Adaptive-Rate Reconstruction of Time-Varying Signals with Application in Compressive Foreground Extraction

João F. C. Mota *Member, IEEE*, Nikos Deligiannis *Member, IEEE*, Aswin C. Sankaranarayanan *Member, IEEE*, Volkan Cevher *Senior Member, IEEE*, Miguel R. D. Rodrigues *Senior Member, IEEE*

***Abstract*—We propose and analyze an online algorithm for reconstructing a sequence of signals from a limited number of linear measurements. The signals are assumed sparse, with unknown support, and evolve over time according to a generic nonlinear dynamical model. Our algorithm, based on recent theoretical results for ℓ_1-ℓ_1 minimization, is recursive and computes the number of measurements to be taken at each time on-the-fly. As an example, we apply the algorithm to online compressive video foreground extraction, a problem stated as follows: given a set of measurements of a sequence of images with a static background, simultaneously reconstruct each image while separating its foreground from the background. The performance of our method is illustrated on sequences of real images. We observe that it allows a dramatic reduction in the number of measurements or reconstruction error with respect to state-of-the-art compressive background subtraction schemes.**

***Index Terms*—State estimation, compressive video, background subtraction, sparsity, ℓ_1 minimization, motion estimation.**

I. Introduction

CONSIDER the problem of reconstructing a sequence of sparse signals from a limited number of measurements. Let $x[k] \in \mathbb{R}^n$ be the signal at time k and $y[k] \in \mathbb{R}^{m_k}$ be the vector of signal measurements at time k, where $m_k \ll n$. Assume the signals evolve according to the dynamical model

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

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

where $\epsilon[k] \in \mathbb{R}^n$ is modeling noise and $A_k \in \mathbb{R}^{m_k \times n}$ is a sensing matrix. In (1a), $f_k : (\mathbb{R}^n)^{k-1} \to \mathbb{R}^n$ is a known, but otherwise arbitrary, map that describes $x[k]$ as a function of past signals. We assume that each $x[k]$ and $\epsilon[k]$ is sparse, i.e., they have a small number of nonzero entries. Our goal is to reconstruct the signal sequence $\{x[k]\}$ from the measurement sequence $\{y[k]\}$. We require the reconstruction scheme to be recursive (or online), i.e., $x[k]$ is reconstructed before acquiring measurements of any future signal $x[i]$, $i > k$, and also to use a minimal number of measurements. We formalize the problem as follows.

Problem statement. *Given two unknown sparse sequences $\{x[k]\}$ and $\{\epsilon[k]\}$ satisfying* (1)*, design an online algorithm that 1) uses a minimal number of measurements m_k at time k, and 2) perfectly reconstructs each $x[k]$ from $y[k]$ acquired as in* (1b)*, and possibly $x[i]$, $i < k$.*

Note that our setting immediately generalizes from the case where each $x[k]$ is sparse to the case where $x[k]$ has a sparse representation in a linear, invertible transform.[1] An aspect that distinguishes our problem from other recursive signal reconstruction problems, such as the ones addressed in [2]–[9], is that the number of measurements m_k varies at each iteration and has to be computed recursively.

A. Applications

Many problems require estimating a sequence of signals from a sequence of measurements satisfying the model in (1). These include classification and tracking in computer vision systems [10], [11], radar tracking [12], dynamic MRI [13], [14] and several tasks in wireless sensor networks [15].

Our application focus, however, is *compressive background subtraction* [16] (a.k.a. foreground subtraction). Background subtraction is a key task for detecting and tracking objects in a video sequence and it has been applied, for example, in video surveillance [17], [18], traffic monitoring [19], [20], and medical imaging [21], [22]. Although there are many background subtraction techniques, e.g., [11], [23], [24], most of

J. Mota, and M. Rodrigues were supported by the EPSRC grant EP/K033166/1. A. C. Sankaranarayanan was supported by the NSF grant CCF-1117939. N. Deligiannis was supported by the VUB strategic research programme M3D2. V. Cevher's was supported 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. Part of this work was presented at the IEEE International Conference on Acoustics, Speech, and Signal Processing, 2015 [1].

J. Mota and M. Rodrigues are with the Department of Electronic and Electrical Engineering, University College London, UK.
E-mail: {j.mota,m.rodrigues}@ucl.ac.uk.

N. Deligiannis is with the Department of Electronics and Informatics, Vrije Universiteit Brussel, 1050 Brussels, Belgium, and also with iMinds, 9050 Ghent, Belgium. He was with the Department of Electronic and Electrical Engineering, University College London, UK.
E-mail: ndeligia@etro.vub.ac.be.

A. Sankaranarayanan is with the Department of Electrical and Computer Engineering, Carnegie Mellon University, PA 15213, USA.
E-mail: saswin@andrew.cmu.edu.

V. Cevher is with the Laboratory for Information and Inference Systems (LIONS), EPFL, Switzerland.
E-mail: volkan.cevher@epfl.ch.

[1] If $x[k]$ is not sparse but $z[k] := \Psi x[k]$ is, where Ψ is an invertible matrix, then redefine f_k as the composition $f_k^z = \Psi^{-1} \circ f_k \circ \Psi$ and A_k as $A_k^z := A_k \Psi^{-1}$. The signal $z[k]$ satisfies (1) with f_k^z and A_k^z.

them assume access to full frames and, thus, are inapplicable in compressive video sensing [25]–[30], a technology used in cameras where sensing is expensive. In particular, devices such as the single-pixel camera [25]–[30] acquire compressive measurements from images using few sensors, as in (1b); this allows expensive sensors that capture images in non-conventional wavelengths, such as infrared or UV, and also allows high-speed imaging [30]. Performing these tasks with conventional camera technology is expensive due to the large number of required sensors/pixels. Medical imaging [13], [14], [21] is another application where measurements are expensive and are acquired as in (1b). Although we do not explore this application, we believe our work can have significant impact in dynamic MRI [13], [14].

In compressive video sensing, one has access not to full frames as in conventional video, but only to a small set of linear measurements of each frame, as in (1b). Cevher et al. [16] noticed that background subtraction is possible in this context if the foreground pixels, i.e., those associated to a moving object, occupy a small area in each frame. Assuming the background image is known beforehand, compressed sensing techniques [31], [32] such as ℓ_1-norm minimization allow reconstructing each foreground. This not only reconstructs the original frame (if we add the reconstructed foreground to the known background), but also performs background subtraction as a by-product [16].

With the exception of [33], [34], most approaches to compressive video sensing and compressive background subtraction assume a fixed number of measurements for all frames [16], [25], [27]–[30], [35]. If this number is too small, reconstruction fails. If it is too large, reconstruction succeeds, but at the cost of unnecessary measurements in some or all frames. The work in [33], [34] addresses this problem with an online scheme that uses cross-validation to compute the number of required measurements. Given a reconstructed foreground, [33], [34] estimates the area of the true foreground using extra cross-validation measurements. Then, assuming that consecutive frames have equal foreground areas, the phase diagram of the sensing matrix, computed beforehand, gives the number of measurements for the next frame. This approach, however, fails to use information from past frames in the reconstruction process, information that, as we will see, can be used to significantly reduce the number of measurements.

B. Overview of our approach and contributions

Overview. Our approach to adaptive-rate signal reconstruction is based on the recent theoretical results of [36], [37]. These characterize the performance of sparse reconstruction schemes in the presence of side information. The scheme we are most interested in is ℓ_1-ℓ_1 *minimization*:

$$\begin{array}{ll} \underset{x}{\text{minimize}} & \|x\|_1 + \beta\|x - w\|_1 \\ \text{subject to} & Ax = y\,, \end{array} \tag{2}$$

where $x \in \mathbb{R}^n$ is the optimization variable and $\|x\|_1 := \sum_{i=1}^n |x_i|$ is the ℓ_1-norm. In (2), $y \in \mathbb{R}^m$ is a vector of measurements, β a positive parameter. The vector $w \in \mathbb{R}^n$ is assumed known and is the so-called *prior* or *side information*: a vector similar to the vector that we want to reconstruct, say $x^\star$. Note that if we set $\beta = 0$ in (2), we obtain *basis pursuit* (BP) [38], a well-known sparse reconstruction problem at the core of *compressed sensing* [31], [32]. Problem (2) generalizes BP by integrating the side information w. The work in [36], [37] shows that, if w has reasonable quality (in a precise sense defined below) and the entries of A are drawn from an i.i.d. Gaussian distribution, the number of measurements required by (2) to reconstruct $x^\star$ is much smaller than the number of measurements required by BP. Furthermore, the theory in [36], [37] establishes that $\beta = 1$ is an optimal choice, irrespective of any problem parameter. This makes the reconstruction problem (2) parameter-free.

We address the problem of recursively reconstructing a sequence of sparse signals satisfying (1) as follows. Assuming the measurement matrix is Gaussian,[2] we propose an algorithm that uses (2) with $w = f_k\big(\{x[i]\}_{i=1}^{k-1}\big)$ to reconstruct each signal $x[k]$. And, building upon the results of [36], [37], we equip our algorithm with a mechanism to automatically compute an estimate on the number of required measurements. As application, we consider compressive background subtraction and show how to generate side information from past frames.

Contributions. We summarize our contributions as follows:

i) We propose an adaptive-rate algorithm for reconstructing sparse sequences satisfying the model in (1).
ii) We establish conditions under which our algorithm reconstructs a finite sparse sequence $\{x[i]\}_{i=1}^k$ with large probability.
iii) We describe how to apply the algorithm to compressive background subtraction problems, using motion-compensated extrapolation to predict the next image to be acquired. In other words, we show how to generate side information.
iv) Given that images predicted by motion-compensated extrapolation are known to exhibit Laplacian noise, we then characterize the performance of (2) under this model.
v) Finally, we show the impressive performance of our algorithm for performing compressive background subtraction on several sequences of real images.

Besides the incorporation of a scheme to compute a minimal number of measurements on-the-fly, there is another aspect that makes our algorithm fundamentally different from prior work. As overviewed in Section II, most prior algorithms for reconstructing dynamical sparse signals work well only when the sparsity pattern of $x[k]$ varies slowly with time. Our algorithm, in contrast, operates well even when the sparsity pattern of $x[k]$ varies arbitrarily between consecutive time instants, as shown by our theory and experiments. What is required to vary slowly is the "quality" of the prediction given by each f_k (i.e., the quality of the side information) and, to a lesser extent, not the sparsity pattern of $x[k]$ but only its sparsity, i.e., the *number of* nonzero entries.

[2]Although Gaussian matrices are hard to implement in practical systems, they have optimal performance. There are, however, other more practical matrices with a similar performance, e.g., [39], [40].

C. Organization

Section II overviews related work. In Section III, we state the results from [36], [37] that are used by our algorithm. Section IV describes the algorithm and establishes reconstruction guarantees. Section V concerns the application to compressive background subtraction. Experimental results illustrating the performance of our algorithm are shown in section VI; and section VII concludes the paper. The appendix contains the proofs of our results.

II. Related work

There is extensive literature on reconstructing time-varying signals from limited measurements. Here, we provide an overview by referring a few landmark papers.

The Kalman filter. The classical solution to estimate a sequence of signals satisfying (1) or, in the control terminology, the state of a dynamical system, is the Kalman filter [41]. The Kalman filter is an online algorithm that is least-squares optimal when the model is linear, i.e., $f_k\big(\{x[i]\}_{i=0}^{k-1}\big) = Fx[k-1]$, and the sequence $\{\epsilon[k]\}$ is Gaussian and independent across time. Several extensions are available when these assumptions do not hold [42]–[44]. The Kalman filter and its extensions, however, are inapplicable to our scenario, as they do not easily integrate the additional knowledge that the state is sparse.

Dynamical sparse signal reconstruction. Some prior work incorporates signal structure, such as sparsity, into online sparse reconstruction procedures. For example, [3]–[5] adapts a Kalman filter to estimate a sequence of sparse signals. Roughly, we have an estimate of the signal's support at each time instant and use the Kalman filter to compute the (nonzero) signal values. When a change in the support is detected, the estimate of the support is updated using compressed sensing techniques. The work in [3]–[5], however, assumes that the support varies very slowly and does not provide any strategy to update (or compute) the number of measurements; indeed, the number of measurements is assumed constant along time. Also assuming the support varies slowly and using a fixed number of measurements, the work in [45]–[47] addresses the same problem using a different approach: it integrates an estimate of the signal's support into the sparse reconstruction scheme using a technique known as Modified-CS [48]. Note that the methods in [3]–[5], [45]–[47] come with stability guarantees, in the sense that if the number of (fixed) measurements is large enough and consecutive signals have approximately the same support, then these methods have smaller restricted isometry constants (and thereby milder reconstruction guarantees) than simple sparse reconstruction methods.

Related work that also assumes a fixed number of measurements includes [49], which uses approximate belief propagation, and [50], which integrates sparsity knowledge into a Kalman filter via a pseudo-measurement technique. The works in [6], [7] and [8] propose online algorithms named GROUSE and PETRELS, respectively, for estimating signals that lie on a low-dimensional subspace. Their model can be seen as a particular case of (1), where each map f_k is linear and depends only on the previous signal. Akin to most prior work, both GROUSE and PETRELS assume that the rank of the underlying subspace (i.e., the sparsity of $x[k]$) varies slowly with time, and fail to provide a scheme to compute the number of measurements.

