DYNAMIC SPARSE STATE ESTIMATION USING ℓ_1-ℓ_1 MINIMIZATION: ADAPTIVE-RATE MEASUREMENT BOUNDS, ALGORITHMS AND APPLICATIONS

João Mota[⋆], *Nikos Deligiannis*[⋆], *Aswin C. Sankaranarayanan*[†], *Volkan Cevher*[♯], *Miguel Rodrigues*[⋆]

[⋆] Electronic and Electrical Engineering Dept., University College London, UK
[†] Electrical and Computer Engineering Dept., Carnegie Mellon University, USA
[♯] Laboratory for Information and Inference Systems (LIONS), EPFL, Switzerland

ABSTRACT

We propose a recursive algorithm for estimating time-varying signals from a few linear measurements. The signals are assumed sparse, with unknown support, and are described by a dynamical model. In each iteration, the algorithm solves an ℓ_1-ℓ_1 minimization problem and estimates the number of measurements that it has to take at the next iteration. These estimates are computed based on recent theoretical results for ℓ_1-ℓ_1 minimization. We also provide sufficient conditions for perfect signal reconstruction at each time instant as a function of an algorithm parameter. The algorithm exhibits high performance in compressive tracking on a real video sequence, as shown in our experimental results.

Index Terms— State estimation, sparsity, background subtraction, motion estimation, online algorithms

1. INTRODUCTION

We study the reconstruction of sparse, time-varying signals from a limited number of linear measurements. Let $x[k] \in \mathbb{R}^n$ denote the target signal at time k and let $y[k] \in \mathbb{R}^{m_k}$ denote m_k measurements of $x[k]$. We consider the dynamical model

$$x[k] = f_k(x[k-1]) + \epsilon[k] \tag{1a}$$

$$y[k] = A_k\, x[k] + \eta[k]\,, \tag{1b}$$

where $f_k : \mathbb{R}^n \to \mathbb{R}^n$ describes the evolution of the signals $x[k]$, $k = 1, 2, \ldots$, between consecutive time instants, and $A_k \in \mathbb{R}^{m_k \times n}$ is the matrix of measurements at time k. The quantities $\epsilon[k]$ and $\eta[k]$ capture model inaccuracies and measurement noise, respectively.

One of the oldest problems in control theory is to estimate the state sequence $\{x[k]\}_{k\geq 1}$ from the measurements $\{y[k]\}_{k\geq 1}$. The classical solution is the Kalman filter [1], a recursive algorithm, known to be least-squares optimal when the model is linear ($f_k(x) = F_k\, x$) and $\epsilon[k]$ and $\eta[k]$ are zero-mean Gaussian. Several extensions have been proposed for the case where these two assumptions are not met, e.g., [2–4]. The Kalman filter and these extensions, however, cannot easily integrate additional knowledge of the signal's structure, e.g., sparsity, and suffer from lack of observability when the number of measurements is limited, i.e., $m_k \ll n$.

Contributions. Our goal is to reconstruct each signal $x[k] \in \mathbb{R}^n$ from a small number of measurements $m_k \ll n$, when $x[k]$ is sparse and has unknown support. Furthermore, we assume sparse model inaccuracies $\epsilon[k]$ and bounded measurement noise $\eta[k]$, i.e., there exists $\sigma_k \geq 0$ such that $\|\eta[k]\|_2 \leq \sigma_k$. Under these conditions, the Kalman filter has poor performance [5] (especially for non-Gaussian $\epsilon[k]$'s). Assuming that the entries of each A_k are i.i.d. Gaussian[1], we propose estimating $x[k]$ recursively as follows:

$$\begin{array}{lll} \hat{x}[k] \in & \underset{x}{\text{Argmin}} & \|x\|_1 + \big\|x - f_k(\hat{x}[k-1])\big\|_1 \\ & \text{s.t.} & \big\|A_k x - y[k]\big\|_2 \leq \sigma_k\,, \end{array} \tag{2}$$

where $\hat{x}[k-1]$ is the signal estimate at time $k-1$. In (2), $\|x\|_1 := \sum_i |x_i|$ is the ℓ_1-norm and $\|x\|_2 := \sqrt{\sum_i x_i^2}$ is the ℓ_2-norm. Note that, in general, (2) may have more than one solution. Based on the results in [8, 9], we propose a recursive mechanism to compute the number of measurements m_k at each time k. This scheme minimizes m_k while guaranteeing perfect reconstruction in the noiseless scenario, $\eta[k] = 0$, or stable reconstruction (i.e., $\|\hat{x}[k] - x[k]\|_2 \leq 2\sigma_k/\epsilon$, for some $0 < \epsilon < 1$) in the noisy scenario. Furthermore, note that there are no parameters (weights) to tune in (2).

