Compressed Randomized UTV Decompositions for Low-Rank Approximations and Big Data Applications

Maboud F. Kaloorazi†, and Rodrigo C. de Lamare‡,‡

† Centre for Telecommunications Studies (CETUC)
Pontifical Catholic University of Rio de Janeiro, Brazil
‡ Department of Electronics, University of York, United Kingdom
E-mail: {kaloorazi,delamare}@cetuc.puc-rio.br

Abstract—Low-rank matrix approximations play a fundamental role in numerical linear algebra and signal processing applications. This paper introduces a novel rank-revealing matrix decomposition algorithm termed Compressed Randomized UTV (CoR-UTV) decomposition along with a CoR-UTV variant aided by the power method technique. CoR-UTV is primarily developed to compute an approximation to a low-rank input matrix by making use of random sampling schemes. Given a large and dense matrix of size \( m \times n \) with numerical rank \( k \), where \( k \ll \min\{m,n\} \), CoR-UTV requires a few passes over the data, and runs in \( O(\text{rank}) \) floating-point operations. Furthermore, CoR-UTV can exploit modern computational platforms and, consequently, can be optimized for maximum efficiency. CoR-UTV is simple and accurate, and outperforms reported alternative methods in terms of efficiency and accuracy. Simulations with synthetic data as well as real data in image reconstruction and robust principal component analysis applications support our claims.

Index Terms—Matrix computations, low-rank approximations, UTV decomposition, randomized algorithms, dimension reduction, matrix decomposition, image reconstruction, robust PCA.

I. INTRODUCTION

LOW-RANK matrix approximations, that is, approximating a given matrix by one of lower rank, play an increasingly important role in signal processing and its applications. Such compact representation which retains most important information of a high-dimensional matrix can provide a significant reduction in memory requirements, and more importantly, computational costs when the latter scales, e.g., according to a high-degree polynomial, with the dimensionality. Matrices with low-rank structures have found many applications in background subtraction [1], [2], [3], [4], system identification [5], IP network anomaly detection [6], [7], latent variable graphical modeling, [8], ranking and collaborative filtering, [9], subspace clustering [10], [11], [12], adaptive, sensor and multichannel signal processing [13], [14], [15], [16], [17], [18], [19], [20], [21], [22], [23], [24], [25], [26], [27], [28], [29], [30], [31], biometrics [32], [33], statistical process control and multidimensional fault identification [34], [35], quantum state tomography [36], and DNA microarray data [37].

Singular value decomposition (SVD) [38] and the rank-revealing QR (RRQR) decomposition [39], [40] are among the most commonly used algorithms for computing a low-rank approximation of a matrix. On the other hand, a UTV decomposition [41], [42] is a compromise between the SVD and the RRQR decomposition having the virtues of both: UTV \( i) \) is computationally more efficient than the SVD, and \( ii) \) provides information on the numerical null space of the matrix (RRQR does not explicitly furnish the null space information) [41], [42], [43], [44]. Given a matrix \( A \), the UTV algorithm computes a decomposition \( A = UTV^T \), where \( U \) and \( V \) have orthonormal columns, and \( T \) is triangular (either upper or lower triangular). These deterministic algorithms, however, are computationally expensive for large data sets. Furthermore, standard techniques for their computation are challenging to parallelize in order to utilize advanced computer architectures [45], [46], [47]. Recently developed algorithms for low-rank approximation based on random sampling schemes, however, have been shown to be remarkably computationally efficient, highly accurate and robust, and are known to outperform the traditional algorithms in many practical situations [46], [47], [48], [49], [50], [51]. The power of randomized algorithms lies in the facts that \( i) \) they are computationally efficient, and \( ii) \) their main operations can be optimized for maximum efficiency on modern computational platforms.

A. Contributions

Inspired by recent developments, this paper presents a novel randomized rank-revealing algorithm termed compressed randomized UTV (CoR-UTV) decomposition [52]. Given a large and dense rank-\( k \) matrix \( A \) of size \( m \times n \), the CoR-UTV algorithm computes a low-rank approximation \( \hat{A}_{\text{CoR}} \) of \( A \) such that

\[
\hat{A}_{\text{CoR}} = UTV^T, \tag{1}
\]

where \( U \) and \( V \) have orthonormal columns, and \( T \) is triangular (either upper or lower, whichever is preferred). CoR-UTV only requires a few passes through data, for a matrix stored externally, and runs in \( O(mnk) \) floating-point operations (flops). The operations of the algorithm involve matrix-matrix
multiplication, the QR and RRQR decompositions. Due to recently developed Communication-Avoiding QR algorithms [53], [54], [55], which can perform the computations with optimal/minimum communication costs, CoR-UTV can be optimized for peak machine performance on modern architectures. We provide a theoretical analysis for CoR-UTV, that is, the rank-revealing property of the algorithm is proved, and upper bounds on the error of the low-rank approximation are given.

Furthermore, we apply CoR-UTV to treat an image reconstruction problem, as well as to solve the robust principal component analysis (robust PCA) problem [1], [56], [57], i.e., to decompose a given matrix with grossly corrupted entries into a low-rank matrix plus a sparse matrix of outliers, in applications of background subtraction in surveillance video, and shadow and specularity removal from face images.

B. Notation

Bold-face upper-case letters are used to denote matrices. For a matrix \( A \), \( \| A \|_0, \| A \|_1, \| A \|_2, \| A \|_\infty, \) and \( \| A \|_F \) denote the \( \ell_0 \)-norm, the \( \ell_1 \)-norm, the spectral norm, the Frobenius norm, and the nuclear norm, respectively. \( \sigma_j(A) \) and \( \sigma_{\min}(A) \) denote the \( j \)-th largest and the smallest singular value of \( A \), respectively. The numerical range and numerical null space of \( A \) are denoted by \( \mathcal{R}(A) \) and \( \mathcal{N}(A) \), respectively. The symbol \( \mathbb{E} \) denotes expected value with respect to random variables, and the dagger \( \dagger \) denotes the Moore-Penrose pseudo-inverse.

The remainder of this paper is structured as follows. In Section III we introduce the mathematical model of the data and discuss related works. In Section III we describe our proposed method, which also includes a variant that uses the power iteration scheme in detail. Section IV presents our theoretical analysis. In Section V we develop an algorithm for robust PCA using CoR-UTV. In Section VI we present and discuss our numerical experimental results, and our concluding remarks are given in Section VII.

II. MATHEMATICAL MODEL AND RELATED WORKS

Given a matrix \( A \in \mathbb{R}^{m \times n} \), where \( m \geq n \), with numerical rank \( k \), its singular value decomposition (SVD) [38] is defined as:

\[
A = U \Sigma V^T = \begin{bmatrix} U_k & U_0 \end{bmatrix} \begin{bmatrix} \Sigma_k & 0 \\ 0 & \Sigma_0 \end{bmatrix} \begin{bmatrix} V_k \\ V_0 \end{bmatrix}^T,
\]

where \( U_k \in \mathbb{R}^{m \times k}, U_0 \in \mathbb{R}^{m \times n-k} \) have orthonormal columns, \( \Sigma_k \in \mathbb{R}^{k \times k} \) and \( \Sigma_0 \in \mathbb{R}^{n-k \times n-k} \) are diagonal matrices containing the singular values, i.e., \( \Sigma_k = \text{diag}(\sigma_1, ..., \sigma_k) \) and \( \Sigma_0 = \text{diag}(\sigma_{k+1}, ..., \sigma_n) \), and \( V_k \in \mathbb{R}^{n \times k} \) and \( V_0 \in \mathbb{R}^{n \times n-k} \) have orthonormal columns. \( A \) can be written as \( A = A_k + A_0 \), where \( A_k = U_k \Sigma_k V_k^T \) and \( A_0 = U_0 \Sigma_0 V_0^T \). The SVD constructs the optimal rank-\( k \) approximation \( A_k \) to \( A \), [58], [59] i.e.,

\[
\min_{\text{rank}(B) \leq k} \| A - B \|_F = \| A - A_k \|_F = \sqrt{\sum_{j=k+1}^{n} \sigma_j^2}.
\]

In this paper we focus on the matrix \( A \) defined above. The SVD is highly accurate and yields detailed information on singular subspaces and singular values. However, it is prohibitive to compute for large data sets. Moreover, standard techniques for its computation are challenging to parallelize in order to take advantage of modern computational environments [45], [46], [47]. An economic version of the SVD is the partial SVD based on Krylov subspace methods, such as the Lanczos and Arnoldi algorithms, which constructs an approximate SVD of an input matrix, for instance \( A \), at a cost \( \mathcal{O}(mnk) \). The partial SVD, however, suffers from two drawbacks. First, inherently, it is numerically unstable [38], [39], [60]. Second, it does not lend itself to parallel implementations [46], [47], which makes it unsuitable for modern computational architectures.