We briefly overview the work in [9], which is close to ours, since it uses a related reconstruction problem. However, perhaps due to lack of theory for that problem, it assumes a fixed number of measurements for all signals (as most prior work) and, thus, does not solve the problem we consider in this paper. Three dynamical reconstruction schemes are studied in [9]. The one with the best performance is

$$\underset{x}{\text{minimize}}\ \|x\|_1 + \beta\|x-w\|_1 + \beta_2\|Ax-y\|_2^2\,, \qquad (3)$$

where $\beta_2 > 0$ and $\|\cdot\|_2$ is the Euclidean ℓ_2-norm. Problem (3) is the Lagrangian version of the noise-robust version of (2):

$$\begin{aligned} \underset{x}{\text{minimize}} \quad & \|x\|_1 + \beta\|x-w\|_1 \\ \text{subject to} \quad & \|Ax-y\|_2 \le \sigma\,, \end{aligned} \qquad (4)$$

where σ is a bound on the measurement noise. For β_2 in a given range, the solutions of (3) and (4) coincide. This makes the approach in [9] related to ours. Nevertheless, using (4) has two important advantages: first, in practice, it is easier to obtain bounds on the measurement noise σ than it is to tune β_2; second, and more importantly, the problem in (4) has well-characterized reconstruction guarantees [36], [37]. It is exactly those guarantees that enable our scheme for computing of the number of measurements online.

Robust PCA. A technique that has been successfully applied to perform background subtraction is Robust PCA (RPCA) [24], [51], [52]. RPCA decomposes a data matrix into the sum of a sparse and a low-rank matrix. In the context of background subtraction, a column of the data matrix corresponds to a video frame, which is decomposed into foreground (sparse matrix) and background (low-rank matrix). In RPCA, both the foreground and background are unknown, and the latter may vary slowly across time. Also, it either requires access to full frames or, in the case of compressive RPCA [53], each measurement may contain information from several different frames, e.g., the first and last frames, making it a batch algorithm: all frames are processed together, not online as in our algorithm.

There are, however, online extensions of RPCA, for example, the stochastic optimization approach of [54], and an algorithm called Prac-ReProCS [55]. The online algorithm in [54] is shown to converge to the same solution as the batch RPCA, but it requires access to full frames. Prac-ReProCS [55] is also backed up by theory [56] and can handle compressive measurements. It requires a training sequence of background images, which is akin to our assumption of knowing the background image, and works under the assumption that the foreground objects move slowly. The version of Prac-ReProCS that handles compressive measurements, however, allows to reconstruct only the foreground sequence and not the (slow-changing) background and, in addition, assumes that the measurement matrix is the same for all frames. This implies, in particular, that the number of measurements is fixed, which is the reason why Prac-ReProCS fails to solve our problem.

III. Preliminaries: Static signal reconstruction using ℓ_1-ℓ_1 minimization

This section reviews some results from [36], namely reconstruction guarantees for (2) in a static scenario, i.e., when we estimate just one signal, not a sequence. As mentioned before, $\beta = 1$ is an optimal choice: it not only minimizes the bounds in [36], but also leads to the best results in practice. This is the reason why we use $\beta = 1$ henceforth.

ℓ_1-ℓ_1 minimization. Let $x^\star \in \mathbb{R}^n$ be a sparse vector, and assume we have m linear measurements of $x^\star$: $y = Ax^\star$, where $A \in \mathbb{R}^{m\times n}$. Denote the *sparsity* of $x^\star$ with $s := |\{i : x_i^\star \neq 0\}|$, where $|\cdot|$ is the cardinality of a set. Assume we have access to a signal $w \in \mathbb{R}^n$ similar to $x^\star$ (in the sense that $\|x^\star - w\|_1$ is small) and suppose we attempt to reconstruct $x^\star$ by solving the ℓ_1-ℓ_1 minimization problem (2) with $\beta = 1$:

$$\begin{array}{ll} \underset{x}{\text{minimize}} & \|x\|_1 + \|x - w\|_1 \\ \text{subject to} & Ax = y\,. \end{array} \tag{5}$$

The number of measurements that problem (5) requires to reconstruct $x^\star$ is a function of the "quality" of the side information w. Quality in [36] is measured in terms of the following parameters:

$$\xi := \big|\{i : w_i \neq x_i^\star = 0\}\big| - \big|\{i : w_i = x_i^\star \neq 0\}\big|\,, \tag{6a}$$

$$\overline{h} := \big|\{i : x_i^\star > 0,\ x_i^\star > w_i\} \cup \{i : x_i^\star < 0,\ x_i^\star < w_i\}\big|\,. \tag{6b}$$

Note that the number of components of w that contribute to $\overline{h}$ are the ones defined on the support of $x^\star$; thus, $0 \le \overline{h} \le s$.

Theorem 1 (Th. 1 in [36])**.** *Let $x^\star, w \in \mathbb{R}^n$ be the vector to reconstruct and the side information, respectively. 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 $1/m$. If*

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

then, with probability at least $1 - \exp\big(-\frac{1}{2}(m - \sqrt{m})^2\big)$, $x^\star$ is the unique solution of (5)*.*

Theorem 1 establishes that if the number of measurements is larger than (7) then, with high probability, (5) reconstructs $x^\star$ perfectly. The bound in (7) is a function of the signal dimension n and sparsity s, and of the quantities ξ and $\overline{h}$, which depend on the signs of the entries of $x^\star$ and $w - x^\star$, but not on their magnitudes. When w is such that $\overline{h}$ is small, the bound in (7) is much smaller than the one for BP[3] in [57]:

$$m \ge 2s\log\Big(\frac{n}{s}\Big) + \frac{7}{5}s + 1\,. \tag{8}$$

Namely, [57] establishes that if (8) holds and if $A \in \mathbb{R}^{m\times n}$ has i.i.d. Gaussian entries with zero mean and variance $1/m$ then, with probability similar to the one in Theorem 1, $x^\star$ is the unique solution to BP. Indeed, if $\overline{h} \ll s$ and ξ is larger than a small negative constant, then (7) is much smaller than (8). Note that, in practice, the quantities s, ξ, and $\overline{h}$ are unknown, since they depend on the unknown signal $x^\star$. In the next section, we propose an online scheme to estimate them using past signals.

[3]Recall that Basis Pursuit (BP) is (2) with $\beta = 0$.

Noisy case. Theorem 1 has a counterpart for noisy measurements, which we state informally; see [36] for details. Let $y = Ax^\star + \eta$, where $\|\eta\|_2 \le \sigma$. Let also $A \in \mathbb{R}^{m\times n}$ be as in Theorem 1 with

$$m \ge \frac{1}{(1-\tau)^2}\left[2\overline{h}\log\Big(\frac{n}{s+\xi/2}\Big) + \frac{7}{5}\Big(s + \frac{\xi}{2}\Big) + \frac{3}{2}\right], \tag{9}$$

where $0 < \tau < 1$. Let $\hat{x}_{\text{noisy}}$ be any solution of (4). Then, with overwhelming probability, $\|\hat{x}_{\text{noisy}} - x^\star\|_2 \le 2\sigma/\tau$, i.e., (4) reconstructs $x^\star$ stably. Our algorithm, described in the next section, adapts easily to the noisy scenario, but we provide reconstruction guarantees only for the noiseless case.

Algorithm 1 Adaptive-Rate Sparse Signal Reconstruction

Input: $0 \le \alpha \le 1$, a positive sequence $\{\delta_k\}$, and estimates $\hat{s}_1$ and $\hat{s}_2$ of the sparsity of $x[1]$ and $x[2]$, respectively.

Part I: Initialization

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

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

end for

Set $w[2] = f_2(\hat{x}[1])$ and compute

$$\hat{\xi}_2 := \big|\{i : w_i[2] \neq \hat{x}_i[2] = 0\}\big| - \big|\{i : w_i[2] = \hat{x}_i[2] \neq 0\}\big|$$

$$\hat{\overline{h}}_2 := \big|\{i : \hat{x}_i[2] > 0,\ \hat{x}_i[2] > w_i[2]\} \cup \{i : \hat{x}_i[2] < 0,\ \hat{x}_i[2] < w_i[2]\}\big|\,.$$

Set $\hat{\overline{m}}_2 = 2\hat{\overline{h}}_2 \log\big(n/(\hat{s}_2 + \hat{\xi}_2/2)\big) + (7/5)\big(\hat{s}_2 + \hat{\xi}_2/2\big) + 1$

Set $\phi_3 = \hat{\overline{m}}_2$

Part II: Online estimation

for each time instant $k = 3, 4, 5, \ldots$ **do**
  Set $m_k = (1 + \delta_k)\phi_k$
  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 $w[k] = f_k(\{\hat{x}[i]\}_{i=1}^{k-1})$ and find $\hat{x}[k]$ such that

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

  Compute

$$\hat{s}_k = |\{i : \hat{x}[k] \neq 0\}|$$

$$\hat{\xi}_k = \big|\{i : w_i[k] \neq \hat{x}_i[k] = 0\}\big| - \big|\{i : w_i[k] = \hat{x}_i[k] \neq 0\}\big|$$

$$\hat{\overline{h}}_k = \big|\{i : \hat{x}_i[k] > 0,\ \hat{x}_i[k] > w_i[k]\} \cup \{i : \hat{x}_i[k] < 0,\ \hat{x}_i[k] < w_i[k]\}\big|\,.$$

  Set $\hat{\overline{m}}_k = 2\hat{\overline{h}}_k \log\big(n/(\hat{s}_k + \hat{\xi}_k/2)\big) + (7/5)\big(\hat{s}_k + \hat{\xi}_k/2\big) + 1$
  Update $\phi_{k+1} = (1 - \alpha)\phi_k + \alpha\, \hat{\overline{m}}_k$

end for

IV. Online sparse signal estimation

Algorithm 1 describes our online scheme for reconstructing a sparse sequence $\{x[k]\}$ satisfying (1), i.e.,

$$x[k] = f_k\big(\{x[i]\}_{i=1}^{k-1}\big) + \epsilon[k]$$
$$y[k] = A_k\, x[k]\,.$$

Intuitively, given estimates $\hat{x}[i]$, $i < k$, of all the signals prior to time k, Algorithm 1 reconstructs $x[k]$ by solving (5) with $A = A_k$, $y = y[k]$, and $w = f_k\big(\{x[i]\}_{i=1}^{k-1}\big)$. Although described for a noiseless scenario, the algorithm easily adapts to the noisy scenario, as discussed later. Such adaptation is essential on a real system, e.g., a single-pixel camera [26].

A. Algorithm description

The algorithm consists of two parts: the initialization, where the first two signals $x[1]$ and $x[2]$ are reconstructed using BP, and the online estimation, where the remaining signals are reconstructed using ℓ_1-ℓ_1 minimization.

Part I: Initialization. Steps 1-6 compute the number of measurements m_1 and m_2 according to the bound in (8), and then reconstruct $x[1]$ and $x[2]$ via BP. The expressions for m_1 and m_2 in step 2 require estimates $\hat{s}_1$ and $\hat{s}_2$ of the sparsity of $x[1]$ and $x[2]$, which are given as input to the algorithm. Henceforth, variables with hats refer to estimates. Steps 7-9 initialize the estimator ϕ_k: during Part II of the algorithm, ϕ_k should approximate the right-hand side of (7) for $x[k]$, i.e., with $s = s_k$, $\overline{h} = \overline{h}_k$, and $\xi = \xi_k$, where the subscript k indicates that these are parameters associated with $x[k]$.

Part II: Online estimation. The loop in Part II starts by computing the number of measurements as $m_k = (1+\delta_k)\phi_k$, where δ_k, an input to the algorithm, is a (positive) safeguard parameter. We take more measurements from $x[k]$ than the ones prescribed by ϕ_k, because ϕ_k is only an approximation to the bound in (7), as explained next. After acquiring measurements from $x[k]$, we reconstruct it as $\hat{x}[k]$ via ℓ_1-ℓ_1 minimization with $w[k] = f_k(\{\hat{x}[i]\}_{i=1}^{k-1})$ (step 14). Next, in step 15, we compute the sparsity $\hat{s}_k$ and the quantities in (6), $\hat{\xi}_k$ and $\hat{\overline{h}}_k$, for $\hat{x}[k]$. If the reconstruction of $x[k]$ is perfect, i.e., $\hat{x}[k] = x[k]$, then all these quantities match their true values. In that case, $\hat{\overline{m}}_k$ in step 16 will also match the true value of the bound in (7). Note, however, that the bound for $x[k]$, $\hat{\overline{m}}_k$, is computed only after $x[k]$ is reconstructed. Consequently, the number of measurements used in the acquisition of $x[k]$, $k > 2$, is a function of the bound (7) for $x[k-1]$. Since the bounds for $x[k]$ and $x[k-1]$ might differ, we take more measurements than the ones specified by ϕ_k by a factor δ_k, as in step 11. Also, we mitigate the effect of failed reconstructions by filtering $\hat{\overline{m}}_k$ with an exponential moving average filter, in step 17. Indeed, if reconstruction fails for some $x[k]$, the resulting $\hat{\overline{m}}_k$ might differ significantly from the true bound in (7). The role of the filter is to smooth out such variations.

Extension to the noisy case. Algorithm 1 can be easily extended to the scenario where the acquisition process is noisy, i.e., $y[k] = A_k x[k] + \eta_k$. Assume η_k is arbitrary noise, but has bounded magnitude, i.e., we know σ_k such that $\|\eta_k\|_2 \leq \sigma_k$. In that case, the constraint in the reconstruction problems in steps 5 and 14 should be replaced by $\|A_k x - y[k]\|_2 \leq \sigma_k$. The other modification is in steps 8 and 16, whose expressions for $\hat{\overline{m}}_k$ are multiplied by $1/(1-\tau)^2$ as in (9). Our reconstruction guarantees, however, hold only for the noiseless case.