Applications. The model in (1) is actually applicable to non-sparse signals, provided they have sparse representations in a suitable domain. Let $z[k] \in \mathbb{R}^n$ be a non-sparse signal that has a sparse representation $x[k] = \Psi z[k]$, where $\Psi : \mathbb{R}^{n\times n}$ is the sparsifying transform. Suppose $z[k]$ evolves as $z[k] = \hat{f}_k(z[k-1]) + \epsilon[k]$, and we observe $y[k] = \hat{A}_k z[k] + \eta[k]$. Then, the sparse coefficients $x[k]$ evolve as in (1) with $f_k(x) = \Psi \hat{f}_k(\Psi^{-1}x)$ and $A_k = \hat{A}_k\Psi$. Thus, the class of signals described by our model is quite broad, and the applications are diverse. They include, for example, dynamic MRI [10, 11], radar [12], and background subtraction [13].

2. RELATED WORK

Prior work that incorporates signal structure in state estimation problems includes [14–16]. This work splits the problem of estimating a dynamic sparse signal into the problems of estimating its support, which is addressed with compressed sensing techniques, and estimating its values, which is addressed with a Kalman filter. This method, however, assumes that the support of the signal varies slowly in time. Other approaches assume the coefficients of the support also vary slowly [17, 18], or the signal varies smoothly [18, 19], including with an evolution governed by a linear dynamical system [20]. Instead of assuming smoothness or slow-varying supports, our scheme assumes that the quality of the prediction given by f_k does not vary much between consecutive instants.

J. Mota, N. Deligiannis, and M. Rodrigues were supported by the EPSRC grant EP/K033166/1. A. C. Sankaranarayanan was supported in part by the NSF grant CCF-1117939. V. Cevher's work is supported in part by the European Commission under grants MIRG-268398 and ERC Future Proof and by the Swiss Science Foundation under grants SNF 200021-132548, SNF 200021-146750, and SNF CRSII2-147633.

[1]This is common in systems designed to sample signals according to the compressed sensing paradigm [6, 7].

The work in [5] studies three reconstruction schemes, the best of which is a Lagrangian version of (2), i.e., there are no constraints and the objective has the additional term $\beta_2\big\|A_k x - y[k]\big\|_2$. It was experimentally shown in [5] that the Lagrangian version of (2) has an excellent performance and outperforms Kalman filtering, even when the model inaccuracies $\epsilon[k]$ are Gaussian (and thus not sparse). Note that the solutions of (2) and its Lagrangian version coincide when β_2 is chosen properly. However, the advantages of solving (2) w.r.t. its Lagrangian version are twofold. First, in practice, it is easier to obtain bounds on the magnitude of $\eta[k]$ than it is to tune the parameter β_2. Second, the recent results in [8,9] establish reconstruction guarantees for (2) in the case of static signals; those results also establish an optimal value for the parameter β (equal to 1), making (2) parameter-free.

Finally, while prior work studies reconstruction schemes where the number of measurements is the same in all time instants [5,14,15, 17–21] (a notable exception is [22], where cross-validation is used to estimate the required number of measurements), our reconstruction scheme adapts the number of measurements recursively.

3. BACKGROUND: STATIC SIGNAL RECONSTRUCTION

Our scheme is motivated by the recent results in [8]. This reference studies problem (2) in a static scenario, i.e., when only one iteration of (2) is performed. We summarize those results next.

Let $x^\star \in \mathbb{R}^n$ be a sparse signal of which we have m linear noisy measurements $y = Ax^\star + \eta$, where $A \in \mathbb{R}^{m\times n}$ and $\|\eta\|_2 \leq \sigma$, for a known $\sigma \geq 0$. We assume access to a signal $w \in \mathbb{R}^n$ similar to $x^\star$, in the sense that $\|x^\star - w\|_1$ is expected to be small. Suppose we attempt to reconstruct $x^\star$ by solving the ℓ_1-ℓ_1 minimization problem:

$$\begin{array}{ll} \underset{x}{\text{minimize}} & \|x\|_1 + \beta\|x - w\|_1 \\ \text{subject to} & \|Ax - y\|_2 \leq \sigma\,, \end{array} \tag{3}$$

where $\beta > 0$. The following result from [8] establishes reconstruction guarantees for (3). To state it, we need the following quantities:

$$\overline{h} := \big|\{i\,:\, x_i^\star > 0,\ x_i^\star > w_i\} \cup \{i\,:\, x_i^\star < 0,\ x_i^\star < w_i\}\big| \tag{4a}$$

$$\xi := \big|\{i\,:\, w_i \neq x_i^\star = 0\}\big| - \big|\{i\,:\, w_i = x_i^\star \neq 0\}\big|\,, \tag{4b}$$

where $|\cdot|$ denotes the cardinality of a set. Note that $0 \leq \overline{h} \leq s$, where s is the sparsity of $x^\star$.

