GENERALIZED TENSOR COMPRESSIVE SENSING

Shmuel Friedland, Qun Li, Student Member, IEEE, and Dan Schonfeld, Fellow, IEEE

Abstract—Compressive sensing (CS) has triggered enormous research activity since its first appearance. CS exploits the signal’s sparseness or compressibility in a particular domain and integrates data compression and acquisition, thus allowing exact reconstruction through relatively few non-adaptive linear measurements. While conventional CS theory relies on data representation in the form of vectors, many data types in various applications such as color imaging, video sequences, and multi-sensor networks, are intrinsically represented by higher-order tensors. Application of CS to higher-order data representation is typically performed by conversion of the data to very long vectors that must be measured using very large sampling matrices, thus imposing a huge computational and memory burden. In this paper, we propose Generalized Tensor Compressive Sensing (GTCS)—a unified framework for compressive sensing of higher-order tensors. GTCS offers an efficient means for representation of multidimensional data by providing simultaneous acquisition and compression from all tensor modes. In addition, we propound two reconstruction procedures, a serial method (GTCS-S) and a parallelizable method (GTCS-P). We then compare the performance of the proposed method with Kronecker compressive sensing (KCS) and multi-way compressive sensing (MWCS). We demonstrate experimentally that GTCS outperforms KCS and MWCS in terms of both accuracy and speed.

Index Terms—Compressive sensing, convex optimization, multilinear algebra, higher-order tensor, generalized tensor compressive sensing.

I. INTRODUCTION

Recent literature has witnessed an explosion of interest in sensing that exploits prior structure knowledge in the general form of sparsity, meaning that signals can be represented by only a few coefficients in some domain. Central to much of this recent work is the paradigm of compressive sensing (CS), also known under the terminology of compressed sensing, compressive sampling or compress sensing [1]–[3]. CS theory permits relatively few linear measurements of the signal while still allowing exact reconstruction via nonlinear recovery process. The key idea is that the sparsity helps in isolating the original vector. The first intuitive approach to a reconstruction algorithm consists in searching for the sparsest vector that is consistent with the linear measurements. However, this \( \ell_0 \)-minimization problem is NP-hard in general and thus computationally infeasible. There are essentially two approaches for tractable alternative algorithms. The first is convex relaxation, leading to \( \ell_1 \)-minimization [4], also known as basis pursuit [5], whereas the second constructs greedy algorithms. Besides, in image processing, the use of total-variation minimization which is closely connected to \( \ell_1 \)-minimization first appears in [6] and is widely applied later on. By now basic properties of the measurement matrix which ensure sparse recovery by \( \ell_1 \)-minimization are known: the null space property (NSP) [7] and the restricted isometry property (RIP) [8].

An intrinsic limitation in conventional CS theory is that it relies on data representation in the form of vector. In fact, many data types do not lend themselves to vector data representation. For example, images are intrinsically matrices. As a result, great efforts have been made to extend traditional CS to CS of data in matrix representation. A straightforward implementation of CS on 2D images recasts the 2D problem as traditional 1D CS problem by converting images to long vectors, such as in [9]. However, despite of considerably huge memory and computational burden imposed by long vector data and large sampling matrix, the sparse solutions produced by straightforward \( \ell_1 \)-minimization often incur visually unpleasant, high-frequency oscillations. This is due to the neglect of attributes known to be widely possessed by images, such as smoothness. In [10], instead of seeking sparsity in the transformed domain, they proposed a total variation-based minimization to promote smoothness of the reconstructed image. Later, as an alternative for alleviating the huge computational and memory burden associated with image vectorization, block-based CS (BCS) was proposed in [11]. In BCS, an image is divided into non-overlapping blocks and acquired using an appropriately-sized measurement matrix.

Another direction in the extension of CS to matrix CS generalizes CS concept and outlines a dictionary relating concepts from cardinality minimization to those of rank minimization [12]–[14]. The affine rank minimization problem consists of finding a matrix of minimum rank that satisfies a given set of linear equality constraints. It encompasses commonly seen low-rank matrix completion problem [14] and low-rank matrix approximation problem as special cases. [12] first introduced recovery of the minimum-rank matrix via nuclear norm minimization. [13] generalized the RIP in [8] to matrix case and established the theoretical condition under which the nuclear norm heuristic can be guaranteed to produce the minimum-rank solution.

Real-world signals of practical interest such as color imaging, video sequences and multi-sensor networks, are usually generated by the interaction of multiple factors or multimedia and thus can be intrinsically represented by higher-order tensors. Therefore, the higher-order extension of CS theory for multidimensional data has become an emerging topic. One direction attempts to find the best rank-R tensor...
approximation as a recovery of the original data tensor as in [15], they also proved the existence and uniqueness of the best rank-r tensor approximation in the case of 3rd order tensors under appropriate assumptions. In [16], multi-way compressed sensing (MWCS) for sparse and low-rank tensors suggests a two-step recovery process: fitting a low-rank model in compressed domain, followed by per-mode decompression. However, the performance of MWCS relies highly on the estimation of the tensor rank, which is an NPhard problem. The other direction [17], [18] uses Kronecker product, we use \( \otimes \) to denote tensor product of two vectors. They can be related by \( \mathbf{u} \otimes \mathbf{v} = \mathbf{u} \mathbf{v}^T \). In this paper, we won’t differentiate between outer product and tensor product.