Remarks. We will see in the next section that Algorithm 1 works well when each δ_k is chosen according to the prediction quality of f_k: the worse the prediction quality, the larger δ_k should be. In practice, it may be more convenient to make δ_k constant, as we do in our experiments in Section VI. Note that the conditions under which our algorithm performs well differ from the majority of prior work. For example, the algorithms in [3], [4], [6]–[8], [16], [33], [34], [49], [50], [58] work well when the sparsity pattern of $x[k]$ varies slowly between consecutive time instants. Our algorithm, in contrast, works well when the quality parameters ξ_k and $\overline{h}_k$ and also the sparsity s_k vary slowly; in other words, when the quality of the prediction of f_k varies slowly. On the other hand, if f_k gives a bad quality prediction at iteration k, the number of measurements taken at iteration $k+1$ will approach (8), for BP. To see that, assume $x[k]$ and $\hat{x}[k]$ have the same support (i.e., $\hat{x}[k]$ was reconstructed successfully), which is disjoint from the support of $w[k]$. In that case, $\hat{\overline{h}}_k = s_k$ and, when n is large enough, the dominant terms in (7) and (8) both equal $2s_k \log n$. We finally remark that Algorithm 1 can possibly be adapted to reconstruction problems that integrate prior information in a different way (e.g., modified-CS, modified-BPDN [48], and ℓ_1-ℓ_2 minimization [36]); the only requirement is that sample complexity bounds as in Theorem 1 are available.

B. Reconstruction guarantees

The following result bounds the probability with which Algorithm 1 with $\alpha = 1$ perfectly reconstructs a finite-length sequence $\{x[i]\}_{i=1}^{k}$. The idea is to rewrite the condition that (7) applied to $x[i-1]$ is $(1+\delta_i)$ times larger than (7) applied to $x[i]$. If that condition holds for the entire sequence then, using Theorem 1 and assuming that the matrices A_k are drawn independently, we can bound the probability of successful reconstruction. The proof is in Appendix A.

Lemma 2. *Let $\alpha = 1$, $\underline{m} := \min\big\{m_1, m_2, \min_{i=3,\ldots,k} \hat{\overline{m}}_i\big\}$, and fix $k \geq 3$. Let also, for all $i = 3, \ldots, k$,*

$$\delta_i \geq \frac{2\big[\overline{h}_i \log(\frac{n}{u_i}) - \overline{h}_{i-1}\log(\frac{n}{u_{i-1}})\big] + \frac{7}{5}(u_i - u_{i-1})}{2\overline{h}_{i-1}\log(\frac{n}{u_{i-1}}) + \frac{7}{5}u_{i-1} + 1}\,, \quad (10)$$

where $u_i := s_i + \xi_i/2$. Assume $\hat{s}_q \geq s_q := |\{j : x_j[q] \neq 0\}|$, for $q = 1, 2$, i.e., that the initial sparsity estimates $\hat{s}_1$ and $\hat{s}_2$ are not smaller than the true sparsity of $x[1]$ and $x[2]$. Assume also that the matrices $\{A_i\}_{i=1}^{k}$ are drawn independently. Finally, assume that the estimates $w[i]$ produced by f_i are such that, for all $i = 3, \ldots, k$, $\overline{h}_i > 0$ and $x_j[i] = w_j[i] = 0$ for at least one index j. Then, the probability (over the sequence of matrices $\{A_i\}_{i=1}^{k}$) that Algorithm 1 reconstructs $x[i]$ perfectly in all time instants $1 \leq i \leq k$ is at least

$$\Big(1 - \exp\Big[-\frac{1}{2}(\underline{m} - \sqrt{\underline{m}})^2\Big]\Big)^k\,. \quad (11)$$

When the conditions of Lemma 2 hold, the probability of perfect reconstruction decreases with the length k of the

sequence, albeit at a very slow rate: for example, if $\underline{m}$ is as small as 8, then (11) equals 0.9998 for $k = 10^2$, and 0.9845 for $k = 10^4$. If $\underline{m}$ is larger, these numbers are even closer to 1.

Interpretation of (10)**.** As shown in the proof, condition (10) is equivalent to $(1 + \delta_i)\overline{m}_{i-1} \geq \overline{m}_i$, where $\overline{m}_i$ is (7) applied to $x[i]$. To get more insight about this condition, rewrite it as

$$\delta_i \geq \frac{\overline{h}_i - \overline{h}_{i-1} + c_1(n)}{\overline{h}_{i-1} + c_2(n)}\,, \tag{12}$$

where

$$c_1(n) := \frac{2\overline{h}_{i-1}\log u_{i-1} - 2\overline{h}_i \log u_i + \frac{7}{5}(u_i - u_{i-1})}{2\log n}$$

$$c_2(n) := \frac{\frac{7}{5}u_{i-1} + 1 - 2\overline{h}_{i-1}\log u_{i-1}}{2\log n}\,.$$

Suppose $\{x[i]\}$ and $\{\epsilon[i]\}$ are signals for which $n \gg u_i, \overline{h}_i$. In that case, $c_1(n), c_2(n) \simeq 0$, and condition (12) tells us that the oversampling factor δ_i should be larger than the relative variation of $\overline{h}_i$ from time $i-1$ to time i. In general, the magnitude of $c_1(n)$ and $c_2(n)$ can be significant, since they approach zero at a relatively slow rate, $o(1/\log n)$. Hence, those terms should not be ignored.

Remarks on the noisy case. There is an inherent difficulty in establishing a counterpart of Lemma 2 for the noisy measurement scenario: namely, the quality parameters ξ and $\overline{h}$ in (6) are not continuous functions of x. So, no matter how close a reconstructed signal is from the original one, their quality parameters can differ arbitrarily. And, for the noisy measurement case, we can never guarantee that the reconstructed and the original signals are equal; at most, if (9) holds, they are within a distance $2\sigma/\tau$, for $0 < \tau < 1$.

So far, we have considered $\{x[k]\}$ and $\{\epsilon[k]\}$ to be deterministic sequences. In the next section, we will model $\{\epsilon[k]\}$ (and thus $\{x[k]\}$) as a Laplacian stochastic process.

V. Compressive Video Background subtraction

We now consider the application of our algorithm to compressive video background subtraction. We start by modeling the problem of compressive background subtraction as the estimation of a sequence of sparse signals satisfying (1). Our background subtraction system, based on Algorithm 1, is then introduced. Finally, we establish reconstruction guarantees for our scheme when $\epsilon[k]$ in (1a) is Laplacian noise.

A. Model

Let $\{Z[k]\}_{k\geq 1}$ be a sequence of images with resolution $N_1 \times N_2$, and let $z[k] \in \mathbb{R}^n$ with $n := N_1 \cdot N_2$ be the column-major vectorization of the kth image. At time k, we collect m_k linear measurements of $Z[k]$: $u[k] = A_k z[k]$, where $A_k \in \mathbb{R}^{m_k \times n}$ is a measurement matrix. We decompose each image $Z[k]$ as $Z[k] = X[k] + B$, where $X[k]$ is the kth foreground image, typically sparse, and B is the background image, assumed known and the same in all images. Let $x[k]$ and b be vectorizations of $X[k]$ and B. Because the background image is known, we take measurements from it using A_k: $u^b[k] = A_k b$. Then, as suggested in [16], we subtract $u^b[k]$ from $u[k]$:

$$y[k] := u[k] - u^b[k] = A_k(z[k] - b) = A_k x[k]\,. \tag{13}$$

This equation tells us that, although we cannot measure the foreground image $x[k]$ directly, we can still construct a vector measurements, $y[k]$, as if we would. Given that $x[k]$ is usually sparse, the theory of compressed sensing tells us that it can be reconstructed by solving, for example, BP [31], [32]. Specifically, if $x[k]$ has sparsity s_k and the entries of A_k are realizations of i.i.d. zero-mean Gaussian random variables with variance $1/m_k$, then $2s_k \log(n/s_k) + (7/5)s_k + 1$ measurements suffice to reconstruct $x[k]$ perfectly [57] [cf. (8)].

Notice that (13) is exactly the equation of measurements in (1b). Regarding equation (1a), we will use it to model the estimation of the foreground of each frame, $x[k]$, from previous foregrounds, $\{x[i]\}_{i=1}^{k-1}$. We use a motion-compensated extrapolation technique, as explained in Subsection V-C. This technique is known to produce image estimates with an error well modeled as Laplacian and, thus, each $\|\epsilon[k]\|_1$ is expected to be small. This perfectly aligns with the way we integrate side information in our reconstruction scheme: namely, the second term in the objective of the optimization problem in step 14 of Algorithm 1 is exactly $\|\epsilon[k]\|_1$.

B. Our background subtraction scheme

Fig. 1 shows the block diagram of our compressive background subtraction scheme and, essentially, translates Algorithm 1 into a diagram. The scheme does not apply to the reconstruction of the first two frames, which are reconstructed as in [16], i.e., by solving BP. This corresponds to Part I of Algorithm 1. The scheme in Fig. 1 depicts Part II of Algorithm 1. The motion extrapolation module constructs a motion-compensated prediction $e[k]$ of the current frame, $z[k]$, by using the two past (reconstructed) frames, $\hat{z}[k-2]$ and $\hat{z}[k-1]$. Motion estimation is performed in the image domain ($z[k]$) rather than in the foreground domain ($x[k]$), as the former contains more texture, thereby yielding a more accurate motion field. Next, the background frame b is subtracted from $e[k]$ to obtain a prediction of the foreground $x[k]$, i.e., the side information $w[k]$. These two operations are modeled in Algorithm 1 with the function f_k, which takes a set of past reconstructed signals (in our case, $\hat{x}[k-2]$ and $\hat{x}[k-1]$, to which we add b, obtaining $\hat{z}[k-2]$ and $\hat{z}[k-1]$, respectively), and outputs the side information $w[k]$. This is one of the inputs of the ℓ_1-ℓ_1 block, which solves the optimization problem (5). To obtain the other input, i.e., the set of foreground measurements $y[k]$, we proceed as specified in equation (13): we take measurements $u[k] = A_k z[k]$ of the current frame and, using the same matrix, we take measurements of the background $u[k] = A_k b$. Subtracting them we obtain $y[k] = u[k] - u^b[k]$. The output of the ℓ_1-ℓ_1 module is the estimated foreground $\hat{x}[k]$, from which we obtain the estimate of the current frame as $\hat{z}[k] = \hat{x}[k] + b$.

C. Motion-compensated extrapolation

To obtain an accurate predition $e[k]$, we use a motion-compensated extrapolation technique similar to what is used

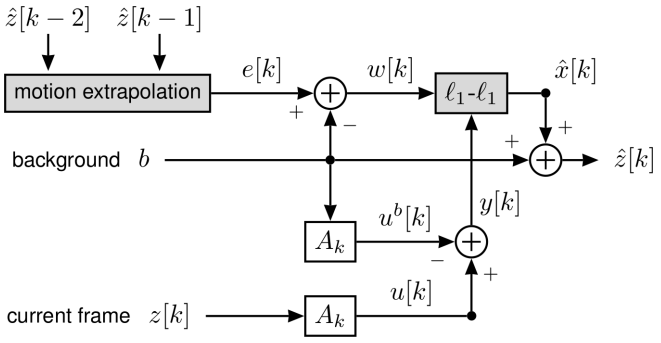


Figure 1. Block diagram of Algorithm 1 when applied to background subtraction. The main blocks are highlighted.

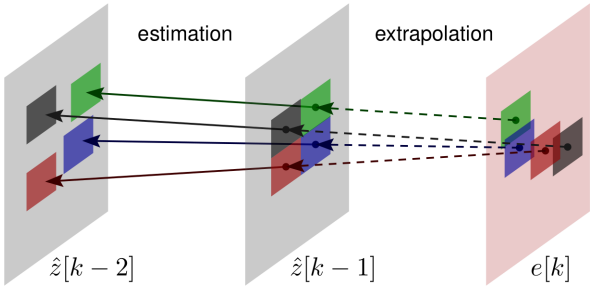


Figure 2. Scheme of motion-compensated extrapolation. We use the motion between matching blocks in $\hat{z}[k-2]$ and $\hat{z}[k-1]$ to create an estimate $e[k]$ of frame $z[k]$.

in distributed video coding for generating decoder-based motion-compensated predictions [59]–[61]. Our technique is illustrated in Fig. 2. In the first stage, we perform forward block-based motion estimation between the reconstructed frames $\hat{z}[k-2]$ and $\hat{z}[k-1]$. The block matching algorithm is performed with half-pel accuracy and considers a block size of $\gamma \times \gamma$ pixels and a search range of ρ pixels. The required interpolation for half-pel motion estimation is performed using the 6-tap filter of H.264/AVC [62]. In addition, we use the ℓ_1-norm (or sum of absolute differences, SAD) as error metric. The resulting motion vectors are then spatially smoothed by applying a weighted vector-median filter [63]. The filtering improves the spatial coherence of the resulting motion field by removing outliers (i.e., motion vectors that are far from the true motion field). Assuming linear motion between $\hat{z}[k-2]$ and $\hat{z}[k-1]$, and $\hat{z}[k-1]$ and $\hat{z}[k]$, we linearly project the motion vectors between $\hat{z}[k-2]$ and $\hat{z}[k-1]$ to obtain $e[k]$, our estimate of $z[k]$; see Fig. 2. During motion compensation, pixels in $e[k]$ that belong to overlapping prediction blocks are estimated as the average of their corresponding motion-compensated pixel predictors in $\hat{z}[k-1]$. Pixels in uncovered areas (i.e., no motion-compensated predictor is available) are estimated by averaging the three neighbor pixel values in $e[k]$ (up, left and up-left pixel positions, following a raster scan of the frame) and the corresponding pixel in $\hat{z}[k-1]$.

Remarks. In the above scheme, the motion extrapolation block creates an estimate $e[k]$ of the current frame $z[k]$ by assuming linear motion between the past two frames, $z[k-2]$ and $z[k-1]$. Hence, it gives good predictions when foreground objects move linearly. This is the case of small displacements, since nonlinear motion is locally well approximated by linear motion. When displacements are large (i.e., objects move fast), the above scheme still gives good predictions if the motion is linear. This contrasts with prior methods for background subtraction and dynamical sparse signal estimation, such as [3], [4], [6]–[8], [16], [33], [34], [49], [50], [58], which work well only in the slow-motion case, i.e., when the foreground area between consecutive frames is approximately constant.

In practice, the background may change during the operation of the algorithm due to, for example, illumination change or camera movement. As in video coding [60]–[62], this indicates a scene cut, which can be detected by our algorithm via a dramatic increase in the measurement rate. In that case, the algorithm should take enough measurements to reconstruct an entire frame, the new background.

