SPARSE AND LOW-RANK OPTIMIZATION FOR PLIABLE INDEX CODING

Tao Jiang, Yuanming Shi

School of Information Science and Technology, ShanghaiTech University, Shanghai, China. E-mail: {jiangtao1,shiym}@shanghaitech.edu.cn

ABSTRACT

In this paper, we investigate the pliable index coding problem, where clients are interested in receiving any messages (instead of specific messages) that they do not have. The motivating applications including caching networks, recommendation systems and distributed computing systems, where the clients are happy to receive any messages not available in them. However, the pliable index coding problem turns out to be computationally intractable, for which we propose a novel sparse and low-rank optimization framework to assist efficient algorithms design in real field, thereby minimizing the number of channel uses for message delivery. To address the nonconvex challenges in this framework, we further propose the alternating projection algorithm to solve the sparse and low-rank optimization problem with local convergence guarantees. Simulation results demonstrate that the number of channel uses can be significantly reduced for message delivery via the sparse and low-rank optimization.

Index Terms— Pliable index coding, sparse and low-rank optimization, alternating projection method.

1. INTRODUCTION

With the explosive growth of mobile handsets, as well as diversified services and data intensive intelligent applications, enabling ultra-high data rate and ultra-low latency content delivery is a crucial requirement for future communication networks [1]. Meanwhile, the services of communication networks are increasingly becoming entertainment-oriented, moving away from satisfying a specific request towards satisfying the content-type traffic (e.g., caching networks [2], recommendation systems [3]). This content-type communication network pervades tremendous applications in smart search engines and recommendation systems (e.g., Google and Pandora) [4]. For example, if we search for the latest news on the Internet, we do not care which specific news we receive and are happy to obtain any latest news we do not have. Furthermore, this type of communication scenario also exists in distributed computing systems with data shuffling to

improve the statistical performance for distributed training of machine models [5].

However, communication networks today are mainly used to serve for conveying specific messages to the receivers. This is fundamentally different from the content-type traffic networks. Furthermore, side information also plays a vital role for reducing the number of transmissions in communication networks [6]. Based on these observations, we investigate a new communication network with content-type traffic to provide high data rate and low latency communication services by leveraging the side information at receivers. Specifically, this communication scenario is formulated as a pliable index coding problem [7], which consists of a server with m messages and n receivers, where each receiver has the side information with a subset of the messages. The receivers are satisfied if they receive any new messages they do not have. This differs from the index coding problem [6, 8], where each receiver requests specific messages it does not have from the server. Our goal is to find minimum number of broadcasting transmissions so that all the receivers are satisfied in the pliable index coding scenario.

However the pliable index coding problem turns out to be computationally intractable to find the optimal solutions [7]. Recently, a polynomial time algorithm building on an algebraic decoding criterion for pliable index coding has been proposed [9] in finite field. Instead, inspired by the recent success of generalized low-rank models for index coding problem [10, 11] in real field for wireless communications, in this paper, we propose a novel sparse and low rank optimization framework for pliable index coding to minimize the number of channel uses for message delivery. Specifically, the sparsity of this model represents the messages request pattern at receivers, while the low-rankness of this model represents the number of channel uses for messages delivery. The proposed framework can assist efficient algorithms in real field with applications in wireless communication scenarios.

Unfortunately, the resulting sparse and low-rank optimization model raises a unique challenge due to a non-convex objective function (rank) and a non-convex constraint (ℓ_0). One of the popular approach is convex relaxation based on the convex surrogates ℓ_1 -norm and nuclear norm, yielding polynomial time complexity algorithms [12]. However, this approach is inapplicable in our problem as it always returns

This work was supported in part by the National Nature Science Foundation of China under Grant 61601290 and in part by the Shanghai Sailing Program under Grant 16YF1407700.

the full-rank identity matrix in some cases. To address this issue, we propose an alternating projection algorithm to solve the sparse and low-rank optimization problem, which leads to a feasibility detection problem between two non-convex sets. Although the convergence of alternating projection algorithm for convex sets is well established [13], the convergence results for non-convex sets are rather limited. Based on the observations that the semi-algebraic property in the sparse and low-rank models, we establish the local convergence guarantees for our proposed algorithm. Simulation results demonstrate that the number of channel uses for message delivery can be significantly reduced via the sparse and low-rank optimization.

