SIAM J. OPTIM.
Vol. 32, No. 3, pp. 2180–2207

ORTHOGONAL TRACE-SUM MAXIMIZATION: TIGHTNESS OF THE SEMIDEFINITE RELAXATION AND GUARANTEE OF LOCALLY OPTIMAL SOLUTIONS*

JOONG-HO WON†, TENG ZHANG‡, AND HUA ZHOU§

Abstract. This paper studies an optimization problem on the sum of traces of matrix quadratic forms in m semiorthogonal matrices, which can be considered as a generalization of the synchronization of rotations. While the problem is nonconvex, this paper shows that its semidefinite programming relaxation solves the original nonconvex problems exactly with high probability under an additive noise model with small noise in the order of $O(m^{1/4})$. In addition, it shows that, with high probability, the sufficient condition for global optimality considered in Won, Zhou, and Lange [*SIAM J. Matrix Anal. Appl.*, 2 (2021), pp. 859–882] is also necessary under a similar small noise condition. These results can be considered as a generalization of existing results on phase synchronization.

Key words. semidefinite programming, tightness of convex relaxation, estimation error, locally optimal solutions

MSC codes. 68Q25, 68R10, 68U05

DOI. 10.1137/21M1422707

1. Introduction.

This paper considers the orthogonal trace-sum maximization (OTSM) problem [35] of estimating m matrices $\boldsymbol{O}_1, \ldots, \boldsymbol{O}_m$ with $\boldsymbol{O}_i \in \mathbb{R}^{d_i \times r}$ from the optimization problem:

$$\text{(OTSM)} \quad \text{maximize} \sum_{1\le i,j\le m} \operatorname{tr}(\boldsymbol{O}_i^T \boldsymbol{S}_{ij} \boldsymbol{O}_j) \ \text{ subject to } \ \boldsymbol{O}_i \in \mathcal{O}_{d_i,r},\ i = 1, \ldots, m,$$

where $\boldsymbol{S}_{ij} = \boldsymbol{S}_{ji}^T \in \mathbb{R}^{d_i \times d_j}$ for $i, j = 1, \ldots, m$, $r \le \min_{i=1,\ldots,m} d_i$, and $\mathcal{O}_{d,r} = \{O \in \mathbb{R}^{d\times r} : \boldsymbol{O}^T \boldsymbol{O} = \mathbf{I}_r\}$ is the Stiefel manifold of semiorthogonal matrices; $\mathbf{I}_r$ denotes the identity matrix of order r.

The OTSM problem has applications in generalized canonical correlation analysis (CCA) [18] and Procrustes analysis [17, 30]. Furthermore, if $d_1 = \cdots = d_m = r$, then (OTSM) reduces to the problem of synchronization of rotations [5], which has wide applications in multireference alignment [4], cryogenic electron microscopy (cryo-EM) [29, 36], 2D/3D point set registration [19, 12, 9], and multiview structure from motion [2, 3, 32].

1.1. Related works.

While the OTSM problem is proposed recently in [35], it is closely related to many well-studied problems. In particular, its special cases have been studied in the name of angular synchronization, which can be considered

*Received by the editors June 1, 2021; accepted for publication (in revised form) June 8, 2022; published electronically August 30, 2022.
https://doi.org/10.1137/21M1422707
Funding: JW was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (2019R1A2C1007126). TZ was partly supported by the NSF grants CNS-1818500. HZ was partly supported by NIH grants HG006139, GM141798 and NSF grants DMS-2054253 and IIS-2205441.
†Department of Statistics, Seoul National University, Seoul 08826, Korea (wonj@stats.snu.ac.kr).
‡Department of Mathematics, University of Central Florida, Orlando, FL 32816 USA (Teng.Zhang@ucf.edu).
§Departments of Biostatistics and Computational Medicine, University of California, Los Angeles, CA 90095 USA (huazhou@ucla.edu).

as a special case of (OTSM) in the complex-valued setting, $\mathbb{Z}_2$ synchronization, and synchronization of rotations. The OTSM problem itself can also be considered as a special case of the group synchronization problem.

Angular synchronization. The complex-valued OTSM problem with $d_1 = \cdots = d_m = 1$ is equivalent to a problem called angular synchronization or phase synchronization, which estimates angles $\theta_1, \ldots, \theta_m \in [0, 2\pi)$ from the observation of relative offsets $(\theta_i - \theta_j)$ mod 2π. The problem has applications in cryo-EM [28], comparative biology [16], and many others. To address this problem, Singer [28] formulates the problem as a nonconvex optimization problem

$$\max_{\boldsymbol{x} \in \mathbb{C}^m} \boldsymbol{x}^* \boldsymbol{C} \boldsymbol{x} \text{ subject to } |x_1| = \cdots = |x_m| = 1, \tag{1.1}$$

where $x_k = e^{i\theta_k}$ for all $1 \leq k \leq m$. In fact, (1.1) can be considered as the special case of (OTSM) when $d_1 = \cdots = d_m = r = 2$.

The angular synchronization problem (1.1) has been studied extensively. For example, Singer [28] proposes two methods, by eigenvectors and semidefinite programming, respectively. The performance of the method is analyzed using random matrix theory and information theory. In [4], Bandeira, Boumal, and Singer assume the model $\boldsymbol{C} = \boldsymbol{z}\boldsymbol{z}^* + \sigma \boldsymbol{W}$, where $\boldsymbol{z} \in \mathbb{C}^m$ satisfies $|z_1| = \cdots = |z_m| = 1$ and $\boldsymbol{W} \in \mathbb{C}^{m \times m}$ is a Hermitian Gaussian Wigner matrix, and show that if $\sigma \leq \frac{1}{18} m^{\frac{1}{4}}$, then the solution of the semidefinite programming approach is also the solution to (1.1) with high probability. Using a more involved argument and a modified power method, Zhong and Boumal [37] improve the bound in [4] to $\sigma = O(\sqrt{\frac{m}{\log m}})$.

There is another line of works that solves (1.1) using power methods. In particular, Boumal [6] investigates a modified power method and shows that the method converges to the solution of (1.1) when $\sigma = O(m^{\frac{1}{6}})$; Liu, Yue, and Man-Cho So [24] investigate another generalized power method and prove the convergence for $\sigma = O(m^{\frac{1}{4}})$; and Zhong and Boumal [37] improve the rate to $\sigma = O(\sqrt{\frac{m}{\log m}})$.

There are some other interesting works for the angular synchronization problem that are not based on semidefinite programming or power method. [23] assumes that the pairwise differences are only observed over a graph, studies the landscape of a proposed objective function, and shows that the global minimizer is unique when the associated graph is incomplete and follows the Erdös–Rényi model. [27] proposes an approximate message passing (AMP) algorithm and analyzes its behavior by identifying phases where the problem is easy, computationally hard, and statistically impossible.

$\mathbb{Z}_2$ synchronization. The real-valued OTSM problem with $d_1 = \cdots = d_m = 1$ is called the $\mathbb{Z}_2$-synchronization problem [11] for $\mathbb{Z}_2 = \{1, -1\}$. For this problem, [14] shows that the solution of the semidefinite programming method matches the minimax lower bound on the optimal Bayes error rate for the original problem (1.1).

Synchronization of rotations. The OTSM problem with $d_1 = \cdots = d_m = r > 2$ is called "synchronization of rotations" in some literature. This special case has wide applications in graph realization and point cloud registration, multiview structure from motion [2, 3, 32], common lines in cryo-EM [29], orthogonal least squares [36], and 2D/3D point set registration [19]. [8] studies the problem from the perspective of manifold optimization and derives the Cramér–Rao bounds, which are the lower bounds of the variance of any unbiased estimator. [31] proposes a distributed algorithm with theoretical guarantees on convergence. [33] discusses a method to make the estimator in (OTSM) more robust to outlying observations. Another robust algorithm based on the maximum likelihood estimator is proposed in [7]. As for the theoretical properties, [5] analyzes a semidefinite program approach that solves the problem approximately and studies its approximation ratio. [25] investigates a generalized power method for this problem. A recent manuscript [22] follows the line of [4, 6, 24, 37] and proves that the original problem and the relaxed problem have the same solution when $\sigma \leq O(\frac{\sqrt{m}}{d+\sqrt{d}\log m})$.

Group synchronization. The OTSM problem can also be considered as a special case of the group synchronization problem, which recovers a vector of elements in a group, given noisy pairwise measurements of the relative elements $g_u g_v^{-1}$. The OTSM problem is the special case when the group is $\mathcal{O}_{d,r}$, the set of orthogonal matrices. [1] studies the properties of weak recovery when the elements are from a generic compact group and the underlying graph of pairwise observations is the d-dimensional grid. [27] proposes an AMP algorithm for solving synchronization problems over a class of compact groups. [26] generates the estimation from compact groups to the class of Cartan motion groups, which includes the important special case of rigid motions by applying the compactification process. [10] assumes that the measurement graph is sparse and there are corrupted observations and shows that minimax recovery rate depends almost exclusively on the edge sparsity of the measurement graph irrespective of other graphical metrics.

1.2. Our contribution.

The main contribution of this work is the study of the OTSM problem under an additive noise model. The main results are threefold: First, we propose a semidefinite programming approach for solving (OTSM) and show that it solves (OTSM) exactly when the size of noise is bounded. Second, we show that, under a similar bounded noise condition, the sufficient condition for global optimality of a critical point, studied in [35], is in fact necessary and sufficient. Finally, these noise boundedness conditions are satisfied with high probability under Gaussianity. These results can be considered as a generalization of [4] from angular synchronization to the OTSM problem.

2. The OTSM problem.

2.1. Model assumption.

In this work, we assume the MAXBET model of generating $\boldsymbol{S}_{ij}$, which postulates the existence of $\{\boldsymbol{\Theta}_i\}_{1\leq i\leq m}$ and $\{\boldsymbol{W}_{ij}\}_{1\leq i\neq j\leq m}$ such that $\boldsymbol{\Theta}_i \in \mathcal{O}_{d_i,r}$ for all $1 \leq i \leq m$, and

$$\text{(MAXBET)} \qquad \boldsymbol{S}_{ij} = \boldsymbol{\Theta}_i\boldsymbol{\Theta}_j^T + \boldsymbol{W}_{ij}, \text{ where } \boldsymbol{W}_{ij} = \boldsymbol{W}_{ji}^T \text{ for all } 1 \leq i,j \leq m.$$

In this model, $\boldsymbol{\Theta}_i\boldsymbol{\Theta}_j^T$ is considered as the "clean measurement of relative elements," and $\boldsymbol{W}_{ij}$ is considered as an additive noise. This is a natural model for the generalized CCA in [35]. Consider a latent variable model in which a latent variable $\boldsymbol{z} \in \mathbb{R}^r$ has zero mean and covariance matrix $\mathbf{I}_r$, and an observation in the ith group is given by $\boldsymbol{a}_i = \boldsymbol{\Theta}_i\boldsymbol{z} + \boldsymbol{\epsilon}_i \in \mathbb{R}^{d_i}$, $i = 1, \ldots, m$, with the noise $\boldsymbol{\epsilon}_i$ uncorrelated with $\boldsymbol{z}$ and $\boldsymbol{\epsilon}_j$, $j \neq i$. If the noise covariance is $\tau\mathbf{I}_{d_i}$, then the auto-covariance of group i is $\boldsymbol{\Sigma}_{ii} + \tau\mathbf{I}_{d_i}$. The (population) cross-covariance matrix between groups i and j is $\boldsymbol{\Sigma}_{ij} = \boldsymbol{\Theta}_i\boldsymbol{\Theta}_j^T$. The generalized CCA [30, 35] seeks (semi)orthogonal matrices $\{\boldsymbol{O}_i \in \mathcal{O}_{d_i,r}\}$ such that the expected inner product between matrices $\boldsymbol{O}_i^T\boldsymbol{a}_i$ and $\boldsymbol{O}_j^T\boldsymbol{a}_j$ is summed and maximized for each pair (i,j), which is $\sum_{i,j}\operatorname{tr}(\boldsymbol{O}_i^T\boldsymbol{\Sigma}_{ij}\boldsymbol{O}_j)$. Also note that $\mathbb{E}[\langle \boldsymbol{O}_i^T\boldsymbol{a}_i, \boldsymbol{O}_i^T\boldsymbol{a}_i\rangle] = \operatorname{tr}(\boldsymbol{O}_i^T\boldsymbol{\Sigma}_{ii}\boldsymbol{O}_i) + \text{const}$. If we assume that $\{\boldsymbol{\Theta}_i\}$ is (semi)orthogonal, then this problem is precisely (OTSM), and the forthcoming Proposition 2.1 shows that the population version of this generalized CCA recovers precisely the transformations $\{\boldsymbol{\Theta}_i\}$ of the latent variable $\boldsymbol{z}$. Now let us turn to the practical setting. Appealing to the law of large numbers, the sample estimate of $\boldsymbol{\Sigma}_{ij}$ can then be written as $\boldsymbol{S}_{ij} = \boldsymbol{\Sigma}_{ij} + \boldsymbol{W}_{ij} = \boldsymbol{\Theta}_i \boldsymbol{\Theta}_j^T + \boldsymbol{W}_{ij}$. A statistical interest is whether $\{\boldsymbol{\Theta}_i\}$ can be precisely estimated by solving the sample version of (OTSM). Model (MAXBET) is also standard for synchronization problems, such as synchronization of rotations [33, 8] and group synchronization [1, 27].

In some applications [30, 18], it is also natural to assume the MAXDIFF model that ignores the auto-covariance terms:

$$\text{(MAXDIFF)} \qquad \boldsymbol{S}_{ii} = \boldsymbol{0} \quad \text{and} \quad \boldsymbol{S}_{ij} = \boldsymbol{\Theta}_i \boldsymbol{\Theta}_j^T + \boldsymbol{W}_{ij},\ i \neq j.$$

In this work, we will present our main results based on the MAXBET model and discuss the MAXDIFF model in the remarks.

When there is no noise in either the MAXBET or MAXDIFF model, setting $\boldsymbol{O}_i = \boldsymbol{\Theta}_i$, $i = 1, \dots, m$, solves problem (OTSM). The proof is deferred to section 5.1.

PROPOSITION 2.1. *In the noiseless case ($\boldsymbol{W}_{ij} = \boldsymbol{0}$ for all i, j), $(\boldsymbol{O}_1, \dots, \boldsymbol{O}_m) = (\boldsymbol{\Theta}_1, \dots, \boldsymbol{\Theta}_m)$ globally solves* (OTSM) *under the model* (MAXBET) *or* (MAXDIFF).

However, in the presence of noise, Proposition 2.1 does not hold, and problem (OTSM) is difficult to solve. To establish theoretical guarantees for the noisy setting, we investigate two approaches; one is based on semidefinite programming, and the other one is based on finding locally optimal solutions of (OTSM).

2.2. Approach 1: Semidefinite programming relaxation. While the problem (OTSM) is nonconvex and difficult to solve, we can relax it to a convex optimization problem based on semidefinite programming that can be solved efficiently. In fact, semidefinite programming–based approaches have been proposed and analyzed for the problem of angular synchronization [28, 4, 37] and synchronization of rotations [5], and our proposed method can be considered as a generalization of these existing methods.

The argument of the relaxation is as follows. Let $D = \sum_{i=1}^m d_i$,

$$\text{(2.1)} \qquad \boldsymbol{S} = \begin{bmatrix} \boldsymbol{S}_{11} & \boldsymbol{S}_{12} & \dots & \boldsymbol{S}_{1m} \\ \boldsymbol{S}_{21} & \boldsymbol{S}_{22} & & \boldsymbol{S}_{2m} \\ \vdots & & \ddots & \vdots \\ \boldsymbol{S}_{m1} & \boldsymbol{S}_{m2} & \cdots & \boldsymbol{S}_{mm} \end{bmatrix} \in \mathbb{R}^{D \times D}, \text{ and } \boldsymbol{O} = \begin{bmatrix} \boldsymbol{O}_1 \\ \vdots \\ \boldsymbol{O}_m \end{bmatrix} \in \mathbb{R}^{D \times r};$$

then by setting $\boldsymbol{U} = \boldsymbol{O}\boldsymbol{O}^T$, the problem (OTSM) is equivalent to finding

$$\text{(2.2)} \quad \tilde{\boldsymbol{U}} = \arg\max\{\operatorname{tr}(\boldsymbol{S}\boldsymbol{U}) : \boldsymbol{U} \succcurlyeq \boldsymbol{0},\ \operatorname{rank}(\boldsymbol{U}) = r,\ \boldsymbol{U}_{ii} \preccurlyeq \mathbf{I},\ \operatorname{tr}(\boldsymbol{U}_{ii}) = r,\ i = 1, \dots, m\}$$

for $\boldsymbol{U} \in \mathbb{R}^{D \times D}$ such that $\boldsymbol{U} = \boldsymbol{U}^T$, which can be relaxed to solving

$$\text{(SDP)} \qquad \max_{\boldsymbol{U} \in \mathbb{R}^{D \times D}, \boldsymbol{U} = \boldsymbol{U}^T} \langle \boldsymbol{S}, \boldsymbol{U} \rangle \ \text{ subject to } \boldsymbol{U} \succcurlyeq \boldsymbol{0}, \boldsymbol{U}_{ii} \preccurlyeq \mathbf{I}, \operatorname{tr}(\boldsymbol{U}_{ii}) = r,$$

where $\boldsymbol{M} \succcurlyeq \boldsymbol{0}$ (resp., $\boldsymbol{M} \preccurlyeq \boldsymbol{0}$) means that a matrix $\boldsymbol{M}$ is positive (resp., negative) semidefinite. If a solution $\hat{\boldsymbol{U}}$ to problem (SDP) has rank-r, then we can set $\tilde{\boldsymbol{U}} = \hat{\boldsymbol{U}}$, which can be decomposed to $\hat{\boldsymbol{U}} = \tilde{\boldsymbol{V}}\tilde{\boldsymbol{V}}^T$, where $\tilde{\boldsymbol{V}} \in \mathbb{R}^{D \times r}$. Write $\tilde{\boldsymbol{V}} = [\tilde{\boldsymbol{V}}_1^T, \dots, \tilde{\boldsymbol{V}}_m^T]^T$; then $\tilde{\boldsymbol{V}}_i \in \mathcal{O}_{d_i,r}$ for all $1 \leq i \leq m$ and $(\tilde{\boldsymbol{V}}_1, \dots, \tilde{\boldsymbol{V}}_m)$ *globally* solves problem (OTSM).

This work shows that if the noises $\boldsymbol{W}_{ij}$ are "small," then the solutions of problems (OTSM) and (SDP) are equivalent in the sense that $\hat{\boldsymbol{U}} = \tilde{\boldsymbol{V}}\tilde{\boldsymbol{V}}^T$ with $\tilde{\boldsymbol{V}}$ rank-r; hence the convex relaxation is tight. Furthermore, each $\tilde{\boldsymbol{V}}_i$ converges to $\boldsymbol{\Theta}_i$ as $m \to \infty$, as desired for CCA applications.

2.3. Approach 2: Characterization of critical points. While the semidefinite programming (SDP) approach is convex and can be solved with high accuracy, it has a large computational cost when D is large, and solving the original nonconvex problem (OTSM) without lifting the variable (from $\boldsymbol{O}$ to $\boldsymbol{U}$) is more efficient. A natural question is, Is there any guarantee on whether a critical point of problem (OTSM), which local nonconvex optimization algorithms usually deliver, is globally optimal?

