Low-rank Tensor Recovery via Iterative Hard Thresholding

Holger Rauhut
RWTH Aachen University
Templergraben 55,
52056 Aachen, Germany
Email: rauhut@mathc.rwth-aachen.de

Reinhold Schneider
Technische Universität Berlin
Straße des 17. Juni 136,
10623 Berlin, Germany
Email: schneidr@math.tu-berlin.de

Željka Stojanac
Hausdorff Center for Mathematics
Institut für Numerische Simulation
Endenicher Allee 62,
53115 Bonn, Germany
Email: zeljka.stojanac@hcm.uni-bonn.de

***Abstract*—We study recovery of low-rank tensors from a small number of measurements. A version of the iterative hard thresholding algorithm (TIHT) for the higher order singular value decomposition (HOSVD) is introduced. As a first step towards the analysis of the algorithm, we define a corresponding tensor restricted isometry property (HOSVD-TRIP) and show that Gaussian and Bernoulli random measurement ensembles satisfy it with high probability.**

I. Introduction and Motivation

Low-rank recovery builds on ideas from the theory of compressive sensing which predicts that sparse vectors can be recovered efficiently from incomplete measurements via efficient algorithms including ℓ_1-minimization. Given a matrix $\mathbf{X} \in \mathbb{R}^{n_1 \times n_2}$ of rank at most $r \ll \min\{n_1, n_2\}$, the goal of the low-rank matrix recovery is to reconstruct $\mathbf{X}$ from linear measurements $\mathbf{y} = \mathcal{A}(\mathbf{X})$, where $\mathcal{A} : \mathbb{R}^{n_1 \times n_2} \to \mathbb{R}^m$ with $m \ll n_1 n_2$. Unfortunately, the natural approach of finding the solution of the optimization problem

$$\min_{\mathbf{Z} \in \mathbb{R}^{n_1 \times n_2}} \operatorname{rank}(\mathbf{Z}) \quad \text{s.t.} \quad \mathcal{A}(\mathbf{Z}) = \mathbf{y}, \tag{1}$$

is NP-hard. Nevertheless, it has been shown that solving the convex optimization problem

$$\min_{\mathbf{Z} \in \mathbb{R}^{n_1 \times n_2}} \|\mathbf{Z}\|_* \quad \text{s.t.} \quad \mathcal{A}(\mathbf{Z}) = \mathbf{y}, \tag{2}$$

reconstructs $\mathbf{X}$ exactly under suitable conditions on $\mathcal{A}$. The required number of measurements scales as $m \geq Cr \max\{n_1, n_2\}$ for Gaussian measurement ensembles [11], [2].

In this note, we go one step further and consider the recovery of low-rank tensors $\mathbf{X} \in \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_d}$ from a small number of linear measurements $\mathbf{y} = \mathcal{A}(\mathbf{X})$, where $\mathcal{A} : \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_d} \to \mathbb{R}^m$ and $m \ll n_1 n_2 \cdots n_d$. Again, we are led to consider the rank-minimization problem

$$\min_{\mathbf{Z} \in \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_d}} \operatorname{rank}(\mathbf{Z}) \quad \text{s.t.} \quad \mathbf{y} = \mathcal{A}(\mathbf{Z}). \tag{3}$$

Different notions of the tensor rank have been introduced, which correspond to different decompositions. One possibility is to define the rank of an arbitrary tensor $\mathbf{X} \in \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_d}$, analogously to the matrix rank, as the smallest number of rank one tensors that sum up to $\mathbf{X}$, where a rank one tensor is of the form $\mathbf{A} = \mathbf{u}_1 \otimes \mathbf{u}_2 \otimes \cdots \otimes \mathbf{u}_d$ or elementwise $\mathbf{A}(i_1, i_2, \ldots, i_d) = \mathbf{u}_1(i_1)\, \mathbf{u}_2(i_2) \cdots \mathbf{u}_d(i_d)$. Expectedly, the problem (3) is NP hard [8]. Although it is possible to define an analog of the nuclear norm $\|\cdot\|_*$ for tensors and consider the minimization problem

$$\min_{\mathbf{Z} \in \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_d}} \|\mathbf{Z}\|_* \quad \text{s.t.} \quad \mathbf{y} = \mathcal{A}(\mathbf{Z}),$$

the computation of $\|\cdot\|_*$ and thereby this problem is NP hard [8] as well for tensors of order $d \geq 3$.

