# Extreme Point Pursuit—Part II: Further Error Bound Analysis and Applications

Junbin Liu, Ya Liu, Wing-Kin Ma, Mingjie Shao and Anthony Man-Cho So

***Abstract*— In the first part of this study, a convex-constrained penalized formulation was studied for a class of constant modulus (CM) problems. In particular, the error bound techniques were shown to play a vital role in providing exact penalization results. In this second part of the study, we continue our error bound analysis for the cases of partial permutation matrices, size-constrained assignment matrices and non-negative semi-orthogonal matrices. We develop new error bounds and penalized formulations for these three cases, and the new formulations possess good structures for building computationally efficient algorithms. Moreover, we provide numerical results to demonstrate our framework in a variety of applications such as the densest $k$-subgraph problem, graph matching, size-constrained clustering, non-negative orthogonal matrix factorization and sparse fair principal component analysis.**

***Index Terms*— constant modulus optimization, non-convex optimization, error bound, densest subgraph problem, PCA, graph matching, clustering, ONMF**

## I. Introduction

In Part I of this paper [1], we considered a convex-constrained minimization framework for a class of constant modulus (CM) problems. Named extreme point pursuit (EXPP), this framework gives simple well-structured reformulations of the CM problems. This allows us to apply basic methods, such as the projected gradient (or subgradient) method, to build algorithms for CM problems; some underlying assumptions with the objective function are required, but they are considered reasonable in a wide variety of applications. As a requirement, EXPP chooses the constraint set as the convex hull of the CM set. When the projected gradient method is used, the computational efficiency will depend on whether the projection onto the convex hull of the CM set is easy to compute. We examined a number of CM examples that have such a benign property. But we also encountered some CM sets, namely, the partial permutation matrix (PPM) set and the size-constrained assignment matrix (SAM) set, whose convex hull projections may be inefficient to compute. Another difficult set is the non-negative semi-orthogonal matrix (NSOM) set, whose convex hull is not even known.

As Part II of this paper, we continue our analysis with the PPM set, the SAM set, and the NSOM set. Our focus is on the error bound principle for exact penalization, and we want to see if we can relax the EXPP constraint set for the above three cases. We will consider some efficient-to-project constraint sets for the three cases, and we will show that, by suitably modifying the penalty function, we can once again achieve exact penalization. These case-specific results are more powerful versions of their counterparts in Part I of this paper. In the graph matching application, for example, we will numerically illustrate that the new method has much better computational efficiency than the state of the arts. Also, our results for NSOMs draw connections to some recently emerged results [2]–[4], as we shall see.

In addition to error bound analysis, we will numerically demonstrate EXPP in a variety of applications—such as the densest $k$-subgraph problem, graph matching, size-constrained clustering, ONMF, and PCA with sparsity and/or fairness. We will see that EXPP is a working method across different applications, offering reasonable performance and computational efficiency. The organization is as follows. Section II recapitulates some key concepts in Part I of this study. Section III provides further error bound analysis and devises new EXPP formulations. Section IV provides numerical results for different applications. Section V concludes this paper.

## II. A Recap of Part I

We give a summary of Part I of this study, with a focus on the error bound principle. Let $\mathcal{D} \subseteq \mathbb{R}^n$ be a set which will be used to denote the domain of a problem. Let $\mathcal{X} \subseteq \mathcal{D}$ be a set which will often be chosen as a convex compact set. We consider a class of CM problems in the form of

$$\min_{\boldsymbol{x}\in\mathcal{V}} f(\boldsymbol{x}), \tag{1}$$

where $f : \mathcal{D} \to \mathbb{R}$ is $K$-Lipschitz continuous on $\mathcal{X}$; $\mathcal{V} \subseteq \mathcal{X}$ is a non-empty closed CM set with modulus $\sqrt{C}$. Consider a general penalized formulation of (1):

$$\min_{\boldsymbol{x}\in\mathcal{X}} F_\lambda(\boldsymbol{x}) := f(\boldsymbol{x}) + \lambda\, h(\boldsymbol{x}), \tag{2}$$

where $h : \mathbb{R}^n \to \mathbb{R}$ is a penalty function; $\lambda > 0$ is a given scalar. We want to find an $\mathcal{X}$ and an $h$ such that the penalized formulation (2) is "easy" to build an algorithm and has exact penalization guarantees. By exact penalization, we mean that the solution set of (2) is equal to that of (1). According to the error bound principle, formulation (2) is an exact penalization formulation of (1) if $h$ is effectively an error bound function of $\mathcal{V}$ relative to $\mathcal{X}$, i.e.,

$$\begin{aligned} \operatorname{dist}(\boldsymbol{x}, \mathcal{V}) \le \nu\, h(\boldsymbol{x}), &\quad \forall \boldsymbol{x} \in \mathcal{X}, \\ 0 = h(\boldsymbol{x}), &\quad \forall \boldsymbol{x} \in \mathcal{V}, \end{aligned}$$

The work of J. Liu, Y. Liu, W.-K. Ma and M. Shao was supported by the General Research Fund (GRF) of Hong Kong Research Grant Council (RGC) under Project ID CUHK 14208819. The work of A. M.-C. So was supported by the GRF of Hong Kong RGC under Project ID CUHK 14205421.

Junbin Liu and Ya Liu have equal contributions.

for some constant $\nu > 0$. More specifically, the exact penalization result holds when $\lambda$ is sufficiently large such that $\lambda > K\nu$.

We studied a convex-constrained minimization formulation of problem (1), called EXPP, which is an instance of the formulation (2) with

$$\mathcal{X} = \mathrm{conv}(\mathcal{V}), \quad h(\boldsymbol{x}) = C - \|\boldsymbol{x}\|_2^2. \tag{3}$$

We showed that the choice in (3) leads to an error bound, and consequently exact penalization, for many practical CM sets of interest. A merit with EXPP is that its penalty function $h$ in (3) is simple. If the constraint set $\mathcal{X}$ in (3) is friendly to handle, then we may build algorithms for EXPP without much difficulty. There are various possibilities for one to build algorithms for EXPP, and here we focus our attention on the projected gradient method or related methods. Assuming that $f$ is differentiable, the projected gradient method for problem (2) is given by

$$\boldsymbol{x}^{l+1} = \Pi_{\mathcal{X}}(\boldsymbol{x}^l - \eta_l \nabla F_\lambda(\boldsymbol{x}^l)), \quad l = 0, 1, \ldots, \tag{4}$$

where $\eta_l > 0$ is the step size. The computational efficiency of the projected gradient method depends on whether the projection $\Pi_{\mathcal{X}}$ is efficient to compute. We examined a collection of CM sets that have efficient-to-compute projections. But we also encounter CM sets whose projections can be expensive, or has no known way, to compute.

1. *partial permutation matrix (PPM) set:*

$$\mathcal{U}^{n,r} = \{\boldsymbol{X} \in \{0,1\}^{n\times r} \mid \boldsymbol{X}^\top \mathbf{1} = \mathbf{1}, \boldsymbol{X}\mathbf{1} \le \mathbf{1}\},$$

where $n \ge r$. The convex hull of $\mathcal{U}^{n,r}$ is

$$\mathrm{conv}(\mathcal{U}^{n,r}) = \{\boldsymbol{X} \in [0,1]^{n\times r} \mid \boldsymbol{X}^\top \mathbf{1} = \mathbf{1}, \boldsymbol{X}\mathbf{1} \le \mathbf{1}\}. \tag{5}$$

There is no known easy way to compute the projection onto $\mathrm{conv}(\mathcal{U}^{n,r})$. Solving the projection using a numerical solver can in practice be expensive for large $n$ and $r$.

2. *size-constrained assignment matrix (SAM) set:*

$$\mathcal{U}_{\boldsymbol{\kappa}}^{n,r} = \{\boldsymbol{X} \in \{0,1\}^{n\times r} \mid \boldsymbol{X}^\top \mathbf{1} = \boldsymbol{\kappa}, \boldsymbol{X}\mathbf{1} \le \mathbf{1}\}, \tag{6}$$

where $n \ge r$, $\boldsymbol{\kappa} \in \{1, \ldots, n\}^r$, $\sum_{j=1}^r \kappa_j \le n$. The convex hull of $\mathcal{U}_{\boldsymbol{\kappa}}^{n,r}$ is

$$\mathrm{conv}(\mathcal{U}_{\boldsymbol{\kappa}}^{n,r}) = \{\boldsymbol{X} \in [0,1]^{n\times r} \mid \boldsymbol{X}^\top \mathbf{1} = \boldsymbol{\kappa}, \boldsymbol{X}\mathbf{1} \le \mathbf{1}\}. \tag{7}$$

The projection onto $\mathrm{conv}(\mathcal{U}_{\boldsymbol{\kappa}}^{n,r})$ has the same issue as that in the PPM case.

3. *non-negative semi-orthogonal matrix (NSOM) set:*

$$\mathcal{S}_+^{n,r} = \mathcal{S}^{n,r} \cap \mathbb{R}_+^{n\times r},$$

where $n \ge r$, and $\mathcal{S}^{n,r} = \{\boldsymbol{X} \in \mathbb{R}^{n\times r} \mid \boldsymbol{X}^\top \boldsymbol{X} = \boldsymbol{I}\}$ is the semi-orthogonal matrix set. There are no known expressions with the convex hull of $\mathcal{S}_+^{n,r}$ and the associated projection.

In this second part of the study, we continue to study these CM sets.

## III. Further Error Bound Analysis

In this section we perform error bound analysis for the above three sets. The first and second subsections consider the PPM case. We will provide an application example to motivate the study, and then we will derive a new error bound and EXPP to overcome the computational issue described in the last section. Following the same genre, the SAM case is tackled in the third and fourth subsections, and the NSOM case studied in the fifth, sixth, and seventh subsections.

### A. Example: Graph Matching

Consider the following problem for a given $\boldsymbol{A}, \boldsymbol{B} \in \mathbb{R}^{n\times n}$:

$$\min_{\boldsymbol{X}\in\mathcal{U}^{n,n}} \|\boldsymbol{A} - \boldsymbol{X}^\top \boldsymbol{B}\boldsymbol{X}\|_{\mathrm{F}}^2.$$

This problem is called the graph matching (GM) problem in the context of computer vision. The goal is to match the nodes of two equal-size graphs by using the graph edge information. The GM problem does so by finding a set of one-to-one node associations, represented by $\boldsymbol{X}$, such that the associated Euclidean error between the two graph's adjacency matrices, represented by $\boldsymbol{A}$ and $\boldsymbol{B}$, is minimized. Since a feasible $\boldsymbol{X}$ is orthogonal, the GM problem can be rewritten as

$$\min_{\boldsymbol{X}\in\mathcal{U}^{n,n}} f(\boldsymbol{X}) := \|\boldsymbol{X}\boldsymbol{A} - \boldsymbol{B}\boldsymbol{X}\|_{\mathrm{F}}^2. \tag{8}$$

Following the review in the last section, the EXPP formulation of (8) is

$$\min_{\boldsymbol{X}\in\mathrm{conv}(\mathcal{U}^{n,n})} F_\lambda(\boldsymbol{X}) = f(\boldsymbol{X}) - \lambda\|\boldsymbol{X}\|_{\mathrm{F}}^2. \tag{9}$$

It can be expensive to apply the projected gradient method (4) to the EXPP–GM problem (9). The projection operation $\Pi_{\mathrm{conv}(\mathcal{U}^{n,n})}$ in the projected gradient method requires us to solve an optimization problem, namely, minimization of $\|\boldsymbol{Z} - \boldsymbol{X}\|_{\mathrm{F}}^2$ over $\boldsymbol{X} \in \mathrm{conv}(\mathcal{U}^{n,n})$, where $\boldsymbol{Z}$ is given. As mentioned, there is no easy-to-compute solution for this problem. The problem is nevertheless convex, and there is a specialized numerical solver for this problem based on gradient ascent of the dual problem [5]. Still, one would anticipate that solving an optimization problem at each step of the projected gradient method (4) would be computationally burdensome particularly when the problem size $n$ is large.

At this point it is worthwhile to review an important prior study in GM [6]. The authors in that study developed a similar formulation as (8); they essentially put forth the same fundamental idea as EXPP under the concave minimization principle, though not under the error bound principle. They built an algorithm that exploits the problem structure for efficient computations. Specifically they considered the Frank-Wolfe method. The Frank-Wolfe method for (9) is given by

$$\boldsymbol{X}^{l+1} = \boldsymbol{X}^l + \alpha_l(\mathsf{LO}_{\mathrm{conv}(\mathcal{U}^{n,n})}(\nabla F_\lambda(\boldsymbol{X}^l)) - \boldsymbol{X}^l),$$

where $\alpha_l \in [0,1]$ is a step size; $\mathsf{LO}_{\mathcal{X}}(\boldsymbol{g}) \in \arg\min_{\boldsymbol{x}\in\mathcal{X}} \boldsymbol{g}^\top \boldsymbol{x}$ is the linear optimization (LO) oracle for $\mathcal{X}$. To efficiently compute each Frank-Wolfe iteration, we need an efficient way to compute $\mathsf{LO}_{\mathrm{conv}(\mathcal{U}^{n,n})}$. The authors of [6] did so by using the Hungarian algorithm [7], which is a specialized algorithm for solving linear optimization over the set of doubly stochastic

matrices and can efficiently compute $\mathsf{LO}_{\mathrm{conv}(\mathcal{U}^{n,n})}$ with a complexity of $\mathcal{O}(n^3)$.

We have a different proposition. To put into context, recall that $\Delta^n = \{\boldsymbol{x} \in \mathbb{R}^n_+ \mid \boldsymbol{x}^\top \mathbf{1} = 1\}$ denotes the unit simplex. Define

$$\begin{aligned}\tilde{\Delta}^{n\times r} &= \{\boldsymbol{X} \in [0,1]^{n\times r} \mid \boldsymbol{X}^\top \mathbf{1} = \mathbf{1}\} \\ &= \{\boldsymbol{X} \in \mathbb{R}^{n\times r} \mid \boldsymbol{x}_j \in \Delta^n,\ \forall j\}.\end{aligned}$$

This set is a relaxation of $\mathrm{conv}(\mathcal{U}^{n,r})$ in (5) by taking out the row constraint $\boldsymbol{X}\mathbf{1} \le \mathbf{1}$. Our idea is to derive an error bound function $h$ of $\mathcal{U}^{n,r}$ relative to $\tilde{\Delta}^{n\times r}$. If we can do so, then we will have an exact penalization formulation (2) for the GM problem, and more generally, CM problems for the PPM case. Consequently there is a possibility for us to apply the projected gradient method (4) in a computationally efficient fashion. To be specific, we can compute the projection $\Pi_{\tilde{\Delta}^{n\times r}}(\boldsymbol{Z})$ of a given matrix $\boldsymbol{Z} \in \mathbb{R}^{n\times r}$ with a complexity of $\mathcal{O}(nr\log(n))$: the projection $\Pi_{\tilde{\Delta}^{n\times r}}(\boldsymbol{Z})$ corresponds to the unit simplex projections $\Pi_{\Delta^n}(\boldsymbol{z}_j)$'s for $j = 1,\ldots,r$, and the unit simplex projection can be computed with a complexity of $\mathcal{O}(n\log(n))$ [8]. In the GM application, the complexity of the projection $\Pi_{\tilde{\Delta}^{n\times n}}$ is $\mathcal{O}(n^2\log(n))$—which seems more attractive than its Frank-Wolfe counterpart, $\mathcal{O}(n^3)$.

To illustrate the efficiencies of the above described projection and LO operations, we ran a numerical experiment. We tested the runtimes of (i) the projection onto $\mathrm{conv}(\mathcal{U}^{n,n})$, implemented either by CVX or by the specialized dual gradient method [5]; (ii) the LO oracle for $\mathrm{conv}(\mathcal{U}^{n,n})$, implemented by the Hungarian algorithm; and (iii) the projection onto $\tilde{\Delta}^{n\times n}$, done by column-wise unit-simplex projection. The results are shown in Fig. 1. It is seen that the projection onto $\tilde{\Delta}^{n\times n}$ is much faster than the other operations.

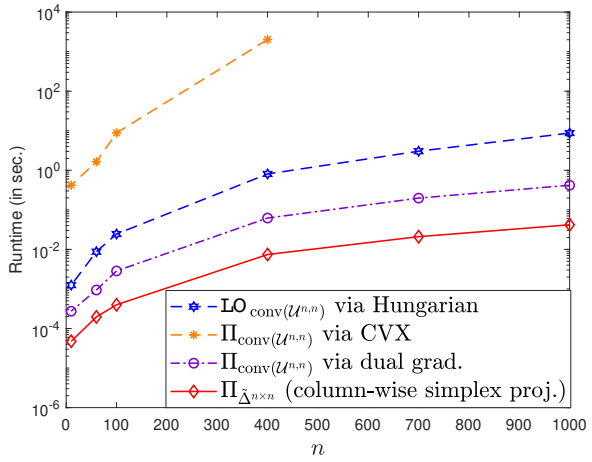



Fig. 1: Runtime comparison with the projections and LO oracle for the full PPM case ($n = r$). We used 200 randomly generated trials to evaluate the average runtimes, except for $\Pi_{\mathrm{conv}(\mathcal{U}^{n,n})}$ via CVX which we tested only 20 trials due to the long runtime.