2. PROBLEM STATEMENT

2.1. Pliable Index Coding Problem

We consider the communication network with content-type traffic (e.g., caching network [2], recommendation systems [14] and distributed computing systems [5]) and exploit the side information to help message delivery. Specifically, the pliable index coding problem consists of a set of K independent messages $\{W_1, W_2, \dots, W_K\}$ and a set of K receivers. Let \mathcal{V}_k denote the index set of messages at the k-th receiver. The k-th receiver has a subset of messages $W_{[\mathcal{V}_k]} = \{W_j, j \in \mathcal{V}_k\}$ as the side information and is satisfied to receive any message it does not have.

Let S be the choice of a finite alphabet set. The coding function $f(W_1, \dots, W_K) = x$ maps all the messages to the sequence of transmitted symbols, where $x \in S^N$ is the sequence of symbols transmitted over N channel uses. Here, each message W_i is a random variable uniformly distributed over the set $\{1, 2, \dots, |S|^{NR_i}\}$ with $|S|^{NR_i}$ as an integer and $R_i \in \mathbb{R}_+$. At the k-th receiver, the decoding function for some requested message W_j with $j \notin \mathcal{V}_k$ is given by $g_{k,j}(x, \mathcal{V}_k) = \hat{W}_j$. The probability of decoding error is given by $p_e = 1 - \Pr\{\hat{W}_j = W_j, \forall j\}$. Let $(S, N, (R_1, \dots, R_K))$ denote the above coding scheme. The rate tuple (R_1, \dots, R_K) is said to be achievable if for every $\epsilon, \delta > 0$ there exists a $(S, N, (\bar{R}_1, \dots, \bar{R}_K))$ coding scheme, for some S, n, such that $\bar{R}_i \geq R_i - \delta, \forall i$, and the error probability $P_e \leq \epsilon$.

2.2. Linear Coding Schemes

We focus on designing the linear encoding and decoding schemes in real field. Specifically, let $v_i \in \mathbb{R}^N$ and $u_i \in \mathbb{R}^N$ be the precoding vector and the decoding vector for message W_i , respectively. Without loss of generality, we represent each message W_i by a scalar symbol $s_i \in \mathbb{R}$. The received signal at the k-th receiver is given by

$$\boldsymbol{y}_k = \sum_{j=1}^K \boldsymbol{v}_j s_j, \quad \forall k = 1, \cdots, K.$$
 (1)

The k-th receiver is happy to receive any messages it does not have. Therefore, the decoding operation at the k-th receiver for some requested message W_j with $j \notin \mathcal{V}_k$ is given by

$$\hat{s}_j = (\boldsymbol{u}_k^T \boldsymbol{v}_j)^{-1} \boldsymbol{u}_k^T \left(\boldsymbol{y}_k - \sum_{i \in \mathcal{V}_k} \boldsymbol{v}_i s_i \right), \quad j \notin \mathcal{V}_k.$$
 (2)

The above decoding operation is achieved by the following interference alignment condition

$$\boldsymbol{u}_k^T \boldsymbol{v}_j \neq 0, \quad \text{for some } j \notin \mathcal{V}_k,$$
 (3)

$$\boldsymbol{u}_k^T \boldsymbol{v}_i = 0, \quad \forall i \neq j, i \notin \mathcal{V}_k, \tag{4}$$

which implies that the desired messages W_j with $j \notin \mathcal{V}_k$ is preserved and all the other messages not in \mathcal{V}_k are eliminated. If the interference alignment condition (3) and (4) are satisfied, then the data rate tuple $(\frac{1}{N}, \dots, \frac{1}{N})$ for each message can be achieved [15, 10]. Therefore, the overall achievable data rate is given by K/N. Our goal, in this paper, is to find a linear coding scheme to maximize the achievable data rate 1/N (normalized by K) by minimizing the channel uses N. Note that, in the index coding scenario for topological interference alignment, condition (3) becomes $u_k^T v_k \neq 0, \forall k$ [10].