D. Reconstruction guarantees for Laplacian modeling noise

It is well known that the noise produced by a motion-compensated prediction module, as the one just described, is Laplacian [59], [64]. In our model, that corresponds to each $\epsilon[k]$ in (1a) being Laplacian. We assume each $\epsilon[k]$ is independent from the matrix of measurements A_k.

Model for $\epsilon[k]$. As in [59], [64], [65] (and references therein), we assume that $\epsilon[k]$ is independent from $\epsilon[l]$, for $k \neq l$, and that the entries of each $\epsilon[k]$ are independent and have zero-mean. The probability distribution of $\epsilon[k]$ is then

$$\begin{aligned}\mathbb{P}(\epsilon[k] \leq u) &= \mathbb{P}(\epsilon_1[k] \leq u_1,\ \epsilon_2[k] \leq u_2, \ldots,\ \epsilon_n[k] \leq u_n) \\ &= \prod_{j=1}^{n} \mathbb{P}(\epsilon_j[k] \leq u_j) \\ &= \prod_{j=1}^{n} \int_{-\infty}^{u_j} \frac{\lambda_j}{2} \exp\left[-\lambda_j |\epsilon_j| \right] d\epsilon_j \,, \end{aligned} \tag{14}$$

where $u \in \mathbb{R}^n$ and $\lambda_j \geq 0$. In words, each $\epsilon_j[k]$ has Laplace distribution with parameter λ_j. The entries of $\epsilon[k]$, although independent, are not identically distributed, since they have possibly different parameters λ_j. The variance σ_j^2 of each component $\epsilon_j[k]$ is given by $\sigma_j^2 = 2/\lambda_j^2$.

Resulting model for $x[k]$. The sequence $\{\epsilon[k]\}$ being stochastic implies that $\{x[k]\}$ is also stochastic. Indeed, if each f_k in (1) is measurable, then $\{x[k]\}_{k\geq 2}$ is a sequence of random variables. Given the independence across time and across components of the sequence $\{\epsilon[k]\}$, the distribution of $x[k]$ given $\{x[i]\}_{i=1}^{k-1}$ is also Laplacian, yet not necessarily with zero-mean. That is, for $u \in \mathbb{R}^n$ and $k \geq 2$,

$$\begin{aligned}&\mathbb{P}\Big(x[k] \leq u \;\Big|\; \{x[i]\}_{i=1}^{k-1}\Big) \\ &= \mathbb{P}\Big(f_k(\{x[i]\}_{i=1}^{k-1}) + \epsilon[k] \leq u \;\Big|\; \{x[i]\}_{i=1}^{k-1}\Big) \\ &= \mathbb{P}\Big(\epsilon[k] \leq u - f_k(\{x[i]\}_{i=1}^{k-1}) \;\Big|\; \{x[i]\}_{i=1}^{k-1}\Big) \\ &= \prod_{j=1}^{n} \int_{-\infty}^{u_j - [f_k(\{x[i]\}_{i=1}^{k-1})]_j} \frac{\lambda_j}{2} \exp\left[-\lambda_j |\epsilon_j| \right] d\epsilon_j \\ &= \prod_{j=1}^{n} \int_{-\infty}^{u_j} \frac{\lambda_j}{2} \exp\Big[-\lambda_j \big| z_j - [f_k(\{x[i]\}_{i=1}^{k-1})]_j \big| \Big] dz_j \end{aligned} \tag{15}$$

where $[f_k(\{x[i]\}_{i=1}^{k-1})]_j$ is the jth component of $f_k(\{x[i]\}_{i=1}^{k-1})$. In words, the distribution of each component of $x[k]$ conditioned on all past realizations $x[i]$, $1 \leq i < k$, is Laplacian with mean $[f_k(\{x[i]\}_{i=1}^{k-1})]_j$ and parameter λ_j. Furthermore, it is independent from the other components.

Reconstruction guarantees. Note that $\{x[k]\}$ and $\{\epsilon[k]\}$ being stochastic processes implies that the quantities in (6), which we will denote by ξ_k and $\overline{h}_k$ for signal $x[k]$, are random variables. Hence, at each time k, the conditions of Theorem 1, namely that $\overline{h}_k > 0$ and that there is at least one index i such that $x_i[k] = w_i[k] = 0$, become events, and may or may not hold. We now impose conditions on the variances $\sigma_j^2 = 2/\lambda_j$ that guarantee the conditions of Theorem 1 are satisfied and, thus, that ℓ_1-ℓ_1 minimization reconstructs $x[k]$ perfectly, with high probability. Given $S \in \{1, \ldots, n\}$, S^c denotes its complement in $\{1, \ldots, n\}$.

Theorem 3. *Let $w \in \mathbb{R}^n$ be given. Let ϵ have distribution* (14)*, where the variance of component ϵ_j is $\sigma_j^2 = 2/\lambda_j^2$. Define $x^\star := w + \epsilon$, and the sets $\Sigma := \{j : \sigma_j^2 \neq 0\}$ and $\mathcal{W} := \{j : w_j \neq 0\}$. Assume $\Sigma^c \cap \mathcal{W}^c \neq \emptyset$, that is, there exists j such that $\sigma_j^2 = 0$ and $w_j = 0$. Assume $A \in \mathbb{R}^{m \times n}$ is generated as in Theorem 1 with a number of measurements*

$$m \geq 2(\mu + t) \log \left(\frac{n}{|\Sigma| + \frac{1}{2}|\Sigma^c \cap \mathcal{W}|} \right) + \frac{7}{5}\Big(|\Sigma| + \frac{1}{2}|\Sigma^c \cap \mathcal{W}| \Big) + 1 \,, \quad (16)$$

for some $t > 1$, where $\mu := \frac{1}{2}\sum_{j \in \Sigma} \left[1 + \exp\left(-\sqrt{2}|w_j|/\sigma_j\right)\right]$. Let $\hat{x}$ denote the solution of ℓ_1-ℓ_1 minimization (5)*. Then,*

$$\mathbb{P}\big(\hat{x} = x^\star\big) \geq \left[1 - \exp\Big(- \frac{(m - \sqrt{m})^2}{2} \Big)\right] \times \times \left[1 - \exp\Big(- \frac{2\mu^2}{|\Sigma|} \Big) - \exp\Big(- \frac{2(t-1)^2}{|\Sigma|} \Big)\right]. \quad (17)$$

The proof is in Appendix B. By assuming each component ϵ_j is Laplacian with parameter $\lambda_j = \sqrt{2}/\sigma_j$ (independent from the other components), Theorem 3 establishes a lower bound on the number of measurements that guarantee perfect reconstruction of $x^\star$ with probability as in (17). Note that all the quantities in (16) are deterministic. This contrasts with the direct application of Theorem 1 to the problem, since the right-hand side of (7) is a function of the random variables s, $\overline{h}$, and ξ. The assumption $\Sigma^c \cap \mathcal{W}^c \neq \emptyset$ implies $\Sigma^c \neq \emptyset$, which means that some components of ϵ have zero variance and, hence, are equal to zero with probability 1. Note that, provided the variances σ_j^2 are known, all the quantities in (16), and consequently in (17), are known.

The proof of Theorem 3 uses the fact that the sparsity of $x^\star$ is $s = |\Sigma| + |\Sigma^c \cap \mathcal{W}|/2$ with probability 1. This implies that the bound in (16) is always smaller than the one for BP in (8) whenever $\mu + t < s = |\Sigma| + |\Sigma^c \cap \mathcal{W}|$. Since $\mu \leq |\Sigma|$, this holds if $t < |\Sigma^c \cap \mathcal{W}|/2$.

We state without proof a consequence of Theorem 3 that is obtained by reasoning as in Lemma 2:

Corollary 4. *Let $\{\epsilon[k]\}$ be a stochastic process where $\epsilon[k]$ has distribution* (14) *and each $\epsilon[k]$ is independent from $\epsilon[l]$, $k \neq l$. Assume that $\{x[k]\}$ is generated as in* (1a) *and consider Algorithm 1 with $\alpha = 1$ at iteration $k > 2$. Assume that $\epsilon[k]$ and A_k are independent. Assume also, for $i = 3, \ldots, k$, that*

$$\delta_i \geq \Bigg\{ 2\Big[(\mu_i + t_i)\log\Big(\frac{n}{u_i}\Big) - (\mu_{i-1} + t_{i-1})\log\Big(\frac{n}{u_{i-1}}\Big)\Big] + \frac{7}{5}(u_i - u_{i-1}) \Bigg\} \Big/ \Bigg\{ 2(\mu_{i-1} + t_{i-1})\log\Big(\frac{n}{u_{i-1}}\Big) + \frac{7}{5}u_{i-1} + 1 \Bigg\}, \quad (18)$$

where $u_i := |\Sigma_i| + |\Sigma_i^c \cap \mathcal{W}_i|/2$, and the quantities μ_i, t_i, Σ_i, and $\mathcal{W}_i$ are defined as in Theorem 3 for signal $x[i]$. Assume the initial sparsity estimates satisfy $\hat{s}_1 \geq s_1$ and $\hat{s}_2 \geq s_2$ with probability 1, where s_1 and s_2 are the sparsity of $x[1]$ and $x[2]$. Then, the probability over the sequences $\{A_i\}_{i=1}^k$ and $\{\epsilon[i]\}_{i=1}^k$ that Algorithm 1 reconstructs $x[i]$ perfectly in all time instants $1 \leq i \leq k$ is at least

$$\prod_{i=1}^{k} \left[1 - \exp\Big(- \frac{(m_i - \sqrt{m_i})^2}{2} \Big)\right] \left[1 - \exp\Big(- \frac{2\mu_i^2}{|\Sigma_i|} \Big) - \exp\Big(- \frac{2(t_i - 1)^2}{|\Sigma_i|} \Big)\right].$$

Corollary 4 establishes reconstruction guarantees of Algorithm 1 when the modeling noise $\epsilon[k]$ in (1a) is Laplacian. In contrast with Lemma 2, the bound in (18) is a function of known parameters, but it requires the variances $\sigma_j^2[i]$ of each $\epsilon_j[i]$, which can be estimated from the past frame in a block-based way [60], [64]. For some insight on (18), assume $u_i \simeq u_{i-1}$, $t_i = t_{i-1}$, and that n is large enough so that terms not depending on it are negligible. Then, (18) becomes $\delta_i \gtrsim (\mu_i - \mu_{i-1})/(\mu_{i-1} + t_{i-1})$, and we can select

$$\delta_i = 2\kappa - 1 \simeq \frac{\kappa - \frac{1}{2}}{\frac{1}{|\Sigma_{i-1}|} + \frac{1}{2}} = \frac{|\Sigma_i| - \frac{1}{2}|\Sigma_{i-1}|}{1 + \frac{1}{2}|\Sigma_{i-1}|} \geq \frac{\mu_i - \mu_{i-1}}{\mu_{i-1} + t_{i-1}}, \quad (19)$$

where $\kappa := |\Sigma_i|/|\Sigma_{i-1}|$, and the inequality is due to $|\Sigma_i|/2 \leq \mu_i \leq |\Sigma_i|$. The approximation in (19) holds if $|\Sigma_{i-1}| \gg 2$, which is often the case in practice. The expression in (19) tells us that, for large n, δ_i is mostly determined by the ratio κ: if $\kappa > 1$ (resp. < 1), then we should select $\delta_i > 1$ (resp. < 1). We observe that, in practice, (18) and (19) give conservative estimates for δ_i. We will see in the next section that selecting a small, constant δ_i (namely 0.1 and 0.01) leads to excellent results without compromising perfect reconstruction.

VI. Experimental results

We applied the scheme described in the previous section to several video sequences.[4] The sequences are described in Table I and were obtained from [66][5], [34][6], [67][7], [18][8], and

[4]The code to reproduce these experiments is available in http://www.ee.ucl.ac.uk/~jmota/AdaptiveRateBackSubtrCode.zip.

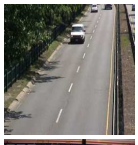
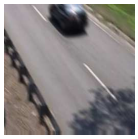
[5]http://trace.eas.asu.edu/yuv/hall_monitor/hall_qcif.7z

[6]http://garrettwarnell.com/ARCS-1.0.zip; for the original, see view 5 of ftp://ftp.cs.rdg.ac.uk/pub/PETS2009/Crowd_PETS09_dataset/a_data/Crowd_PETS09/S2_L1.tar.bz2

[7]http://www.changedetection.net/

[8]http://www.vis.uni-stuttgart.de/~hoeferbn/bse/dataset/SABS-Basic.rar

Table I
VIDEO SEQUENCES USED IN THE EXPERIMENTS.

Acronym	Ref.	Frames	Resolution	Sample image
Hall	[66]	18-300	128 × 128	
PETS2009	[34]	1-171	116 × 116	
Highway	[67]	1-500	120 × 160	
PETS2006	[67]	1-1200	144 × 180	
Canoe	[67]	820-1100	160 × 120	
Stuttgart	[68]	11-310	144 × 192	
Park	[67]	20-600	144 × 176	
Caviar	[69]	344-1235	144 × 192	
Traffic	[67]	1-250	120 × 160	

[69][9]. The table shows the acronyms we refer each sequence by, their source, the frame numbers and resolution we used, and a sample image from each sequence. We performed two types of experiments, Type I and Type II, each with a different purpose. The Type I experiments validate our algorithm, which was designed for Gaussian measurement matrices. Note that Gaussian matrices require explicit storage and their multiplication with vectors is expensive; therefore, in Type I experiments, we use only the first two sequences of Table I, Hall and PETS2009. The goal of Type II experiments is, in turn, to compare our algorithm with the state-of-the-art in compressive background subtraction [34], using all sequences in Table I. Here, instead of Gaussian matrices, we use partial DFT matrices, which do not require explicit storage and use the efficient FFT algorithm for matrix-vector multiplication. Our algorithm still works extremely well for DFT matrices, reconstructing video frames with a very small error, but it may require more measurements with respect to Gaussian matrices.