### *B. A New Error Bound and a New EXPP for PPMs*

Motivated by the GM problem, we perform error bound analysis for the PPM set $\mathcal{U}^{n,r}$ relative to the column-wise unit simplex $\tilde{\Delta}^{n\times r}$. Our result is as follows.

**Theorem 1 (error bound for partial permutation matrices)** *For any $\boldsymbol{X} \in \tilde{\Delta}^{n\times r}$ we have the error bound*

$$\begin{aligned}\mathrm{dist}(\boldsymbol{X},\mathcal{U}^{n,r}) &\le \nu\|\boldsymbol{X}^\top\boldsymbol{X} - \boldsymbol{I}\|_{\ell_1} &\text{(10a)}\\ &= \nu(r + \|\boldsymbol{X}\mathbf{1}\|_2^2 - 2\|\boldsymbol{X}\|_{\mathrm{F}}^2), &\text{(10b)}\end{aligned}$$

*where $\nu = 11\sqrt{r}$.*

The proof of Theorem 1 uses the same proof approach as that for the error bound for $\mathcal{U}^{n,r}_{\boldsymbol{\kappa}}$ in Part I, Section IV.F, of this paper. But the latter is considered easier to show because of the presence of the row constraint $\boldsymbol{X}\mathbf{1} \le \mathbf{1}$. To derive (10) we need to go further, analyzing the singular values of $\boldsymbol{X}$. In view of its complexity, the proof of Theorem 1 is relegated to Appendix B.

Theorem 1 gives rise to a new EXPP formulation for $\mathcal{U}^{n,r}$. By applying the error bound (10b) to the penalized formulation (2), we have an exact penalization formulation for the PPM case:

$$\min_{\boldsymbol{X}\in\tilde{\Delta}^{n\times r}} F_\lambda(\boldsymbol{X}) = f(\boldsymbol{X}) + \lambda(\|\boldsymbol{X}\mathbf{1}\|_2^2 - 2\|\boldsymbol{X}\|_{\mathrm{F}}^2). \quad (11)$$

This new EXPP formulation resembles our previous EXPP, with addition of a friendly penalty term $\|\boldsymbol{X}\mathbf{1}\|_2^2$. Assuming a differentiable $f$, the new EXPP problem (11) can be efficiently handled by the projected gradient method (4). In particular, without counting the complexity of computing the gradient $\nabla F_\lambda$, the per-iteration complexity of the projected gradient method is $\mathcal{O}(nr\log(n))$; see the discussion in the last subsection.

We should provide insight into the error bound in Theorem 1. The PPM set can be characterized as

$$\mathcal{U}^{n,r} = \tilde{\Delta}^{n\times r} \cap \mathcal{S}^{n,r}.$$

The error bound (10a) reveals that, by using $\tilde{\Delta}^{n\times r}$ as the constraint set and $\|\boldsymbol{X}^\top\boldsymbol{X} - \boldsymbol{I}\|_{\ell_1}$ as the penalty, we can achieve exact penalization results. Particularly, $\|\boldsymbol{X}^\top\boldsymbol{X} - \boldsymbol{I}\|_{\ell_1}$ appears as a penalty for promoting $\boldsymbol{X}$ to be semi-orthogonal. Moreover, the error bound (10b) is an equivalent form of $\|\boldsymbol{X}^\top\boldsymbol{X} - \boldsymbol{I}\|_{\ell_1}$. Consider the following result.

**Lemma 1** *Let $\boldsymbol{d} \in \mathbb{R}^r_{++}$ be given. Let $\boldsymbol{D} = \mathrm{Diag}(\boldsymbol{d})$. For any $\boldsymbol{X} \in \mathbb{R}^{n\times r}_+$ with $\|\boldsymbol{x}_j\|_2^2 \le d_j$ for all $j$, it holds that*

$$\|\boldsymbol{X}^\top\boldsymbol{X} - \boldsymbol{D}\|_{\ell_1} = \mathbf{1}^\top\boldsymbol{d} + \|\boldsymbol{X}\mathbf{1}\|_2^2 - 2\|\boldsymbol{X}\|_{\mathrm{F}}^2.$$

*Proof of Lemma 1:* For any $\boldsymbol{X} \in \mathbb{R}^{n\times r}_+$ with $\|\boldsymbol{x}_j\|_2^2 \le d_j$ for all $j$,

$$\begin{aligned}\|\boldsymbol{X}^\top\boldsymbol{X} - \boldsymbol{D}\|_{\ell_1} &= \sum_{j=1}^r (d_j - \|\boldsymbol{x}_j\|_2^2) + \sum_{j=1}^r\sum_{\substack{k=1\\ j\ne k}}^r \boldsymbol{x}_j^\top\boldsymbol{x}_k \\ &= \sum_{j=1}^r (d_j - 2\|\boldsymbol{x}_j\|_2^2) + \left(\sum_{j=1}^r \boldsymbol{x}_j\right)^\top \left(\sum_{j=1}^r \boldsymbol{x}_j\right) \\ &= \mathbf{1}^\top\boldsymbol{d} - 2\|\boldsymbol{X}\|_{\mathrm{F}}^2 + \|\boldsymbol{X}\mathbf{1}\|_2^2.\end{aligned}$$

The proof is complete. ■

Applying Lemma 1 to (10a) gives the error bound (10b); note that any $\boldsymbol{x} \in \Delta^n$ has $\|\boldsymbol{x}\|_2 \leq \|\boldsymbol{x}\|_1 = \boldsymbol{1}^\top \boldsymbol{x} = 1$. The error bound (10b) also reveals interesting insight. For any $\boldsymbol{X} \in \mathbb{R}_+^{n\times r}$ with $\|\boldsymbol{x}_j\|_2 \leq 1$ for all $j$, we can rewrite (10b) as

$$r + \|\boldsymbol{X}\boldsymbol{1}\|_2^2 - 2\|\boldsymbol{X}\|_{\rm F}^2 = \sum_{j=1}^r c_2(\boldsymbol{x}_j) + \sum_{i=1}^n \rho_2(\bar{\boldsymbol{x}}_i), \quad (12)$$

where

$$c_2(\boldsymbol{x}) = 1 - \|\boldsymbol{x}\|_2^2, \quad \rho_2(\boldsymbol{x}) = \|\boldsymbol{x}\|_1^2 - \|\boldsymbol{x}\|_2^2$$

appear as penalties for columns and rows, respectively. According to Part I of this paper, $c_2$ is effectively an error bound function for the unit vector set $\mathcal{U}^n = \{\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n\}$ relative to the unit simplex $\Delta^n$; this means that $c_2$ promotes every column $\boldsymbol{x}_j$ to lie in $\mathcal{U}^n$. As for $\rho_2$, we note the basic norm result that $\|\boldsymbol{x}\|_1 \geq \|\boldsymbol{x}\|_2$ for any $\boldsymbol{x}$, and that $\|\boldsymbol{x}\|_1 = \|\boldsymbol{x}\|_2$ if and only if $\boldsymbol{x}$ is a scaled unit vector, i.e., $\boldsymbol{x} = \alpha \boldsymbol{e}_i$ for some $\alpha$ and $i$. This means that $\rho_2$ promotes every row $\bar{\boldsymbol{x}}_i$ to be a scaled unit vector. Putting the column and row penalties together, we have the interpretation that (12) promotes $\boldsymbol{X}$ to be a PPM.

### C. Example: Size-Constrained Clustering

We turn our interest to the SAM set $\mathcal{U}_{\boldsymbol{\kappa}}^{n,r}$ in (6). Let $\boldsymbol{\kappa} \in \{1, \ldots, n\}^r$, $\sum_{j=1}^r \kappa_j = n$, be given. Given a matrix $\boldsymbol{Y} \in \mathbb{R}^{m\times n}$, we consider the following problem

$$\min_{\boldsymbol{A}\in\mathbb{R}^{m\times r}, \boldsymbol{X}\in\mathcal{U}_{\boldsymbol{\kappa}}^{n,r}} \|\boldsymbol{Y} - \boldsymbol{A}\boldsymbol{X}^\top\|_{\rm F}^2. \quad (13)$$

This problem appears in size-constrained clustering and was used, for instance, in paper-to-session assignment [9]. In size-constrained clustering, we want to cluster a given set of data points $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_n$ into $r$ clusters, and the constraint is that each cluster has size $\kappa_j$. In the size-constrained clustering problem (13), the $j$th column $\boldsymbol{a}_j$ of $\boldsymbol{A}$ describes the cluster center of cluster $j$. From (6) it is seen that $\boldsymbol{X}$ takes the form

$$\boldsymbol{X}^\top = \left[\boldsymbol{e}_{l_1}, \ldots, \boldsymbol{e}_{l_n}\right],$$

for some $l_i \in \{1, \ldots, r\}$. In particular, $\boldsymbol{e}_j$ is constrained to appear in the rows of $\boldsymbol{X}$ for $\kappa_j$ times.

A natural way to handle the size-constrained clustering problem (13) is to apply alternating minimization; see, e.g., [9]. In this study, we use the following reformulation of the size-constrained clustering problem

$$\min_{\boldsymbol{X}\in\mathcal{U}_{\boldsymbol{\kappa}}^{n,r}} f(\boldsymbol{X}) := -{\rm tr}(\boldsymbol{D}^{-1}\boldsymbol{X}^\top \boldsymbol{R}\boldsymbol{X}), \quad (14)$$

where $\boldsymbol{D} = {\rm Diag}(\boldsymbol{\kappa})$; $\boldsymbol{R} = \boldsymbol{Y}^\top\boldsymbol{Y}$. Problem (14) is obtained by substituting the solution to $\boldsymbol{A}$ given an $\boldsymbol{X} \in \mathcal{U}_{\boldsymbol{\kappa}}^{n,r}$, specifically, $\boldsymbol{A} = \boldsymbol{Y}(\boldsymbol{X}^\top)^\dagger = \boldsymbol{Y}\boldsymbol{X}\boldsymbol{D}^{-1}$, to problem (13). Here, $\dagger$ denotes the pseudo inverse. We encounter the same difficulty as the PPM case in Section III-A: The EXPP formulation can be applied to (14), but the projected gradient method for EXPP can be expensive to implement due to the computational cost of numerically solving $\Pi_{{\rm conv}(\mathcal{U}_{\boldsymbol{\kappa}}^{n,r})}$. We want to replace the EXPP constraint set ${\rm conv}(\mathcal{U}_{\boldsymbol{\kappa}}^{n,r})$ with

$$\begin{aligned}\tilde{\mathcal{U}}_{\boldsymbol{\kappa}}^{n,r} &= \{\boldsymbol{X} \in [0,1]^{n\times r} \mid \boldsymbol{X}^\top\boldsymbol{1} = \boldsymbol{\kappa}\} \\ &= \{\boldsymbol{X} \in \mathbb{R}^{n\times r} \mid \boldsymbol{x}_j \in {\rm conv}(\mathcal{U}_{\kappa_j}^n)\ \forall j\}.\end{aligned}$$

Recall that $\mathcal{U}_\kappa^n = \{\boldsymbol{x} \in \{0,1\}^n \mid \boldsymbol{1}^\top\boldsymbol{x} = \kappa\}$ and ${\rm conv}(\mathcal{U}_\kappa^n) = \{\boldsymbol{x} \in [0,1]^n \mid \boldsymbol{1}^\top\boldsymbol{x} = \kappa\}$, where $\kappa \in \{1, \ldots, n\}$; and that $\Pi_{{\rm conv}(\mathcal{U}_\kappa^n)}$ can be efficiently computed by a bisection algorithm with a complexity of $\mathcal{O}(n\log(n))$, given a solution precision [10, Algorithm 1].

### D. A New Error Bound and a New EXPP for SAMs

We have the following result.

**Theorem 2 (error bound for size-constrained assignment matrices)** *For any $\boldsymbol{X} \in \tilde{\mathcal{U}}_{\boldsymbol{\kappa}}^{n,r}$, we have the error bound*

$$\begin{aligned}{\rm dist}(\boldsymbol{X}, \mathcal{U}_{\boldsymbol{\kappa}}^{n,r}) &\leq \nu\|\boldsymbol{X}^\top\boldsymbol{X} - {\rm Diag}(\boldsymbol{\kappa})\|_{\ell_1} &(15a)\\ &= \nu(\boldsymbol{1}^\top\boldsymbol{\kappa} + \|\boldsymbol{X}\boldsymbol{1}\|_2^2 - 2\|\boldsymbol{X}\|_{\rm F}^2), &(15b)\end{aligned}$$

*where $\nu = 3\sum_{i=1}^r (1+2\kappa_i)(1+\sqrt{\kappa_i})\sqrt{\boldsymbol{1}^\top\boldsymbol{\kappa}}$. Note that $\nu \leq 18\kappa_{\max}^2 r^{3/2}$, where $\kappa_{\max} = \max\{\kappa_1, \ldots, \kappa_r\}$.*

The idea behind the proof of Theorem 2 is the same as that of the PPM case in Theorem 1. The actual proof of Theorem 2 is however more tedious. The reader can find the proof in Appendix C. The error bound (15) shares the same insight as its PPM counterpart (see Section III-B), and we shall not repeat.

Applying the error bound (15b) to the penalized formulation (2) gives rise to the following new EXPP formulation for the SAM case:

$$\min_{\boldsymbol{X}\in\tilde{\mathcal{U}}_{\boldsymbol{\kappa}}^{n,r}} F_\lambda(\boldsymbol{X}) = f(\boldsymbol{X}) + \lambda(\|\boldsymbol{X}\boldsymbol{1}\|_2^2 - 2\|\boldsymbol{X}\|_{\rm F}^2). \quad (16)$$

This new EXPP is identical to the new EXPP for the partial permutation matrix case in (11). Assuming a differentiable $f$, we can use the projected gradient method (4) to efficiently handle problem (16). Specifically, the projection $\Pi_{\tilde{\mathcal{U}}_{\boldsymbol{\kappa}}^{n,r}}$, which contributes to the bulk of complexity with the projected gradient method, can be done with a complexity of $\mathcal{O}(nr\log(n))$; the idea is the same as that of the PPM case (see the last paragraph of Section III-A).

### E. Example: Orthogonal Non-Negative Matrix Factorization

As an application example for the NSOM set $\mathcal{S}_+^{n,r}$, consider the orthogonal non-negative matrix factorization (ONMF) problem

$$\min_{\boldsymbol{A}\in\mathbb{R}_+^{m\times r}, \boldsymbol{X}\in\mathcal{S}_+^{n,r}} \|\boldsymbol{Y} - \boldsymbol{A}\boldsymbol{X}^\top\|_{\rm F}^2, \quad (17)$$

where the given matrix $\boldsymbol{Y}$ is non-negative. The ONMF problem is, in essence, a non-negatively scaled clustering problem. It is known by researchers that any point $\boldsymbol{X}$ in $\mathcal{S}_+^{n,r}$ can be characterized as

$$\boldsymbol{X}^\top = \left[\alpha_1\boldsymbol{e}_{l_1}, \ldots, \alpha_n\boldsymbol{e}_{l_n}\right], \quad (18)$$

where $\alpha_i \geq 0$ for all $i$, $l_i \in \{1, \ldots, r\}$, $\sum_{i:l_i=j}\alpha_i^2 = 1$ for all $j$; see, e.g., [2]–[4], [11]. From this characterization, we

see that, fixing an assignment $l_1, \ldots, l_n$, each $\boldsymbol{a}_j$ is a cluster center, obtained by minimizing $\sum_{i:l_i=j} \|\boldsymbol{y}_i - \alpha_i \boldsymbol{a}_j\|_2^2$ over the non-negative $\boldsymbol{a}_j$ and the non-negative scale-compensating scalars $\alpha_i$'s. We consider the following reformulation of the ONMF problem

$$\min_{\boldsymbol{X} \in \mathcal{S}_+^{n,r}} f(\boldsymbol{X}) := -\mathrm{tr}(\boldsymbol{X}^\top \boldsymbol{R} \boldsymbol{X}), \tag{19}$$

where $\boldsymbol{R} = \boldsymbol{Y}^\top \boldsymbol{Y}$. Problem (19) is obtained by putting the solution to $\boldsymbol{A}$ given an $\boldsymbol{X} \in \mathcal{S}_+^{n,r}$ and a $\boldsymbol{Y} \in \mathbb{R}_+^{m \times n}$, i.e., $\boldsymbol{A} = \boldsymbol{Y}(\boldsymbol{X}^\top)^\dagger = \boldsymbol{Y}\boldsymbol{X}$, to (17). As discussed in Part I of this paper, there is no known expression for the convex hull of $\mathcal{S}_+^{n,r}$.

We turn to the error bound principle, seeing if we can obtain an error bound of $\mathcal{S}_+^{n,r}$ relative to a friendly constraint set $\mathcal{X}$. We should mention a recently-emerged related work [3]. Consider the following result.