The previous approaches to low-rank tensor recovery and tensor completion [3] and [9] are based on the sum of nuclear norms of matrices obtained as unfoldings of the tensor (see below for the notion of unfolding). Only numerical experiments have been performed in these papers and at least from a theoretical point of view, we do not believe this to be the right approach since the tensor structure is lost.

We consider a generalization of the singular value decomposition, called HOSVD (higher order singular value decomposition). This decomposition is used in e.g. data mining for handwritten digit classification [12], in signal processing to extend Wiener filters [10], in computer vision [13] and in other applications.

As a proxy for (3) we propose an extension of the IHT algorithm already used for recovery of sparse signals [1] and low-rank matrices [5]. The tensor iterative hard thresholding algorithm (TIHT algorithm) is presented in Section IV. In the last section, we introduce the tensor restricted isometry property (HOSVD-TRIP) and also show that random linear mappings satisfy the HOSVD-TRIP with high probability, under suitable conditions.

The version for the tensor train decomposition (TT decomposition) and hierarchical tucker format (HT decomposition) will be treated in a journal paper in preparation.

II. Notation

We work with tensors $\mathbf{X} \in \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_d}$ of order d. With $\mathbf{X}_{i_k = p}$, for all $p \in [n_k]$, where $[n_k] = \{1, 2, \ldots, n_k\}$, we denote the $(d-1)$-dimensional tensor (called subtensor) that is obtained by fixing the k-th component of a tensor $\mathbf{X}$ to p i.e., $\mathbf{X}_{i_k = p}(i_1, \ldots, i_{k-1}, i_{k+1}, \ldots, i_d) = \mathbf{X}(i_1, \ldots, i_{k-1}, p, i_{k+1}, \ldots, i_d)$, for all $i_l \in [n_l]$ and for all $l \in [d] \setminus \{k\}$. A matrix obtained by taking the first r_k

columns of the matrix $\mathbf{U}$ is denoted by $\mathbf{U}(:,[r_k])$. Similary, $\mathbf{S}([r_1],[r_2],\ldots,[r_d]) \in \mathbb{R}^{r_1\times r_2\times\cdots\times r_d}$ is defined elementwise as $\mathbf{S}([r_1],[r_2],\ldots,[r_d])(i_1,i_2,\ldots,i_d) = \mathbf{S}(i_1,i_2,\ldots,i_d)$, for all $i_k \in [r_k]$ and for all $k \in [d]$.

Matrices will be denoted with capital bold letters, linear mappings with caligraphic capital letters and vectors with small bold letters.

The inner product of two tensors $\mathbf{X},\mathbf{Y} \in \mathbb{R}^{n_1\times n_2\times\cdots\times n_d}$ is defined as

$$\langle \mathbf{X},\mathbf{Y}\rangle = \sum_{i_1=1}^{n_1}\sum_{i_2=1}^{n_2}\ldots\sum_{i_d=1}^{n_d}\mathbf{X}(i_1,i_2,\ldots,i_d)\,\mathbf{Y}(i_1,i_2,\ldots,i_d).$$

The (Frobenius) norm of a tensor $\mathbf{X} \in \mathbb{R}^{n_1\times n_2\times\cdots\times n_d}$, $\|\mathbf{X}\|_F$, induced by this inner product, is

$$\|\mathbf{X}\|_F = \sqrt{\sum_{i_1=1}^{n_1}\sum_{i_2=1}^{n_2}\ldots\sum_{i_d=1}^{n_d}\mathbf{X}^2(i_1,i_2,\ldots,i_d)}.$$

Matricization (unfolding) is the operation that transforms a tensor into a matrix. The mode-k matricization of a tensor $\mathbf{X} \in \mathbb{R}^{n_1\times n_2\times\cdots\times n_d}$ is denoted by $\mathbf{X}_{(k)}$, $\mathbf{X}_{(k)} \in \mathbb{R}^{n_k\times n_1\cdots n_{k-1}n_{k+1}\cdots n_d}$. The rows of the matrix $\mathbf{X}_{(k)}$ are determined by the k-th component of the tensor $\mathbf{X}$, whereas all the remaining components determine its column, i.e.,

$$\mathbf{X}_{(k)}(i_k;(i_1,\ldots,i_{k-1},i_{k+1},\ldots,i_d)) = \mathbf{X}(i_1,\ldots,i_d).$$