Theorem 1 ([8]). *Let $x^\star, w \in \mathbb{R}^n$ and suppose we take m linear measurements $y = Ax^\star + \eta$, where $\|\eta\|_2 \leq \sigma$, for $\sigma \geq 0$. Assume $\overline{h} > 0$ and that there exists at least one index i for which $x_i^\star = w_i = 0$. Let the entries of $A \in \mathbb{R}^{m\times n}$ be i.i.d. Gaussian with zero mean and variance $1/m$.*

1. *If $\sigma = 0$ or, equivalently, $y = Ax^\star$, and*

$$m \geq 2\overline{h}\log\Big(\frac{n}{s+\xi/2}\Big) + \frac{7}{5}\Big(s + \frac{\xi}{2}\Big) + 1\,, \tag{5}$$

then, with probability at least $1 - \exp\big(-\frac{1}{2}(m - \sqrt{m})^2\big)$, $x^\star$ is the unique solution of (3) with $\beta = 1$.

2. *If $\sigma > 0$, define $0 < \epsilon < 1$ and let*

$$m \geq \frac{1}{(1-\epsilon)^2}\Big[2\overline{h}\log\Big(\frac{n}{s+\xi/2}\Big) + \frac{7}{5}\Big(s + \frac{\xi}{2}\Big) + \frac{3}{2}\Big] + 1\,. \tag{6}$$

Then, any solution $\hat{x}$ of (3) with $\beta = 1$ satisfies $\|\hat{x} - x^\star\|_2 \leq 2\sigma/\epsilon$ with probability at least $1 - \exp\big(-\frac{1}{2}(m - (1-\epsilon)\sqrt{m})^2\big)$.

Theorem 1 establishes lower bounds on the number of measurements that guarantees that (3) with $\beta = 1$ recovers $x^\star$ perfectly (resp. stably) in a noiseless (resp. noisy) measurement scenario, with high probability. The bounds in (5) and (6) are a function of the signal dimension n, the signal sparsity s, and the quantities $\overline{h}$ and ξ. Note that $\overline{h}$ and ξ depend only on the signs of each entry of the vectors $x^\star$ and $w - x^\star$, but not on their values. As these quantities are not known in practice (they depend on the unknown signal $x^\star$), we propose in section 4 an adaptive scheme to estimate them using previous signals. Note that $\overline{h}$ and ξ measure the quality of the approximation of $x^\star$ by w. When this approximation is reasonable, problem (3) requires much less measurements than standard ℓ_1 minimization, i.e., (3) with $\beta = 0$. For example, in a noiseless acquisition scenario, standard ℓ_1 minimization requires $2s\log(n/s) + (7/5)s + 1$ measurements for perfect reconstruction with the same probability as in Theorem 1 [23]. When the dominant terms are the log's, (5) can be much smaller than this bound, since $\overline{h} \leq s$.

Finally, we mention that [8] also provides bounds for the case $\beta \neq 1$, but they are significantly more complex than (5) and (6). Interestingly, those bounds are minimized for $\beta = 1$, a value that leads to a practical performance close to optimal. For this reason, we set $\beta = 1$ henceforth.

Algorithm 1 Adaptive-Rate Sparse State Estimation

Initialization: choose $0 \leq \alpha \leq 1$, $\delta > 0$, and estimate s_1 and s_2, the sparsity of $x[1]$ and $x[2]$, respectively.

for the first two time instants $k = 1, 2$ **do**
  Set $m_k = 2s_k\log(n/s_k) + (7/5)s_k + 1$
  Generate Gaussian matrices $A_k \in \mathbb{R}^{m_k\times n}$
  Acquire m_k measurements of $x[k]$: $y[k] = A_k\, x[k]$
  Find $\hat{x}[k]$ such that

$$\hat{x}[k] \in \begin{array}{ll} \underset{x}{\text{Argmin}} & \|x\|_1 \\ \text{s.t.} & A_k\, x = y[k] \end{array}$$

end for
Set $w[2] = f_2(\hat{x}[1])$ and compute $\overline{h}_2$ and ξ_2 as in (4) with $\hat{x}[2]$ and $w[2]$ in place of $x^\star$ and w, respectively.
Set $\overline{m}_2 = 2\overline{h}_2\log(n/(s_2 + \xi_2/2)) + (7/5)(s_2 + \xi_2/2) + 1$
Set $\phi_3 = \overline{m}_2$
for each time instant $k = 3, 4, 5, \ldots$ **do**
  Choose $m_k = (1 + \delta)\phi_k$
  Generate Gaussian matrix $A_k \in \mathbb{R}^{m_k\times n}$
  Acquire m_k measurements of $x[k]$: $y[k] = A_k\, x[k]$
  Set $w[k] = f_k(\hat{x}[k-1])$ and find $\hat{x}[k]$ such that

$$\hat{x}[k] \in \begin{array}{ll} \underset{x}{\text{Argmin}} & \|x\|_1 + \big\|x - w[k]\big\|_1 \\ \text{s.t.} & A_k\, x = y[k] \end{array}$$

