha}{1 - \beta}) + \frac{\alpha}{1 - \beta} \leq b$$

holds. Simplifying,

$$\frac{\alpha(1 - \beta^n)}{1 - \beta} \leq (1 - \beta^n)b \rightarrow \frac{\alpha}{1 - \beta} \leq b,$$

or in terms of the model and system parameters,

$$\frac{(1 + \kappa)\gamma\sqrt{q} + \eta\sqrt{1 + \delta\bar{\epsilon}} + \eta\kappa\bar{\nu}}{(1 + \kappa)(1 - |\eta - 1|) - \eta\delta - \eta\kappa f^*} \leq b.$$

Setting  $\eta = (1 + \delta)^{-1} \leq 2(1 + \delta)^{-1}$  completes the proof.

## 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] E. Candès, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *IEEE Trans on Information Theory*, vol. 52, no. 2, Feb 2006.
- [3] R. Baraniuk, "Compressive sensing," *IEEE Signal Processing Magazine*, vol. 24, no. 4, pp. 118–121, Jul 2007.
- [4] M. Davies and L. Daudet, "Sparse audio representations using the mclt," *Signal processing*, vol. 86, no. 3, pp. 457–470, 2006.
- [5] 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.
- [6] 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.

- [7] D. Donoho and J. Tanner, "Sparse nonnegative solution of underdetermined linear equations by linear programming," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 102, no. 27, p. 9446, 2005.
- [8] E. Candes and J. Romberg, " $\ell^1$ -Magic: Recovery of sparse signals via convex programming," 2005, <http://www.acm.caltech.edu/l1magic/>.
- [9] S.-J. Kim, K. Koh, M. Lustig, S. Boyd, and D. Gorinevsky, "An interior-point method for large scale  $\ell_1$ -regularized least squares," *IEEE Journal on Selected Topics in Signal Processing*, vol. 1, no. 4, pp. 606–617, Dec 2007.
- [10] D. M. Malioutov, M. Cetin, and A. S. Willsky, "Homotopy continuation for sparse signal representation," *IEEE Proceedings of ICASSP*, 2005.
- [11] M. A. T. Figueiredo, R. D. Nowak, and S. J. Wright, "Gradient projection for sparse reconstruction: Application to compressed sensing and other inverse problems," *IEEE Journal of Selected Topics in Signal Processing*, vol. 1, no. 4, pp. 586–597, 2007.
- [12] C. J. Rozell, D. H. Johnson, R. G. Baraniuk, and B. A. Olshausen, "Sparse coding via thresholding and local competition in neural circuits," *Neural Computation*, vol. 20, no. 10, pp. 2526–2563, Oct 2010.
- [13] 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*, vol. 2, no. 3, pp. 530–541, 2012.
- [14] E. Candès and Y. Plan, "Near-ideal model selection by  $\ell_1$  minimization," California Institute of Technology, Tech. Rep., Dec 2007.
- [15] T. Zhang, "Some sharp performance bounds for least squares regression with  $\ell_1$  regularization," *The Annals of Statistics*, vol. 37, no. 5A, pp. 2109–2144, 2009.
- [16] A. Balavoine, C. J. Rozell, and J. Romberg, "Discrete and continuous-time soft-thresholding with dynamic inputs," vol. 63, no. 12, pp. 3165–3176, 2015.
- [17] E. Candès, M. Wakin, and S. Boyd, "Enhancing sparsity by reweighted  $\ell_1$  minimization," *Journal of Fourier Analysis and Applications*, vol. 14, no. 5, pp. 877–905, Dec 2008, special Issue on Sparsity.
- [18] P. Garrigues and B. Olshausen, "Group sparse coding with a laplacian scale mixture prior," *Advances in Neural Information Processing Systems*, vol. 24, 2010.
- [19] D. Wipf and S. Nagarajan, "Iterative reweighted  $\ell_1$  and  $\ell_2$  methods for finding sparse solutions," *Selected Topics in Signal Processing, IEEE Journal of*, vol. 4, no. 2, pp. 317–329, 2010.
- [20] R. Chartrand and W. Yin, "Iteratively reweighted algorithms for compressive sensing," in *Acoustics, Speech and Signal Processing, 2008. ICASSP 2008. IEEE International Conference on*. IEEE, 2008, pp. 3869–3872.
- [21] D. Needell, "Noisy signal recovery via iterative reweighted  $\ell_1$ -minimization," in *Forty-Third Asilomar Conference on Signals, Systems and Computers*. IEEE, 2009, pp. 113–117.
- [22] M. Khajehnejad, W. Xu, S. Avestimehr, and B. Hassibi, "Improved sparse recovery thresholds with two-step reweighted  $\ell_1$  minimization," pp. 1603–1607, 2010.
- [23] W. Xu, M. Khajehnejad, S. Avestimehr, and B. Hassibi, "Breaking through the thresholds: an analysis for iterative reweighted  $\ell_1$  Minimization via the Grassman Angle Framework," *Proceedings of the International Conference on Acoustics, Speech, and Signal Processing*, 2010.
- [24] S. Haykin, "Kalman filters," in *Kalman Filtering and Neural Networks*, S. Haykin, Ed. John Wiley & Sons, Inc., 2001, pp. 1–22.
- [25] 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.
- [26] D. Angelosante, S. I. Roumeliotis, and G. B. Giannakis, "Lasso-Kalman smoother for tracking sparse signals," *Proc of the Asilomar Conference on Signals, Systems and Computers*, 2009.
- [27] 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.



