

TWO-WAY SPECTRUM PURSUIT FOR *CUR* DECOMPOSITION AND ITS APPLICATION IN JOINT COLUMN/ROW SUBSET SELECTION

Ashkan Esmaeili*, Mohsen Joneidi*, Mehrdad Salimitari, Umar Khalid, and Nazanin Rahnavard

Department of Electrical & Computer Engineering
University of Central Florida, Orlando, USA

Emails: {ashkan.esmaeili@, joneidi@ece, m.salimi@, umarkhalid@, nazanin@ece}.ucf.edu

ABSTRACT

The problem of simultaneous column and row subset selection is addressed in this paper. The column space and row space of a matrix are spanned by its left and right singular vectors, respectively. However, the singular vectors are not within actual columns/rows of the matrix. In this paper, an iterative approach is proposed to capture the most structural information of columns/rows via selecting a subset of actual columns/rows. This algorithm is referred to as two-way spectrum pursuit (TWSP) which provides us with an efficient solution for the CUR matrix decomposition. TWSP is applicable in a wide range of applications since it enjoys a linear complexity w.r.t. number of original columns/rows. We demonstrated the application of TWSP for joint channel and sensor selection in cognitive radio networks and efficient supervised data reduction.

Index Terms— Column and rows subset selection, CUR matrix decomposition, spectrum pursuit

1. INTRODUCTION

Matrix factorization provides a concise representation of data. Despite desirable uniqueness conditions and computational simplicity of the well-known singular value decomposition (SVD), it comes with some fundamental shortcomings. The intrinsic structure of data is not inherited to the singular components. Moreover, SVD implies orthogonality on the components which is irrelevant to the underlying structure of the original data. This enforced structure makes the bases, a.k.a. singular vectors, hard to interpret [8]. On the other hand, it is shown that borrowing basic vectors from the actual samples of a dataset provides a robust representation, which can be employed in interesting applications where sampling is their key factor [19]. This problem is studied under the literature of column subset selection problem (CSSP) [5], [2] and CUR decomposition [15], [3]. A general problem for CSSP and CUR can be written in the following form:

$$(\mathbb{S}^c, \mathbb{S}^r) = \underset{\mathbb{S}^c, \mathbb{S}^r}{\operatorname{argmin}} \|\mathbf{X} - \pi_{\mathbb{S}^r}^r(\pi_{\mathbb{S}^c}^c(\mathbf{X}))\|_F^2, \quad (1)$$

where, $\mathbf{X} \in \mathbb{R}^{N \times M}$ is the data matrix containing M data points of dimension N . Here, $\pi_{\mathbb{S}^c}^c(\cdot)$ and $\pi_{\mathbb{S}^r}^r(\cdot)$ indicate column

space projection and row space projection, respectively. More specifically, these operators project all columns (rows) to a low-dimensional subspace spanned by selected columns (rows) of matrix $\mathbf{X}$ indexed by the set $\mathbb{S}^c$ ($\mathbb{S}^r$). The chronological order in applying $\pi_{\mathbb{S}^c}^c(\cdot)$ and $\pi_{\mathbb{S}^r}^r(\cdot)$ does not affect the problem since these operators are linear. Moreover, substituting $\pi_{\mathbb{S}^r}^r$ with the identity projection simplifies the problem to CSSP. Fig. 1 illustrates the structure of the CUR matrix decomposition as a self-representative approach.

A versatile metric for evaluating the performance of a data subset selection algorithm can be defined by the approximation error resulted from the projection of the entire data to the span of selected rows/columns. How close to the optimal selection an algorithm can reach, is determined by comparing its approximation error to the best low-rank approximation error specified by the spectral decomposition. Recently, a fast and accurate algorithm for solving CSSP has been proposed which is called spectrum pursuit (SP) [11]. Inspired by SP, we propose a new algorithm to address the more general case of the CUR matrix decomposition. The main contributions of our paper are summarized as follows:

- A novel algorithm for CUR decomposition, referred to as two-way spectrum pursuit (TWSP), is proposed. TWSP provides an efficient solution for CUR decomposition.
- TWSP enjoys a linear complexity w.r.t. the number of columns and the number of rows of a matrix.
- TWSP is a parameter-free algorithm that only requires the number of desired columns and rows for selection. Thus, TWSP does not require any parameter fine-tuning.
- The TWSP algorithm is put to the test and investigated in a set of synthetic and real experiments.
- The role of the core matrix $\mathbf{U}$ in *CUR* decomposition is illustrated which shows the connection between selected columns and rows. Based on analysis of $\mathbf{U}$, an interesting application for joint sensor/channel selection for spectrum sensing in cognitive radio networks is presented.

In this paper, the m^{th} column of matrix $\mathbf{X}$ and the n^{th} row of matrix $\mathbf{X}$ are denoted by $\mathbf{X}(:, m)$ and $\mathbf{X}(n, :)$ respectively.

The rest of paper is organized as follows. Sec. 2 studies the problem of column subset selection as the conventional one-way selection scheme. In Sec. 3, our main work is presented for joint column and row subset selection. Experimental results are exhibited in Sec. 4 and Sec. 5 concludes the paper.

* equal contribution. This material is supported by the National Science Foundation under Grant No. ECCS-1810256 and CCF-1718195.

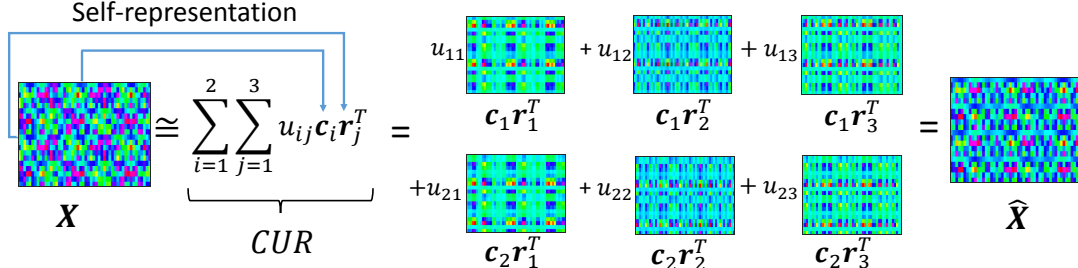


Fig. 1: Two columns and three rows of matrix $\mathbf{X}$ are selected and organized in the matrices $\mathbf{C}$ and $\mathbf{R}$, respectively. The outer product of each pair of a selected column and a selected row constructs a rank-1 matrix, i.e., $\mathbf{c}_i \mathbf{r}_j^T$. The contribution amount of each pair is reflected in variable u_{ij} . The core matrix $\mathbf{U}$ is the collection of all u_{ij} 's. The goal is to minimize $\|\mathbf{X} - \hat{\mathbf{X}}\|_F$, where $\hat{\mathbf{X}} = \mathbf{CUR}$.

2. COLUMN SUBSET SELECTION PROBLEM

Selecting the most diverse subset of data in an optimal sense has been studied vastly [5], [13], [12]. However, these methods do not guarantee that the unselected columns are *well represented* by the selected ones. Further, outliers are selected with a high probability using such algorithms due to their diversity [11]. A more effective approach is selecting some *representatives* which are able to approximate the rest of data accurately [7] as defined as a special case of (1). This is an NP-hard problem [17] and there are several efforts for solving this problem [1], [5], [16], [20]. There are computationally expensive approaches based on convex relaxation [7], [6] that are not computationally feasible for large datasets since their complexity is of order $O(M^3)$, where M is the number of original columns. Recently, we proposed a new algorithm for solving CSSP with a linear complexity, which is called Spectrum Pursuit (SP) [11]. The SP algorithm finds K columns of $\mathbf{X}$ such that their span is close to that of the best rank- K approximation of $\mathbf{X}$. SP is an iterative approach where at each iteration one sample selection is optimized such that the ensemble of selected samples describes the whole dataset more accurately. SP finds representatives such that the column space is spanned accurately via consecutive rank-1 approximations as theoretically analyzed in [10]. In the present paper, we extend the SP algorithm for a joint selection of columns and rows such that their outer product can represent the whole matrix accurately. A naive approach is applying the SP algorithm on the matrix of interest to select a subset of columns and then applying SP on its transpose in order to find a subset of rows. However, this approach is not efficient and we will compare it with our proposed approach which is optimized through a joint representation of selected columns and rows.

3. TWO-WAY SPECTRUM PURSUIT

The introduced joint column/row subset selection in (1) can be written as a CUR decomposition in the following form in which factor matrices must be drawn from actual columns/rows of the original matrix as

$$\begin{aligned} (\mathbf{C}, \mathbf{U}, \mathbf{R}) &= \underset{\mathbf{C}, \mathbf{U}, \mathbf{R}}{\operatorname{argmin}} \|\mathbf{X} - \mathbf{CUR}\|_F^2, \\ \text{s.t. } &\mathbf{c}_k \in \mathbb{X}_c \text{ and } \mathbf{r}_k \in \mathbb{X}_r. \end{aligned} \quad (2)$$

In this problem, $\mathbb{X}_c$ and $\mathbb{X}_r$ indicate the set of normalized columns and rows of matrix $\mathbf{X}$, respectively. Here, $\mathbf{c}_k$ and $\mathbf{r}_k^T$

denote the k^{th} column and the k^{th} row of $\mathbf{C}$ and $\mathbf{R}$, respectively. In other words, $\mathbb{X}_c = \{\mathbf{X}(:, m) / \|\mathbf{X}(:, m)\|_2; \forall m\}$ for all columns and $\mathbb{X}_r = \{\mathbf{X}(n, :) / \|\mathbf{X}(n, :)\|_2; \forall n\}$ for all rows. Please note that replacing constraints in (2) with orthogonality constraint on $\mathbf{c}_k$'s and on $\mathbf{r}_k$'s results in the truncated singular value decomposition (SVD) with K most significant components. In this case, $\mathbf{C}$ and $\mathbf{R}$ contain the first K left singular vectors and the first K right singular vectors of $\mathbf{X}$, respectively. Moreover, the core matrix will be diagonal and the entries will be singular values with diagonal entries as singular values. However, the underlying constraints in (2) turn the problem into a joint subset of row and column selection problem instead of matrix low-rank approximation problem.

To solve this complicated problem, we split it into two consecutive problems for optimization on the k^{th} selected column/row. Our optimization approach is alternative, i.e., a random subset of columns and rows are picked. Then, one column or row is considered to be replaced with a more efficient one at each iteration. Since scaling a vector does not change its span, without loss of generality we assume that the column or the row subject of the optimization lie on the unit sphere. At each iteration, a rank-1 component is optimized characterized by $\mathbf{c} \mathbf{g}^T$ or $\mathbf{h} \mathbf{r}^T$ given by

$$\underset{\mathbf{c}, \mathbf{W}, \mathbf{g}}{\operatorname{argmin}} \underbrace{\|\mathbf{X} - \mathbf{C}_k \mathbf{W} \mathbf{R} - \mathbf{c} \mathbf{g}^T\|_F^2}_{\mathbf{E}^c} \text{ s.t. } \|\mathbf{c}\|_2 = 1 \quad (3a)$$

$$\underset{\mathbf{h}, \mathbf{Y}, \mathbf{r}}{\operatorname{argmin}} \underbrace{\|\mathbf{X} - \mathbf{C} \mathbf{Y} \mathbf{R}_k - \mathbf{h} \mathbf{r}^T\|_F^2}_{\mathbf{E}^r} \text{ s.t. } \|\mathbf{r}\|_2 = 1 \quad (3b)$$

$\mathbf{R}$ and $\mathbf{C}$ are the CUR components to include the selected rows and columns, respectively. These are initialized randomly as presented in Alg. 1. The selected samples replace the initialization through the iterations governed by Eqs. (3a), (3b), (4a), (4b) to be presented as follows. Matrix $\mathbf{C}_k$ is the set of selected columns except the k^{th} one and $\mathbf{R}_k$ is the set of selected rows except the k^{th} row. The first subproblem can be solved easily w.r.t. $\mathbf{c}$ using singular value decomposition. In other words, $\mathbf{c} \mathbf{g}^T$ and $\mathbf{h} \mathbf{r}^T$ are the best rank-1 approximations of the residual $\mathbf{E}^c$ and $\mathbf{E}^r$, respectively. The obtained $\mathbf{c}/\mathbf{r}$ is the best column/row that can be added to the pool of selected columns/rows. However, the obtained vector is not available in the given dataset as a column or row since it is a singular vector which is a function of all columns/rows. The following step re-imposes the underlying constraints at each iteration,

$$\mathbb{S}_k^c = \underset{m}{\operatorname{argmax}} |\mathbf{x}_m^T \mathbf{c}|, \quad \forall \mathbf{x}_m \in \mathbb{X}_c, \quad (4a)$$

$$\mathbb{S}_k^r = \underset{n}{\operatorname{argmax}} |\mathbf{x}_n^T \mathbf{r}|, \quad \forall \mathbf{x}_n \in \mathbb{X}_r. \quad (4b)$$

Here, $\mathbb{S}_k^c$ indicates a singleton that contains the k^{th} selected column and $\mathbb{S}_k^r$ corresponds to the k^{th} selected row. At each iteration, one column or one row is the subject of optimization. The impact of the latest estimation for that column/row on the representation is neglected. Then, an optimized replacement is found. At each iteration of TWSP, sub-problems in (3) are solved and their solutions are matched to the accessible column samples and row samples through matching equations in (4). The new selected column or row is stored in $\mathbb{S}_k^c$ or $\mathbb{S}_k^r$, respectively. At each iteration we need to compute only the first singular vector and there are fast methods to do so [4]. The pair of (3a) and (4a) optimizes and matches a column. Similarly, performing (3b) and (4b) provide us with an optimized row. However, we do not update both of them per each iteration. In fact, which update (column or row) is more effective in each iteration in order to minimize the main cost function should be determined. To this aim, first, we choose a random previously selected column and a random previously selected row. Then, we find the best possible replacement column and the best possible replacement row. Accordingly, we choose whichever that minimizes the cost function more. The best modified column-wise subset is denoted by $\tilde{\mathbb{S}}^c$ and $\tilde{\mathbb{S}}^r$ denotes the best row-wise modified subset. Alg. 1 indicates the steps of the TWSP algorithm. Here, $\dagger$ refers to the Moore–Penrose pseudo-inverse operator. Iterations can be terminated either once CUR decomposition error is saturated or when a maximum number of iterations is reached.

The proposed TWSP provides the selected columns and selected rows in order to form matrix $\mathbf{C}$ and $\mathbf{R}$ in CUR decomposition. The core matrix $\mathbf{U}$ can be estimated as:

$$\mathbf{U} = \mathbf{C}^\dagger \mathbf{X} \mathbf{R}^\dagger. \quad (5)$$

This matrix is a two-way compressed replica of the whole dataset and it contains valuable information in practice as will be discussed in Sec. 4.2. In general, the number of selected columns may differ from the desired number of rows. Here, K_1 refers to the number of columns and K_2 points to the number of rows. It is worthwhile to mention that the complexity order of TWSP is bottle-necked by computational burden for two pseudo-inverses and two singular vectors computation. Thus, the complexity can be expressed as $O(NK_1^2 + MK_2^2 + MN)$ per iteration and the algorithm needs $O(K_1 + K_2)$ iterations (the number of samples we desire to select). Moreover, operator $\operatorname{rnd}(K)$ refers to a random positive integer number less than or equal to K . In the next section, we evaluate the performance of our proposed algorithm.

The CSSP and CUR decomposition problems are NP-hard, i.e., a combinatorial search is required to find the best columns and rows. The proposed TWSP algorithm minimizes the main cost function (1) in practice. However, there is no theoretical guarantee for the convergence of the proposed TWSP Alg. 1.

This algorithm is composed of two phases relating to the columns and the rows, respectively. Per each phase of the TWSP algorithm (we focus on the columns phase w.l.o.g), one column is chosen to have the largest correlation with the minimizer of the Eq. (3a). To facilitate the method with the ability to escape from local minima caused from the greediness in the selection procedure, TWSP is equipped with the option for revising the selection by repeating the procedure for certain times. In order to improve the convergence behavior of TWSP, we evade updating both columns and rows in each iteration. Rather, we prioritize updating a row or a column depending on which exhibits a smaller projection error, i.e., which one is a better minimizer for the cost function per each iteration. The implementation steps of TWSP are summarized in Alg. 1. As seen in Alg. 1, the method alleviates the greediness associated with consecutive selection by removing a previously selected column or row and optimizing it with better replacement. $\mathbf{W}$ and $\mathbf{Y}$ are the core tensors corresponding to the CUR decomposition in Alg. 1.

4. EXPERIMENTAL RESULTS

In order to evaluate TWSP on machine-learning tasks in terms of CUR decomposition accuracy, we apply the proposed TWSP on synthetic data as well as three real applications.

4.1. CUR Decomposition on Synthetic Data

In order to evaluate the general performance of TWSP, we compared it with the state-of-the-art methods for selecting columns and rows. In this regards, we created a 1000×2000 synthetic dataset. The dataset is generated by a rank-30 matrix contaminated with random noise. In Fig. 2, we have illustrated the CUR decomposition error for selecting a subset of rows and columns in the range of 2 to 20 for several selection algorithms. A more accurate algorithm for solving CUR decomposition results in a bigger blue region in Fig. 2. The reconstruction error of CUR is normalized by $\|\mathbf{X}\|_F^2$. We employ SP as the state-of-the-art algorithm for column subset selection [11]. We perform SP on the data matrix $\mathbf{X}$ and its transpose to, respectively, select a subset of columns and rows independently. Then, employing (5) results in a CUR decomposition. We refer to the algorithm in [15] as adaptive CUR. TWSP exhibits the best performance in this experiment. The convergence behavior of TWSP for this experiment is shown in Fig. 3 for selecting 20 columns and 20 rows. The final solution of the TWSP algorithm depends on the initial selected columns and selected rows. However, regardless of the initial condition, the TWSP algorithm minimizes the cost function of CUR decomposition. TWSP prioritizes in selection of columns or rows such that the largest decrease in the projection error is obtained. It is the main feature to obtain convergence in practice.

4.2. Joint Sensor Selection and Channel Assignment

The output products of CUR decomposition are not limited to a subset of columns and rows. In some applications, interestingly, matrix $\mathbf{U}$ is the most important output of a CUR decomposition. Entry (i, j) in $\mathbf{U}$ indicates how important the contribution of the i^{th} column and the j^{th} row is to reconstruct the whole matrix $\mathbf{X}$. This interesting property is utilized for the problem

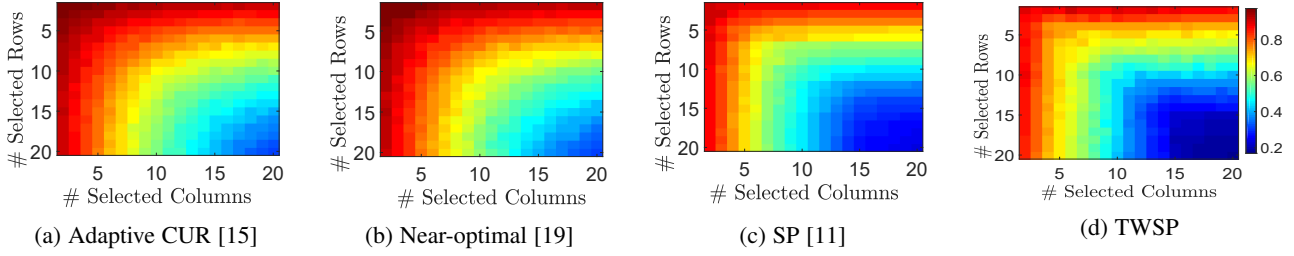


Fig. 2: (a)-(d) Performance comparison in terms of the normalized error of CUR decomposition.

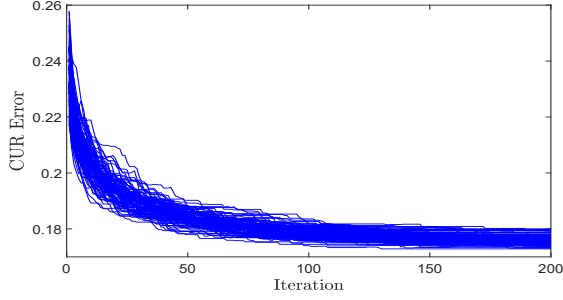


Fig. 3: The convergence behavior of the proposed algorithm w.r.t. the initial condition of selected subset. The initial cost function is corresponding to the initial set which is drawn randomly. The blue curves indicate the optimization path alongside iterations of the TWSP algorithm. Here, 100 different realizations are studied.

of joint sensor selection and channel assignment in a cognitive radio network. To this aim the exact setup in [21] is considered with 900 grid points and 32 frequency channels. The received power magnitudes are organized in a 900×32 matrix. The only difference here is that the uniform sampling pattern of sensing is replaced by the selection based on the CUR decomposition. Our proposed TWSP algorithm provides a fast and accurate solution for CUR decomposition. We select between 20 and 80 locations for spectrum sensing and all 32 channels. Each row of matrix U corresponds to a selected location and it has 32 entries. We are to assign F channels for each selected sensor. In other words, each location does not sense the whole spectrum and only F frequency channels are assigned to each sensor. The top- F entries in each row with the highest absolute value show the most important channels for the corresponding locations to be sensed.

Fig. 4 shows the cartography error of spectrum sensing for the conventional random selection as introduced in [21] and our proposed optimized joint sensors and channels selection. For each sampled locations $F = 8$ channels out of 32 channels are sensed. The sampled spectrum map is interpolated using the Thin plate splines method [18] for both sampling methods. In addition to visual superiority in reconstruction of the spectrum map, channel assignment based on TWSP provides a better quantitative error as observed in Fig. 4

4.3. Supervised Sampling

The proposed TWSP algorithm is an unsupervised data selection algorithm. In supervised settings, labels can infuse information to perform a more viable joint selection. A naive approach is to select representatives from each class independently. However,

Algorithm 1 Two way spectrum pursuit (TWSP)

Require: $X \in \mathbb{R}^{N \times M}$, K_1 and K_2 .

Output: S^c and S^r .

Initialization:

$S^c \leftarrow$ A random subset of $\{1, \dots, M\}$ with $|S^c| = K_1$

$S^r \leftarrow$ A random subset of $\{1, \dots, N\}$ with $|S^r| = K_2$

$\{S_k^c\}_{k=1}^{K_1} \leftarrow$ Partition S^c into K_1 subsets.

$\{S_k^r\}_{k=1}^{K_2} \leftarrow$ Partition S^r into K_2 subsets

$i = \text{rnd}(K_1)$ and $j = \text{rnd}(K_2)$

$C \leftarrow$ columns of X indexed by set S^c

$R \leftarrow$ rows of X indexed by set S^r

while a stopping criterion is not met

$$S_i^c = S^c \setminus S_i^c$$

$$C_i \leftarrow \text{remove column } i \text{ in matrix } C$$

$$W = C_i^\dagger X R^\dagger$$

$$E^c = X - C_i W R \quad (\text{Null space projection})$$

$$c \leftarrow \text{find the first left singular vector of } E^c \quad (3a)$$

$$S_i^c \leftarrow \text{the most correlated column of } E^c \text{ with } c \quad (4a)$$

$$\tilde{S}^c \leftarrow \bigcup_{i=1}^{K_1} S_i^c$$

$$C = X(:, \tilde{S}^c)$$

$$e^c = \min_U \|X - CUR\|_F$$

$$S_j^r = S^r \setminus S_j^r$$

$$R_j \leftarrow \text{remove row } j \text{ in matrix } R$$

$$Y = C^\dagger X R_j^\dagger$$

$$E^r = X - C Y R_j \quad (\text{Null space projection})$$

$$r \leftarrow \text{find the first right singular vector of } E^r \quad (3b)$$

$$S_j^r \leftarrow \text{the most correlated row of } E^r \text{ with } r \quad (4b)$$

$$\tilde{S}^r \leftarrow \bigcup_{j=1}^{K_2} S_j^r$$

$$R = X(\tilde{S}^r, :)$$

$$e^r = \min_U \|X - CUR\|_F$$

$$\text{IF } e^c < e^r$$

$$S^c \leftarrow \tilde{S}^c$$

$$i = \text{rnd}(K_1)$$

else

$$S^r \leftarrow \tilde{S}^r$$

$$j = \text{rnd}(K_2)$$

considering classes jointly is more effective for data reduction. Assume two data classes are given as $X_1 \in \mathbb{R}^{N \times M_1}$ and $X_2 \in \mathbb{R}^{N \times M_2}$. The goal is to select K_1 samples from class 1 and K_2 samples from class 2. To this aim, we obtain the cross correlation of two classes (as a kernel representation). Matrix

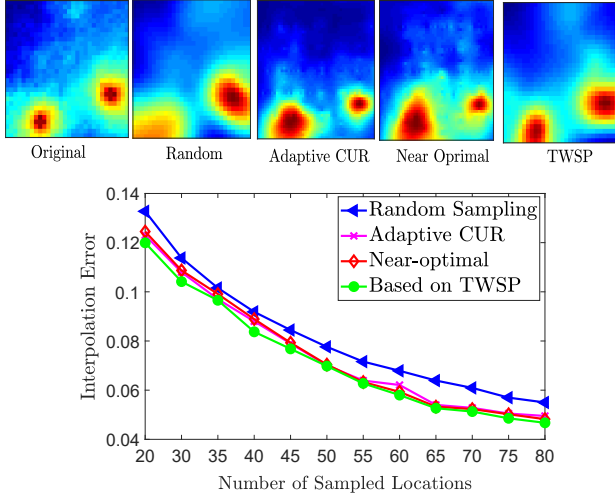


Fig. 4: (top) The original spectrum map and its comparison with the interpolated map using random sampling, adaptive CUR, Near Optimal, and our proposed method (TWSP). (bottom) The interpolation error versus the number of sensed locations is depicted.

$\mathbf{X} = \mathbf{X}_2^T \mathbf{X}_1$, which has M_1 columns and M_2 rows, is fed to TWSP to select K_1 columns and K_2 rows jointly.

As an initial experiment, supervised sampling is performed on Kaggle cats and dogs dataset. The features are obtained by a trained Resnet-18 deep learning model as explained in [9]. Three mutually exclusive data subsets for training, validation, and testing are partitioned randomly from 2000 images of each class. The classification accuracy of 97.5% is achieved from a fine-tuned Resnet-18 using the whole training set containing 1000 samples for each class. Afterwards, samples are selected by applying TWSP on the kernel feature matrix and the Resnet-18 is fine-tuned by using the sampled data. The model's accuracy is compared on the testing set with other sampling methods. Fig. 5 shows the performance of selection algorithms for different numbers of representatives per class. Using only two samples from each class, a classification accuracy of 82.3% can be achieved which is more than 15% improvement compared to random selection and more than 5% improvement compared to other competitors.

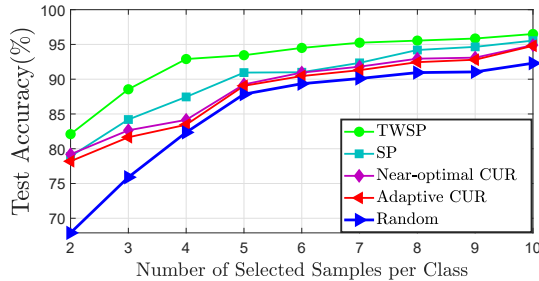


Fig. 5: The classification accuracy of a fine-tuned Resnet-18 network using a few selected data per each class.

We have conducted further experiments to study the effectiveness of the proposed algorithm in the multi-class image classification problem. For this study, we use the Resnet-34 deep learning model pre-trained on CIFAR10 and trained on a subset of ImageNet Dataset comprising of 10 classes. The classes used

for this experiment are Tench, Goldfish, Great white shark, Tiger shark, Hammerhead, Electric ray, Stingray, Cock, Hen, and Ostrich. The original training data consists of 1300 images of each class, and the idea is to use TWSP to select the data samples such that K samples of each class are used for training. After training Resnet-34 for 10 classes of CIFAR10, the feature vectors of all the training images are extracted such that a 1300×512 matrix is obtained for each class. Since TWSP is applicable for the two-class problem, we employ the one-versus-all approach. In other words, a separate cross correlation matrix is generated for each class such that $\mathbf{X}_1$ has the dimensions 512×1300 and $\mathbf{X}_2$ has the dimensions 512×11700 where $\mathbf{X}_1$ represents the feature vector of the class for which samples will be selected while $\mathbf{X}_2$ represents the feature vectors of the remaining 9 classes. Hence, the number of cross correlation matrices is equal to the number of the classes. Then, K_1 rows and K_2 columns are selected separately by applying TWSP on each cross correlation matrix.

This step of generating correlation matrices is different from binary classification where only one correlation matrix was formed. Pre-trained Resnet-34 on CIFAR10 has been refined again by using the sampled data from the ImageNet dataset. The model's accuracy is subsequently compared on the test set with other sampling methods in Fig. 6. As it can be seen, our proposed TWSP algorithm shows a superior performance compared to the state-of-the-art methods in sampling.

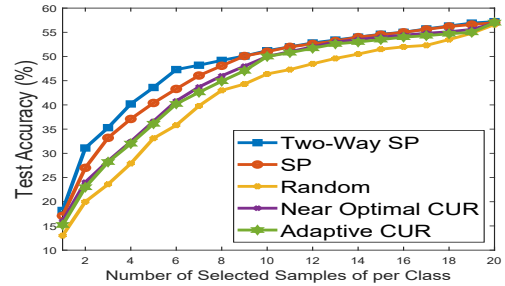


Fig. 6: Test accuracy for 10 classes of CIFAR10 dataset.

4.4. Informative Users/Contents Detection

Another problem gaining a lot of interest by streaming service providers is choosing a set of users to decisively reflect their feedbacks about different products. Therefore, it is crucial for such companies to find a subset of users and media products, reviews of which contain information about other users' unknown behavior. Each user has a limited scope of interest. For example, a user preferring romance and action genres is of a specific personality, able to represent a cluster of users accurately. Moreover, such reviews are more valuable for their areas of interest, not for all genres. Moreover, there exist movies well representing their genres.

As a result, there exists a demand for a reliable algorithm to simultaneously choose the most informative subset of users and movies. Such a subset is desirable for streaming companies to the extent that they are willing to give the users incentives to leave comprehensive reviews for certain products. In this regards, we have evaluated our algorithm on Netflix Prize dataset

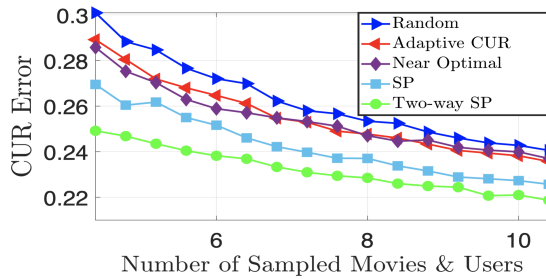


Fig. 7: Comparison of the normalized prediction error with state-of-the-art algorithms obtained by CUR decomposition for simultaneous movies and users subset selection from Netflix dataset.

containing 17,770 movies and 480,189 users. We have reduced the dataset to 990 movies and 4,727 users by considering only movies and users with most reviews. Then, we completed the dataset by Lin et al. method to have a ground truth [14]. We select a subset of rows and columns (users and media) from the completed dataset. Fig. 7 reveals that TWSP shows the best performance in terms of predicting scores for all users/movies based on a few selected users/movies.

5. CONCLUSION

Two-way spectrum pursuit is proposed as an accurate and efficient algorithm for solving CUR decomposition. TWSP can be employed for joint selection of columns and rows such that their outer products is able to reconstruct the whole matrix as accurately as possible. Some applications of the proposed algorithm are presented to show the efficacy of the proposed method. However, they are not limited to the mentioned applications. Moreover, the proposed algorithm can be extended to n -way spectrum pursuit for efficient tensor subset selection. The efficacy in joint selection on different datasets with TWSP is shown against other methods.

6. REFERENCES

- [1] Christos Boutsidis, Petros Drineas, and Malik Magdon-Ismail. Near-optimal column-based matrix reconstruction. *SIAM Journal on Computing*, 43(2):687–717, 2014.
- [2] Christos Boutsidis, Michael W Mahoney, and Petros Drineas. An improved approximation algorithm for the column subset selection problem. In *Proceedings of the twentieth annual ACM-SIAM symposium on Discrete algorithms*, pages 968–977. SIAM, 2009.
- [3] Christos Boutsidis and David Woodruff. Optimal CUR matrix decompositions. *SIAM Journal on Computing*, 2017.
- [4] Pierre Comon and Gene H Golub. Tracking a few extreme singular values and vectors in signal processing. *Proceedings of the IEEE*, 78(8):1327–1343, 1990.
- [5] Amit Deshpande and Luis Rademacher. Efficient volume sampling for row/column subset selection. In *2010 IEEE 51st annual symposium on foundations of computer science*, pages 329–338. IEEE, 2010.
- [6] Ehsan Elhamifar, Guillermo Sapiro, and S Shankar Sastry. Dissimilarity-based sparse subset selection. *IEEE transactions on pattern analysis and machine intelligence*, 2015.
- [7] Ehsan Elhamifar, Guillermo Sapiro, and Rene Vidal. See all by looking at a few: Sparse modeling for finding representative objects. In *2012 IEEE conference on computer vision and pattern recognition*. IEEE, 2012.
- [8] Keaton Hamm and Longxiu Huang. Perspectives on cur decompositions. *Applied and Computational Harmonic Analysis*, 48(3):1088–1099, 2020.
- [9] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.
- [10] Mohsen Joneidi, Saeed Vahidian, Ashkan Esmaeili, and Siavash Khodadadeh. Optimality of spectrum pursuit for column subset selection problem: Theoretical guarantees and applications in deep learning. 2020.
- [11] Mohsen Joneidi, Saeed Vahidian, Ashkan Esmaeili, Weijia Wang, Nazanin Rahnavard, Bill Lin, and Mubarak Shah. Select to better learn: Fast and accurate deep learning using data selection from nonlinear manifolds. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 7819–7829, 2020.
- [12] Mohsen Joneidi, Alireza Zaeemzadeh, Behzad Shahrabi, Guojun Qi, and Nazanin Rahnavard. E-optimal sensor selection for compressive sensing-based purposes. *IEEE Transactions on Big Data*, 2018.
- [13] Chengtao Li, Stefanie Jegelka, and Suvrit Sra. Polynomial time algorithms for dual volume sampling. In *Advances in Neural Information Processing Systems*, 2017.
- [14] Zhouchen Lin, Minming Chen, and Yi Ma. The augmented lagrange multiplier method for exact recovery of corrupted low-rank matrices. *arXiv preprint arXiv:1009.5055*, 2010.
- [15] Michael W Mahoney and Petros Drineas. Cur matrix decompositions for improved data analysis. *Proceedings of the National Academy of Sciences*, 106(3):697–702, 2009.
- [16] Saurabh Paul, Malik Magdon-Ismail, and Petros Drineas. Column selection via adaptive sampling. In *Advances in neural information processing systems*, 2015.
- [17] Yaroslav Shitov. Column subset selection is np-complete. *Linear Algebra and its Applications*, 2020.
- [18] Suzan Üreten, Abbas Yongaçoğlu, and Emil Petriu. A comparison of interference cartography generation techniques in cognitive radio networks. In *2012 IEEE International Conference on Communications (ICC)*. IEEE, 2012.
- [19] Shusen Wang and Zhihua Zhang. Improving cur matrix decomposition and the nyström approximation via adaptive sampling. *The Journal of Machine Learning Research*, 14(1):2729–2769, 2013.
- [20] Alireza Zaeemzadeh, Mohsen Joneidi, Nazanin Rahnavard, and Mubarak Shah. Iterative projection and matching: Finding structure-preserving representatives and its application to computer vision. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2019.
- [21] Guoyong Zhang, Xiao Fu, Jun Wang, Xi-Le Zhao, and Mingyi Hong. Spectrum cartography via coupled block-term tensor decomposition. *IEEE Transactions on Signal Processing*, 2020.