For $\mathbf{X} \in \mathbb{R}^{n_1\times n_2\times\cdots\times n_d}$, $\mathbf{A} \in \mathbb{R}^{J\times n_k}$ and $k \in [d]$, the k-mode multiplication, $\mathbf{X}\times_k\mathbf{A} \in \mathbb{R}^{n_1\times\cdots\times n_{k-1}\times J\times n_{k+1}\times\cdots\times n_d}$ is defined elementwise as

$$(\mathbf{X}\times_k\mathbf{A})(i_1,\ldots,i_{k-1},j,i_{k+1},\ldots,i_d) = \\ = \sum_{i_k=1}^{n_k}\mathbf{X}(i_1,\ldots,i_d)\,\mathbf{A}(j,i_k).$$

Remark 1: Notice that the SVD decomposition of a matrix $\mathbf{X} \in \mathbb{R}^{n_1\times n_2}$ can be written using the above notation as $\mathbf{X} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T = \mathbf{\Sigma}\times_1\mathbf{U}\times_2\mathbf{V}$.

III. HOSVD DECOMPOSITION

The Tucker decomposition, and in particular the HOSVD decomposition [7], decomposes a tensor into a set of matrices and one tensor.

Definition 1 (Tucker decomposition): Given a tensor $\mathbf{X} \in \mathbb{R}^{n_1\times n_2\times\cdots\times n_d}$ the decomposition

$$\mathbf{X} = \mathbf{S}\times_1\mathbf{U}_1\times_2\mathbf{U}_2\times\cdots\times_d\mathbf{U}_d,$$

or elementwise

$$\mathbf{X}(i_1,i_2,\ldots,i_d) = \sum_{j_1=1}^{n_1}\sum_{j_2=1}^{n_2}\ldots\sum_{j_d=1}^{n_d}\mathbf{S}(j_1,\ldots,j_d)\cdot \\ \cdot\,\mathbf{U}_1(i_1,j_1)\,\mathbf{U}_2(i_2,j_2)\cdots\mathbf{U}_d(i_d,j_d)$$

is called Tucker decomposition. The tensor $\mathbf{S} \in \mathbb{R}^{r_1\times r_2\times\cdots\times r_d}$ is called the core tensor and $\mathbf{U}_i \in \mathbb{R}^{n_i\times n_i}$, for all $i \in [d]$, are matrices.

Remark 2: Given invertible matrices $\mathbf{U}_i \in \mathbb{R}^{n_i\times n_i}$, the Tucker decomposition of a tensor $\mathbf{X}$ always exists since

$$\mathbf{S} = \mathbf{X}\times_1\mathbf{U}_1^{-1}\times_2\mathbf{U}_2^{-1}\times\cdots\times\mathbf{U}_d^{-1}$$

defines the core tensor.

Definition 2 (HOSVD decomposition): The HOSVD is a special case of the Tucker decomposition where

- the $\mathbf{U}_k$ are unitary $n_k\times n_k$-matrices, for all $k \in [d]$,
- any two subtensors of the core tensor $\mathbf{S}$ are orthogonal, i.e., $\langle S_{i_k=p}, S_{i_k=q}\rangle = 0$, for all $k \in [d]$ and for all $p \neq q$,
- the subtensors of the core tensor $\mathbf{S}$ are ordered according to their Frobenius norm, i.e., $\|S_{i_k=1}\|_F \geq \|S_{i_k=2}\|_F \geq \ldots \geq \|S_{i_k=n_k}\|_F \geq 0$, for all $k \in [d]$.

Definition 3 (HOSVD-rank): Let $\mathbf{X} \in \mathbb{R}^{n_1\times n_2\times\cdots\times n_d}$. The k-rank of $\mathbf{X}$, denoted by $R_k = \operatorname{rank}_k(\mathbf{X})$, is the rank of the k-th unfolding, i.e.,

$$\operatorname{rank}_k(\mathbf{X}) = \operatorname{rank}\left(\mathbf{X}_{(k)}\right).$$

The HOSVD-rank of a tensor $\mathbf{X}$ is the vector $\mathbf{r}_{\text{HOSVD}}(\mathbf{X}) = (R_1,R_2,\ldots,R_d)$.

Remark 3 ([7]): Let the HOSVD of a tensor $\mathbf{X} \in \mathbb{R}^{n_1\times n_2\times\cdots\times n_d}$ be given as in Definition 1 and let r_k be equal to the highest index for which $\|S_{i_k=r_k}\|_F > 0$. Then

$$r_k = \operatorname{rank}_k(\mathbf{X}) = R_k.$$

Remark 4: Let $\mathbf{X} \in \mathbb{R}^{n_1\times n_2\times\cdots\times n_d}$ be a tensor of HOSVD-rank $(r_1,r_2,\ldots,r_d)$ and let $\mathbf{X} = \mathbf{S}\times_1\mathbf{U}_1\times_2\mathbf{U}_2\times\cdots\times_d\mathbf{U}_d$ be its HOSVD decomposition. Then $\mathbf{X}$ can be written as

$$\mathbf{X} = \overline{\mathbf{S}}\times_1\overline{\mathbf{U}}_1\times_2\overline{\mathbf{U}}_2\times\cdots\times_d\overline{\mathbf{U}}_d,$$

where $\overline{\mathbf{S}} = \mathbf{S}([r_1],[r_2],\ldots,[r_d]) \in \mathbb{R}^{r_1\times r_2\times\cdots\times r_d}$, $\overline{\mathbf{U}}_k = \mathbf{U}_k(:,[r_k]) \in \mathbb{R}^{n_k\times r_k}$, for all $k \in [d]$. From now on, we will assume that the HOSVD decomposition of an arbitrary tensor is of this form.

IV. TENSOR ITERATIVE HARD THRESHOLDING ALGORITHM

In this section we present the tensor iterative hard thresholding algorithm (TIHT) and the corresponding numerical results.

In the TIHT algorithm, $\mathcal{H}_{\mathbf{r}}(\mathbf{X})$ denotes the rank-$\mathbf{r}$ approximation of the tensor $\mathbf{X}$ obtained by restricting the components of its HOSVD decomposition. To be more precise, if $\mathbf{X} = \mathbf{S}\times_1\mathbf{U}_1\times\cdots\times_d\mathbf{U}_d$ is its HOSVD decomposition, then $\mathcal{H}_{\mathbf{r}}(\mathbf{X}) = \overline{\mathbf{S}}\times_1\overline{\mathbf{U}}_1\times\cdots\times_d\overline{\mathbf{U}}_d$ where $\overline{\mathbf{S}} = \mathbf{S}([r_1],\ldots,[r_d]) \in \mathbb{R}^{r_1\times r_2\times\cdots\times r_d}$ and $\overline{\mathbf{U}}_k = \mathbf{U}_k(:,[r_k]) \in \mathbb{R}^{n_k\times r_k}$ for all $k \in [d]$.

Remark 5: In the case of sparse vector recovery and of low-rank matrix recovery, the operator $\mathcal{H}_r$ returns the best r-sparse approximation [1] and best rank-r approximation [5], respectively. This fact is often used in the analysis of the algorithm. However, the rank-$\mathbf{r}$ approximation $\mathcal{H}_{\mathbf{r}}(\mathbf{X})$ of an arbitrary d-th order tensor $\mathbf{X}$ is not necessarily its best rank-$\mathbf{r}$ approximation $\mathbf{X}_{\text{BEST}}$ [4]. To be more precise,

$$\|\mathbf{X}-\mathcal{H}_{\mathbf{r}}(\mathbf{X})\|_F \leq \sqrt{d}\,\|\mathbf{X}-\mathbf{X}_{\text{BEST}}\|_F.$$

Tensor iterative hard thresholding algorithm (TIHT algorithm)

Input: measurement ensemble $\mathcal{A}$, measurement vector $\mathbf{y} = \mathcal{A}(\mathbf{X})$, rank level $\mathbf{r}$
$\mathbf{X}^0 = \mathcal{H}_{\mathbf{r}}(\mathcal{A}^*(\mathbf{y}))$, $j = 0$.
repeat
Compute $\mu_j = \frac{\|\mathcal{A}^*(\mathbf{y}-\mathcal{A}(\mathbf{X}^j))\|_F^2}{\|\mathcal{A}(\mathcal{A}^*(\mathbf{y}-\mathcal{A}(\mathbf{X}^j)))\|_2^2}$.
Set $\mathbf{X}^{j+1} = \mathcal{H}_{\mathbf{r}}\left(\mathbf{X}^j + \mu_j \mathcal{A}^*\left(\mathbf{y} - \mathcal{A}\left(\mathbf{X}^j\right)\right)\right)$.
j=j+1
until a stopping criterion is met at $\bar{j} = j$.
Output: the $\mathbf{r}$-rank tensor $\mathbf{X}^{\#} = \mathbf{X}^j$

This fact causes significant obstacles in the theoretical anaysis of the TIHT. Nevertheless, as we present in the following, the algorithm still works quite well in practice.

We present the numerical results only for 3rd order tensors $\mathbf{X} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$. In the first two experiments we consider a cubic tensor, i.e., $n_1 = n_2 = n_3 = 10$, with equal and unequal ranks of its unfoldings, respectively. In the last case we consider a non-cubic tensor $\mathbf{X} \in \mathbb{R}^{6 \times 10 \times 15}$ with equal ranks of the unfoldings, i.e., $r_1 = r_2 = r_3 = r$.

For fixed tensor dimensions $n_1 \times n_2 \times n_3$, fixed HOSVD-rank $\mathbf{r} = (r_1, r_2, r_3)$ and a fixed number of measurements m we performed 200 simulations.

We consider an algorithm to successfully recover the sensed tensor $\mathbf{X}_0$ if it returns a tensor $\mathbf{X}^{\#}$ s.t. $\left\|\mathbf{X}_0 - \mathbf{X}^{\#}\right\|_F < 10^{-3}$.

The algorithm stops if $\left\|\mathbf{X}^j - \mathbf{X}^{j-1}\right\|_F < 10^{-4}$ in which case we say that the algorithm converged, or it stops if it reached 5000 iterations.

A linear mapping $\mathcal{A} : \mathbb{R}^{n_1 \times n_2 \times n_3} \to \mathbb{R}^m$ is defined by tensors $\mathbf{A}_k \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ via $[\mathcal{A}(\mathbf{X})](k) = \langle \mathbf{X}, \mathbf{A}_k \rangle$, for all $k \in [m]$. The entries of the tensors $\mathbf{A}_k$ are i.i.d. Gaussian $\mathcal{N}\left(0, \frac{1}{m}\right)$. We generate tensors $\mathbf{X}^0 \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ of rank $\mathbf{r} = (r_1, r_2, r_3)$ via its Tucker decomposition. If $\mathbf{X}^0 = \mathbf{S} \times_1 \mathbf{U}_1 \times_2 \mathbf{U}_2 \times_3 \mathbf{U}_3$ is its Tucker decomposition, each of the elements of the tensor $\mathbf{S}$ is taken independently from the normal distribution, $\mathcal{N}(0,1)$, and the components $\mathbf{U}_k \in \mathbb{R}^{n_k \times r_k}$ are the first r_k left singular vectors of a matrix $\mathbf{M}_k \in \mathbb{R}^{n_k \times n_k}$ whose elements are also drawn independently from the normal distribution $\mathcal{N}(0,1)$.

In Figure 1 and Figure 2 we present the recovery results for low-rank tensors of size $10 \times 10 \times 10$. The horizontal axis represents the number of measurements taken with respect to the number of degrees of freedom of an arbitrary tensor of this size. To be more precise, for a tensor of size $n_1 \times n_2 \times n_3$, the number $\bar{n}$ on the horizontal axis represents $m = \left\lceil n_1 n_2 n_2 \frac{\bar{n}}{100} \right\rceil$ measurements. The vertical axis represents the percentage of the successful recovery. The numerical results for tensors of rank $\mathbf{r} = (1,1,1)$, $\mathbf{r} = (2,2,2)$, $\mathbf{r} = (5,5,5)$ and $\mathbf{r} = (7,7,7)$ are presented in Figure 1. Notice that only for the rank $\mathbf{r} = (7,7,7)$, 33% of measurements were not enough for a full recovery. In this case 54% of the measurements and on average 1107 iterations were needed. For tensors of rank $\mathbf{r} = (1,1,1)$ already with 9% of measurements we obtain a full recovery in 321 iterations on average. The algorithm ended on average in 185, 337 and 547 iterations for 20%, 21% and 33% of measurements for ranks $\mathbf{r} = (2,2,2)$, $\mathbf{r} = (3,3,3)$ and $\mathbf{r} = (5,5,5)$, respectively.

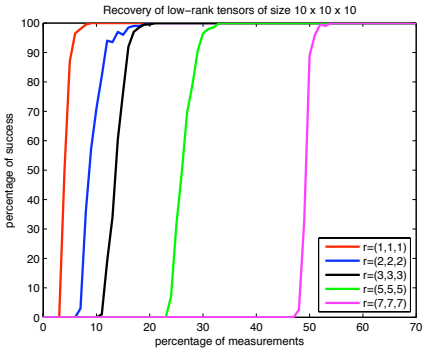

Fig. 1. Recovery of low rank 10 x 10 x 10 tensors of the same rank

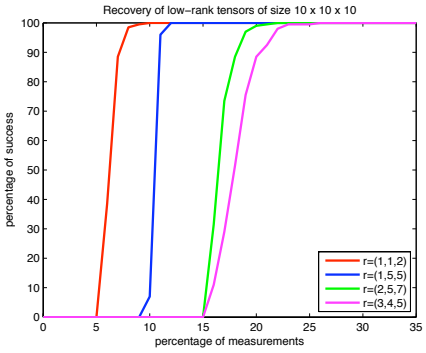

Fig. 2. Recovery of low rank $10 \times 10 \times 10$ tensors of a different rank

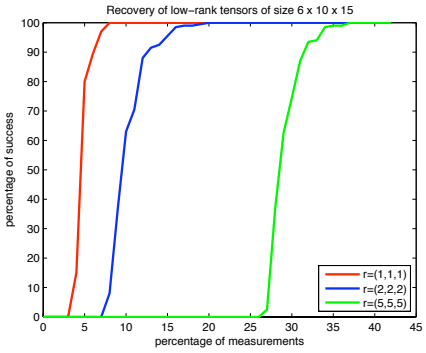

Fig. 3. Recovery of low rank $6 \times 10 \times 15$ tensors of a different rank

In Figure 2 we present the results for tensors of rank $\mathbf{r} = (1,2,2)$, $\mathbf{r} = (1,5,5)$, $\mathbf{r} = (2,5,7)$ and $\mathbf{r} = (3,4,5)$. Only 26% of measurements were enough for a full recovery. For 10%, 12%, 22% and 26% of measurements, the algorithm converged on average in 588, 1912, 696, 384 iterations, for the various ranks respectively.

We obtained similar results for recovery of low-rank tensors of size $6 \times 10 \times 15$ and ranks $\mathbf{r} = (1,1,1)$, $\mathbf{r} = (2,2,2)$ and $\mathbf{r} = (5,5,5)$ - see Figure 3. We managed to get a full recovery from 8% of measurements for the rank $\mathbf{r} = (1,1,1)$, and 20% and 37% of measurements for the remaining ranks. The algorithm ended on average in 511, 214 and 501 iterations, for

the various ranks and number of measurements, respectively.

V. HOSVD TENSOR RIP

The analysis of the IHT algorithm for recovery of sparse vectors [1] and low-rank matrices [5] is based on the corresponding notion of restricted isometry property (RIP). Therefore, we start by introducing an analog for tensors, which we call the tensor restricted isometry property (HOSVD-TRIP).

Definition 4 (HOSVD-TRIP): Let $\mathcal{A}\colon \mathbb{R}^{n_1\times n_2\times\cdots\times n_d} \to \mathbb{R}^m$ be a measurement ensemble. Then for each d-tuple of the integers $\mathbf{r}$, $\mathbf{r} = (r_1, r_2, \ldots, r_d)$, where $r_i \in [n_i]$, for all $i \in [d]$, the tensor restricted isometry constant $\delta_{\mathbf{r}}$ of $\mathcal{A}$ is the smallest quantity such that

$$(1-\delta_{\mathbf{r}})\left\|\mathbf{X}\right\|_F^2 \leq \left\|\mathcal{A}\left(\mathbf{X}\right)\right\|_2^2 \leq (1+\delta_{\mathbf{r}})\left\|\mathbf{X}\right\|_F^2 \qquad (4)$$

holds for all tensors of HOSVD-rank at most $\mathbf{r}$.

We say that $\mathcal{A}$ satisfies the HOSVD-TRIP at rank $\mathbf{r}$ if $\delta_{\mathbf{r}}$ is bounded by a sufficiently small constant between 0 and 1.

A random variable X is called L-subgaussian if there exists a constant $L > 0$ s.t. $\mathbb{E}\left[\exp\left(tX\right)\right] \leq \exp\left(L^2t^2/2\right)$ holds for all $t \in \mathbb{R}$. We call $\mathcal{A} : \mathbb{R}^{n_1\times n_2\times\cdots\times n_d} \to \mathbb{R}^m$ an L-subgaussian measurement ensemble if all elements of $\mathcal{A}$, interpreted as a tensor in $\mathbb{R}^{n_1\times n_2\times\cdots\times n_d\times m}$, are independent mean-zero, variance one, L-subgaussian variables. Gaussian and Bernoulli random measurement ensembles where the entries are standard normal distributed random variables and Rademacher ± 1 variables, respectedly, are special cases of 1-subgaussian measurement ensembles.

Theorem 1: Let $\delta, \varepsilon \in (0,1)$. A random draw of an L-subgaussian measurement ensemble $\mathcal{A} : \mathbb{R}^{n_1\times n_2\times\cdots\times n_d} \to \mathbb{R}^m$ satisfies $\delta_{\mathbf{r}} \leq \delta$ with probability at least $1-\varepsilon$ provided

$$m \geq C\delta^{-2}\max\left\{\left(r^d + dnr\right)\log\left(d\right), \log\left(\varepsilon^{-1}\right)\right\},$$

where $n = \max\left\{n_i : i \in [d]\right\}$ and $r = \max\left\{r_i : i \in [d]\right\}$. The constant $C > 0$ depends only on subgaussian parameter L.

The proof of Theorem 1 uses ε-nets.

Definition 5: A set $\mathcal{N}_\varepsilon \subset X$ is called an ε-net of X with respect to the norm $\left\|\cdot\right\|$ if for each $v \in X$, there exists $v_0 \in \mathcal{N}_\varepsilon$ with $\left\|v_0 - v\right\| \leq \varepsilon$. The minimal cardinality of an ε-net of X with respect to the norm $\left\|\cdot\right\|$, if finite, is denoted $\mathcal{N}\left(X, \left\|\cdot\right\|, \varepsilon\right)$ and is called the covering number of X (at scale ε).

Lemma 1 (Covering number of low-HOSVD-rank tensors): Let

$$S_{\mathbf{r}} = \left\{\mathbf{X} \in \mathbb{R}^{n_1\times n_2\times\cdots\times n_d} : \mathrm{rank}_{\mathrm{HOSVD}}\left(\mathbf{X}\right) \leq \mathbf{r}, \left\|\mathbf{X}\right\|_F = 1\right\}.$$

Then there exists an ε-net $\mathcal{N}_\varepsilon$ of $S_{\mathbf{r}}$ with respect to the Frobenius norm obeying

$$\mathcal{N}\left(S_{\mathbf{r}}, \left\|\cdot\right\|_F, \varepsilon\right) \leq \left(3\left(d+1\right)/\varepsilon\right)^{r_1r_2\cdots r_d + \sum_{i=1}^d n_i r_i}. \qquad (5)$$

The proof of the above lemma follows a similar strategy as in [2] and will be presented in a forthcoming journal paper.

Sketch of the proof of the Theorem 1: We use a tool developed in [6]. We write

$$\mathcal{A}\left(\mathbf{X}\right) = \mathbf{V}_{\mathbf{X}}\boldsymbol{\xi},$$

where $\boldsymbol{\xi}$ is an L-subgaussian random vector of length $n_1n_2\cdots n_dm$ and $\mathbf{V}_{\mathbf{X}}$ is the $m \times n_1n_2\cdots n_dm$ block-diagonal matrix

$$\mathbf{V}_{\mathbf{X}} = \frac{1}{\sqrt{m}}\begin{bmatrix} \mathbf{x}^T & 0 & \cdots & 0 \\ 0 & \mathbf{x}^T & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & \mathbf{x}^T \end{bmatrix},$$

where $\mathbf{x}$ is the vectorized version of the tensor $\mathbf{X}$. With this notation the restricted isometry constant is given by

$$\delta_{\mathbf{r}} = \sup_{\mathbf{X}\in S_{\mathbf{r}}}\left|\left\|\mathbf{V}_{\mathbf{X}}\boldsymbol{\xi}\right\|_2^2 - \mathbb{E}\left\|\mathbf{V}_{\mathbf{X}}\boldsymbol{\xi}\right\|_2^2\right|.$$

Theorem 3.1 in [6] provides a general probabilistic bound for expressions in the form of the right hand side above in terms of the diameters $d_F(\mathcal{B})$ and $d_{2\to 2}(\mathcal{B})$ of the set $\mathcal{B} := \{\mathbf{V}_{\mathbf{X}} : \mathbf{X} \in S_{\mathbf{r}}\}$ with respect to the Frobenius norm and the operator norm, as well as in terms of Talagrand's functional $\gamma_2(\mathcal{B}, \left\|\cdot\right\|_{2\to 2})$. It is straightforward to see that $d_F(\mathcal{B}) = 1$ and $d_{2\to 2}(\mathcal{B}) = \frac{1}{\sqrt{m}}$. The bound of the γ_2-functional via a Dudley type integral [6] yields

$$\gamma_2\left(\mathcal{B}, \left\|\cdot\right\|_{2\to 2}\right) \leq C\frac{1}{\sqrt{m}}\int_0^1 \sqrt{\log\left(\mathcal{N}\left(S_{\mathbf{r}}, \left\|\cdot\right\|_2, u\right)\right)}du.$$

Using (5) for $\mathcal{N}\left(S_{\mathbf{r}}, \left\|\cdot\right\|_F, u\right)$ we reach

$$\gamma_2\left(\mathcal{B}, \left\|\cdot\right\|_{2\to 2}\right) \leq \tilde{C}\sqrt{\frac{\left(r_1r_2\cdots r_d + \sum_{i=1}^d n_ir_i\right)\log\left(d\right)}{m}}.$$

The claim follows then from Theorem 3.1 in [6]. ∎

REFERENCES

[1] T. Blumensath, M. E. Davis, *Iterative hard thresholding for compressed sensing*. Appl. Comput. Harmon. Anal., 27(3):265–274, 2009.

[2] E. J. Candès , Y. Plan, *Tight oracle bounds for low-rank matrix recovery from a minimal number of random measurements*. IEEE Transactions on Information Theory, 57(4):2342–2359, 2009.

[3] S. Gandy, B. Recht, I. Yamada, *Tensor completion and low-n-rank tensor recovery via convex optimization*. Inverse Problems, 27(2):19pp, 2011.

[4] L. Grasedyck, *Hierarchical Singular Value Decomposition of Tensors*. SIAM J. Matrix Anal. Appl., 31(4):2029–2054, 2010.

[5] P. Jain, R. Meka, I. S. Dhillon, *Guaranteed Rank Minimization via Singular Value Projection*. Neural Information Processing Systems, pp. 937–945, 2010.

[6] F. Krahmer, S. Mendelson, H. Rauhut *Suprema of Chaos Processes and the Restricted Isometry Property*. Comm. Pure Appl. Math., to appear

[7] L. de Lathauwer, B. de Moor, J. Vandewalle, *A multilinear singular value decomposition*. SIAM J. Matrix Anal. Appl., 21(4):1253–1278, 2000.

[8] L.-H. Lim, C. J. Hillar, *Most Tensor Problems are NP-Hard*. http://arxiv.org/pdf/0911.1393v3.pdf, 2012.

[9] J. Liu, P. Musiaski, P. Wonka, J. Ye, *Tensor Completion for Estimating Missing Values in Visual Data*. IEEE Transactions on Pattern Analyis and Machine Intelligence, 35(1):208–220, 2013.

[10] D. Muti, S. Bourennane, *Multidimensional filtering based on a tensor approach*. Signal Process., 85(12):2338–2353, 2005.

[11] B. Recht, M. Fazel, P. A. Parillo, *Guaranteed minimum-rank solution of linear matrix equations via nuclear norm minimization*. SIAM Rev., 52(3):471–501, 2010.

[12] B. Savas, L. Eldén, *Handwritten digit classification using higher order singular value decomposition*. Pattern Recog., 40(3):993–1008, 2003.

[13] M. A. O. Vasilescu, D. Terzopoulos, *Multilinear analysis of image ensembles: TensorFaces*. Lecture Notes in Computer Science, 2350:447–460, Springer-Verlag, 2002.