- [28] 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.
- [29] A. Sankaranarayanan, P. Turaga, R. Baraniuk, and R. Chellappa, "Compressive acquisition of dynamic scenes," *Computer Vision–ECCV 2010*, pp. 129–142, 2010.
- [30] M. S. Asif, L. Hamilton, M. Brummer, and J. Romberg, "Motion-adaptive spatio-temporal regularization (master) for accelerated dynamic mri," *Magnetic Resonance in Medicine*, vol. 70, no. 3, pp. 800–812, 2013.
- [31] A. Y. Aravkin, J. V. Burke, and G. Pillonetto, "Optimization viewpoint on kalman smoothing with applications to robust and sparse estimation," in *Compressed Sensing & Sparse Filtering*, 2014, pp. 237–280.
- [32] —, "Sparse/robust estimation and kalman smoothing with nonsmooth log-concave densities: Modeling, computation, and theory," *The Journal of Machine Learning Research*, vol. 14, no. 1, pp. 2689–2728, 2013.
- [33] M. S. Asif and J. Romberg, "Dynamic updating for  $\ell_1$  minimization," *IEEE Journal of Selected Topics in Signal Processing*, vol. 4, no. 2, pp. 421–434, Apr 2010.
- [34] M. S. Asif, A. Charles, J. Romberg, and C. Rozell, "Estimation and dynamic updating of time-varying signals with sparse variations," *ICASSP*, 2011.
- [35] A. Carmi, P. Gurfil, and D. Kanevsky, "Methods for sparse signal recovery using Kalman filtering pseudo-measuremnt norms and quasi-norms," *IEEE Transactions on Signal Processing*, vol. 58, no. 4, pp. 2405–2409, Apr 2010.
- [36] N. Vaswani, "Kalman filtered compressed sensing," *Proc of ICIP 2008*, pp. 893–896, 2008.
- [37] C. Qiu and N. Vaswani, "Recursive sparse recovery in large but low dimensional noise," *48th Allerton Conference on Communication Control and Computing*, 2011.
- [38] D. Zachariah, S. Chatterjee, and M. Jansson, "Dynamic iterative pursuit," *IEEE Transactions on Signal Processing*, vol. 60, no. 9, pp. 4967–4972, 2012.
- [39] J. Ziniel, L. C. Potter, and P. Schniter, "Tracking and smoothing of time-varying sparse signals via approximate belief propagation," *Proceesings of the Asilomar Conference on Signals, Systems and Computers*, 2010.
- [40] 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.
- [41] 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.
- [42] 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.
- [43] A. Charles, M. S. Asif, J. Romberg, and C. Rozell, "Sparsity penalties in dynamical system estimation," *Proc of the CISS*, March 2011.
- [44] E. Hall and R. Willett, "Dynamical models and tracking regret in online convex programming," in *Proceedings of the International Conference on Machine Learning*, 2013, pp. 579–587.
- [45] R. Baraniuk, V. Cevher, and M. Wakin, "Low-dimensional models for dimensionality reduction and signal recovery: A geometric perspective," *Proceedings of the IEEE*, vol. 98, no. 6, pp. 959–971, 2010.
- [46] A. S. Charles and C. J. Rozell, "Convergence of basis pursuit de-noising with dynamic filtering," *Proceedings the IEEE GlobalSIP*, Dec 2014.
- [47] K. Bredies and D. A. Lorenz, "Linear convergence of iterative soft-thresholding," *Journal of Fourier Analysis and Applications*, vol. 14, no. 5, pp. 813–837, 2008.
- [48] M. A. Khajehnejad, W. Xu, S. Avestimehr, and B. Hassibi, "Weighted  $\ell_1$  minimization for sparse recovery with prior information," pp. 483–487, 2009.