  Compute $\overline{h}_k$ and ξ_k as in (4) with $\hat{x}[k]$ and $w[k]$
  Set $s_k = |\{i\,:\, \hat{x}[k] \neq 0\}|$
  Set $\overline{m}_k = 2\overline{h}_k\log(n/(s_k + \xi_k/2)) + (7/5)(s_k + \xi_k/2) + 1$
  Update $\phi_{k+1} = (1 - \alpha)\phi_k + \alpha\,\overline{m}_k$
end for

4. DYNAMIC SIGNAL RECONSTRUCTION

Algorithm 1 describes the scheme we propose for recursive estimation of $x[k]$. For simplicity, we consider only the noiseless measurement scenario, but its adaptation to the noisy one is immediate.

The algorithm is meant to be run on a real-time system, since the measurements taken at each iteration are determined on-the-fly. In steps 1-6, the first two signals, $x[1]$ and $x[2]$, are reconstructed using standard ℓ_1 minimization. The number of measurements m_1 and m_2 are computed as in [23], and require an estimate of the signals' sparsity. Steps 7-9 initialize our "estimator" ϕ_k of the true bound on the number of measurements. That is, during the recursive part of the algorithm, i.e., steps 10-19, ϕ_k should approximate the right-hand side of (5) for $s = s_k$, $\overline{h} = \overline{h}_k$, and $\xi = \xi_k$, where the subscript k indicates that these are parameters associated with $x[k]$. Since ϕ_k is only an approximation, we take more measurements than the ones it prescribes, as in step 11, where δ is a safeguard parameter. Steps 12-14 describe the measurement process and the reconstruction of $\hat{x}[k]$ using ℓ_1-ℓ_1 minimization. Next, steps 15-16 compute the quantities $\overline{h}_k$, ξ_k, and s_k, and step 17 computes $\overline{m}_k$ which, if the reconstruction was perfect, equals the right-hand side of (5) applied to $x[k]$. Note, however, that $\overline{m}_k$ is computed only after the measurements of $x[k]$ have been taken and the reconstruction of $\hat{x}[k]$ has occurred. The value $\overline{m}_k$ is then used in step 18 to update ϕ_k as an exponential moving average filter with parameter α.

To explain the rationale for the filtering step and the safeguard parameter δ, suppose there is no filtering, i.e., $\alpha = 1$. In that case, the estimator ϕ_k of (5) applied to $x[k]$ is simply $\overline{m}_{k-1}$ which, if $\hat{x}[k-1] = x[k-1]$ (perfect reconstruction at $k-1$), equals (5) applied to $x[k-1]$. Since (5) applied to $x[k-1]$ might differ from (5) applied to $x[k]$, we take more measurements for $x[k]$ than the ones specified by $\overline{m}_{k-1}$, that is, $m_k = (1+\delta)\overline{m}_{k-1}$ (step 11). So, even when there is perfect reconstruction at time $k-1$, δ should be large enough to account for variations of (5) from $x[k-1]$ to $x[k]$; see Lemma 1 below for a lower bound. If reconstruction fails at time $k-1$, i.e., $\hat{x}[k-1] \neq x[k-1]$, $\overline{m}_{k-1}$ may be very different from (5) applied to $x[k-1]$ and to $x[k]$. The reason for filtering $\overline{m}_k$ in step 18 is to mitigate the effect of these failed reconstructions. We therefore recommend setting $\alpha < 1$.

Reconstruction guarantees. The following lemma considers $\alpha = 1$ and derives a lower bound on the probability of reconstruction success at each time, provided δ is large enough.

Lemma 1. *Let $\alpha = 1$, $k > 2$, and $\underline{m} = \min_{i=1,\ldots,k} \overline{m}_k$. Let also*

$$\delta \geq \max_{3\leq i\leq k} \frac{2\big[\overline{h}_i \log(\frac{n}{u_i}) - \overline{h}_{i-1}\log(\frac{n}{u_{i-1}})\big] + \frac{7}{5}(u_i - u_{i-1})}{1 + 2\overline{h}_{i-1}\log(\frac{n}{u_{i-1}}) + \frac{7}{5}u_{i-1}}, \tag{7}$$

where $u_i := s_i + \xi_i/2$. Assume $s_q \geq |\{j : x_j[q] \neq 0\}|$, for $q = 1, 2$, i.e., that the initial sparsity estimates s_1 and s_2 are not smaller than the true sparsity of $x[1]$ and $x[2]$. Then, the probability that Algorithm 1 reconstructs $x[i]$ perfectly in all time instants $1 \leq i \leq k$ is not smaller than

$$\Big(1 - \exp\big[-\frac{1}{2}(\underline{m} - \sqrt{\underline{m}})^2\big]\Big)^k. \tag{8}$$