**Theorem 3 (a special case of Lemma 3.1 in [3])** *Define $\tilde{\mathcal{S}}_+^{n \times r} = \{\boldsymbol{X} \in \mathbb{R}_+^{n \times r} \mid \|\boldsymbol{x}_j\|_2 = 1 \ \forall j\}$ as the column-wise non-negative unit sphere. For any $\boldsymbol{X} \in \tilde{\mathcal{S}}_+^{n \times r}$, we have the error bound*

$$\mathrm{dist}(\boldsymbol{X}, \mathcal{S}_+^{n,r}) \leq \sqrt{2r} \sqrt{\|\boldsymbol{X}\boldsymbol{1}\|_2^2 - r}. \tag{20}$$

The authors of [3] actually derived a more general result than the above, but the above result is considered the most representative and was the standard choice in the authors' numerical demonstrations. From the perspective of this study, the most interesting question lies in how the above error bound was shown. We will come back to this later. Using the error bound principle (applying (20) to the penalized formulation (2)), the authors of [3] gave the following exact penalization formulation

$$\min_{\boldsymbol{X} \in \tilde{\mathcal{S}}_+^{n,r}} f(\boldsymbol{X}) + \lambda \sqrt{\|\boldsymbol{X}\boldsymbol{1}\|_2^2 - r} \tag{21}$$

for the ONMF problem (19) or for CM problems for the NSOM case. The presence of a square root in the penalty of (21) makes the objective function non-smooth, which adds some challenge from the viewpoint of building algorithms. The authors of [3] handled problem (21) by using manifold optimization to deal with the manifold $\tilde{\mathcal{S}}_+^{n \times r}$ and by applying smooth approximation to deal with the penalty term. Taking inspiration from the above error bound result, another group of researchers derived a general error bound result as follows.

**Theorem 4 (Theorem 5 in [4])** *For any $\boldsymbol{X} \in \mathbb{R}^{n \times r}$, we have the error bound*

$$\mathrm{dist}(\boldsymbol{X}, \mathcal{S}_+^{n,r}) \leq 5 r^{\frac{3}{4}} (\|\boldsymbol{X}_-\|_{\mathrm{F}}^{\frac{1}{2}} + \|\boldsymbol{X}^\top \boldsymbol{X} - \boldsymbol{I}\|_{\mathrm{F}}^{\frac{1}{2}}), \tag{22}$$

*where $\boldsymbol{X}_- = \max\{-\boldsymbol{X}, \boldsymbol{0}\}$.*

### *F. A Modified Error Bound and a New EXPP for NSOMs*

We derive a modified error bound result for $\mathcal{S}_+^{n,r}$. Let

$$\tilde{\mathcal{B}}_+^{n \times r} = \{\boldsymbol{X} \in \mathbb{R}_+^{n \times r} \mid \|\boldsymbol{x}_j\|_2 \leq 1 \ \forall j\}$$

define the non-negative column-wise unit $\ell_2$-norm ball. Let

$$\psi_p(\boldsymbol{X}) = \left[ \sum_{j=1}^r c_1(\boldsymbol{x})^p + \sum_{i=1}^n \rho_1(\bar{\boldsymbol{x}}_i)^p \right]^{\frac{1}{p}} \tag{23}$$

be a penalty function, where $p \in \{1, 2\}$,

$$c_1(\boldsymbol{x}) = 1 - \|\boldsymbol{x}\|_2, \quad \rho_1(\boldsymbol{x}) = \|\boldsymbol{x}\|_1 - \|\boldsymbol{x}\|_\infty.$$

The penalty $\psi_p$ shares a similar rationale as the one in (12): for $\boldsymbol{X} \in \mathbb{R}^{n \times r}$ with $\|\boldsymbol{x}_j\|_2 \leq 1$ for all $j$, $c_1(\boldsymbol{x}_j)$ promotes $\boldsymbol{x}_j$ to have unit $\ell_2$ norm, while $\rho_1(\bar{\boldsymbol{x}}_i)$ promotes $\bar{\boldsymbol{x}}_i$ to be a scaled unit vector. We should note that $\|\boldsymbol{x}\|_1 \geq \|\boldsymbol{x}\|_\infty$, and that $\|\boldsymbol{x}\|_1 = \|\boldsymbol{x}\|_\infty$ if and only if $\boldsymbol{x}$ is a scaled unit vector. We should also recall from (18) that an NSOM has rows being scaled unit vectors. Our result is as follows.

**Theorem 5 (error bound for non-negative semi-orthogonal matrices)** *For any $\boldsymbol{X} \in \tilde{\mathcal{B}}_+^{n \times r}$, we have the error bounds*

$$\mathrm{dist}(\boldsymbol{X}, \mathcal{S}_+^{n,r}) \leq \nu \, \psi_2(\boldsymbol{X}) \leq \nu \, \psi_1(\boldsymbol{X}), \tag{24}$$

*where $\nu = \max\{\sqrt{6}, 2\sqrt{r}\}$.*

The proof of Theorem 5 is provided in Appendix D, and we will give insight into these error bounds in the next subsection. Theorem 5 can be used to build a new EXPP formulation. From (24) we have a further error bound

$$\nu \, \psi_1(\boldsymbol{X}) \leq \nu \left[ r - \|\boldsymbol{X}\|_{\mathrm{F}}^2 + \sum_{i=1}^n (\boldsymbol{1}^\top \bar{\boldsymbol{x}}_i - s_1(\bar{\boldsymbol{x}}_i)) \right],$$

where we use the fact that $\|\boldsymbol{x}_j\|_2 \geq \|\boldsymbol{x}_j\|_2^2$ whenever $\|\boldsymbol{x}_j\|_2 \leq 1$; we recall $s_1(\boldsymbol{x}) = x_{[1]}$; $\boldsymbol{1}^\top \bar{\boldsymbol{x}}_i - s_1(\bar{\boldsymbol{x}}_i)$ is an alternate form of $\rho_1(\bar{\boldsymbol{x}}_i)$ for non-negative $\bar{\boldsymbol{x}}_i$. The above error bound leads us to the following new EXPP formulation for the NSOM case:

$$\min_{\boldsymbol{X} \in \tilde{\mathcal{B}}_+^{n \times r}} F_\lambda(\boldsymbol{X}) = f(\boldsymbol{X}) + \lambda \left[ \sum_{i=1}^n (\boldsymbol{1}^\top \bar{\boldsymbol{x}}_i - s_1(\bar{\boldsymbol{x}}_i)) - \|\boldsymbol{X}\|_{\mathrm{F}}^2 \right] \tag{25}$$

Unlike the previous EXPP formulations, the above formulation has non-smooth penalty terms $-s_1(\bar{\boldsymbol{x}}_i)$'s. But since they are concave, they are considered not difficult to handle when we use majorization minimization techniques to build algorithms. Specifically, by Jensen's inequality, we have

$$-s_1(\boldsymbol{x}) \leq -s_1(\boldsymbol{x}') + x'_{l'} - x_{l'}, \quad \text{for any } \boldsymbol{x}, \boldsymbol{x}',$$

where $l'$ is such that $x'_{l'} = s_1(\boldsymbol{x}')$. The above inequality can be used to conveniently construct a majorant for the penalty function. Moreover, it should be mentioned that the projection onto $\tilde{\mathcal{B}}_+^{n \times r}$ has a closed form. The projection $\Pi_{\tilde{\mathcal{B}}_+^{n \times r}}(\boldsymbol{Z})$ for a given matrix $\boldsymbol{Z} \in \mathbb{R}^{n \times r}$ corresponds to the projections $\Pi_{\mathcal{B}_+^n}(\boldsymbol{z}_j)$'s, where $\mathcal{B}_+^n = \mathcal{B}^n \cap \mathbb{R}_+^n$. The projection $\Pi_{\mathcal{B}_+^n}(\boldsymbol{z})$ for a given vector $\boldsymbol{z} \in \mathbb{R}^n$ equals $\boldsymbol{z}_+$ if $\|\boldsymbol{z}_+\|_2 \leq 1$, and $\boldsymbol{z}_+ / \|\boldsymbol{z}_+\|_2$ if $\|\boldsymbol{z}_+\|_2 > 1$; here, $\boldsymbol{z}_+ = \max\{\boldsymbol{0}, \boldsymbol{z}\}$.

### G. Insights and Further Remarks for NSOMs

We should describe the insight behind the proof of Theorem 5. Some of the key steps of our proof are actually the same as those of the previous results in Theorems 3 and 4. We however change one important ingredient. Recall from (18) that every NSOM has rows being scaled unit vectors. This motivates us to consider the scaled unit vector set

$$\mathcal{W}^n := \{\boldsymbol{x} \in \mathbb{R}^n \mid \boldsymbol{x} = \alpha \boldsymbol{e}_i,\ \alpha \in \mathbb{R}, i \in \{1, \ldots, n\}\}.$$

In particular we consider the error bound for $\mathcal{W}^n$:

**Lemma 2** *For any $\boldsymbol{x} \in \mathbb{R}^n$ we have the error bound*

$$\operatorname{dist}(\boldsymbol{x}, \mathcal{W}^n) \leq \|\boldsymbol{x}\|_1 - \|\boldsymbol{x}\|_\infty = \rho_1(\boldsymbol{x}). \tag{26}$$

*Proof of Lemma 2:* Let $\boldsymbol{x} \in \mathbb{R}^n$ be given. Let $\boldsymbol{y} = x_l \boldsymbol{e}_l$, where $l$ is such that $|x_l| = \max\{|x|_1, \ldots, |x_n|\}$. It can be verified that $\boldsymbol{y} = \Pi_{\mathcal{W}^n}(\boldsymbol{x})$ and $\operatorname{dist}(\boldsymbol{x}, \mathcal{W}^n) = \|\boldsymbol{x} - \boldsymbol{y}\|_2 \leq \|\boldsymbol{x} - \boldsymbol{y}\|_1 = \rho_1(\boldsymbol{x})$. ■

Our proof differs in that we use the error bound in Lemma 2, while the prior results used a different bound. Interested readers are referred to Appendix E, wherein we delineate the difference.

In fact, from the proof of Theorem 5, we notice the following generalization.

**Corollary 1** *Let $\mathcal{S}_{\mathsf{e}}^{n,r} = \{\boldsymbol{X} \in \mathbb{R}^{n\times r} \mid \|\boldsymbol{x}_j\|_2 = 1, \bar{\boldsymbol{x}}_i \in \mathcal{W}^r,\ \forall i, j\}$. For any $\boldsymbol{X} \in \mathbb{R}^{n\times r}$ with $\|\boldsymbol{x}_j\|_2 \leq 1$ for all $j$, we have error bounds $\operatorname{dist}(\boldsymbol{X}, \mathcal{S}_{\mathsf{e}}^{n,r}) \leq \nu\, \psi_2(\boldsymbol{X}) \leq \nu\, \psi_1(\boldsymbol{X})$, where $\nu$ is given by the one in Theorem 5.*

We omit the proof as it is a trivial variation of the proof of Theorem 5. Note that $\mathcal{S}_+^{n,r} \subset \mathcal{S}_{\mathsf{e}}^{n,r} \subset \mathcal{S}^{n,r}$, and $\mathcal{S}_{\mathsf{e}}^{n,r}$ permits negative components.

It is also worth noting that Theorem 5 has connections with the previous results in Theorems 3 and 4. To describe it, we first derive the following result.

**Lemma 3** *For any $\boldsymbol{x} \in \mathbb{R}^n$, it holds that $\rho_1(\boldsymbol{x})^2 \leq \|\boldsymbol{x}\|_1^2 - \|\boldsymbol{x}\|_2^2 = \rho_2(\boldsymbol{x})$.*

*Proof of Lemma 3:* Without loss of generality, assume $|x_1| \geq |x_i|$ for all $i$. We have

$$\begin{aligned}\rho_1(\boldsymbol{x})^2 &= (\|\boldsymbol{x}\|_1 - |x_1|)^2 = \|\boldsymbol{x}\|_1^2 - 2|x_1|\|\boldsymbol{x}\|_1 + |x_1|^2 \\ &\leq \|\boldsymbol{x}\|_1^2 - 2(|x_1|^2 + |x_2|^2 + \cdots + |x_n|^2) + |x_1|^2 \\ &\leq \|\boldsymbol{x}\|_1^2 - \|\boldsymbol{x}\|_2^2.\end{aligned}$$

The proof is complete. ■

From Lemma 3 we have the following further result.

**Lemma 4** *For any $\boldsymbol{X} \in \tilde{\mathcal{B}}_+^{n\times r}$, it holds that*

$$\psi_2(\boldsymbol{X}) \leq \sqrt{r + \|\boldsymbol{X}\boldsymbol{1}\|_2^2 - 2\|\boldsymbol{X}\|_{\mathrm{F}}^2} \tag{27a}$$

$$= \|\boldsymbol{X}^\top \boldsymbol{X} - \boldsymbol{I}\|_{\ell_1}^{\frac{1}{2}}. \tag{27b}$$

$$\leq r\|\boldsymbol{X}^\top \boldsymbol{X} - \boldsymbol{I}\|_{\mathrm{F}}^{\frac{1}{2}} \tag{27c}$$

*Proof of Lemma 4:* Let $\boldsymbol{X} \in \tilde{\mathcal{B}}_+^{n\times r}$ be given. Since $\|\boldsymbol{x}_j\|_2 \leq 1$, we have $c_1(\boldsymbol{x}_j)^2 \leq 1 - \|\boldsymbol{x}_j\|_2 \leq 1 - \|\boldsymbol{x}_j\|_2^2 = c_2(\boldsymbol{x}_j)$. By Lemma 3 we have $\rho_1(\bar{\boldsymbol{x}}_i)^2 \leq \rho_2(\bar{\boldsymbol{x}}_i)$. Applying these results to (23), and observing (12), we have (27a). Eq. (27b) is a direct consequence of Lemma 1. Eq. (27c) is the consequence of the norm result $\|\boldsymbol{a}\|_1 \leq \sqrt{n}\|\boldsymbol{a}\|_2$ for $\boldsymbol{a} \in \mathbb{R}^n$. ■

By Theorem 5, all the functions in (27) are effective error bound functions of $\mathcal{S}_+^{n,r}$ relative to $\tilde{\mathcal{B}}_+^{n\times r}$. If we limit $\boldsymbol{X}$ to lie in $\tilde{\mathcal{S}}_+^{n\times r}$, then the effective error bound function in (27a) becomes that in (20) in Theorem 3. The effective error bound function in (27c) is identical to that in (22) in Theorem 4 for the special case of $\boldsymbol{X} \in \tilde{\mathcal{B}}_+^{n\times r}$. Hence, simply speaking, Theorem 5 can produce the error bound results in the prior studies.

We end with a further comment. The penalty functions $\rho_1$ and $\rho_2$ for scaled unit vectors were considered in [2]. The authors of that work studied a variation of the ONMF problem (17) wherein they remove the unit $\ell_2$-norm column constraints with $\boldsymbol{X}$. They showed stationarity results with the use of $\rho_1$ and $\rho_2$. Their analysis is not based on error bounds, and the problem nature with $\mathcal{S}_+^{n,r}$ is considered more challenging than that with the scaled unit vectors.

## IV. Numerical Results in Different Applications

In this section we numerically demonstrate EXPP in different applications.

### A. The Algorithm for EXPP

We should specify the algorithm used to implement EXPP in our experiments. The algorithm used is an extrapolated variant of the projected gradient (PG) method, which was numerically found to have faster convergence; see, e.g., [12]. It is also the same algorithm used in our prior studies [13], [14]. The algorithm is shown in Algorithm 1. To explain it, let us write down the EXPP formulation

$$\min_{\boldsymbol{x}\in\mathcal{X}} F_\lambda(\boldsymbol{x}) = f(\boldsymbol{x}) + \lambda\, h(\boldsymbol{x}), \tag{28}$$

where $\mathcal{X} = \operatorname{conv}(\mathcal{V})$; $h(\boldsymbol{x}) = -\|\boldsymbol{x}\|_2^2$; $f$ is assumed to be differentiable and have Lipschitz continuous gradient on $\mathcal{X}$. Line 6–7 of Algorithm 1 is the extrapolated PG step. If we replace $\boldsymbol{z}$ by $\tilde{\boldsymbol{x}}^l$ and $\nabla G_{\lambda_k}(\boldsymbol{z}|\tilde{\boldsymbol{x}}^l)$ by $\nabla F_{\lambda_k}(\tilde{\boldsymbol{x}}^l)$, we return to the baseline PG step. The point $\boldsymbol{z}$ is an extrapolated point, and the extrapolation sequence $\{\alpha_l\}$ is chosen as the FISTA sequence [15]. The function $G_\lambda(\boldsymbol{x}|\boldsymbol{x}')$ is a majorant of $F_\lambda(\boldsymbol{x})$ at $\boldsymbol{x}'$, and we apply majorization before the PG step. Specifically we majorize the concave $h$ by $u(\boldsymbol{x}|\boldsymbol{x}') = -\|\boldsymbol{x}'\|_2^2 - 2(\boldsymbol{x}-\boldsymbol{x}')^\top \boldsymbol{x}'$ (it is the result of Jensen's inequality), and then we construct the majorant $G_\lambda(\boldsymbol{x}|\boldsymbol{x}') = f(\boldsymbol{x}) + \lambda u(\boldsymbol{x}|\boldsymbol{x}')$ (for $\lambda > 0$). With this choice we choose the step size $\eta$ as the reciprocal of the Lipschitz constant of $\nabla G_\lambda(\boldsymbol{x}|\boldsymbol{x}')$. Furthermore, we apply a homotopy strategy wherein we gradually increase penalty parameter $\lambda$ so that we start with a possibly convex problem and end with the target EXPP problem (with exact penalization).