- [49] C. Miosso, R. von Borries, M. Argaez, L. Velazquez, C. Quintero, and C. M. Potes, “Compressive sensing reconstruction with prior information by iteratively reweighted least-squares,” *IEEE Transactions on Signal Processing*, vol. 57, no. 6, pp. 2424–2431, 2009.
- [50] R. E. Carrillo, L. F. Polania, and K. E. Barner, “Iterative algorithms for compressed sensing with partially known support,” in *IEEE International Conference on Acoustics Speech and Signal Processing (ICASSP)*, 2010, pp. 3654–3657.
- [51] S. Chrétien and A. O. H. III, “Kullback proximal algorithms for maximum-likelihood estimation,” *IEEE Transactions on Information Theory*, vol. 46, no. 5, pp. 1800–1810, 2000.
- [52] C. F. Cadieu and B. A. Olshausen, “Learning intermediate-level representations of form and motion from natural movies,” *Neural computation*, vol. 24, no. 4, pp. 827–866, 2012.
- [53] J. Ziniel and P. Schniter, “Dynamic compressive sensing of time-varying signals via approximate message passing,” *IEEE Transaction on Signal Processing*, vol. 61, no. 21, pp. 5270–5284, 2013.
- [54] I. Selesnick, R. Baraniuk, and N. Kingsbury, “The dual-tree complex wavelet transform,” *IEEE Signal Processing Magazine*, vol. 22, no. 6, pp. 123–151, 2005.
- [55] R. Coifman, F. Geshwind, and Y. Meyer, “Noiselets,” *Applied Computational Harmonic Analysis*, vol. 10, no. 1, pp. 27–44, 2001.
- [56] J. Romberg, “Imaging via compressive sampling,” *IEEE Signal Processing Magazine*, vol. 25, no. 2, pp. 14–20, 2008.
- [57] S. Becker, E. Candès, and M. Grant, “Templates for convex cone problems with applications to sparse signal recovery,” *Mathematical Programming Computation*, pp. 1–54, 2011.
- [58] A. Charles and C. Rozell, “Spectral superresolution of hyperspectral imagery using reweighted  $\ell_1$  spatial filtering,” *IEEE Geoscience and Remote Sensing Letters*, vol. 11, no. 3, pp. 602–606, March 2014.
- [59] I. Daubechies, M. Defrise, and C. De Mol, “An iterative thresholding algorithm for linear inverse problems with a sparsity constraint,” *Communications on pure and applied mathematics*, vol. 57, no. 11, pp. 1413–1457, 2004.