Another widely used algorithm for low-rank approximations considered as a relatively economic alternative to the SVD is the RRQR decomposition [39]. The RRQR is a special QR decomposition with column pivoting (QRCP), which reveals the numerical rank of the input matrix. Given the matrix \( A \), it takes the following form:

\[
AP = QR = Q \begin{bmatrix} R_{11} & R_{12} \\ 0 & R_{22} \end{bmatrix},
\]

where \( P \) is a permutation matrix, \( Q \in \mathbb{R}^{m \times n} \) has orthonormal columns, \( R \in \mathbb{R}^{n \times n} \) is upper triangular where \( R_{11} \in \mathbb{R}^{k \times k} \) is well-conditioned with \( \sigma_{\min}(R_{11}) = O(\sigma_k) \), and the \( \ell_2 \)-norm of \( R_{22} \in \mathbb{R}^{n-k \times n-k} \) is sufficiently small, i.e., \( \| R_{22} \|_2 = O(\sigma_{k+1}) \) (here we have written the reduced QR decomposition, where the silent columns and rows of \( Q \) and \( R \), respectively, have been removed). If there is an additional requirement that the \( \ell_2 \)-norm of \( R_{11}^{-1}R_{12} \) is small, i.e., a low order polynomial in \( n \), this decomposition is called “strong RRQR decomposition” [40]. The rank-\( k \) approximation to \( A \) is then computed as follows:

\[
\hat{A}_{\text{RRQR}} = Q(:, 1 : k) R(1 : 1, :) P^T,
\]

where we have used MATLAB notation to indicate submatrices, i.e., \( Q(:, 1 : k) \) denotes the first \( k \) columns of \( Q \), and \( R(1 : 1, :) \) denotes the first \( k \) rows of \( R \).

A UTV decomposition [41], [42] is a compromise between the SVD and QRCP, which has the virtues of both. For the matrix \( A \), it takes the form:

\[
A = UTV^T \quad (7)
\]

where \( U \in \mathbb{R}^{m \times n} \) and \( V \in \mathbb{R}^{n \times n} \) have orthonormal columns, and \( T \) is triangular. If \( T \) is upper triangular, the decomposition is called URV decomposition [41]:

\[
A = U \begin{bmatrix} T_{11} & T_{12} \\ 0 & T_{22} \end{bmatrix} V^T. \quad (8)
\]

If \( T \) is lower triangular, the decomposition is called ULV decomposition [42]:

\[
A = U \begin{bmatrix} T_{11} & 0 \\ T_{21} & T_{22} \end{bmatrix} V^T. \quad (9)
\]
The URV and ULV decompositions are collectively referred to as UTV decompositions and are performed by reduction of the matrix $A$ using unitary transformations to upper and lower triangular forms, respectively. If there is a well-defined gap in the singular value spectrum of $A$, i.e., $\sigma_k \gg \sigma_{k+1}$, the UTV decompositions are said to be rank-revealing in the sense that the numerical rank $k$ is revealed in the triangular submatrices $T_{11} \in \mathbb{R}^{k \times k}$, and the $\ell_2$-norm of off-diagonal submatrices, $[T_{12}^T T_{22}^T]^T$ and $[T_{21} \ T_{22}]$, are of the order $O(\sigma_{k+1})$.

QR and UTU decompositions, however, provide highly accurate approximations to $A$, but they suffer from two drawbacks. First, they are expensive to compute in terms of arithmetic costs, i.e., $O(mn^2)$. Second, methods for their computation are challenging to parallelize and, as a result, they cannot exploit modern architectures.

Recently developed algorithms for low-rank approximations based on randomization have attracted significant attention due to the facts that they can not exploit modern architectures. The works in [66], [67] proposed to apply a random Gaussian embedding matrix $R$ to a low-rank matrix. The rank approximation was then obtained through computations in a compressed form. Further computations are then performed on the reduced-size matrix through the QR decomposition with column pivoting to obtain the final low-rank approximation. Sarlós [50] proposed a randomized algorithm termed subspace-orbit randomized SVD (SOR-SVD), projects the matrix onto a low-dimensional subspace using a random matrix, capturing most attributes of the data. Further computations are then performed on the reduced-size matrix through the QR decomposition and the SVD to give the approximation. Gu [47] applied a slightly modified version of the R-SVD algorithm to improve subspace iteration methods, and presents a new error analysis. The second method proposed by Halko et al. [46] Section 5.5] was a single-pass algorithm, i.e., it required only one pass over the data, to compute a low-rank approximation. For the matrix $A$, the decomposition, which we call two-sided randomized SVD (TSR-SVD), is computed as described in Alg. 1.

**Algorithm 1 Two-Sided Randomized SVD (TSR-SVD)**

**Input:** Matrix $A \in \mathbb{R}^{m \times n}$, integers $k$ and $\ell$.

**Output:** A rank-$\ell$ approximation.

1. Draw random matrices $\Phi_1 \in \mathbb{R}^{m \times \ell}$ and $\Phi_2 \in \mathbb{R}^{m \times \ell}$.
2. Compute $Y_1 = A\Phi_1$ and $Y_2 = A^T \Phi_2$ in a single pass through $A$;
3. Compute QR decompositions $Y_1 = Q_1 R_1$, $Y_2 = Q_2 R_2$;
4. Compute $B_{\text{approx}} = Q_1^T Y_1 (Q_2^T \Phi_1)^\dagger$;
5. Compute an SVD $B_{\text{approx}} = U \Sigma V$;
6. $A \approx (Q_1 U) \Sigma (Q_2 V)^T$.

In Alg. 1 $Q_1 \bar{U} \in \mathbb{R}^{m \times \ell}$ and $Q_2 \bar{V} \in \mathbb{R}^{n \times \ell}$ are approximations to the left and right singular subspace of $A$, respectively, $\Sigma \in \mathbb{R}^{\ell \times \ell}$ contains an approximation to the first $\ell$ singular values of $A$, and $B_{\text{approx}}$ is an approximation to $B = Q_1 A Q_2$.

TSR-SVD, however, gives a very poor approximation compared to the optimal SVD due to the single-pass strategy. The reason behind is, mainly, poor approximate basis drawn from the row space of $A$, i.e., $Q_1$. Furthermore, for a general input matrix, the authors do not provide neither theoretical error bounds nor numerical experiments for the TSR-SVD.

The work in [3] proposed a rank-revealing decomposition algorithm based on randomized sampling schemes; the matrix $A$ is compressed through pre- and post-multiplication by approximate orthonormal bases for $R(A)$ and $R(A^T)$ obtained via randomization, columns of the reduced matrix and, accordingly, the bases are permuted, the low-rank approximation is then given by projecting the compressed matrix back to the original space. The work in [68] proposed a randomized algorithm termed subspace-orbit randomized SVD (SOR-SVD) to compute a fixed-rank approximation of an input matrix. SOR-SVD, first, alternately projects the matrix onto its column and row space. Next, orthonormal bases for $R(A)$ and $R(A^T)$ are approximated. The matrix is then transformed into a lower dimensional space using the approximate bases. Finally, a truncated SVD is carried out on the transformed data, and the low-rank approximation is given by projecting the small projected data back to the original space.

This work was developed by drawing inspiration from the rank-revealing algorithm proposed in [3], and also SOR-SVD in [68]. Our analysis was inspired by the work in [68].

III. **COMPRESSED RANDOMIZED UTV DECOMPOSITIONS**

In this section, we present a rank-revealing decomposition algorithm powered by the randomized sampling schemes termed compressed randomized UTV (CoR-UTV) decomposition, which computes a low-rank approximation of a given matrix. We focus on the matrix $A$ with $m \geq n$, where the CoR-UTV algorithm, in the form of (1), produces a middle matrix $T$, which is upper triangular, i.e., URV decomposition. The modifications required for a corresponding CoR-UTV algorithm for the other case, where $m < n$ that produces a lower triangular middle matrix $T$, i.e., ULV decomposition, is
The CoR-UTV algorithm is presented in Alg. 2.