3. PROBLEM FORMULATION

In this section, we propose a sparse and low rank optimization approach to find the linear encoding and decoding vectors, thereby minimizing the number of channel uses for messages delivery. This is achieved by rewriting the interference alignment conditions (3) and (4) into a rank minimization problem with sparsity constraints. Specifically, let $X_{ij} =$ $u_i^T v_j, \forall i, j = 1, \dots, K$. Define the $K \times K$ matrix X = $[X_{ij}]$, we have the rank of matrix X as rank(X) = N. To minimize the number of channel uses, we propose to solve the following sparse and low-rank optimization problem:

$$\begin{array}{ll} \underset{\boldsymbol{X} \in \mathbb{R}^{K \times K}}{\text{minimize}} & \operatorname{rank}(\boldsymbol{X}) \\ \text{subject to} & \|\boldsymbol{x}_{[\mathcal{V}_{k}^{c}]}^{k}\|_{0} = 1, \quad k = 1, 2, \cdots, K, \end{array}$$
(5)

where \boldsymbol{x}^k denotes the k-th row of matrix \boldsymbol{X} and $\mathcal{V}_k^c = [K] \setminus \mathcal{V}_k$ with $[K] = \{1, 2, \dots, K\}$.

Unfortunately, problem (5) turns out to be a non-convex and highly intractable problem. One of the popular methods to solve the sparse and low rank optimization problems is using the convex surrogates ℓ_1 norm and nuclear norm [12]:

$$\underset{\boldsymbol{X} \in \mathbb{R}^{K \times K}}{\text{minimize}} \quad \|\boldsymbol{X}\|_* + \lambda \sum_{k=1}^{K} \|\boldsymbol{x}_{[\mathcal{V}_k^c]}^k\|_1, \tag{6}$$

where $\lambda > 0$ is a regularized parameter, $\|\boldsymbol{x}\|_1 = \sum_i |x_i|$ and $\|\boldsymbol{X}\|_*$ is the nuclear norm of \boldsymbol{X} , i.e., the summation of the singular values of \boldsymbol{X} . However, based on the fact that $\|\boldsymbol{X}\|_* \geq \text{Tr}(\boldsymbol{X})$, problem (6) will always return the full-rank solution $\boldsymbol{X} = \boldsymbol{I}_K$ for some cases, e.g., $\mathcal{V}_k = [K] \setminus \{k\}$ for $k = 1, \dots, K$. We thus propose to solve the nonconvex optimization problem (5) directly.

Specifically, to find the minimal rank, we propose to solve a sequence of fixed rank optimization problems as follows:

find
$$X$$

subject to rank $(X) \leq r$, (7)
 $\|\boldsymbol{x}_{[\mathcal{V}_{k}^{L}]}^{k}\|_{0} = 1, \quad k = 1, 2, \cdots, K.$

By increasing r, we shall find the linear coding matrix with the smallest channel uses r to satisfy the sparsity constraints, e.g., users are satisfied to receive any messages that they do not have.

4. ALTERNATING PROJECTION ALGORITHM

In this section, we propose the alternating projection algorithm to solve problem (7) with local convergence guarantees. This is based on the observation that projecting any given matrix onto the low-rank constraint set or the ℓ_0 -ball constrain set can be achieved using simple analytical expressions.

4.1. Alternating Projection Algorithm

Define the rank constraint set $S_r = \{ \boldsymbol{X} : \operatorname{rank}(\boldsymbol{X}) \leq r \}$ and the ℓ_0 -ball constraint set $S_p = \{ \boldsymbol{X} : \| \boldsymbol{x}_{[\mathcal{V}_k^c]}^k \|_0 \leq 1, k = 1, 2, \cdots, K \}$. Problem (7) can be reformulated as finding a common point of the following two sets, i.e.,

find
$$X \in \mathcal{S}_r \cap \mathcal{S}_p$$
, (8)

where we relaxed the fixed-sparsity constrain in problem (7) into a ℓ_0 -ball constrain.