Recall that we address the problem of online compressive background subtraction, not classical background subtraction as in RPCA. In particular, we have no access to full frames, only to a limited number of measurements (decided by the algorithm), which are processed in an online fashion; also, the background image is known. In all experiments, the background image is always the first frame of the sequence, that is, the frame corresponding to the leftmost number in the third column of Table I.

A. *Type I experiments*

The goal of these experiments is to validate Algorithm 1. Hence, we generated measurements matrices as in Theorem 1 and Lemma 2: each entry of $A_k \in \mathbb{R}^{m_k \times n}$ is i.i.d. Gaussian with zero mean and variance $1/m_k$, and A_k and A_j are independent for $j \neq k$.

Experimental setup. As mentioned before, in Type I experiments, we use only the Hall and PETS2009 sequences. We set the oversampling parameters as $\delta := \delta_k = 0.1$, for all k, and the filter parameter as $\alpha = 0.5$. While for the Hall sequence we used the true sparsity of the first two foregrounds, i.e., $\hat{s}_1 = s_1 = 417$ and $\hat{s}_2 = s_2 = 446$, for the PETS2009 sequence we set these parameters to values much smaller than their true values: $10 = \hat{s}_1 \ll s_1 = 194$ and $10 = \hat{s}_2 \ll s_2 = 211$. In spite of this poor initialization, the algorithm was able to quickly adapt, as we will see. We removed isolated pixels from each frame by preprocessing the full sequences. For motion estimation, we used $\gamma \times \gamma = 8 \times 8$ blocks, and a search limit of $\rho = 6$. Finally, we mention that after computing the side information $w[k]$ for frame k, we amplified the magnitude of its components by 30%. This, according to the theory in [36], improves the quality of the side information. To solve BP in the reconstruction of the first two frames we used SPGL1 [70], [71].[10] To solve the ℓ_1-ℓ_1 minimization problem (5) in the reconstruction of the remaining frames we used DECOPT [72], [73].[11]

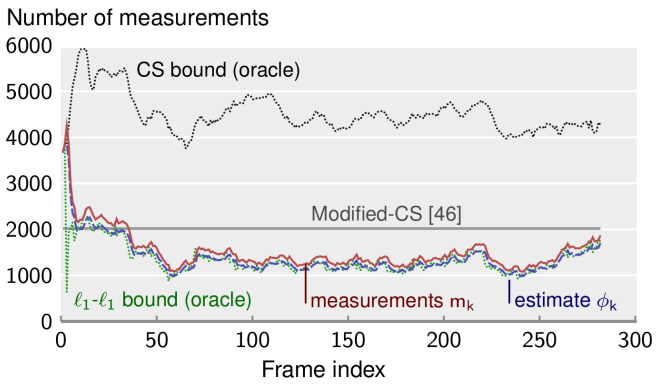
Recall that Algorithm 1 assumes noiseless measurements. To illustrate its extension to noisy measurements, i.e., $y[k] = A_k x[k] + \eta_k$, with $\|\eta_k\|_2 \leq \sigma_k$, we also applied it to the case where the Hall sequence is acquired with noise. Specifically, we generated η_k as a vector of i.i.d. Gaussian entries with zero mean and variance $4/m_k$, and used $\sigma_k = 2$ in all frames. The number of measurements was computed as in (9) with $\tau = 0.1$.

For reference, in the noiseless case of the Hall sequence, we compared our algorithm with time-adapted Modified-CS [47], [48], a method for reconstructing a sparse sequence that uses a fixed number of measurements m. In these experiments, we set $m = 2000$, as it was the smallest (round) number that allowed reconstructing most of the frames without being too wasteful.

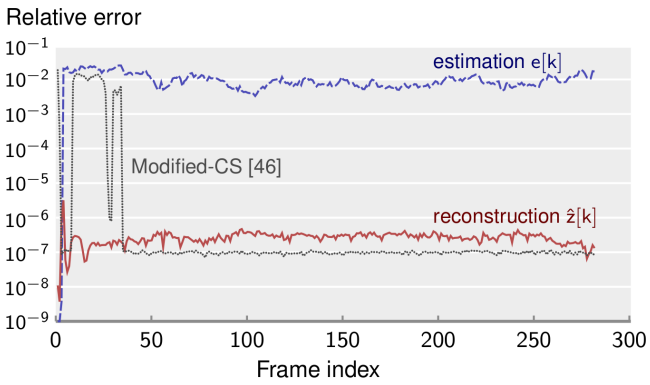
[9] http://groups.inf.ed.ac.uk/vision/CAVIAR/CAVIARDATA1/

[10] Available at http://www.math.ucdavis.edu/~mpf/spgl1/

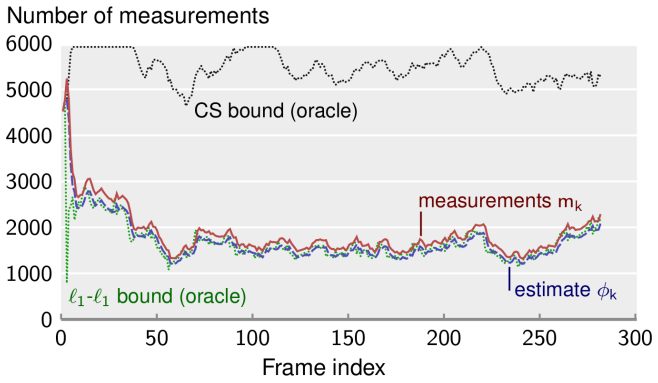
[11] Available at http://lions.epfl.ch/decopt/



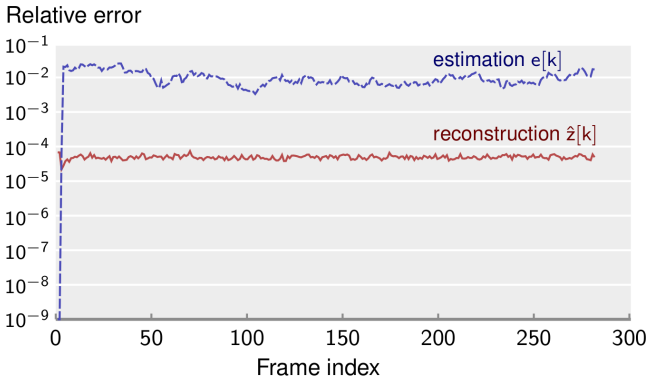
(a) Hall: Number of measurements.



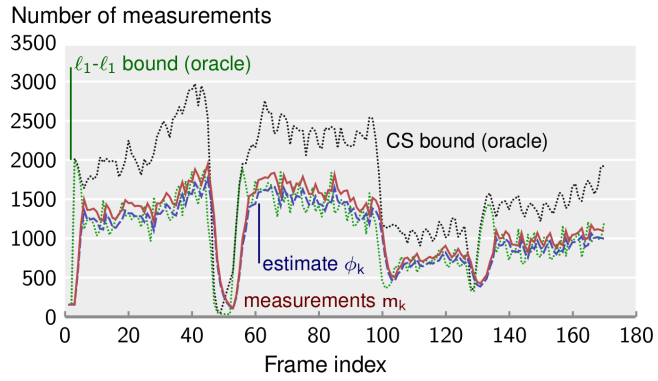
(b) Hall: Relative errors of estimation and reconstruction.



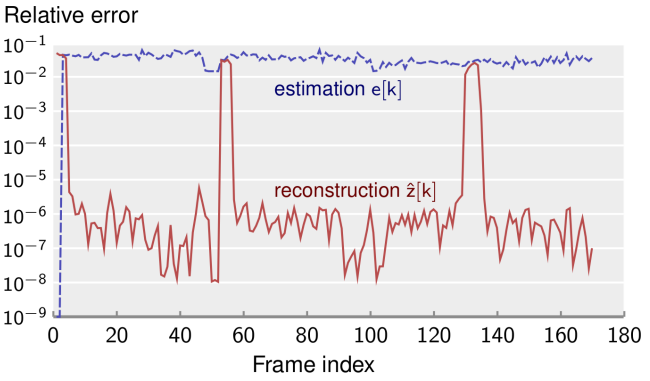
(c) Hall (noisy case): Number of measurements.



(d) Hall (noisy case): Relative errors of estimation and reconstruction.



(e) PETS2009: Number of measurements.



(f) PETS2009: Relative errors of estimation and reconstruction.

Figure 3. Results of Type I experiments. The left-hand side plots show the number of measurements m_k taken from each frame (solid red line), the estimate ϕ_k (dashed blue line), and the right-hand side of (7) and (8) (dotted green and black lines, respectively). The right-hand plots side show the relative errors of estimation $\|e[k]-z[k]\|_2/\|z[k]\|_2$ and reconstruction $\|\hat{z}[k]-z[k]\|_2/\|z[k]\|_2$. (a) and (b) Hall sequence with noiseless acquisition; (c) and (d) Hall sequence with noisy acquisition; (e) and (f) PETS2009 sequence with noiseless acquisition.

Time-adapted Modified-CS [47] reconstructs each frame using the support of the previously reconstructed frame as an aid. In our implementation, the support of a signal was computed as the set of indices containing at least 90% of its energy.

Results. Fig. 3 shows the results of Type I experiments. The plots on the left-hand side show the number of measurements m_k Algorithm 1 took from each frame and the corresponding estimate ϕ_k of (7); the same plots also show the bounds (7) and (8) as if an oracle told us the true values of s_k, $\overline{h}_k$, and ξ_k. The plots on the right-hand side show the relative errors of the estimated image $e[k]$ and the reconstruction image $\hat{z}[k]$, i.e., $\|e[k]-z[k]\|_2/\|z[k]\|_2$ and $\|\hat{z}[k]-z[k]\|_2/\|z[k]\|_2$.

We observe that in all the left-hand side plots, m_k and ϕ_k are always below the CS bound (8), except at a few frames in Fig. 3(e) (PETS2009 sequence). In those frames, there is no foreground and thus the number of required measurements approaches zero. Since there are no such frames in the Hall sequence, all quantities in Figs. 3(a) and 3(c) do not exhibit such large fluctuations. In all the right-hand side plots, the

estimation errors were approximately constant, around 0.01 for the Hall sequence (both in the noisy and noiseless cases) and around 0.93 for the PETS sequence. The reconstruction error varied between 3.8×10^{-9} and 3.5×10^{-6} for the noiseless Hall sequence [Fig. 3(b)] and between 2.4×10^{-5} and 8.3×10^{-5} for the noisy Hall sequence [Fig. 3(d)]. For the PETS sequence [Fig. 3(f)], it was always below 10^{-5} except at three instances, where the reconstruction error approached the estimation error. These correspond to the frames with no foreground [making the bounds in (7) and (8) approach zero] and to the initial frames, where the number of measurements was much smaller than (7). In spite of these "ill-conditioned" frames, our algorithm was able to quickly adapt in the next frames and follow the ℓ_1-ℓ_1 bound curve closely. In Figs. 3(a) and 3(b), we can also see that Modified-CS [47], which took 2000 measurements from each frame, failed to reconstruct the first 34 frames for lack of measurements. Algorithm 1, in contrast, reconstructed all the frames perfectly and used 25% less measurements than Modified-CS did.

These experiments show three key features of Algorithm 1. The first and most important is that the estimate ϕ_k of the bound (7) is always very close to the true value of the bound. The second one is that the algorithm uses significantly less measurements than if we were using BP [16], [34], even if we knew the true foreground sparsity, and than Modified-CS. The third feature is that there is not much difference between the noiseless and noisy cases in terms of the number of measurements; the most noticeable difference is in the reconstruction error: in Fig. 3(d), the reconstruction error is about three orders of magnitude larger than the one in Fig. 3(b).

B. Type II experiments

The goal of Type II experiments is comparing Algorithm 1 (Alg. 1) with the state-of-the-art in adaptive-rate background subtraction, the ARCS-CS algorithm in [34]. For memory and time constraints, we obtained measurements from each frame using partial DFT matrices instead of Gaussian matrices, a modification that, as we will see, causes Alg. 1 to take more measurements without compromising reconstruction.

Experimental setup. In these experiments, we set δ to an even smaller value, 0.01, and the search range ρ to 4. Furthermore, the sparsity of the first two frames, $\hat{s}_1$ and $\hat{s}_2$, was set to 150 in all sequences, independently of their true value. We used a simpler solver for (5), based on ADMM. All the other parameters were the same as in the Type I experiments. We compared Alg. 1 with the algorithm ARCS-CV in [34].[12] Its parameters were set as in [34]: $\tau = 0.1$ and $\sigma_b = 0.0157$. The sparsity estimator $\hat{s}_1$ was initialized with 150 in all sequences (as in Alg. 1), and we modified the code so that its internal solver, SPGL1, had no constraints on the number of iterations. Recall that [34] presented experimental results for DFT matrices as well.