**Mode-i Product** The mode-i product of a tensor \( \mathbf{X} = [x_{\alpha_1 \ldots \alpha_d}] \in \mathbb{R}^{N_1 \times \ldots \times N_d} \) and a matrix \( \mathbf{U} = [u_{j,\alpha_i}] \in \mathbb{R}^{J \times N_i} \) is denoted by \( \mathbf{X} \times_i \mathbf{U} \) and is of size \( N_1 \times \ldots \times N_i-1 \times J \times N_{i+1} \times \ldots \times N_d \). By element, we have \( (\mathbf{X} \times_i \mathbf{U})_{\alpha_1 \ldots \alpha_{i-1} j \alpha_{i+1} \ldots \alpha_d} = \sum_{\alpha_i = 1}^{N_i} x_{\alpha_1 \ldots \alpha_i j \alpha_{i+1} \ldots \alpha_d} \).

**Mode-i Fiber and Mode-i Unfolding** The mode-i fiber of a tensor \( \mathbf{X} = [x_{\alpha_1 \ldots \alpha_d}] \in \mathbb{R}^{N_1 \times \ldots \times N_d} \) is obtained by fixing every index but \( \alpha_i \). The mode-i unfolding \( \mathbf{X}(i) \) of \( \mathbf{X} \) arranges the mode-i fibers to be the columns of the resulting \( N_i \times (N_1 \cdot \ldots \cdot N_i-1 \cdot N_{i+1} \cdot \ldots \cdot N_d) \) matrix.

\[ \mathbf{Y} = \mathbf{X}(i) \begin{bmatrix} U_1 & \cdots & U_d \end{bmatrix} \]

is equivalent to \( \mathbf{Y}(i) = U_i \mathbf{X}(i)(U_d \otimes \cdots \otimes U_{i+1} \otimes U_{i-1} \otimes \cdots \otimes U_1)^T \).

**Core Tucker Decomposition** [20] Let \( \mathbf{X} \in \mathbb{R}^{N_1 \times \ldots \times N_d} \) with mode-i unfolding \( \mathbf{X}(i) \in \mathbb{R}^{N_i \times (N_1 \cdot \ldots \cdot N_{i-1} \cdot N_{i+1} \cdot \ldots \cdot N_d)} \). Denote by \( R_i(\mathbf{X}) \subseteq \mathbb{R}^{N_i} \) the column space of \( \mathbf{X}(i) \) whose rank is \( r_i \). Let \( \mathbf{c}_{i_1, \ldots, i_d} \) be a basis in \( R_i(\mathbf{X}) \). Then the subspace \( \mathbf{V}(\mathbf{X}) := R_1(\mathbf{X}) \circ \cdots \circ R_d(\mathbf{X}) \subseteq \mathbb{R}^{N_1 \times \ldots \times N_d} \) contains \( \mathbf{X} \). Clearly a basis in \( \mathbf{V} \) consists of the vectors \( \mathbf{c}_{i_1, \ldots, i_d} \) where \( i_j \in [r_j] := \{1, \ldots, r_j\} \) and \( j \in [d] \). Hence the core Tucker decomposition of \( \mathbf{X} \) is

\[ \mathbf{X} = \sum_{i_j \in [r_j], j \in [d]} \xi_{i_1, \ldots, i_d} \mathbf{c}_{i_1, \ldots, i_d} \circ \cdots \circ \mathbf{c}_{i_d, \ldots, i_d}. \]  

A special case of core Tucker decomposition is the higher-order singular value decomposition (HOSVD). Every tensor \( \mathbf{X} \in \mathbb{R}^{N_1 \times \ldots \times N_d} \) can be written as

\[ \mathbf{X} = \mathbf{S} \times_1 U_1 \times_2 \cdots \times_d U_d, \]

where \( U_i = [u_{1, \ldots, u_{N_i}}] \) is orthogonal for \( i = 1, \ldots, d \) and \( \mathbf{S} \) is called the core tensor which can be obtained easily by \( \mathbf{S} = \mathbf{X}(1) \times_U 1 \times_2 \cdots \times_d U_d \).

There are many ways to get a weaker decomposition as

\[ \mathbf{X} = \sum_{i=1}^{K} \mathbf{a}_i^{(1)} \circ \cdots \circ \mathbf{a}_i^{(d)} \]

A simple constructive way is as follows. First decompose \( \mathbf{X}(1) \) as \( \mathbf{X}(1) = \sum_{j=1}^{r_1} c_{j,1} \mathbf{g}_{j,1} \) (e.g., by singular value decomposition (SVD)). Now each \( \mathbf{g}_{j,1} \) can be viewed as a tensor of order \( d-1 \). Let \( \mathbf{a}_{i,2} \) denote the core tensor \( \mathbf{X}(2) \) which is the output of the second SD. Continuing in this way we get a decomposition of type [3]. Note that if \( \mathbf{X} \) is s-sparse then each vector in \( R_i(\mathbf{X}) \) is s-sparse and each rank \( r_i \) is at most \( s \). So \( K \leq s^{d-1} \).

**CANDDECOMP/PARAFAC Decomposition** [21] For a tensor \( \mathbf{X} \in \mathbb{R}^{N_1 \times \ldots \times N_d} \), the CANDDECOMP/PARAFAC (CP) decomposition is \( \mathbf{X} \approx \sum_{i=1}^{R} \mathbf{a}_i^{(1)} \circ \cdots \circ \mathbf{a}_i^{(d)} \), where \( \lambda = [\lambda_1, \ldots, \lambda_R]^T \in \mathbb{R}^R \) and \( \mathbf{A}^{(i)} = [\mathbf{a}_i^{(1)} \cdots \mathbf{a}_i^{(d)}] \in \mathbb{R}^{N_i \times R} \) for \( i = 1, \ldots, d \).
B. Compressive sensing