Using the optimality conditions for the convex relaxation (SDP), Won, Zhou, and Lange [35] study sufficient conditions for a critical point of problem (OTSM) to be globally optimal. Specifically, the first-order necessary condition for local optimality of (OTSM) is

$$\boldsymbol{O}_i\boldsymbol{\Lambda}_i = \sum_{j=1}^{m} \boldsymbol{S}_{ij}\boldsymbol{O}_j, \quad i = 1, \ldots, m, \tag{2.3}$$

for some symmetric matrix $\boldsymbol{\Lambda}_i$. The latter matrix is the Lagrange multiplier associated with the constraint $\boldsymbol{O}_i \in \mathcal{O}_{d_i,r}$ and has a representation $\boldsymbol{\Lambda}_i = \sum_{j=1}^{m} \boldsymbol{O}_i^T\boldsymbol{S}_{ij}\boldsymbol{O}_j$. In what follows, a critical point is defined as $(\boldsymbol{O}_1, \ldots, \boldsymbol{O}_m)$ with $\boldsymbol{O}_i \in \mathcal{O}_{d_i,r}$ satisfying (2.3). If τ_i is the smallest eigenvalue of $\boldsymbol{\Lambda}_i$, then a critical point is a global optimum of (OTSM) if

(2.4)

$$\boldsymbol{L}(\boldsymbol{O}, \boldsymbol{\Lambda}) \succcurlyeq \boldsymbol{0}, \quad \text{where } \boldsymbol{O} = [\boldsymbol{O}_1^T, \ldots, \boldsymbol{O}_m^T]^T,\ \boldsymbol{\Lambda} = (\boldsymbol{\Lambda}_1, \ldots, \boldsymbol{\Lambda}_m), \text{ and}$$

$$\boldsymbol{L}(\boldsymbol{O}, \boldsymbol{\Lambda}) = \begin{bmatrix} \boldsymbol{O}_1\boldsymbol{\Lambda}_1\boldsymbol{O}_1^T + \tau_1(\mathbf{I}_{d_1} - \boldsymbol{O}_1\boldsymbol{O}_1^T) & & \\ & \ddots & \\ & & \boldsymbol{O}_m\boldsymbol{\Lambda}_m\boldsymbol{O}_m^T + \tau_m(\mathbf{I}_{d_m} - \boldsymbol{O}_m\boldsymbol{O}_m^T) \end{bmatrix} - \boldsymbol{S}.$$

A block relaxation-type algorithm that converges to a critical point is also proposed in [35]. However, characterization of such a point that does *not* satisfy condition (2.4) has remained an open question.

This paper shows that, if the noises $\boldsymbol{W}_{ij}$ are "small" in a similar sense to that of Approach 1, the sufficient condition (2.4) is also necessary for global optimality. Thus, under this regime we can fully determine whether or not a critical point, which can be found by a simple local algorithm, is globally optimal. Furthermore, each $\boldsymbol{O}_i$ converges to $\boldsymbol{\Theta}_i$ as $m \to \infty$, up to a common phase shift, as desired for CCA applications.

2.4. Notation. This work sometimes divides a matrix $\boldsymbol{X}$ of size $D \times D$ into m^2 submatrices such that the (i, j) block is a $d_i \times d_j$ submatrix. We use $\boldsymbol{X}_{ij}$ or $[\boldsymbol{X}]_{ij}$ to denote this submatrix. Similarly, sometimes we divide a matrix of $\boldsymbol{Y} \in \mathbb{R}^{D\times r}$ or a vector $\boldsymbol{y} \in \mathbb{R}^D$ into m submatrices or an m vector, where the ith component, denoted by $\boldsymbol{Y}_i$, $[\boldsymbol{Y}]_i$ or $\boldsymbol{y}_i$, $[\boldsymbol{y}]_i$, is a matrix of size $d_i \times r$ or a vector of length d_i.

For any matrix $\boldsymbol{X}$, we use $\|\boldsymbol{X}\|$ to represent its operator norm and $\|\boldsymbol{X}\|_F$ to represent its Frobenius norm. In addition, $\boldsymbol{P}_{\boldsymbol{X}}$ represents an orthonormal matrix whose column space is the same as $\boldsymbol{X}$, $\boldsymbol{P}_{\boldsymbol{X}^\perp}$ is an orthonormal matrix whose column space is the orthogonal complement of the column space of $\boldsymbol{X}$, $\Pi_{\boldsymbol{X}} = \boldsymbol{P}_{\boldsymbol{X}}\boldsymbol{P}_{\boldsymbol{X}}^T$ is the projector to the column space of $\boldsymbol{X}$, and $\Pi_{\boldsymbol{X}^\perp}$ is the projection matrix to the orthogonal complement of the column space of $\boldsymbol{X}$. If $\boldsymbol{Y} \in \mathbb{R}^{n\times n}$ is symmetric, we use $\lambda_1(\boldsymbol{Y}) \geq \lambda_2(\boldsymbol{Y}) \geq \cdots \geq \lambda_n(\boldsymbol{Y})$ to denote its eigenvalues in descending order.

3. Main results. In this section, we present our main results. The first main result, Theorem 3.1, shows that if the noises $\boldsymbol{W}_{ij}$ are "small," then the convex relaxation in (SDP) solves the original problem (OTSM) exactly. The second main result, Theorem 3.9, shows that if the noises $\boldsymbol{W}_{ij}$ are "small," then a critical point is globally optimal if and only if condition (2.4) holds.

3.1. Theoretical guarantees on the SDP approach. This section provides conditions that if the noises $\boldsymbol{W}_{ij}$ are "small," then the solution of problem (SDP) has rank-r and is equivalent to the solution of the problem (OTSM) in the sense that $\hat{\boldsymbol{U}} = \tilde{\boldsymbol{V}}\tilde{\boldsymbol{V}}^T$ with $\tilde{\boldsymbol{V}}$ rank-r; hence the convex relaxation is tight.

We begin with two deterministic conditions on $\boldsymbol{W}$ in Theorem 3.1 and Corollary 3.3, with showing that the condition holds with high probability under a Gaussian model in Corollary 3.4, and with a statement on the consistency of the solution in Corollary 3.7. The statement of the first deterministic theorem is as follows.

THEOREM 3.1. *If* $m \geq \|\boldsymbol{W}\|(4\sqrt{r}+1)+1$ *and* $\boldsymbol{W} \in \mathbb{R}^{D\times D}$ *is small in the sense that*

$$m > 4m\frac{2\left(\max_{1\leq i\leq m}\|[\boldsymbol{W\Theta}]_i\|_F + 4\|\boldsymbol{W}\|^2\sqrt{\frac{r}{m}}\right)}{m-\|\boldsymbol{W}\|(4\sqrt{r}+1)-1}$$
$$+2\left(\max_{1\leq i\leq m}\|[\boldsymbol{W\Theta}]_i\|_F + 4\|\boldsymbol{W}\|^2\sqrt{\frac{r}{m}}\right) + 8\|\boldsymbol{W}\|\sqrt{\frac{r}{m}} + 2\|\boldsymbol{W}\|, \tag{3.1}$$

then the solutions of (OTSM) *and relaxation* (SDP) *are equivalent in the sense that a solution* $\hat{\boldsymbol{U}}$ *to* (SDP) *also solves* (2.2).

The proof of Theorem 3.1 will be presented in section 4.1. While the condition (3.1) is rather complicated, we expect that it holds for large m when $\|\boldsymbol{W}\|$ and $\max_{i=1,\dots,m}\|(\boldsymbol{W\Theta})_i\|_F$ grow slowly as m increases. To prove this idea rigorously, we introduce the notion of $\boldsymbol{\Theta}$-discordant noise, which is inspired by the notion of "z-discordant matrix" in [4, Definition 3.1].

DEFINITION 3.2 ($\boldsymbol{\Theta}$-discordance). *Let* $\boldsymbol{\Theta} = (\boldsymbol{\Theta}_1,\dots,\boldsymbol{\Theta}_m) \in \times_{i=1}^m \mathcal{O}_{d_i,r}$. *Recall* $D = \sum_{i=1} d_i$. *A matrix* $\boldsymbol{W}$ *is said to be* $\boldsymbol{\Theta}$*-discordant if it is symmetric and satisfies* $\|\boldsymbol{W}\| \leq 3\sqrt{D}$ *and* $\max_{i=1,\dots,m}\|[\boldsymbol{W\Theta}]_i\|_F \leq 3\sqrt{Dr\log m}$.

Based on the definition of $\boldsymbol{\Theta}$-discordant noise, The next corollary is a deterministic, nonasymptotic statement that simplifies the condition (3.1) in Theorem 3.1. Its proof is deferred to section 4.2.

COROLLARY 3.3. *Let* $d = D/m$. *If* $m \geq 8$ *and* $\sigma^{-1}\boldsymbol{W}$ *is* $\boldsymbol{\Theta}$*-discordant for*

$$\sigma \leq \frac{m^{1/4}}{60\sqrt{dr}}, \tag{3.2}$$

then condition (3.1) *holds, and the solutions of* (OTSM) *and* (SDP) *are equivalent.*

Next, we apply a natural probabilistic model and investigate the $\boldsymbol{\Theta}$-discordant property. In particular, we follow [6, 4, 37] and use an additive Gaussian noise model to generate the symmetric noise matrix $\boldsymbol{W}$:

$$\text{Upper triangular part of } W \in \mathbb{R}^{D\times D} \text{ is elementwisely}$$
$$\text{independent and identically distributed (i.i.d.) sampled from } N(0,\sigma^2). \tag{3.3}$$

For this model, we have the following corollary that shows if $\sigma \leq O(\frac{m^{1/4}}{\sqrt{dr}})$, then (3.1) holds with high probability. Its proof is deferred to section 4.3.

COROLLARY 3.4. *Assume the additive Gaussian noise model in* (3.3), $m \geq 3$ *or* $m \geq 2$ *and* $\min_{i=1}^m d_i \geq 6$*; then with probability at least* $1 - 1/m - 2\exp(-\frac{(3-2\sqrt{2})^2}{4}D)$, $\boldsymbol{W}$ *satisfies the* $\boldsymbol{\Theta}$*-discordant property.*

As a result, if $\sigma \leq \frac{m^{1/4}}{60\sqrt{dr}}$ *and* $m \geq 8$*, then with the same probability, the condition* (3.1) *holds, and the solutions of* (OTSM) *and* (SDP) *are equivalent.*

Remark 3.5. The assumption $m \geq 8$ in Corollary 3.3 can be relaxed but with a different constant factor in the upper bound of σ in (3.2). For example, if $m \geq 3$ is assumed, then we need $\sigma \leq \frac{m^{1/4}}{124\sqrt{dr}}$.

Remark 3.6. The result in this section can be naturally adapted to the MAXDIFF model. The main intermediate results for the proof of Theorem 3.1 given in section 4.1, including Lemma 4.1 and Lemma 4.2, still hold with $\boldsymbol{W}_{ii} = 0$. While the estimations in Lemma 4.3 do not hold, following the steps given at the end of section 5.2.1, we are still able to obtain similar bounds on the difference between $\tilde{\boldsymbol{V}}$ and $\boldsymbol{\Theta}$. In summary, we are able to obtain parallel results to Theorem 3.1 and Corollary 3.3 for the MAXDIFF setting. In particular, if $\boldsymbol{W}$ is generated using the model in Corollary 3.3, then the solutions of (OTSM) and (SDP) with the MAXDIFF model are equivalent with probability at least $1 - 1/m - 2\exp(-\frac{(3-2\sqrt{2})^2}{4}D)$ if $\sigma \leq \frac{m^{1/4}}{120\sqrt{dr}}$ and $m \geq 10$. This more restrictive bound under the MAXDIFF model is expected since (MAXDIFF) utilizes *less* information on the clean signal $\boldsymbol{\Theta}$ for the same number of measurements.

Following the proof of Theorem 3.1, we have a consistency result, i.e., that the solution of (SDP) recovers the true signal $\boldsymbol{\Theta}$ if m is sufficiently large.

COROLLARY 3.7. *Assuming the conditions in Corollary* 3.3*, then the solution of* (SDP)*,* $\tilde{\boldsymbol{U}}$*, admits a decomposition* $\tilde{\boldsymbol{U}} = \tilde{\boldsymbol{V}}\tilde{\boldsymbol{V}}^T$ *with* $\tilde{\boldsymbol{V}} \in \mathbb{R}^{D\times r}$ *such that*

$$\max_{i=1,\dots,m} \|\tilde{\boldsymbol{V}}_i - \boldsymbol{\Theta}_i\|_F \leq \frac{2\left(3\sigma\sqrt{dmr\log m} + 36\sigma^2 d\sqrt{rm}\right)}{m - 3\sigma\sqrt{dm}(4\sqrt{r}+1) - 1}. \tag{3.4}$$

Thus, if $\sigma = o(\frac{m^{1/4}}{\sqrt{dr}})$*, then* $\max_{i=1,\dots,m} \|\tilde{\boldsymbol{V}}_i - \boldsymbol{\Theta}_i\|_F \to 0$ *as* $m \to \infty$*.*

Remark 3.8. For the MAXDIFF model, (3.4) is replaced with

$$\max_{i=1,\dots,m} \|\tilde{\boldsymbol{V}}_i - \boldsymbol{\Theta}_i\|_F \leq \frac{6\sigma\sqrt{dmr\log m} + \frac{72\sigma^2 dm\sqrt{rm}}{m-2}}{m - \frac{12\sigma\sqrt{dm^3 r}}{m-2}}.$$

The bound follows from the discussion of Lemma 4.3 in the MAXDIFF setting. If $\sigma = o(\frac{m^{1/4}}{\sqrt{dr}})$, then $\max_{i=1,\dots,m} \|\tilde{\boldsymbol{V}}_i - \boldsymbol{\Theta}_i\|_F \to 0$ as $m \to \infty$.

3.2. Theoretical guarantees on critical points. This section presents the condition on the size of the noise $\boldsymbol{W}$ that ensures condition (2.4) holds for any globally optimal point $(\boldsymbol{O}_1, \dots, \boldsymbol{O}_m)$ and its associated Lagrange multipliers $(\boldsymbol{\Lambda}_1, \dots, \boldsymbol{\Lambda}_m)$. We begin with two deterministic conditions on $\boldsymbol{W}$ in Theorem 3.9 and Corollary 3.10, show that the condition in Corollary 3.10 holds with high probability under the additive Gaussian model (3.3) in Corollary 3.11, and establish the consistency in Corollary 3.14.

Recall that the first-order necessary condition for local optimality of (OTSM) is given in (2.3). The associated Lagrange multiplier is symmetric:
(3.5)

$$\boldsymbol{\Lambda}_i = \boldsymbol{O}_i^T \left(\sum_{j=1}^m \boldsymbol{S}_{ij} \boldsymbol{O}_j \right) = \left(\sum_{j=1}^m \boldsymbol{S}_{ij} \boldsymbol{O}_j \right)^T \boldsymbol{O}_i = \frac{1}{2} \sum_{j=1}^m \boldsymbol{O}_i^T \boldsymbol{S}_{ij} \boldsymbol{O}_j + \frac{1}{2} \sum_{j=1}^m \boldsymbol{O}_j^T \boldsymbol{S}_{ji} \boldsymbol{O}_i.$$

It is also known that a necessary condition for global optimality of a critical point is that the $\boldsymbol{\Lambda}_i$ in (3.5) is symmetric and positive semidefinite for all i [35, Proposition 3.1]. Note this result does not imply condition (2.4). The first deterministic result implying condition (2.4) is given in the following.

THEOREM 3.9. *Suppose noise $\boldsymbol{W}$ is small in the sense that*

$$\begin{aligned} m \geq \|\boldsymbol{W}\|(4\sqrt{r}+1) + \max_{1 \leq i \leq m} \|[\boldsymbol{W}\boldsymbol{\Theta}]_i\|_F + 4\|\boldsymbol{W}\|^2 \sqrt{\frac{r}{m}} \\ + \frac{2m(\max_{1 \leq i \leq m} \|[\boldsymbol{W}\boldsymbol{\Theta}]_i\|_F + 4\|\boldsymbol{W}\|^2 \sqrt{\frac{r}{m}})}{m - 4\|\boldsymbol{W}\|\sqrt{r}} + 16\|\boldsymbol{W}\|^2 \frac{r}{m}. \end{aligned} \tag{3.6}$$

If $(\boldsymbol{O}_1, \ldots, \boldsymbol{O}_m)$ *is a global optimum of* (OTSM)*, then* $(\boldsymbol{O}_1, \ldots, \boldsymbol{O}_m)$ *and its associated Lagrange multipliers* $(\boldsymbol{\Lambda}_1, \ldots, \boldsymbol{\Lambda}_m)$ *satisfy condition* (2.4)*.*

The proof of this theorem is deferred to section 4.5. Theorem 3.9 implies that, under the small noise regime quantified by inequality (3.6), condition (2.4) is *necessary and sufficient* for global optimality.

The following corollary is a deterministic, nonasymptotic statement that simplifies condition (3.6) using the notion of $\boldsymbol{\Theta}$-discordance (Definition 3.2). The idea is similar to (3.1). The left-hand side of condition (3.6) dominates the right-hand side (RHS) as $m \to \infty$ if $\|\boldsymbol{W}\|$ and $\max_{i=1,\ldots,m} \|(\boldsymbol{W}\boldsymbol{\Theta})_i\|_F$ are bounded or increase slowly as m increases. Thus, we can expect that inequality (3.6) is satisfied if noise variance σ is small and the number of observations m is large.

COROLLARY 3.10. *Let $d = D/m$. Suppose that $m \geq 2$,*

$$\sigma \leq \frac{m^{1/4}}{31\sqrt{dr}}, \tag{3.7}$$

and $\sigma^{-1}\boldsymbol{W}$ is $\boldsymbol{\Theta}$-discordant; then (3.6) *holds. Thus if* $(\boldsymbol{O}_1, \ldots, \boldsymbol{O}_m)$ *is a global optimum of* (OTSM)*, then* $(\boldsymbol{O}_1, \ldots, \boldsymbol{O}_m)$ *and its associated Lagrange multipliers* $(\boldsymbol{\Lambda}_1, \ldots, \boldsymbol{\Lambda}_m)$ *satisfy condition* (2.4)*.*

The proof is deferred to section 4.6.

Finally, since Corollary 3.4 shows that $\boldsymbol{W}$ in the additive Gaussian noise model (3.3) is $\boldsymbol{\Theta}$-discordant after scaling by σ, Corollary 3.10 implies the following result on the probabilistic model.

COROLLARY 3.11. *Suppose the additive Gaussian noise model in* (3.3) *holds. If* $\sigma \leq \frac{m^{1/4}}{31\sqrt{rd}}$ *and* $m \geq 3$ *or* $m \geq 2$ *and* $\min_{i=1,\ldots,m} d_i \geq 6$*, then with probability at least* $1 - 1/m - 2\exp(-\frac{(3-2\sqrt{2})^2}{4}D)$*, any global optimum* $(\boldsymbol{O}_1, \ldots, \boldsymbol{O}_m)$ *of* (OTSM) *and its associated Lagrange multipliers* $(\boldsymbol{\Lambda}_1, \ldots, \boldsymbol{\Lambda}_m)$ *satisfy condition* (2.4)*.*

Remark 3.12. The upper bound of σ in the RHS of (3.7) can be made smaller if m increases. For example, if we have $m \geq 4$, then (3.7) can be relaxed to $\sigma \leq \frac{m^{1/4}}{29\sqrt{dr}}$; if $m \geq 9$, $\sigma \leq \frac{m^{1/4}}{26\sqrt{dr}}$ suffices.

Remark 3.13. If instead the MAXDIFF model is assumed, the present analysis holds for $m \geq 4$ and (3.7) replaced with $\sigma \leq \frac{m^{1/4}}{64\sqrt{dr}}$. This is a worse bound as opposed to $\frac{m^{1/4}}{29\sqrt{dr}}$ for (MAXBET) (See Remark 3.12). To obtain the same bound as (3.7), we need $m \geq 9$; see section 4.7. Similar to the SDP relaxation, the more restrictive bound in the MAXDIFF model is expected since (MAXDIFF) utilizes less information on the clean signal $\mathbf{\Theta}$ for the same number of measurements.

The following consistency result is a by-product of the proof of Theorem 3.9. Recall that problem (OTSM) is invariant to "simultaneous rotation," i.e., postmultiplying a fully orthogonal matrix $\boldsymbol{Q} \in \mathcal{O}_{r,r}$ to $\boldsymbol{O}_i$'s (see, e.g., [34, equation (8.2)]).