Algorithm 1 is also used to implement the new EXPP formulations (11), (16) and (25) for the PPM, SAM and

**Algorithm 1** A PG-type algorithm for (28), with homotopy

1: **given:** a sequence $\{\lambda_k\}$, an extrapolation sequence $\{\alpha_l\}$, and a starting point $\boldsymbol{x}^0$
2: $k \leftarrow 0$
3: **repeat**
4: $\quad \tilde{\boldsymbol{x}}^0 \leftarrow \boldsymbol{x}^k$, $\boldsymbol{z} \leftarrow \boldsymbol{x}^k$, $l \leftarrow 0$
5: $\quad$ **repeat**
6: $\quad\quad \tilde{\boldsymbol{x}}^{l+1} \leftarrow \Pi_{\mathcal{X}}(\boldsymbol{z} - \eta \nabla G_{\lambda_k}(\boldsymbol{z}|\tilde{\boldsymbol{x}}^l))$ for some $\eta > 0$
7: $\quad\quad \boldsymbol{z} \leftarrow \tilde{\boldsymbol{x}}^{l+1} + \alpha_l(\tilde{\boldsymbol{x}}^{l+1} - \tilde{\boldsymbol{x}}^l)$
8: $\quad\quad l \leftarrow l + 1$
9: $\quad$ **until** a stopping rule is met
10: $\quad \boldsymbol{x}^{k+1} \leftarrow \tilde{\boldsymbol{x}}^l$
11: $\quad k \leftarrow k + 1$
12: **until** a stopping rule is met
13: **output:** $\boldsymbol{x}^k$

NSOM cases, respectively; we will call them "EXPP–II" to avoid confusion with the previous EXPP. For the PPM and SAM cases we have $h(\boldsymbol{X}) = \|\boldsymbol{X}\mathbf{1}\|_2^2 - 2\|\boldsymbol{X}\|_{\rm F}^2$. We choose $u(\boldsymbol{X}|\boldsymbol{X}') = \|\boldsymbol{X}\mathbf{1}\|_2^2 - 4{\rm tr}(\boldsymbol{X}^\top \boldsymbol{X}') + 2\|\boldsymbol{X}'\|_{\rm F}^2$; we keep the convex term $\|\boldsymbol{X}\mathbf{1}\|_2^2$ and majorize the concave term $-2\|\boldsymbol{X}\|_{\rm F}^2$. For the NSOM case we have $h(\boldsymbol{X}) = -\|\boldsymbol{X}\|_{\rm F}^2 + \sum_{i=1}^n (\mathbf{1}^\top \bar{\boldsymbol{x}}_i - s_1(\bar{\boldsymbol{x}}_i))$. We choose $u(\boldsymbol{X}|\boldsymbol{X}') = -2{\rm tr}(\boldsymbol{X}^\top \boldsymbol{X}') + \|\boldsymbol{X}'\|_{\rm F}^2 + \sum_{i=1}^n (\mathbf{1}^\top \bar{\boldsymbol{x}}_i - x_{i,l_i'})$, where $l_i'$ is such that $x'_{i,l_i'} = \max\{x'_{i,1}, \ldots, x'_{i,r}\}$; again we keep the convex terms and majorize the concave terms. The rest of the operations are identical.

We should mention that Algorithm 1 is equipped with guarantees of convergence to a stationary point. To be specific, the majorization and extrapolated PG loop in Line 5–9 of Algorithm 1 was shown to be able to converge to a stationary point of problem (28) (with $\lambda = \lambda_k$) [13]. The premise of this is that $h$ is differentiable and has Lipschitz continuous gradient. This premise is satisfied in most of the above described cases, with the NSOM case being the only exception. The critical-point convergence for the NSOM case (nonsmooth $h$) is a subject of future study.

We may deal with applications that have non-differentiable objective function $f$. For such cases we replace the extrapolated PG step in Line 6–7 of Algorithm 1 with the projected subgradient method.

Some details with the stopping rules of Algorithm 1 are as follows. Unless specified otherwise, the stopping rule for the inner loop is either $\|\tilde{\boldsymbol{x}}^{l+1} - \tilde{\boldsymbol{x}}^l\|_2 / \|\tilde{\boldsymbol{x}}^l\|_2 < \varepsilon_1$ or $l > \bar{L}$ for some given $\varepsilon_1$ and $\bar{L}$. Unless specified otherwise, the stopping rule for the outer loop is $\|\boldsymbol{x}^{k+1} - \boldsymbol{x}^k\|_2 / \|\boldsymbol{x}^k\|_2 < \varepsilon_2$, ${\rm dist}(\boldsymbol{x}^k, \mathcal{V}) \le \varepsilon_3$, or $\lambda_k > \bar{\lambda}$ for some given $\varepsilon_2$, $\varepsilon_3$ and $\bar{\lambda}$.

### *B. MIMO detection with MPSK Constellations*

We consider MIMO detection. In this problem, we have a received signal model $\boldsymbol{y} = \boldsymbol{H}\boldsymbol{x} + \boldsymbol{v}$, where $\boldsymbol{y} \in \mathbb{C}^m$ is the received signal; $\boldsymbol{H} \in \mathbb{C}^{m \times n}$ is the channel; $\boldsymbol{x} \in \mathbb{C}^n$ is the transmitted symbol vector; $\boldsymbol{v} \in \mathbb{C}^m$ is noise. The goal is to detect $\boldsymbol{x}$ from $\boldsymbol{y}$. Assuming that every $x_i$ is drawn from an $M$-ary phase shift keying (PSK) constellation $\Theta_M = \{x \in \mathbb{C} \mid x = e^{{\rm j}\frac{2\pi l}{M} + {\rm j}\frac{\pi}{M}},\ l \in \{0, 1, \ldots, M-1\}\}$, we consider the maximum-likelihood (ML) detector

$$\hat{\boldsymbol{x}} = \arg \min_{\boldsymbol{x} \in \Theta_M^n} \|\boldsymbol{y} - \boldsymbol{H}\boldsymbol{x}\|_2^2.$$

We use EXPP to handle the ML detector.

The benchmarked algorithms are (i) MMSE: the minimum mean square detector; (ii) SDR: the semidefinite relaxation detector [16]; (iii) LAMA: the approximate message passing method in [17] with incorporation of damping [18] (with damping factor 0.7). The parameter settings of EXPP are $\lambda_0 = 0.01$, $\lambda_{k+1} = 5\lambda_k$, $\varepsilon_1 = 10^{-4}$, $\bar{L} = 100$, $\bar{\lambda} = 10^4$.

The simulation settings are as follows. A number of $10,000$ Monte-Carlo trials were run. The channel $\boldsymbol{H}$ was generated based on a correlated MIMO channel model $\boldsymbol{H} = \boldsymbol{R}_r^{\frac{1}{2}} \tilde{\boldsymbol{H}} \boldsymbol{R}_t^{\frac{1}{2}}$, where the components of $\tilde{\boldsymbol{H}}$ are independent and identically distributed (i.i.d.) and follow a circular Gaussian distribution with mean zero and variance 1, and $\boldsymbol{R}_t$ and $\boldsymbol{R}_r$ follow the exponential model with parameter $\rho = 0.2$ [19]. The vector $\boldsymbol{v}$ was generated as a component-wise i.i.d. circular Gaussian noise with mean 0 and variance $\sigma^2$.

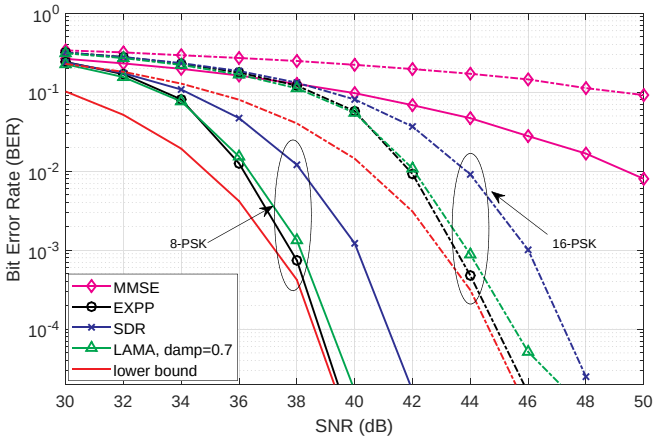



Fig. 2: MIMO detection BER performance. Solid lines: 8-PSK; dashed lines: 16-PSK.

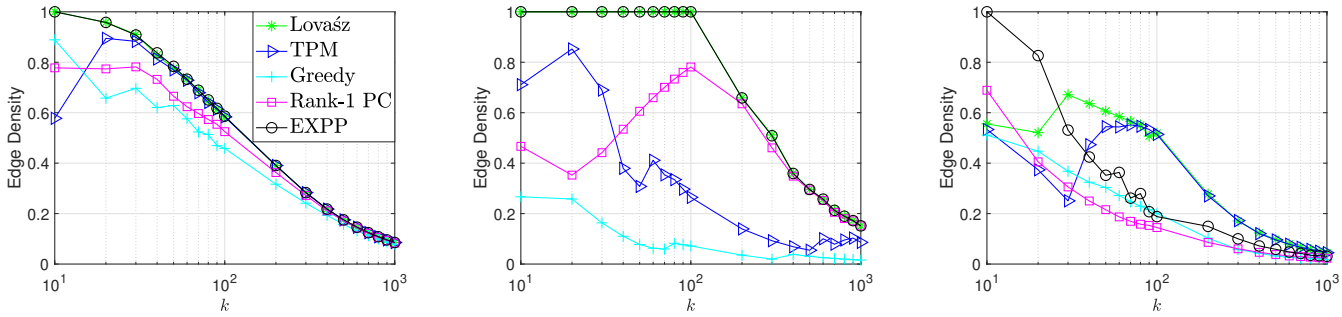
| Algorithms | SDR | EXPP | LAMA |
|---|---|---|---|
| time (in sec.) | 1.017 | 0.026 | 0.021 |

TABLE I: Average runtime performance in MIMO detection. 8-PSK, SNR= 38dB.

Fig. 2 and Table I display the bit-error-rate (BER) and runtime performances of the various detectors, respectively. The problem size is $m = n = 80$. "Lower bound" is the performance baseline when there is no MIMO interference. EXPP gives the best BER performance and is closely followed by LAMA. The runtime performance of EXPP is slightly worse than that of LAMA, but is still competitive.

### *C. Densest $k$-Subgraph Problem*

We consider the densest $k$-subgraph (D$k$S) problem. The problem is to identify the most connected subgraph from a




(a) Loc-Gowalla (b) NotreDame (c) Google

(d) YouTube (e) Talk (f) LiveJournal

Fig. 3: Performance with the D$k$S problem.

graph. Let $\boldsymbol{W} \in \mathbb{R}^{n\times n}$ be the adjacency matrix of a given graph. Given a subgraph size $k$, the D$k$S problem is given by

$$\min_{\boldsymbol{x}\in\mathcal{U}_k^n} -\boldsymbol{x}^\top \boldsymbol{W}\boldsymbol{x},$$

where $\mathcal{U}_k^n = \{\boldsymbol{x} \in \{0,1\}^n \mid \mathbf{1}^\top\boldsymbol{x} = k\}$. We use EXPP to handle this problem.

The benchmarked algorithms are (i) Greedy: the greedy algorithm in [20]; (ii) TPM: the truncated power method [21]; (iii) Rank-1 PC: the rank-one binary principal component approximation method [22]; (iv) Lovaśz: Lovaśz relaxation, implemented by the ADMM method and a Frank-Wolfe post-processing step [10]. The parameter settings of EXPP are $\lambda_0 = -\sigma_1(\boldsymbol{W})$, $\lambda_{k+1} = \lambda_k + 0.2\sigma_1(\boldsymbol{W})$, $\varepsilon_1 = 10^{-3}$, $\bar{L} = 50$, $\varepsilon_3 = 10^{-2}$, $\bar{\lambda} = 1.1\sigma_1(\boldsymbol{W})$. Note that the above choice of $\lambda_0$ is to make the corresponding EXPP problem convex; and that if $\lambda_k < 0$ then we choose $u(\boldsymbol{x}|\boldsymbol{x}') = h(\boldsymbol{x})$ (no majorization).

We performed our test on large real-world graphs [23]. Table II describes the graph sizes of the tested datasets, which range from nearly $200,000$ nodes to nearly $4,000,000$ nodes. We adopt the same data pre-processing as that in [10]. Fig. 3 shows the edge densities, $\boldsymbol{x}^\top\boldsymbol{W}\boldsymbol{x}/(k^2-k)$, achieved by the various algorithms. Lovaśz gives the best performance in all the datasets. EXPP is the second best. Except for the Google dataset, EXPP achieves similar performance as Lovaśz. Table II shows the runtimes. We see that EXPP runs faster than Lovaśz.

### *D. Graph Matching*

We consider the GM problem described in Section III-A. The benchmarked algorithms are (i) PATH [6]; (ii) GNCCP [24]; (iii) LAGSA [25]. PATH was concisely reviewed in Section III-A, and GNCCP and LAGSA are related methods. We employ EXPP–II in (11). The parameter settings of EXPP are $\lambda_0 = 10^{-5}$, $\lambda_{k+1} = 4\lambda_k$, $\varepsilon_1 = 10^{-4}$, $\varepsilon_2 = 10^{-5}$, $\bar{L} = 100$, $\bar{\lambda} = 10^3$.

We performed our test on graphs constructed from various image datasets. We follow a standard protocol to extract graphs from a pair of images; see e.g., [26]. We tested several image sets from 5 datasets: CMU-House [27], PASCAL [28], DTU [29], DAISY [30], and SUIRD [31]. CMU-House and PASCAL are small graphs and are commonly used in GM papers; DTU, DAISY, and SUIRD are large graphs. For each pair of images, we performed 10 independent graph constructions and used them for experiments.

Table III shows the average matching accuracies and runtimes of the different algorithms; the matching accuracy is defined as the ratio of the number of correctly matched nodes to the node size $n$. We did not test PATH and LAGSA on large-size graphs because they run too slowly. EXPP–II gives the best matching accuracy performance for all the datasets. More importantly, EXPP–II runs much faster than the other algorithms. This is particularly so for large-size graphs.

We also want to examine sensitivities with outliers. We used the same protocol to construct graphs, except that we added 10 outliers to the graphs. Table IV shows the average matching accuracies and runtimes. EXPP–II is seen to show better matching accuracies than the benchmarked algorithms.

For visual illustration, we provide matching examples in Fig. 4–5. An example for the case of $n = 500$ is enlarged and shown in Fig. 6 for clear visualization.

| Datasets | $n$ | $m$ | Runtime (in sec.) | | | | |
|---|---|---|---|---|---|---|---|
| | | | Lovaśz | TPM | Greedy | Rank-1 PC | EXPP |
| Lcc-Gowalla | 196,591 | 950,327 | 41.32 | 0.10 | 0.02 | 0.56 | 2.40 |
| NotreDame | 325,729 | 1,497,134 | 18.95 | 0.07 | 0.03 | 0.91 | 3.11 |
| Google | 875,713 | 5,105,039 | 274.13 | 0.50 | 0.13 | 5.91 | 25.36 |
| YouTube | 1,134,890 | 2,987,624 | 1.05e3 | 1.04 | 0.13 | 2.94 | 18.27 |
| Talk | 2,394,385 | 5,021,410 | 1.21e3 | 1.50 | 0.33 | 5.02 | 36.92 |
| LiveJournal | 3,997,962 | 34,681,189 | 1.29e3 | 3.82 | 0.66 | 22.78 | 124.68 |

TABLE II: Average runtime with the D$k$S problem. $n$ = number of nodes, $m$ = number of edges.