Compressive sensing is one of the ways to encode sparse information. A vector \( x \in \mathbb{R}^N \) is called \( s \)-sparse if it has at most \( s \) nonzero coordinates. The CS measurement protocol measures the signal \( x \) with the measurement matrix \( A \in \mathbb{R}^{m \times N} \) where \( m < N \) and transmits the encoded information \( y \in \mathbb{R}^m \) where \( y = Ax \). The receiver knows \( A \) and attempts to recover \( x \) from \( y \). Since \( m < N \), there are usually infinitely many solutions for such under-constrained problem. However, if \( x \) is known to be sufficiently sparse, then exact recovery of \( x \) is possible, which establishes the fundamental tenet of CS theory. The recovery is done by finding a solution \( z^* \in \mathbb{R}^N \) satisfying

\[
    z^* = \arg\min \{ \|z\|_1, \; Az = y \}. \tag{4}
\]

Such \( z^* \) coincides with \( x \). The following well known result states that each \( s \)-sparse solution can be recovered uniquely if \( A \) satisfies the null space property of order \( s \), denoted as NSPs. That is, if \( Aw = 0, w \in \mathbb{R}^N \setminus \{0\} \), then for any subset \( S \subseteq \{1, \ldots, N\} \) with cardinality \( |S| = s \) it holds that \( \|v_S\|_1 = 0 \) for each vector \( v \) on the index set \( S \) and is set to zero on \( S^c \).

A simple way to generate such \( A \) is to use \( A \) sampled randomly from Gaussian or Bernoulli distributions. Then there exists a universal constant \( c \) such that if

\[
    m \geq 2cs \ln \frac{N}{s} \tag{5}
\]

then the recovery of \( x \) using \( \|z\|_1 \) is successful with probability at least \( 1 - \exp(-\frac{m}{2}) \).

In fact, most signals of practical interest are not really sparse in any domain. Instead, they are only compressible, meaning that in some particular domain, the coefficients, when sorted by magnitude, decay very rapidly, typically like a power law \([22]\). Given a signal \( x \in \mathbb{R}^N \) which can be represented by \( \theta \in \mathbb{R}^N \) in some transformed domain, i.e. \( x = \Phi \theta \), with sorted coefficients such that \( |\theta|_{(1)} \geq \ldots \geq |\theta|_{(N)} \), it obeys that \( |\theta|_{(n)} \leq Rn^{-p} \) for each \( n \in [N] \), where \( 0 < p < 1 \) and \( R \) is some constant. According to \( [22] \), with probability \( 1 \), the solution \( g^* \in \mathbb{R}^N \) to

\[
    g^* = \arg\min \{ \|g\|_1, \; A\Phi g = y \} \tag{6}
\]

is unique. Furthermore, denote by \( x_j \) the recovered signal via \( x_j = \Phi g \), with a very large probability we have the approximation

\[
    \|x - x_j\|_2 \leq CR \left( \frac{m}{\ln N} \right)^{\frac{1}{2} - \frac{1}{p}} \tag{7}
\]

where \( A \in \mathbb{R}^{m \times N} \) is sampled randomly, \( m \) is the number of measurements and \( C \) is some constant. This provides theoretical foundation for CS of compressible signals.

Moreover, when the observation \( y \) is noisy with bounded error \( \epsilon \), an approximation of the signal \( f^* \in \mathbb{R}^N \) also with bounded error can be obtained by solving the following relaxed recovery problem \([23]\).

\[
    f^* = \arg\min \{ \|f\|_1, \; \|Af - y\|_2 \leq \epsilon \}. \tag{7}
\]

Recently, the extension of CS theory for multidimensional signals has become an emerging topic. The objective of our paper is to consider the case where the \( s \)-sparse vector \( x \) is represented as an \( s \)-sparse tensor \( \mathcal{X} = [x_{i_1, \ldots, i_d}] \in \mathbb{R}^{N_1 \times \ldots \times N_d} \).

If we ignore the structure of \( \mathcal{X} \) as a tensor, and view it as a vector of size \( N = N_1 \cdots N_d \), clearly, we can transmit \( \mathcal{X} \) as \( x \) by using \( y = Ax \). If we use a random \( A \) as described above, we need \( m \) to be at least of order

\[
    m \geq 2cs(- \ln s + \sum_{i=1}^{d} \ln N_i). \tag{8}
\]

In \([18]\), Kronecker compressive sensing (KCS) constructs \( A \) from Kronecker product \( A := U_1 \otimes \ldots \otimes U_d \), where \( U_i \in \mathbb{R}^{m_i \times N_i} \) for \( i = 1, \ldots, d \) and each \( U_i \) has NSPs property. Then \( x \) is recovered uniquely from \( y = Ax \) by \( \ell_1 \)-minimization in \( \|z\|_1 \).

In this paper, we analyze the compression and reconstruction of tensor \( \mathcal{X} \) from the tensor \( y = \mathcal{X} \odot U_1 \times U_2 \times \cdots \times U_d \in \mathbb{R}^{m_1 \times m_2 \times \cdots \times m_d} \) using a sequence of \( \ell_1 \)-minimizations similar to the minimization in \( \|z\|_1 \).

More specifically, we propose two reconstruction procedures, a serial method (GTCS-S) and a parallelizable method (GTCS-P) in terms of recovery of each tensor mode. One advantage of our method is that our recovery problems are in terms of each \( U_i \), which is much smaller comparing with the recovery related to \( A \) as given by the minimization in \( \|z\|_1 \). This means that the amount of computations of our method is much less than that given by \( \|z\|_1 \).