Proof. Since $\alpha = 1$, step 11 becomes $m_i = (1+\delta)\overline{m}_{i-1}$, for all $3 \leq i \leq k$. According to Theorem 1, if $(1+\delta)\overline{m}_{i-1}$ is not smaller than the right-hand side of (5) applied to $x[i]$, that is,

$$(1+\delta)\overline{m}_{i-1} \geq 2\overline{h}_i \log\Big(\frac{n}{u_i}\Big) + \frac{7}{5}u_i + 1, \tag{9}$$

then the probability of perfect reconstruction at time i is not smaller than $1 - \exp\big(-\frac{1}{2}(m_i - \sqrt{m_i})^2\big)$. In other words,

$$\mathbb{P}(S_i|E_i) \geq 1 - \exp\Big[-\frac{1}{2}(m_i - \sqrt{m_i})^2\Big], \tag{10}$$

where S_i is the event "perfect reconstruction at time i" and E_i is the event in (9). Simple algebraic manipulation shows that if we replace the expression for $\overline{m}_{i-1}$ (in step 17) in (9) , we obtain

$$\delta \geq \frac{2\big[\overline{h}_i \log(\frac{n}{u_i}) - \overline{h}_{i-1}\log(\frac{n}{u_{i-1}})\big] + \frac{7}{5}(u_i - u_{i-1})}{1 + 2\overline{h}_{i-1}\log(\frac{n}{u_{i-1}}) + \frac{7}{5}u_{i-1}}. \tag{11}$$

That is, (11) is event E_i. Therefore, condition (7) corresponds to the event $E := E_3 \wedge E_4 \wedge \cdots \wedge E_k$. And we have

$$\begin{aligned}
&\mathbb{P}\big(S_1 \wedge S_2 \wedge \cdots \wedge S_k | E\big) && (12)\\
&= \mathbb{P}(S_1)\mathbb{P}(S_2)\prod_{i=3}^{k} \mathbb{P}(S_i|S_1 \wedge \cdots \wedge S_{i-1} \wedge E) && (13)\\
&= \mathbb{P}(S_1)\mathbb{P}(S_2)\prod_{i=3}^{k} \mathbb{P}(S_i|E_i)\,. && (14)
\end{aligned}$$

From (12) to (13), we used the fact that S_1 and S_2 are independent. From (13) to (14), we used the fact that $S_i|E = S_i|E_i$, for $3 \leq i \leq k$, and that the events S_i conditioned on E_i (i.e., (9)) are independent, for $3 \leq i \leq k$. Now note that, since $\overline{m}_i \geq \underline{m}$ and $1 - \exp(-(1/2)(x - \sqrt{x})^2)$ is an increasing function, (10) implies

$$\mathbb{P}(S_i|E_i) \geq 1 - \exp\Big[-\frac{1}{2}(\underline{m} - \sqrt{\underline{m}})^2\Big]. \tag{15}$$

The right-hand side of (15) also lower bounds $\mathbb{P}(S_1)$ and $\mathbb{P}(S_2)$ [23]. From (14) and (15), we obtain

$$\mathbb{P}\big(S_1 \wedge S_2 \wedge \cdots \wedge S_k|E\big) \geq \bigg(1 - \exp\Big[-\frac{1}{2}(\underline{m} - \sqrt{\underline{m}})^2\Big]\bigg)^k,$$

and the lemma is proved. □

When the conditions of the lemma hold, the probability of successful reconstruction decreases with time, albeit with a very slow rate: for example, if $\underline{m} = 8$, which is very small for applications, the right-hand side of (8) gives 0.9998 for $k = 10^2$, and 0.9845 for $k = 10^4$. Larger $\underline{m}$ give even smaller rates.