### Compressed Randomized UTV (CoR-UTV)

**Input:** A rank-$\ell$ approximation.

1. Draw a standard Gaussian matrix $\Psi \in \mathbb{R}^{n \times \ell}$; 
2. Compute $C_1 = A\Psi$; 
3. Compute $C_2 = A^T C_1$; 
4. Compute QR decompositions $C_1 = Q_1 R_1$, $C_2 = Q_2 R_2$; 
5. Compute $D = Q_2^T A Q_2$; 
6. Compute the QRCP $D = \tilde{Q} \tilde{R} \tilde{P}^T$; 
7. Form the CoR-UTV-based low-rank approximation of $A$: $A_{\text{CoR}} = U T \Psi^T$; 

where $U = Q_1 \tilde{Q} \in \mathbb{R}^{m \times \ell}$ and $V = Q_2 \tilde{P} \in \mathbb{R}^{n \times \ell}$ construct approximations to the $\ell$ leading left and right singular vectors of $A$, respectively, and $T = R \in \mathbb{R}^{\ell \times \ell}$ is upper triangular with diagonals approximating the first $\ell$ singular values of $A$.

The CoR-UTV algorithm is presented in Alg. 2.

CoR-UTV requires three passes over the data, for a matrix stored externally, but it can be altered to revisit the data only once. To this end, the compressed matrix $D$, equation (14), can be computed by making use of available matrices as follows:

$$D Q_2^T \Psi = Q_1^T A Q_2 Q_2^T \Psi.$$  

where both sides of currently unknown $D$ are postmultiplied by $Q_2^T \Psi$. Having defined $A \approx A Q_2 Q_2^T$ and $C_1 = A \Psi$, an approximation to $D$ is obtained by:

$$D_{\text{approx}} = Q_1^T C_1 (Q_2^T \Psi)^\dagger.$$  

The key differences between CoR-UTV and TSR-SVD are as follows:

- CoR-UTV uses a sketch of the input matrix in order to project it onto its row space, i.e., equation (12). This $i)$ significantly improves the quality of the approximate basis $Q_2$, and, as a result, the quality of the approximate right singular subspace of $A$ compared to that of TSR-SVD, which uses a random matrix for the projection, and $ii)$ allows (15) to provide a highly accurate approximation to (14).
- CoR-UTV applies a column-pivoted QR decomposition to $D$, i.e., equation (15), whereas TSR-SVD uses an SVD to factor the compressed matrix. This, as explained later on, reduces the computational costs of CoR-UTV compared to TSR-SVD.

The key difference between CoR-UTV and SOR-SVD is that SOR-SVD applies a truncated SVD and, as result, gives a rank-$k$ approximation to $A$, while CoR-UTV employs a column-pivoted QR decomposition and returns a rank-$\ell$ approximation. Nevertheless, the SVD is computationally more expensive than the column-pivoted QR, and standard techniques to computing it are challenging to parallelize. While recently developed column-pivoted QR algorithms use randomization, which can factor a matrix with optimal/minimum communication cost, this can substantially reduce the computational costs of decomposing the compressed matrix, compared to the SVD, when it does not fit into fast memory.

CoR-UTV may be sufficiently accurate for matrices whose singular values display some decay, however, in applications where the data matrix has a slowly decaying singular values, it may produce singular vectors and singular values that deviate significantly from the exact ones (computed by the SVD).

Notice that to compute CoR-UTV when the power method is employed, a non-updated $C_2$ must be used to form $D_{\text{approx}}$.

### Analysis of CoR-UTV Decompositions

In this section, we provide an analysis of the performance of CoR-UTV, the basic version in Alg. 2 and the one that uses the power method in Alg. 3. In particular, we discuss the rank-revealing property of the algorithm, and provide upper bounds on the error of the low-rank approximation for CoR-UTV.

We borrow material from [68] since the two algorithms, CoR-UTV and SOR-SVD, have a few steps similar. However,
Algorithm 3 CoR-UTV with Power Method

**Input:** Matrix $A \in \mathbb{R}^{m \times n}$, integers $k$, $\ell$ and $q$.

**Output:** A rank-$k$ approximation.

1. Draw a standard Gaussian matrix $C_2 \in \mathbb{R}^{n \times \ell}$;
2. for $i = 1: q + 1$ do
   3. Compute $C_1 = AC_2$;
   4. Compute $C_2 = A^T C_1$;
5. end for
6. Compute QR decompositions $C_1 = Q_1 R_1$, $C_2 = Q_2 R_2$;
7. Compute $D = Q_1^T A Q_2$ or $D_{\text{approx}} = Q_1^T C_1 (Q_2^T C_2)^\dagger$;
8. Compute a QRCP $D = QR \tilde{P}$ or $D_{\text{approx}} = QR \tilde{P}^\dagger$;
9. Form the CoR-UTV-based low-rank approximation of $A$: $A_{\text{CoR}} = UTV^T$; $U = Q_1 Q, T = R, V = Q_2 P^T$.

The key difference is that these randomized algorithms employ different deterministic decomposition methods to factor the input matrix. We discuss that CoR-UTV is computationally cheaper and, moreover, can exploit advanced computer architectures better than SOR-SVD.

### A. Rank-Revealing Property

To prove that CoR-UTV is rank-revealing, it is required to show that i) the $T$ factor of the decomposition reveals the rank of $A$, and ii) the trailing off-diagonal block of $T$ is small in $\ell_2$-norm. Furthermore, the relation between the Gaussian random matrix used and the $A$ rank-$k$ approximation to an input matrix $A$ is called an oversampling parameter $[46]$, $[47]$. Since $\Psi$ has interaction with the right singular vectors $V$ of $A$, i.e., equation (11), we have

$$\Psi = \sqrt{\tilde{A}} \Psi \approx [\tilde{A}_1 \quad \tilde{A}_2]^T$$

where $p$ is called an oversampling parameter $[46]$, $[47]$. Since $\Psi$ has interaction with the right singular vectors $V$ of $A$, i.e., equation (11), we have

$$\Psi = \sqrt{\tilde{A}} \Psi \approx [\tilde{A}_1 \quad \tilde{A}_2]^T$$

and when the matrix $D$ is formed through step 1 to step 7 of Alg. 3 i.e., the power method is used, we have

$$\sigma_k \geq \sigma_k(D) \geq \sqrt{1 + \frac{\sigma_k}{\sqrt{1 + \nu^2 \gamma_k}}}$$

and when the power method is used. Alg. 3 we have

$$\nu = \nu_1 \nu_2, \nu_1 = \sqrt{n - \ell + p + \sqrt{\ell} + 7}, \nu_2 = \frac{4p}{\sqrt{\ell}}.$$ This completes the discussion on the rank-revealing property of the CoR-UTV algorithm.

### B. Low-Rank Approximation

CoR-UTV efficiently constructs an accurate low-rank approximation to an input matrix $A$. We provide theoretical guarantees on the accuracy of these approximations in terms of the Frobenius and spectral norm. To this end, we first state a theorem from $[63]$.

**Theorem 4:** Let the matrix $A$ have an SVD as defined in (2), and $D_1 \in \mathbb{R}^{m \times \ell}$ and $D_2 \in \mathbb{R}^{n \times \ell}$ be orthonormal matrices constructed by means of CoR-UTV, where $1 \leq k \leq \ell$. Let, furthermore, $D_k$ be the best rank-$k$ of $D = Q_1 A Q_2$. Then, we have

$$\|A - Q_1 D_k Q_2^T\|_F \leq \|A_0\|_F + \|A_k - Q_1 Q_2^T A_k\|_F$$

and

$$\|A - Q_1 D_k Q_2^T\|_2 \leq \|A_0\|_2 + \|A_k - Q_1 Q_2^T A_k\|_2.$$
Now, we rewrite the CoR-UTV-based low-rank approximation \( \hat{A}_{\text{CoR}} = Q_1 D Q_2^T \). (30)

This perfectly makes sense since the column-pivoted QR decomposition, which factors \( D \) is a numerically stable deterministic method \([38]\). Thus, for \( \xi = 2, F \), it follows that

\[
\| A - \hat{A}_{\text{CoR}} \|_\xi \leq \| A - Q_1 D_k Q_2^T \|_\xi. \tag{31}
\]

This relation holds because \( D_k \) is the rank-\( k \) approximation of \( D \).

**Theorem 5:** With the notation of Theorem 2 for \( \xi = 2, F \), we have

\[
\| A - \hat{A}_{\text{CoR}} \|_\xi \leq \| A_0 \|_\xi + \| A_k - Q_1 Q_1^T A_k \|_F + \| A_k - A_k Q_2 Q_2^T \|_F. \tag{32}
\]

Having stated the connection between CoR-UTV and SOR-SVD, we now obtain upper bounds for the CoR-UTV-based low-rank approximation error.