If we choose our matrices \( U_i \) at random then we have the condition

\[
    m_i \geq 2cs \ln \frac{N_i}{s}, \quad i = 1, \ldots, d. \tag{9}
\]

A similar idea to GTCS-P, namely multi-way compressed sensing (MWCS) \([16]\) for sparse and low-rank tensors, also suggests a two-step recovery process: fitting a low-rank model \( \mathcal{X} \) in compressed domain, followed by per-mode decompression. However, the performance of MWCS relies highly on the estimation of the tensor rank, which is an NP-hard problem. The proposed GTCS manages to get rid of tensor rank estimation and thus considerably reduces the computational complexity in comparison to MWCS.

III. GENERALIZED TENSOR COMPRESSIVE SENSING

A. Generalized tensor compressive sensing with serial recovery (GTCS-S)

We first discuss our method for matrices, i.e. \( d = 2 \) and then for tensors \( d \geq 3 \).

**Theorem 3.1:** Let \( X = [x_{ij}] \in \mathbb{R}^{N_1 \times N_2} \) be \( s \)-sparse. Let \( U_i \in \mathbb{R}^{m_i \times N_i} \) and assume that \( U_i \) has NSPs property for \( i = 1, 2 \). Define

\[
    Y = [y_{pq}] = U_1 X U_2^T \in \mathbb{R}^{m_1 \times m_2}. \tag{10}
\]

Then \( X \) can be recovered uniquely using the following procedure. Let \( y_1, \ldots, y_{m_2} \in \mathbb{R}^{m_1} \) be the columns of \( Y \). Let \( z_i^* \in \mathbb{R}^{N_1} \) be a solution of

\[
    z_i^* = \min \{ \|z_i\|_1, \; U_i z_i = y_i \}, \quad i = 1, \ldots, m_2. \tag{11}
\]

Then each \( z_i^* \) is unique and \( s \)-sparse. Let \( Z \in \mathbb{R}^{N_1 \times m_2} \) be the matrix with columns \( z_1^*, \ldots, z_{m_2}^* \). Let \( w_1^T, \ldots, w_{N_2}^T \) be the
rows of $Z$. Then the $j^{th}$ row of $X$ is the solution $u_j^* \in \mathbb{R}^{N_2}$ of
\[
u_j^* = \min\{\|u_j\|_1, U_2u_j = w_j\}, \quad j = 1, \ldots, N_1. \tag{12}
\]

**Proof:** Let $Z = XU_2^T \in \mathbb{R}^{N_1 \times m_2}$. Assume that $z_{i1}, \ldots, z_{im_2}$ are the columns of $Z$. Note that $z_i^*$ is a linear combination of the $N_2$ columns of $X$, given by the $i^{th}$ row of $U_2$. Since $X$ is $s$-sparse, $z_i^*$ has at most $s$ nonzero entries. Note that $Y = U_1Z$, it follows that $y_i = U_1z_i^*$. Since $U_1$ has NSP$_s$, we deduce the equality (11). Observe next that $Z^T = U_2X^T$. Hence the column $w_j$ of $Z^T$ is $w_j = U_2u_j^*$. Since $X$ is $s$-sparse, each $u_j^*$ is $s$-sparse. The assumption that $U_2$ has NSP$_s$ property implies (12). This completes the proof. ■

If we choose $U_1, U_2$ to be random, then we need the assumption (9). We now make the following observation. Suppose we know that each column of $XU_2^T$ is $s$-sparse and each row of $X$ is $s_d$-sparse. Then from the proof of Theorem 3.1, it follows that we can recover $X$, on the assumption that $U_1$ has NSP$_{s_1}$ and $U_2$ has NSP$_{s_2}$.

**Theorem 3.2 (GTCS-S):** Let $\mathcal{X} = [x_{i1}, \ldots, x_{id}] \in \mathbb{R}^{N_1 \times \ldots \times N_d}$ be $s$-sparse. Let $U_i \in \mathbb{R}^{m_i \times N_i}$ and assume that $U_i$ has NSP$_{s_i}$ property for $i = 1, \ldots, N$. Define
\[
Y = [y_{j1}, \ldots, y_{jd}] = X \times U_1 \times \ldots \times U_d U \in \mathbb{R}^{m_1 \times \ldots \times m_d}. \tag{13}
\]

Then $\mathcal{X}$ can be recovered uniquely using the following recursive procedure. Unfold $Y$ in mode 1,
\[
Y_{(1)} = U_1X_{(1)}[\otimes^2_k = u_k] \in \mathbb{R}^{m_1 \times (m_2 \cdots m_d)}. \tag{14}
\]

Let $y_{11}, \ldots, y_{m_2 \cdots m_d}$ be the columns of $Y_{(1)}$. Then $y_{1i} = U_1 z_{1i}$ where each $z_{1i} \in \mathbb{R}^{N_1}$ is $s$-sparse. Recover each $z_{1i}$ using (4). Let $Z = X \times U_2 U_3 \times \ldots \times U_d U \in \mathbb{R}^{m_1 \times m_2 \cdots m_d}$ with its mode-$1$ fibers being $z_{i1}, \ldots, z_{im_2 \cdots m_d}$. Unfold $Z$ in mode 2,
\[
Y_{(2)} = U_2X_{(2)}[\otimes^3_k = u_k \otimes I] \in \mathbb{R}^{m_3 \times (N_1 m_2 \cdots m_d)}. \tag{15}
\]

Let $w_{11}, \ldots, w_{N_1 \cdots m_d}$ be the columns of $Z_{(2)}$. Then $w_{ij} = U_2 v_{ij}$ where each $v_{ij} \in \mathbb{R}^{N_2}$ is $s$-sparse. Recover each $v_{ij}$ using (4). Continue the above procedure for mode 3, $\ldots$, $d$ and $\mathcal{X}$ can be reconstructed in series.

The proof follows directly that of Theorem 3.1 and hence is skipped here.

