A Fast Matrix Completion Method for Index Coding

Ehsan Asadi*, Shayan Aziznejad*, Mohammad H. Amerimehr† and Arash Amini*
*Department of Electrical Engineering, Sharif University of Technology, Tehran, Iran.
Emails: ehsan_asadik@yahoo.com, sh.aziznejad@gmail.com and aamini@sharif.edu
†Iran Telecommunication Research Center, Tehran, Iran
Email: amerimehr@gmail.com

***Abstract*—We investigate the problem of index coding, where a sender transmits distinct packets over a shared link to multiple users with side information. The aim is to find an encoding scheme (linear combinations) to minimize the number of transmitted packets, while providing each user with sufficient amount of data for the recovery of the desired parts. It has been shown that finding the optimal linear index code is equivalent to a matrix completion problem, where the observed elements of the matrix indicate the side information available for the users. This modeling results in an incomplete square matrix with all ones on the main diagonal (and some other parts), which needs to be completed with minimum rank. Unfortunately, this is a case in which conventional matrix completion techniques based on nuclear-norm minimization are proved to fail [Huang, Rouayheb 2015]. Instead, an alternating projection (the AP algorithm) method is proposed in [Huang, Rouayheb 2015]. In this paper, in addition to proving the convergence of the AP algorithm under certain conditions, we introduce a modification which considerably improves the run time of the method.**

***Index Terms*—Alternating projection, coded caching, index coding, matrix completion, network caching, rank minimization.**

I. Introduction

Let us consider a communication scenario in which a sender wishes to transmit a set of packets to a number of receivers over an error-free broadcast channel. Each receiver might have a priori access to some of the packets (side information) and is interested to receive a subset of the remaining packets. Knowing the available side information at the receiver ends, the sender wishes to broadcast a minimal number of coded packets that enable all the receivers to obtain their desired packets. This scenario is known as the index coding problem.

Index coding is motivated by many applications such as network caching, video on demand, and cloud distributed storage. For example in a caching network, each client opportunistically stores contents that might be needed in the future (called cashed data). This way, a server receives multiple file requests from the clients. A simple strategy is to separately respond to each request. However, it might be possible to reduce the transmission rate by combing requests and taking into account the already cached data by the clients. In the latter case, the server needs to combine the requested data via encoding; thus, this problem is referred to as *coded caching*. It is shown that under certain conditions, the coded caching problem induces an index coding problem [1].

To better clarify the index coding problem, let us consider a scenario with a single transmitter and three receivers $\{U_i\}_{i=1}^3$. Further, let the transmitter have the packets X_1, X_2, X_3 from which X_2, X_3 and $\{X_1, X_2\}$ are already available to U_1, U_2 and U_3, respectively. Next, each receiver U_i requests the packet X_i. To respond to these requests, instead of transmitting three simple packets, the transmitter has the option of sending two coded packets of the form $X_1 + X_2$ and $X_1 + X_3$. It is easy to verify that each receiver is able to retrieve its desired packet from the broadcasted and cashed data.

The statement of the index coding problem was originally introduced in [2]. One can observe many conceptual similarities between index coding and network coding. In particular, similar to network coding, linear methods are the dominant approaches in index coding. In other words, it is common to limit the transmitting packets to linear combinations of the original packets. The equivalence of the linear index coding to a matrix completion problem was shown in [3]. The goal in a matrix completion problem is to find an $m \times n$ matrix $\mathbf{A}$ with minimum rank by knowing a subset of its elements [4]. Equivalently, we look for the solution to

$$\min_{\mathbf{M}\in\mathbb{R}^{m\times n}} \operatorname{rank}(\mathbf{M}) \quad s.t : \|\mathcal{P}_\Omega(\mathbf{M}) - \mathcal{P}_\Omega(\mathbf{A})\|_F < \epsilon, \quad (1)$$

where Ω is the index set of known entries of $\mathbf{A}$ and $\mathcal{P}_\Omega(\mathbf{M})$ is an m by n matrix which is defined as

$$\mathcal{P}_\Omega(\mathbf{M})_{i,j} \triangleq \begin{cases} m_{i,j} & (i,j) \in \Omega, \\ 0 & (i,j) \notin \Omega. \end{cases} \quad (2)$$

For the index coding problem, we assume n users $\{U_i\}_{i=1}^n$, and n packets $\{X_i\}_{i=1}^n$, where U_i requests to have the packet X_i and has already cached a subset of other packets. In the equivalent matrix completion problem, we are dealing with an n by n matrix $\mathbf{A}$ with the known entries consisting of ones on its main diagonal (indicating that user i wants the i^{th} packet) and zero (i, j) entries where user i does not have the j^{th} packet on its cash. Other entries of $\mathbf{A}$ are considered as missing. If $\mathbf{A}$ is the matrix with minimum rank k that satisfies these constraints, each of its rows can be represented as a linear combination of k independent vectors $\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_k$. Therefore, it is sufficient for the transmitter to broadcast k linear combinations of the packets, where the coefficients of the j^{th} linear combination are the elements of $\mathbf{v}_j$.