COROLLARY 3.14. *Let* $(\boldsymbol{O}_1, \ldots, \boldsymbol{O}_m) \in \times_{i=1}^m \mathcal{O}_{d_i,r}$ *be a global optimum of* (OTSM). *If the noise* $\sigma^{-1}\boldsymbol{W}$ *is* $\mathbf{\Theta}$*-discordant and* $m > 144\sigma^2 dr$*, we have an estimation error*

$$\min_{\boldsymbol{Q}\in\mathcal{O}_{r,r}} \max_{1\leq i\leq m} \|\boldsymbol{O}_i\boldsymbol{Q} - \mathbf{\Theta}_i\|_F \leq \frac{2\left(3\sigma\sqrt{\frac{dr\log m}{m}} + 36\sigma^2 d\sqrt{\frac{r}{m}}\right)}{1 - 12\sigma\sqrt{\frac{dr}{m}}}.$$

Thus if $\sigma = o(\frac{m^{1/4}}{\sqrt{dr}})$, then we have $\min_{\boldsymbol{Q}\in\mathcal{O}_{r,r}} \max_{1\leq i\leq m} \|\boldsymbol{O}_i\boldsymbol{Q} - \mathbf{\Theta}_i\|_F \to 0$ as $m \to \infty$, as desired.

Remark 3.15. If the MAXDIFF model is assumed, $m > 2$, and $m^{3/2} - 2m^{1/2} - 12\sigma\sqrt{dr}m - 3 > 0$, then under $\mathbf{\Theta}$-discordance

$$\min_{\boldsymbol{Q}\in\mathcal{O}_{r,r}} \max_{1\leq i\leq m} \|\boldsymbol{O}_i\boldsymbol{Q} - \mathbf{\Theta}_i\|_F \leq \frac{2\left(3\sigma\sqrt{\frac{dr\log m}{m}} + 36\sigma^2 \frac{d\sqrt{r}}{\sqrt{m}-2/\sqrt{m}}\right)}{1 - 12\sigma\frac{\sqrt{dr}}{\sqrt{m}-2/\sqrt{m}} - \frac{3}{m}}.$$

3.3. Comparison with existing works. Our results generalize the work [4] on angular synchronization, which analyzes the setting $d = r = 1$ with complex values. In particular, Theorem 3.1, Corollary 3.3, Corollary 3.4, and Corollary 3.11 are generalizations of Lemma 3.2, Theorem 2.1, and Proposition 4.5 in [4], respectively. Corollary 3.3 is similar to Lemma 3.2 in [4] in the sense that both results establish deterministic conditions such that the original problem and the relaxed problem have the same solutions under a "discordant" condition. In addition, Corollary 3.4 is a generalization of [4, Theorem 2.1] in the sense that both results establish upper bounds on the size of noise σ under an additive Gaussian model. At last, both Corollary 3.11 and Proposition 4.5 in [4] show that local solutions satisfying an assumption are global optima.

Theorem 2.1 and Proposition 4.5 in [4] require $\sigma \leq \frac{1}{18}m^{\frac{1}{4}}$. In comparison, Corollary 3.4 and Corollary 3.11 require $\sigma \leq \frac{1}{60}m^{\frac{1}{4}}$ and $\sigma \leq \frac{1}{31}m^{\frac{1}{4}}$ under the setting $d = r = 1$, so our result is only worse by a constant factor.

The upper bound $\sigma \leq \frac{1}{18}m^{\frac{1}{4}}$ in [4, Theorem 2.1] is later improved to $\sigma \leq O(\sqrt{\frac{m}{\log m}})$ in [37], based on a much more complicated argument and an algorithmic implementation. After finishing this work, we became aware of a recent manuscript [22], which investigates the synchronization-of-rotations problem using the method in [37], and proves that the original problem and the relaxed problem have the same solution when $\sigma \leq O(\frac{\sqrt{m}}{d+\sqrt{d}\log m})$. While it is better than our rate $\sigma \leq O(\frac{m^{1/4}}{d})$ when $r = d$, our analysis investigates a more generic problem where r could be smaller than d and establishes deterministic conditions that can be verified for a variety of probabilistic models. In comparison, the method in [22] is specifically designed for the additive Gaussian noise model.

While the results in this section are generalizations of the results in [4] to the group of semiorthogonal matrices, we remark that the generalization is nontrivial in two aspects. First, as commented in the conclusion of [4], the noncommutative nature of semiorthogonal matrices renders the analysis more difficult. For example, the derivation in (5.29) is more difficult than the corresponding equation in [4, equation (4.3)]. Second, to analyze the more generic problem, we introduce a novel optimality certificate in Lemma 4.1, which is very different from the corresponding certificate in [4, Lemma 4.4]. In particular, our certificate concerns three variables, c, $\boldsymbol{T}^{(1)}$, and $\boldsymbol{T}^{(2)}$, while [4, Lemma 4.4] only depends on a single variable. More importantly, the certificate in [4, Lemma 4.4] has an explicit formula, but there is no explicit formula for the certificates $(c, \boldsymbol{T}^{(1)}, \boldsymbol{T}^{(2)})$ in our work. To address this issue, we let $c = m/2$ and define $\boldsymbol{T}^{(1)}$ and $\boldsymbol{T}^{(2)}$ in a constructive way in (5.10).

Ling [21] also proposes a generalization of [4] to the group of orthogonal matrices, which can be considered as our setting with $r = d$. Similar to [4, Lemma 4.4], the certificate in [21, Proposition 5.1] is based on a single variable with an explicit formula. While $-\boldsymbol{T}^{(1)}$ in our work serves a similar purpose as the certificates in [4, Lemma 4.4] and [21, Proposition 5.1], $\boldsymbol{T}^{(2)}$ and c are required for our setting and do not have an explicit formula. In comparison, under the setting of orthogonal matrices (i.e., $r = d$), our rate is in the order of $\sigma = O(\frac{m^{1/4}}{d})$, which is slightly worse than the rate of $O(\frac{m^{1/4}}{d^{3/4}})$ in [21] by a factor of $d^{1/4}$. We suspect that this is due to the more generic problem that we analyze, and our rate could be improved with a different way of constructing the certificates than (5.10), but we will leave it as a possible future direction. Related, in the simulation study presented in Appendix A, it is numerically demonstrated that the certificate (2.4) of global optimality is satisfied by the critical points generated by the proximal block ascent algorithm in [35] for a wide range of noise variances, even if condition (3.6) or (3.7) is not satisfied. This observation also suggests that condition (3.7) may be further improved.

4. Proof of main results.

4.1. Proof of Theorem 3.1. Recall that (OTSM) and (2.2) are equivalent in the sense that $\tilde{\boldsymbol{U}}_{ij} = \hat{\boldsymbol{O}}_i\hat{\boldsymbol{O}}_j^T$ for all $1 \leq i, j \leq m$, where $\tilde{\boldsymbol{U}} = (\tilde{\boldsymbol{U}}_{ij})$ is a solution to (2.2) and $\hat{\boldsymbol{O}} = (\hat{\boldsymbol{O}}_i)$ is a solution to (OTSM). It is sufficient to show that (2.2) and its relaxation (SDP) have the same solution. Then, the proof of Theorem 3.1 can be divided into three components as follows.

1. Lemma 4.1 shows that if $\boldsymbol{S}$ admits a decomposition $\boldsymbol{T}^{(1)} + \boldsymbol{T}^{(2)} + c\mathbf{I}$, where $\boldsymbol{T}^{(1)}$, $\boldsymbol{T}^{(2)}$, and a solution of (2.2) satisfy the conditions (4.1)–(4.2), then this solution is also the unique solution to the relaxed problem (SDP).

2. By constructing the certificates $\boldsymbol{T}^{(1)}$ and $\boldsymbol{T}^{(2)}$, Lemma 4.2 establishes (4.3), a sufficient condition such that (4.1)–(4.2) hold.

3. Lemma 4.3 establishes a perturbation result for the solution of (2.2). When $\boldsymbol{W}$ is small, the perturbation result can be used to verify (4.3).

We first present our lemmas and a short proof of Theorem 3.1 based on these lemmas and leave the technical proofs of the lemmas to section 5.

LEMMA 4.1 (a condition for the equivalence between (2.2) and (SDP)). *Let* $\tilde{\boldsymbol{U}}$ *be a solution to* (2.2), *and assume that it admits a decomposition* $\tilde{\boldsymbol{U}} = \tilde{\boldsymbol{V}}\tilde{\boldsymbol{V}}^T$ *with* $\tilde{\boldsymbol{V}} \in \mathbb{R}^{D\times r}$. *If there exists a decomposition* $\boldsymbol{S} = \boldsymbol{T}^{(1)} + \boldsymbol{T}^{(2)} + c\mathbf{I}$ *such that*

$$\boldsymbol{T}^{(1)} = \Pi_{\tilde{\boldsymbol{V}}^\perp}\boldsymbol{T}^{(1)}\Pi_{\tilde{\boldsymbol{V}}^\perp}, \quad \boldsymbol{T}^{(2)}_{ii} = \Pi_{\tilde{\boldsymbol{V}}_i}\boldsymbol{T}^{(2)}_{ii}\Pi_{\tilde{\boldsymbol{V}}_i} \text{ for all } 1 \le i \le m, \tag{4.1}$$

$$\{P^T_{\tilde{\boldsymbol{V}}_i}\boldsymbol{T}^{(2)}_{ii}P_{\tilde{\boldsymbol{V}}_i}\}_{i=1}^m \text{ and } -P^T_{\tilde{\boldsymbol{V}}^\perp}\boldsymbol{T}^{(1)}P_{\tilde{\boldsymbol{V}}^\perp} \text{ are positive definite matrices,} \tag{4.2}$$

then $\tilde{\boldsymbol{U}}$ *is also the unique solution to the relaxed problem* (SDP). *Therefore,* (2.2) *and* (SDP) *have the same unique solution.*

LEMMA 4.2 (a simplified condition in terms of the solution of (2.2)). *Let* $\tilde{\boldsymbol{U}}$ *be a solution to* (2.2)*, and assume that it admits a decomposition* $\tilde{\boldsymbol{U}} = \tilde{\boldsymbol{V}}\tilde{\boldsymbol{V}}^T$ *with* $\tilde{\boldsymbol{V}} \in \mathbb{R}^{D\times r}$*. If*

$$\frac{m}{2} \ge \max_{1\le i\le m}\left\|\textstyle\sum_{j=1}^m \boldsymbol{W}_{ij}\tilde{\boldsymbol{V}}_i\right\| + 2m\max_{1\le i\le m}\|\tilde{\boldsymbol{V}}_i - \boldsymbol{\Theta}_i\| + \|\boldsymbol{\Theta}^T\tilde{\boldsymbol{V}} - m\mathbf{I}\| + \|\boldsymbol{W}\|, \tag{4.3}$$

then there exist $\boldsymbol{T}^{(1)}$ *and* $\boldsymbol{T}^{(2)}$ *such that* $\boldsymbol{S} = \boldsymbol{T}^{(1)} + \boldsymbol{T}^{(2)} + \frac{m}{2}\mathbf{I}$*, and* (4.1)–(4.2) *hold with* $c = m/2$.

LEMMA 4.3 (perturbation bounds of the solutions of (2.2)). *If* $m > \|\boldsymbol{W}\|(4\sqrt{r}+1)+1$*, then for* $\tilde{\boldsymbol{U}}$*, any solution to* (2.2)*, there is a decomposition* $\tilde{\boldsymbol{U}} = \tilde{\boldsymbol{V}}\tilde{\boldsymbol{V}}^T$ *with* $\tilde{\boldsymbol{V}} \in \mathbb{R}^{D\times r}$ *such that*

$$\begin{aligned} &\|\tilde{\boldsymbol{V}} - \boldsymbol{\Theta}\|_F \le 4\|\boldsymbol{W}\|\sqrt{\tfrac{r}{m}}, \\ &\max_{1\le i\le m}\|[\boldsymbol{W}\tilde{\boldsymbol{V}}]_i\|_F \le \max_{1\le i\le m}\|[\boldsymbol{W}\boldsymbol{\Theta}]_i\|_F + 4\|\boldsymbol{W}\|^2\sqrt{\tfrac{r}{m}}, \end{aligned} \tag{4.4}$$

and

$$\max_{1\le i\le m}\|\tilde{\boldsymbol{V}}_i - \boldsymbol{\Theta}_i\|_F \le \frac{2\left(\max_{1\le i\le m}\|[\boldsymbol{W}\boldsymbol{\Theta}]_i\|_F + 4\|\boldsymbol{W}\|^2\sqrt{\frac{r}{m}}\right)}{m - \|\boldsymbol{W}\|(4\sqrt{r}+1) - 1}. \tag{4.5}$$

Proof of Theorem 3.1. Lemma 4.1 and Lemma 4.2 imply that, to prove Theorem 3.1, it is sufficient to prove (4.3), which can be verified by application of Lemma 4.3. ∎

4.2. Proof of Corollary 3.3.

Proof of Corollary 3.3. Under the **Θ**-discordant property, inequality (3.1) is satisfied if m is greater than

$$\frac{8m[3\sigma\sqrt{drm\log m} + 36\sigma^2 d\sqrt{rm}]}{m - 2 - 6\sigma\sqrt{dm}(2\sqrt{r}+1)} + 2[3\sigma\sqrt{drm\log m} + 36\sigma^2 d\sqrt{rm}] + 24\sigma\sqrt{dr} + 6\sigma\sqrt{dm}$$

or, by dividing the above expression by m,

$$1 > \left(2 + \frac{8}{1 - \frac{2}{m} - \frac{6\sigma\sqrt{d}(2\sqrt{r}+1)}{\sqrt{m}}}\right)\left[\frac{3\sigma\sqrt{dr\log m}}{\sqrt{m}} + \frac{36\sigma^2 d\sqrt{r}}{\sqrt{m}}\right] + \frac{24\sigma\sqrt{dr}}{m} + \frac{6\sigma\sqrt{d}}{\sqrt{m}}.$$

If $\sigma \le \frac{m^{1/4}}{60\sqrt{dr}}$, then the RHS of the above inequality is upper bounded by

$$\begin{aligned} &\left(2 + \frac{8}{1 - \frac{2}{m} - \frac{6}{60}\frac{2\sqrt{r}+1}{\sqrt{r}}\frac{1}{m^{1/4}}}\right)\left[\frac{3}{60}\frac{\sqrt{\log m}}{m^{1/4}} + \frac{36}{3600}\frac{1}{\sqrt{r}}\right] + \frac{24}{60}\frac{1}{m^{3/4}} + \frac{6}{60}\frac{1}{\sqrt{r}}\frac{1}{m^{1/4}} \\ &\le \left(2 + \frac{8}{1 - \frac{2}{m} - \frac{18}{60}\frac{1}{m^{1/4}}}\right)\left[\frac{3}{60}\frac{\sqrt{\log m}}{m^{1/4}} + \frac{36}{3600}\right] + \frac{24}{3600}\frac{1}{m^{3/4}} + \frac{6}{3600}\frac{1}{m^{1/4}} \end{aligned}$$

since $r \geq 1$ and $\frac{2\sqrt{r}+1}{\sqrt{r}} \leq 3$. The last line is decreasing in m if $m \geq 8$. At $m = 8$, the denominator in the last line is $1 - \frac{2}{8} - \frac{18}{60}\frac{1}{8^{1/4}} > 0$, and the value of the whole line is less than 1. ∎

4.3. Proof of Corollary 3.4.

Proof of Corollary 3.4. Considering Corollary 3.3, it is sufficient to show that Gaussian noise $\boldsymbol{W}$ satisfies the $\boldsymbol{\Theta}$-discordance with high probability under the MAXBET model. Assume $\sigma^{-1}\boldsymbol{W}_{ij}$ has i.i.d. standard normal entries. Then from $[\boldsymbol{W\Theta}]_i = \sum_{j=1}^m \boldsymbol{W}_{ij}\boldsymbol{\Theta}_j \in \mathbb{R}^{d_i \times r}$, it is obvious that this matrix has zero-mean normal entries. To see the variance, note

$$\operatorname{vec}(\boldsymbol{W}_{ij}\boldsymbol{\Theta}_j) = \operatorname{vec}(\mathbf{I}_{d_i}\boldsymbol{W}_{ij}\boldsymbol{\Theta}_j) = (\boldsymbol{\Theta}_j^T \otimes \mathbf{I}_{d_i})\operatorname{vec}(\boldsymbol{W}_{ij}).$$

Then $\operatorname{Cov}(\operatorname{vec}(\boldsymbol{W}_{ij})) = \sigma^2 \mathbf{I}_{d_i d_j}$ and

$$\begin{aligned}\operatorname{Cov}(\operatorname{vec}(\boldsymbol{W}_{ij}\boldsymbol{\Theta}_j)) &= \sigma^2(\boldsymbol{\Theta}_j^T \otimes \mathbf{I}_{d_i})(\boldsymbol{\Theta}_j^T \otimes \mathbf{I}_{d_i})^T = \sigma^2(\boldsymbol{\Theta}_j^T \otimes \mathbf{I}_{d_i})(\boldsymbol{\Theta}_j \otimes \mathbf{I}_{d_i})\\ &= \sigma^2(\boldsymbol{\Theta}_j^T\boldsymbol{\Theta}_j \otimes \mathbf{I}_{d_i}\mathbf{I}_{d_i}) = \sigma^2(\mathbf{I}_r \otimes \mathbf{I}_{d_i}) = \sigma^2\mathbf{I}_{rd_i};\end{aligned}$$

i.e., $\boldsymbol{W}_{ij}\boldsymbol{\Theta}_j$ has i.i.d. normal entries with variance σ^2. Then $[\boldsymbol{W\Theta}]_i$ is the sum of m i.i.d. copies of $\boldsymbol{W}_{ij}\boldsymbol{\Theta}_j$; hence entries have variance $m\sigma^2$. Now from Theorem 9.26 of [15],

$$\Pr\left(\frac{1}{\sigma\sqrt{m}}\|[\boldsymbol{W\Theta}]_i\| \geq \sqrt{d_i} + \sqrt{r} + t\right) \leq e^{-t^2/2}$$

for $t \geq 0$. Applying the union bound and noting that $\frac{1}{\sqrt{r}}\|[\boldsymbol{W\Theta}]_i\|_F \leq \|(\boldsymbol{W\Theta})_i\|_2$, we obtain

$$\Pr\left(\max_{i=1,\dots,m}\|[\boldsymbol{W\Theta}]_i\|_F \leq \sigma(\sqrt{\underline{d}rm} + r\sqrt{m} + t\sqrt{r})\right) > 1 - me^{-t^2/2},$$

where $\underline{d} = \min_{i=1,\dots,m} d_i$. Now choose t such that $e^{-t^2/2} = 1/m^2$, i.e., $t = 2\sqrt{\log m}$. Then,

$$\Pr\left(\max_{i=1,\dots,m}\|[\boldsymbol{W\Theta}]_i\|_F \leq \sigma(\sqrt{\underline{d}rm} + r\sqrt{m} + 2\sqrt{r\log m})\right) > 1 - \frac{1}{m}. \tag{4.6}$$

Since $\underline{d} \geq \max\{r, 2\}$ and $m \geq 2$, we have $r \leq \sqrt{\underline{d}r}$ and $\sqrt{\underline{d}m} \geq 2$. Furthermore, if $m \geq 3$, then $m \leq m\log m$. Thus

$$\sqrt{\underline{d}rm} + r\sqrt{m} + 2\sqrt{r\log m} \leq 3\sqrt{\underline{d}rm\log m} \leq 3\sqrt{Dr\log m}. \tag{4.7}$$

If $m = 2$ and $\underline{d} \geq 6$,

$$\sqrt{2\underline{d}r} + \sqrt{2r^2} + 2\sqrt{r\log 2} \leq 3\sqrt{2\underline{d}r\log 2} \leq 3\sqrt{Dr\log 2}.$$

Thus if $m \geq 3$ or $m \geq 2$ and $\underline{d} \geq 6$, then $\max_{i=1,\dots,m}\|[\frac{1}{\sigma}\boldsymbol{W\Theta}]_i\|_F \leq 3\sqrt{Dr\log m}$ with probability at least $1 - 1/m$.