Define the projection of a point $X \notin S$ onto a given set S as

$$\Pi_{\mathcal{S}}(\boldsymbol{Z}) := \underset{\boldsymbol{X} \in \mathcal{S}}{\arg\min} \|\boldsymbol{X} - \boldsymbol{Z}\|_{F}^{2}.$$
(9)

We adopt the strategy of alternating projection onto S_r and S_p to find a common point in the intersection of the two sets, which is presented in Algorithm 1.

By Eckart-Young theorem [16], the projection of $X \notin S_r$ onto S_r can be computed via truncated SVD of X

$$\Pi_{\mathcal{S}_r}(\boldsymbol{X}) = \sum_{i=1}^r \sigma_i \boldsymbol{u}_i \boldsymbol{v}_i^T, \qquad (10)$$

where $\{\sigma_i\}_{i=1}^r$, $\{u_i\}_{i=1}^r$ and $\{v_i\}_{i=1}^r$ are the *r* largest singular values and the corresponding left and right singular vectors of X. The projection of $Y \notin S_p$ onto S_p can be computed by

$$\boldsymbol{x}_{[\mathcal{V}_{k}^{c}]}^{k} = \Pi_{\ell_{0} \leq 1}(\boldsymbol{y}_{[\mathcal{V}_{k}^{c}]}^{k}), \quad k = 1, 2, \cdots, K,$$
 (11)

A	Algorithm I: Alternating Projection for Problem
(8)	
	Input : Side information \mathcal{V}_k , $k = 1, \cdots K$, rank r .
1	Initialization: Let X_0 be a random matrix with
	$\operatorname{rank}(\boldsymbol{X}_0) = r.$
2	for $i = 0, 1, 2, \cdots$ do
3	$egin{array}{c} m{Y}_i = \Pi_{m{\mathcal{S}}_r}(m{X}_i) \end{array}$
4	$oldsymbol{X}_{i+1} = \Pi_{\mathcal{S}_p}(oldsymbol{Y}_i)$
5	end

where the projection of a vector \boldsymbol{y} onto the set $\ell_0 \leq 1 := \{\boldsymbol{x} : \|\boldsymbol{x}\|_0 \leq 1\}$ is the hard thresholding operator. That is, we sort the coefficients of \boldsymbol{y} in decreasing magnitude and keep the largest one and set the remaining entries to be zeros.

4.2. Convergence Guarantees

The fact that alternating projection onto two convex sets converges to a point in the intersection of these two sets (if their intersection is non-empty) was previously established in [13]. Recently, the local convergence results of the alternating projection for semi-algebraic sets have be established in [17], and the convergence rate results under a further regularity condition are provided in [18]. Please refer to [19] for a comprehensive review. Exploiting the fact that the rank constraint set S_r satisfies the prox regularity and both S_r and S_p are semi-algebraic sets, we shall establish local convergence of the alternating projection algorithm in Algorithm 1, as presented in the following theorem.

Theorem 1. Let $S^* \in S_r \cap S_p$, then there exists a neighborhood U of S^* such that every sequence of alternating projections X_i, Y_i which enters U converges to some $X^* \in S_r \cap S_p$. Furthermore, suppose $\operatorname{rank}(S^*) = r$, then $||X_i - X^*||_F = O(i^{-\rho})$ for some $\rho \in (0, \infty)$.

Proof. Please refer to Section 7 for the details.

5. NUMERICAL RESULTS

In this section, we simulate the proposed sparse and low-rank optimization approach for the pliable index coding problem. This needs to solve a sequence of problem (8) via increasing rank r to find the minimal channel uses based on the proposed alternating projection algorithm.

For problem (8) with K = 15, the side information at each receiver is generated uniformly at random with the same size. The maximum number of iterations of the proposed alternating projection algorithm is set to be 10^4 and we terminate the algorithm when the error $\|X_i - Y_i\|_F \le 10^{-12}$. We compare the pliable index coding based transmission scheme with the index coding based transmission scheme [20]. We plot the channel uses versus the size of side information in Fig.1 (a) and each point is averaged for 200 realizations.