The linear index coding problem and its matrix completion equivalent are in general NP hard [5]. In the context of matrix rank minimization, it is known that under certain constraints, relaxing the non-convex rank function with the nuclear norm

ISBN 978-0-9928626-7-1

of the matrix might not change the solution [6]; hence, the problem can be solved with efficient convex optimization methods. Unfortunately, when the known entries of the matrix include the main diagonal this relaxation is gauranteed to fail [7]. This explains why the proposed matrix completion methods developed for the index coding problem such as those in [7], [8] deviate from the standard nuclear-norm-based techniques and usually lack rigorous convergence results.

The contribution of this paper is twofold: we first introduce a modification on the AP-Index-coding algorithm of [7], which considerably reduces its computational complexity and enhances its speed. Next, we prove the convergence of the AP Index coding algorithm under some conditions.

The paper is organized as follows: in Section II we review the AP Index coding algorithm and introduce our method. We provide the convergence analysis of the AP Index coding algorithm in Section III, and present the simulation results in Section IV. Finally, we conclude the paper in Section V.

II. Algorithm

A. The AP Index coding algorithm

Assume $\mathbf{A}$ is an m by n matrix with $\text{rank}(\mathbf{A}) = r \ll \min(m,n)$. Let us define the hyperplane $H_\Omega \in \mathbb{R}^{m\times n}$ as $H_\Omega = \{\mathbf{M} \in \mathbb{R}^{m\times n} | \mathcal{P}_\Omega(\mathbf{M}) = \mathcal{P}_\Omega(\mathbf{A})\}$. Suppose that $\mathbf{L}$ is the solution of (1). Define the hyperplane $H_\mathbf{L} \in \mathbb{R}^{m\times n}$ as $H_\mathbf{L} = \{\mathbf{M} \in \mathbb{R}_{m\times n} | \text{Span}(\mathbf{M}) = \text{Span}(\mathbf{L})\}$. Obviously, $\mathbf{L}$ lies on both H_Ω and $H_\mathbf{L}$, and therefore, on their intersection. If $\text{Span}(\mathbf{L})$ was known, we could have estimated $\mathbf{L}$ by starting from an arbitrary matrix and alternatively projecting it onto H_Ω and $H_\mathbf{L}$ (until we reach an intersection). However, $\text{Span}(\mathbf{L})$ is unknown and its estimation is part of the problem. The main idea in the AP Index coding algorithm is to sequentially estimate $\text{Span}(\mathbf{L})$ and apply the alternating projection technique. For $\mathbf{M} \in H_\Omega$ and $\mathbf{Q} \in O_{m\times r}$, where $O_{m\times r}$ is the set of matrices with unit-length and orthogonal columns ($\text{Span}(\mathbf{Q})$ represents an estimate of $\text{Span}(\mathbf{L})$), let us define $f(\mathbf{M},\mathbf{Q})$ as

$$f(\mathbf{M},\mathbf{Q}) = \|\mathbf{Q}\mathbf{Q}^T\mathbf{M} - \mathbf{M}\|_2^2. \tag{3}$$

Indeed, $f(\mathbf{M},\mathbf{Q})$ is the distance between $\mathbf{M}$ and the hyperplane $H_\mathbf{Q} = \{\mathbf{N} \in \mathbb{R}^{m\times n} | \text{Span}(\mathbf{N}) = \text{Span}(\mathbf{Q})\}$. Now consider the minimization problem

$$\underset{\mathbf{M}\in H_\Omega, \mathbf{Q}\in O_{m\times r}}{\text{argmin}} \quad f(\mathbf{M},\mathbf{Q}). \tag{4}$$

We find the solution to (4) by sequentially updating $\mathbf{M}$ and $\mathbf{Q}$. More precisely, we show in Lemma 1 that for a fixed $\mathbf{M}$, the best choice of $\mathbf{Q}$ that minimizes $f(\mathbf{M},\mathbf{Q})$ is $\mathbf{U}_r$, which is the $n \times r$ matrix formed by the first r columns of $\mathbf{U}$ in the singular value decomposition (SVD) of $\mathbf{M} = \mathbf{U}\Sigma\mathbf{V}^T$. Similarly, in Lemma 2 we show that for a fixed $\mathbf{Q}$, the best choice of $\mathbf{M}$ that minimizes $f(\mathbf{M},\mathbf{Q})$ is the intersection of H_Ω and $H_\mathbf{Q}$, which is found by the alternating projection method.