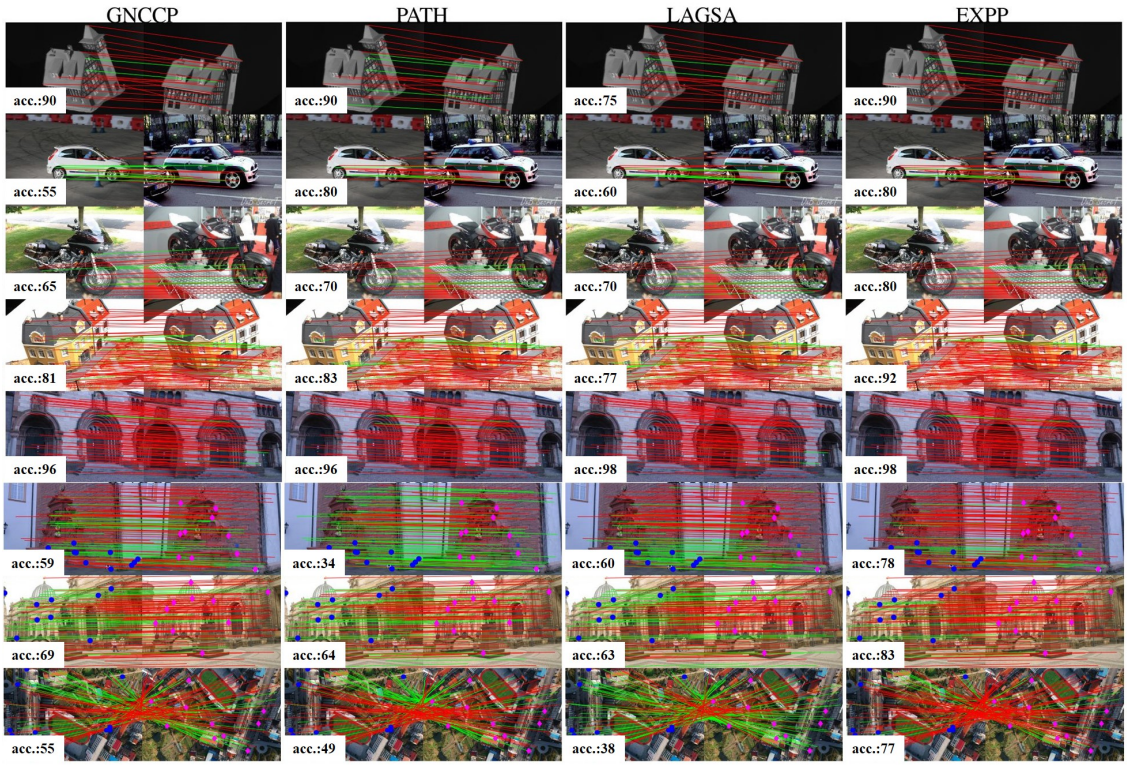



Fig. 4: Illustration of graph matching instances. First row to third row: CMU-House, Car, Motorbike, with $n = 20$; fourth row to fifth row: DTU-House, HerzJesu, with $n = 100$; sixth row to eighth row: Fountain, Semper and Stadium, $n = 110$, 10 outliers. Red lines: correctly matched nodes, green lines: mismatched nodes. Blue dots and magenta dots represent outliers in the two figures respectively.

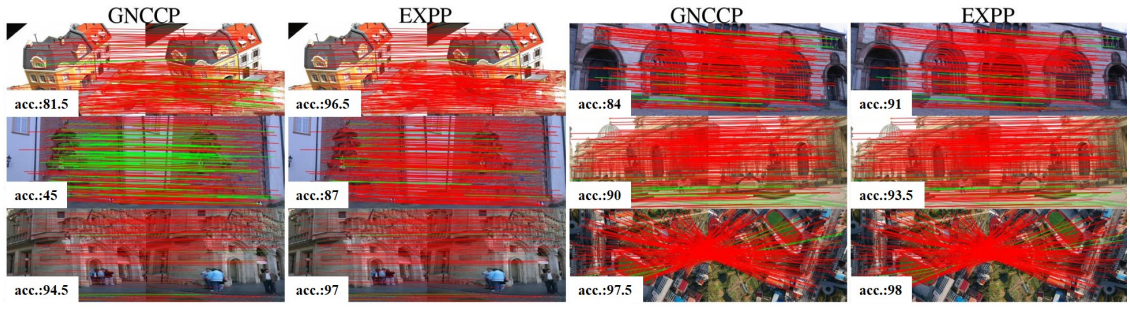



Fig. 5: Illustration of graph matching instances with $n = 200$. First row: DTU-House, HerzJesu; second row: Fountain, Semper; third row: Brussels, Stadium. Red lines: correctly matched nodes, green lines: mismatched nodes.

### *E. Size-Constrained Clustering*

We consider the size-constrained clustering problem described in Section III-C. The benchmarked algorithms are (i) AM: an alternating minimization method [9]; (ii) SC: $K$-means followed by a post-processing based on mixed integer linear programming [32]; (iii) E-Kmeans: an extension of $K$-means [33]. We employ EXPP–II in (16). We revise the objective function (cf. (14)) as $f(\boldsymbol{x}) + \gamma\|\boldsymbol{X}\|_{\mathrm{F}}^2$, with $\gamma = 1.1\sigma_1(\boldsymbol{R})/\kappa_{[r]}$, such that the revised objective function is convex and EXPP–II starts with a convex problem. The parameter

| Data | $n$ | Metric | EXPP–II | GNCCP | PATH | LAGSA |
|---|---|---|---|---|---|---|
| CMU-House | 20 | acc. | **77.00** | 74.00 | 54.50 | 43.00 |
| | | time | **0.03** | 0.09 | 1.21 | 0.41 |
| Car in PASCAL | 20 | acc. | **82.50** | 65.00 | 59.50 | 61.00 |
| | | time | **0.02** | 0.04 | 1.22 | 0.76 |
| Motorbike in PASCAL | 20 | acc. | **96.00** | 80.50 | 70.00 | 73.50 |
| | | time | **0.02** | 0.08 | 1.16 | 0.64 |
| DTU-House | 500 | acc. | **83.62** | 60.64 | - | - |
| | | time | **16** | 711 | - | - |
| HerzJesu in DAISY | 500 | acc. | **82.56** | 49.18 | - | - |
| | | time | **11** | 786 | - | - |
| Semper in DAISY | 500 | acc. | **92.04** | 54.46 | - | - |
| | | time | **13** | 619 | - | - |
| Stadium in SUIRD | 500 | acc. | **91.68** | 40.46 | - | - |
| | | time | **10** | 435 | - | - |

TABLE III: Graph matching results. $n$ = number of nodes, acc. = matching accuracy in percent, time = runtime in seconds.

| Data | Metric | EXPP–II | GNCCP | PATH | LAGSA |
|---|---|---|---|---|---|
| DTU-House | acc. | **75.70** | 60.10 | 29.10 | 41.60 |
| | time | **0.51** | 3.57 | 234.55 | 283.32 |
| HerzJesu in DAISY | acc. | **78.50** | 58.90 | 22.90 | 35.50 |
| | time | **0.50** | 3.71 | 209.89 | 307.61 |
| Fountain in DAISY | acc. | **79.50** | 59.00 | 30.90 | 45.10 |
| | time | **0.43** | 2.85 | 325.34 | 278.73 |
| Semper in DAISY | acc. | **79.50** | 59.00 | 30.90 | 45.10 |
| | time | **0.50** | 4.42 | 236.76 | 371.07 |
| Brussels in DAISY | acc. | **77.80** | 48.40 | 37.80 | 40.40 |
| | time | **0.48** | 3.08 | 207.81 | 302.79 |
| Stadium in SUIRD | acc. | **82.10** | 63.10 | 43.00 | 58.20 |
| | time | **0.49** | 3.79 | 213.74 | 308.04 |

TABLE IV: Graph matching results in the presence of outliers. $n = 110$ nodes, 10 outliers, acc. = matching accuracy in percent, time = runtime in seconds.

settings of EXPP–II are $\lambda_0 = 10^{-3}K\nu$, $\lambda_{k+1} = \lambda_k + 0.1K\nu$, $\varepsilon_1 = 10^{-4}$, $\varepsilon_3 = 10^{-2}$, $\bar{L} = 200$, $\bar{\lambda} = K\nu$, where $\nu$ is specified in Theorem 2 and $K$ is the Lipschitz constant of $f(\boldsymbol{X})$ on $\tilde{\mathcal{U}}_{\boldsymbol{\kappa}}^{n,r}$.

We evaluate the algorithms on several real-world datasets, some of which are used by the prior work [32]. The results are shown in Table V. We see that the clustering accuracies, defined as the number of correctly clustered data points normalized by the total number of data points, of EXPP–II are generally good compared to the other algorithms. Note that SC fails to work for the Fashion MNIST dataset, due possibly to the large data size. EXPP–II cannot compete with E-Kmeans in terms of runtimes, but otherwise EXPP–II is much faster than the other algorithms.

### *F. Orthogonal Non-Negative Matrix Factorization*

We consider the ONMF problem described in Section III-E. The benchmarked algorithms are (i) ONPMF [11]; (ii) NSNCP [2]: an alternating minimization algorithm for a penalized ONMF formulation; (iii) EP4ORTH [3]: a manifold optimization algorithm for the formulation reviewed in (21). We consider both the EXPP–II formulation in (25) and the EXPP formulation in Part I of this paper. In EXPP we have $h(\boldsymbol{X}) = -\|\boldsymbol{X}\|_{\mathrm{F}}^2$ and $\mathcal{X} = \mathcal{B}^{n,r} \cap \mathbb{R}_+^{n\times r}$, and we employ Dykstra's projection algorithm [38] to perform projection onto $\mathcal{X}$. The parameter settings of EXPP–II (respectively [resp.] EXPP) are $\lambda_0 = 10^{-15}$ (resp. $10^{-5}$), $\lambda_{k+1} = 10\lambda_k$ (resp. $5\lambda_k$), $\varepsilon_1 = 10^{-9}$, $\bar{L} = 50$ (resp. 100), $\bar{\lambda} = K\nu$, where $\nu$ is specified in Theorem 5 and $K$ is the Lipschitz constant of $f$ on $\tilde{\mathcal{B}}_+^{n,r}$. Unlike the previous applications, we found that EXPP and EXPP–II do not work if we start with a $\lambda_0$ such that the problem is convex (the corresponding solution is $\mathbf{0}$ which is not a good starting point). We initialize EXPP and EXPP–II with the NNDSVD method [39] which is commonly adopted by other ONMF methods such as ONPMF and EP4ORTH.

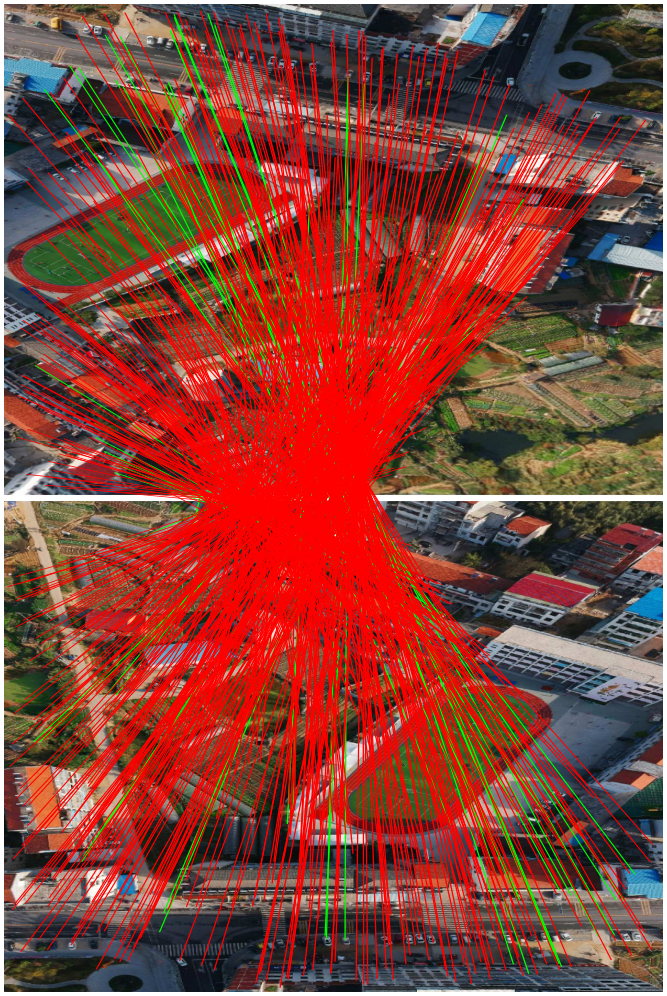


Fig. 6: A graph matching instance of Stadium with $n = 500$. Red lines: correctly matched nodes, green lines: mismatched nodes.

We evaluate the algorithms on various real-world datasets: face images [40], hyperspectral images [41], and text [42]. Table VI describes the dimensions of these datasets. The accuracies and runtimes of the various algorithms are shown in Table VII and Table VIII, respectively. In general, EXPP and EXPP–II perform comparably with the other methods. EXPP–II has faster runtimes while EXPP has higher clustering accuracies.

### *G. Sparse and Fair Principal Component Analysis*

We consider sparse and fair PCAs. Let $\boldsymbol{Y} \in \mathbb{R}^{m\times n}$ be a data matrix. Sparse PCA considers

$$\min_{\boldsymbol{X}\in\mathcal{S}^{m\times r}} f(\boldsymbol{X}) = -\mathrm{tr}(\boldsymbol{X}^\top \boldsymbol{R}\boldsymbol{X}) + \mu\|\boldsymbol{X}\|_{\ell_1},$$

| Datasets | $m$ | $n$ | $\boldsymbol{\kappa}$ | Metric | AM | EXPP–II | SC | E-Kmeans |
|---|---|---|---|---|---|---|---|---|
| Fashion MNIST [34] | 784 | $70,000$ | $r = 10$; 7K samples/cluster | acc. | 0.485 | **0.525** | - | 0.450 |
| | | | | time | 7792 | 125 | - | **51** |
| Iris [35] | 4 | 150 | $(50, 50, 50)$ | acc. | 92.00 | **93.33** | 81.33 | **93.33** |
| | | | | time | 0.153 | 0.100 | 0.283 | **0.006** |
| Glass Identification [36] | 9 | 214 | $(70, 76, 17, 13, 9, 29)$ | acc. | 37.38 | 48.60 | **52.34** | 45.33 |
| | | | | time | 0.390 | 0.025 | 0.286 | **0.007** |
| Balance Scale [37] | 4 | 625 | $(288, 288, 49)$ | acc. | 60.00 | **67.52** | 65.60 | 61.44 |
| | | | | time | 0.226 | 0.077 | 0.295 | **0.012** |

TABLE V: Size-constrained clustering results. acc. = clustering accuracy in percent, time = runtime in seconds.

| | TDT2 | Reuters | YALE | ORL | Jasper Ridge | Samson |
|---|---|---|---|---|---|---|
| $m$ | 34,642 | 5,423 | 1024 | 1,024 | 198 | 156 |
| $n$ | 7,809 | 748 | 165 | 150 | 10,000 | 9,025 |
| $r$ | 25 | 10 | 15 | 15 | 4 | 3 |

TABLE VI: Datasets for ONMF.

| Algorithms | TDT2 | Reuters | YALE | ORL | Jasper Ridge | Samson |
|---|---|---|---|---|---|---|
| DTPP | 67.20 | 49.20 | **43.64** | 48.00 | 80.28 | 89.86 |
| ONPMF | 69.47 | 50.40 | 35.15 | 42.67 | 79.71 | 94.81 |
| NSNCP | 67.44 | 49.20 | 42.42 | **56.00** | 81.49 | 94.23 |
| EP4ORTH | 70.48 | 50.27 | 37.58 | 40.67 | 82.86 | 93.65 |
| EXPP–II | 67.38 | 52.14 | 37.58 | 46.67 | 90.27 | 92.72 |
| EXPP | **74.30** | **64.84** | 37.58 | 42.67 | **94.46** | **95.86** |

TABLE VII: ONMF clustering accuracies in percent.

| Algorithms | TDT2 | Reuters | YALE | ORL | Jasper Ridge | Samson |
|---|---|---|---|---|---|---|
| DTPP | 87.46 | 2.19 | **0.20** | **0.11** | **0.63** | **0.63** |
| ONPMF | 1008.67 | 31.94 | 5.02 | 4.38 | 25.10 | 15.90 |
| NSNCP | 79.95 | 1.96 | 1.79 | 0.77 | 5.38 | 2.25 |
| EP4ORTH | 40.66 | 1.37 | 1.10 | 1.08 | 2.21 | 2.14 |
| EXPP–II | **10.11** | **0.36** | 0.25 | 0.16 | 2.68 | 2.34 |
| EXPP | 63.60 | 2.60 | 0.88 | 0.68 | 8.68 | 6.57 |

TABLE VIII: ONMF runtimes in seconds.

where $\mu > 0$ is given; $\boldsymbol{R} = \boldsymbol{Y}\boldsymbol{Y}^\top$ is the data correlation matrix. The aim is to recover sparse principal components. Let $\boldsymbol{Y}_t \in \mathbb{R}^{m\times n_t}$, $t = 1, \ldots, T$, be data matrices of different groups. Fair PCA considers

$$\min_{\boldsymbol{X}\in\mathcal{S}^{m\times r}} f(\boldsymbol{X}) = \max_{t=1,\ldots,T} -\mathrm{tr}(\boldsymbol{X}^\top \boldsymbol{R}_t \boldsymbol{X}),$$

where $\boldsymbol{R}_t = \boldsymbol{Y}_t\boldsymbol{Y}_t^\top$ is the data correlation matrix associated with the $t$th group. The aim of fair PCA is to reduce possible bias to individual groups. We also consider sparse fair PCA

$$\min_{\boldsymbol{X}\in\mathcal{S}^{m\times r}} f(\boldsymbol{X}) = \max_{t=1,\ldots,T} -\mathrm{tr}(\boldsymbol{X}^\top \boldsymbol{R}_t \boldsymbol{X}) + \mu\|\boldsymbol{X}\|_{\ell_1}.$$

We use EXPP to handle all the above PCAs.