(a) Channel uses versus different size of side information



(b) The convergence rates of alternating projection algorithm

Fig. 1. Numerical Results

Fig.1 (a) demonstrates that the pliable index coding based transmission scheme needs fewer channel uses compared to the index coding based scheme for message delivery, especially when the size of side information at each receiver is small. Fig.1 (b) illustrates the convergence behavior of the alternating projection algorithm for problem (8) with K = 17, r = 12 and the size of the side information is set to be 8. For all the realization with different random initial points, we can see that the algorithm converges to the intersection of the two sets rapidly and linearly.

6. CONCLUSIONS

In this paper, we propose a novel sparse and low rank optimization framework to solve the pliable index coding problem in real field. To address the non-convex challenges, we propose the alternating projection algorithm to solve the sparse and low-rank optimization problem with local convergence guarantees. Simulation results demonstrated that he number of channel uses can be significantly reduced for message delivering using pliable index coding.

7. PROOF OF THEOREM 1

To prove theorem 1, we first provide the following lemma.

Lemma 1 (Semi-algebraic intersections [17]). Consider two nonempty closed semi-algebraic sets X, Y in a Euclidean space with X bounded. If the method of alternating projections starts in Y and near X, then the distance of the iterates to the intersection $X \cap Y$ converges to zero, and hence every limit point lies in $X \cap Y$.

We will first give the definition of the semi-algebraic sets, followed by the proof that the sets S_p and S_r are both semi-algebraic sets.

Definition 1 (Semi-algebraic set). A subset S of \mathbb{R}^n is called the semi-algebraic set if there exists a finite number of real polynomial functions g_{ij} and h_{ij} such that

$$S = \bigcup_{j} \bigcap_{i} \{ \boldsymbol{u} \in \mathbb{R}^{n} : g_{ij}(\boldsymbol{u}) = 0, h_{ij}(\boldsymbol{u}) < 0 \}.$$

The next lemma establishes that the set S_p and S_r are both semi-algebraic sets.

Lemma 2. The sets $S_p = \{ \boldsymbol{X} : \|\boldsymbol{x}_{[\mathcal{V}_k^c]}^k\|_0 \leq 1, k = 1, 2, \cdots, K \}$ and $S_r = \{ \boldsymbol{X} : \operatorname{rank}(\boldsymbol{X}) \leq r \}$ are both semialgebraic.

Proof. For the set S_p , we can represent it as

$$S_p = \bigcap_{k=1}^{K} \bigcup_{j=0,1} L_k \triangleq \left\{ \boldsymbol{X} : \|\boldsymbol{x}_{[\mathcal{V}_k^c]}^k\|_0 = j \right\}.$$
(12)

Let $\mathcal{I} = \{I : I \subset \{1, \cdots, |\mathcal{V}_k^c|\}, |I| = 1\}$, then

$$L_k = \bigcup_{I \in \mathcal{I}} \left\{ \boldsymbol{X} : (\boldsymbol{x}_{[\mathcal{V}_k^c]}^k)_{I^c} = 0 \right\},\,$$

which is a semi-algebraic set. Therefore, S_p is a semi-algebraic set.

For the set S_r , we define the polynomial map $\psi_r : (\mathbb{R}^N \times \mathbb{R}^N)^r \to \mathbb{R}^{N \times N}$ of the form

$$\psi_r(\boldsymbol{u}_1, \boldsymbol{v}_1; \dots; \boldsymbol{u}_r, \boldsymbol{v}_r) = \sum_{i=1}^r \boldsymbol{u}_i \times \boldsymbol{v}_i^T.$$
(13)

It is clear that the image of ψ_r is exactly $S_r = \{ X : rank(X) \leq r \}$. Thus S_r is semi-algebraic according to the Tarski—Seidenberg Theorem [21], which states that the image of the polynomial map from semi-algebraic set is also a semi-algebraic set.

It has been show that the rank constraint set S_r is proxregular at all points X^* with rank $(X^*) = r$ [22, Proposition 3.8]. This together with [18, Proposition 4] and [18, Corollary 9] gives the convergence rates results stated in theorem 1.