To get some insight about (7), let j be an index for which the maximum is achieved in the right-hand side of (7). Also, let n be much larger than s_j and ξ_j. Then, (7) becomes

$$\delta \geq \frac{2(\overline{h}_j - \overline{h}_{j-1})\log n + o(\log n)}{2\overline{h}_{j-1}\log n + o(\log n)},$$

and, for a large n,

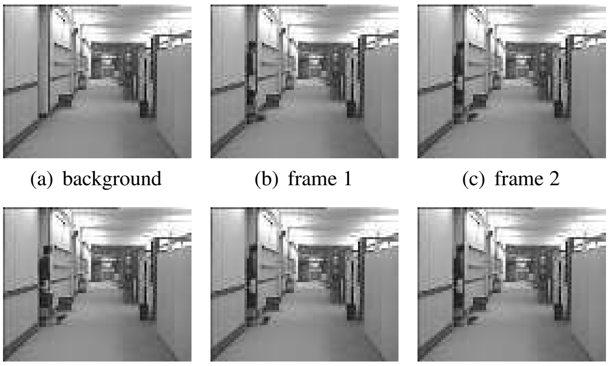
$$\delta \gtrsim \frac{\overline{h}_j - \overline{h}_{j-1}}{\overline{h}_{j-1}}. \tag{16}$$

Equation (16) tells us that, for $\alpha = 1$ and for very sparse signals, the oversampling factor δ in Algorithm 1 should be greater than the largest relative increase between two consecutive $\overline{h}_k$'s. Writing

$$\overline{h}_k = \big|\{i : x_i[k] > 0,\, \epsilon_i[k] > 0\} \cup \{i : x_i[k] < 0,\, \epsilon_i[k] < 0\}\big|$$

(see (4a)), we conclude that $\overline{h}_k$ increases if and only if there is a new index i for which $x_i[k]$ and $\epsilon_i[k]$ have the same sign.

Variations of Algorithm 1. For example, rather than generating a matrix A_k at each iteration, one can generate a single (Gaussian) matrix $A \in \mathbb{R}^{n\times n}$ at the beginning and, at each iteration, select m_k rows of A randomly. Another variation, motivated by Lemma 1, sets $\alpha = 1$ and recursively updates δ applying, e.g., an exponential moving average filter to the expression in the right-hand side of (7).



(a) background (b) frame 1 (c) frame 2

(d) frame 3 (e) frame 3: prediction (f) frame 3: reconstr.

Fig. 1. (a)-(d) background image and first three frames; (e) predicted image using (reconstructed) frames 1 and 2; and (f) reconstruction of frame 3 by ℓ_1-ℓ_1 minimization.

5. EXPERIMENTAL RESULTS

We assessed the performance of Algorithm 1 by applying it to compressive background subtraction [13], which we explain next.

Compressive background subtraction. Let $\{z[k]\}$ be a sequence of (vectorized) images with the same background $b \in \mathbb{R}^n$, assumed known. We have access only to a set of m_k linear measurements $u[k] = A_k z[k]$ from each image $z[k]$, where $A_k \in \mathbb{R}^{m_k \times n}$ is a measurement matrix. Each $z[k]$ can then be decomposed as $z[k] = x[k] + b$, where $x[k]$ is the image *foreground*. As noticed in [13], foregrounds are typically sparse and thus can be reconstructed using standard ℓ_1 minimization. To do it, we need access to foreground measurements, which can be obtained as follows [13]: given $u[k]$, take measurements of the known background b using the same measurement matrix, $u^0 := A_k b$, and subtract them from $u[k]$, i.e., $y[k] := u[k] - u^0 = A_k(z[k] - b) = A_k x[k]$.

Our approach. In our experiments, we modified the model in (1) by assuming that each $x[k]$ is generated by the *two* previous signals, i.e., $x[k] = f_k(x[k-1], x[k-2])$. This modification has no implications on our algorithm or on the associated reconstruction guarantees. However, it allows us to model the action of a motion-compensated extrapolation algorithm [24–26]: given two (consecutive) images, $z[k-1]$ and $z[k-2]$, predict the next one, $z[k]$, assuming linear motion. We perform extrapolation on the image domain rather than on the foreground domain, because the texture of the former is richer and improves the estimation performance. The side information fed to ℓ_1-ℓ_1 minimization is, of course, in the foreground domain: $w[k] = e[k] - b$, where $e[k]$ is the image extrapolated by the motion-compensated algorithm.

Experimental setup. We used the *Hall video sequence* (`http://trace.eas.asu.edu/yuv/`), from which we removed the first 18 frames, as they had no foreground. Each image was downsampled to a resolution of 128×128. For the motion-compensated extrapolation, we used sub-pel motion estimation with a block size of 8×8 pixels and a search range of 6 pixels. The parameters α and δ were 0.5 and 0.1, respectively. The parameters s_1 and s_2 were initialized with the true sparsity of the first two foregrounds. To solve each ℓ_1-ℓ_1 minimization problem in step 14 of Algorithm 1, we used ADMM [27, 28], where one term of the objective function is $\|x\|_1 + \|x - w\|_1$ and the other term is the indicator function of the linear system $Ax = y$. It can be shown that both terms have closed-form proximal operators. The augmented Lagrangian parameter was updated as suggested in [29].