The benchmarked algorithms for sparse PCA are: (i) IMRP: iterative minimization of rectangular Procrustes [43]; (ii) Gpower-$\ell_1$: generalized power method with $\ell_1$ penalty [44]. The benchmarked algorithms for fair PCA are (i) ARPGDA: the alternating Riemannian projected gradient descent ascent algorithm [45]; (ii) F-FPCA: a subgradient-type algorithm [46]. There is no algorithm to benchmark for sparse fair PCA.

EXPP is implemented by the projected subgradient method since the objective functions are non-smooth. The parameter settings of EXPP are $\lambda_0 = 0$, $\lambda_{k+1} = \lambda_k + 0.1K$, $\varepsilon_1 = 10^{-3}$, $\varepsilon_3 = 10^{-6}$, $\bar{L} = 300$, $\bar{\lambda} = K$, where $K$ is the Lipschitz constant of $f$ over $\mathcal{B}^{n,r}$. The step size of the projected subgradient method is set as $c/\sqrt{l+1}$ for some $c > 0$. We initialize EXPP with PCA. The EXPP for sparse PCA, fair PCA, and sparse fair PCA are called EXPP–S, EXPP–F, and EXPP–SF, respectively.

We use the following performance metrics to evaluate the performance of the various algorithms: explained variance and cardinality, which indicate the trade-off between sparsity and variance; and minimum variance, which measures fairness. Explained variance is measured as

$$\mathrm{tr}(\boldsymbol{X}^\top \boldsymbol{R}\boldsymbol{X})/\mathrm{tr}(\hat{\boldsymbol{X}}^\top \boldsymbol{R}\hat{\boldsymbol{X}}),$$

where $\boldsymbol{X}$ is the principal component matrix recovered by an algorithm; $\hat{\boldsymbol{X}}$ is that of PCA. Minimum variance is measured as

$$\min_{t=1,\ldots,T} \mathrm{tr}(\boldsymbol{X}^\top \boldsymbol{R}_t \boldsymbol{X}).$$

As a minor step, we project $\boldsymbol{X}$ onto $\mathcal{S}^{n,r}$ before we evaluate the above performance metrics; some algorithms such as Gpower-$\ell_1$ do not guarantee semi-orthogonality exactly. The cardinality is measured as the number of elements that are larger than $0.01\max_{i,j}|x_{ij}|$. We normalize minimum variance and cardinality to $[0, 1]$.

We test the algorithms on three real-world datasets: MNIST, Fashion MNIST, and CIFAR10. Each dataset consists of 10 different classes. The information about the datasets and the corresponding EXPP parameters are provided in Table IX. To create a data matrix, we randomly selected 5 classes. Then, for each class, we randomly selected data points to form one data matrix $\boldsymbol{Y}_t$. The data lengths of the different classes are $\{n_t\} = \{5, 10, 500, 1000, 5000\}$; we consider unbalanced data groups. We conducted 100 Monte Carlo runs for each dataset.

The results are shown in Table X. For sparse PCA, we see that EXPP–S performs better than Gpower-$\ell_1$ while IMRP outperforms EXPP–S. We note that IMPR adopts a different sparsity penalty. For fair PCA, EXPP–F performs comparably with ARPGDA and works better than F-FPCA. EXPP–SF is seen to be able to strike a balance between sparsity and fairness. The numerical results demonstrate the versatility of EXPP to handle different formulations.

## V. Conclusion

As the second part of our study, we developed new EXPP formulations for the cases of PPMs, SAMs and NSOMs. They have lower requirements with constraints and can be efficiently handled from the viewpoint of building algorithms. We also demonstrated the utility of EXPP by performing numerical experiments on a variety of applications.

| Datasets | Image Size | Number of Images | Data Type | EXPP–S | EXPP–SF | EXPP–F |
|---|---|---|---|---|---|---|
| MNIST [47] | $28 \times 28$ | 70000 | Digits | $c = 0.025$, $\mu = 0.3$ | $c = 10$, $\mu = 0.2$ | $c = 0.5$ |
| Fashion MNIST [34] | $28 \times 28$ | 70000 | Clothing | $c = 0.05$, $\mu = 0.25$ | $c = 100$, $\mu = 0.2$ | $c = 20$ |
| CIFAR10 [48] | $32 \times 32$ | 60000 | real-world objects | $c = 0.1$, $\mu = 0.25$ | $c = 20$, $\mu = 0.35$ | $c = 20$ |

TABLE IX: Datasets information and EXPP parameter setups. $c$: the constant in stepsize; $\mu$: sparsity penalty parameter.

| Algorithms | MNIST, $r = 10$. | | | | Fashion MNIST, $r = 15$. | | | | CIFAR10, $r = 20$. | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Expl. Var. | Min. Var. | Card. | Time /s | Expl. Var. | Min. Var. | Card. | Time /s | Expl. Var. | Min. Var. | Card. | Time /s |
| IMRP | **0.7733** | 0.3149 | 0.1037 | 0.6130 | **0.7890** | 0.3093 | **0.0642** | 1.6011 | **0.7126** | 0.5665 | 0.0462 | 4.0329 |
| Gpower-$\ell_1$ | 0.6252 | 0.3010 | 0.0998 | **0.4039** | 0.7070 | 0.3141 | 0.1056 | 1.8938 | 0.4876 | 0.3796 | **0.0445** | 4.0616 |
| EXPP–S | 0.6439 | 0.2427 | **0.0834** | 0.7227 | 0.7579 | 0.2696 | 0.0769 | **0.4915** | 0.6250 | 0.4222 | 0.0517 | **0.6302** |
| EXPP–SF | 0.4117 | 0.5639 | 0.1505 | 4.9375 | 0.3336 | 0.5164 | 0.0813 | 3.9784 | 0.4676 | 0.5221 | 0.1868 | 3.3913 |
| EXPP–F | 0.7967 | **1** | 0.9912 | 1.9908 | 0.7685 | **1** | 0.9755 | 1.9273 | 0.9354 | 0.9996 | 0.9938 | **1.0820** |
| ARPGDA | 0.8143 | 0.9895 | 0.9947 | **0.6747** | 0.8460 | 0.9895 | 0.9948 | **0.7944** | 0.9494 | **1** | 0.9981 | 1.5184 |
| F-FPCA | 0.9560 | 0.8189 | 1 | 6.3386 | 0.9490 | 0.8133 | 1 | 6.9849 | 0.9781 | 0.9484 | 1 | 4.2289 |

TABLE X: Sparse and/or fair PCA results.

## VI. Acknowledgment

The authors would like to thank Prof. Aritra Konar, Prof. Nikos Sidiropoulos, and Prof. Jiang Bo for sharing their source codes and data sets for the D$k$S, size-constrained clustering, and ONMF problems, respectively.

## Appendix

### A. Some Basic Results

We describe some basic results which will be used in this Appendix. The following lemma encapsulates the proof procedure of Theorems 1, 2 and 5 in a generic form.

**Lemma 5** *Let $\mathcal{X} \subseteq \mathbb{R}^n$ be a compact set. Let $\mathcal{V} \subseteq \mathcal{X}$ a set. Suppose that there exists a mapping $\boldsymbol{y} : \mathcal{X} \to \mathbb{R}^n$, a function $h : \mathcal{X} \to \mathbb{R}$, and a scalar $\delta > 0$ such that, for any $\boldsymbol{x} \in \mathcal{X}$,*

$$\|\boldsymbol{x} - \boldsymbol{y}(\boldsymbol{x})\|_2 \leq h(\boldsymbol{x}), \tag{29}$$

$$h(\boldsymbol{x}) < \delta \quad \Longrightarrow \quad \boldsymbol{y}(\boldsymbol{x}) \in \mathcal{V}. \tag{30}$$

*Then we have the inequality*

$$\mathrm{dist}(\boldsymbol{x}, \mathcal{V}) \leq \max\left\{1, \frac{B}{\delta}\right\} h(\boldsymbol{x}), \quad \forall \boldsymbol{x} \in \mathcal{X}, \tag{31}$$

*where $B$ is any constant such that*

$$B \geq \sup\{\|\boldsymbol{x} - \boldsymbol{z}\|_2 \mid \boldsymbol{x} \in \mathcal{X}, \boldsymbol{z} \in \mathcal{V}\}. \tag{32}$$

*Proof of Lemma 5:* Let $\boldsymbol{x} \in \mathcal{X}$ be given. Suppose that $h(\boldsymbol{x}) \leq \delta$. Then $\mathrm{dist}(\boldsymbol{x}, \mathcal{V}) \leq \|\boldsymbol{x} - \boldsymbol{y}(\boldsymbol{x})\|_2 \leq h(\boldsymbol{x})$. On the other hand, suppose that $h(\boldsymbol{x}) > \delta$. Then $\mathrm{dist}(\boldsymbol{x}, \mathcal{V}) \leq B \leq Bh(\boldsymbol{x})/\delta$. Combining the two cases leads to (31). ■

It should be noted that the error bound analyses in [3], [4] essentially introduced the procedure in Lemma 5 in a case-specific manner for $\mathcal{S}_+^{n,r}$. Our error bound analysis for $\mathcal{U}_{\boldsymbol{\kappa}}^{n,r}$ in Part I of this paper also used this procedure. The following results were previously shown and will be used.

1. unit sphere $\mathcal{S}^n = \{\boldsymbol{x} \in \mathbb{R}^n \mid \|\boldsymbol{x}\|_2 = 1\}$: Let $\boldsymbol{x} \in \mathbb{R}^n$, $\|\boldsymbol{x}\|_2 \leq 1$, be given. Let $\boldsymbol{y} = \Pi_{\mathcal{S}^n}(\boldsymbol{x})$. It holds that
$$\mathrm{dist}(\boldsymbol{x}, \mathcal{S}^n) = \|\boldsymbol{x} - \boldsymbol{y}\|_2 \leq 1 - \|\boldsymbol{x}\|_2 \tag{33}$$
2. unit vector set $\mathcal{U}^n = \{\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n\} \subseteq \mathbb{R}^n$: Let $\boldsymbol{x} \in \Delta^n$ be given. Let $\boldsymbol{y} = \Pi_{\mathcal{U}^n}(\boldsymbol{x})$. It holds that
$$\mathrm{dist}(\boldsymbol{x}, \mathcal{U}^n) = \|\boldsymbol{x} - \boldsymbol{y}\|_2 \leq 2(1 - \|\boldsymbol{x}\|_2^2). \tag{34}$$
3. selection vector set $\mathcal{U}_\kappa^n = \{\boldsymbol{x} \in \{0,1\}^n \mid \mathbf{1}^\top \boldsymbol{x} = \kappa\}$, $\kappa \in \{1, \ldots, n\}$: Let $\boldsymbol{x} \in \mathrm{conv}(\mathcal{U}_\kappa^n)$ be given. Let $\boldsymbol{y} = \Pi_{\mathcal{U}_\kappa^n}(\boldsymbol{x})$. It holds that
$$\mathrm{dist}(\boldsymbol{x}, \mathcal{U}_\kappa^n) = \|\boldsymbol{x} - \boldsymbol{y}\|_2 \leq 2(\kappa - \|\boldsymbol{x}\|_2^2). \tag{35}$$
4. scaled unit vector set $\mathcal{W}^n = \{\boldsymbol{x} \in \mathbb{R}^n \mid \boldsymbol{x} = \alpha \boldsymbol{u}, \alpha \in \mathbb{R}, \boldsymbol{u} \in \mathcal{U}^n\}$: Let $\boldsymbol{x} \in \mathbb{R}^n$ be given. Let $\boldsymbol{y} = \Pi_{\mathcal{W}^n}(\boldsymbol{x}) = x_l \boldsymbol{e}_l$, where $l$ is such that $|x_l| = \|\boldsymbol{x}\|_\infty$. It holds that
$$\mathrm{dist}(\boldsymbol{x}, \mathcal{W}^n) = \|\boldsymbol{x} - \boldsymbol{y}\|_2 \leq \|\boldsymbol{x}\|_1 - \|\boldsymbol{x}\|_\infty. \tag{36}$$

### B. Proof of Theorem 1

Let $\boldsymbol{X} \in \mathbb{R}^{n \times r}$, with $\boldsymbol{x}_j \in \Delta^n$ for all $j$, be given. Let $\boldsymbol{Y} \in \mathbb{R}^{n \times r}$ be given by

$$\boldsymbol{y}_j = \Pi_{\mathcal{U}^n}(\boldsymbol{x}_j), \quad j = 1, \ldots, r. \tag{37}$$

Let

$$h(\boldsymbol{X}) = 2\|\boldsymbol{X}^\top \boldsymbol{X} - \boldsymbol{I}\|_{\ell_1}. \tag{38}$$

We want to apply Lemma 5. We first show the first condition (29) of Lemma 5. Using the error bound result (34) for $\mathcal{U}^n$, the error $\|\boldsymbol{X} - \boldsymbol{Y}\|_{\mathrm{F}}$ is bounded as

$$\|\boldsymbol{X} - \boldsymbol{Y}\|_{\mathrm{F}} \leq \sum_{j=1}^r \|\boldsymbol{x}_j - \boldsymbol{y}_j\|_2 \leq 2(r - \|\boldsymbol{X}\|_{\mathrm{F}}^2).$$

From the proof of Lemma 1 (see Section III-B), it can be seen that $\|\boldsymbol{X}^\top \boldsymbol{X} - \boldsymbol{I}\|_{\ell_1} \geq r - \|\boldsymbol{X}\|_{\mathrm{F}}^2$. It follows that the choice in (37)–(38) satisfies the first condition (29) of Lemma 5.

Second we show that the second condition (30) of Lemma 5 is satisfied for some $\delta$. Consider the following lemma.

**Lemma 6** *Let $\boldsymbol{Y} \in \{0,1\}^{n \times r}$ be a matrix with $\boldsymbol{y}_j \in \{\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n\}$ for all $j$, and with $n \geq r$. The singular values of $\boldsymbol{Y}$ satisfy $\sigma_i(\boldsymbol{Y})^2 \in \{0, 1, \ldots, r\}$ for all $i$ and $\sum_{i=1}^r \sigma_i(\boldsymbol{Y})^2 = r$.*

*Proof of Lemma 6:* Represent each $\boldsymbol{y}_j$ by $\boldsymbol{y}_j = \boldsymbol{e}_{l_j}$ for some $l_j \in \{1,\ldots,n\}$. We have

$$\boldsymbol{Y}\boldsymbol{Y}^\top = \textstyle\sum_{j=1}^r \boldsymbol{e}_{l_j}\boldsymbol{e}_{l_j}^\top = \mathrm{Diag}(\boldsymbol{d}),$$

where $d_i \in \{0,1,\ldots,r\}$ counts the number of times $\boldsymbol{e}_i$ appears in $\boldsymbol{y}_1,\ldots,\boldsymbol{y}_r$. As a basic SVD result, the above equation implies that $\sigma_i(\boldsymbol{Y})^2 = d_{[i]}$ for all $i$. Since $\sum_{i=1}^r d_i = r$, it follows that $\sum_{i=1}^r \sigma_i(\boldsymbol{Y})^2 = r$. ■

Lemma 6 implies that, if $\sigma_i(\boldsymbol{Y}) < \sqrt{2}$ for all $i$, then $\sigma_1(\boldsymbol{Y}) = \cdots = \sigma_r(\boldsymbol{Y}) = 1$ and consequently $\boldsymbol{Y}$ lies in $\mathcal{U}^{n,r}$. Suppose that $h(\boldsymbol{X}) < \delta$ for some $\delta > 0$. By the Weyl inequality,

$$\begin{aligned}\sigma_i(\boldsymbol{Y}) &\le \sigma_i(\boldsymbol{X}) + \sigma_1(\boldsymbol{Y}-\boldsymbol{X}) \le \sigma_i(\boldsymbol{X}) + \|\boldsymbol{Y}-\boldsymbol{X}\|_{\mathrm{F}} \\ &\le \sigma_i(\boldsymbol{X}) + h(\boldsymbol{X}) \\ &< \sigma_i(\boldsymbol{X}) + \delta. \end{aligned} \tag{39}$$

Also, it is seen from (38) that

$$\begin{aligned}\tfrac{1}{2}h(\boldsymbol{X}) &\ge \|\boldsymbol{X}^\top\boldsymbol{X} - \boldsymbol{I}\|_{\mathrm{F}} = \|\boldsymbol{\sigma}(\boldsymbol{X})^2 - \boldsymbol{1}\|_2 \\ &\ge |\sigma_i(\boldsymbol{X})^2 - 1|,\end{aligned}$$

for any $i$. The above equation implies $\sigma_i(\boldsymbol{X}) < \sqrt{1+\delta/2}$. Applying this to (39) yields

$$\sigma_i(\boldsymbol{Y}) < \sqrt{1+\delta/2} + \delta \le 1 + \delta/2 + \delta.$$

By choosing $\delta = 2(\sqrt{2}-1)/3$ such that $\sigma_i(\boldsymbol{Y}) < \sqrt{2}$ for all $i$, the matrix $\boldsymbol{Y}$ must lie in $\mathcal{U}^{n,r}$. Hence, the second condition (30) of Lemma 5 is satisfied for the choice of $h$ and $\boldsymbol{y}$ in (37)–(38) and for $\delta = 2(\sqrt{2}-1)/3$.