As for matrices, assume mode-$i$ fibers of $X$ are $s_{i+1} U_{i+1} \times \ldots \times U_d$ is $s_i$-sparse for $i = 1, \ldots, d - 1$ and mode-$d$ fibers of $X$ is $s_d$-sparse, then we can relax the condition such that $U_{i}$ only has to satisfy NSP$_{s_i}$ for $i = 1, \ldots, d$.

**B. Generalized tensor compressive sensing with parallelizable recovery (GTCS-P)**

**Theorem 3.3 (GTCS-P):** Let $\mathcal{X} = [x_{i1}, \ldots, x_{id}] \in \mathbb{R}^{N_1 \times \ldots \times N_d}$ be $s$-sparse. Let $U_i \in \mathbb{R}^{m_i \times N_i}$ and assume that $U_i$ has NSP$_{s_i}$ property for $i = 1, \ldots, d$. Define $Y$ as in (13), then $\mathcal{X}$ can be recovered uniquely using the following procedure. Consider a decomposition of $Y$ such that,
\[
Y = \sum_{i=1}^{K} b_i^{(1)} \circ \cdots \circ b_i^{(d)}, \quad b_i^{(j)} \in R_j(Y) \subseteq U_j R_j(X),
\]
\[
j \in [d]. \tag{14}
\]

Let $w_i^{(j)*} \in R_j(X) \subseteq \mathbb{R}^{N_j}$ be a solution of
\[
w_i^{(j)*} = \min\{\|w_i^{(j)}\|_1, U_j w_i^{(j)} = b_i^{(j)}\}, \quad i = 1, \ldots, K, j = 1, \ldots, d. \tag{15}
\]

Thus each $w_i^{(j)*}$ is unique and $s$-sparse. Then,
\[
X = \sum_{i=1}^{K} w_i^{(1)} \circ \cdots \circ w_i^{(d)}, \quad w_i^{(j)} \in R_j(X), j \in [d]. \tag{16}
\]

**Proof:** Since $\mathcal{X}$ is $s$- sparse, each vector in $R_j(X)$ is $s$- sparse. So if each $U_i$ has NSP$s_i$, we can get a unique $s$-sparse vector $w_i^{(j)} \in R_j(X)$ such that $U_j w_i^{(j)} = b_i^{(j)}$ and obtain a tensor
\[
Z = \sum_{i=1}^{K} w_i^{(1)} \circ \cdots \circ w_i^{(d)}, \quad w_i^{(j)} \in R_j(X), j \in [d]. \tag{17}
\]

We have
\[
(X - Z) \times U_1 \times \ldots \times U_d U_d = 0. \tag{18}
\]

We next show $Z = \mathcal{X}$. For that we are going to assume a slightly more general scenario. Namely each $R_j(X) \subseteq V_j \subseteq \mathbb{R}^{N_j}$ such that each nonzero vector in $V_j$ is $s$-sparse. So $R_j(Y) \subseteq U_j R_j(X) \subseteq U_j V_j$ for $j \in [d]$. Assume $\mathcal{X} \neq Z$. We next prove by induction on mode $m$ which yields contradiction to this assumption.

Suppose we have that
\[
(X - Z) \times U_m \times \ldots \times U_d U_d = 0. \tag{19}
\]

Unfold $\mathcal{X}$ and $Z$ in mode $m$, the column spaces of $X(m)$ and $Z(m)$ are contained in $V_m$ while the row spaces are contained in $V_m := V_1 \circ \cdots \circ V_{m-1} \circ V_{m+1} \circ \cdots \circ V_d$. Since we assume that $\mathcal{X} \neq Z$, thus $X(m) - Z(m) \neq 0$. Then $X(m) - Z(m) = \sum_{i=1}^{p} u_i v_i^T$ where rank $(X(m) - Z(m)) = p$, $u_1, \ldots, u_p \in V_m, v_1, \ldots, v_p \in V_m$ are linearly independent. Observe next that $U_m u_1, \ldots, U_m u_p$ are linearly independent. To show this, $0 = \sum_{i=1}^{p} a_i U_m u_i = U_m u, \quad u = \sum_{i=1}^{p} a_i u_i \in V_m$. Since $u$ is $s$-sparse and $U_m$ has NSP$_s$ we deduce that $u = 0$, so $a_1 = \ldots = a_p = 0$.

Since $(X - Z) \times U_m \times \ldots \times U_d U_d = 0$, it is equivalent to
\[
0 = U_m (X(m) - Z(m))(U_d \circ \cdots \circ U_{m+1} \circ I)^T = U_m (X(m) - Z(m)) U_m^T = \sum_{i=1}^{p} (U_m u_i)^T U_m^T = 0.
\]

Since $(X - Z) \times U_m \times \ldots \times U_d U_d = 0$, it is equivalent to
\[
0 = U_m (X(m) - Z(m))(U_d \circ \cdots \circ U_{m+1} \circ I)^T = U_m (X(m) - Z(m)) U_m^T = \sum_{i=1}^{p} (U_m u_i)^T U_m^T = 0.
\]

Fold back to tensor form, this is equivalent to
\[
(X - Z) \times I_m \times U_{m+1} \times \ldots \times U_d U_d = (X - Z) \times U_m \times \ldots \times U_d U_d = 0,
\]
where $I_m$ is an $N_m \times N_m$ identity matrix. Hence we show by mode-$m$ unfolding, we are able to replace $U_m$ with $I_m$ in (19).
Similarly, when \( m = 1 \), we can replace \( U_1 \) with \( I_1 \) in (18). Continue in this way replacing \( U_m \) with \( I_m \) for \( 2 \leq m \leq d \), we will arrive at

\[
(X - Z) \times_1 U_1 \times \ldots \times_d U_d = (X - Z) \times_1 I_1 \times \ldots \times_d I_d = X - Z = 0.
\]

This brings contradiction to our assumption that \( X \neq Z \). Thus, it proves that \( X = Z \). This completes the proof.

In fact, if all vectors \( \in R_s(X) \) are \( s_t \)-sparse, then \( U_i \) only has to satisfy NSP\(_s\). The above recovery procedure consists of two stages: the decomposition stage and the reconstruction stage where the latter for each tensor mode can be implemented in parallel.

Observe that if we are satisfied with recovering a rank-\( R \) approximation of \( X \), we only need to fit a rank-\( R \) approximation of \( Y = \sum_{r=1}^{R} b_{r}^{(1)} \otimes \ldots \otimes b_{r}^{(d)} \) (e.g. by CP decomposition) and subsequently recover each \( w_r^{(j)} \) for \( j = 1, \ldots, d \) and \( r = 1, \ldots, R \). Then a rank-\( R \) approximation of \( X \) would be \( \hat{X} = \sum_{r=1}^{R} w_r^{(1)} \otimes \ldots \otimes w_r^{(d)} \).

IV. EXPERIMENTAL RESULTS

We experimentally demonstrate the performance of GTCS methods on sparse and compressible images and video sequences. In [19], KCS has shown its outstanding performance for compression of multidimensional signals in comparison with several other methods such as independent measurements and partitioned measurements. Therefore, we choose KCS as a comparison to the proposed GTCS methods. Another method we compare our method to is MWCS. Our experiments use the \( \ell_1 \)-minimization solvers from [24]. We set the same threshold to determine the termination of \( \ell_1 \)-minimization in all subsequent experiments. All simulations are executed on a desktop with 2.4 GHz Intel Core i5 CPU and 8GB RAM.

A. Sparse image representation

As shown in Figure 3, we use the cameraman image, which is a 64 × 64 image (\( N = 4096 \) pixels). Its columns are 14-sparse and rows are 18-sparse. The image itself is 178-sparse. We let the number of measurements evenly split among the two modes, that is, for each mode, the randomly constructed Gaussian matrix \( U \) is of size \( K \times 64 \). Therefore the KCS measurement matrix \( U \otimes U \) is of size \( K^2 \times 4096 \). Thus the total number of samples is \( K^2 \). We define the normalized number of samples to be \( \frac{K^2}{N} \). For GTCS, both the serial recovery method GTCS-S and the parallelizable recovery method GTCS-P are implemented. In the matrix case, GTCS-P coincides with MWCS and we simply conduct SVD on the compressed image in the decomposition stage of GTCS-P. Although the reconstruction stage of GTCS-P is parallelizable, we here recover each vector in series. We comprehensively examine the performance of all the above methods by varying \( K \) from 1 to 45.

Figure 1(a) and 1(b) compare the peak signal to noise ratio (PSNR) and the recovery time respectively. Both KCS and GTCS methods achieve PSNR over 30dB when \( K = 39 \). As \( K \) increases, GTCS-S tends to outperform KCS in terms of both accuracy and efficiency. Although PSNR of GTCS-P is the lowest among the three methods, it is most time efficient. Moreover, with parallelization of GTCS-P, the recovery procedure can be further accelerated considerably. The reconstructed images when \( K = 38 \), that is, using 0.35 normalized number of samples, are shown in Figure 2(b), 2(c) and 2(d). Though GTCS-P usually recovers much noisier image, it is good at recovering the non-zero structure of the original image.

B. Compressible image representation

As shown in Figure 4(a), the cameraman image is resized to 64 × 64 (\( N = 4096 \) pixels). The image itself is non-sparse.
However, in some transformed domain, such as discrete cosine transformation (DCT) domain in this case, the magnitudes of the coefficients decay by power law in both directions (see Figure 3(b)), thus are compressible. We let the number of measurements evenly split among the two modes. Again, in matrix data case, MWCS concurs with GTCS-P. We exhaustively vary $K$ from 1 to 64.

Figure 4(a) and 4(b) compare the PSNR and the recovery time respectively. Unlike the sparse image case, GTCS-P shows outstanding performance in comparison with all other methods, in terms of both accuracy and speed, followed by KCS and then GTCS-S. The reconstructed images when $K = 46$, using 0.51 normalized number of samples and when $K = 63$, using 0.96 normalized number of samples are shown in Figure 5.

![Figure 3](Image)

Fig. 3. The original cameraman image (resized to $64 \times 64$ pixels) in space domain and DCT domain.

![Figure 4](Images)

Fig. 4. PSNR and reconstruction time comparison on compressible image.

![Figure 5](Images)

Fig. 5. Reconstructed cameraman images. In this two-dimensional case, GTCS-P is equivalent to MWCS.

C. Sparse video representation

We next compare the performance of GTCS and KCS on video data. Each frame of the video sequence is preprocessed to have size $24 \times 24$ and we choose the first 24 frames. The video data together is represented by a $24 \times 24 \times 24$ tensor and has $N = 13824$ voxels in total. To obtain a sparse tensor, we manually keep only $6 \times 6 \times 6$ nonzero entries in the center of the video tensor data and the rest are set to zero. Therefore, the video tensor itself is $216$-sparse and its mode-$i$ fibers are all $6$-sparse for $i = 1, \ldots, 3$. The randomly constructed Gaussian measurement matrix for each mode is now of size $K \times 24$ and the total number of samples is $K^3$. Therefore, the normalized number of samples is $\frac{K^3}{K^3}$. In the decomposition stage of GTCS-P, we employ a decomposition described in Section II-A to obtain a weaker form of the core Tucker decomposition. We vary $K$ from 1 to 13.