Lemma 1. *Suppose $\mathbf{M} \in H_\Omega$ with SVD $\mathbf{M} = \mathbf{U}\Sigma\mathbf{V}^T$ is given. Then, the matrix formed by the first r columns of $\mathbf{U}$ denoted by $\mathbf{U}_r$ minimizes $f(\mathbf{M},\mathbf{Q})$ among all rank r orthonormal matrices.*

Algorithm 1 AP Index Coding Algorithm

input : Ω, $\mathcal{P}_\Omega(\mathbf{A})$, ϵ, *iternum*
output : completed matrix $\mathbf{B}$

procedure AP Index coding
  $r \leftarrow n$
  while $error \le \epsilon$ **do**
    $\mathbf{A}_0 \leftarrow \mathcal{P}_\Omega(\mathbf{A})$
    $r \leftarrow r - 1$
    for $k = 0 : iternum - 1$ **do**
      $[\mathbf{U},\Sigma,\mathbf{V}^T] = SVD(\mathbf{A}_k)$
      $\mathbf{U}_r \leftarrow$ first r columns of $\mathbf{U}$
      $\mathbf{B} = \mathbf{U}_r \times \mathbf{U}_r^T \times \mathbf{A}_k$
      $\mathbf{A}_{k+1} = \mathcal{P}_\Omega(\mathbf{A}) + P_{\Omega^c}(\mathbf{B})$
    end
    $error = \|\mathcal{P}_\Omega(\mathbf{A}_{k+1}) - \mathcal{P}_\Omega(\mathbf{A})\|_2$
  end

Proof. The operator $\mathbf{Q}\mathbf{Q}^T$ projects $\mathbf{M}$ onto $H_\mathbf{Q}$. Also note that $\text{rank}(\mathbf{Q}\mathbf{Q}^T\mathbf{M}) \le rank(\mathbf{Q}) = r$. Therefore $\mathbf{Q}\mathbf{Q}^T\mathbf{M}$ is a rank r approximation of $\mathbf{M}$, and it is known that the best rank r approximation of $\mathbf{M}$ is $\mathbf{U}_r\mathbf{U}_r^T\mathbf{M}$. □

Lemma 2. *For a fixed $\mathbf{Q}$, the minimizer of $f(\mathbf{M},\mathbf{Q})$ can be found by the alternating projection method.*

Proof. As $f(\mathbf{M},\mathbf{Q})$ reveals the distance between $\mathbf{M} \in H_\mathbf{Q}$ and the hyperplane $H_\mathbf{Q}$, solving (4) can be seen as the problem of finding the shortest distance between the two hyperplanes $H_\mathbf{Q}$ and H_Ω. This problem can be solved by starting from a point in H_Ω and alternatively projecting it onto $H_\mathbf{Q}$ and H_Ω. □

Algorithm 1 depicts the AP Index coding algorithm. This technique was proposed in [7] for the index coding problem, and also in [9] for the general problem of matrix completion. However, both works heuristically derive the method without proving the convergence of the method.

B. Our proposed algorithm

The singular value decomposition is a fundamental block in most of the matrix completion techniques. However, SVD is a costly algorithm with a complexity that scales like $O(n^3)$ with the dimension n. In applications that are connected to high-dimensional signals such as big-data, this imposes a great challenge.

Here, we propose a method that uses the QR decomposition instead of the SVD in the projection step of the AP Index coding algorithm. By this modification, we will find a better projecting direction with a less complicated algorithm that will significantly decrease the run-time of the program. The steps of our proposed algorithm and its pseudo-code are depicted in Algorithm 2.

III. Convergence Analysis of the Algorithm

In this section, we shall show that under some mild conditions, the AP Index coding algorithm converges. First, note

ISBN 978-0-9928626-7-1 © EURASIP 2017

Algorithm 2 QRS

input : Ω, $\mathcal{P}_\Omega(\mathbf{A})$, ϵ, *iternum*
output : completed matrix B

procedure QRS
  $r \leftarrow 0$
  while *error* $\geq \epsilon$ **do**
    $\mathbf{A}_0 \leftarrow \mathcal{P}_\Omega(\mathbf{A})$
    $r \leftarrow r + 1$
    Choose a random vector $\mathbf{v}_{new} \in \mathbb{R}^n$
    $\mathbf{V} \leftarrow [\mathbf{V}|\mathbf{v}_{new}]$
    for $k = 0 : iternum - 1$ **do**
      $\mathbf{V} \leftarrow \mathbf{A}_k \times \mathbf{A}_k^T \times \mathbf{V}$
      $[\mathbf{R}, \mathbf{Q}] \leftarrow QR(\mathbf{V})$
      $\mathbf{B} = \mathbf{Q} \times \mathbf{Q}^T \times \mathbf{A}_k$
      $\mathbf{A}_{k+1} = \mathcal{P}_\Omega(\mathbf{A}) + P_{\Omega^c}(\mathbf{B})$
    end
    $error = \|\mathcal{P}_\Omega(\mathbf{A}_{k+1}) - \mathcal{P}_\Omega(\mathbf{A})\|_2$
  end

that if the rank r of $\mathbf{A}$ is known, then, the k^{th} iteration in the AP Index coding algorithm can be rewritten as

$$\mathbf{B}_k = \mathcal{L}_r(\mathcal{P}_\Omega(\mathbf{A}) + \mathcal{P}_{\Omega^c}(\mathbf{B}_{k-1})), \tag{5}$$

where the initializing matrix is $\mathbf{B}_0 = \mathcal{P}_\Omega(\mathbf{A})$. Here, $\mathcal{L}_r(\mathbf{M})$ stands for the best rank r approximation of $\mathbf{M}$. We observe that the distance of the space $\mathcal{P}_{\Omega^c}(\mathbf{A})$ from Span($\mathbf{A}$) and Span($\mathbf{A}^T$) plays a significant role in our convergence analysis. Thus, if $\mathbf{A} = \mathbf{U}\Sigma\mathbf{V}$ represents the SVD of $\mathbf{A}$, we assume the followings to proceed:

H1: for every $\mathbf{X} \in \mathbb{R}^{m\times n}$, we have $\|\mathbf{U}^T\mathcal{P}_{\Omega^c}(\mathbf{X})\|_F \leq \alpha\|\mathbf{X}\|_F$,
H2: for every $\mathbf{X} \in \mathbb{R}^{m\times n}$, we have $\|\mathcal{P}_{\Omega^c}(\mathbf{X})\mathbf{V}\|_F \leq \alpha\|\mathbf{X}\|_F$,
H3: $\|\mathcal{P}_{\Omega^c}(\mathbf{A})\|_F < r\sigma_{\min}(\mathbf{A})^2$.

In the above assumptions, $k = \text{rank}(\mathbf{A})$, $\sigma_{\min}(\mathbf{A})$ is the minimum singular value of $\mathbf{A}$ and $\alpha > 0$ is a constant satisfying the following inequality

$$\alpha < \left(1 + \frac{(2\sigma_{\max}(\mathbf{A}) + \|\mathcal{P}_{\Omega^c}(\mathbf{A})\|_F)}{\sigma_{\min}(\mathbf{A})^2 - \frac{\|\mathcal{P}_{\Omega^c}(\mathbf{A})\|_F^2}{r}}\|\mathcal{P}_{\Omega^c}(\mathbf{A})\|_F\right)^{-1}. \tag{6}$$

With these assumptions, we prove the following convergence result:

Theorem 1. *If H1-H3 hold, then, the AP Index coding algorithm converges to* $\mathbf{A}$.

The proof is postponed to the appendix.

IV. Simulation Results

The simulation results in [7] indicate that the AP Index coding algorithm has a superior performance to some other competing methods. We perform our simulation scenarios on a work station equipped with an intel core i7-6500U and 8GB of RAM. Further, each of the figures show the average results over 100 realizations. For implementing network cashing setup, we assume that each user caches any given packet with probability p independent of other packets. We further set the tolerance parameter in the Algorithm 2 as ϵ. We have investigated the performance of our algorithm for different number of users and different number of iterations. We have also investigated the effect of tolerance on the resulting rank. Fig 1 shows that the performance of our algorithm improves as n increases. Fig 6 shows that our algorithm is less complex compared to AP method and Fig 3 reveals that there is not much difference in the performance of AP and QRS algorithms.

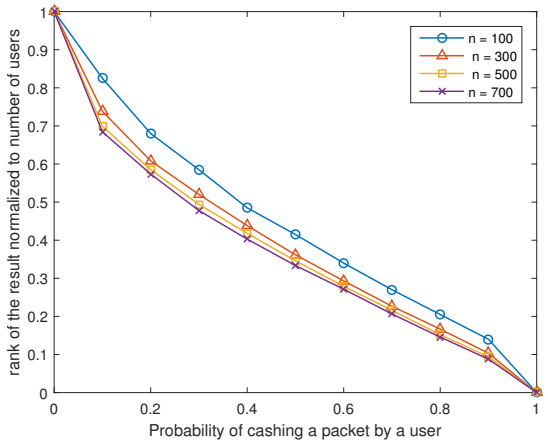

Fig. 1. Average rank of the result with number of iterations = 25 and tolerance = 0.001

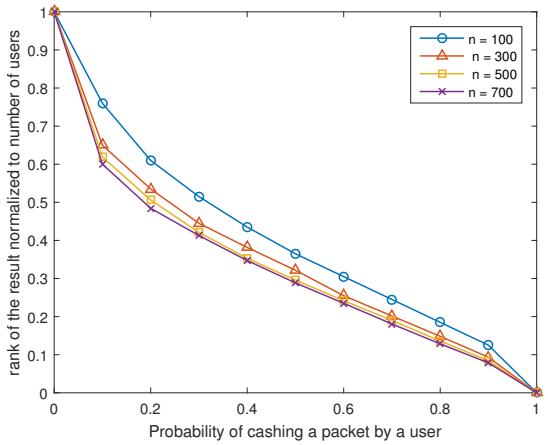

Fig. 2. Average rank of the result with number of iterations = 25 and tolerance = 0.01

V. Conclusion

In this paper we consider the AP Index coding algorithm which is matrix completion algorithm for the index coding problem. We presented the convergence analysis of this method. Then we introduced a new and faster method and investigated its performance. The simulation results showed a much better convergence time while the performance still remains satisfactory. Moreover, as the number of users increase