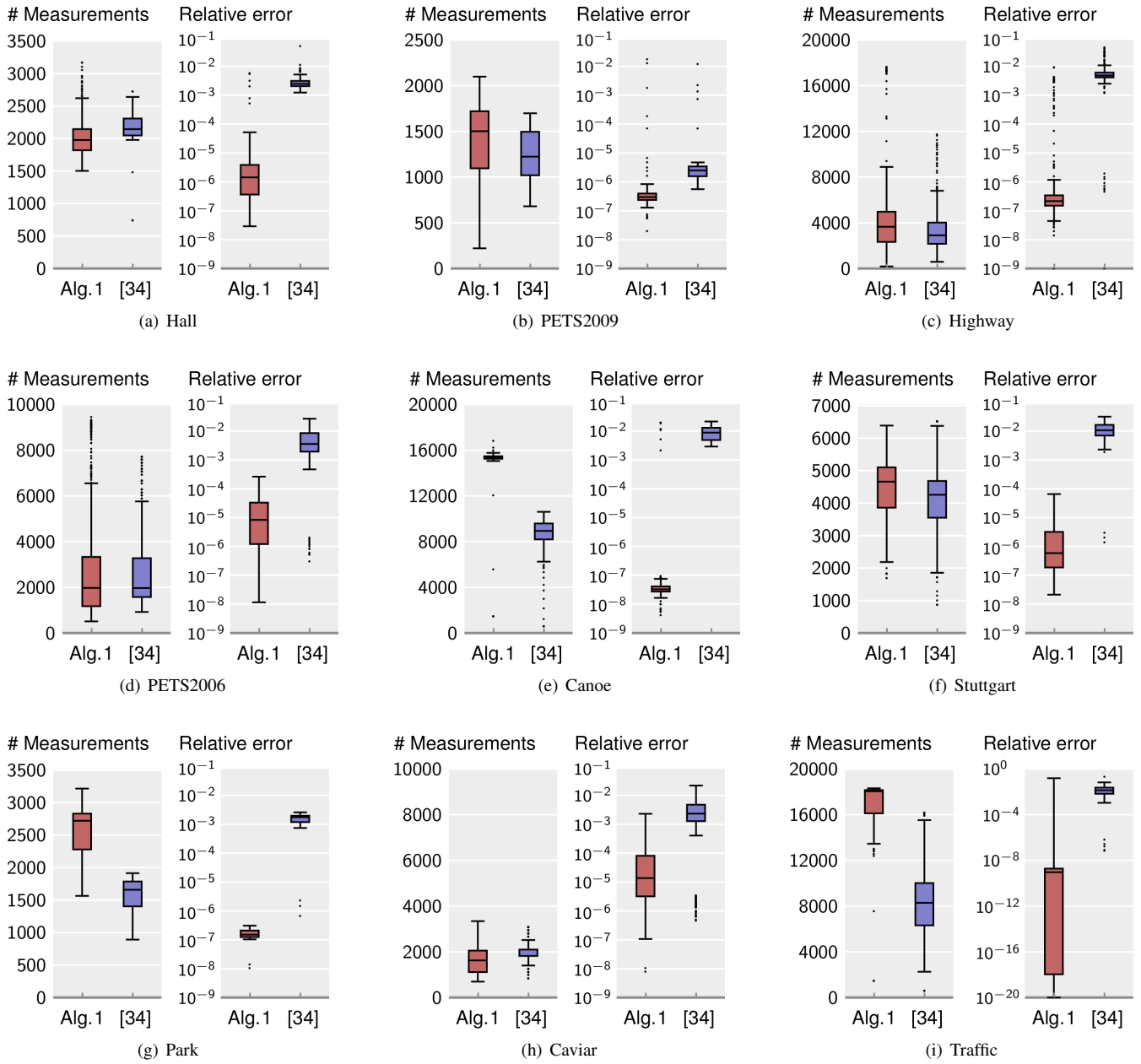
Results. Fig. 4 shows the results of these experiments for all the sequences of Table I. The results are presented as boxplots. In a boxplot, the data points are divided into four equal-sized groups, the quartiles. The median is represented with a horizontal line, and the second and third quartiles are enclosed in a box. The smallest (resp. largest) data point that is at a distance of the lower (resp. upper) quartile smaller than $3/2$ times the box height is marked with the extremity of a "whisker." Data points at farther distances, the outliers, are represented with dots. This type of plot provides a compact representation of data with large variability, which is the case of the data we are presenting: the number of measurements taken by Alg. 1 and ARCS-CV [34], and the respective reconstruction errors. Each subfigure of Fig. 4 contains two plots; the left-hand side plot depicts the number of measurements, the right-hand plot the relative error $\|\hat{z}[k] - z[k]\|_2 / \|z[k]\|_2$. Within each plot, the left boxplot corresponds to Alg. 1 (in red), the right boxplot to ARCS-CV [34] (in blue).

For example, in the Hall sequence [Fig. 4(a)], we see that the number of measurements taken by Alg. 1 had a median of 1976, an increase of 45% with respect to the Gaussian case [Fig. 3(a)]. In spite of this increase, Alg. 1 was slightly more efficient than ARCS-CV [34] in terms of the number of measurements, but much more efficient in terms of the reconstruction error (right-hand side plot of Fig. 4(a)). In fact, except for the PETS2009 sequence [Fig. 4(b)], the median of the reconstruction error of ARCS-CV was always larger than 10^{-3}. Given that the internal solver of ARCS-CV, SPGL1 [70], [71], is a high-precision solver, this indicates that the number of measurements taken by ARCS-CV was not enough for reconstruction, in most frames and for all sequences (except PETS2009). Taking this into account, it is then surprising that Alg. 1 required less measurements than ARCS-CV for some sequences. For example, the median of the number of measurements of Alg. 1 was smaller than the median of ARCS-CV in the Hall [Fig. 4(a)], PETS2006 [Fig. 4(d)], and Caviar [Fig. 4(h)] sequences; note that for these sequences the reconstruction error of Alg 1 was also much smaller. In the remaining sequences, Alg. 1 took more measurements than ARCS-CV, but its reconstruction error was significantly smaller than the reconstruction error of ARCS-CV. We remark that the Traffic sequence exhibits jittering, causing some frames to have practically no background: all pixels comprise the foreground. As a consequence, Alg. 1 required $n = 19200$ measurements from most frames [Fig. 4(i)], that is, as many measurements as the image dimension. This, in turn, caused the reconstruction error to be extremely small. Overall, these experiments show that Alg. 1 outperforms the prior state-of-the-art algorithm ARCS-CV in [34] in terms of the number of measurements for perfect reconstruction. Still, despite the fact that the background is not constant or that the motion is rapid, our algorithm performs extremely well, often surpassing the state of the art in terms of number of measurements, while always reconstructing with a smaller error.

VII. Conclusions

We proposed and analyzed an online algorithm for reconstructing sparse sequences of signals from a limited number of measurements. The signals vary across time according to a

[12] We used the implementation in http://garrettwarnell.com/ARCS-1.0.zip.



(a) Hall (b) PETS2009 (c) Highway

(d) PETS2006 (e) Canoe (f) Stuttgart

(g) Park (h) Caviar (i) Traffic

Figure 4. Results of Type II experiments. Each subfigure corresponds to a line of Table I and contains two boxplots: one for the number of measurements and another for the reconstruction error. In a boxplot, the horizontal line is the median, the box extremes are the Q_1 and Q_3 quartiles, and the whiskers are the minimum and maximum of data that is within $3/2$ of the interquartile range of Q_1 and Q_3; points outside this range (outliers) are represented with dots.

nonlinear dynamical model, and the measurements are linear. Our algorithm is based on ℓ_1-ℓ_1 minimization and, assuming Gaussian measurement matrices, it estimates the required number of measurements to perfectly reconstruct each signal, automatically and on-the-fly. We also explored the application of our algorithm to compressive video background subtraction and tested its performance on sequences of real images. Our experiments show that it also works for DFT matrices and that it reduces the number of required measurements with respect to prior compressive video background subtraction schemes by a large margin. Interesting research directions include extending the algorithm to handle more complex priors, such as block sparsity and spatial foreground contiguity, adapting the algorithm to perform background subtraction in uncompressed videos, and exploring applications in medical imaging.

Appendix A
Proof of Lemma 2

First, note that condition (10) is a function only of the parameters of the sequences $\{x[k]\}$ and $\{\epsilon[k]\}$ and, therefore, is a deterministic condition. Simple algebraic manipulation shows that it is equivalent to

$$(1+\delta_i)\left[2\overline{h}_{i-1}\log\left(\frac{n}{u_{i-1}}\right)+\frac{7}{5}u_{i-1}+1\right]\geq 2\overline{h}_i\log\left(\frac{n}{u_i}\right)$$

$$+\frac{7}{5}u_i+1\,,$$

or

$$(1+\delta_i)\overline{m}_{i-1}\geq\overline{m}_i\,, \tag{20}$$

where $\overline{m}_i$ is the right-hand side of (7) applied to $x[i]$, that is,

$$\overline{m}_i:=2\overline{h}_i\log\Big(\frac{n}{s_i+\xi_i/2}\Big)+\frac{7}{5}\Big(s_i+\frac{\xi_i}{2}\Big)+1\,. \tag{21}$$

Notice that the source of randomness in Algorithm 1 is the set of matrices (random variables) A_k, generated in steps 3 and 12. Define the event S_i as "perfect reconstruction at time i." Since we assume that $\hat{s}_1$ and $\hat{s}_2$ are larger than the true sparsity of $x[1]$ and $x[2]$, there holds [57]

$$\begin{aligned}\mathbb{P}(S_i)&\geq 1-\exp\Big[-\frac{1}{2}(m_i-\sqrt{m_i})^2\Big]\\&\geq 1-\exp\Big[-\frac{1}{2}(\underline{m}-\sqrt{\underline{m}})^2\Big]\,,\end{aligned} \tag{22}$$

for $i=1,2$, where the second inequality is due to $m_i\geq\underline{m}$ and $1-\exp(-(1/2)(x-\sqrt{x})^2)$ being an increasing function.

Next, we compute the probability of the event "perfect reconstruction at time i" given that there was "perfect reconstruction at all previous time instants $l<i$," i.e., $\mathbb{P}(S_i|\bigwedge_{l<i}S_l)$, for all $i=3,\dots,k$. Since we assume $\alpha=1$, we have $\phi_i=\hat{\overline{m}}_{i-1}$ and step 11 of Algorithm 1 becomes $m_i=(1+\delta_i)\hat{\overline{m}}_{i-1}$, for all $i\geq 3$. Under the event S_{i-1}, i.e., $\hat{x}[i-1]=x[i-1]$, we have $\hat{\overline{h}}_{i-1}=\overline{h}_{i-1}$, $\hat{\xi}_{i-1}=\xi_{i-1}$, and $\hat{\overline{m}}_{i-1}=\overline{m}_{i-1}$, where $\overline{m}_{i-1}$ is defined in (21). (The hat variables are random variables.) Consequently, due to our assumption (20), step 11 can be written as $m_i=(1+\delta_i)\hat{\overline{m}}_{i-1}=(1+\delta_i)\overline{m}_{i-1}\geq\overline{m}_i$. This means (7) is satisfied. By assumption, all the other conditions of Theorem 1 are satisfied and, hence, for $i\geq 3$,

$$\begin{aligned}\mathbb{P}\Big(S_i\,\Big|\bigwedge_{l<i}S_l\Big)&\geq 1-\exp\Big[-\frac{1}{2}(m_i-\sqrt{m_i})^2\Big]\\&\geq 1-\exp\Big[-\frac{1}{2}(\underline{m}-\sqrt{\underline{m}})^2\Big]\,,\end{aligned} \tag{23}$$

where, again, we used the fact that $m_i\geq\underline{m}$ and that $1-\exp(-(1/2)(x-\sqrt{x})^2)$ is an increasing function.

Finally, we bound the probability that there is perfect reconstruction at all time instants $1\leq i\leq k$:

$$\begin{aligned}&\mathbb{P}(S_1\wedge S_2\wedge\cdots\wedge S_k)\\&\quad=\mathbb{P}(S_1)\mathbb{P}(S_2|S_1)\prod_{i=3}^{k}\mathbb{P}(S_i|S_1\wedge\cdots\wedge S_{i-1}) &(24)\\&\quad=\mathbb{P}(S_1)\mathbb{P}(S_2)\prod_{i=3}^{k}\mathbb{P}\Big(S_i\,\Big|\bigwedge_{l<i}S_l\Big) &(25)\\&\quad\geq\left(1-\exp\Big[-\frac{1}{2}(\underline{m}-\sqrt{\underline{m}})^2\Big]\right)^k. &(26)\end{aligned}$$

From (24) to (25) we used the independence between S_1 and S_2. From (25) to (26), we used (22) and (23). □

Appendix B
Proof of Theorem 3

Recall the definitions of ξ and $\overline{h}$ in (6):

$$\begin{aligned}\xi&=\big|\{j\,:\,w_j\neq x_j^\star=0\}\big|-\big|\{j\,:\,w_j=x_j^\star\neq 0\}\big|\\\overline{h}&=\big|\{j\,:\,x_j^\star>0,\ \epsilon_j>0\}\cup\{j\,:\,x_j^\star<0,\epsilon_j<0\}\big|\,,\end{aligned}$$

where we rewrote $\overline{h}$ using $x^\star=w+\epsilon$. Define the events $\mathcal{A}:=$ "$\exists_j\,:\,x_j=w_j=0$", $\mathcal{B}:=$ "$\overline{h}>0$", and

$$\mathcal{C}:=\text{"}m\geq 2\overline{h}\log\Big(\frac{n}{s+\xi/2}\Big)+\frac{7}{5}\Big(s+\frac{\xi}{2}\Big)+1\,\text{,"}$$

which are the assumptions of Theorem 1. In $\mathcal{C}$, m and n are deterministic, whereas s, $\overline{h}$, and ξ are random variables. Then,

$$\begin{aligned}\mathbb{P}\big(\hat{x}=x^\star\big)&\geq\mathbb{P}\Big(\hat{x}=x^\star\,\Big|\,\mathcal{A}\wedge\mathcal{B}\wedge\mathcal{C}\Big)\cdot\mathbb{P}\big(\mathcal{A}\wedge\mathcal{B}\wedge\mathcal{C}\big)\\&\geq\left[1-\exp\Big(-\frac{(m-\sqrt{m})^2}{2}\Big)\right]\cdot\mathbb{P}\big(\mathcal{A}\wedge\mathcal{B}\wedge\mathcal{C}\big)\,,\end{aligned} \tag{27}$$

where we used Theorem 1. The rest of the proof consists of lower bounding $\mathbb{P}\big(\mathcal{A}\wedge\mathcal{B}\wedge\mathcal{C}\big)$.