8. REFERENCES

- Y. Shi, J. Zhang, W. Chen, and K. B. Letaief, "Generalized sparse and low-rank optimization for ultra-dense networks," *IEEE Commun. Mag.*, vol. 56, pp. 42–48, Jun. 2018.
- [2] P. Sermpezis, T. Giannakas, T. Spyropoulos, and L. Vigneri, "Soft cache hits: Improving performance through recommendation and delivery of related content," *IEEE J. Sel. Areas Commun.*, 2018.
- [3] L. Song and C. Fragouli, "Making recommendations bandwidth aware," *IEEE Trans. Inf. Theory*, 2018.
- [4] L. Song and C. Fragouli, "Content-type coding," in Network Coding (NetCod), 2015 International Symposium on, pp. 31–35, IEEE, 2015.
- [5] L. Song, C. Fragouli, and T. Zhao, "A pliable index coding approach to data shuffling," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, pp. 2558–2562, Jun. 2017.
- [6] Z. Bar-Yossef, Y. Birk, T. Jayram, and T. Kol, "Index coding with side information," *IEEE Trans. Inf. Theory*, vol. 57, no. 3, pp. 1479–1494, 2011.
- [7] S. Brahma and C. Fragouli, "Pliable index coding," *IEEE Trans. Inf. Theory*, vol. 61, no. 11, pp. 6192–6203, 2015.
- [8] H. Maleki, V. R. Cadambe, and S. A. Jafar, "Index coding—an interference alignment perspective," *IEEE Trans. Inf. Theory*, vol. 60, no. 9, pp. 5402–5432, 2014.
- [9] L. Song and C. Fragouli, "A polynomial-time algorithm for pliable index coding," *IEEE Trans. Inf. Theory*, vol. 64, no. 2, pp. 979–999, 2018.
- [10] Y. Shi, J. Zhang, and K. B. Letaief, "Low-rank matrix completion for topological interference management by riemannian pursuit," *IEEE Trans. Wireless Commun.*, vol. 15, no. 7, pp. 4703–4717, 2016.
- [11] Y. Shi, B. Mishra, and W. Chen, "Topological interference management with user admission control via riemannian optimization," *IEEE Trans. Wireless Commun.*, vol. PP, no. 99, pp. 1–1, 2016.
- [12] S. Oymak, A. Jalali, M. Fazel, Y. C. Eldar, and B. Hassibi, "Simultaneously structured models with application to sparse and low-rank matrices," *IEEE Trans. Inf. Theory*, vol. 61, no. 5, pp. 2886–2908, 2015.
- [13] L. Gubin, B. Polyak, and E. Raik, "The method of projections for finding the common point of convex sets," USSR Comput. Mathematics and Math. Physics, vol. 7, no. 6, pp. 1–24, 1967.

- [14] L. Song and C. Fragouli, "Making recommendations bandwidth aware," in *Proc. IEEE Int. Symp. Inf. Theory* (*ISIT*), pp. 2243–2247, Jun. 2017.
- [15] S. A. Jafar, "Topological interference management through index coding," *IEEE Trans. Inf. Theory*, vol. 60, no. 1, pp. 529–568, 2014.
- [16] C. Eckart and G. Young, "The approximation of one matrix by another of lower rank," *Psychometrika*, vol. 1, no. 3, pp. 211–218, 1936.
- [17] D. Drusvyatskiy, A. D. Ioffe, and A. S. Lewis, "Transversality and alternating projections for nonconvex sets," *Found. Comput. Math.*, vol. 15, no. 6, pp. 1637–1651, 2015.
- [18] D. Noll and A. Rondepierre, "On local convergence of the method of alternating projections," *Found. Comput. Math.*, vol. 16, no. 2, pp. 425–455, 2016.
- [19] A. Y. Kruger, D. R. Luke, and N. H. Thao, "Set regularities and feasibility problems," *Math. Programming*, vol. 168, no. 1-2, pp. 279–311, 2018.
- [20] X. Huang and S. El Rouayheb, "Index coding and network coding via rank minimization," in *Proc. IEEE Int. Symp. Inf. theory Workshop (ITW)*, pp. 14–18, 2015.
- [21] M. Coste, "An introduction to semialgebraic geometry," *RAAG network school*, vol. 145, p. 30, 2002.
- [22] D. R. Luke, "Prox-regularity of rank constraint sets and implications for algorithms," *J. of Math. Imag. and Vision*, vol. 47, no. 3, pp. 231–238, 2013.