ISBN 978-0-9928626-7-1 © EURASIP 2017

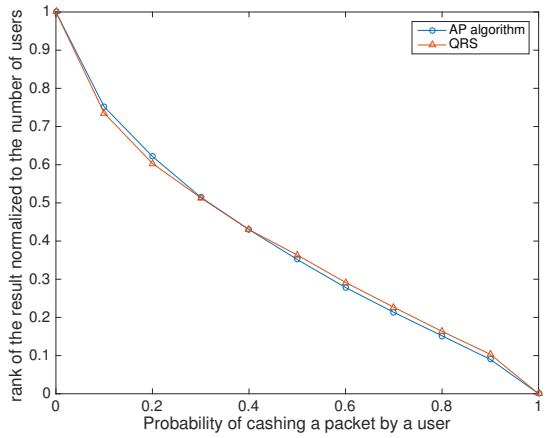

Fig. 3. Average rank of the outputs of QRS and AP algorithms for one run with number of iterations = 25 and tolerance = 0.001

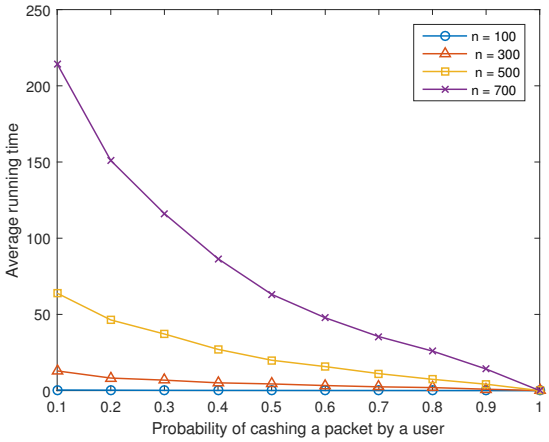

Fig. 5. Average time (in seconds) for one run with number of iterations = 25 and tolerance = 0.01

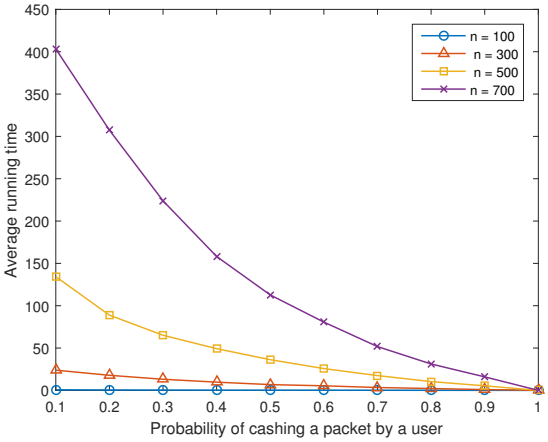

Fig. 4. Average time (in seconds) for one run with number of iterations = 25 and tolerance = 0.001

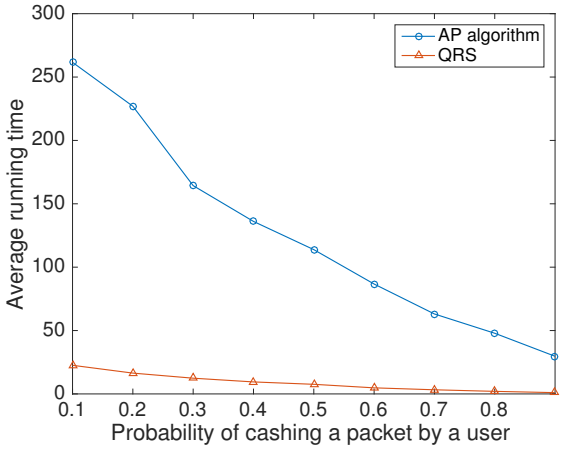

Fig. 6. Average runtime comparison between AP and QRS algorithms for one run with number of iterations = 25 and tolerance = 0.001

we saw that the performance of our algorithm improves. This is an important point, since the complexity of other matrix completion methods would not let us to increase the number of user as much as desired.

APPENDIX

proof of Theorem 1. In this section, the aim is to find a constant like $0 < \mu < 1$ such that $\|\mathbf{A}-\mathbf{B}_{k+1}\|_F \leq \mu\|\mathbf{A}-\mathbf{B}_k\|_F$. Define $\mathbf{C}_k = \mathcal{P}_\Omega(\mathbf{A}) + \mathcal{P}_{\Omega^c}(\mathbf{B}_k)$. Let's denote the orthogonal projection matrix into Span($\mathbf{A}$) as $\mathbf{T}$. It is known that $\mathbf{T} = \mathbf{A}\mathbf{A}^\dagger$ where $\mathbf{A}^\dagger$ is the Moore-Penrose pseudo inverse of $\mathbf{A}$. Then we have

$$\|\mathbf{A} - \mathbf{B}_{k+1}\|_F = \|\mathbf{A} - \mathbf{T}\mathbf{C}_k + \mathbf{T}\mathbf{C}_k - \mathcal{L}_r(\mathbf{C}_k)\|_F \tag{7}$$

$$\leq \|\mathbf{A} - \mathbf{T}\mathbf{C}_k\|_F + \|(\mathbf{T} - \mathcal{L}_r)(\mathbf{C}_k)\|_F \tag{8}$$

$$= \|\mathbf{T}\mathcal{P}_{\Omega^c}(\mathbf{A} - \mathbf{B}_k)\|_F + \|(\mathbf{T} - \mathcal{L}_r)(\mathbf{C}_k)\|_F. \tag{9}$$

Now we will use the Davis-Kahan theorem [10] in order to give an upper bound to the right hand of (9).