To bound $\|\boldsymbol{W}\|$, observe that $\boldsymbol{W} \overset{d}{=} \boldsymbol{W}^{(1)} + \boldsymbol{W}^{(2)}$, where $\boldsymbol{W}^{(1)} \in \mathbb{R}^{D\times D}$ has entries i.i.d. from $N(0, \sigma^2/2)$, and $\boldsymbol{W}^{(2)}$ is generated as follows: $[\boldsymbol{W}^{(2)}]_{ij} = [\boldsymbol{W}^{(1)}]_{ji}^T$ for $i \neq j$, and $[\boldsymbol{W}^{(2)}]_{ii}$ has entries i.i.d. from $N(0, \sigma^2/2)$ under (MAXBET), or

$[\boldsymbol{W}^{(2)}]_{ii} = -[\boldsymbol{W}^{(2)}]_{ii}$ under (MAXDIFF). Marginally both $\boldsymbol{W}^{(1)}$ and $\boldsymbol{W}^{(2)}$ have entries i.i.d. from $N(0, \sigma^2/2)$. Then, [13, Theorem II.13] implies that

$$\Pr\left(\frac{\sqrt{2}}{\sigma}\|\boldsymbol{W}^{(1)}\| \geq 2\sqrt{D} + t\right) = \Pr\left(\frac{\sqrt{2}}{\sigma}\|\boldsymbol{W}^{(2)}\| \geq 2\sqrt{D} + t\right) < e^{-t^2/2}.$$

Applying the union bound and $\|\boldsymbol{W}\| \leq \|\boldsymbol{W}^{(1)}\| + \|\boldsymbol{W}^{(2)}\|$ yields

$$\Pr\left(\|\boldsymbol{W}\| \leq \sigma\sqrt{2}(2\sqrt{D} + t)\right) > 1 - 2e^{-t^2/2}$$

for $t \geq 0$. Choose $t = (\frac{3}{\sqrt{2}} - 2)\sqrt{D}$ to have $\Pr(\|\boldsymbol{W}\| \leq 3\sigma\sqrt{D}) > 1 - 2e^{-\frac{(3-2\sqrt{2})^2}{4}D}$. ◻

4.4. Proof of Corollary 3.7. The proof follows from (4.5) in Lemma 4.3 and Corollary 3.3.

4.5. Proof of Theorem 3.9. As a preparation, we provide intermediate results first. Proofs of these results are provided in section 5.3.

LEMMA 4.4. *Let $\boldsymbol{\Lambda}_i$ be the Lagrange multiplier of a critical point $(\boldsymbol{O}_1, \ldots, \boldsymbol{O}_m)$ of problem* (OTSM). *That is, it is a symmetric $r \times r$ matrix satisfying $\boldsymbol{O}_i\boldsymbol{\Lambda}_i = \sum_{j=1}^m \boldsymbol{S}_{ij}\boldsymbol{O}_j$. Then, for block matrices $\boldsymbol{O} = [\boldsymbol{O}_1^T, \ldots, \boldsymbol{O}_m^T]^T$ and $\boldsymbol{\Theta} = [\boldsymbol{\Theta}_1^T, \ldots, \boldsymbol{\Theta}_m^T]^T$, the following holds under* (MAXBET)*:*

$$\|\boldsymbol{\Lambda}_i - m\mathbf{I}\| \leq \|\textstyle\sum_{j=1}^m \boldsymbol{W}_{ij}\boldsymbol{O}_j\| + m\|\boldsymbol{O}_i^T\boldsymbol{\Theta}_i - \mathbf{I}_r\| + \|\boldsymbol{\Theta}^T\boldsymbol{O} - m\mathbf{I}_r\|.$$

Under (MAXDIFF), *we have*

$$\|\boldsymbol{\Lambda}_i - (m-1)\mathbf{I}\| \leq \|\textstyle\sum_{j\neq i} \boldsymbol{W}_{ij}\boldsymbol{O}_j\| + m\|\boldsymbol{O}_i^T\boldsymbol{\Theta}_i - \mathbf{I}_r\| + \|\boldsymbol{\Theta}^T\boldsymbol{O} - m\mathbf{I}_r\|.$$

Results parallel to Lemma 4.3 are also obtained.

LEMMA 4.5. *Let $(\tilde{\boldsymbol{O}}_1, \ldots, \tilde{\boldsymbol{O}}_m) \in \times_{i=1}^m \mathcal{O}_{d,r}$ be a a global optimum of* (OTSM)*. If we build a block matrix $\tilde{\boldsymbol{O}} = [\tilde{\boldsymbol{O}}_1^T, \ldots, \tilde{\boldsymbol{O}}_m^T]^T$, then there exists an orthogonal matrix $\boldsymbol{R} \in \mathcal{O}_{r,r}$ such that $(\tilde{\boldsymbol{O}}_1\boldsymbol{R}, \ldots, \tilde{\boldsymbol{O}}_m\boldsymbol{R})$ is also a global optimum and for $\boldsymbol{O} = \tilde{\boldsymbol{O}}\boldsymbol{R}$ the following error estimates hold:*

$$\|\boldsymbol{O} - \boldsymbol{\Theta}\|_F \leq \begin{cases} 4\|\boldsymbol{W}\|\frac{\sqrt{r}}{\sqrt{m}} & \textit{under (MAXBET)}, \\ 4\|\boldsymbol{W}\|\frac{\sqrt{r}}{\sqrt{m}-2/\sqrt{m}} & \textit{under (MAXDIFF)}, \end{cases} \tag{4.8}$$

$$\|\boldsymbol{\Theta}^T\boldsymbol{O} - m\mathbf{I}_r\| \leq \begin{cases} 4\|\boldsymbol{W}\|\sqrt{r} & \textit{under (MAXBET)}, \\ 4\|\boldsymbol{W}\|\frac{\sqrt{r}}{1-2/m} & \textit{under (MAXDIFF)}, \end{cases} \tag{4.9}$$

$$\max_{1\leq i\leq m}\|[\boldsymbol{W}\boldsymbol{O}]_i\|_F \leq \max_{1\leq i\leq m}\|[\boldsymbol{W}\boldsymbol{\Theta}]_i\|_F + \begin{cases} 4\|\boldsymbol{W}\|^2\frac{\sqrt{r}}{\sqrt{m}} & \textit{under (MAXBET)}, \\ 4\|\boldsymbol{W}\|^2\frac{\sqrt{r}}{\sqrt{m}-2/\sqrt{m}} & \textit{under (MAXDIFF)}, \end{cases} \tag{4.10}$$

$$\max_{1\leq i\leq m}\|\boldsymbol{O}_i - \boldsymbol{\Theta}_i\|_F \leq \begin{cases} \frac{2\left(\max_{1\leq i\leq m}\|[\boldsymbol{W}\boldsymbol{\Theta}]_i\|_F + 4\|\boldsymbol{W}\|^2\frac{\sqrt{r}}{\sqrt{m}}\right)}{m - 4\|\boldsymbol{W}\|\sqrt{r}} & \textit{under (MAXBET)}, \\ \frac{2\left(\max_{1\leq i\leq m}\|[\boldsymbol{W}\boldsymbol{\Theta}]_i\|_F + 4\|\boldsymbol{W}\|^2\frac{\sqrt{r}}{\sqrt{m}-2/\sqrt{m}}\right)}{m - 4\|\boldsymbol{W}\|\frac{\sqrt{r}}{1-2/m} - 3} & \textit{under (MAXDIFF)}, \end{cases} \tag{4.11}$$

where

$$m > \begin{cases} 4\|\boldsymbol{W}\|\sqrt{r} & \textit{under } \text{(MAXBET)}, \\ \frac{5+4\|\boldsymbol{W}\|\sqrt{r}+\sqrt{16\|\boldsymbol{W}\|^2 r+40\|\boldsymbol{W}\|\sqrt{r}+1}}{2} & \textit{under } \text{(MAXDIFF)}. \end{cases}$$

Assume the data model (MAXBET). We want a condition on the noise matrices $\boldsymbol{W}_{ij}$ that guarantees the certificate (2.4) to hold. Let $(\tilde{\boldsymbol{O}}_1, \ldots, \tilde{\boldsymbol{O}}_m)$ be a global optimum of (OTSM) and $\tilde{\boldsymbol{O}} = [\tilde{\boldsymbol{O}}_1^T, \ldots, \tilde{\boldsymbol{O}}_m^T]$. Since $\boldsymbol{L}\hat{\boldsymbol{O}} = \boldsymbol{0}$ for $\hat{\boldsymbol{O}} = [\hat{\boldsymbol{O}}_1^T, \ldots, \hat{\boldsymbol{O}}_m^T]$ whenever $(\hat{\boldsymbol{O}}_1, \ldots, \hat{\boldsymbol{O}}_m)$ is a critical point, it suffices to find a condition that

$$\boldsymbol{x}^T \boldsymbol{L}(\tilde{\boldsymbol{O}}, \tilde{\boldsymbol{\Lambda}})\boldsymbol{x} \geq 0 \quad \text{for all } \boldsymbol{x} = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_m), \boldsymbol{x}_i \in \mathbb{R}^{d_i} \text{ such that } \tilde{\boldsymbol{O}}^T \boldsymbol{x} = \boldsymbol{0},$$

where $\tilde{\boldsymbol{\Lambda}} = (\tilde{\boldsymbol{\Lambda}}_1, \ldots, \tilde{\boldsymbol{\Lambda}}_m)$, $\tilde{\boldsymbol{\Lambda}}_i = \sum_{j=1}^m \tilde{\boldsymbol{O}}_i^T \boldsymbol{S}_{ij} \tilde{\boldsymbol{O}}_j$, is the collection of the associated Lagrange multipliers.

Let $(\boldsymbol{O}_1, \ldots, \boldsymbol{O}_m)$ be a critical point and $\boldsymbol{O} = [\boldsymbol{O}_1^T, \ldots, \boldsymbol{O}_m^T]^T$. Then, for any $\boldsymbol{x}$ satisfying $\boldsymbol{O}^T \boldsymbol{x} = \boldsymbol{0}$,

$$\begin{aligned} \boldsymbol{x}^T \boldsymbol{L}(\boldsymbol{O}, \boldsymbol{\Lambda})\boldsymbol{x} &= \sum_{i=1}^m \left(\boldsymbol{x}_i^T \boldsymbol{O}_i \boldsymbol{\Lambda}_i \boldsymbol{O}_i^T \boldsymbol{x}_i + \tau_i \boldsymbol{x}_i^T \boldsymbol{O}_i^\perp \boldsymbol{O}_i^{\perp T} \boldsymbol{x} \right) - \boldsymbol{x}^T \boldsymbol{S} \boldsymbol{x} \\ &\geq \sum_{i=1}^m \left(\tau_i \boldsymbol{x}_i^T \boldsymbol{O}_i \boldsymbol{O}_i^T \boldsymbol{x}_i + \tau_i \boldsymbol{x}_i^T \boldsymbol{O}_i^\perp \boldsymbol{O}_i^{\perp T} \boldsymbol{x}_i \right) - \boldsymbol{x}^T \boldsymbol{S} \boldsymbol{x} \\ &= \sum_{i=1}^m \tau_i \|\boldsymbol{x}_i\|^2 - \boldsymbol{x}^T \boldsymbol{S} \boldsymbol{x}. \end{aligned}$$

The block matrix (2.1) can be written as

$$\boldsymbol{S} = \boldsymbol{\Theta}\boldsymbol{\Theta}^T + \boldsymbol{W}, \tag{4.12}$$

where $\boldsymbol{W}$ is a block matrix constructed from $\boldsymbol{W}_{ij}$ in a similar fashion to (2.1). Then

$$\begin{aligned} \boldsymbol{x}^T \boldsymbol{S} \boldsymbol{x} &= \boldsymbol{x}^T \boldsymbol{\Theta}\boldsymbol{\Theta}^T \boldsymbol{x} + \boldsymbol{x}^T \boldsymbol{W} \boldsymbol{x} = \boldsymbol{x}^T (\boldsymbol{\Theta} - \boldsymbol{O})(\boldsymbol{\Theta} - \boldsymbol{O})^T \boldsymbol{x} + \boldsymbol{x}^T \boldsymbol{W} \boldsymbol{x} \\ &= \|(\boldsymbol{\Theta} - \boldsymbol{O})^T \boldsymbol{x}\|^2 + \boldsymbol{x}^T \boldsymbol{W} \boldsymbol{x} \leq \|\boldsymbol{\Theta} - \boldsymbol{O}\|^2 \|\boldsymbol{x}\|^2 + \|\boldsymbol{W}\| \|\boldsymbol{x}\|^2. \end{aligned} \tag{4.13}$$

The second equality is due to $\boldsymbol{O}^T \boldsymbol{x} = \boldsymbol{0}$. Hence we have

$$\boldsymbol{x}^T \boldsymbol{L}(\boldsymbol{O}, \boldsymbol{\Lambda})\boldsymbol{x} \geq \sum_{i=1}^m \tau_i \|\boldsymbol{x}_i\|^2 - \|\boldsymbol{\Theta} - \boldsymbol{O}\|^2 \|\boldsymbol{x}\|^2 - \|\boldsymbol{W}\| \|\boldsymbol{x}\|^2. \tag{4.14}$$

Combining Weyl's inequality and Lemma 4.4, we obtain a lower bound on τ_i:

$$\tau_i \geq m - \|[\boldsymbol{W}\boldsymbol{O}]_i\| - m\|\boldsymbol{O}_i^T \boldsymbol{\Theta}_i - \mathbf{I}\| - \|\boldsymbol{\Theta}^T \boldsymbol{O} - m\mathbf{I}\|.$$

Substituting this with inequality (4.15), we see

$$\begin{aligned} \boldsymbol{x}^T \boldsymbol{L}(\boldsymbol{O}, \boldsymbol{\Lambda})\boldsymbol{x} &\geq (m - \|\boldsymbol{\Theta}^T \boldsymbol{O} - m\mathbf{I}\|)\|\boldsymbol{x}\|^2 - \|\boldsymbol{\Theta} - \boldsymbol{O}\|^2 \|\boldsymbol{x}\|^2 \\ &\quad - \sum_{i=1}^m \left(\|[\boldsymbol{W}\boldsymbol{O}]_i\| \|\boldsymbol{x}_i\|^2 + m\|\boldsymbol{O}_i^T \boldsymbol{\Theta}_i - \mathbf{I}\| \|\boldsymbol{x}_i\|^2 \right) - \|\boldsymbol{W}\| \|\boldsymbol{x}\|^2 \\ &\geq \Big(m - \|\boldsymbol{\Theta}^T \boldsymbol{O} - m\mathbf{I}\| - \|\boldsymbol{\Theta} - \boldsymbol{O}\|^2 \\ &\quad - \max_{i=1,\ldots,m} \|[\boldsymbol{W}\boldsymbol{O}]_i\| - m \max_{i=1,\ldots,m} \|\boldsymbol{O}_i^T \boldsymbol{\Theta}_i - \mathbf{I}\| - \|\boldsymbol{W}\| \Big) \|\boldsymbol{x}\|^2. \end{aligned} \tag{4.15}$$

Thus if

$$m \geq \|\mathbf{\Theta}^T \boldsymbol{O} - m\mathbf{I}\| + \|\mathbf{\Theta} - \boldsymbol{O}\|^2 + \max_{i=1,\dots,m} \|[\boldsymbol{W}\boldsymbol{O}]_i\| + m \max_{i=1,\dots,m} \|\boldsymbol{O}_i^T \mathbf{\Theta}_i - \mathbf{I}\| + \|\boldsymbol{W}\|, \tag{4.16}$$

then we have $\boldsymbol{L}(\boldsymbol{O}, \mathbf{\Lambda}) \succcurlyeq \mathbf{0}$.

Now suppose $(\tilde{\boldsymbol{O}}_1, \dots, \tilde{\boldsymbol{O}}_m)$ is a global optimum and $\tilde{\mathbf{\Lambda}} = (\tilde{\mathbf{\Lambda}}_1, \dots, \tilde{\mathbf{\Lambda}}_m)$ is the collection of the associated Lagrange multipliers. Let $\tilde{\boldsymbol{O}} = [\tilde{\boldsymbol{O}}_1^T, \dots, \tilde{\boldsymbol{O}}_m^T]^T$. Then, by Lemma 4.5 there exists $\boldsymbol{R} \in \mathcal{O}_{r,r}$ such that $\boldsymbol{O} = \tilde{\boldsymbol{O}}\boldsymbol{R}$ satisfies inequalities (4.8)–(4.11). Then, for this $\boldsymbol{O}$ the RHS of inequality (4.16) can be bounded:

$$\begin{aligned} &\|\mathbf{\Theta}^T \boldsymbol{O} - m\mathbf{I}\| + \max_{i=1,\dots,m} \|(\boldsymbol{W}\boldsymbol{O})_i\| \quad + m \max_{i=1,\dots,m} \|\boldsymbol{O}_i^T \mathbf{\Theta}_i - \mathbf{I}\| + \|\mathbf{\Theta} - \boldsymbol{O}\|^2 + \|\boldsymbol{W}\| \\ &\leq 4\|\boldsymbol{W}\|\sqrt{r} + \max_{1 \leq i \leq m} \|[\boldsymbol{W}\mathbf{\Theta}]_i\|_F + 4\|\boldsymbol{W}\|^2 \sqrt{\tfrac{r}{m}} \\ &\quad + \frac{2m(\max_{1 \leq i \leq m} \|[\boldsymbol{W}\mathbf{\Theta}]_i\|_F + 4\|\boldsymbol{W}\|^2 \sqrt{\frac{r}{m}})}{m - 4\|\boldsymbol{W}\|\sqrt{r}} + 16\|\boldsymbol{W}\|^2 \frac{r}{m} + \|\boldsymbol{W}\|. \end{aligned}$$

If this bound is less than or equal to m, the resulting inequality is precisely (3.6), and then condition (4.16) is satisfied. In other words, $\boldsymbol{L}(\boldsymbol{O}, \mathbf{\Lambda}) \succcurlyeq \mathbf{0}$, where $\mathbf{\Lambda} = (\mathbf{\Lambda}_1, \dots, \mathbf{\Lambda}_m)$ and $\mathbf{\Lambda}_i = \sum_{j=1}^m \boldsymbol{O}_i^T \boldsymbol{S}_{ij} \boldsymbol{O}_j = \boldsymbol{R}\tilde{\mathbf{\Lambda}}_i \boldsymbol{R}^T$, $i = 1, \dots, m$, are the associated Lagrange multipliers.

Finally, observing that

$$\boldsymbol{L}\big(\boldsymbol{A}, (\boldsymbol{B}_1, \dots, \boldsymbol{B}_m)\big) = \boldsymbol{L}\big(\boldsymbol{A}\boldsymbol{Q}, (\boldsymbol{Q}^T \boldsymbol{B}_1 \boldsymbol{Q}, \dots, \boldsymbol{Q}^T \boldsymbol{B}_m \boldsymbol{Q})\big)$$

for any $\boldsymbol{Q} \in \mathcal{O}_{r,r}$, $\boldsymbol{A} = [\boldsymbol{A}_1^T, \dots, \boldsymbol{A}_m^T]^T$ with $\boldsymbol{A}_i \in \mathbb{R}^{d_i \times r}$, and $\boldsymbol{B}_i \in \mathbb{R}^{r \times r}$ shows that $\boldsymbol{L}(\tilde{\boldsymbol{O}}, \tilde{\mathbf{\Lambda}}) \succcurlyeq \mathbf{0}$.

For the similar result under the model (MAXDIFF), see section 4.7.

4.6. Proof of Corollary 3.10. Under the $\mathbf{\Theta}$-concordance of $\frac{1}{\sigma}\boldsymbol{W}$, the RHS of inequality (3.6) in Theorem 3.9 is upper bounded by

$$12\sigma\sqrt{Dr} + 3\sigma\sqrt{Dr\log m} + 36\sigma^2 D\sqrt{\frac{r}{m}} + \frac{2m(3\sigma\sqrt{Dr\log m} + 36\sigma^2 D\sqrt{\frac{r}{m}})}{m - 12\sigma\sqrt{Dr}} + 144\sigma^2 \frac{Dr}{m} + 3\sigma\sqrt{D} \tag{4.17}$$

if $\sigma < \frac{m}{12\sqrt{Dr}}$. If (4.17) is less than or equal to m or equivalently

$$1 \geq 12\sigma\sqrt{\frac{dr}{m}} + 3\sigma\sqrt{\frac{dr\log m}{m}} + 36\sigma^2 d\sqrt{\frac{r}{m}} + \frac{2(3\sigma\sqrt{drm\log m} + 36\sigma^2 d\sqrt{rm})}{m - 12\sigma\sqrt{drm}} + 144\sigma^2 \frac{dr}{m} + 3\sigma\sqrt{\frac{d}{m}} \tag{4.18}$$

for $\sigma < \frac{m^{1/2}}{12\sqrt{dr}}$, then from the proof of Theorem 3.9, we see that condition (4.16) is satisfied, and thus the claim is proved.