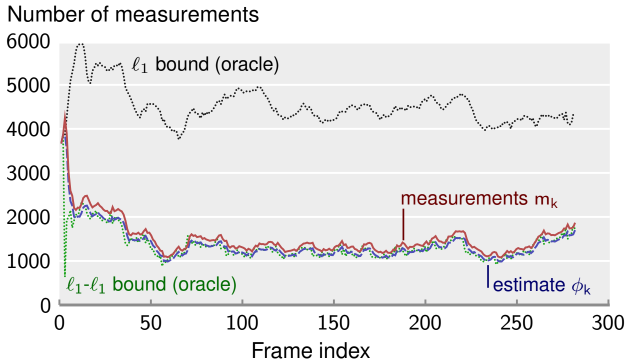


Fig. 2. Number of measurements m_k, estimate ϕ_k, and right-hand side (5) for $x[k]$ per frame. It is also shown the bound for standard ℓ_1 minimization [23].

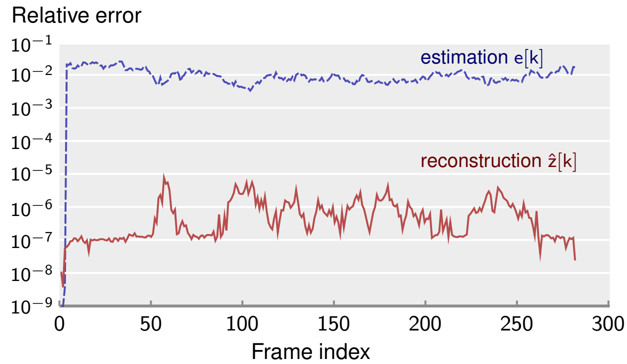


Fig. 3. Relative errors for estimation and reconstruction per frame.

Results. Fig. 2 shows the number of measurements m_k taken by Algorithm 1 (red), and the estimator ϕ_k (blue). It also shows the oracle bounds for standard ℓ_1 minimization in [23] (black) and ℓ_1-ℓ_1 minimization (5) (green). These are called "oracle" because they are computed assuming the signal to be reconstructed is known. We remark that all previous approaches to compressive background subtraction, e.g., [13, 22], require always more measurements than the standard ℓ_1 bound, i.e., their performance curves are always above the black line. We thus see that our algorithm allows a dramatic reduction in the number of required measurements. Furthermore, its performance is close to optimal, since the number of measurements (red line) follows the ℓ_1-ℓ_1 bound (green line) very closely. Fig. 3 shows the relative reconstruction and estimation errors, respectively, $\|\hat{z}[k] - z[k]\|_2 / \|z[k]\|_2$ and $\|e[k] - z[k]\|_2 / \|z[k]\|_2$. It can be seen that each frame was reconstructed almost perfectly. In fact, the reconstruction error was determined mostly by the precision of the solver we used for ℓ_1-ℓ_1 minimization.

6. CONCLUSIONS

We proposed a recursive ℓ_1-ℓ_1 minimization algorithm for reconstructing time varying sparse signals from a limited number of linear measurements. Based on recent theoretical results on ℓ_1-ℓ_1 minimization, the algorithm estimates, on-the-fly, the number of measurements required to reconstruct the signal in the next time instant. Experimental results on compressive background subtraction using real test video data demonstrate the validity of our estimation scheme and the high reconstruction performance of ℓ_1-ℓ_1 minimization.

7. REFERENCES

[1] R. Kalman, "A new approach to linear filtering and prediction problems," *Trans. ASME-J. Basic Engineering*, vol. 82, no. D, pp. 35–45, 1960.

[2] S. Haykin, *Kalman Filtering and Neural Networks*, John Wiley and Sons, 2001.

[3] J. Geromel, "Optimal linear filtering under parameter uncertainty," *IEEE Trans. Sig. Proc.*, vol. 47, no. 1, pp. 168–175, 1999.

[4] L. El Ghaoui and G. Calafiore, "Robust filtering for discrete-time systems with bounded noise and parametric uncertainty," *IEEE Trans. Aut. Control*, vol. 46, no. 7, pp. 1084–1089, 2001.

[5] A. Charles, M. Asif, J. Romberg, and C. Rozell, "Sparsity penalties in dynamical system estimation," in *IEEE Conf. Information Sciences and Systems*, 2011, pp. 1–6.

[6] E. Candès, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *IEEE Trans. Info. Th.*, vol. 52, no. 2, pp. 489–509, 2006.

[7] D. Donoho, "Compressed sensing," *IEEE Trans. Info. Th.*, vol. 52, no. 4, pp. 1289–1306, 2006.

[8] J. Mota, N. Deligiannis, and M. Rodrigues, "Compressed sensing with prior information: Optimal strategies, geometry, and bounds," submitted to IEEE Trans. Info. Th., `http://arxiv.org/abs/1408.5250`, 2014.