Theorem 2 (Davis-Kahan)**.** *Assume that* $\mathbf{X}$, $\mathbf{Y}$ *are Hermitian* n *by* n *matrices and we have the decomposition*

$$\mathbf{X} = \mathbf{E}_0\Sigma_0\mathbf{E}_0^T + \mathbf{E}_1\Sigma_1\mathbf{E}_1^T \tag{10}$$

$$\mathbf{X} + \mathbf{Y} = \mathbf{F}_0\Lambda_0\mathbf{F}_0^T + \mathbf{F}_1\Lambda_1\mathbf{F}_1^T \tag{11}$$

where $\mathbf{E} = [\mathbf{E}_0|\mathbf{E}_1]$ *and* $\mathbf{F} = [\mathbf{F}_0|\mathbf{F}_1]$ *are orthonormal bases for the* n *dimensional space. Then*

$$\|\mathbf{E}_0\mathbf{E}_0^T - \mathbf{F}_0\mathbf{F}_0^T\|_F = \|\mathbf{F}_1^T\mathbf{E}_0\|_F = \|\mathbf{E}_1^T\mathbf{F}_0\|_F \leq \frac{\|\mathbf{F}_1^T\mathbf{Y}\mathbf{E}_0\|_F}{\delta}.$$

where delta is chosen such that none of the eigenvalues of Λ_1 *are in the interval* $[\sigma_{\min} - \delta, \sigma_{\max} + \delta]$ *where* $\sigma_{\min}$ *and* $\sigma_{\max}$ *are the smallest and largest eigenvalue of* Σ_0*, respectively.*

Consider the SVD decomposition of $\mathbf{A}$, $\mathbf{A} = \mathbf{U}\mathbf{S}\mathbf{V}^T$. Since rank($\mathbf{A}$) = r, $\mathbf{S}$ has only r non zero entry on its diagonal. So we can write

$$\mathbf{A}\mathbf{A}^T = \mathbf{U}\mathbf{S}^2\mathbf{U}^T = \mathbf{U}_r\mathbf{S}_r^2\mathbf{U}_r^T \tag{12}$$

ISBN 978-0-9928626-7-1 © EURASIP 2017

Where $\mathbf{U}_r$ is the r first columns of $\mathbf{U}$, i.e. $\mathbf{U} = [\mathbf{U}_r|\mathbf{U}_{n-r}]$. So, we set $\mathbf{X} = \mathbf{A}\mathbf{A}^T$, $\mathbf{E}_0 = \mathbf{U}_r$, $\mathbf{E}_1 = \mathbf{U}_{n-r}$, $\Sigma_0 = \mathbf{S}_r$ and $\Sigma_1 = 0$. We also set $\mathbf{Y} = \mathbf{C}_k\mathbf{C}_k^T - \mathbf{X}$, $\mathbf{F}_0 = \mathbf{W}_r$, $\mathbf{F}_1 = \mathbf{W}_{n-r}$, $\Lambda_0 = \text{diag}(\sigma_1(\mathbf{Y}), \sigma_2(\mathbf{Y}), \cdots, \sigma_r(\mathbf{Y}))$ and $\Lambda_1 = \text{diag}(\sigma_{r+1}(\mathbf{Y}), \sigma_{r+2}(\mathbf{Y}), \cdots, \sigma_n(\mathbf{Y}))$ (Note that $\mathbf{W} = [\mathbf{W}_r|\mathbf{W}_{n-r}]$ diagonalize Y i.e. $\mathbf{W}^T\mathbf{Y}\mathbf{W}$ is a diagonal matrix).

Note that using above notation, we have $T = \mathbf{U}_r\mathbf{U}_r^T$ and $\mathcal{L}_r = \mathbf{W}_r\mathbf{W}_r^T$. So using Davis-Kahan theorem we have

$$\|\mathbf{T} - \mathcal{L}_r\|_F \leq \frac{\|\mathbf{W}_{n-r}^T(\mathbf{C}_k\mathbf{C}_k^T - \mathbf{A}\mathbf{A}^T)\mathbf{U}_r\|_F}{\delta} \tag{13}$$

To have and lower bound for δ, we use the Weyl's theorem [11].

Theorem 3 (Weyl's theorem)**.** *Assume δ_r and σ_r are the r^{th} singular values of matrices $\mathbf{A}$ and $\mathbf{B}$, respectively. We have*

$$|\delta_r - \sigma_r| \leq \|\mathbf{A} - \mathbf{B}\|_2^2 \tag{14}$$

Define $\mathbf{E}_k = \mathbf{A} - \mathbf{C}_k = -\mathcal{P}_{\Omega^c}(\mathbf{A} - \mathbf{B}_k)$. Using this theorem and the fact that $\mathbf{A}\mathbf{A}^T$ is a rank r matrix, we know

$$|\sigma_i(\mathbf{C}_k)| \leq \|\mathbf{E}_k\|_2 \quad i = r+1, r+2, \cdots, n. \tag{15}$$

Therefore we can see that

$$\delta \geq \sigma\min(\Sigma_0) - \sigma_{\max}(\Lambda_1) \geq \sigma_{\min}(\mathbf{A})^2 - \|\mathbf{E}_k\|_2^2. \tag{16}$$

Substitution of above inequality into (13) gives us

$$\|\mathbf{T} - \mathcal{L}_r\|_F \leq \frac{\|\mathbf{W}_{n-r}^T(\mathbf{A}\mathbf{E}_k^T + \mathbf{E}_k\mathbf{A}^T + \mathbf{E}_k\mathbf{E}_k^T)\mathbf{U}_r\|_F}{\sigma_{\min}(\mathbf{A})^2 - \|\mathbf{E}_k\|_2^2} \tag{17}$$

Using the inequality $\|\mathbf{A}\mathbf{B}\|_F \leq \|\mathbf{A}\|_2\|\mathbf{B}\|$ and knowing $\|\mathbf{W}_{n-r}^T\|_2 = 1$, we have

$$\|\mathbf{T} - \mathcal{L}_r\|_F \leq \frac{\|\mathbf{A}\mathbf{E}_k^T\mathbf{U}_r + \mathbf{E}_k\mathbf{V}_r\Sigma_r + \mathbf{E}_k\mathbf{E}_k^T\mathbf{U}_r\|_F}{\sigma_{\min}(\mathbf{A})^2 - \|\mathbf{E}_k\|_2^2} \tag{18}$$

By the theorem's assumptions we have

$$\|\mathbf{U}_r^T\mathbf{E}_k\|_F \leq \alpha\|\mathbf{E}_k\|_F \tag{19}$$

$$\|\mathbf{E}_k\mathbf{V}_r\|_F \leq \alpha\|\mathbf{E}_k\|_F. \tag{20}$$

By the triangle inequality, it can be seen that

$$\|\mathbf{T} - \mathcal{L}_r\|_F \leq \frac{\|\mathbf{A}\mathbf{E}_k^T\mathbf{U}_r\|_F + \|\mathbf{E}_k\mathbf{V}_r\Sigma_r\|_F + \|\mathbf{E}_k\mathbf{E}_k^T\mathbf{U}_r\|_F}{\sigma_{\min}(\mathbf{A})^2 - \|\mathbf{E}_k\|_2^2}. \tag{21}$$

Knowing $\mathbf{A}^T\mathbf{U}_r = \mathbf{V}_r\Sigma_r$ and (18), implies

$$\|\mathbf{T} - \mathcal{L}_r\|_F \leq \frac{(2\sigma_{\max} + \|E_k\|_F)\,\alpha\|E_k\|_F}{\sigma_{\min}(\mathbf{A})^2 - \|\mathbf{E}_k\|_2^2}. \tag{22}$$

Moreover, using $\mathbf{C}_k = \mathbf{A} + \mathbf{E}_k$, $\mathbf{T}\mathbf{A} = \mathbf{A}$ and $\mathcal{L}_r\mathbf{A} = \mathbf{A}$, we obtain that

$$(\mathbf{T} - \mathcal{L}_r)\mathbf{C}_k = (\mathbf{T} - \mathcal{L}_r)\mathbf{A} + (\mathbf{T} - \mathcal{L}_r)\mathbf{E}_k = (\mathbf{T} - \mathcal{L}_r)\mathbf{E}_k \tag{23}$$

and therefore

$$\|(\mathbf{T} - \mathcal{L}_r)\mathbf{C}_k\|_F \leq \|(\mathbf{T} - \mathcal{L}_r)\|_F\|\mathbf{E}_k\|_2. \tag{24}$$

Now, by (22) and (24)

$$\|(\mathbf{T} - \mathcal{L}_r)\mathbf{C}_k\|_F \leq \frac{(2\sigma_{\max} + \|\mathbf{E}_k\|_F)\,\alpha\|\mathbf{E}_k\|_F}{\sigma_{\min}(\mathbf{A})^2 - \|\mathbf{E}_k\|_2^2}\|\mathbf{E}_k\|_2. \tag{25}$$

Using (19) and knowing $\mathbf{T} = \mathbf{U}_r\mathbf{U}_r^T$, it can be seen that

$$\|\mathbf{T}\mathcal{P}_{\Omega^c}(\mathbf{A} - \mathbf{B}_k)\|_F \leq \alpha\|\mathbf{E}_k\|_F. \tag{26}$$