With the two conditions of Lemma 5 satisfied, we are ready to apply Lemma 5. It can be verified that $B = \sqrt{2r}$ satisfies the requirement (32) in Lemma 5 (see (36) in Part I of this paper). Applying the above $h$, $\delta$ and $B$ to (31) in Lemma 5, we get

$$\mathrm{dist}(\boldsymbol{X},\mathcal{U}^{n,r}) \le \frac{2\sqrt{2}}{2(\sqrt{2}-1)/3}\sqrt{r}\|\boldsymbol{X}^\top\boldsymbol{X} - \boldsymbol{I}\|_{\ell_1}.$$

The above inequality leads to the final result (10a); note $(2\sqrt{2})/[2(\sqrt{2}-1)/3] = 10.2426$. The accompanied result (10b) is obtained by applying Lemma 1 to (10a).

### C. Proof of Theorem 2

Assume $\kappa_1 \ge \cdots \ge \kappa_r$ without loss of generality. Let $\boldsymbol{X} \in \mathbb{R}^{n\times r}$, with $\boldsymbol{x}_j \in \mathrm{conv}(\mathcal{U}^n_{\kappa_j})$ for all $j$, be given. Let $\boldsymbol{Y} \in \mathbb{R}^{n\times r}$ be given by

$$\boldsymbol{y}_j = \Pi_{\mathcal{U}^n_{\kappa_j}}(\boldsymbol{x}_j), \quad j = 1,\ldots,r. \tag{40}$$

Let

$$h(\boldsymbol{X}) = 2\|\boldsymbol{X}^\top\boldsymbol{X} - \mathrm{Diag}(\boldsymbol{\kappa})\|_{\ell_1}. \tag{41}$$

Once again we apply Lemma 5. Following the same proof as Theorem 1, it can be shown that

$$\|\boldsymbol{X}-\boldsymbol{Y}\|_{\mathrm{F}} \le 2(\boldsymbol{1}^\top\boldsymbol{\kappa} - \|\boldsymbol{X}\|_{\mathrm{F}}^2) \le h(\boldsymbol{X}),$$

where the first inequality is due to the error bound result (35) for $\mathcal{U}^n_\kappa$; the second inequality can be seen from the proof of Lemma 1. The first condition (29) of Lemma 5 is thereby satisfied. The second condition (30) of Lemma 5 is more challenging to show. Consider the following lemma.

**Lemma 7** *Let $\boldsymbol{Y} \in \{0,1\}^{n\times r}$ be a matrix with $\boldsymbol{y}_j \in \mathcal{U}^n_{\kappa_j}$ for all $j$, and with $n \ge r$. If $\sum_{i=1}^r \sigma_i(\boldsymbol{Y})^4 < \boldsymbol{1}^\top(\boldsymbol{\kappa}^2) + 2$, then it must be true that $\boldsymbol{y}_i^\top\boldsymbol{y}_j = 0$ for all $i \ne j$. Consequently, the matrix $\boldsymbol{Y}$ lies in $\mathcal{U}^{n,r}_{\boldsymbol{\kappa}}$.*

*Proof of Lemma 7:* Consider the matrix product $\boldsymbol{Y}^\top\boldsymbol{Y}$. Let $i,j \in \{1,\ldots,r\}$, $i \ne j$, be any two distinct indices. On the one hand,

$$\begin{aligned}\|\boldsymbol{Y}^\top\boldsymbol{Y}\|_{\mathrm{F}}^2 &\ge \textstyle\sum_{l=1}^r \|\boldsymbol{y}_l\|_2^4 + 2(\boldsymbol{y}_i^\top\boldsymbol{y}_j)^2 \\ &= \boldsymbol{1}^\top(\boldsymbol{\kappa}^2) + 2(\boldsymbol{y}_i^\top\boldsymbol{y}_j)^2.\end{aligned}$$

On the other hand,

$$\|\boldsymbol{Y}^\top\boldsymbol{Y}\|_{\mathrm{F}}^2 = \|\boldsymbol{\sigma}(\boldsymbol{Y})^2\|_2^2 = \textstyle\sum_{i=1}^r \sigma_i(\boldsymbol{Y})^4.$$

The above two equations imply that

$$(\boldsymbol{y}_i^\top\boldsymbol{y}_j)^2 \le \tfrac{1}{2}[\textstyle\sum_{i=1}^r \sigma_i(\boldsymbol{Y})^4 - \boldsymbol{1}^\top(\boldsymbol{\kappa}^2)].$$

Since $\boldsymbol{Y} \in \{0,1\}^{n\times r}$, we have $\boldsymbol{y}_i^\top\boldsymbol{y}_j \in \{0,1,\ldots,n\}$. If $\sum_{i=1}^r \sigma_i(\boldsymbol{Y})^4 < \boldsymbol{1}^\top(\boldsymbol{\kappa}^2)+2$, then we are left with $\boldsymbol{y}_i^\top\boldsymbol{y}_j = 0$. The proof is complete. ■

Suppose that $h(\boldsymbol{X}) < \delta$ for some $\delta > 0$. By the same proof as before (cf. (39)),

$$\sigma_i(\boldsymbol{Y}) < \sigma_i(\boldsymbol{X}) + \delta. \tag{42}$$

Also, from (41),

$$\begin{aligned}\tfrac{1}{2}h(\boldsymbol{X}) &\ge \|\boldsymbol{X}^\top\boldsymbol{X} - \mathrm{Diag}(\boldsymbol{\kappa})\|_{\mathrm{F}} \ge \|\boldsymbol{\sigma}(\boldsymbol{X})^2 - \boldsymbol{\kappa}\|_2 \\ &\ge |\sigma_i(\boldsymbol{X})^2 - \kappa_i|,\end{aligned}$$

for any $i$. Here, the second inequality is due to the von Neumann trace inequality result $\|\boldsymbol{A}-\boldsymbol{B}\|_{\mathrm{F}} \ge \|\boldsymbol{\sigma}(\boldsymbol{A}) - \boldsymbol{\sigma}(\boldsymbol{B})\|_2$. The above equation implies $\sigma_i(\boldsymbol{X}) < \sqrt{\kappa_i + \delta/2}$. Putting it into (42) leads to

$$\begin{aligned}\sigma_i(\boldsymbol{Y})^2 &< \kappa_i + \delta/2 + 2\delta\sqrt{\kappa_i+\delta/2} + \delta^2 \\ &\le \kappa_i + \delta/2 + 2\delta[\sqrt{\kappa_i} + \delta/(2\sqrt{\kappa_i})] + \delta^2 \\ &= \kappa_i + (1/2 + 2\sqrt{\kappa_i})\delta + (1 + 1/\sqrt{\kappa_i})\delta^2, \end{aligned} \tag{43}$$

where the second equation is due to the inequality $\sqrt{a+b} \le \sqrt{a} + b/\sqrt{a}$ for $a > 0$ and $b \ge 0$. Suppose that

$$(1 + 1/\sqrt{\kappa_i})^{\frac{1}{2}}\delta \le 1, \quad \forall i. \tag{44}$$

Eq. (43) can be further bounded as

$$\sigma_i(\boldsymbol{Y})^2 < \kappa_i + \underbrace{[1/2 + 2\sqrt{\kappa_i} + (1+1/\sqrt{\kappa_i})^{\frac{1}{2}}]}_{:=\alpha_i}\delta, \tag{45}$$

Taking square on (45) gives

$$\sigma_i(\boldsymbol{Y})^4 < \kappa_i^2 + 2\kappa_i\alpha_i\delta + \alpha_i^2\delta^2. \tag{46}$$

Suppose that

$$\alpha_i\delta \le 1, \quad \forall i. \tag{47}$$

Then we further bound (46) as

$$\sigma_i(\boldsymbol{Y})^4 < \kappa_i^2 + \underbrace{(2\kappa_i\alpha_i + \alpha_i)}_{:=\beta_i}\delta, \tag{48}$$

and consequently,

$$\sum_{i=1}^r \sigma_i(\boldsymbol{Y})^4 < \mathbf{1}^\top(\boldsymbol{\kappa}^2) + (\mathbf{1}^\top\boldsymbol{\beta})\delta.$$

Let us choose

$$\delta = 2/(\mathbf{1}^\top\boldsymbol{\beta}).$$

It can be verified that this choice of $\delta$ satisfies the requirements (44) and (47). By invoking Lemma 7, we see that $\boldsymbol{Y}$ lies in $\mathcal{U}^{n,r}_{\boldsymbol{\kappa}}$. Hence we have the second condition (30) of Lemma 5 satisfied.

Finally we assemble the components together to obtain the error bound. Let $B = \sqrt{2\mathbf{1}^\top\boldsymbol{\kappa}}$, which can be verified to satisfy (32). We express $\beta_i$ as

$$\begin{aligned}\beta_i &= (1+2\kappa_i)[1/2 + 2\sqrt{\kappa_i} + \underbrace{(1+1/\sqrt{\kappa_i})^{\frac{1}{2}}}_{\leq\sqrt{2}\leq 1.5}] \\ &\leq 2(1+2\kappa_i)(1+\sqrt{\kappa_i}).\end{aligned}$$

We have

$$\frac{B}{\delta} \leq \sqrt{2\mathbf{1}^\top\boldsymbol{\kappa}}\underbrace{(\textstyle\sum_{j=1}^r (1+2\kappa_j)(1+\sqrt{\kappa_j}))}_{:=\gamma}.$$

Note that the right-hand side of the above equation is greater than 1. Applying the above $h$, $\delta$ and $B$ to (31) in Lemma 5 leads to

$$\operatorname{dist}(\boldsymbol{X}, \mathcal{U}^{n,r}) \leq 2\sqrt{2\mathbf{1}^\top\boldsymbol{\kappa}}\gamma\|\boldsymbol{X}^\top\boldsymbol{X} - \operatorname{Diag}(\boldsymbol{\kappa})\|_{\ell_1},$$

and consequently the final result (15a). In addition, applying Lemma 1 to (15a) gives (15b).

### D. Proof of Theorem 5

Let $\boldsymbol{X} \in \mathbb{R}_+^{n\times r}$, with $\|\boldsymbol{x}_j\|_2 \leq 1$ for all $j$, be given. Let $\boldsymbol{W} \in \mathbb{R}^{n\times r}$ be given by $\bar{\boldsymbol{w}}_i = x_{i,l_i}\boldsymbol{e}_{l_i}$ for all $i$, where $l_i$ is such that $x_{i,l_i} = \max\{x_{i,1}, \ldots, x_{i,r}\}$. Let $\boldsymbol{Y} \in \mathbb{R}^{n\times r}$ be given by

$$\boldsymbol{y}_j = \begin{cases} \boldsymbol{w}_j/\|\boldsymbol{w}_j\|_2, & \boldsymbol{w}_j \neq \mathbf{0} \\ \boldsymbol{u}, & \boldsymbol{w}_j = \mathbf{0} \end{cases}$$

for all $j$, where $\boldsymbol{u}$ denotes any non-negative unit $\ell_2$ norm vector. It is worth noting that

$$\bar{\boldsymbol{w}}_i = \Pi_{\mathcal{W}^r}(\bar{\boldsymbol{x}}_i), \quad \boldsymbol{y}_j = \Pi_{\mathcal{S}^n}(\boldsymbol{w}_j),$$

for all $i, j$. Let

$$h(\boldsymbol{X}) = \sqrt{6}\left[\sum_{j=1}^r c_1(\boldsymbol{x}_j)^2 + \sum_{i=1}^n \rho_1(\bar{\boldsymbol{x}}_i)^2\right]^{\frac{1}{2}}, \quad (49)$$

where $c_1(\boldsymbol{x}) = 1 - \|\boldsymbol{x}\|_2$, $\rho_1(\boldsymbol{x}) = \|\boldsymbol{x}\|_1 - \|\boldsymbol{x}\|_\infty$. As before we apply Lemma 5. We begin by showing the first condition (29) of Lemma 5. Using the error bounds (36) and (33) for $\mathcal{W}^r$ and $\mathcal{S}^n$, respectively, it holds that

$$\|\bar{\boldsymbol{x}}_i - \bar{\boldsymbol{w}}_i\|_2 \leq \rho_1(\bar{\boldsymbol{x}}_i), \quad (50)$$

$$\begin{aligned}\|\boldsymbol{w}_j - \boldsymbol{y}_j\|_2 &\leq 1 - \|\boldsymbol{w}_j\|_2 \\ &\leq 1 - \|\boldsymbol{x}_j\|_2 + \|\boldsymbol{w}_j - \boldsymbol{x}_j\|_2 \quad (51) \\ &\leq \sqrt{2}\left[c_1(\boldsymbol{x}_j)^2 + \|\boldsymbol{w}_j - \boldsymbol{x}_j\|_2^2\right]^{\frac{1}{2}}, \quad (52)\end{aligned}$$

where (51) is due to the triangle inequality; (52) is due to $(|a|+|b|)^2 \leq 2(|a|^2+|b|^2)$. This gives rise to

$$\begin{aligned}\|\boldsymbol{X} - \boldsymbol{Y}\|_{\mathrm{F}} &\leq \|\boldsymbol{X} - \boldsymbol{W}\|_{\mathrm{F}} + \|\boldsymbol{W} - \boldsymbol{Y}\|_{\mathrm{F}} \\ &\leq \sqrt{2}\left(\|\boldsymbol{X} - \boldsymbol{W}\|_{\mathrm{F}}^2 + \|\boldsymbol{W} - \boldsymbol{Y}\|_{\mathrm{F}}^2\right)^{\frac{1}{2}} \\ &\leq \sqrt{2}\left[3\|\boldsymbol{X} - \boldsymbol{W}\|_{\mathrm{F}}^2 + 2\textstyle\sum_{j=1}^r c_1(\boldsymbol{x}_j)^2\right]^{\frac{1}{2}} \\ &\leq \sqrt{6}\left[\textstyle\sum_{i=1}^n \rho_1(\bar{\boldsymbol{x}}_i)^2 + \sum_{j=1}^r c_1(\boldsymbol{x}_j)^2\right]^{\frac{1}{2}} \\ &= h(\boldsymbol{X}), \quad (53)\end{aligned}$$

where the third inequality is due to (52); the fourth inequality is due to (50). The first condition (29) of Lemma 5 is obtained.

The proof of the second condition (30) of Lemma 5 is as follows. From the way $\boldsymbol{Y}$ is constructed, we notice that $\boldsymbol{Y}$ is a non-negative semi-orthogonal matrix if $\|\boldsymbol{w}_j\|_2 > 0$ for all $j$. Suppose that $h(\boldsymbol{X}) < \delta$ for some $\delta > 0$. As a reverse version of (53), we have, for any $j$,

$$\begin{aligned}\delta &> \sqrt{6}\left[\|\boldsymbol{X} - \boldsymbol{W}\|_{\mathrm{F}}^2 + \sum_{j=1}^r c_1(\boldsymbol{x}_j)^2\right]^{\frac{1}{2}} \\ &\geq \sqrt{6}\left[\|\boldsymbol{x}_j - \boldsymbol{w}_j\|_2^2 + c_1(\boldsymbol{x}_j)^2\right]^{\frac{1}{2}} \\ &\geq \sqrt{3}\left[\|\boldsymbol{x}_j - \boldsymbol{w}_j\|_2 + c_1(\boldsymbol{x}_j)\right] \\ &= \sqrt{3}\left(\|\boldsymbol{x}_j - \boldsymbol{w}_j\|_2 + 1 - \|\boldsymbol{x}_j\|_2\right) \\ &\geq \sqrt{3}\left(1 - \|\boldsymbol{w}_j\|_2\right).\end{aligned}$$

Here, the third inequality is due to $(|a|+|b|)^2 \leq 2(|a|^2+|b|^2)$ and $c_1(\boldsymbol{x}_j) \geq 0$ for any $\|\boldsymbol{x}_j\|_2 \leq 1$; the last inequality is due to the triangle inequality. The above equation implies that, if $\delta = \sqrt{3}$, then $\|\boldsymbol{w}_j\|_2 > 0$. The second condition (30) of Lemma 5 is shown.