**Theorem 6:** Let the matrix \( A \) have an SVD as defined in (2), \( 2 \leq p + k \leq \ell \), and \( \hat{A}_{\text{CoR}} \) is computed through the basic version of CoR-UTV, Alg. 2. Furthermore, assume that \( \hat{\Psi}_1 \) is full row rank. Then, for \( \xi = 2, F \), we have

\[
\| A - \hat{A}_{\text{CoR}} \|_\xi \leq \| A_0 \|_\xi + \sqrt{\frac{\alpha^2 \| \Psi_2 \|_2^2 \| \Psi_1 \|_2^2}{1 + \beta^2 \| \Psi_2 \|_2^2 \| \Psi_1 \|_2^2}} + \frac{\eta^2 \| \Psi_2 \|_2^2 \| \Psi_1 \|_2^2}{1 + \tau^2 \| \Psi_2 \|_2^2 \| \Psi_1 \|_2^2}, \tag{33}
\]

where \( \alpha = \sqrt{\frac{\ell - p + 1}{\sigma_k}} \), \( \beta = \frac{\sigma_{p-1}}{\sigma_k} \), \( \eta = \sqrt{\frac{\ell - p + 1}{\sigma_k}} \), \( \tau = \frac{\sigma_{p-1}}{\sigma_k} \).

When the power iteration is used, Alg. 3, \( \alpha = \sqrt{\frac{\ell - p + 1}{\sigma_k}} \left( \frac{\sigma_{p-1}}{\sigma_k} \right)^{2q} \), \( \beta = \frac{\sigma_{p-1}}{\sigma_k} \left( \frac{\sigma_{p-1}}{\sigma_k} \right)^{2q} \). (34)

**Theorem 7:** With the notation of Theorem 6 and \( \gamma_k = \frac{\sigma_{p+1}}{\sigma_k} \), for the basic version of CoR-UTV, Alg. 2, we have

\[
\mathbb{E}\| A - \hat{A}_{\text{CoR}} \|_\xi \leq \| A_0 \|_\xi + (1 + \gamma_k) \sqrt{\nu \ell \sigma_{p+1}}, \tag{34}
\]

and when the power method is used, Alg. 3, we have

\[
\mathbb{E}\| A - \hat{A}_{\text{CoR}} \|_\xi \leq \| A_0 \|_\xi + (1 + \gamma_k) \sqrt{\nu \ell \sigma_{p+1} \gamma_k^{2q}}, \tag{35}
\]

where \( \nu \) is defined in Theorem 3.

This completes the discussion on the low-rank approximation error bounds for the CoR-UTV algorithm.

### C. Computational Complexity

The computational cost of any algorithm involves i) arithmetic, i.e., the number of floating-point operations, and ii) communication, i.e., synchronization and data movement either through levels of a memory hierarchy or between parallel processing units \([53]\). On advanced computers, for a large data matrix which is stored externally, the communication cost becomes substantially more expensive compared to the arithmetic \([53]\), \([72]\). Therefore, developing new algorithms or redesigning existing algorithms to solve a problem in hand with minimum communication cost is highly desirable.

The advantage of algorithms based on randomization over their classical counterparts lies in the fact that i) they operate on a reduced-size version of the data matrix rather than a matrix itself, resulting in a reduction of flops, and ii) they can be organized to exploit modern architectures, performing a decomposition with minimum communication cost.

To factor the matrix \( A \), CoR-UTV of Alg. 2 incurs the following costs (we only consider high-order terms):

- **Step 1** (generating \( \Psi \)) costs \( n \ell \).
- **Step 2** (forming \( C_1 \)) costs \( 2mn \ell \).
- **Step 3** (forming \( C_2 \)) costs \( 2mn \ell \).
- **Step 4** (QR decompositions) costs \( 2m\ell^2 + 2n\ell^2 \).
- **Step 5** (forming \( D \)) costs \( 2m\ell^2 + 2mn\ell \). If the matrix \( D \) is approximated by \( \hat{D}_{\text{approx}} \) of equation \( (13) \) in this step, the cost would be \( 2m\ell^2 + 2n\ell^2 + 3\ell^3 \).
- **Step 6** (performing QRCP) costs \( \frac{8\ell^3}{3} \).
- **Step 7** (forming the left and right approximate bases) costs \( 2m\ell^2 + 2n\ell \).

Summing up the costs in Steps 1 to 7, we obtain:

\[
C_{\text{CoR-UTV}} \sim 3\ell C_{\text{mult}} + 6m\ell^2 + n\ell(2\ell + 3) + \frac{8}{3}\ell^3, \tag{36}
\]

or

\[
C_{\text{CoR-UTV}} \sim 2\ell C_{\text{mult}} + 6m\ell^2 + n\ell(4\ell + 3) + \frac{17}{3}\ell^3, \tag{37}
\]

when the compressed matrix \( D \) is approximated by \( \hat{D}_{\text{approx}} \), where \( C_{\text{mult}} \) is the cost of a matrix-vector multiplication with \( A \) or \( A^T \). The first terms of the right-hand sides of \( (36) \) and \( (37) \), resulting from multiplying \( A \) and \( A^T \) with the corresponding matrices dominate the costs, and the sample size parameter \( \ell \) is typically close to the minimal rank \( k \). When CoR-UTV employs the power method, it requires \( 2q + 3 \) passes over the data (for a matrix stored out-of-core) with the following operation count:

\[
C_{\text{CoR-UTV}} \sim (2q + 3)\ell C_{\text{mult}} + 6m\ell^2 + n\ell(2\ell + 3) + \frac{8}{3}\ell^3. \tag{38}
\]

For the case in which the compressed matrix \( D \) is approximated by \( \hat{D}_{\text{approx}} \), CoR-UTV requires \( 2q + 2 \) passes over the data, and the flop count satisfies

\[
C_{\text{CoR-UTV}} \sim (2q + 2)\ell C_{\text{mult}} + 6m\ell^2 + n\ell(4\ell + 3) + \frac{17}{3}\ell^3. \tag{39}
\]

The Co-R-UTV, TFR-SVD and SOR-SVD algorithms except for matrix-matrix multiplications, which are readily parallelizable perform two QR decompositions on matrices of
size $m \times \ell$ and $n \times \ell$. Demmel et al. [53] recently developed communication-avoiding sequential and parallel QR decomposition algorithms that perform the computations with optimal communication costs. Hence, this step of all three algorithms can be implemented efficiently. In addition, CoR-UTV performs one QRCP on an $\ell \times \ell$ matrix, however TSR-SVD and SOR-SVD perform an SVD on the $\ell \times \ell$ matrix, which is more expensive than QRCP. Furthermore, recently developed QRCP algorithms based on randomization can perform the factorization with minimum communication costs [54], [55], [70], while standard techniques to compute an SVD are challenging for parallelization [45], [46], [47]. As a result, for very large matrices to be factored on high performance computing architectures, where the compressed $\ell \times \ell$ matrix does not fit into fast memory, the execution time to compute CoR-UTV can be substantially less than those of TSR-SVD and SOR-SVD. This is an advantage of CoR-UTV over TSR-SVD and SOR-SVD. See [53], [54] for a comprehensive discussion on communication costs.

V. ROBUST PCA WITH CoR-UTV

This section describes how to solve the robust PCA problem using the proposed CoR-UTV method. Principal component analysis (PCA) [34] is a widely-used linear dimensionality reduction technique that transforms a high-dimensional data to a low-dimensional subspace which contains most features of the original data. PCA, however, is known to be very sensitive to grossly perturbed observations. In order to robustify PCA against gross corruption, robust PCA [1], [56], [57] was proposed. Robust PCA represents an input low-rank matrix $M \in \mathbb{R}^{m \times n}$, whose a fraction of entries being corrupted, as a linear superposition of a low-rank matrix $L$ and a sparse matrix of outliers $S$ such as $M = L + S$, by solving the following convex program:

$$\min_{(L,S)} \|L\|_*, \|S\|_1 + \lambda \|S\|_1,$$

subject to $M = L + S,$

(40)

where $\|B\|_* \triangleq \sum_i \sigma_i(B)$ is the nuclear norm of any matrix $B$, $\|B\|_1 \triangleq \sum_{i,j} |B_{i,j}|$ is the $l_1$-norm of $B$, and $\lambda > 0$ is a tuning parameter [73], [74], [75], [76], [77], [78], [79], [80], [81], [82], [83], [84], [85], [86], [87], [88], [89], [90], [91], [92], [93], [94], [95], [96], [97], [98], [99], [100], [101], [102], [103], [104]. One efficient method to solve (40) is the method of augmented Lagrange multipliers (ALM) [105], [106], which iteratively minimizes the following augmented Lagrangian function with respect to either variable $L$ or $S$ with the other one being fixed:

$$L(L, S, Y, \mu) \triangleq \|L\|_* + \lambda \|S\|_1 + (Y, M - L - S)$$

$$+ \frac{\mu}{2} \|M - L - S\|_F^2,$$

(41)

where $Y \in \mathbb{R}^{m \times n}$ is the Lagrange multiplier matrix, and $\mu > 0$ is a penalty parameter. The robust PCA solved by the ALM method is given in Alg. [4]

In Alg. [4] for any matrix $B$ with an SVD defined as $B = U_B \Sigma_B V_B^T$, $D_\delta(B)$ refers to a singular value thresholding operator defined as $D_\delta(B) = U_B \Sigma_B \delta(\Sigma_B) V_B^T$, where $\delta(x) = \text{sgn}(x) \max(|x| - \delta, 0)$ is a shrinkage operator [107], and $\lambda$, $\mu$, $Y_0$, and $S_0$ are initial values.

The ALM method yields the optimal solution, however its serious bottleneck is computing a computationally demanding SVD at each iteration to approximate the low-rank component $L$ of $M$. To address this concern and to speed up the convergence of the ALM method, the work in [108] proposes a few techniques including predicting the principal singular space dimension, a continuation technique [109], and a truncated SVD by using PROPACK package [110]. The modified algorithm [108] substantially improves the convergence speed, however the bottleneck is that the truncated SVD [110] employed uses the lanczos algorithm that is inherently unstable and, moreover, due to the limited data reuse in its operations, it has very poor performance on modern architectures [38], [46], [47], [60].

To address this issue, we thus, by retaining the original objective function proposed in [1], [56], [57], [108], apply CoR-UTV as a surrogate to the truncated SVD to solve the robust PCA problem. We adopt the continuation technique [108], [109], which increases $\mu$ in each iteration. The proposed method which is called ALM-CoRUTV is given in Alg. [5]

Algorithm 4 Robust PCA solved by ALM

**Input:** Matrix $M$, $\lambda$, $\mu$, $Y_0 = S_0 = 0$, $j = 0$.

**Output:** Low-rank plus sparse matrix.

1: while the algorithm does not converge do
2: Compute $L_{j+1} = D_{\mu^{-1}}(M - S_j + \mu^{-1}Y_j)$;
3: Compute $S_{j+1} = S_{\lambda^{-1}}(M - L_{j+1} + \mu^{-1}Y_j)$;
4: Compute $Y_{j+1} = Y_j + \mu(M - L_{j+1} - S_{j+1})$;
5: end while
6: return $L^*$ and $S^*$.

Algorithm 5 Robust PCA solved by CoR-UTV

**Input:** Matrix $M$, $\lambda$, $\mu$, $Y_0 = S_0 = 0$, $j = 0$.

**Output:** Low-rank plus sparse matrix.

1: while the algorithm does not converge do
2: Compute $L_{j+1} = D_{\mu^{-1}}(M - S_j + \mu^{-1}Y_j)$;
3: Compute $S_{j+1} = S_{\lambda^{-1}}(M - L_{j+1} + \mu^{-1}Y_j)$;
4: Compute $Y_{j+1} = Y_j + \mu(M - L_{j+1} - S_{j+1})$;
5: Update $\mu_{j+1} = \max(\mu_j, \mu)$;
6: end while
7: return $L^*$ and $S^*$.

In Alg. [5] for any matrix $B$ having a CoR-UTV decomposition described in Section III $C_\delta(B)$ refers to a CoR-UTV thresholding operator defined as:

$$C_\delta(B) = U(:, 1 : r) \Sigma(1 : r, :) V^T,$$

(42)

where $r$ is the number of diagonals of $T$ greater than $\delta$, and $\lambda$, $\mu_0$, $\mu$, $\rho$, $Y_0$, and $S_0$ are initial values. The main operation of the ALM-CoRUTV algorithm is computing CoR-UTV in each iteration, which is efficient in terms of flops, $O(mnk)$, and can be computed with minimum communication costs; see subsection IV-C. In subsection IV-C we experimentally verify that ALM-CoRUTV converges to the exact optimal solution.
VI. NUMERICAL EXPERIMENTS

In this section, we present the results of some numerical experiments conducted to evaluate the performance of the CoR-UTV algorithm for approximating a low-rank input matrix. We show that CoR-UTV provides highly accurate singular values and low-rank approximations, and compare our algorithm against several other algorithms from the literature. We furthermore employ CoR-UTV for solving the robust PCA problem. The experiments were run in MATLAB on a desktop PC with a 3 GHz intel Core i5-4430 processor and 8 GB of memory.

A. Comparison of Rank-Revealing Property & Singular Values

We first show that CoR-UTV i) is rank revealer, i.e., the gap in the singular value spectrum of the input matrix is revealed, and ii) provides highly accurate singular values that with remarkable fidelity track singular values of the matrix. For the sake of simplicity, we focus on square matrices.

For the randomized algorithms considered, namely CoR-UTV, TSR-SVD, and SOR-SVD, here and in the next subsection, the results presented are averaged over 20 trials. Each trial was run with the same input matrix with an independent draw of the test matrix (or matrices for TSR-SVD).

We construct two types of matrices of order $10^3$:

- Matrix 1 (Noisy Low-rank). This matrix is formed by a linear superposition of two matrices $A = A_1 + E$, $A_1 = UV^T$, where $U$ and $V$ are random orthonormal matrices, $\Sigma$ is a diagonal matrix containing the singular values $\sigma_j$ that decrease linearly from 1 to $10^{-9}$, and $\sigma_j = 0$ for $j = k+1, ..., 10^3$. The matrix $E$ is a Gaussian matrix normalized to have $\ell_2$-norm gap $\sigma_k$. We set the numerical rank $k = 20$, and consider two cases:
  - gap = 0.01: NoisyLowRank-I
  - gap = 0.1: NoisyLowRank-II

- Matrix 2 (Fast Decay). This matrix is formed in a similar way as $A_1$ of Matrix 1, but now the diagonals of $\Sigma$ take a different form such that $\sigma_j = 1$ for $j = 1, ..., k$, and $\sigma_j = (j-k+1)^{-2}$ for $j = k+1, ..., 10^3$. We set the numerical rank $k = 10$.

We compare the quality of singular values of the matrices considered computed by our method, described in Section III against that of alternative rank-revealing methods such as the SVD, QR with column pivoting (QRCP), UTV, described in Section II, and TSR-SVD (Alg. 1).

For CoR-UTV and TSR-SVD, we arbitrarily set the sample size parameter to $\ell = 2k$. Both algorithms require the same number of passes over $A$, either two or $2q+2$ when the power method is used, to perform a factorization. To compute a UTV decomposition, we implement the $\text{lurv}$ function from [61].

The results are shown in Figs. 1 and 2. We make the following observations:

- CoR-UTV, without making use of the power iteration scheme, i.e., $q = 0$, provides an excellent approximation to singular values for NoisyLowRank-I and Matrix 2. For NoisyLowRank-II, CoR-UTV outperforms TSR-SVD when $q = 0$, in approximating both leading and trailing singular values, and it only requires two steps of the power iteration to deliver singular values as accurate as the optimal SVD. The QRCP algorithm, however, gives a fuzzy approximation to singular values of the input matrices.

- For all matrices (NoisyLowRank-I, NoisyLowRank-II, Matrix 2), CoR-UTV strongly reveals the numerical rank $k$, as do the SVD, UTV and TSR-SVD, while QRCP weakly reveals the rank of NoisyLowRank-I and Matrix 2, and only suggests the gap in the singular values of NoisyLowRank-II.

B. Comparison of Low-Rank Approximation

1) Rank-$\ell$ Approximation: Since CoR-UTV computes a rank-$\ell$ approximation of a given matrix, we first investigate how accurate this approximation is. To this end, we compute a rank-$\ell$ approximation $\hat{A}_{\text{CoR}}$ for NoisyLowRank-I, NoisyLowRank-II, and Matrix 2 using Alg. 2 and Alg. 3 for each sample size parameter $\ell$, and calculate the approximation error as:

$$e_{\ell} = \|A - \hat{A}_{\text{CoR}}\|_2.$$  \hspace{1cm} (43)

We compare the approximation errors against those produced by the rank-$\ell$ approximation using the SVD, i.e., minimal error $\sigma_{\ell+1}$.