Therefore, putting (25) and (26) into (9) result to

$$\|\mathbf{A} - \mathbf{B}_{k+1}\|_F \leq \alpha\|\mathbf{E}_k\|_F + \frac{(2\sigma_{\max} + \|\mathbf{E}_k\|_F)\,\alpha\|\mathbf{E}_k\|_F}{\sigma_{\min}(\mathbf{A})^2 - \|\mathbf{E}_k\|_2^2}\|\mathbf{E}_k\|_2. \tag{27}$$

Finally, using $\|\mathbf{E}_{k+1}\|_F \leq \|\mathbf{A} - \mathbf{B}_{k+1}\|_F$ and $\|E_k\|_2 \leq \|E_k\|_F \leq \sqrt{r}\|E_k\|_2$, it can be concluded that

$$\|\mathbf{E}_{k+1}\|_F \leq \alpha\|\mathbf{E}_k\|_F\left(1 + \frac{(2\sigma_{\max} + \|\mathbf{E}_k\|_F)}{\sigma_{\min}(\mathbf{A})^2 - \frac{\|\mathbf{E}_k\|_F^2}{r}}\|\mathbf{E}_k\|_F\right). \tag{28}$$

The function $f(x) = 1 + \frac{(2\sigma_{\max}+x)}{\sigma_{\min}(\mathbf{A})^2 - \frac{x}{r}}x$ is a strictly increasing function of x in the interval $[0, \sigma_{\min}(\mathbf{A})^2]$. Since $\mathbf{E}_0 = \mathcal{P}_{\Omega^c}(A)$, by the theorem's assumption $\alpha f(\|\mathbf{E}_0\|_F) < 1$. Therefore, by defining $\mu_k = \alpha f(\|\mathbf{E}_k\|_F)$, it can be easily seen that $\mu_0 < 1$ and

$$\|\mathbf{E}_1\|_F \leq \mu_0\|\mathbf{E}_0\|_F \leq \|\mathbf{E}_0\|_F. \tag{29}$$

Therefore $\mu_1 < \mu_0$ and so $\|\mathbf{E}_2\|_F \leq \mu_1\|\mathbf{E}_1\|_F \leq \mu_0^2\|\mathbf{E}_0\|_F$.

Using a simple induction yields

$$\|E_k\|_F \leq \mu_0^k\|\mathbf{E}_0\|_F \to 0 \quad k \to \infty. \tag{30}$$

combining (30) with (28) give $\lim_{k\to\infty}\|\mathbf{A} - \mathbf{B}_k\|_F = 0$, which guarantees the convergence of the algorithm. □

References

[1] M. A. Maddah-Ali and U. Niesen, "Decentralized coded caching attains order-optimal memory-rate tradeoff," *IEEE/ACM Transactions on Networking*, vol. 23, no. 4, pp. 1029–1040, 2015.

[2] Y. Birk and T. Kol, "Informed-source coding-on-demand (ISCOD) over broadcast channels," in *Proc. 17th Annual IEEE Conference Comput. Commun. (INFOCOM), San Francisco, CA*, vol. 3, 1998, pp. 1257–1264.

[3] Z. Bar-Yossef, Y. Birk, T. S. Jayram, and T. Kol, "Index coding with side information," *IEEE Transactions on Information Theory*, vol. 57, no. 3, pp. 1479–1494, 2011.

[4] B. Recht, M. Fazel, and P. A. Parrilo, "Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization," *SIAM review*, vol. 52, no. 3, pp. 471–501, 2010.

[5] M. Langberg and A. Sprintson, "On the hardness of approximating the network coding capacity," *IEEE Transactions on Information Theory*, vol. 57, no. 2, pp. 1008–1014, 2011.

[6] E. Candès and B. Recht, "Exact matrix completion via convex optimization," *ACM Communications*, vol. 55, no. 6, pp. 111–119, 2012.

[7] X. Huang and S. E. Rouayheb, "Index coding and network coding via rank minimization," in *2015 IEEE Information Theory Workshop*, 2015, pp. 14–18.

[8] H. Esfahanizadeh, F. Lahouti, and B. Hassibi, "A matrix completion approach to linear index coding problem," in *2014 IEEE Information Theory Workshop*, 2014, pp. 531–535.

[9] A. Esmaeili, E. Asadi, and F. Marvasti, "Iterative null-space projection method with adaptive thresholding in sparse signal recovery and matrix completion," *arXiv preprint arXiv:1610.00287*, 2016.

[10] Y. Yu, T. Wang, R. J. Samworth *et al.*, "A useful variant of the davis–kahan theorem for statisticians," *Biometrika*, vol. 102, no. 2, pp. 315–323, 2015.

[11] J. N. Franklin, *Matrix theory*. Courier Corporation, 2012.

ISBN 978-0-9928626-7-1 © EURASIP 2017