The fourth term on the RHS of inequality (4.18) is

$$\frac{2\left(3\sigma\sqrt{\frac{dr\log m}{m}} + 36\sigma^2 d\sqrt{\frac{r}{m}}\right)}{1 - 12\sigma\sqrt{\frac{dr}{m}}} \leq \frac{2}{1 - \frac{12}{31}\frac{1}{m^{1/4}}}\left(3\sigma\sqrt{\frac{dr\log m}{m}} + 36\sigma^2 d\sqrt{\frac{r}{m}}\right)$$

if $\sigma \le \frac{m^{1/4}}{31\sqrt{dr}}$. Thus, by replacing σ with $\frac{m^{1/4}}{31\sqrt{dr}}$, the RHS of (4.18) is upper bounded by

$$\frac{12}{31}\frac{1}{m^{1/4}} + \left(1 + \frac{2}{1 - \frac{12}{31}\frac{1}{m^{1/4}}}\right)\left(\frac{3}{31}\frac{\sqrt{\log m}}{m^{1/4}} + \frac{36}{961}\frac{1}{\sqrt{r}}\right) + \frac{144}{961}\frac{1}{m^{1/2}} + \frac{3}{31}\frac{1}{\sqrt{r}m^{1/4}}.$$

Since $r \ge 1$, $\frac{\sqrt{\log m}}{m^{1/4}} \le \sqrt{\frac{2}{e}}$, and the rest of the terms are decreasing in m, the above quantity is less than 1 for $m \ge 2$.

4.7. Theorem 3.9, Corollary 3.10, and Corollary 3.11 under (MAXDIFF). Under the MAXDIFF model, inequality (4.13) is replaced by

$$\begin{aligned} \boldsymbol{x}^T\boldsymbol{S}\boldsymbol{x} &\le \|\boldsymbol{\Theta} - \boldsymbol{O}\|^2\|\boldsymbol{x}\|^2 - \sum_{i=1}^m \|\boldsymbol{\Theta}_i\boldsymbol{\Theta}_i^T\boldsymbol{x}_i\|^2 + \|\boldsymbol{W}\|\|\boldsymbol{x}\|^2 \\ &\le \left(\|\boldsymbol{\Theta} - \boldsymbol{O}\|^2 - \min_{1\le i\le m}\|\boldsymbol{\Theta}_i\|^2 + \|\boldsymbol{W}\|\right)\|\boldsymbol{x}\|^2 \end{aligned}$$

and (4.15) by

$$\begin{aligned} \boldsymbol{x}^T\boldsymbol{L}(\boldsymbol{O},\boldsymbol{\Lambda})\boldsymbol{x} \ge \Big(&m - 1 - \|\boldsymbol{\Theta}^T\boldsymbol{O} - m\mathbf{I}\| - \max_{i=1,\dots,m}\|[\boldsymbol{W}\boldsymbol{O}]_i\| \\ &-m\max_{i=1,\dots,m}\|\boldsymbol{O}_i^T\boldsymbol{\Theta}_i - \mathbf{I}\| - \|\boldsymbol{\Theta} - \boldsymbol{O}\|^2 - \|\boldsymbol{W}\| + \min_{1\le i\le m}\|\boldsymbol{\Theta}_i\|^2\Big)\|\boldsymbol{x}\|^2 \\ = \Big(&m - \|\boldsymbol{\Theta}^T\boldsymbol{O} - m\mathbf{I}\| - \max_{i=1,\dots,m}\|[\boldsymbol{W}\boldsymbol{O}]_i\| \\ &-m\max_{i=1,\dots,m}\|\boldsymbol{O}_i^T\boldsymbol{\Theta}_i - \mathbf{I}\| - \|\boldsymbol{\Theta} - \boldsymbol{O}\|^2 - \|\boldsymbol{W}\|\Big)\|\boldsymbol{x}\|^2 \end{aligned}$$

since $\|\boldsymbol{\Theta}_i\| = 1$ for all i. Thus condition (4.16) for $\boldsymbol{L}(\boldsymbol{O},\boldsymbol{\Lambda}) \succcurlyeq \mathbf{0}$ to hold remains unchanged. Applying Lemma 4.5, we obtain

$$\begin{aligned} m \ge\ & 4\|\boldsymbol{W}\|\tfrac{\sqrt{r}}{1-2/m} + \max\nolimits_{1\le i\le m}\|[\boldsymbol{W}\boldsymbol{\Theta}]_i\|_F + 4\|\boldsymbol{W}\|^2\tfrac{\sqrt{r}}{\sqrt{m}-2/\sqrt{m}} + \|\boldsymbol{W}\| \\ &+ \frac{2m(\max_{1\le i\le m}\|[\boldsymbol{W}\boldsymbol{\Theta}]_i\|_F + 4\|\boldsymbol{W}\|^2\tfrac{\sqrt{r}}{\sqrt{m}-2/\sqrt{m}})}{m - 4\|\boldsymbol{W}\|\tfrac{\sqrt{r}}{1-2/m} - 3} + 16\|\boldsymbol{W}\|^2\tfrac{r}{(\sqrt{m}-2/\sqrt{m})^2}. \end{aligned}$$

Proceeding as above for (MAXBET), we obtain the bound on σ as stated in Remark 3.13.

Furthermore, inequality (4.6) is replaced by

$$\Pr\left(\max_{i=1,\dots,m}\|[\boldsymbol{W}\boldsymbol{\Theta}]_i\|_F \le \sigma(\sqrt{\underline{d}r(m-1)} + r\sqrt{m-1} + 2\sqrt{r\log m})\right) > 1 - \frac{1}{m}$$

(recall that $\underline{d} = \min_{m=1,\dots,m} d_i$), and inequality (4.7) holds for $m \ge 2$ for all $\underline{d}$ since $m - 1 \le m\log m$ for all $m \ge 2$. Thus the conclusion of Corollary 3.11 holds without modification, provided that $m \ge 4$ and $\sigma \le \frac{m^{1/4}}{64\sqrt{dr}}$ as stated in Remark 3.13.

4.8. Proof of Corollary 3.14. The desired results follow immediately from inequality (4.11) of Lemma 4.5 and Definition 3.2.

5. Proofs of technical lemmas and propositions.

5.1. Proof of Proposition 2.1.

Proof of Proposition 2.1. First consider model (MAXBET). We have $\boldsymbol{S}_{ij} = \boldsymbol{\Theta}_i\boldsymbol{\Theta}_j^T$ for all i, j. Then the objective of (OTSM) is

$$\sum_{i,j} \operatorname{tr}(\boldsymbol{O}_i^T\boldsymbol{\Theta}_i\boldsymbol{\Theta}_j^T\boldsymbol{O}_j) = \sum_{i,j} \operatorname{tr}[(\boldsymbol{\Theta}_i^T\boldsymbol{O}_i)^T(\boldsymbol{\Theta}_j^T\boldsymbol{O}_j)].$$

Each term is bounded by the von Neumann–Fan inequality [20, Example 2.8.7]

$$\operatorname{tr}[(\boldsymbol{\Theta}_i^T\boldsymbol{O}_i)^T(\boldsymbol{\Theta}_j^T\boldsymbol{O}_j)] \le \sum_{k=1}^{r} \sigma_k(\boldsymbol{\Theta}_i^T\boldsymbol{O}_i)\sigma_k(\boldsymbol{\Theta}_j^T\boldsymbol{O}_j), \tag{5.1}$$

where $\sigma_k(\boldsymbol{M})$ is the kth largest singular value of matrix $\boldsymbol{M}$. Since $\boldsymbol{O}_i^T\boldsymbol{\Theta}_i\boldsymbol{\Theta}_i^T\boldsymbol{O}_i \preccurlyeq \boldsymbol{O}_i^T\boldsymbol{O}_i = \mathbf{I}_r$, we see $\max_{k=1,\dots,r} \sigma_k(\boldsymbol{\Theta}_i^T\boldsymbol{O}_i) \le 1$ for all $\boldsymbol{O}_i \in \mathcal{O}_{d_i,r}$, $i = 1,\dots,m$. Thus the largest possible value of the RHS of inequality (5.1) is r, and (OTSM) has maximum m^2r. This is achieved by $\boldsymbol{O}_i = \boldsymbol{\Theta}_i$ for $i = 1,\dots,m$ since $\boldsymbol{\Theta}_i^T\boldsymbol{\Theta}_i = \mathbf{I}_r$.

It is straightforward to modify the above proof for model (MAXDIFF). The maximum is $m(m-1)r$. ∎

5.2. Proofs of Lemmas for Theorem 3.1.

Proof of Lemma 4.1. For any $\boldsymbol{U}$ in the constraint set of (SDP) such that $\boldsymbol{U} \neq \tilde{\boldsymbol{U}}$ and $\boldsymbol{X} = \boldsymbol{U} - \tilde{\boldsymbol{U}}$, we have $P_{\tilde{\boldsymbol{V}}^\perp}^T\boldsymbol{X}P_{\tilde{\boldsymbol{V}}^\perp} = P_{\tilde{\boldsymbol{V}}^\perp}^T\boldsymbol{U}P_{\tilde{\boldsymbol{V}}^\perp} - P_{\tilde{\boldsymbol{V}}^\perp}^T\tilde{\boldsymbol{U}}P_{\tilde{\boldsymbol{V}}^\perp} = P_{\tilde{\boldsymbol{V}}^\perp}^T\boldsymbol{U}P_{\tilde{\boldsymbol{V}}^\perp} \succcurlyeq \mathbf{0}$ and $P_{\tilde{\boldsymbol{V}}_i}^T\boldsymbol{X}_{ii}P_{\tilde{\boldsymbol{V}}_i} = P_{\tilde{\boldsymbol{V}}_i}^T\boldsymbol{U}_{ii}P_{\tilde{\boldsymbol{V}}_i} - P_{\tilde{\boldsymbol{V}}_i}^T\tilde{\boldsymbol{U}}_{ii}P_{\tilde{\boldsymbol{V}}_i} = P_{\tilde{\boldsymbol{V}}_i}^T\boldsymbol{U}_{ii}P_{\tilde{\boldsymbol{V}}_i} - \mathbf{I} \preccurlyeq \mathbf{0}$. In summary, $\boldsymbol{X}$ has the properties of

$$P_{\tilde{\boldsymbol{V}}^\perp}^T\boldsymbol{X}P_{\tilde{\boldsymbol{V}}^\perp} \succcurlyeq \mathbf{0},\ \operatorname{tr}(\boldsymbol{X}_{ii}) = 0 \text{ and } P_{\tilde{\boldsymbol{V}}_i}^T\boldsymbol{X}_{ii}P_{\tilde{\boldsymbol{V}}_i} \preccurlyeq \mathbf{0} \text{ for all } 1 \le i \le m. \tag{5.2}$$

In addition, either $P_{\tilde{\boldsymbol{V}}^\perp}^T\boldsymbol{X}P_{\tilde{\boldsymbol{V}}^\perp}$ is nonzero or $P_{\tilde{\boldsymbol{V}}_i}^T\boldsymbol{X}_{ii}P_{\tilde{\boldsymbol{V}}_i}$ is nonzero for some i. If they are all zero matrices, then we have

$$P_{\tilde{\boldsymbol{V}}^\perp}^T\boldsymbol{U}P_{\tilde{\boldsymbol{V}}^\perp} = P_{\tilde{\boldsymbol{V}}^\perp}^T\tilde{\boldsymbol{U}}P_{\tilde{\boldsymbol{V}}^\perp} = \mathbf{0}, \tag{5.3}$$

$$P_{\tilde{\boldsymbol{V}}_i}^T\boldsymbol{U}_{ii}P_{\tilde{\boldsymbol{V}}_i} = P_{\tilde{\boldsymbol{V}}_i}^T\tilde{\boldsymbol{U}}_{ii}P_{\tilde{\boldsymbol{V}}_i} = \mathbf{I}_r. \tag{5.4}$$

Since $\boldsymbol{U}_{ii} \succcurlyeq 0$, we have $\tilde{\boldsymbol{V}}_i^T\boldsymbol{U}_{ii}\tilde{\boldsymbol{V}}_i \succcurlyeq \mathbf{0}$. Combining it with $\operatorname{tr}(P_{\tilde{\boldsymbol{V}}_i}^T\boldsymbol{U}_{ii}P_{\tilde{\boldsymbol{V}}_i}) = r$ (due to (5.4)) and $r = \operatorname{tr}(\boldsymbol{U}_{ii}) = \operatorname{tr}(P_{\tilde{\boldsymbol{V}}_i}^T\boldsymbol{U}_{ii}P_{\tilde{\boldsymbol{V}}_i}) + \operatorname{tr}(\tilde{\boldsymbol{V}}_i^T\boldsymbol{U}_{ii}\tilde{\boldsymbol{V}}_i)$, we have $\tilde{\boldsymbol{V}}_i^T\boldsymbol{U}_{ii}\tilde{\boldsymbol{V}}_i = \mathbf{0}$. Combining it with $\boldsymbol{U}_{ii} \succcurlyeq \mathbf{0}$, we have $\tilde{\boldsymbol{V}}_i^T\boldsymbol{U}_{ii} = \mathbf{0}$ and $\boldsymbol{U}_{ii}^T\tilde{\boldsymbol{V}}_i = \mathbf{0}$. It implies that $\boldsymbol{U}_{ii} = \tilde{V}_iZ_i\tilde{V}_i^T$ for some positive semidefinite Z_i. That $\boldsymbol{U}_{ii} \preccurlyeq \mathbf{I}$ and $\operatorname{tr}(\boldsymbol{U}_{ii}) = r$ in turn implies that $Z_i = \mathbf{I}_r$. Thus,

$$\boldsymbol{U}_{ii} = \tilde{\boldsymbol{V}}_i\tilde{\boldsymbol{V}}_i^T. \tag{5.5}$$

In addition, (5.3) and $\boldsymbol{U} \succcurlyeq \mathbf{0}$ mean that $\boldsymbol{U} = \Pi_{\tilde{\boldsymbol{V}}}^T\boldsymbol{U}\Pi_{\tilde{\boldsymbol{V}}}$; that is, there exists a matrix $\boldsymbol{Z} \in \mathbb{R}^{r\times r}$ such that $\boldsymbol{U} = \tilde{\boldsymbol{V}}\boldsymbol{Z}\tilde{\boldsymbol{V}}^T$, and as a result, $\boldsymbol{U}_{ii} = \tilde{\boldsymbol{V}}_i\boldsymbol{Z}\tilde{\boldsymbol{V}}_i^T$. Combining it with (5.5), we have $\boldsymbol{Z} = \mathbf{I}$ and $\boldsymbol{U} = \tilde{\boldsymbol{V}}\tilde{\boldsymbol{V}}^T = \tilde{\boldsymbol{U}}$, which is a contradiction to assumption $\boldsymbol{U} \neq \tilde{\boldsymbol{U}}$.

Combining the property of $\boldsymbol{X}$ in (5.2) with the assumption of $\boldsymbol{T}$ in (4.2) that $\{P_{\tilde{\boldsymbol{V}}_i}^T \boldsymbol{T}_{ii}^{(2)} P_{\tilde{\boldsymbol{V}}_i}\}_{i=1}^m$ and $-P_{\tilde{\boldsymbol{V}}^\perp}^T \boldsymbol{T}^{(1)} P_{\tilde{\boldsymbol{V}}^\perp}$ are positive definite matrices, we have

$$\begin{aligned}\operatorname{tr}(\boldsymbol{X}\boldsymbol{S}) &= \operatorname{tr}(\boldsymbol{X}\boldsymbol{T}^{(1)}) + \operatorname{tr}(\boldsymbol{X}\boldsymbol{T}^{(2)}) + c\operatorname{tr}(\boldsymbol{X}) \\ &= \operatorname{tr}[(P_{\tilde{\boldsymbol{V}}^\perp}^T \boldsymbol{X} P_{\tilde{\boldsymbol{V}}^\perp})(P_{\tilde{\boldsymbol{V}}^\perp}^T \boldsymbol{T}^{(1)} P_{\tilde{\boldsymbol{V}}^\perp})] + \textstyle\sum_{i=1}^m \operatorname{tr}(\boldsymbol{X}_{ii}\boldsymbol{T}_{ii}^{(2)}) \\ &= \operatorname{tr}[(P_{\tilde{\boldsymbol{V}}^\perp}^T \boldsymbol{X} P_{\tilde{\boldsymbol{V}}^\perp})(P_{\tilde{\boldsymbol{V}}^\perp}^T \boldsymbol{T}^{(1)} P_{\tilde{\boldsymbol{V}}^\perp})] \\ &\quad + \textstyle\sum_{i=1}^m \operatorname{tr}[(P_{\tilde{\boldsymbol{V}}_i}^T \boldsymbol{X}_{ii} P_{\tilde{\boldsymbol{V}}_i})(P_{\tilde{\boldsymbol{V}}_i}^T \boldsymbol{T}_{ii}^{(2)} P_{\tilde{\boldsymbol{V}}_i})] < 0.\end{aligned} \tag{5.6}$$

The first equality uses assumption (4.1). The last inequality is strict because either $P_{\tilde{\boldsymbol{V}}^\perp}^T \boldsymbol{X} P_{\tilde{\boldsymbol{V}}^\perp}$ is nonzero or $P_{\tilde{\boldsymbol{V}}_i}^T \boldsymbol{X}_{ii} P_{\tilde{\boldsymbol{V}}_i}$ is nonzero for some $1 \le i \le m$. Then (5.6) implies that $\operatorname{tr}(\boldsymbol{S}\boldsymbol{U}) < \operatorname{tr}(\boldsymbol{S}\tilde{\boldsymbol{U}})$ for all $\boldsymbol{U} \ne \tilde{\boldsymbol{U}}$, and as a result, $\tilde{\boldsymbol{U}}$ is the unique solution to (SDP). ∎

Proof of Lemma 4.2. In this proof, we aim to construct the certificate in Lemma 4.1. The process can be divided into three steps:

- Find a decomposition of $\boldsymbol{S} = \boldsymbol{S}^{(1)} + \boldsymbol{S}^{(2)}$ based on the first-order optimality.
- Construct the certificate $\boldsymbol{T}^{(1)}$ and $\boldsymbol{T}^{(2)}$ from the decomposition $\boldsymbol{S}^{(1)}$ and $\boldsymbol{S}^{(2)}$. The explicit expression is given in (5.10).
- Verify that the certificate satisfies the conditions in Lemma 4.1.

Step 1: Decomposition of $\boldsymbol{S}$ based on the first-order optimality. We investigate the first-order condition for any solution of (2.2) and summarize the result in Lemma 5.1 as below.

LEMMA 5.1. *Let $\tilde{\boldsymbol{U}} = \tilde{\boldsymbol{V}}\tilde{\boldsymbol{V}}^T$ be a solution to* (2.2) *with $\tilde{\boldsymbol{V}} \in \mathbb{R}^{D\times r}$. Then the input matrix $\boldsymbol{S}$ can be decomposed into $\boldsymbol{S} = \boldsymbol{S}^{(1)} + \boldsymbol{S}^{(2)}$, where $\boldsymbol{S}^{(1)}$ and $\boldsymbol{S}^{(2)}$ are such that*

$$[\boldsymbol{S}^{(1)}]_{ij} = \begin{cases} \boldsymbol{S}_{ij}, & i \ne j, \\ \boldsymbol{S}_{ii} - \sum_{j=1}^m \boldsymbol{S}_{ij}\tilde{\boldsymbol{V}}_j\tilde{\boldsymbol{V}}_i^T, & i = j, \end{cases} \tag{5.7}$$

$$[\boldsymbol{S}^{(2)}]_{ij} = \begin{cases} \boldsymbol{0}, & i \ne j, \\ \sum_{j=1}^m \boldsymbol{S}_{ij}\tilde{\boldsymbol{V}}_j\tilde{\boldsymbol{V}}_i^T, & i = j, \end{cases} \tag{5.8}$$

and satisfy that

$$\boldsymbol{S}^{(1)} = \Pi_{\tilde{\boldsymbol{V}}^\perp}\boldsymbol{S}^{(1)}\Pi_{\tilde{\boldsymbol{V}}^\perp} \quad \textit{and} \quad \boldsymbol{S}_{ii}^{(2)} = \Pi_{\tilde{\boldsymbol{V}}_i}\boldsymbol{S}_{ii}^{(2)}\Pi_{\tilde{\boldsymbol{V}}_i} \quad \textit{for all } 1 \le i \le m. \tag{5.9}$$

The properties of $\boldsymbol{S}^{(1)}$ and $\boldsymbol{S}^{(2)}$ in (5.9) are exactly the same as the condition (4.1) for certificates $\boldsymbol{T}^{(1)}$ and $\boldsymbol{T}^{(2)}$ in Lemma 4.1. As a result, it is convenient to construct our certificates $\boldsymbol{T}^{(1)}$ and $\boldsymbol{T}^{(2)}$ based on $\boldsymbol{S}^{(1)}$ and $\boldsymbol{S}^{(2)}$. In fact, the explicit expression of (5.10) in step 2 shows that $\boldsymbol{T}^{(1)}$ is derived from $\boldsymbol{S}^{(1)}$ and $\boldsymbol{T}^{(2)}$ is derived from $\boldsymbol{S}^{(2)}$.