[9] J. Mota, N. Deligiannis, and M. Rodrigues, "Compressed sensing with side information: Geometrical interpretation and performance bounds," accepted at IEEE Global Conf. on Signal and Information Processing (GlobalSIP), 2014.

[10] M. Lustig, D. Donoho, and J. Pauly, "Sparse MRI: The application of compressed sensing to rapid MR imaging," *Magnetic Resonance in Medicine*, vol. 58, pp. 1182–1195, 2007.

[11] G.-H. Chen, J. Tang, and S. Leng, "Prior image constrained compressed sensing (PICCS): a method to accurately reconstruct dynamic CT images from highly undersampled projection data sets," *Med Phys.*, vol. 35, no. 2, pp. 660–663, 2008.

[12] M. Herman and T. Strohmer, "High-resolution radar via compressed sensing," *IEEE Trans. Sig. Proc.*, vol. 57, no. 6, pp. 2275–2284, 2009.

[13] V. Cevher, A. Sankaranarayanan, M. Duarte, D. Reddy, R. Baraniuk, and R. Chellappa, "Compressive sensing for background subtraction," in *Europ. Conf. Computer Vision (ECCV)*, 2008.

[14] N. Vaswani, "Kalman filtered compressed sensing," in *IEEE Conf. Image Proc. (ICIP)*, 2008, pp. 893–896.

[15] N. Vaswani, "Analyzing least squares and Kalman filtered compressed sensing," in *IEEE Conf. Acoustics, Speech and Signal Processing (ICASSP)*, 2009, pp. 3013–3016.

[16] J. Zhan and N. Vaswani, "Robust PCA with partial subspace knowledge," in *IEEE Int. Symp. Inf. Theory (ISIT)*, 2014.

[17] J. Ziniel, L. Potter, and P. Schniter, "Tracking and smoothing of time-varying sparse signals via approximate belief propagation," in *Proceedings of the Asilomar Conf. on Signals, Systems and Computers*, 2010, pp. 808–812.

[18] D. Angelosante, G. Giannakis, and E. Grossi, "Compressed sensing of time-varying signals," in *IEEE Inter. Conf. on Digital Signal Processing*, 2009, pp. 1–8.

[19] M. Asif and J. Romberg, "Dynamic updating for ℓ_1 minimization," *IEEE J. Selected Topics in Signal Processing*, vol. 4, no. 2, pp. 421–434, 2010.

[20] A. Sankaranarayanan, P. Turaga, R. Chellapa, and R. Baranuik, "Compressive acquisition of linear dynamical systems," *SIAM J. Imaging Sciences*, vol. 6, no. 4, pp. 2109–2133, 2013.

[21] D. Sejdinović, C. Andrieu, and R. Piechocki, "Bayesian sequential compressed sensing in sparse dynamical systems," in *48th Annual Allerton Conf.*, 2010, pp. 1730–1736.

[22] G. Warnell, D. Reddy, and R. Chellappa, "Adaptive rate compressive sensing for background subtraction," in *IEEE Conf. Acoustics, Speech and Signal Processing (ICASSP)*, 2012, pp. 1477–1480.

[23] V. Chandrasekaran, B. Recht, P. Parrilo, and A. Willsky, "The convex geometry of linear inverse problems," *Found. Comp. Mathematics*, vol. 12, pp. 805–849, 2012.

[24] B. Girod, A. Aaron, S. Rane, and D. Rebollo-Monedero, "Distributed video coding," *Proceedings of the IEEE*, vol. 93, no. 1, pp. 71–83, 2005.

[25] R. Puri, A. Majumdar, and K. Ramchandran, "PRISM: A video coding paradigm with motion estimation at the decoder," *IEEE Trans. Image Processing*, vol. 16, no. 10, pp. 2436–2448, 2007.

[26] N. Deligiannis, J. Barbarien, M. Jacobs, A. Munteanu, A. Skodras, and P. Schelkens, "Side-information-dependent correlation channel estimation in hash-based distributed video coding," *IEEE Trans. Image Processing*, vol. 21, no. 4, pp. 1934–1949, 2012.

[27] R. Glowinski and A. Marrocco, "Sur l'approximation, par éléments finis d'ordre un, et la résolution, par pénalisation-dualité, d'une classe de problèmes de dirichelet non linéaires," *Revue Française d'Automatique, Informatique, et Recherche Opérationelle*, vol. 9, no. 2, pp. 41–76, 1975.

[28] D. Gabay and B. Mercier, "A dual algorithm for the solution of nonlinear variational problems via finite element approximations," *Computers and Mathematics with Applications*, vol. 2, no. 1, pp. 17–40, 1976.

[29] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating method of multipliers," *Foundations and Trends in Machine Learning*, vol. 3, no. 1, pp. 1–122, 2011.