Judging from the figures, i) when $q = 0$, which corresponds to the basic version of CoR-UTV (Alg. 2), the approximation
is rather poor. ii) The error incurred by Alg. 2 produces an upper bound for the minimal error. iii) With only one step of the power iteration \( q = 1 \), Alg. 3 the accuracy of the approximation substantially improves, resulting in an approximation as accurate as the optimal SVD for all three matrices.

2) Rank-\( k \) Approximation: We now compare the low-rank approximations constructed by our method against those of the SVD, QRCP, and TSR-SVD. We also include SOR-SVD [68] in our comparison. We have excluded the UTV algorithm because it has, by far, the worst performance among the algorithms discussed. This allows us to display the behavior of other algorithms clearly in the graphs.

To make a fair comparison, we construct a rank-\( k \) approximation \( \hat{A}_{out} \) to \( A \) by each algorithm, and calculate the error:

\[
\varepsilon_k = \| A - \hat{A}_{out} \|_\xi ,
\]

where \( \xi = F \) for the Frobenius-norm error, and \( \xi = 2 \) for the spectral-norm error.

A rank-\( k \) approximation for the SVD, QRCP is computed as described in [2] and [6], respectively. For the randomized algorithms, TSR-SVD, SOR-SVD, CoR-UTV, however, we construct a rank-\( k \) approximation by varying the sample size parameter \( \ell \), since, as shown, this parameter colors the quality of approximations. The rank-\( k \) approximation by TSR-SVD, Alg. 1 is constructed by selecting the first \( k \) approximate singular vectors and corresponding singular values. SOR-SVD constructs a rank-\( k \) approximation of an input matrix, see [68], and the rank-\( k \) approximation by CoR-UTV is computed as:

\[
\hat{A}_{CoR-k} = U(:,1:k)T(1:k,:)V^T.
\]

For the randomized algorithms, we run the experiment with no power method \( (q = 0) \), and \( q = 2 \). We make two observations: (1) When \( q = 0 \), for matrices NoisyLowRank-I and NoisyLowRank-II, as the number of samples increases the performance of CoR-UTV exceeds that of QRCP, becoming close to optimal performance of the SVD. For these two matrices, CoR-UTV and SOR-SVD show similar performances, exceeding the performance of TSR-SVD. For Matrix 2, by increasing the sample size parameter TSR-SVD and SOR-SVD show slightly better performance than CoR-UTV, while CoR-UTV outperforms QRCP. (2) When \( q = 2 \), the errors resulting from CoR-UTV show no loss of accuracy compared to the optimal SVD. In this case, QRCP has the poorest performance for all examples.

3) Image Reconstruction: We assess the quality of low-rank approximation by reconstructing a gray-scale image of a differential gear of size 1280 × 804, taken from [55], using CoR-UTV, truncated QRCP, and the truncated SVD by using (widely recommended) PROPACK package [110]. The PROPACK function provides an efficient algorithm to compute a specified number of largest singular values and corresponding singular vectors of a given matrix by making use of the Lanczos bidiagonalization algorithm with partial reorthogonalization, which is suitable for approximating large low-rank matrices.

The results display the Frobenius-norm approximation error against the corresponding approximation rank, where the error is calculated as:

\[
\varepsilon_{approx} = \| A - \hat{A}_{approx} \|_F ,
\]

where \( \hat{A}_{approx} \) is the approximation computed by each algorithm.

For the rank-20 approximation, truncated QRCP and CoR-UTV without power iteration technique produce the poorest reconstruction qualities. CoR-UTV with one step of power iteration produces a better result. Truncated SVD and CoR-UTV with two steps of power iteration, however, appear to have reconstructed images that are visually identical. For the rank-90 approximation, with a careful scrutiny, fine defects appear in reconstructions by truncated QRCP and CoR-UTV with \( q = 0 \), while reconstructed images by truncated SVD, CoR-UTV with \( q = 1 \) and \( q = 2 \) are visually indistinguishable from the original.

C. Robust Principal Component Analysis

In this subsection, we experimentally investigate the efficiency and efficacy of ALM-CoRUTV, described in Table 5 in recovering the low-rank and sparse components of synthetic and real data. We compare the results obtained with those of the efficiently implemented inexact ALM method by [108], called InexactALM hereafter.

1) Synthetic Matrix Recovery: We form a rank-\( k \) matrix \( M = L + S \) as a linear combination of a low-rank matrix \( L \in \mathbb{R}^{n \times n} \) and a sparse error matrix \( S \in \mathbb{R}^{n \times n} \). The matrix \( L \) is generated as \( L = UV^T \), where \( U, V \in \mathbb{R}^{n \times k} \) have standard Gaussian distributed entries. The error matrix \( S \) has \( s \) non-zero entries independently drawn from the set \( \{-80, 80\} \).

We apply the ALM-CoRUTV and InexactALM algorithms to \( M \) to recover \( L \) and \( S \). The numerical results are summarized in Tables 1 and 1. Table 1 presents the results where the rank of \( L \) \( r(L) = 0.05 \times n \) and \( s = \| S \|_0 = 0.05 \times n^2 \), and Table 1 presents the results for a more challenging scenario where \( r(L) = 0.05 \times n \) and \( s = \| S \|_0 = 0.10 \times n^2 \).

In our experiments, we adopt the initial values suggested in [108], and both algorithms are terminated when the following stopping condition holds:

\[
\frac{\| M - L^{out} - S^{out} \|_F}{\| M \|_F} < 10^{-5} ,
\]

where \( (L^{out}, S^{out}) \) is the pair of output of either algorithm. The results of ALM-CoRUTV are reported in the numerators, and those of InexactALM in the denominators. In the Tables, \( Time(s) \) refers to the computational time in seconds, \( Iter. \) refers to the number of iterations, and \( \zeta \) refers to the relative error defined as \( \frac{\| M - L^{out} - S^{out} \|_F}{\| M \|_F} \).

CoR-UTV requires a prespecified rank \( \ell \) to perform the factorization. Thus, we set \( \ell = 2k \), as a random start, and \( q = 1 \) (one step of a power iteration). Judging from the results in Tables 1 and 1 we make several observations on ALM-CoRUTV:

- It successfully detects the exact numerical rank \( k \) of the input matrix in all cases.


**TABLE I: Numerical results for synthetic matrix recovery for the case** \( r(L) = 0.05 \times n \) and \( s = 0.05 \times n^2 \).

| Dataset | InexactALM | ALM-CoRUTV |
|---------|------------|------------|
| Highway | 50e5       | 1e5        |
| Escalator | 100e5     | 1e5        |
| Airport | 150e4      | 50e5       |
| Train stop | 150e4  | 50e5       |

**TABLE II: Numerical results for synthetic matrix recovery for the case** \( r(L) = 0.05 \times n \) and \( s = 1 \times n^2 \).

| Dataset | InexactALM | ALM-CoRUTV |
|---------|------------|------------|
| Highway | 50e5       | 1e5        |
| Escalator | 100e5     | 1e5        |
| Airport | 150e4      | 50e5       |
| Train stop | 150e4  | 50e5       |

- It provides the exact optimal solution, having the same number of iterations for the first test case, while it requires one more iteration for the second challenging test case, compared to InexactALM.
- It outperforms InexactALM in terms of runtime, with speedups of up to 8.6 times.

In summary, ALM-CoRUTV exactly recovers the low-rank and sparse matrices from a grossly corrupted matrix at a much lower cost compared to InexactALM. However, we expect ALM-CoRUTV to be faster on multicore and accelerator-based computers, since CoR-UTV can be computed with minimum communication cost.

2) **Background Modeling in Surveillance Video:** Extracting the foreground from the background in a video stream is an increasingly important task in video analysis. This task can be formulated as a robust PCA problem, where the background is represented by a sparse matrix and the foreground is represented by a low-rank matrix.

Here, we apply ALM-CoRUTV to four different surveillance videos. The first two videos are from [111], and the other two are from [112]. The first video consists of 200 grayscale frames of size 176 × 144, taken in a hall of an airport. The frames are stacked as columns of a matrix \( M \), forming \( M \in \mathbb{R}^{25344 \times 200} \). This video has a relatively static background. The second video consists of 200 grayscale frames of size 130 × 160, taken from an escalator at an airport. We form a matrix \( M \in \mathbb{R}^{20800 \times 200} \) by stacking individual frames as its columns. The background of this video changes due to the moving escalator. The third video has 200 grayscale frames of size 240 × 320, taken from a highway. We thus form \( M \in \mathbb{R}^{76800 \times 200} \). The fourth video consists of 200 grayscale frames of size 288 × 432, taken in a tram stop. Therefore, \( M \in \mathbb{R}^{124416 \times 200} \).