Lower bound on $\mathbb{P}\big(\mathcal{A}\wedge\mathcal{B}\wedge\mathcal{C}\big)$. Recall that w is fixed and that each component $x_j^\star$ is determined by $x_j^\star=w_j+\epsilon_j$. Due to the continuity of the distribution of ϵ, with probability 1, no component $j\in\Sigma$ (i.e., $\sigma_j^2\neq 0$) contributes to ξ. When $j\in\Sigma^c$, we have two cases:

- $j\in\Sigma^c\cap\mathcal{W}$ (i.e., $\sigma_j^2=0$ and $w_j\neq 0$): in this case, we have $x_j^\star=w_j$ with probability 1. Hence, these components contribute to the second term of ξ.
- $j\in\Sigma^c\cap\mathcal{W}^c$ (i.e., $\sigma_j^2=0$ and $w_j=0$): in this case, we also have $x_j^\star=w_j$ with probability 1. However, these components do not contribute to ξ.

We conclude $\mathbb{P}\big(\mathcal{D}\big)=\mathbb{P}\big(\xi=-\big|\Sigma^c\cap\mathcal{W}\big|\big)=1$, where $\mathcal{D}$ is the event "$\xi=-\big|\Sigma^c\cap\mathcal{W}\big|$." From the second case above we also conclude that our assumption $\Sigma^c\cap\mathcal{W}^c\neq\emptyset$ implies $\mathbb{P}\big(\mathcal{A}\big)=1$. We can then write

$$\begin{aligned}\mathbb{P}\big(\mathcal{A}\wedge\mathcal{B}\wedge\mathcal{C}\big)&=\mathbb{P}\big(\mathcal{A}\big)\cdot\mathbb{P}\big(\mathcal{B}\wedge\mathcal{C}\,\big|\,\mathcal{A}\big)=\mathbb{P}\big(\mathcal{B}\wedge\mathcal{C}\,\big|\,\mathcal{A}\big)\\&\geq\mathbb{P}\big(\mathcal{B}\wedge\mathcal{C}\,\big|\,\mathcal{A},\,\mathcal{D}\big)\cdot\mathbb{P}\big(\mathcal{D}\,\big|\,\mathcal{A}\big) &(28)\\&=\mathbb{P}\big(\mathcal{B}\wedge\mathcal{C}\,\big|\,\mathcal{A},\,\mathcal{D}\big)\cdot\mathbb{P}\big(\mathcal{D}\big) &(29)\\&=\mathbb{P}\big(\mathcal{B}\wedge\mathcal{C}\,\big|\,\mathcal{A},\,\mathcal{D}\big) &(30)\\&=\mathbb{P}\big(0<\overline{h}\leq\mu+t\,\big|\,\mathcal{A},\,\mathcal{D}\big)\,. &(31)\end{aligned}$$

From (28) to (29), we used the fact that the events $\mathcal{A}=$ "$\Sigma^c\cap\mathcal{W}^c\neq\emptyset$" and $\mathcal{D}=$ "$\xi=-|\Sigma^c\cap\mathcal{W}|$" are independent. This follows from the independence of the components of ϵ and the disjointness of $\Sigma^c\cap\mathcal{W}^c$ and $\Sigma^c\cap\mathcal{W}$. From (30) to (31), we used the fact that event $\mathcal{C}$ conditioned on $\mathcal{D}$ is equivalent to $\overline{h}\leq\mu+t$. To see why, note that the sparsity of $x^\star$ is given by $s=|\Sigma|+|\Sigma^c\cap\mathcal{W}|$; thus, given $\mathcal{D}$, $s+\xi/2$ equals $|\Sigma|+|\Sigma^c\cap\mathcal{W}|/2$; now, subtract the expression in assumption (16) from the expression that defines event $\mathcal{C}$:

$$0\geq 2(\overline{h}-\mu-t)\log\left[\frac{n}{|\Sigma|+\frac{1}{2}|\Sigma^c\cap\mathcal{W}|}\right].$$

Using the fact that $n=|\Sigma|+|\Sigma^c|\geq|\Sigma|+|\Sigma^c\cap\mathcal{W}|\geq|\Sigma|+|\Sigma^c\cap\mathcal{W}|/2$, we conclude that $\mathcal{C}$ is equivalent to the event "$\overline{h}\leq\mu+t$." We now bound (31) as follows:

$$\mathbb{P}\big(0<\overline{h}\leq\mu+t\,\big|\,\mathcal{A},\,\mathcal{D}\big)$$

$$\geq \mathbb{P}\big(0 < \overline{h} < \mu + t - 1 \,\big|\, \mathcal{A},\, \mathcal{D}\big)$$
$$= 1 - \mathbb{P}\big(\overline{h} \leq 0 \,\big|\, \mathcal{A},\, \mathcal{D}\big) - \mathbb{P}\big(\overline{h} \geq \mu + t - 1 \,\big|\, \mathcal{A},\, \mathcal{D}\big)$$
$$= 1 - \mathbb{P}\big(\overline{h} - \mu \leq -\mu \,\big|\, \mathcal{A},\, \mathcal{D}\big) - \mathbb{P}\big(\overline{h} - \mu \geq t - 1 \,\big|\, \mathcal{A},\, \mathcal{D}\big) \tag{32}$$
$$\geq 1 - \exp\Big[-\frac{2\mu^2}{|\Sigma|}\Big] - \exp\Big[-\frac{2(t-1)^2}{|\Sigma|}\Big], \tag{33}$$

where the last step, explained below, is due to Hoeffding's inequality [74]. Note that once this step is proven, (33) together with (27) and (31) give (17), proving the theorem.

Proof of step (32)**-**(33)**.** Hoeffding's inequality states that if $\{Z_j\}_{j=1}^L$ is a sequence of independent random variables and $\mathbb{P}(0 \leq Z_j \leq 1) = 1$ for all j, then [74, Th.4]:

$$\mathbb{P}\Big(\sum_{j=1}^{L} Z_j - \sum_{j=1}^{L} \mathbb{E}\big[Z_j\big] \geq \tau\Big) \leq \exp\Big[-\frac{2\tau^2}{L}\Big] \tag{34}$$
$$\mathbb{P}\Big(\sum_{j=1}^{L} Z_j - \sum_{j=1}^{L} \mathbb{E}\big[Z_j\big] \leq -\tau\Big) \leq \exp\Big[-\frac{2\tau^2}{L}\Big], \tag{35}$$

for any $\tau > 0$. We apply (35) to the second term in (32) and (34) to the third term. This is done by showing that $\overline{h}$ is the sum of $|\Sigma|$ independent random variables, taking values in $[0, 1]$ with probability 1, and whose expected values sum to μ. Note that $\mu > 0$ by definition, and $t > 1$ by assumption.

We start by noticing that the components of ϵ that contribute to $\overline{h}$ are the ones for which $\sigma_j^2 \neq 0$, i.e., $j \in \Sigma$ (otherwise, $\epsilon_j = 0$ with probability 1). Using $x_j^\star = w_j + \epsilon_j$ [cf. (1a)], we have $\overline{h} = \sum_{j\in\Sigma} Z_j$, where Z_j is the indicator of the event

$$\epsilon_j > \max\big\{0,\, -w_j\big\} \quad \text{or} \quad \epsilon_j < \min\big\{0,\, -w_j\big\}, \tag{36}$$

that is, $Z_j = 1$ if (36) holds, and $Z_j = 0$ otherwise. By construction, $0 \leq Z_j \leq 1$ for all j. Furthermore, because the components of ϵ are independent, so are the random variables Z_j. All we have left to do is to show that the sum of the expected values of Z_j conditioned on $\mathcal{A}$ and $\mathcal{D}$ equals μ. This involves just simple integration. Let $j \in \Sigma$. Then,

$$\mathbb{E}\Big[Z_j \,\Big|\, \mathcal{A},\, \mathcal{D}\Big] = \mathbb{P}\big(Z_j = 1 \,\big|\, \mathcal{A},\, \mathcal{D}\big) \tag{37}$$
$$= \mathbb{P}\big(\epsilon_j > \max\big\{0,\, -w_j\big\}\big) + \mathbb{P}\big(\epsilon_j < \min\big\{0,\, -w_j\big\}\big) \tag{38}$$
$$= \frac{1 + \exp\big(-\lambda_j |w_j|\big)}{2} \tag{39}$$
$$= \frac{1 + \exp\big(-\sqrt{2}|w_j|/\sigma_j\big)}{2}\,. \tag{40}$$

From (37) to (38), we used the fact that the events in (36) are disjoint for any w_j. From (38) to (39), we used the fact that λ_j is finite for $j \in \Sigma$, and

$$\mathbb{P}\big(\epsilon_j > \max\big\{0,\, -w_j\big\}\big) = \int_{\max\{-w_j,0\}}^{+\infty} \frac{\lambda_j}{2} \exp(-\lambda_j |u|)\, du$$
$$= \begin{cases} \frac{1}{2} & ,\ w_j > 0 \\ \frac{1}{2}\exp\big(\lambda_j w_j\big) & ,\ w_j < 0 \end{cases}$$
$$\mathbb{P}\big(\epsilon_j < \min\big\{0,\, -w_j\big\}\big) = \int_{-\infty}^{\min\{-w_j,0\}} \frac{\lambda_j}{2} \exp(\lambda_j |u|)\, du$$
$$= \begin{cases} \frac{1}{2}\exp\big(-\lambda_i w_j\big) & ,\ w_j > 0 \\ \frac{1}{2} & ,\ w_j < 0\,. \end{cases}$$

And from (39) to (40) we simply replaced $\lambda_j = \sqrt{2}/\sigma_j$. The expected value of $\overline{h}$ conditioned on $\mathcal{A}$ and $\mathcal{D}$ is then

$$\mathbb{E}\Big[\overline{h} \,\Big|\, \mathcal{A},\, \mathcal{D}\Big] = \mathbb{E}\Big[\sum_{j\in\Sigma} Z_j \,\Big|\, \mathcal{A},\, \mathcal{D}\Big] = \sum_{j\in\Sigma} \mathbb{E}\Big[Z_j \,\Big|\, \mathcal{A},\, \mathcal{D}\Big]$$
$$= \frac{1}{2} \sum_{j\in\Sigma} \Big[1 + \exp\big(-\sqrt{2}|w_j|/\sigma_j\big)\Big] =: \mu\,,$$

where we used (40).

□

REFERENCES

[1] J. Mota, N. Deligiannis, A. C. Sankaranarayanan, V. Cevher, and M. Rodrigues, "Dynamic sparse state estimation using ℓ_1-ℓ_1 minimization: Adaptive-rate measurement bounds, algorithms, and applications," in *IEEE Intern. Conf. Acoustics, Speech, and Sig. Proc. (ICASSP)*, 2015, pp. 3332–3336.

[2] R. Kalman and R. Bucy, "New results in linear filtering and prediction theory," *Trans. ASME-J. Basic Engineering*, vol. 83, no. 1, pp. 95–108, 1961.

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

[4] ——, "Analyzing least squares and Kalman filtered compressed sensing," in *IEEE Intern. Conf. Acoustics, Speech, and Sig. Proc. (ICASSP)*, 2009, pp. 3013–3016.

[5] ——, "LS-CS-Residual (LS-CS): Compressive sensing on least squares residual," *IEEE Trans. Sig. Proc.*, vol. 58, no. 8, pp. 4108–4120, 2010.

[6] L. Balzano, R. Nowak, and B. Recht, "Online identification and tracking of subspaces from highly incomplete information," in *Allerton Conf. Communications, Control, and Computing*, 2010, pp. 704–711.

[7] L. Balzano and S. J. Wright, "Local convergence of an algorithm for subspace identification from partial data," *Found. Computational Mathematics*, pp. 1–36, 2014.

[8] Y. Chi, Y. C. Eldar, and R. Calderbank, "PETRELS: Parallel subspace estimation and tracking by recursive least squares from partial observations," *IEEE Trans. Sig. Proc.*, vol. 61, no. 23, pp. 5947–5959, 2013.

[9] 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.

[10] D. A. Forsyth and J. Ponce, *Computer Vision: A Modern Approach*. Prentice Hall, 2002.

[11] R. Chellappa, A. C. Sankaranarayanan, A. Veeraraghavan, and P. Turaga, "Statistical methods and models for video-based tracking, modeling, and recognition," *Foundations and Trends in Signal Processing*, vol. 3, no. 1-2, pp. 1–151, 2009.

[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] U. Gamper and S. Boesiger, P. Kozerke, "Compressed sensing in dynamic MRI," *Magnetic Resonance in Medicine*, vol. 59, no. 2, pp. 365–373, 2008.

[14] L. Weizman, Y. C. Eldar, and D. Ben-Bashat, "Compressed sensing for longitudinal MRI: An adaptive-weighted approach," *Medical Physics*, vol. 42, no. 9, pp. 5195–5207, 2015.

[15] A. Ribeiro, I. D. Schizas, S. I. Roumeliotis, and G. B. Giannakis, "Kalman filtering in wireless sensor networks," *IEEE Control Syst. Mag.*, vol. 30, no. 2, pp. 66–86, 2010.

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

[17] L. Maddalena and A. Petrosino, "A self-organizing approach to background subtraction for visual surveillance applications," *IEEE Trans. Image Processing*, vol. 17, no. 7, pp. 1168–1177, 2008.

[18] S. Brutzer, B. Höferlin, and G. Heidemann, "Evaluation of background subtraction techniques for video surveillance," in *IEEE Conf. Comp. Vision and Pattern Recognition (CVPR)*, 2011, pp. 1937–1944.

[19] B. Tseng, C.-Y. Lin, and J. Smith, "Real-time video surveillance for traffic monitoring using virtual line analysis," in *IEEE Inter. Conf. Multimedia and Expo*, 2002, pp. 541–544.

[20] S. Cheung and C. Kamath, "Robust techniques for background subtraction in urban traffic video," in *Symposium on Electronic Imaging (SPIE)*, 2003, pp. 881–892.

[21] A. Profio, O. Balchum, and F. Carstens, "Digital background subtraction for fluorescence imaging," *Med. Phys.*, vol. 13, no. 5, pp. 717–721, 1986.

[22] R. Otazo, E. Candès, and D. Sodickson, "Low-rank plus sparse matrix decomposition for accelerated dynamic MRI with separation of background and dynamic components," *Magnetic Resonance in Medicine*, vol. 73, pp. 1125–1136, 2015.

[23] M. Piccardi, "Background subtraction techniques: a review," in *IEEE Inter. Conf. Systems, Man and Cybernetics*, 2004, pp. 3099–3104.

[24] E. Candès, X. Li, Y. Ma, and J. Wright, "Robust principal component analysis?" *Journal of the ACM*, vol. 58, no. 3, pp. 11:1–11:37, 2011.

[25] M. Wakin, J. Laska, M. Duarte, D. Baron, S. Sarvotham, D. Takhar, K. Kelly, and R. Baraniuk, "Compressive imaging for video representation and coding," in *Picture Coding Symposium*, 2006.

[26] M. Duarte, M. Davenport, D. Takhar, J. Laska, T. Sun, K. Kelly, and R. Baraniuk, "Single-pixel imaging via compressive sampling," *IEEE Sig. Proc. Mag.*, vol. 25, no. 2, pp. 83–91, 2008.

[27] 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.

[28] A. Sankaranarayanan, C. Studer, and R. Baraniuk, "CS-MUVI: video compressive sensing for spatial multiplexing cameras," in *Intern. Conf. Computation Photography*, 2012, pp. 1–10.

[29] A. C. Sankaranarayanan, L. Xu, C. Studer, Y. Li, K. Kelly, and R. G. Baraniuk, "Video compressive sensing for spatial multiplexing cameras using motion-flow models," 2014, under review at SIAM J. Imaging Sciences.

[30] D. Reddy, A. Veeraraghavan, and R. Chellappa, "P2C2: programmable pixel compressive camera for high speed imaging," in *IEEE Conf. Comp. Vision and Pattern Recognition (CVPR)*, 2011, pp. 329–336.

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

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

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

[34] G. Warnell, S. Bhattacharya, R. Chellappa, and T. Basar, "Adaptive-rate compressive sensing using side information," 2014, preprint: http://arxiv.org/abs/1401.0583.

[35] J. Park and M. Wakin, "A multiscale framework for compressive sensing of video," in *Picture Coding Symposium*, 2009, pp. 1–4.

[36] J. Mota, N. Deligiannis, and M. Rodrigues, "Compressed sensing with prior information: Optimal strategies, geometry, and bounds," 2014, preprint: http://arxiv.org/abs/1408.5250.

[37] ——, "Compressed sensing with side information: Geometrical interpretation and performance bounds," in *IEEE Global Conf. Sig. and Inf. Proc. (GlobalSIP)*, 2014, pp. 675–679.

[38] S. Chen, D. Donoho, and M. Saunders, "Atomic decomposition by basis pursuit," *SIAM J. Sci. Comp.*, vol. 20, no. 1, pp. 33–61, 1998.

[39] R. Berinde and P. Indyk, "Sparse recovery using sparse matrices," MIT-CSAIL-TR-2008-001, Tech. Rep., 2008.

[40] A. Liutkis, D. Martina, S. Popoff, G. Chardon, O. Katz, G. Lerosey, S. Gigan, L. Daudet, and I. Carron, "Imaging with nature: Compressive imaging using a multiply scattering medium," *Nature*, vol. 4, no. 5552, pp. 1–7, 2014.

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

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

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

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

[45] N. Vaswani, "Stability (over time) of Modified-CS for recursive causal sparse reconstruction," in *Allerton Conf. Communications, Control, and Computing*, 2010, pp. 1722–1729.

[46] J. Zhan and N. Vaswani, "Robust PCA with partial subspace knowledge," in *IEEE Intern. Symposium on Information Theory (ISIT)*, 2014, pp. 2192–2196.

[47] ——, "Time invariant error bounds for modified-cs-based sparse signal sequence recovery," *IEEE Trans. Inf. Th.*, vol. 61, no. 3, pp. 1389–1409, 2015.

[48] N. Vaswani and W. Lu, "Modified-CS: Modifying compressive sensing for problems with partially known support," *IEEE Trans. Sig. Proc.*, vol. 58, no. 9, pp. 4595–4607, 2010.

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

[50] A. Carmi, P. Gurfil, and D. Kanevsky, "Methods for sparse signal recovery using Kalman filtering with embedded pseudo-measurement norms and quasi-norms," *IEEE Trans. Sig. Proc.*, vol. 58, no. 4, pp. 2405–2409, 2010.

[51] V. Chandrasekaran, S. Sanghavi, P. A. Parrilo, and A. S. Willsky, "Rank-sparsity incoherence for matrix decomposition," *SIAM J. Optim.*, vol. 21, no. 2, pp. 572–596, 2011.

[52] X. Cui, J. Huang, S. Zhang, and D. A. Metaxas, "Background subtraction using low rank and group sparsity constraints," in *European Conf. on Computer Vision (ECCV)*, 2012, pp. 612–625.

[53] J. Wright, A. Ganesh, K. Min, and Y. Ma, "Compressive principal component pursuit," *Information and Inference: A Journal of the IMA*, vol. 2, pp. 32–68, 2013.

[54] J. Feng, H. Xu, and S. Yan, "Online robust pca via stochastic optimization," in *Proceedings of Neural Information Processing Systems Foundation (NIPS)*, 2013.

[55] H. Guo, C. Qiu, and N. Vaswani, "An online algorithm for separating sparse and low-dimensional signal sequences from their sum," *IEEE Trans. Sig. Proc.*, vol. 62, no. 16, pp. 4284–4297, 2014.

[56] C. Qiu, N. Vaswani, B. Lois, and L. Hogben, "Recursive robust PCA or recursive sparse recovery in large but structured noise," *IEEE Trans. Inf. Th.*, vol. 60, no. 8, pp. 5007–5039, 2014.

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

[58] D. Kanevsky, A. Carmi, L. Horesh, P. Gurfil, B. Ramabhadran, and T. Sainath, "Kalman filtering for compressed sensing," in *Conf. Information Fusion (FUSION)*, 2010, pp. 1–8.

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

[60] 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.

[61] L. Natário, C. Brites, J. Ascenso, and F. Pereira, "Extrapolating side information for low-delay pixel-domain distributed video coding," in *Visual Content Processing and Representation*, 2006, pp. 16–21.

[62] T. Wiegand, G. J. Sullivan, G. Bjontegaard, and A. Luthra, "Overview of the H. 264/AVC video coding standard," *IEEE Trans. Circuits Sys. for Video Tech.*, vol. 13, no. 7, pp. 560–576, 2003.

[63] L. Alparone, M. Barni, F. Bartolini, and V. Cappellini, "Adaptively weighted vector-median filters for motion-fields smoothing," in *IEEE Intern. Conf. Acoustics, Speech, and Sig. Proc. (ICASSP)*, 1996, pp. 2267–2270.

[64] N. Deligiannis, A. Munteanu, S. Wang, S. Cheng, and P. Schelkens, "Maximum likelihood Laplacian correlation channel estimation in layered Wyner-Ziv coding," *IEEE Trans. Sig. Proc.*, vol. 62, no. 4, pp. 892–904, 2014.

[65] X. Fan, O. C. Au, and N. M. Cheung, "Transform-domain adaptive correlation estimation (TRACE) for Wyner-Ziv video coding," *IEEE Trans. Circuits Sys. for Video Tech.*, vol. 20, no. 11, pp. 1423–1436, 2010.

[66] P. Seeling and M. Reisslein, "Video traffic characteristics of modern encoding standards: H.264/AVC with SVC and MVC extensions and H.265/HEVC," *The Scientific World Journal*, vol. 2014, pp. 1–16, 2014.

[67] N. Goyette, P.-M. Jodoin, F. Porikli, J. Konrad, and P. Ishwar, "changedetection.net: A new change detection benchmark dataset," in *IEEE Conf. Comp. Vision and Pattern Recognition (CVPR)*, 2012, pp. 16–21.

[68] S. Bruzer, B. Höferlin, and G. Heidemann, "Evaluation of background subtraction techniques for video surveillance," in *IEEE Conf. Comp. Vision and Pattern Recognition (CVPR)*, 2011, pp. 1937–1944.

[69] R. Fisher, J. Santos-Victor, and J. Crowley, "CAVIAR: Context aware vision using image-based active recognition," 2005, EC Funded CAVIAR project/IST 2001 37540; link: http://groups.inf.ed.ac.uk/vision/CAVIAR/CAVIARDATA2/OneShopOneWait1front/OneShopOneWait1front.tar.gz.

[70] E. van den Berg and M. Friedlander, "Probing the Pareto frontier for basis pursuit solutions," *SIAM J. Sci. Comput.*, vol. 31, no. 2, pp. 890–912, 2008.

[71] ——, "Sparse optimization with least-squares constraints," *SIAM J. Optim.*, vol. 21, no. 4, pp. 1201–1229, 2011.

[72] Q. Tran-Dinh and V. Cevher, "Constrained convex minimization via model-based excessive gap," in *Proceedings of Neural Information Processing Systems Foundation (NIPS)*, 2014.
[73] ——, "A primal-dual algorithmic framework for constrained convex minimization," 2014, preprint: http://arxiv.org/abs/1406.5403.
[74] G. Lugosi, "Concentration-of-measure inequalities," 2009, lecture notes.

João F. C. Mota received the B.Sc. and M.Sc. degrees in Electrical and Computer Engineering from Instituto Superior Técnico, Technical University of Lisbon, in 2008, and the Ph.D. degree in Electrical and Computer Engineering from both Instituto Superior Técnico, Technical University of Lisbon, and Carnegie Mellon University, PA, in 2013. He is currently a Senior Research Associate in the Department of Electronic and Electrical Engineering at University College London, UK. His research interests include signal processing, optimization theory, machine learning, data science, and distributed information processing and control. Dr. Mota received the 2015 IEEE Signal Processing Society Young Author Best Paper Award.

Nikos Deligiannis (S'08, M'10) is assistant professor with the Electronics and Informatics department at Vrije Universiteit Brussel. He received the Diploma in electrical and computer engineering from University of Patras, Greece, in 2006, and the PhD in applied sciences (awarded with Highest Honors and congratulations from the Jury) from Vrije Universiteit Brussel, Belgium, in 2012. From Jun. 2012 to Sep. 2013, he was a postdoctoral researcher with the Department of Electronics and Informatics, Vrije Universiteit Brussel. From Oct. 2013 to Feb. 2015, he was working as postdoctoral research associate at the Department of Electronic and Electrical Engineering at University College London, UK. During that period, he was also acting as technical consultant on big visual data technologies at the British Academy of Film and Television Arts (BAFTA), UK. His current research interests include big data processing, compressed sensing, ad hoc networking for the internet-of-things, distributed computing, and visual search. Prof. Deligiannis has authored over 60 journal and conference publications, book chapters, and two patent applications (one owned by iMinds, Belgium and the other by BAFTA, UK). He was a recipient of the 2011 ACM/IEEE International Conference on Distributed Smart Cameras Best Paper Award and the 2013 Scientific Prize FWO-IBM Belgium.

Aswin Sankaranarayanan is an Assistant Professor at Carnegie Mellon University (CMU). He received his Ph.D. from University of Maryland, College Park where he was awarded the distinguished dissertation fellowship for his thesis work by the ECE department in 2009. He was a post-doctoral researcher at the DSP group at Rice University before joining the faculty at the ECE Department at CMU. Aswin is currently the PI of the Image Science Lab at CMU, and his research consists of a diverse portfolio including imaging, machine vision, and image processing. He has received best paper awards at the CVPR Workshop on Computational Cameras and Displays (2015) and Analysis and Modeling of Faces and Gestures (2010).

Volkan Cevher (SM'10) received a B.Sc. (valedictorian) in Electrical Engineering in 1999 from Bilkent University in Ankara, Turkey, and he received a Ph.D. in Electrical and Computer Engineering in 2005 from the Georgia Institute of Technology in Atlanta, GA. He held research scientist positions at the University of Maryland, College Park from 2006-2007 and at Rice University in Houston, TX, from 2008-2009. Currently, he is an Associate Professor at the Swiss Federal Institute of Technology Lausanne and a Faculty Fellow in the Electrical and Computer Engineering Department at Rice University. His research interests include signal processing theory, machine learning, graphical models, and information theory. Dr. Cevher received a Best Paper Award at SPARS in 2009, a Best Paper Award at CAMSAP in 2015, and an ERC StG in 2011.

Miguel Rodrigues (S'98, M'02) received the Licenciatura degree in Electrical Engineering from the Faculty of Engineering of the University of Porto, Portugal in 1998 and the Ph.D. degree in Electronic and Electrical Engineering from University College London, UK in 2002. He carried out postdoctoral research work both at Cambridge University, UK, as well as Princeton University, USA, in the period 2003 to 2007.

Dr. Rodrigues is a Senior Lecturer with the Department of Electronic and Electrical Engineering, University College London, U.K. He was previously with the Department of Computer Science, University of Porto, Portugal, where he also led the Information Theory and Communications Research Group at Instituto de Telecomunicações Porto. He has also held visiting research appointments at Princeton University, USA, Duke University, USA, Cambridge University, UK, and University College London, UK, in the period 2007 to 2013. His research interests are in the general areas of information theory, communications theory and signal processing.

Dr. Rodrigues was the recipient of the IEEE Communications and Information Theory Societies Joint Paper Award in 2011 for the work on Wireless Information-Theoretic Security (with M. Bloch, J. Barros and S. W. McLaughlin). He was also the recipient of the the Prize Engenheiro António de Almeida, the Prize Engenheiro Cristiano Spratley, and the Merit Scholarship from the University of Porto, and the best student poster prize at the 2nd IMA Conference on Mathematics in Communications. He was also awarded doctoral and postdoctoral research fellowships from the Portuguese Foundation for Science and Technology, and research fellowships from Foundation Calouste Gulbenkian.