Proof of Lemma 5.1. Since $\tilde{\boldsymbol{V}}$ must satisfy the first-order local optimality condition (2.3), that is, $\tilde{\boldsymbol{V}}_i\boldsymbol{\Lambda}_i = \sum_j \boldsymbol{S}_{ij}\tilde{\boldsymbol{V}}_j$, we can construct the block diagonal matrix $\boldsymbol{S}^{(2)}$ by letting $\boldsymbol{S}_{ii}^{(2)} = \tilde{\boldsymbol{V}}_i\boldsymbol{\Lambda}_i\tilde{\boldsymbol{V}}_i^T = \sum_j \boldsymbol{S}_{ij}\tilde{\boldsymbol{V}}_j\tilde{\boldsymbol{V}}_i^T$. Then it follows that

$$\Pi_{\tilde{\boldsymbol{V}}_i}\boldsymbol{S}_{ii}^{(2)}\Pi_{\tilde{\boldsymbol{V}}_i} = \tilde{\boldsymbol{V}}_i\boldsymbol{\Lambda}_i\tilde{\boldsymbol{V}}_i^T = \boldsymbol{S}_{ii}^{(2)}.$$

Furthermore,

$$[\boldsymbol{S}^{(2)}\tilde{\boldsymbol{V}}]_i = \boldsymbol{S}_{ii}^{(2)}\tilde{\boldsymbol{V}}_i = \tilde{\boldsymbol{V}}_i\boldsymbol{\Lambda}_i = \sum_j \boldsymbol{S}_{ij}\tilde{\boldsymbol{V}}_j = [\boldsymbol{S}\tilde{\boldsymbol{V}}]_i.$$

Thus $\boldsymbol{S}^{(2)}\tilde{\boldsymbol{V}} = \boldsymbol{S}\tilde{\boldsymbol{V}}$, and for $\boldsymbol{S}^{(1)} = \boldsymbol{S} - \boldsymbol{S}^{(2)}$, we see $\boldsymbol{S}^{(1)}\tilde{\boldsymbol{V}} = 0$ and $\tilde{\boldsymbol{V}}^T\boldsymbol{S}^{(1)} = 0$ (by symmetry). This implies $\Pi_{\tilde{\boldsymbol{V}}^\perp}\boldsymbol{S}^{(1)}\Pi_{\tilde{\boldsymbol{V}}^\perp} = \boldsymbol{S}^{(1)}$. Hence condition (5.9) is satisfied. ∎

Step 2: Construction and verification of a certificate. We construct the certificates $\boldsymbol{T}^{(1)}$ and $\boldsymbol{T}^{(2)}$ based on $\boldsymbol{S}^{(1)}$ and $\boldsymbol{S}^{(2)}$ as follows:

$$\text{(5.10)} \qquad \boldsymbol{T}_{ij}^{(1)} = \begin{cases} \boldsymbol{S}_{ij}^{(1)}, & i \neq j, \\ \boldsymbol{S}_{ii}^{(1)} - c\Pi_{\tilde{\boldsymbol{V}}_i^\perp}, & i = j \end{cases}, \qquad \boldsymbol{T}_{ij}^{(2)} = \begin{cases} \boldsymbol{S}_{ij}^{(2)}, & i \neq j, \\ \boldsymbol{S}_{ii}^{(2)} - c\Pi_{\tilde{\boldsymbol{V}}_i}, & i = j. \end{cases}$$

It remains to verify that the certificate satisfies the assumptions in Lemma 4.1.

Step 2a: Proof of (4.1). From the properties of $\boldsymbol{S}^{(1)}$ and $\boldsymbol{S}^{(2)}$ from step 1, it is clear that $\boldsymbol{S} = \boldsymbol{T}^{(1)} + \boldsymbol{T}^{(2)} + c\mathbf{I}$, $\Pi_{\tilde{\boldsymbol{V}}^\perp}\boldsymbol{T}^{(1)}\Pi_{\tilde{\boldsymbol{V}}^\perp} = \boldsymbol{T}^{(1)}$, and $\boldsymbol{T}_{ii}^{(2)} = \Pi_{\tilde{\boldsymbol{V}}_i}\boldsymbol{T}_{ii}^{(2)}\Pi_{\tilde{\boldsymbol{V}}_i}$.

Step 2b: Prove that $\{P_{\tilde{\boldsymbol{V}}_i}^T\boldsymbol{T}_{ii}^{(2)}P_{\tilde{\boldsymbol{V}}_i}\}_{i=1}^m$ is positive definite. Applying, for all $1 \le i \le m$,

$$\text{(5.11)} \quad \|\tilde{\boldsymbol{V}}_i^T[\boldsymbol{S}^{(2)}]_{ii}\tilde{\boldsymbol{V}}_i - m\mathbf{I}\| \le \left\| \textstyle\sum_{j=1}^m \boldsymbol{W}_{ij}\tilde{\boldsymbol{V}}_i \right\| + m\|\tilde{\boldsymbol{V}}_i^T\boldsymbol{\Theta}_i - \mathbf{I}\| + \|\boldsymbol{\Theta}^T\tilde{\boldsymbol{V}} - m\mathbf{I}\|$$

(which will be proved in step 3) and Weyl's inequality for perturbation of eigenvalues and noting that $\|\tilde{\boldsymbol{V}}_i^T\boldsymbol{\Theta}_i - \mathbf{I}\| \le \|\tilde{\boldsymbol{V}}_i - \boldsymbol{\Theta}_i\|$, we see $P_{\tilde{\boldsymbol{V}}_i}^T\boldsymbol{T}_{ii}^{(2)}P_{\tilde{\boldsymbol{V}}_i}$ is positive definite for all $1 \le i \le m$ if

$$\text{(5.12)} \qquad m > c + \max_{1\le i\le m} \left\| \textstyle\sum_{j=1}^m \boldsymbol{W}_{ij}\tilde{\boldsymbol{V}}_i \right\| + m \max_{1\le i\le m} \|\tilde{\boldsymbol{V}}_i - \boldsymbol{\Theta}_i\| + \|\boldsymbol{\Theta}^T\tilde{\boldsymbol{V}} - m\mathbf{I}\|,$$

which follows from (4.3) with $c = m/2$.

Step 2c: Prove that $-P_{\tilde{\boldsymbol{V}}^\perp}^T\boldsymbol{T}^{(1)}P_{\tilde{\boldsymbol{V}}^\perp}$ is positive definite. Let $\mathrm{Sp}(\boldsymbol{X})$ be the column space of the matrix $\boldsymbol{X}$, and define the subspaces $L_1 = \mathrm{Sp}(\boldsymbol{\Theta})$, $L_2 = \{\boldsymbol{x} \in \mathbb{R}^D : \boldsymbol{x}_i \in \mathrm{Sp}(\boldsymbol{\Theta}_i)\}$, and $L_3 = L_2^\perp = \{\boldsymbol{x} \in \mathbb{R}^D : \boldsymbol{x}_i \in \mathrm{Sp}(\boldsymbol{\Theta}_i^\perp)\}$, and let $\boldsymbol{S}^{(1)*} = -m\Pi_{L_2\cap L_1^\perp}$ and $\boldsymbol{T}^{(1)*} = \boldsymbol{S}^{(1)*} - c\Pi_{L_3} = -m\Pi_{L_2\cap L_1^\perp} - c\Pi_{L_3}$. More specifically, we have

$$\text{(5.13)} \qquad [\boldsymbol{S}^{(1)*}]_{ij} = \boldsymbol{\Theta}_i\boldsymbol{\Theta}_j^T \text{ for } i \neq j, [\boldsymbol{S}^{(1)*}]_{ii} = -(m-1)\boldsymbol{\Theta}_i\boldsymbol{\Theta}_i^T$$

and $\boldsymbol{T}^{(1)*}$ as follows: $\boldsymbol{T}_{ij}^{(1)*} = \boldsymbol{S}_{ij}^{(1)*}$, $\boldsymbol{T}_{ii}^{(1)*} = \boldsymbol{S}_{ii}^{(1)*} - c\Pi_{\boldsymbol{\Theta}_i^\perp}$.

Considering that $\dim(L_2 \cap L_1^\perp) = \dim(L_2) - \dim(L_1) = rm - r$ and $\dim(L_3) = D - \dim(L_2) = D - rm$, we have $\lambda_{r+1}(\boldsymbol{T}^{(1)*}) = -c$. Applying Weyl's inequality and noting $\|\boldsymbol{\Theta}_i\boldsymbol{\Theta}_i^T - \tilde{\boldsymbol{V}}_i\tilde{\boldsymbol{V}}_i^T\| = \|\boldsymbol{\Theta}_i(\boldsymbol{\Theta}_i - \tilde{\boldsymbol{V}}_i)^T + (\tilde{\boldsymbol{V}}_i - \boldsymbol{\Theta}_i)\tilde{\boldsymbol{V}}_i^T\| \le 2\|\boldsymbol{\Theta}_i - \tilde{\boldsymbol{V}}_i\|$, we have

$$\begin{aligned} |\lambda_{r+1}(\boldsymbol{T}^{(1)*}) - \lambda_{r+1}(\boldsymbol{T}^{(1)})| &\le \|\boldsymbol{T}^{(1)*} - \boldsymbol{T}^{(1)}\| \\ &\le \|\boldsymbol{S}^{(1)*} - \boldsymbol{S}^{(1)}\| + c\max_{1\le i\le m}\|\Pi_{\boldsymbol{\Theta}_i^\perp} - \Pi_{\tilde{\boldsymbol{V}}_i^\perp}\| \\ &= \|\boldsymbol{S}^{(1)*} - \boldsymbol{S}^{(1)}\| + c\max_{1\le i\le m}\|\boldsymbol{\Theta}_i\boldsymbol{\Theta}_i^T - \tilde{\boldsymbol{V}}_i\tilde{\boldsymbol{V}}_i^T\| \\ &\le \|\boldsymbol{S}^{(1)*} - \boldsymbol{S}^{(1)}\| + 2c\max_{1\le i\le m}\|\boldsymbol{\Theta}_i - \tilde{\boldsymbol{V}}_i\|. \end{aligned}$$

Combining it with

$$\|\boldsymbol{S}^{(1)*} - \boldsymbol{S}^{(1)}\| \leq m \max_{1\leq i\leq m} \|\tilde{\boldsymbol{V}}_i - \boldsymbol{\Theta}_i\| + \max_{1\leq i\leq m} \|\textstyle\sum_{j=1}^m \boldsymbol{W}_{ij}\tilde{\boldsymbol{V}}_j\| + \|\boldsymbol{\Theta}^T\tilde{\boldsymbol{V}} - m\mathbf{I}\| + \|\boldsymbol{W}\| \tag{5.14}$$

(which will be proved in step 3) and

$$c > (m+2c) \max_{1\leq i\leq m} \|\tilde{\boldsymbol{V}}_i - \boldsymbol{\Theta}_i\| + \max_{1\leq i\leq m} \|\sum_{j=1}^m \boldsymbol{W}_{ij}\tilde{\boldsymbol{V}}_j\| + \|\boldsymbol{\Theta}^T\tilde{\boldsymbol{V}} - m\mathbf{I}\| + \|\boldsymbol{W}\| \tag{5.15}$$

(which follows from (4.3) with $c = m/2$), $\lambda_{r+1}(\boldsymbol{T}^{(1)})$ is negative, which means that $\boldsymbol{T}^{(1)}$ has at least $D - r$ negative eigenvalues. By definition, $\boldsymbol{T}^{(1)}$ has r zero eigenvalues with eigenvectors spanning the column space of $\tilde{\boldsymbol{V}}$, so $P_{\tilde{\boldsymbol{V}}^\perp}^T \boldsymbol{T}^{(1)} P_{\tilde{\boldsymbol{V}}^\perp}$ is negative definite.

Step 3: Proof of auxiliary inequalities (5.11) and (5.14).

Step 3a: Proof of (5.11). Combining (5.8) with

$$\textstyle\sum_{j=1}^m \boldsymbol{\Theta}_j^T\tilde{\boldsymbol{V}}_j = \boldsymbol{\Theta}^T\tilde{\boldsymbol{V}}, \tag{5.16}$$

we see

$$\begin{aligned}
\left\|\tilde{\boldsymbol{V}}_i^T[\boldsymbol{S}^{(2)}]_{ii}\tilde{\boldsymbol{V}}_i - m\mathbf{I}\right\| &= \left\|\tilde{\boldsymbol{V}}_i^T\left(\textstyle\sum_{j=1}^m \boldsymbol{S}_{ij}\tilde{\boldsymbol{V}}_j\right) - m\mathbf{I}\right\| \\
&\leq \left\|\textstyle\sum_{j=1}^m \boldsymbol{W}_{ij}\tilde{\boldsymbol{V}}_j\right\| + \left\|\tilde{\boldsymbol{V}}_i^T\boldsymbol{\Theta}_i\left(\textstyle\sum_{j=1}^m \boldsymbol{\Theta}_j^T\tilde{\boldsymbol{V}}_j\right) - m\mathbf{I}\right\| \\
&\leq \left\|\textstyle\sum_{j=1}^m \boldsymbol{W}_{ij}\tilde{\boldsymbol{V}}_i\right\| + \left\|\tilde{\boldsymbol{V}}_i^T\boldsymbol{\Theta}_i - \mathbf{I}\right\|\left\|\textstyle\sum_{j=1}^m \boldsymbol{\Theta}_j^T\tilde{\boldsymbol{V}}_j\right\| + \left\|\textstyle\sum_{j=1}^m \boldsymbol{\Theta}_j^T\tilde{\boldsymbol{V}}_j - m\mathbf{I}\right\| \\
&\leq \left\|\textstyle\sum_{j=1}^m \boldsymbol{W}_{ij}\tilde{\boldsymbol{V}}_i\right\| + m\left\|\tilde{\boldsymbol{V}}_i^T\boldsymbol{\Theta}_i - \mathbf{I}\right\| + \left\|\boldsymbol{\Theta}^T\tilde{\boldsymbol{V}} - m\mathbf{I}\right\|,
\end{aligned}$$

where $\boldsymbol{S}_{ij} = \boldsymbol{W}_{ij} + \boldsymbol{\Theta}_i\boldsymbol{\Theta}_j^T$ when $i \neq j$ is used for the first inequality.

Step 3b: Proof of (5.14). Applying (5.7), (5.8), and (5.13), we have that, for both MAXBET and MAXDIFF models,

$$[\boldsymbol{S}^{(1)} - \boldsymbol{S}^{(1)*}]_{ij} = \begin{cases} \boldsymbol{S}_{ij} - \boldsymbol{\Theta}_i\boldsymbol{\Theta}_j^T = \boldsymbol{W}_{ij}, & i \neq j, \\ \boldsymbol{W}_{ii} - (\sum_{j=1}^m \boldsymbol{S}_{ij}\tilde{\boldsymbol{V}}_j)\tilde{\boldsymbol{V}}_i^T + m\boldsymbol{\Theta}_i\boldsymbol{\Theta}_i^T, & i = j. \end{cases} \tag{5.17}$$

As a result,

$$\|\boldsymbol{S}^{(1)} - \boldsymbol{S}^{(1)*}\| \leq \|\boldsymbol{W}\| + \max_{1\leq i\leq m} \left\|\left(\textstyle\sum_{j=1}^m \boldsymbol{S}_{ij}\tilde{\boldsymbol{V}}_j\right)\tilde{\boldsymbol{V}}_i^T - m\boldsymbol{\Theta}_i\boldsymbol{\Theta}_i^T\right\|. \tag{5.18}$$

Using (5.16), we have

$$\begin{aligned}
&\|(\textstyle\sum_{j=1}^m \boldsymbol{\Theta}_i\boldsymbol{\Theta}_j^T\tilde{\boldsymbol{V}}_j)\tilde{\boldsymbol{V}}_i^T - m\boldsymbol{\Theta}_i\boldsymbol{\Theta}_i^T\| = \|(\textstyle\sum_{j=1}^m \boldsymbol{\Theta}_j^T\tilde{\boldsymbol{V}}_j)\tilde{\boldsymbol{V}}_i^T - m\boldsymbol{\Theta}_i^T\| \\
&\leq \|(\textstyle\sum_{j=1}^m \boldsymbol{\Theta}_j^T\tilde{\boldsymbol{V}}_j) - m\mathbf{I}\| + m\|\tilde{\boldsymbol{V}}_i - \boldsymbol{\Theta}_i\| = \|\boldsymbol{\Theta}^T\tilde{\boldsymbol{V}} - m\mathbf{I}\| + m\|\tilde{\boldsymbol{V}}_i - \boldsymbol{\Theta}_i\|.
\end{aligned} \tag{5.19}$$

Applying (5.18), (5.19), and $\boldsymbol{S}_{ij} = \boldsymbol{W}_{ij} + \boldsymbol{\Theta}_i\boldsymbol{\Theta}_j^T$ when $i \neq j$, (5.14) is proved. ∎

Proof of Lemma 4.3. First, we remark that the choice of $\tilde{\boldsymbol{V}} \in \mathbb{R}^{D\times r}$ is only unique up to an $r \times r$ orthogonal matrix. That is, for any orthogonal matrix $\boldsymbol{O} \in \mathbb{R}^{r\times r}$, $\tilde{\boldsymbol{V}}\boldsymbol{O}$ is also a potential choice. In this proof, we choose $\tilde{\boldsymbol{V}}$ such that $\boldsymbol{\Theta}^T\tilde{\boldsymbol{V}} \in \mathbb{R}^{r\times r}$ is a symmetric, positive semidefinite matrix, and as a result, $\operatorname{tr}(\boldsymbol{\Theta}^T\tilde{\boldsymbol{V}}) = \|\boldsymbol{\Theta}^T\tilde{\boldsymbol{V}}\|_*$.

Then we have that

$$
\begin{aligned}
\|\tilde{\boldsymbol{V}} - \boldsymbol{\Theta}\|_F^2 &= \textstyle\sum_{i=1}^m \|\tilde{\boldsymbol{V}}_i - \boldsymbol{\Theta}_i\|_F^2 = \sum_{i=1}^m \|\tilde{\boldsymbol{V}}_i\|_F^2 + \|\boldsymbol{\Theta}_i\|_F^2 - 2\operatorname{tr}(\tilde{\boldsymbol{V}}_i\boldsymbol{\Theta}_i^T) \\
&= \sum_{i=1}^m \|\tilde{\boldsymbol{V}}_i\|_F^2 + \|\boldsymbol{\Theta}_i\|_F^2 - 2\operatorname{tr}(\boldsymbol{\Theta}_i^T\tilde{\boldsymbol{V}}_i) = 2rm - 2\operatorname{tr}\left(\sum_{i=1}^m \boldsymbol{\Theta}_i^T\tilde{\boldsymbol{V}}_i\right) \\
&= 2rm - 2\operatorname{tr}(\boldsymbol{\Theta}^T\tilde{\boldsymbol{V}}) = 2rm - 2\|\boldsymbol{\Theta}^T\tilde{\boldsymbol{V}}\|_*,
\end{aligned}
\tag{5.20}
$$

where $\|\cdot\|_*$ represents the nuclear norm that is the summation of all singular values (and since $\boldsymbol{V}^T\tilde{\boldsymbol{V}}$ is positive semidefinite, it is also the summation of its eigenvalues).