In order for CoR-UTV, used in ALM-CoRUTV, to approximate the low-rank component of real data, we determine the prespecified rank \( \ell \) by making use of the following bound that relates the numerical rank \( k \) of any matrix \( B \) with the nuclear and Frobenius norms [38]:

\[
\|B\|_* \leq \sqrt{k} \|B\|_F.
\] (48)

We set \( \ell = k + p \), where \( k \) is the minimum value satisfying (48), and \( p = 2 \) is an oversampling parameter. Again, we set \( q = 1 \) for CoR-UTV.

**TABLE III: Numerical results for real matrix recovery.**

| Dataset | InexactALM | ALM-CoRUTV |
|---------|------------|------------|
| Airpot | 15.4       | 5.1        |
| Escalator | 11.9      | 4.2        |
| Highway | 53.6       | 16.2       |
| Train stop | 83.6   | 25.2       |
| Yale B01 | 4.2       | 2.1        |
| Yale B02 | 4.2       | 2.1        |

3) **Shadow and Specularity Removal From Face Images:**

Another task in computer vision that fits nicely into the robust PCA model is removing shadows and specularities from face images; images of the same face taken under varying illumination can be modeled as a superposition of a low-rank and sparse components.

In this experiment, we use face images taken from the Yale B face database [113]. Table III summarizes the numerical results.

We conclude that ALM-CoRUTV can successfully recover the face images under different illuminations from the dataset studied two times faster than InexactALM.

**VII. CONCLUSION**

In this paper, we have presented CoR-UTV, a rank-revealing algorithm based on the randomized sampling paradigm, for computing a low-rank approximation of an input matrix. We have presented theoretical analysis for CoR-UTV, and have shown through numerical experiments on two classes of matrices that CoR-UTV reveals the numerical rank better than QRCP, and provides results as good as those of the optimal SVD. CoR-UTV outperforms QRCP in low-rank approximation, and when the power method is incorporated, provides results as accurate as those of the SVD. CoR-UTV is more efficient than the deterministic SVD, QRCP, UTV, and competing randomized TSR-SVD and SOR-SVD in terms of arithmetic cost and, moreover, can exploit advanced computational platforms better by exposing higher levels of parallelism.
than all algorithms mentioned. We also applied CoR-UTV to solve the robust PCA problem via the ALM method. Our studies demonstrate that the resulting ALM-CoRUTV provides the exact optimal solution and, moreover, is substantially faster than efficiently implemented InexactALM.

REFERENCES

[1] J. Wright, Y. Peng, Y. Ma, A. Ganesh, and S. Rao, “Robust Principal Component Analysis: Exact Recovery of Corrupted Low-Rank Matrices,” NIPS, pp. 2080–2088, 2009.
[2] T. Bouwmans and E. H. Zahzah, “Robust PCA via Principal Component Pursuit: A review for a comparative evaluation in video surveillance,” Computer Vision and Image Understanding, vol. 122, pp. 22–34, 2014.
[3] M. F. Kaloorazi and R. C. de Lamare, “Low-Rank and Sparse Matrix Recovery Based on a Randomized Rank-revealing Decomposition,” in 22nd Intl Conf. on DSP 2017, UK, Aug 2017.
[4] M. Rahmani and G. Atia, “High Dimensional Low Rank Plus Sparse Matrix Decomposition,” IEEE Trans. Signal Process., vol. 65, no. 8, pp. 2004–2019, Jan 2017.
[5] M. F. Kaloorazi, M. F. Kaloorazi, and G. B. Giannakis, “Decentralized Sparsity-Regularized Rank Minimization: Algorithms and Applications,” IEEE Trans. Signal Process., vol. 61, no. 21, pp. 5374–5388, 2013.
[6] M. F. Kaloorazi and R. C. de Lamare, “Anomaly Detection in IP Networks Based on Randomized Subspace Methods,” in ICASSP.
[7] V. Chandrasekaran, P. Parrilo, and A. Willsky, “Latent variable graphical model selection via convex optimization,” The Ann. of Stat., vol. 40, no. 4, pp. 1935–1970, 2012.
[8] N. Srebro, N Alon and T. Jaakkola, “Generalization error bounds for collaborative prediction with low-rank matrices,” in NIPS’04, pp. 5–27.
[9] M. Soltanolkotabi, E. Elhamifar, and E. J. Candès, “Robust subspace clustering,” Annals of Statistics, vol. 42, no. 2, pp. 669–699, 2014.
[10] M. Rahmani and G. Atia, “Coherence Pursuit: Fast, Simple, and Robust Principal Component Analysis,” IEEE Trans. Signal Process., vol. 65, no. 23, pp. 6260–6275, Dec 2017.
[11] T.-H. Oh, Y. Matsushita, Y.-W. Tai, and I. S. Kweon, “Fast randomized singular value thresholding for low-rank optimization,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 40, no. 2, pp. 376–391, Mar. 2017.
[12] R. C. de Lamare and R. Sampaio-Neto, “Reduced-rank interference suppression for ds-cdma based on interpolated fir filters,” IEEE Communications Letters, vol. 9, no. 3, pp. 213–215, March 2005.
[13] ——, “Adaptive reduced-rank mmse filtering with interpolated fir filters and adaptive interpolators,” IEEE Signal Processing Letters, vol. 12, no. 3, pp. 177–180, March 2005.
[14] ——, “Reduced-rank adaptive filtering based on joint iterative optimization of adaptive filters,” IEEE Signal Processing Letters, vol. 14, no. 12, pp. 980–983, Dec 2007.
[15] R. C. de Lamare, “Adaptive reduced-rank lcmv beamforming algorithms based on joint iterative optimization of filters,” Electronics Letters, vol. 44, no. 9, pp. 565–566, April 2008.
[16] R. C. de Lamare, M. Haardt, and R. Sampaio-Neto, “Blind adaptive constrained reduced-rank parameter estimation based on constant modulus design for cma interference suppression,” IEEE Transactions on Signal Processing, vol. 56, no. 6, pp. 2470–2482, June 2008.
[17] R. de Lamare and R. Sampaio-Neto, “Adaptive Reduced-Root Processing Based on Joint and Iterative Interpolation, Decimation, and Filtering,” IEEE Trans. Signal Process., vol. 57, no. 7, pp. 2503–2514.
[18] R. C. de Lamare and R. Sampaio-Neto, “Reduced-rank spacetime adaptive interference suppression with joint iterative least squares algorithms for spread-spectrum radar,” IEEE Transactions on Vehicular Technology, vol. 59, no. 3, pp. 1217–1228, March 2010.
[19] L. Wang, R. C. de Lamare, and M. Yukawa, “Adaptive reduced-rank constrained constant modulus algorithms based on joint iterative optimization of filters for beamforming,” IEEE Transactions on Signal Processing, vol. 58, no. 6, pp. 2983–2997, June 2010.
[20] R. C. de Lamare, P.-G. Magnusson, and J. A. Tropp, “Finding structure with randomness: Probabilistic algorithms for constructing approximate matrix decompositions,” SIAM Review, vol. 53, no. 2, pp. 217–288, June 2011.
[21] R. C. de Lamare and R. Sampaio-Neto, “Adaptive reduced-rank equalization algorithms based on alternating optimization design techniques for mimo systems,” IEEE Transactions on Vehicular Technology, vol. 60, no. 6, pp. 2482–2494, July 2011.
[22] R. Fa and R. C. D. Lamare, “Reduced-rank stap algorithms using joint iterative optimization of filters,” IEEE Transactions on Aerospace and Electronic Systems, vol. 47, no. 3, pp. 1668–1684, July 2011.
[23] R. C. de Lamare, R. Sampaio-Neto, and M. Haardt, “Blind adaptive constrained constant-modulus reduced-rank interference suppression algorithms based on interpolation and switched decimation,” IEEE Transactions on Signal Processing, vol. 59, no. 2, pp. 681–695, Feb 2011.
[24] S. Li and R. C. de Lamare, “Blind reduced-rank adaptive receivers for ds-uwb systems based on joint iterative optimization and the constrained constant modulus criterion,” IEEE Transactions on Vehicular Technology, vol. 60, no. 6, pp. 2505–2518, July 2011.
[25] N. Song, R. C. de Lamare, M. Haardt, and M. Wolf, “Adaptive widely linear reduced-rank interference suppression based on the multistage wiener filter,” IEEE Transactions on Signal Processing, vol. 60, no. 8, pp. 4003–4016, Aug 2012.
[26] N. Song, W. U. Alokoi, R. C. de Lamare, and M. Haardt, “Adaptive widely linear reduced-rank beamforming based on joint iterative optimization,” IEEE Signal Processing Letters, vol. 21, no. 3, pp. 265–269, March 2014.
[27] L. Wang, R. C. de Lamare, and M. Haardt, “Direction finding algorithms based on joint iterative subspace optimization,” IEEE Transactions on Aerospace and Electronic Systems, vol. 51, no. 1, pp. 270–289, January 2015.
[28] Y. Cai, R. C. de Lamare, B. Champagne, B. Qin, and M. Zhao, “Adaptive reduced-rank receive processing based on minimum symbol error-rate criterion for large-scale multiple-antenna systems,” IEEE Transactions on Communications, vol. 63, no. 11, pp. 4185–4201, Nov 2015.
[29] L. Qiu, Y. Cai, R. C. de Lamare, and M. Zhao, “Reduced-rank doa estimation algorithms based on alternating low-rank decomposition,” IEEE Signal Processing Letters, vol. 23, no. 5, pp. 565–569, May 2016.
[30] B. Victor, K. Bowyer, and S. Sarkar, “An evaluation of face and ear biometrics,” in IJCB 2002, vol. 1, pp. 420–432.
[31] J. Wright, A. Yang, A. Ganesh, S. Sastry, and Y. Ma, “Robust face recognition via sparse representation,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 31, no. 2, pp. 210–227, Feb. 2009.
[32] J. E. Jackson, A User’s Guide to Principal Components, Wiley (1991).
[33] R. Dunia and S. J. Qin, “A Subspace Approach to Multidimensional Fault Identification and Reconstruction,” AICHE J., vol. 44, no. 8, pp. 1813–1831, Aug 1998.
[34] D. D. Gross, Y.-K. Liu, S. Flammia, S. Becker, and J. Eisert, “Quantum state tomography via compressed sensing,” Phys. Rev. Lett., vol. 105, no. 15, p. 150401, Dec 2010.
[35] O. Troyanskaya, M. Cantor, G. Sherlock, P. Brown, T. Hastie, R. Tibshirani, David, D. Botstein, and R. B. Altman, “Missing value estimation methods for dna microarrays,” Bioinformatics, vol. 17, no. 6, pp. 520–525, Jun 2001.
[36] G. H. Golub and C. F. van Loan, Matrix Computations, 3rd ed., Johns Hopkins Univ. Press, Baltimore, MD, (1996).
[37] T. F. Chan, “Rank revealing QR factorizations,” Linear Algebra and its Applications, vol. 88–89, pp. 67–82, Apr 1987.
[38] M. Gu and S. C. Eisenstat, “Efficient Algorithms for Computing a Strong Rank-Revealing QR Factorization,” SIAM J. Sci. Comput., vol. 17, no. 4, pp. 848–869, 1996.
[39] G. Stewart, “An updating algorithm for subspace tracking,” IEEE Trans. Signal Process., vol. 40, no. 6, pp. 1535–1541, Jun 1992.
[40] G. W. Stewart, “Updating a Rank-Revealing ULV Decomposition,” SIAM J. Matrix Anal. Appl., vol. 14, no. 2, pp. 494–499, 1993.
[41] ——, Matrix Algorithms: Volume 1: Basic Decompositions, SIAM, Philadelphia, PA, (1998).
[42] P. C. Hansen, Rank-deficient and Discrete Ill-posed Problems: Numerical Aspects of Linear Inversion. Philadelphia, PA, USA: SIAM, 1998.
[43] J. Demmel, Applied Numerical Linear Algebra, SIAM, (1997).
[94] R. Fa and R. C. D. Lamare, “Multi-branch successive interference
cancellation for mimo spatial multiplexing systems: Design, analysis
and adaptive implementation,” IET Communications, vol. 5, no. 4, pp.
484–494, March 2011.

[95] A. G. D. Uchoa, C. T. Healy, and R. C. de Lamare, “Iterative detection
and decoding algorithms for mimo systems in block-fading channels
using ldpc codes,” IEEE Transactions on Vehicular Technology, vol. 65,
no. 4, pp. 2735–2741, April 2016.

[96] L. Zhang, Y. Cai, R. C. de Lamare, and M. Zhao, “Robust multibranch
tomlinson-harashima precoding design in amplify-and-forward mimo
relay systems,” IEEE Transactions on Communications, vol. 62, no. 10,
pp. 3476–3490, Oct 2014.

[97] P. Li and R. C. de Lamare, “Distributed iterative detection with
reduced message passing for networked mimo cellular systems,” IEEE
Transactions on Vehicular Technology, vol. 63, no. 6, pp. 2947–2954,
July 2014.

[98] R. C. D. Lamare, “Joint iterative power allocation and linear interfer-
ence suppression algorithms for cooperative ds-cdma networks,” IET
Communications, vol. 6, no. 13, pp. 1930–1942, Sept 2012.

[99] T. Peng and R. C. de Lamare, “Adaptive buffer-aided distributed space-
time coding for cooperative wireless networks,” IEEE Transactions on
Communications, vol. 64, no. 5, pp. 1888–1900, May 2016.

[100] W. Zhang, R. C. de Lamare, C. Pan, M. Chen, J. Dai, B. Wu, and
X. Bao, “Widely linear precoding for large-scale mimo with iqi:
Algorithms and performance analysis,” IEEE Transactions on Wireless
Communications, vol. 16, no. 5, pp. 3298–3312, May 2017.

[101] R. C. de Lamare, “Massive mimo systems: Signal processing chal-
genesis and future trends,” URSI Radio Science Bulletin, vol. 2013,
no. 347, pp. 8–20, Dec 2013.

[102] L. T. N. Landau and R. C. de Lamare, “Branch-and-bound precoding
for multiuser mimo systems with 1-bit quantization,” IEEE Wireless
Communications Letters, vol. 6, no. 6, pp. 770–773, Dec 2017.

[103] Z. Shao, R. C. de Lamare, and L. T. N. Landau, “Iterative detection
and decoding for large-scale multiple-antenna systems with 1-bit adcs,”
IEEE Wireless Communications Letters, vol. 7, no. 3, pp. 476–479, June
2018.

[104] J. Gu, R. C. de Lamare, and M. Huemer, “Buffer-aided physical-
layer network coding with optimal linear code designs for cooperative
networks,” IEEE Transactions on Communications, vol. 66, no. 6, pp.
2560–2575, June 2018.

[105] D. Bertsekas, Constrained Optimization and Lagrange Multiplier
Method, Academic Press, (1982).

[106] X. Yuan and J. Yang, “Sparse and low-rank matrix decomposition via
alternating direction methods,” Pacific Journal of Opt., vol. 9, no. 1,
pp. 167–180, 2009.

[107] E. Hale, W. Yin, and Y. Zhang, “Fixed-Point Continuation for \ell_1-
Minimization: Methodology and Convergence,” SIAM J. Opt., vol. 19,
no. 3, pp. 1107–1130, 2008.

[108] Z. Lin, R. Liu, and Z. Su, “Linearized Alternating Direction Method
with Adaptive Penalty for Low-Rank Representation,” in NIPS, no. 1,
2011, pp. 1–9.

[109] K. C. Toh and S. Yun, “An accelerated proximal gradient algorithm
for nuclear norm regularized linear least squares problems,” Pac. J. Opt.
vol. 6, no. 3, pp. 615–640, 2010.

[110] R. M. Larsen, Efficient algorithms for helioseismic inversion, PhD
Thesis, University of Aarhus, Denmark (1998).

[111] L. Li, W. Huang, I.-H. Gu, and Q. Tian, “Statistical Modeling of
Complex Backgrounds for Foreground Object Detection,” IEEE Trans.
Image Process., vol. 13, no. 11, pp. 1459–1472, Nov 2004.

[112] Y. Wang, P.-M. Jodoin, F. Porikli, J. Konrad, Y. Benezech, and P. Ishwar,
“Cidnet 2014: An expanded change detection benchmark dataset,” in
IEEE Conference on Computer Vision and Pattern Recognition
Workshops, 2014, pp. 387–394.

[113] A. S. Georghiades, P. N. Belhumeur, and D. J. Kriegman, “From few
to many: illumination cone models for face recognition under variable
lighting and pose,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 23,
no. 6, pp. 643–660, 2001.