Figure 6(a) depicts PSNR of the first non-zero frame recovered by all three methods. Please note that the PSNR values of different video frames recovered by the same method are the same. All methods exhibit an abrupt increase in PSNR at $K = 10$ (using 0.07 normalized number of samples). Also, Figure 6(b) summarizes the recovery time. In comparison to the image case, the time advantage of GTCS becomes more important in the reconstruction of higher-order tensor data.

We specifically look into the recovered frames of all three methods when $K = 12$. Since all the recovered frames achieve a PSNR higher than 40 dB, it is hard to visually observe any difference compared to the original video frame. Therefore, we display the reconstruction error image defined as the absolute difference between the reconstructed image and the original image. Figures 7(a)–(c) visualize the reconstruction errors of all three methods. Compared to KCS, GTCS-S achieves much lower reconstruction error using much less time.

To compare the performance of GTCS-P with MWCS, we examine MWCS with various tensor rank estimations and
D. Compressible video representation

We finally examine the performance of GTCS, KCS and MWCS on compressible video data. Each frame of the video sequence is preprocessed to have size $24 \times 24$ and we choose the first 24 frames. The video data together is represented by a $24 \times 24 \times 24$ tensor. The video itself is non-sparse, yet compressible in three-dimensional DCT domain. In the decomposition stage of GTCS-P, we employ a decomposition described in Section II-A to obtain a weaker form of the core Tucker decomposition and denote this method by GTCS-P (CT). We also test the performance of GTCS-P by using HOSVD in the decomposition stage and denote this method by GTCS-P (HOSVD) hereby. $K$ varies from 1 to 21. Note that in GTCS-S, more and more noise will be induced as the recovery by mode continues and the recovery method by $\ell_1$-minimization using (4) would be inappropriate or even has no solution at certain stage. In our experiment, GTCS-S by (4) works for $K$ from 1 to 7. To use GTCS-S for $K = 8$ and higher, relaxed recovery (7) could be employed for reconstruction. Figure 9(a) and Figure 9(b) depict PSNR and reconstruction time of all methods up to $K = 7$. For $K = 8$ to 21, the results are shown in Figure 9(c) and Figure 9(d).

We specifically look into the recovered frames of all methods when $K = 17$ and $K = 21$. Recovered frames 1, 5, 9, 13, 17, 21 (originally as shown in Figure 10) are depicted as an example in Figure 11 and Figure 12 respectively.

As shown in Figure 13(a), the performance of MWCS relies highly on the estimation of the tensor rank. We examine the performance of MWCS with various rank estimations. Experimental results demonstrate that GTCS outperforms MWCS not only in speed, but also in accuracy.

V. CONCLUSION

Real-world signals of practical interest such as color imaging, video sequences and multi-sensor networks, are usually generated by the interaction of multiple factors or multimedia and thus can be intrinsically represented by higher-order tensors. Therefore, the extension of CS theory for multidimensional signals has become an emerging topic. Existing
methods include Kronecker compressive sensing (KCS) for sparse tensors and multi-way compressive sensing (MWCS) for sparse and low-rank tensors. KCS uses Kronecker product matrices to act as sparsifying bases as well as to represent the measurement protocols used in distributed settings. However, due to the vectorization of multidimensional signals, the recovery procedure is rather time consuming and not applicable in practice. Although MWCS achieves more efficient reconstruction by fitting a low-rank model in compressed domain, followed by per-mode decompression, its performance relies highly on tensor rank estimation, which is NP-hard. We propose Generalized Tensor Compressive Sensing (GTCS)-a unified framework for compressive sensing of sparse higher-order tensors. In addition, we propose two reconstruction procedures, a serial method (GTCS-S) and a parallelizable method (GTCS-P). We then compare the performance of GTCS with KCS and MWCS experimentally on various types of data including sparse image, compressible image, sparse video and compressible video. Experimental results show that GTCS outperforms KCS and MWCS in terms of both accuracy and efficiency. The advantage of our method mainly comes from two aspects. First, compared to KCS, our recovery problems are in terms of each tensor mode, which is much smaller comparing with the vectorization of all tensor modes. Such advantage becomes more important as the order of the data increases. Secondly, unlike MWCS, GTCS manages to
get rid of tensor rank estimation, which considerably reduces the computational complexity and at the same time improves the reconstruction accuracy.

REFERENCES

[1] E. Candès, J. Romberg, and T. Tao, “Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information,” Information Theory, IEEE Transactions on, vol. 52, no. 2, pp. 489–509, feb. 2006.

[2] E. Candès and T. Tao, “Near-optimal signal recovery from random projections: Universal encoding strategies?” Information Theory, IEEE Transactions on, vol. 52, no. 12, pp. 5406–5425, dec. 2006.

[3] D. Donoho, “Compressed sensing,” Information Theory, IEEE Transactions on, vol. 52, no. 4, pp. 1289–1306, apr. 2006.

[4] D. Donoho and B. Logan, “Signal recovery and the large sieve,” SIAM Journal on Applied Mathematics, vol. 52, no. 2, pp. 577–591, 1992.

[5] S. Chen, D. Donoho, and M. Saunders, “Atomic decomposition by basis pursuit,” SIAM Journal on Scientific Computing, vol. 20, pp. 33–61, 1996.

[6] L. Rudin, S. Osher, and E. Fatemi, “Nonlinear total variation based noise removal algorithms,” Physica D: Nonlinear Phenomena, vol. 60, no. 1-4, pp. 259–268, 1992.

[7] A. Cohen, W. Dahmen, and R. Devore, “Compressed sensing and best k-term approximation,” J. Amer. Math. Soc, pp. 211–231, 2009.