Finally we put together the components. We choose $B = \sqrt{2r}$, which can be verified to satisfy (32). The inequality (31) associated with the above $\delta$, $h$ and $B$ is

$$\operatorname{dist}(\boldsymbol{X}, \mathcal{S}_+^{n,r}) \leq \nu \underbrace{\left[\sum_{j=1}^r c_1(\boldsymbol{x}_j)^2 + \sum_{i=1}^n \rho_1(\bar{\boldsymbol{x}}_i)^2\right]^{\frac{1}{2}}}_{=\psi_2(\boldsymbol{X})},$$

where $\nu = \max\{\sqrt{6}, 2\sqrt{r}\}$. Furthermore, as a basic norm result, we have $\psi_2(\boldsymbol{X}) \leq \sum_{j=1}^r c_1(\boldsymbol{x}_j) + \sum_{i=1}^n \rho_1(\bar{\boldsymbol{x}}_i) = \psi_1(\boldsymbol{X})$. The proof of Theorem 5 is complete.

### E. Relationship of the Proof of Theorem 5 with Prior Work

A key step of the proof of Theorem 5 in Appendix D is the construction of $\boldsymbol{W}$ and $\boldsymbol{Y}$. This step is identical to that in the prior studies [3], [4]. The important difference lies in the bound in (50). We use the error bound for scaled unit vectors, given in Lemma 2, to derive the bound in (50). The prior studies used another bound: For any $\boldsymbol{X} \in \mathbb{R}_+^{n\times r}$, it holds that

$$\|\boldsymbol{X} - \boldsymbol{W}\|_{\mathrm{F}}^2 \leq \sum_{j=1}^r \sum_{l=1, l\neq j}^r \boldsymbol{x}_j^\top\boldsymbol{x}_l = \|\boldsymbol{X}\mathbf{1}\|_2^2 - \|\boldsymbol{X}\|_{\mathrm{F}}^2; \quad (54)$$

see [3, Lemma 3.1] and [4, Lemma 7]. An oversimplified way to understand the prior studies is as follows: replace our bound

(50) with (54), and then proceed with the error bound proof in Appendix D. There are differences with the fine details and the reader is referred to the prior studies [3], [4] for the details. Our bound (50) may be seen as an improvement over (54). From (50) and the proof of Lemma 4, we have

$$\|\boldsymbol{X}-\boldsymbol{W}\|_{\mathrm{F}}^2 \leq \sum_{i=1}^{n} \rho_1(\bar{\boldsymbol{x}}_i)^2 \leq \|\boldsymbol{X}\boldsymbol{1}\|_2^2 - \|\boldsymbol{X}\|_{\mathrm{F}}^2.$$

## REFERENCES


[1] J. Liu, Y. Liu, W.-K. Ma, M. Shao, and A. M.-C. So, "Extreme point pursuit–Part I: A framework for constant modulus optimization," submitted to *IEEE Trans. Signal Process.*, 2024.

[2] S. Wang, T.-H. Chang, Y. Cui, and J.-S. Pang, "Clustering by orthogonal NMF model and non-convex penalty optimization," *IEEE Trans. Signal Process.*, vol. 69, pp. 5273–5288, 2021.

[3] B. Jiang, X. Meng, Z. Wen, and X. Chen, "An exact penalty approach for optimization with nonnegative orthogonality constraints," *Math Program.*, vol. 198, no. 1, pp. 855–897, 2023.

[4] X. Chen, Y. He, and Z. Zhang, "Tight error bounds for the sign-constrained Stiefel manifold," *arXiv preprint arXiv:2210.05164*, 2022.

[5] B. Jiang, Y.-F. Liu, and Z. Wen, "$\ell_p$-norm regularization algorithms for optimization over permutation matrices," *SIAM J. Optim.*, vol. 26, no. 4, pp. 2284–2313, 2016.

[6] M. Zaslavskiy, F. Bach, and J.-P. Vert, "A path following algorithm for the graph matching problem," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 12, pp. 2227–2242, 2008.

[7] L. F. McGinnis, "Implementation and testing of a primal-dual algorithm for the assignment problem," *Oper. Res.*, vol. 31, no. 2, pp. 277–291, 1983.

[8] L. Condat, "Fast projection onto the simplex and the $\ell_1$ ball," *Math. Program.*, vol. 158, no. 1-2, pp. 575–585, 2016.

[9] N. D. Sidiropoulos and E. Tsakonas, "Signal processing and optimization tools for conference review and session assignment," *IEEE Signal Process. Mag.*, vol. 32, no. 3, pp. 141–155, 2015.

[10] A. Konar and N. D. Sidiropoulos, "Exploring the subgraph density-size trade-off via the Lovaśz extension," in *Proc. Int. Conf. Web Search Data Min.*, pp. 743–751, 2021.

[11] F. Pompili, N. Gillis, P.-A. Absil, and F. Glineur, "Two algorithms for orthogonal nonnegative matrix factorization with application to clustering," *Neurocomput.*, vol. 141, pp. 15–25, 2014.

[12] Y. Xu and W. Yin, "A block coordinate descent method for regularized multiconvex optimization with applications to nonnegative tensor factorization and completion," *J. Imag. Sci.*, vol. 6, no. 3, pp. 1758–1789, 2013.

[13] M. Shao, Q. Li, W.-K. Ma, and A. M.-C. So, "A framework for one-bit and constant-envelope precoding over multiuser massive MISO channels," *IEEE Trans. Signal Process.*, vol. 67, no. 20, pp. 5309–5324, 2019.

[14] M. Shao and W.-K. Ma, "Binary MIMO detection via homotopy optimization and its deep adaptation," *IEEE Trans. Signal Process.*, vol. 69, pp. 781–796, 2020.

[15] A. Beck, *First-order Methods in Optimization*. SIAM, 2017.

[16] W.-K. Ma, P.-C. Ching, and Z. Ding, "Semidefinite relaxation based multiuser detection for M-ary PSK multiuser systems," *IEEE Trans. Signal Process.*, vol. 52, no. 10, pp. 2862–2872, 2004.

[17] C. Jeon, R. Ghods, A. Maleki, and C. Studer, "Optimal data detection in large MIMO," *arXiv preprint arXiv:1811.01917*, 2018.

[18] S. Rangan, P. Schniter, A. K. Fletcher, and S. Sarkar, "On the convergence of approximate message passing with arbitrary matrices," *IEEE Trans. Inf. Theory*, vol. 65, no. 9, pp. 5339–5351, 2019.

[19] S. L. Loyka, "Channel capacity of MIMO architecture using the exponential correlation matrix," *IEEE Commun. Lett.*, vol. 5, no. 9, pp. 369–371, 2001.

[20] U. Feige, D. Peleg, and G. Kortsarz, "The dense $k$-subgraph problem," *Algorithm.*, vol. 29, pp. 410–421, 2001.

[21] X.-T. Yuan and T. Zhang, "Truncated power method for sparse eigenvalue problems.," *J. Mach. Learn. Res.*, vol. 14, no. 4, 2013.

[22] D. Papailiopoulos, I. Mitliagkas, A. Dimakis, and C. Caramanis, "Finding dense subgraphs via low-rank bilinear optimization," in *Proc. Int. Conf. Mach. Learn.*, pp. 1890–1898, 2014.

[23] J. Leskovec and A. Krevl, "SNAP Datasets: Stanford large network dataset collection," 2014.

[24] Z.-Y. Liu and H. Qiao, "GNCCP—graduated nonconvexity and concavity procedure," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 36, no. 6, pp. 1258–1267, 2013.

[25] Y. Xia, "An efficient continuation method for quadratic assignment problems," *Comput. Oper. Res.*, vol. 37, no. 6, pp. 1027–1032, 2010.

[26] F. Zhou and F. De la Torre, "Factorized graph matching," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 38, no. 9, pp. 1774–1789, 2015.

[27] "CMU/VASC Image Database." http://vasc.ri.cmu.edu//idb/html/motion/house/index.html.

[28] M. Leordeanu, R. Sukthankar, and M. Hebert, "Unsupervised learning for graph matching," *Int. J. Comput. Vis.*, vol. 96, pp. 28–45, 2012.

[29] H. Aanæs, R. R. Jensen, G. Vogiatzis, E. Tola, and A. B. Dahl, "Large-scale data for multiple-view stereopsis," *Int. J. Comput. Vis.*, vol. 120, pp. 153–168, 2016.

[30] E. Tola, V. Lepetit, and P. Fua, "DAISY: An efficient dense descriptor applied to wide-baseline stereo," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 5, pp. 815–830, 2009.

[31] Y. Liu, X. Gong, J. Chen, S. Chen, and Y. Yang, "Rotation-invariant siamese network for low-altitude remote-sensing image registration," *IEEE J. Sel. Top. Appl. Earth. Obs. Remote. Sens.*, vol. 13, pp. 5746–5758, 2020.

[32] S. Zhu, D. Wang, and T. Li, "Data clustering with size constraints," *Knowl. Based Syst.*, vol. 23, no. 8, pp. 883–889, 2010.

[33] N. Ganganath, C.-T. Cheng, and K. T. Chi, "Data clustering with cluster size constraints using a modified $k$-means algorithm," in *Proc. Int. Conf. Cyber-Enabled Distrib. Comput. Knowl. Discovery*, pp. 158–161, 2014.

[34] H. Xiao, K. Rasul, and R. Vollgraf, "Fashion-MNIST: a novel image dataset for benchmarking machine learning algorithms," *arXiv preprint arXiv:1708.07747*, 2017.

[35] R. A. Fisher, "Iris." UCI Machine Learning Repository, 1988.

[36] B. German, "Glass Identification." UCI Machine Learning Repository, 1987.

[37] R. Siegler, "Balance Scale." UCI Machine Learning Repository, 1994.

[38] J. P. Boyle and R. L. Dykstra, "A method for finding projections onto the intersection of convex sets in Hilbert spaces," in *Lecture Notes Statist.*, pp. 28–47, Springer, 1986.

[39] C. Boutsidis and E. Gallopoulos, "SVD based initialization: A head start for nonnegative matrix factorization," *Pattern Recognit.*, vol. 41, no. 4, pp. 1350–1362, 2008.

[40] D. Cai, X. He, J. Han, and H.-J. Zhang, "Orthogonal Laplacianfaces for face recognition," *IEEE Trans. Image Process.*, vol. 15, no. 11, pp. 3608–3614, 2006.

[41] F. Zhu, Y. Wang, B. Fan, S. Xiang, G. Meng, and C. Pan, "Spectral unmixing via data-guided sparsity," *IEEE Trans. Image Process.*, vol. 23, no. 12, pp. 5412–5427, 2014.

[42] D. Cai, X. He, and J. Han, "Locally consistent concept factorization for document clustering," *IEEE Trans. Knowl. Data Eng.*, vol. 23, no. 6, pp. 902–913, 2010.

[43] K. Benidis, Y. Sun, P. Babu, and D. P. Palomar, "Orthogonal sparse PCA and covariance estimation via Procrustes reformulation," *IEEE Trans. Signal Process.*, vol. 64, no. 23, pp. 6211–6226, 2016.

[44] M. Journée, Y. Nesterov, P. Richtárik, and R. Sepulchre, "Generalized power method for sparse principal component analysis.," *J. Mach. Learn. Res.*, vol. 11, no. 2, pp. 517–553, 2010.

[45] M. Xu, B. Jiang, W. Pu, Y.-F. Liu, and A. M.-C. So, "An efficient alternating Riemannian/projected gradient descent ascent algorithm for fair principal component analysis," in *Proc. Int. Conf. Acoust. Speech Signal Process.*, pp. 7195–7199, IEEE, 2024.

[46] G. Zalcberg and A. Wiesel, "Fair principal component analysis and filter design," *IEEE Trans. Signal Process.*, vol. 69, pp. 4835–4842, 2021.

[47] Y. LeCun, C. Cortes, and C. Burges, "MNIST handwritten digit database," *ATT Labs [Online]. Available: http://yann.lecun.com/exdb/mnist*, vol. 2, 2010.

[48] A. Krizhevsky, G. Hinton, *et al.*, "Learning multiple layers of features from tiny images," *MIT and NYU, Tech. Rep.*, 2009.

**Junbin Liu** received the B.S. degree from the South China University of Technology, Guangzhou, China, in 2017 and M.S. degree from the University of Chinese Academy of Sciences in 2020. He is currently pursuing the Ph.D. degree with the Department of Electronic Engineering, the Chinese University of Hong Kong, under the supervision of Professor Wing-Kin Ma.

His research interests encompass statistical signal processing methods, optimization theories, and their wide-ranging applications.

**Ya Liu** received the B.S. degree from the University of Electronic Science and Technology of China (UESTC), Chengdu, China, in 2018. She is currently pursuing a Ph.D. degree with the Department of Electronic Engineering at the Chinese University of Hong Kong (CUHK), under the supervision of Professor Wing-Kin Ma.

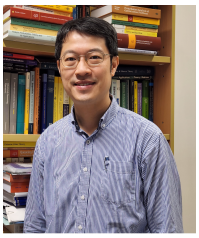
Her research focuses include matrix factorization, signal processing, and optimization, along with their diverse applications.

**Wing-Kin Ma** (Fellow, IEEE) is currently a Professor with the Department of Electronic Engineering, The Chinese University of Hong Kong (CUHK), Hong Kong. His research interests include signal processing, optimization and communications, with recent focus on i) optimization and statistical aspects with structured matrix factorization, with application to remote sensing and data science; and ii) coarsely quantized MIMO transceiver designs.

Dr. Ma has rich experience in editorial service, such as an Associate Editor, and then later, a Senior Area Editor, and then from 2021 to 2023, the Editor-in-Chief of the IEEE TRANSACTIONS ON SIGNAL PROCESSING, and many others. He was a Tutorial Speaker in EUSIPCO 2011 and ICASSP 2014, and was an IEEE Signal Processing Society (SPS) Distinguished Lecturer in 2018–2019. He was the recipient of the Research Excellence Award 2013–2014 by CUHK, the 2015 IEEE Signal Processing Magazine Best Paper Award, the 2016 IEEE Signal Processing Letters Best Paper Award, and the 2018 IEEE SPS Best Paper Award. He served as a Member of the Signal Processing for Communications and Networking Technical Committee (SPCOM-TC) in 2015–2020, a Member of Signal Processing Theory and Methods Technical Committee (SPTM-TC) in 2012–2017, the SPS Regional Director-at-Large for Region 10 in 2020–2021, and a Technical Program Co-Chair of ICASSP 2023. He co-founded and co-organized One World Signal Processing in 2020, a virtual seminar series for signal processing.

**Mingjie Shao** (S'16-M'20) received the B.S. degree from the Xidian University, Xi'an, China, in 2015 and Ph.D. degree from the Chinese University of Hong Kong (CUHK) in 2020. He was a Postdoctoral Fellow with the Department of Electronic Engineering, CUHK from 2020 to 2023. He is currently a Research Professor (Qilu Young Scholar) in the School of Information Science and Engineering, Shandong University, Qingdao, China. He was the recipient of the Hong Kong PhD Fellowship Scheme (HKPFS) from August 2015. He was listed in the Student Best Paper Finalists in ICASSP 2017.

His research interests focus on convex and non-convex optimization, statistical signal processing and machine learning for wireless communication.



**Anthony Man-Cho So** (Fellow, IEEE) received the B.S.E. degree in computer science from Princeton University, Princeton, NJ, USA, with minors in applied and computational mathematics, engineering and management systems, and German language and culture; the M.Sc. degree in computer science and the Ph.D. degree in computer science with a Ph.D. minor in mathematics from Stanford University, Stanford, CA, USA. He is currently Dean of the Graduate School, Deputy Master of Morningside College, and a Professor with the Department of Systems Engineering and Engineering Management at The Chinese University of Hong Kong (CUHK), Hong Kong SAR, China. His research interests include optimization theory and its applications in various areas of science and engineering, including computational geometry, machine learning, signal processing, and statistics.

Dr. So is a Fellow of the Hong Kong Institution of Engineers. He is the recipient of a number of research and teaching awards, including the 2024 INFORMS Computing Society Prize, the SIAM Review SIGEST Award in 2024, the 2022 University Grants Committee Teaching Award, the 2018 IEEE Signal Processing Society Best Paper Award, the 2015 IEEE Signal Processing Society Signal Processing Magazine Best Paper Award, the 2014 IEEE Communications Society Asia-Pacific Outstanding Paper Award, the 2013 CUHK Vice-Chancellor's Exemplary Teaching Award, and the 2010 INFORMS Optimization Society Optimization Prize for Young Researchers. He currently serves on the Editorial Boards of *Journal of Global Optimization*, *Mathematics of Operations Research*, *Mathematical Programming*, *Open Journal of Mathematical Optimization*, *Optimization Methods and Software*, and *SIAM Journal on Optimization*. He was also the Lead Guest Editor of the Special Issue on Non-Convex Optimization for Signal Processing and Machine Learning of the IEEE SIGNAL PROCESSING MAGAZINE and a Guest Editor of the Special Issue on Advanced Optimization Theory and Algorithms for Next Generation Wireless Communication Networks of the IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS.