COMPRESSION ONLINE DECOMPOSITION OF DYNAMIC SIGNALS VIA $N$-\ell_1
MINIMIZATION WITH CLUSTERED PRIORS

Huynh Van Luong*, Nikos Deligiannis†, Sören Forchhammer‡, and André Kaup*

*Multimedia Communications and Signal Processing, University of Erlangen-Nuremberg, 91058 Erlangen, Germany
†Vrije Universiteit Brussel, Pleinlaan 2, B-1050 Brussels, Belgium
‡DTU Fotonik, Technical University of Denmark, 2800 Kgs. Lyngby, Denmark

ABSTRACT

We introduce a compressive online decomposition via solving an $n$-\ell_1 cluster-weighted minimization to decompose a sequence of data vectors into sparse and low-rank components. In contrast to conventional batch Robust Principal Component Analysis (RPCA)—which needs to access full data—our method processes a data vector of the sequence per time instance from a small number of measurements. The $n$-\ell_1 cluster-weighted minimization promotes (i) the structure of the sparse components and (ii) their correlation with multiple previously-recovered sparse vectors via clustering and re-weighting iteratively. We establish guarantees on the number of measurements required for successful compressive decomposition under the assumption of slowly-varying low-rank components. Experimental results show that our guarantees are sharp and the proposed algorithm outperforms the state of the art.

Index Terms— Robust PCA, sparse signal, low-rank model, cluster-weighted minimization, prior information.

1. INTRODUCTION

Robust principal component analysis (RPCA) [1] has emerged as an important method in data analysis with applications ranging from anomaly detection to computer vision [1]. RPCA considers a data matrix $M \in \mathbb{R}^{n \times t}$ as the sum of a sparse component matrix $X$ and a low-rank component matrix $L$ and solves the principal component pursuit (PCP) [1] problem:

$$\min_{L,S} \|L\|_* + \lambda \|X\|_1 \text{ s.t. } M = L + X,$$

where $\|L\|_* = \sum \sigma_i(L)$ is the nuclear norm—sum of singular values $\sigma_i(L)$—of the matrix $L$, $\|X\|_1$ is the $\ell_1$-norm of $X$ (seen as a long vector), and $\lambda$ is a balance parameter. Batch RPCA requires access to the full data $M$, where the low-rank $L$ lies on the low-dimensional subspace and the sparse $X$ accounts for structured discrepancies. In streaming data, where $M$ is formed from a sequence of vectors, batch RPCA induces significant computations as well as high storage and communication costs.

Online RPCA [2, 3] and its compressive counterpart [4, 5, 6] can deal with these shortcomings by processing each column in $M$ (or compressive measurements from each column) sequentially. These approaches [2, 3] assume slowly-changing low-rank components and leverage compressed sensing (CS) [7] to recover the sparse component. The methods in [8, 4, 5, 6] operate on compressive measurements to handle the availability of subsampled or incomplete data so as to reduce the communication and storage costs.

Online RPCA [2] and recursive CS [9, 10] have been proposed to reconstruct a dynamic time-varying sparse signals using prior information. The study in [2] proposed a recursive method to the compressive case and used modified-CS [11] to leverage prior support knowledge under the condition of slowly-varying support. The study in [10] assumed non-varying low-rank components and recovered the sparse component using \ell_1-\ell_1 minimization [12, 9, 13, 14]. More methods are reviewed in [15, 16]. However, these methods do not exploit multiple prior information expressing the temporal correlation between incoming components and decomposed prior vectors.

The use of structural sparse components as prior knowledge has been studied in [17, 18, 19, 20]. Model-based CS [18] showed that exploiting prior information on the signal structure can reduce the number of measurements. The study in [20] introduced approximation algorithms to extend model-based CS to a wider class of signal models, whereas, the work in [19] leveraged the support of periodic clustered sparse signals. Alternatively, the structured sparsity model [17], which constrains signal coefficients into $C$-clusters without assuming prior knowledge of the location and size of the clusters, has provided provable performance guarantees. Motivated by these ideas, we use the clustered-based model for multiple priors via re-weighting the clustered sparse components per iteration.

Problem. We consider the following problem [21]: At time instance $t$, we aim to decompose $M_t = L_t + X_t \in \mathbb{R}^{n \times t}$ into $X_t = [x_1, x_2, ..., x_t]$ and $L_t = [v_1, v_2, ..., v_t]$, where $x_t, v_t \in \mathbb{R}^n$ are column-vectors in $X_t$ and $L_t$, respectively. We assume that $L_{t-1} = [v_1, v_2, ..., v_{t-1}]$ and $X_{t-1} = [x_1, x_2, ..., x_{t-1}]$ have been recovered at $t-1$ and at time instance $t$ we have access to measurements $y_t = \Phi_t(x_t + v_t)$, where $\Phi_t \in \mathbb{R}^{m_t \times n}$ $(m_t < n)$ is a random projection [22] that can vary with $t$. We formulate the problem

$$\min_{x_t, v_t} \left\{ \left\| L_{t-1} - v_t \right\|_1 + \lambda_1 \|x_t\|_1 + \lambda_2 f_{\text{prior}}(x_t, X_{t-1}) \right\},$$

s.t. $y_t = \Phi_t(x_t + v_t),$

where $f_{\text{prior}}(\cdot)$ expresses the correlation of $x_t$ with previously recovered sparse components $X_{t-1}$. Problem (2) aims to exploit the temporal correlation across multiple priors. Moreover, we want to leverage the structure in $x_t$ and $X_{t-1}$ to reduce further the number of measurements. This also raises a question of how many measurements are required to guarantee the successful decomposition.

Contribution. We derive theoretical guarantees on the number of measurements required for the $n$-\ell_1 cluster-weighted minimization. These guarantees serve as lower bounds for a compressive online decomposition algorithm (CODA) [21] that solves Problem (2) via minimizing the $n$-\ell_1 cluster-weighted problem. The algorithm recovers recursively the low-rank and sparse vectors given multiple priors [23] and leverages the structure of the sparse components by iteratively clustering and re-weighting.

2. COMPRESSIVE ONLINE DECOMPOSITION USING $N$-\ell_1 CLUSTER-WEIGHTED MINIMIZATION

We first formulate Problem (2) as the $n$-\ell_1 cluster-weighted minimization problem that is solved by CODA [21] and then, we establish performance guarantees followed up by a proof sketch.

2.1. The Compressive Online Decomposition Algorithm (CODA)

The $n$-\ell_1 Cluster-Weighted Minimization Problem. $n$-\ell_1 cluster-
where a set of indices of elements belonging to cluster across multiple priors. The algorithm computes \( \text{diag} \) number of elements in the cluster for the motion across the frames \([24]\) by means of optical flow \([25]\). 

B as priors for and is enhanced by incorporating the structure of the signals in the given time instance, the bound can be used to predict the number of required to solve Problem (4), which also serves as performance guar-

2.2. Decomposition Guarantees 

Solving Problem (4). CODA [21] solves the \( n \)-\( \ell_1 \) cluster-weighted minimization in (4) by using proximal gradient methods \([26]\), where, at every iteration \( k \), the algorithm updates the weights \( W_j, \gamma_j, \) and \( \beta_j \), and computes \( \tilde{x}_t \) and \( \tilde{v}_t \). In this way, we adaptively weight multiple priors according to their qualities during the iterative process. In this work, the constraints are set as \( \sum_{j=1}^{C} \gamma_j = 1 \) for each cluster, \( \sum_{j=1}^{C} \gamma_j = 1 \) in a given cluster, \( \sum_{j=1}^{C} \gamma_j = 1 \) across multiple priors. The algorithm computes \( \tilde{x}_t^{(k)} \) and \( \tilde{v}_t^{(k)} \) at iteration \( k + 1 \), by keeping \( W_j, \gamma_j, \) and \( \beta_j \) fixed, and using the soft thresholding operator \([26]\) and the single value thresholding operator \([27]\), respectively. For details on CODA, we refer to Algorithm 1 in [21]. CODA is based on our previous CORPICA \([23, 28]\) and operates in two steps: it first solves Problem (4) given \( Z_{t-1} \) and \( B_{t-1} \in \mathbb{R}^{n \times d} \) and the reconstructed \( \hat{x}_t \) and \( \hat{v}_t \) are used to update \( Z_t \) and \( B_t \), which are to be used in the following time instance.

2.2. Decomposition Guarantees 

We establish a bound for the minimum number of measurements required to solve Problem (4), which also serves as performance guarantees for CODA. The purpose of this bound is to determine the compressive rate for successfully recursive decomposition; namely, at a given time instance, the bound can be used to predict the number of measurements required at the next time instance. We aim to derive a bound that depends on the support of the sparse vector \( x_t \) to be recovered and the correlations between \( x_t \) and multiple priors \( z_j \). The correlation is expressed via the supports of the differences \( x_t - z_j \). We assume that \( v_t \) is slowly-changing; as such, by solving Problem (3), we obtain \( \tilde{v}_t \) incurring a certain measurement error that is bounded by \( \sigma \), \( |\| \phi_t(\tilde{v}_t - v_t) \||_2 < \sigma \), where \( \tilde{v}_t \) is a recovered low-rank component. In the rest of the paper, we use \( s_0 \) to denote the dimensionality of the support of the source \( x_t \) and \( s_j \) to denote that of each difference vector \( x_t - z_j \); namely, we can write \( |\| x_t - z_j \||_0 = s_j \) for \( j = 0, \ldots, J \). We establish the following result:

**Theorem 2.1.** In Problem (4), let \( m_t \) be the number of measurements required to successfully recover \( x_t \) \((|\| x_t \||_0 = s_0) \) and \( v_t \) from compressive measurements \( y_t = \Phi_t(x_t + v_t) \) where the elements of \( \Phi_t \in \mathbb{R}^{m \times n} \) are drawn from an i.i.d. Gaussian distribution—and prior information \( z_{t-1} := \{ z_j \}_{j=1}^{C} \) with \( |\| x_t - z_j \||_0 = s_j \) and \( z_j \in \mathbb{R}^n \) and \( B_{t-1} \in \mathbb{R}^{n \times d} \). Let also \( \lambda_{m_0} \) denote the expected length of a zero-mean, unit-variance \( m_1 \)-dimensional random Gaussian vector and \( \omega(x) \) the Gaussian width \([29, \text{Definition 3.1}]\).

1. If \( v_t \) is not changing in time \(^2\), we can recover \( x_t \) without error with probability greater than \( 1 - \exp(-\frac{1}{2}[\lambda_{m_0} - \omega(x)])^2 \) provided that
\[
m_t \geq 2n \exp \left( \frac{n}{\sum_{j=1}^{C} \beta_j \gamma_j^2} \right) + \frac{J}{6} \left( \sum_{j=0}^{J} \beta_j \gamma_j^2 \right) + 1.
\]
2. If \( v_t \) is slowly-changing, that is, \( B_{t-1} \) and \( z_{t-1} \) are low-rank, we obtain \( \hat{x}_t \) and \( \hat{v}_t \) by solving (4). Assuming that the measurement error of \( v_t \) is bounded as \( |\| \phi_t(\tilde{v}_t - v_t) \||_2 < \sigma \), we have that \( |\| x_t - \tilde{x}_t \||_2 \leq \frac{2\sigma}{\sqrt{\rho} \sqrt{n}} \) where \( \sigma < \sigma' \), with probability greater than \( 1 - \exp(-\frac{1}{2}[\lambda_{m_0} - \omega(x)]) - (1 - \frac{1}{\sqrt{\rho}})^2 \) provided that
\[
m_t \geq 2n \exp \left( \frac{n}{\sum_{j=1}^{C} \beta_j \gamma_j^2} \right) + \frac{J}{6} \left( \sum_{j=0}^{J} \beta_j \gamma_j^2 \right) + \frac{3}{2 \rho}. \]

where \( \alpha = \frac{1}{2} \sum_{j=1}^{C} \beta_j \gamma_j^2 + \frac{1}{2} e^2 \sum_{j=0}^{J} \beta_j \gamma_j^2 \) in both (5) and (6), wherein \( \epsilon > 0 \) and \( \tilde{\eta} = \min \left\{ \frac{1}{\sum_{j=1}^{C} (\|W_j\|_{\infty})^2 \|x_j \|_{\infty}^2} \right\} \).

2.3. Proof Sketch of Guarantees 

**Measurement Condition.** For convenience, we use the notation \( x, v, y, \) and skip the index \( t \) in the proof. Following the CS theory \([7]\), let \( x \in \mathbb{R}^m \) be a sparse signal for which we have access to random Gaussian measurements \( y = \Phi x \in \mathbb{R}^m, m < n \). Then \( x \) can be recovered by solving minimization \( \|x\|_1 \) s.t. \( y = \Phi x \), which can be relaxed to
\[
\min_{x} \|f(x) + g(x)\|_1,
\]
where \( f(x) = \frac{1}{2} \|x - y\|_2^2 \) and \( g(x) = \lambda \|x\|_1 \), and \( \lambda > 0 \) is a regularization parameter. Let us denote that the subdifferential \( \partial g(x) \) \([30]\) of a convex function \( g(x) \) at a point \( x \in \mathbb{R}^n \) is given by \( \partial g(x) := \{ u \in \mathbb{R}^n: g(y) \geq g(x) + u^T (y - x) \text{ for all } y \in \mathbb{R}^n \} \). Let \( g \sim N(0, L_0) \) denote a vector of \( n \) independent, zero-mean, and unit-variance Gaussian random variables and \( E[\cdot] \) the expectation with respect to \( g \). The Euclidean distance of \( g \) with respect to a convex cone \( C \) \([29]\) is defined by
\[
\text{dist}(g, C) := \min \{ \|g - u\|_2: u \in C \}. \]

**Proposition 2.2.** \([29, \text{Corollary 3.3, Proposition 3.6}]\) Let \( \Phi \in \mathbb{R}^{m \times n} \) be a random projection, whose elements are drawn from the i.i.d. Gaussian distribution, and \( \lambda_{m_0} \), \( \omega(x) \) be defined as in Theorem 2.1 and [29, Definition 3.1].

\(^2\)Problem (3) becomes \( \min_{x_t} \left\{ \frac{1}{2} \|\Phi_t x_t - y_t\|_2^2 + \lambda \sum_{j=0}^{J} \beta_j \|W_j(x_t - z_j)\|_1 \} \).
1. By observing \( y = \Phi x \in \mathbb{R}^m \) and solving (7), \( x \in \mathbb{R}^n \) is successfully recovered with probability greater than \( 1 - \exp(-\frac{1}{2}(\Lambda_m - \omega(x))^2) \) provided that \( m \geq U_y + 1 \).

2. We observe \( y = \Phi x + \theta \) with the noise \( \theta \) bounded as \( |\theta| \leq \sigma \). Let \( x \) denote any solution in (7) and \( \theta < \rho < 1 \). We have that \( |x - x_\ell| < \frac{2\gamma}{\sqrt{m}} \) with probability greater than \( 1 - \exp(-\frac{1}{2}(\Lambda_m - \omega(x))^2) \) provided that \( m \geq U_y + 3/\rho \).

The quantity \( U_y \) is calculated given a convex norm function \( g(x) \) by

\[
U_y = \min_{\tau > 0} \mathbb{E}_g \left[ \text{dist}^2(g, \tau \cdot \tilde{g}g(x)) \right].
\]

(9)

**Supporting results.** Recall that the probability density of the normal distribution \( N(0, 1) \) with zero-mean and unit variance is

\[
\psi(x) := \left(1 / \sqrt{2\pi}\right) e^{-x^2/2}.
\]

(10)

We also consider the following inequality [9]:

\[
\frac{1 - x^{-1}}{\sqrt{\pi \log(x)}} \leq 1 \leq \frac{2}{\sqrt{2\pi}} \frac{x}{5},
\]

(11)

for all \( x > 1 \). Moreover, adhering to the formulation in [9], we use the following inequality with \( \epsilon > 0 \) in our derivations:

\[
\frac{1}{\sqrt{2\pi}} \int_0^\epsilon (v - x)^2 e^{-v^2/2} dv \leq \psi(x) / x.
\]

(12)

**Proof of Theorem 2.1.** We firstly derive the bound in (5); the bound in (6) is then derived as its noisy counterpart.

We derive the bounds on Proposition 2.2. We first compute the subdifferential \( \tilde{g}g(x) \) and then the distance, \( \text{dist}(\cdot) \), between the standard normal vector \( g \) and \( \tilde{g}g(x) \). The \( u \in \tilde{g}g(x) \) of \( g(x) \) is derived through the decomposed components of the sum \( g(x) = \sum_{j=1}^J \beta_j g_j(x) \), where \( g_j(x) = \| \gamma_j \| \mathbf{W}_j(x - z_j) \| \). As a result, \( \tilde{g}g(x) = \sum_{j=1}^J \beta_j \tilde{g}g_j(x) \). Considering that the distance from the standard normal vector \( g \) to the subdifferential \( \tilde{g}g(x) \) given by (8), we derive a relation between this distance and all decomposed distances of \( g_j(x) \) as

\[
\text{dist}^2(g, \tau \cdot \tilde{g}g(x)) \leq \sum_{j=1}^J \beta_j^2 \text{dist}^2(g, \tau \cdot \tilde{g}g_j(x)),
\]

where \( \sum_{j=1}^J \beta_j = 1 \). Taking the expectation of (13) gives

\[
\mathbb{E}_g[\text{dist}^2(g, \tau \cdot \tilde{g}g(x))] \leq \sum_{j=1}^J \beta_j \mathbb{E}_g[\text{dist}^2(g, \tau \cdot \tilde{g}g_j(x))].
\]

(14)

**Distance Expectation**

We now consider each component \( \mathbb{E}_g[\text{dist}^2(g, \tau \cdot \tilde{g}g_j(x))] \), where \( g_j(x) = \| \gamma_j \| \mathbf{W}_j(x - z_j) \| \). We firstly consider \( x \in z_j \). We have defined that \( |x - z_j| = s_j \). Without loss of generality, we assume that all the \( n_j \) components of the first cluster of \( x \) are zero, i.e., \( x_{\Omega_j \setminus z_j} = 0 \). The nonzero components are clustered into \( \Omega_j = 1 \) disjoint clusters, \( \Omega_j = \ldots = \Omega_{\ell_j} \), leading to the following to the subdifferential \( u \in \tilde{g}g_j(x) \) of \( g_j(x) \) is given by

\[
u_i = \gamma_j^w s_j \text{sign}(x_i - z_j), \quad \Omega_j = \sum_{i=1}^{\ell_j} \nu_i \leq \gamma_j^w s_j, \quad \ell_j.
\]

(15)

Let us consider the weights \( w_j \) that are defined in CODA [21] for a specific \( z_j \) and denote that \( n_j = \sum_{i=\ell_j}^{\ell_j + \ell_j^*} \nu_i \).

\[
\eta_j = \frac{1}{\sum_{i=\ell_j}^{\ell_j + \ell_j^*} |w_j^i|_2^2}, \quad i \in \Omega_j, \quad \ell_j.
\]

We express \( w_j \) and \( \eta_j \) for the source \( x - z_j \) as

\[
w_j = \frac{\eta_j}{\ell_j^*}, \quad \text{respectively},
\]

(16)

We can then compute the distance from the standard normal vector \( g \) to the subdifferential \( \tilde{g}g_j(x) \) based on (8)

\[
\text{dist}^2(g, \tau \cdot \tilde{g}g_j(x)) = \sum_{i=\ell_j}^{\ell_j + \ell_j^*} \left| \gamma_j^w s_j \right|^2 \left( 1 - \gamma_j^w s_j \right)^2 + \frac{1}{\ell_j^*}
\]

(18)

where \( \max(a, 0) \) returns the maximum value between \( a \in \mathbb{R} \) and 0. Further, after taking the expectation of (18), given \( w_j^i_{\Omega_j} = 1 \) [see (16)], we obtain

\[
\mathbb{E}_g[\text{dist}^2(g, \tau \cdot \tilde{g}g_j(x))] = \sum_{i=\ell_j}^{\ell_j + \ell_j^*} \eta_j^2 \left( 1 - \gamma_j^w s_j \right)^2.
\]

(19)

We apply (12) on the third term in (19) to get

\[
\mathbb{E}_g[\text{dist}^2(g, \tau \cdot \tilde{g}g_j(x))] \leq \eta_j^2 \left( 1 - \gamma_j^w s_j \right)^2 + \frac{1}{\ell_j^*} \left( 1 - \gamma_j^w s_j \right)^2.
\]

(20)

**3.2.2. Bound Derivation**

Inserting (20) in (14) for all functions \( g_j(x) \) gives

\[
\mathbb{E}_g[\text{dist}^2(g, \tau \cdot \tilde{g}g(x))] \leq \sum_{j=0}^{\ell_j} \beta_j s_j + \tau^2 \sum_{j=0}^{\ell_j} \beta_j^2 \sum_{i=\ell_j}^{\ell_j + \ell_j^*} \eta_j^2 \left( 1 - \gamma_j^w s_j \right)^2 + \frac{1}{\ell_j^*} \left( 1 - \gamma_j^w s_j \right)^2.
\]

(21)

where \( \gamma_j^w s_j \). Then, we can write (21) as

\[
\mathbb{E}_g[\text{dist}^2(g, \tau \cdot \tilde{g}g(x))] \leq \sum_{j=0}^{\ell_j} \beta_j + \tau^2 \sum_{j=0}^{\ell_j} \beta_j^2 \sum_{i=\ell_j}^{\ell_j + \ell_j^*} \eta_j^2 \left( 1 - \gamma_j^w s_j \right)^2 + \frac{1}{\ell_j^*} \left( 1 - \gamma_j^w s_j \right)^2.
\]

(22)

Inequality (22) holds due to \( s_j = \sum_{j=0}^{\ell_j} \beta_j s_j \) and from (9), (22), we have

\[
U_y = \min_{\tau > 0} \left( \left( s + \tau^2 \sum_{j=0}^{\ell_j} \beta_j^2 \sum_{i=\ell_j}^{\ell_j + \ell_j^*} \eta_j^2 \left( 1 - \gamma_j^w s_j \right)^2 \right) \right).
\]

(23)

Inserting (10) in (23) gives

\[
U_y = \min_{\tau > 0} \left( \left( \left( \tau^2 \sum_{j=0}^{\ell_j} \beta_j^2 \sum_{i=\ell_j}^{\ell_j + \ell_j^*} \eta_j^2 \left( 1 - \gamma_j^w s_j \right)^2 \right) \right) \right).
\]

(24)

We can select a parameter \( \tau > 0 \) in (24) to obtain a bound; here, we select \( \tau = (\eta / \ell_j^*) \sqrt{\log(n / \bar{s})} \), which gives

\[
U_y \leq \bar{s} + 2 \log(2) \sqrt{\bar{s} \sum_{j=0}^{\ell_j} \beta_j^* \sum_{i=\ell_j}^{\ell_j + \ell_j^*} \eta_j^* \left( 1 - \gamma_j^w \bar{s}_j \right)^2}.
\]

(25)
We denote \( \alpha = \frac{\gamma}{\bar{s}} \sum_{j=0}^{j_0} \sum_{\ell=1}^{C_d} \gamma_j^2 \sum_{i=1}^{\ell-1} u_{j,i}^2 \) in the second term in (25). Finally, applying (11) to the last term of (25) gives
\[
U_{\nu} \leq 2\alpha \log(n/\bar{s}) + (7/5)\bar{s}.
\]
As a result, from Proposition 2.2, we get the bound in (5) as
\[
m_{\ell} \geq 2\alpha \log(n/\bar{s}) + (7/5)\bar{s} + 1.
\]
For slowly-changing \( \nu \), we assume that the measurement error \( \| \Phi(\hat{v} - v) \|_2 < \sigma \). Applying Proposition 2.2 for the noisy case, we get the bound in (6).

3. EXPERIMENTAL RESULTS

The performance of CODA employing \( n-\ell_1 \) cluster-weighted minimization is evaluated and compared to our previous CORPCA [23, 28] with \( n-\ell_1 \) minimization as well as GRASTA [3] and ReProCS [2]. ReProCS [2] recovers the sparse components, while GRASTA recovers the low-rank components [5]. The experimental setup involves both synthetic and real video data, as described in [21, 28].

**Synthetic Data:** We evaluate CODA on a test set of synthetically-generated data vectors \( \mathbf{M} = [\mathbf{x}_{d+1} + \mathbf{v}_{d+1}, \ldots, \mathbf{x}_{d+q} + \mathbf{v}_{d+q}] \), where we set \( n = 500, r = 5 \), the number of vectors for training \( d = 100 \), and the number of testing vectors \( q = 100 \). We vary \( s_0 \) (from 10 to 110) and the number of measurements \( m \), and we assess the probability of success for the sparse component \( \mathbf{P}_{\text{Sparse}(\text{success})} \) and the low-rank component \( \mathbf{P}_{\text{Low-rank}(\text{success})} \) component, averaged over the test vectors. \( \mathbf{P}_{\text{Sparse}(\text{success})} \) (resp. \( \mathbf{P}_{\text{Low-rank}(\text{success})} \)) is defined as the number of times in which the sparse component \( \mathbf{v} \) (resp. the low-rank component \( \mathbf{v} \)) is recovered within an error \( \| \mathbf{v} - \hat{\mathbf{v}} \|_2 \leq 10^{-2} \) (resp. \( \| \mathbf{v} - \hat{\mathbf{v}} \|_2 \leq 10^{-2} \)) divided by the total 50 Monte Carlo simulations. In CODA, we have set \( \epsilon = 0.1, \lambda = 1/\sqrt{\bar{m}}, \mu = 10^{-3} \), and the number of clusters \( C = 7 \). In addition, we evaluate our bound for \( n-\ell_1 \) cluster-weighted minimization in (6) as well as the corresponding bounds for \( n-\ell_1 \) minimization\(^4 \) [28], ReProCS (bounded as weighted-\( \ell_1 \) [31], denoted as \( w\ell_1 \)), \( \ell_1 \), and \( \ell_1-\ell_1 \) minimization\(^5 \).

We set the parameter \( \rho \) to 1.2/3, 1/3, 1/3, 1.5/3, and 0.4/3 for the bounds of \( n-\ell_1 \) cluster-weighted, \( n-\ell_1 \), \( \ell_1-\ell_1 \), and \( w\ell_1 \) minimization, respectively, based on experimentation.

The results in Fig. 1 demonstrate the efficiency of CODA. CODA can recover a 500-dimensional data vector from few measurements \( m/n = 0.25 \) to 0.6, see the white areas in Fig. 1(a)). For values of \( s_0 > 70 \), CORPCA-\( n-\ell_1 \) can not recover the sparse components successfully [see grey areas in Fig. 1(b)], while CODA succeeds. Fig. 1(e) shows that the performance of ReProCS is worse than that of CODA. Moreover, Fig. 1(e) shows that GRASTA delivers lower low-rank recovery performance than CODA. It is also worth observing that the measurement bound for recovering the sparse component by \( n-\ell_1 \) cluster-weighted minimization [red line in Fig. 1(a)] is sharper than the existing bounds for \( n-\ell_1 \), \( \ell_1-\ell_1 \), and \( w\ell_1 \) minimization, depicted in Figs. 1(b)-(e).

**Video Data:** Fig. 2 shows the video foreground-background decomposition performance of CODA against CORPCA [21] for the Bootstrap (rescaled to 60×80 pixels) and Curtain (rescaled to 64×80 pixels) sequences [32], under various measurement rates \( m/n \). The results show that CODA can recover better foreground (upper) and background (lower) images, even at low rates (i.e., \( m/n = 0.5 \) for Bootstrap and \( m/n = 0.4 \) for Curtain).

4. CONCLUSION

The paper introduced a compressive online decomposition algorithm (CODA) employing \( n-\ell_1 \) cluster-based minimization, which decomposes dynamic signals from compressive measurements. CODA incorporates multiple priors in the decomposition problem and leverages the sparse structures via iteratively clustering and re-weighting the sparse components. We also established theoretical bounds on the number of measurements to guarantee successful decomposition for \( n-\ell_1 \) cluster-weighted minimization. Experimental results have shown that the established bounds are sharp and that CODA outperforms existing methods.

---

\(^4\)In [31], the weighted-\( \ell_1 \) bound is obtained as \( m_{\text{modCS}} \geq 2s_0 \log(n/s_0) + 2s_0 + s_0 \).

\(^5\)The bounds for \( \ell_1 \) and \( \ell_1-\ell_1 \) minimization to achieve successful recovery of the sparse component assuming a slow-varying low-rank component are \( m_{\ell_1} \geq 2\frac{\bar{m}}{\rho} \log(\frac{n}{s_0}) + \frac{\gamma}{\bar{s}} \frac{s_0}{\rho} + \frac{\bar{m}}{\rho} \frac{\gamma}{\bar{s}} \) and \( m_{\ell_1-\ell_1} \geq 2\frac{\bar{m}}{\rho} \log(\frac{n}{s_0 + \xi}) + \frac{\gamma}{\bar{s}} \frac{s_0}{\rho} + \frac{\bar{m}}{\rho} \frac{\gamma}{\bar{s}} \frac{s_0}{\rho} + \frac{\bar{m}}{\rho} \frac{\gamma}{\bar{s}} \frac{s_0}{\rho} \) [9] (see Theorem 2.1).
5. REFERENCES