[8] E. Candès, “The restricted isometry property and its implications for compressed sensing,” Comptes Rendus Mathematique, vol. 346, no. 9-10, pp. 589–592, 2008.

[9] M. Wakin, J. Laska, M. Duarte, D. Baron, S. Sarvotham, D. Takhar, K. Kelly, and R. Baraniuk, “An architecture for compressive imaging,” in IEEE ICIP, oct. 2006, pp. 1273–1276.

[10] E. J. Candes, J. K. Romberg, and T. Tao, “Stable signal recovery from incomplete and inaccurate measurements,” Comm. Pure Appl. Math., vol. 59, no. 8, pp. 1207–1223, aug. 2006.

[11] L. Gan, “Block compressed sensing of natural images,” in Digital Signal Processing, 2007 15th International Conference on, jul. 2007, pp. 403–406.

[12] M. Fazel, “Matrix rank minimization with applications,” Ph.D. dissertation, 2002.

[13] B. Recht, M. Fazel, and P. Parrilo, “Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization,” SIAM Rev., vol. 52, no. 3, pp. 471–501, aug. 2010.

[14] E. Candès and B. Recht, “Exact matrix completion via convex optimization,” Commun. ACM, vol. 55, no. 6, pp. 111–119, jun. 2012.

[15] L. Lim and P. Comon, “Multiarray signal processing: Tensor decomposition meets compressed sensing,” CoRR, vol. abs/1002.4935, 2010.

[16] N. Sidiropoulos and A. Kyrillidis, “Multi-way compressed sensing for sparse low-rank tensors,” Signal Processing Letters, IEEE, vol. 19, no. 11, pp. 757–760, 2012.

[17] M. Duarte and R. Baraniuk, “Kronecker product matrices for compressive sensing,” in IEEE ICASSP 2010, mar. 2010, pp. 3650–3653.

[18] ——, “Kronecker compressive sensing,” Image Processing, IEEE Transactions on, vol. 21, no. 2, pp. 494–504, feb. 2012.

[19] Q. Li, D. Schonfeld, and S. Friedland, “Generalized tensor compressive sensing,” IEEE International Conference on Multimedia & Expo, 2013.

[20] L. R. Tucker, “The extension of factor analysis to three-dimensional
Fig. 9. PSNR and reconstruction time comparison on susie video.

Fig. 13. PSNR comparison of GTCS-P with MWCS on compressible video when $K = 17$, using 0.36 normalized number of samples and $K = 21$, using 0.67 normalized number of samples. The highest PSNR of MWCS with estimated tensor rank varying from 1 to 24 appears when Rank = 4 and Rank = 7, respectively.

Shmuel Friedland received all his degrees in Mathematics from Israel Institute of Technology (IIT), Haifa, Israel: B.Sc in 1967, M.Sc. in 1969, D.Sc. in 1971. He held Postdoc positions in Weizmann Institute of Science, Israel; Stanford University; IAS, Princeton. From 1975 to 1983, he was a member of Institute of Mathematics, Hebrew U., Jerusalem, and was promoted to the rank of Professor in 1982. Since 1985 he is a Professor at University of Illinois at Chicago. He was a visiting Professor at University of Wisconsin; Madison; IMA, Minneapolis; IHES, Bures-sur-Yvette; IIT, Haifa; Berlin Mathematical School, Berlin. Friedland contributed to the following fields of mathematics: one complex variable, matrix and operator theory, numerical linear algebra, combinatorics, ergodic theory and dynamical systems, mathematical physics, mathematical biology, algebraic geometry. He authored about 170 papers, with many known coauthors, including one Fields Medal winner. He received the first Hans Schneider prize in Linear Algebra, jointly with M. Fiedler and I. Gohberg, in 1993. He was awarded recently a smoked salmon for solving the set-theoretic version of the salmon problem: \url{http://www.dms.uaf.edu/~eallman}. For more details on Friedland vita and research, see \url{http://www.math.uic.edu/~friedlan}.

Qun Li is currently a PhD student majoring in Electrical Engineering at the University of Illinois at Chicago (UIC), U.S.A. She received the B.S. degree in Communications Engineering from Nanjing University of Science and Technology, China, in 2009 and the M.S. degree in Electrical Engineering from UIC in 2012. Her research interests include machine learning, computer vision and higher-order data analysis.

Dan Schonfeld received the B.S. degree in electrical engineering and computer science from the University of California, Berkeley, and the M.S. and Ph.D. degrees in electrical and computer engineering from the Johns Hopkins University, Baltimore, MD, in 1986, 1988, and 1990, respectively. He joined University of Illinois at Chicago in 1990, where he is currently a Professor in the Departments of Electrical and Computer Engineering, Computer Science, and Bioengineering, and Co-Director of the Multimedia Communications Laboratory (MCL). He has authored over 170 technical papers in various journals and conferences.

matrices,” in Contributions to mathematical psychology, H. Gulliksen and N. Frederiksen, Eds. New York: Holt, Rinehart and Winston, 1964.
[21] T. G. Kolda and B. W. Bader, “Tensor decompositions and applications,” SIAM REVIEW, vol. 51, no. 3, pp. 455–500, 2009.
[22] D. Donoho, M. Vetterli, R. A. DeVore, and I. Daubechies, “Data compression and harmonic analysis,” IEEE Trans. Inform. Theory, vol. 44, pp. 2435–2476, 1998.
[23] F. Santosa and W. Symes, “Linear inversion of band-limited reflection seismograms,” SIAM Journal on Scientific and Statistical Computing, vol. 7, no. 4, pp. 1207–1230, 1986.
[24] E. J. Candes and J. K. Romberg, “The $l_1$ magic toolbox,” Available online: \url{http://www.l1-magic.org}.