Using the definition in (2.2), we have

$$
\operatorname{tr}(\tilde{\boldsymbol{V}}^T\boldsymbol{S}\tilde{\boldsymbol{V}}) \geq \operatorname{tr}(\boldsymbol{\Theta}^T\boldsymbol{S}\boldsymbol{\Theta}). \tag{5.21}
$$

Applying $\boldsymbol{S} = \boldsymbol{\Theta}\boldsymbol{\Theta}^T + \boldsymbol{W}$, (5.21) implies

$$
\begin{aligned}
\operatorname{tr}(\tilde{\boldsymbol{V}}^T\boldsymbol{W}\tilde{\boldsymbol{V}}) + \|\tilde{\boldsymbol{V}}^T\boldsymbol{\Theta}\|_F^2 &= \operatorname{tr}(\tilde{\boldsymbol{V}}^T\boldsymbol{W}\tilde{\boldsymbol{V}}) + \operatorname{tr}(\tilde{\boldsymbol{V}}^T\boldsymbol{\Theta}\boldsymbol{\Theta}^T\tilde{\boldsymbol{V}}) \\
&\geq \operatorname{tr}(\boldsymbol{\Theta}^T\boldsymbol{W}\boldsymbol{\Theta}) + \operatorname{tr}(\boldsymbol{\Theta}^T\boldsymbol{\Theta}\boldsymbol{\Theta}^T\boldsymbol{\Theta}) = \operatorname{tr}(\boldsymbol{\Theta}^T\boldsymbol{W}\boldsymbol{\Theta}) + \|\boldsymbol{\Theta}^T\boldsymbol{\Theta}\|_F^2
\end{aligned}
$$

and

$$
\operatorname{tr}(\tilde{\boldsymbol{V}}^T\boldsymbol{W}\tilde{\boldsymbol{V}}) - \operatorname{tr}(\boldsymbol{\Theta}^T\boldsymbol{W}\boldsymbol{\Theta}) \geq \|\boldsymbol{\Theta}^T\boldsymbol{\Theta}\|_F^2 - \|\tilde{\boldsymbol{V}}^T\boldsymbol{\Theta}\|_F^2 = rm^2 - \|\tilde{\boldsymbol{V}}^T\boldsymbol{\Theta}\|_F^2. \tag{5.22}
$$

Since $\|\boldsymbol{X}\|_F^2 = \sum_i \lambda_i(\boldsymbol{X})^2$, we have

$$
\begin{aligned}
rm^2 - \|\tilde{\boldsymbol{V}}^T\boldsymbol{\Theta}\|_F^2 &= \textstyle\sum_{i=1}^r (m^2 - \lambda_i(\tilde{\boldsymbol{V}}^T\boldsymbol{\Theta})^2) \\
&\geq m \textstyle\sum_{i=1}^r (m - \lambda_i(\tilde{\boldsymbol{V}}^T\boldsymbol{\Theta})) = m(rm - \|\tilde{\boldsymbol{V}}^T\boldsymbol{\Theta}\|_*).
\end{aligned}
\tag{5.23}
$$

The combination of (5.22), (5.23), $\|\tilde{\boldsymbol{V}}\|_F = \|\boldsymbol{\Theta}\|_F = \sqrt{rm}$, $\operatorname{tr}(\boldsymbol{A}\boldsymbol{B}) \leq \|\boldsymbol{A}\|_F\|\boldsymbol{B}\|_F$, and $\|\boldsymbol{A}\boldsymbol{B}\|_F \leq \|\boldsymbol{A}\|\|\boldsymbol{B}\|_F$ implies that

$$
\begin{aligned}
m(rm - \|\tilde{\boldsymbol{V}}^T\boldsymbol{\Theta}\|_*) &\leq \operatorname{tr}(\tilde{\boldsymbol{V}}^T\boldsymbol{W}\tilde{\boldsymbol{V}}) - \operatorname{tr}(\boldsymbol{\Theta}^T\boldsymbol{W}\boldsymbol{\Theta}) \\
&= \operatorname{tr}((\tilde{\boldsymbol{V}} - \boldsymbol{\Theta})^T\boldsymbol{W}\tilde{\boldsymbol{V}}) + \operatorname{tr}(\boldsymbol{\Theta}^T\boldsymbol{W}(\tilde{\boldsymbol{V}} - \boldsymbol{\Theta})) \\
&\leq \|\boldsymbol{W}\|\|\tilde{\boldsymbol{V}} - \boldsymbol{\Theta}\|_F\|\tilde{\boldsymbol{V}}\|_F + \|\boldsymbol{W}\|\|\tilde{\boldsymbol{V}} - \boldsymbol{\Theta}\|_F\|\boldsymbol{\Theta}\|_F \\
&= 2\|\boldsymbol{W}\|\|\tilde{\boldsymbol{V}} - \boldsymbol{\Theta}\|_F\sqrt{rm}.
\end{aligned}
$$

Combining it with (5.20), we have

$$
\tfrac{m}{2}\|\tilde{\boldsymbol{V}} - \boldsymbol{\Theta}\|_F^2 \leq 2\|\boldsymbol{W}\|\|\tilde{\boldsymbol{V}} - \boldsymbol{\Theta}\|_F\sqrt{rm}, \tag{5.24}
$$

which implies

$$
\|\tilde{\boldsymbol{V}} - \boldsymbol{\Theta}\|_F \leq 4\|\boldsymbol{W}\|\sqrt{\tfrac{r}{m}}, \tag{5.25}
$$

proving the first inequality in (4.4). It implies that

$$
\|\tilde{\boldsymbol{V}}^T\boldsymbol{\Theta} - m\mathbf{I}\|_F = \|(\tilde{\boldsymbol{V}} - \boldsymbol{\Theta})^T\boldsymbol{\Theta}\|_F \leq \|\tilde{\boldsymbol{V}} - \boldsymbol{\Theta}\|_F\sqrt{m} \leq 4\|\boldsymbol{W}\|\sqrt{r}. \tag{5.26}
$$

Applying (5.25), the second inequality in (4.4) is proved:

$$
\begin{aligned}
\max_{1\le i\le m}\|[\boldsymbol{W}\tilde{\boldsymbol{V}}]_i\|_F &\le \max_{1\le i\le m}\|[\boldsymbol{W}\boldsymbol{\Theta}]_i\|_F + \max_{1\le i\le m}\|[\boldsymbol{W}(\tilde{\boldsymbol{V}}-\boldsymbol{\Theta})]_i\|_F \\
&\le \max_{1\le i\le m}\|[\boldsymbol{W}\boldsymbol{\Theta}]_i\|_F + \|\boldsymbol{W}\|\|\tilde{\boldsymbol{V}}-\boldsymbol{\Theta}\|_F \\
&\le \max_{1\le i\le m}\|[\boldsymbol{W}\boldsymbol{\Theta}]_i\|_F + 4\|\boldsymbol{W}\|^2\sqrt{\tfrac{r}{m}}.
\end{aligned}
\tag{5.27}
$$

Now let us consider $\bar{\boldsymbol{V}}\in\mathbb{R}^{D\times r}$ defined by $\bar{\boldsymbol{V}}_i=\boldsymbol{\Theta}_i$ and $\bar{\boldsymbol{V}}_j=\tilde{\boldsymbol{V}}_j$ for all $1\le j\le m, j\ne i$. By definition we have $\mathrm{tr}(\tilde{\boldsymbol{V}}^T\boldsymbol{S}\tilde{\boldsymbol{V}})\ge\mathrm{tr}(\bar{\boldsymbol{V}}^T\boldsymbol{S}\bar{\boldsymbol{V}})$, and it is equivalent to $\mathrm{tr}((\tilde{\boldsymbol{V}}-\bar{\boldsymbol{V}})^T\boldsymbol{S}\tilde{\boldsymbol{V}})+\mathrm{tr}(\tilde{\boldsymbol{V}}^T\boldsymbol{S}(\tilde{\boldsymbol{V}}-\bar{\boldsymbol{V}}))-\mathrm{tr}((\tilde{\boldsymbol{V}}-\bar{\boldsymbol{V}})^T\boldsymbol{S}(\tilde{\boldsymbol{V}}-\bar{\boldsymbol{V}}))\ge 0$. By the definition of $\bar{\boldsymbol{V}}$, $\tilde{\boldsymbol{V}}$, and $\boldsymbol{S}$, we have

$$
\begin{aligned}
2\,\mathrm{tr}((\tilde{\boldsymbol{V}}_i-\boldsymbol{\Theta}_i)^T\boldsymbol{\Theta}_i\boldsymbol{\Theta}^T\tilde{\boldsymbol{V}}) &+ 2\,\mathrm{tr}((\tilde{\boldsymbol{V}}_i-\boldsymbol{\Theta}_i)^T[\boldsymbol{W}\tilde{\boldsymbol{V}}]_i) \\
&- \mathrm{tr}((\tilde{\boldsymbol{V}}_i-\boldsymbol{\Theta}_i)^T\boldsymbol{S}_{ii}(\tilde{\boldsymbol{V}}_i-\boldsymbol{\Theta}_i))\ge 0.
\end{aligned}
\tag{5.28}
$$

Recall that $\tilde{\boldsymbol{V}}$ is chosen such that $\boldsymbol{\Theta}^T\tilde{\boldsymbol{V}}$ is symmetric and positive semidefinite, and apply the fact that, when $\boldsymbol{A}$ is positive semidefinite, then $\mathrm{tr}(\boldsymbol{B}\boldsymbol{A})=\mathrm{tr}(\boldsymbol{B}^T\boldsymbol{A})$; and when both $\boldsymbol{A},\boldsymbol{B}$ are positive semidefinite, $\mathrm{tr}(\boldsymbol{A}\boldsymbol{B})\ge\mathrm{tr}(\boldsymbol{A}\lambda_{\min}(\boldsymbol{B})\mathbf{I})\ge\lambda_{\min}(\boldsymbol{B})\,\mathrm{tr}(\boldsymbol{A})$ ($\lambda_{\min}$ represents the smallest eigenvalue), we have

$$
\begin{aligned}
&\mathrm{tr}\big[(\boldsymbol{\Theta}_i-\tilde{\boldsymbol{V}}_i)^T\boldsymbol{\Theta}_i\boldsymbol{\Theta}^T\tilde{\boldsymbol{V}}\big]=\mathrm{tr}\left[(\mathbf{I}-\tilde{\boldsymbol{V}}_i^T\boldsymbol{\Theta}_i)(\boldsymbol{\Theta}^T\tilde{\boldsymbol{V}})\right] \\
&\quad=\tfrac12\,\mathrm{tr}\left[(2\mathbf{I}-\tilde{\boldsymbol{V}}_i^T\boldsymbol{\Theta}_i-\boldsymbol{\Theta}_i^T\tilde{\boldsymbol{V}}_i)(\boldsymbol{\Theta}^T\tilde{\boldsymbol{V}})\right]=\tfrac12\,\mathrm{tr}\left[(\tilde{\boldsymbol{V}}_i-\boldsymbol{\Theta}_i)^T(\tilde{\boldsymbol{V}}_i-\boldsymbol{\Theta}_i)(\boldsymbol{\Theta}^T\tilde{\boldsymbol{V}})\right] \\
&\quad\ge\tfrac12\,\mathrm{tr}\left[(\tilde{\boldsymbol{V}}_i-\boldsymbol{\Theta}_i)^T(\tilde{\boldsymbol{V}}_i-\boldsymbol{\Theta}_i)\right]\lambda_r(\boldsymbol{\Theta}^T\tilde{\boldsymbol{V}})=\tfrac12\|\tilde{\boldsymbol{V}}_i-\boldsymbol{\Theta}_i\|_F^2\lambda_r(\boldsymbol{\Theta}^T\tilde{\boldsymbol{V}}).
\end{aligned}
\tag{5.29}
$$

In addition, we have

$$
\mathrm{tr}((\tilde{\boldsymbol{V}}_i-\boldsymbol{\Theta}_i)^T\boldsymbol{S}_{ii}(\tilde{\boldsymbol{V}}_i-\boldsymbol{\Theta}_i))\ge-\|\boldsymbol{S}_{ii}\|\|\tilde{\boldsymbol{V}}_i-\boldsymbol{\Theta}_i\|_F^2\ge-(1+\|\boldsymbol{W}_{ii}\|)\|\tilde{\boldsymbol{V}}_i-\boldsymbol{\Theta}_i\|_F^2,
\tag{5.30}
$$

and $\mathrm{tr}(\boldsymbol{A}\boldsymbol{B})\le\|\boldsymbol{A}\|_F\|\boldsymbol{B}\|_F$ implies

$$
\mathrm{tr}((\tilde{\boldsymbol{V}}_i-\boldsymbol{\Theta}_i)^T[\boldsymbol{W}\tilde{\boldsymbol{V}}]_i)\le\|\tilde{\boldsymbol{V}}_i-\boldsymbol{\Theta}_i\|_F\|\boldsymbol{W}^T\tilde{\boldsymbol{V}}\|_F.
\tag{5.31}
$$

Combining (5.28), (5.29), (5.30), and (5.31),

$$
\|\tilde{\boldsymbol{V}}_i-\boldsymbol{\Theta}_i\|_F\|[\boldsymbol{W}\tilde{\boldsymbol{V}}]_i\|_F\ge\|\tilde{\boldsymbol{V}}_i-\boldsymbol{\Theta}_i\|_F^2(\lambda_r(\boldsymbol{\Theta}^T\tilde{\boldsymbol{V}})-1-\|\boldsymbol{W}_{ii}\|).
$$

Combining it with (5.27) and (5.26), which implies that $\lambda_r(\boldsymbol{\Theta}^T\tilde{\boldsymbol{V}})\ge m-4\|\boldsymbol{W}\|\sqrt{r}$, and noting that $\|\boldsymbol{W}\|\ge\|\boldsymbol{W}_{ii}\|$, (4.5) is proved. ∎

5.2.1. Lemma 4.3 under (MAXDIFF).

Proof of Lemma 4.3 *under* (MAXDIFF). Following the proof of Lemma 4.3 under (MAXBET), we have

$$
\begin{aligned}
2\|\boldsymbol{W}\|\|\tilde{\boldsymbol{V}}-\boldsymbol{\Theta}\|_F\sqrt{rm} &\ge \mathrm{tr}(\tilde{\boldsymbol{V}}^T\boldsymbol{W}\tilde{\boldsymbol{V}})-\mathrm{tr}(\boldsymbol{\Theta}^T\boldsymbol{W}\boldsymbol{\Theta}) \\
&\ge (rm^2-\|\tilde{\boldsymbol{V}}^T\boldsymbol{\Theta}\|_F^2)-(rm-\textstyle\sum_{i=1}^m\|\tilde{\boldsymbol{V}}_i^T\boldsymbol{\Theta}_i\|_F^2) \\
&\ge \tfrac{m}{2}\|\tilde{\boldsymbol{V}}-\boldsymbol{\Theta}\|_F^2-\textstyle\sum_{i=1}^m\|\tilde{\boldsymbol{V}}_i-\boldsymbol{\Theta}_i\|_F^2=(\tfrac{m}{2}-1)\|\tilde{\boldsymbol{V}}-\boldsymbol{\Theta}\|_F^2,
\end{aligned}
$$

where the first inequality is (5.23), the second inequality is from the definition of $\boldsymbol{S}$ under the MAXDIFF setting, and the third inequality is from $r-\|\tilde{\boldsymbol{V}}_i^T\boldsymbol{\Theta}_i\|_F^2\le\|\tilde{\boldsymbol{V}}_i-\boldsymbol{\Theta}_i\|_F^2=2r-\mathrm{tr}(\tilde{\boldsymbol{V}}_i^T\boldsymbol{\Theta}_i)$ since $\|\tilde{\boldsymbol{V}}_i^T\boldsymbol{\Theta}\|_F^2-2\,\mathrm{tr}(\tilde{\boldsymbol{V}}_i^T\boldsymbol{\Theta})+r=\|\tilde{\boldsymbol{V}}_i^T\boldsymbol{\Theta}_i-\mathbf{I}_r\|_F^2$.

As a result, if $m > 2$,

$$\|\tilde{\boldsymbol{V}} - \boldsymbol{\Theta}\|_F \leq 4\|\boldsymbol{W}\|\tfrac{\sqrt{rm}}{m-2},\ \|\tilde{\boldsymbol{V}}^T\boldsymbol{\Theta} - m\mathbf{I}\|_F \leq 4\|\boldsymbol{W}\|\tfrac{m\sqrt{r}}{m-2},$$
$$\max_{1\leq i\leq m} \|[\boldsymbol{W}\tilde{\boldsymbol{V}}]_i\|_F \leq \max_{1\leq i\leq m} \|[\boldsymbol{W}\boldsymbol{\Theta}]_i\|_F + 4\|\boldsymbol{W}\|^2\tfrac{\sqrt{rm}}{m-2}.$$

In addition, (5.28) is replaced with $2\,\mathrm{tr}((\tilde{\boldsymbol{V}}_i - \boldsymbol{\Theta}_i)^T\boldsymbol{\Theta}_i\boldsymbol{\Theta}^T\tilde{\boldsymbol{V}}) + 2\,\mathrm{tr}((\tilde{\boldsymbol{V}}_i - \boldsymbol{\Theta}_i)^T[\boldsymbol{W}\tilde{\boldsymbol{V}}]_i) \geq 0$. Then we have $\frac{1}{2}\lambda_r(\boldsymbol{\Theta}^T\tilde{\boldsymbol{V}})\|\boldsymbol{\Theta}_i - \tilde{\boldsymbol{V}}_i\|_F^2 \leq \|\boldsymbol{\Theta}_i - \tilde{\boldsymbol{V}}_i\|_F(\|[\boldsymbol{W}\tilde{\boldsymbol{V}}]_i\|_F)$ and

$$\max_{1\leq i\leq m} \|\boldsymbol{\Theta}_i - \tilde{\boldsymbol{V}}_i\|_F \leq \frac{2\max_{1\leq i\leq m}\|[\boldsymbol{W}\boldsymbol{\Theta}]_i\|_F + 8\|\boldsymbol{W}\|^2\frac{\sqrt{rm}}{m-2}}{m - 4\|\boldsymbol{W}\|\frac{m\sqrt{r}}{m-2}}$$

for $m > 4\|\boldsymbol{W}\|\sqrt{r} + 2$. ◻

5.3. Proof of lemmas for Theorem 3.9.

Proof of Lemma 4.4. From $\boldsymbol{O}_i\boldsymbol{\Lambda}_i = \sum_{j=1}^m \boldsymbol{S}_{ij}\boldsymbol{O}_j$, we have $\boldsymbol{\Lambda}_i = \sum_{j=1}^m \boldsymbol{O}_i^T\boldsymbol{S}_{ij}\boldsymbol{O}_j$. Hence, under (MAXBET),

$$\begin{aligned}
\|\boldsymbol{\Lambda}_i - m\mathbf{I}_r\| &= \|\textstyle\sum_{j=1}^m \boldsymbol{O}_i^T\boldsymbol{S}_{ij}\boldsymbol{O}_j - m\mathbf{I}_r\| \\
&\leq \|\boldsymbol{O}_i^T\textstyle\sum_{j=1}^m \boldsymbol{W}_{ij}\boldsymbol{O}_j\| + \|\boldsymbol{O}_i^T\boldsymbol{\Theta}_i\textstyle\sum_{j=1}^m\boldsymbol{\Theta}_j^T\boldsymbol{O}_j - m\mathbf{I}_r\| \\
&\leq \|\textstyle\sum_{j=1}^m \boldsymbol{W}_{ij}\boldsymbol{O}_j\| + \|(\boldsymbol{O}_i^T\boldsymbol{\Theta}_i - \mathbf{I}_r)\textstyle\sum_{j=1}^m\boldsymbol{\Theta}_j^T\boldsymbol{O}_j + \textstyle\sum_{j=1}^m\boldsymbol{\Theta}_j^T\boldsymbol{O}_j - m\mathbf{I}_r\| \\
&\leq \|\sum_{j=1}^m \boldsymbol{W}_{ij}\boldsymbol{O}_j\| + \|(\boldsymbol{O}_i^T\boldsymbol{\Theta}_i - \mathbf{I}_r)\sum_{j=1}^m\boldsymbol{\Theta}_j^T\boldsymbol{O}_j\| + \|\sum_{j=1}^m\boldsymbol{\Theta}_j^T\boldsymbol{O}_j - m\mathbf{I}_r\| \\
&\leq \|\textstyle\sum_{j=1}^m \boldsymbol{W}_{ij}\boldsymbol{O}_j\| + \|\boldsymbol{O}_i^T\boldsymbol{\Theta}_i - \mathbf{I}_r\|\|\textstyle\sum_{j=1}^m\boldsymbol{\Theta}_j^T\boldsymbol{O}_j\| + \|\boldsymbol{\Theta}^T\boldsymbol{O} - m\mathbf{I}_r\| \\
&\leq \|\textstyle\sum_{j=1}^m \boldsymbol{W}_{ij}\boldsymbol{O}_j\| + m\|\boldsymbol{O}_i^T\boldsymbol{\Theta}_i - \mathbf{I}_r\| + \|\boldsymbol{\Theta}^T\boldsymbol{O} - m\mathbf{I}_r\|
\end{aligned}$$

since $\|\boldsymbol{\Theta}_j^T\boldsymbol{O}_j\| \leq 1$. Under the MAXDIFF model,

$$\begin{aligned}
&\|\boldsymbol{\Lambda}_i - (m-1)\mathbf{I}_r\| = \|\textstyle\sum_{j\neq i}\boldsymbol{O}_i^T\boldsymbol{S}_{ij}\boldsymbol{O}_j - (m-1)\mathbf{I}_r\| \\
&\leq \|\boldsymbol{O}_i^T\textstyle\sum_{j\neq i}\boldsymbol{W}_{ij}\boldsymbol{O}_j\| + \|\boldsymbol{O}_i^T\boldsymbol{\Theta}_i\textstyle\sum_{j\neq i}\boldsymbol{\Theta}_j^T\boldsymbol{O}_j - (m-1)\mathbf{I}_r\| \\
&\leq \|\textstyle\sum_{j\neq i}\boldsymbol{W}_{ij}\boldsymbol{O}_j\| + \|(\boldsymbol{O}_i^T\boldsymbol{\Theta}_i - \mathbf{I}_r)\textstyle\sum_{j=1}^m\boldsymbol{\Theta}_j^T\boldsymbol{O}_j + \textstyle\sum_{j\neq i}\boldsymbol{\Theta}_j^T\boldsymbol{O}_j - (m-1)\mathbf{I}_r\| \\
&\leq \|\textstyle\sum_{j\neq i}\boldsymbol{W}_{ij}\boldsymbol{O}_j\| + \|(\boldsymbol{O}_i^T\boldsymbol{\Theta}_i - \mathbf{I}_r)\textstyle\sum_{j\neq i}\boldsymbol{\Theta}_j^T\boldsymbol{O}_j\| + \|\textstyle\sum_{j\neq i}\boldsymbol{\Theta}_j^T\boldsymbol{O}_j - (m-1)\mathbf{I}_r\| \\
&\leq \left\|\sum_{j\neq i}\boldsymbol{W}_{ij}\boldsymbol{O}_j\right\| + \left\|\boldsymbol{O}_i^T\boldsymbol{\Theta}_i - \mathbf{I}_r\right\|\left\|\sum_{j\neq i}\boldsymbol{\Theta}_j^T\boldsymbol{O}_j\right\| + \|\boldsymbol{\Theta}^T\boldsymbol{O} - m\mathbf{I}_r - \boldsymbol{\Theta}_i^T\boldsymbol{O}_i + \mathbf{I}_r\| \\
&\leq \|\textstyle\sum_{j\neq i}\boldsymbol{W}_{ij}\boldsymbol{O}_j\| + (m-1)\|\boldsymbol{O}_i^T\boldsymbol{\Theta}_i - \mathbf{I}_r\| + \|\boldsymbol{\Theta}^T\boldsymbol{O} - m\mathbf{I}_r\| + \|\boldsymbol{O}_i^T\boldsymbol{\Theta}_i - \mathbf{I}_r\| \\
&= \|\textstyle\sum_{j\neq i}\boldsymbol{W}_{ij}\boldsymbol{O}_j\| + m\|\boldsymbol{O}_i^T\boldsymbol{\Theta}_i - \mathbf{I}_r\| + \|\boldsymbol{\Theta}^T\boldsymbol{O} - m\mathbf{I}_r\|.
\end{aligned}$$
◻

The following technical lemma is needed to prove Lemma 4.5.

LEMMA 5.2. *Suppose $\boldsymbol{X}, \boldsymbol{Y} \in \mathcal{O}_{d,r}$, and $\boldsymbol{\Lambda} \in \mathbb{R}^{d\times d}$ is symmetric and positive semidefinite. Then, there holds* $\mathrm{tr}[\boldsymbol{X}\boldsymbol{\Lambda}(\boldsymbol{Y} - \boldsymbol{X})] \leq 0$.

Proof. Note

$$\begin{aligned}
\mathrm{tr}[\boldsymbol{X}\boldsymbol{\Lambda}(\boldsymbol{Y} - \boldsymbol{X})] &\leq \mathrm{tr}(\boldsymbol{\Lambda}^T\boldsymbol{X}^T(\boldsymbol{Y} - \boldsymbol{X})) = \mathrm{tr}(\boldsymbol{\Lambda}(\boldsymbol{X}^T\boldsymbol{Y} - \mathbf{I}_r)) \\
&= \mathrm{tr}(\boldsymbol{\Lambda}\boldsymbol{X}^T\boldsymbol{Y}) - \mathrm{tr}(\boldsymbol{\Lambda}) = \mathrm{tr}(\boldsymbol{Y}^T\boldsymbol{X}\boldsymbol{\Lambda}) - \mathrm{tr}(\boldsymbol{\Lambda}) \\
&= \mathrm{tr}(\boldsymbol{\Lambda}\boldsymbol{Y}^T\boldsymbol{X}) - \mathrm{tr}(\boldsymbol{\Lambda}) = \mathrm{tr}\left[\boldsymbol{\Lambda}\left(\tfrac{1}{2}\boldsymbol{X}^T\boldsymbol{Y} + \tfrac{1}{2}\boldsymbol{Y}^T\boldsymbol{X} - \mathbf{I}_r\right)\right].
\end{aligned}$$

Since $\boldsymbol{X}\boldsymbol{X}^T \preccurlyeq \mathbf{I}_d$, $(\boldsymbol{X}^T\boldsymbol{Y})^T(\boldsymbol{X}^T\boldsymbol{Y}) = \boldsymbol{Y}^T\boldsymbol{X}\boldsymbol{X}^T\boldsymbol{Y} \preccurlyeq \boldsymbol{Y}^T\boldsymbol{Y} = \mathbf{I}_r$. Thus $\|\boldsymbol{X}^T\boldsymbol{Y}\|_2 \le 1$. Likewise $\|\boldsymbol{Y}^T\boldsymbol{X}\|_2 \le 1$. Then, because $\frac{1}{2}\boldsymbol{X}^T\boldsymbol{Y} + \frac{1}{2}\boldsymbol{Y}^T\boldsymbol{X}$ is symmetric,

$$\lambda_{\max}\left(\tfrac{1}{2}\boldsymbol{X}^T\boldsymbol{Y} + \tfrac{1}{2}\boldsymbol{Y}^T\boldsymbol{X}\right) \le \left\|\tfrac{1}{2}\boldsymbol{X}^T\boldsymbol{Y} + \tfrac{1}{2}\boldsymbol{Y}^T\boldsymbol{X}\right\| \le \tfrac{1}{2}\|\boldsymbol{X}^T\boldsymbol{Y}\|_2 + \tfrac{1}{2}\|\boldsymbol{Y}^T\boldsymbol{X}\| \le 1$$

and $\frac{1}{2}\boldsymbol{X}^T\boldsymbol{Y} + \frac{1}{2}\boldsymbol{Y}^T\boldsymbol{X} - \mathbf{I}_r \preccurlyeq \mathbf{0}$. Since $\boldsymbol{\Lambda} \succcurlyeq \mathbf{0}$, it follows that $\operatorname{tr}[\boldsymbol{X}\boldsymbol{\Lambda}(\boldsymbol{Y}-\boldsymbol{X})] \le 0$. ◻

Proof of Lemma 4.5. Let the singular value decomposition of $\boldsymbol{\Theta}^T\tilde{\boldsymbol{O}}$ be $\boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^T$, where $\boldsymbol{U}, \boldsymbol{V} \in \mathcal{O}_{r,r}$ and $\boldsymbol{\Sigma} \in \mathbb{R}^{r\times r}$ is diagonal with nonnegative entries. Let $\boldsymbol{R} = \boldsymbol{V}\boldsymbol{U}^T \in \mathcal{O}_{r\times r}$. Then, for $\boldsymbol{O} = \tilde{\boldsymbol{O}}\boldsymbol{R}$, it holds $\boldsymbol{\Theta}^T\boldsymbol{O} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{U}^T \succcurlyeq \mathbf{0}$.

Clearly, $(\boldsymbol{O}_1, \ldots, \boldsymbol{O}_m) = (\tilde{\boldsymbol{O}}_1\boldsymbol{R}, \ldots, \tilde{\boldsymbol{O}}_m\boldsymbol{R})$ is globally optimal. Therefore,

$$\operatorname{tr}(\boldsymbol{\Theta}^T\boldsymbol{S}\boldsymbol{\Theta}) \le \operatorname{tr}(\boldsymbol{O}^T\boldsymbol{S}\boldsymbol{O}),$$

which is similar to inequality (5.21) in the proof of Lemma 4.3. It immediately follows that, under (MAXBET),

$$2\sqrt{mr}\|\boldsymbol{W}\|\|\boldsymbol{O}-\boldsymbol{\Theta}\|_F \ge \operatorname{tr}(\boldsymbol{O}^T\boldsymbol{W}\boldsymbol{O}) - \operatorname{tr}(\boldsymbol{\Theta}^T\boldsymbol{W}\boldsymbol{\Theta}) \ge \tfrac{m}{2}\|\boldsymbol{O}-\boldsymbol{\Theta}\|_F^2$$

and, under (MAXDIFF),

$$2\sqrt{mr}\|\boldsymbol{W}\|\|\boldsymbol{O}-\boldsymbol{\Theta}\|_F \ge \operatorname{tr}(\boldsymbol{O}^T\boldsymbol{W}\boldsymbol{O}) - \operatorname{tr}(\boldsymbol{\Theta}^T\boldsymbol{W}\boldsymbol{\Theta}) \ge \left(\tfrac{m}{2}-1\right)\|\boldsymbol{O}-\boldsymbol{\Theta}\|_F^2,$$

from which inequality (4.8) holds. Inequality (4.10) follows from

$$\begin{aligned}\|[\boldsymbol{W}\boldsymbol{O}]_i\|_F &\le \|[\boldsymbol{W}(\boldsymbol{O}-\boldsymbol{\Theta})]_i\|_F + \|[\boldsymbol{W}\boldsymbol{\Theta}]_i\|_F = \|\boldsymbol{W}_{i\cdot}(\boldsymbol{O}-\boldsymbol{\Theta})\|_F + \|[\boldsymbol{W}\boldsymbol{\Theta}]_i\|_F \\ &\le \|\boldsymbol{W}_{i\cdot}\|\|\boldsymbol{O}-\boldsymbol{\Theta}\|_F + \|[\boldsymbol{W}\boldsymbol{\Theta}]_i\|_F \le \|\boldsymbol{W}\|\|\boldsymbol{O}-\boldsymbol{\Theta}\|_F + \|[\boldsymbol{W}\boldsymbol{\Theta}]_i\|_F\end{aligned}$$

and inequality (4.8), where $\boldsymbol{W}_{i\cdot} = [\boldsymbol{W}_{i1}, \ldots, \boldsymbol{W}_{im}]$, is the ith row block of $\boldsymbol{W}$.

Inequality (4.8) also implies

$$\|\boldsymbol{\Theta}^T\boldsymbol{O} - m\mathbf{I}_r\| \le \begin{cases} 4\|\boldsymbol{W}\|\sqrt{r} & \text{under (MAXBET)}, \\ 4\|\boldsymbol{W}\|\frac{\sqrt{r}}{1-2/m} & \text{under (MAXDIFF)}. \end{cases} \tag{5.32}$$

We first consider the MAXBET model. The global optimality of $(\boldsymbol{O}_1, \ldots, \boldsymbol{O}_m)$ asserts that the associated Lagrange multiplier $\boldsymbol{\Lambda}_i$ of $\boldsymbol{O}_i$ satisfies $\boldsymbol{O}_i\boldsymbol{\Lambda}_i = \sum_{j=1}^m \boldsymbol{S}_{ij}\boldsymbol{O}_j$ (see (2.3)) and is symmetric and positive semidefinite [35, Proposition 3.1]. Since $\boldsymbol{S}_{ij} = \boldsymbol{\Theta}_i\boldsymbol{\Theta}_j^T + \boldsymbol{W}_{ij}$,

$$\sum_{j=1}^m \boldsymbol{S}_{ij}\boldsymbol{O}_j = \sum_{j=1}^m \boldsymbol{\Theta}_i\boldsymbol{\Theta}_j^T\boldsymbol{O}_j + \sum_{j=1}^m \boldsymbol{W}_{ij}\boldsymbol{O}_j = \boldsymbol{\Theta}_i\boldsymbol{\Theta}^T\boldsymbol{O} + [\boldsymbol{W}\boldsymbol{O}]_i. \tag{5.33}$$

Thus from Lemma 5.2, we have

$$\begin{aligned}0 &\ge \operatorname{tr}[(\boldsymbol{\Theta}_i - \boldsymbol{O}_i)^T \textstyle\sum_{j\neq i}\boldsymbol{S}_{ij}\boldsymbol{O}_j] \\ &= \operatorname{tr}[(\boldsymbol{\Theta}_i - \boldsymbol{O}_i)^T\boldsymbol{\Theta}_i\boldsymbol{\Theta}^T\boldsymbol{O}] + \operatorname{tr}[(\boldsymbol{\Theta}_i - \boldsymbol{O}_i)^T[\boldsymbol{W}\boldsymbol{O}]_i].\end{aligned} \tag{5.34}$$

Using that $\boldsymbol{\Theta}^T\boldsymbol{O}$ is symmetric and positive semidefinite, we have, similar to inequality (5.29),

$$\operatorname{tr}[(\boldsymbol{\Theta}_i - \boldsymbol{O}_i)^T\boldsymbol{\Theta}_i\boldsymbol{\Theta}^T\boldsymbol{O}] \ge \tfrac{1}{2}\lambda_r(\boldsymbol{\Theta}^T\boldsymbol{O})\|\boldsymbol{\Theta}_i - \boldsymbol{O}_i\|_F^2. \tag{5.35}$$

Then the Cauchy–Schwarz inequality and inequality (4.10) entail

$$\frac{1}{2}\lambda_r(\boldsymbol{\Theta}^T\boldsymbol{O})\max_{1\le i\le m}\|\boldsymbol{\Theta}_i-\boldsymbol{O}_i\|_F\le\max_{1\le i\le m}\|[\boldsymbol{W}\boldsymbol{O}]_i\|_F\le\max_{1\le i\le m}\|[\boldsymbol{W}\boldsymbol{\Theta}]_i\|_F+4\|\boldsymbol{W}\|^2\frac{\sqrt{r}}{\sqrt{m}}.$$

Combining inequality (5.32) and Weyl's inequality, $\lambda_r(\boldsymbol{\Theta}^T\boldsymbol{O})\ge m-4\|\boldsymbol{W}\|\sqrt{r}$, and inequality (4.11) is obtained.

Under the MAXDIFF model, (5.33) becomes $\boldsymbol{O}_i\boldsymbol{\Lambda}_i=\sum_{j\neq i}\boldsymbol{S}_{ij}\boldsymbol{O}_j=\boldsymbol{\Theta}_i\boldsymbol{\Theta}^T\boldsymbol{O}-\boldsymbol{\Theta}_i\boldsymbol{\Theta}_i^T\boldsymbol{O}_i+[\boldsymbol{W}\boldsymbol{O}]_i$, and inequality (5.34) is replaced by

$$0\ge\operatorname{tr}[(\boldsymbol{\Theta}_i-\boldsymbol{O}_i)^T\boldsymbol{\Theta}_i\boldsymbol{\Theta}^T\boldsymbol{O}]-\operatorname{tr}[(\boldsymbol{\Theta}_i-\boldsymbol{O}_i)^T\boldsymbol{\Theta}_i\boldsymbol{\Theta}_i^T\boldsymbol{O}_i]+\operatorname{tr}[(\boldsymbol{\Theta}_i-\boldsymbol{O}_i)^T[\boldsymbol{W}\boldsymbol{O}]_i].$$

Inequality (5.35) remains intact, and

$$\begin{aligned}-\operatorname{tr}[(\boldsymbol{\Theta}_i-\boldsymbol{O}_i)^T\boldsymbol{\Theta}_i\boldsymbol{\Theta}_i^T\boldsymbol{O}_i]&=\operatorname{tr}[(\boldsymbol{O}_i-\boldsymbol{\Theta}_i)^T\boldsymbol{\Theta}_i\boldsymbol{\Theta}_i^T(\boldsymbol{O}_i-\boldsymbol{\Theta}_i)]-\operatorname{tr}[(\boldsymbol{\Theta}_i-\boldsymbol{O}_i)^T\boldsymbol{\Theta}_i\boldsymbol{\Theta}_i^T\boldsymbol{\Theta}_i]\\&\ge-\|\boldsymbol{\Theta}_i\boldsymbol{\Theta}_i^T\|\|\boldsymbol{O}_i-\boldsymbol{\Theta}_i\|_F^2-\operatorname{tr}[(\boldsymbol{\Theta}_i-\boldsymbol{O}_i)^T\boldsymbol{\Theta}_i]\\&\ge-\|\boldsymbol{O}_i-\boldsymbol{\Theta}_i\|_F^2-\operatorname{tr}(\mathbf{I}_r-\boldsymbol{O}_i^T\boldsymbol{\Theta}_i)\\&\ge-\|\boldsymbol{O}_i-\boldsymbol{\Theta}_i\|_F^2-\frac{1}{2}[\|\boldsymbol{O}_i\|_F^2+\|\boldsymbol{\Theta}_i\|_F^2-2\operatorname{tr}(\boldsymbol{O}_i^T\boldsymbol{\Theta}_i)]=-\frac{3}{2}\|\boldsymbol{O}_i-\boldsymbol{\Theta}_i\|_F^2,\end{aligned}$$

where the third line is due to $\|\boldsymbol{\Theta}_i\boldsymbol{\Theta}_i^T\|\le 1$. Hence the Cauchy–Schwarz inequality and inequality (4.10) now give

$$\begin{aligned}\frac{1}{2}(\lambda_r(\boldsymbol{\Theta}^T\boldsymbol{O})-3)\max_{1\le i\le m}\|\boldsymbol{\Theta}_i-\boldsymbol{O}_i\|_F&\le\max_{1\le i\le m}\|[\boldsymbol{W}\boldsymbol{O}]_i\|_F\\&\le\max_{1\le i\le m}\|[\boldsymbol{W}\boldsymbol{\Theta}]_i\|_F+4\|\boldsymbol{W}\|^2\frac{\sqrt{r}}{\sqrt{m}-2/\sqrt{m}}.\end{aligned}$$

Inequality (5.32) and Weyl's inequality now result in $\lambda_r(\boldsymbol{\Theta}^T\boldsymbol{O})\ge m-4\|\boldsymbol{W}\|\frac{\sqrt{r}}{1-2/m}$, and inequality (4.11) is obtained. For a valid bound we need $m>4\|\boldsymbol{W}\|\frac{\sqrt{r}}{1-2/m}+3$. Solving the involved quadratic inequality provides the desired lower bound for m. ∎

6. Conclusion.

This paper studies the OSTM problem [35]. It shows two results when the noise is small: first, that while the problem is nonconvex, its solution can be achieved by solving its convex relaxation; second, condition (2.4) is necessary and sufficient for global optimality of a critical point, making the former a genuine certificate.

A future direction is to improve the estimation on maximum noise that this method can handle. While this paper shows that the method succeeds when $\sigma=O(m^{1/4})$, we expect that it would also hold for noise as large as $\sigma=O(m^{1/2})$, which has been proven in [37] for phase synchronization and in [22] for synchronization of rotations. We suspect that the suboptimality of this result arises from the estimation of $\max_{1\le i\le m}\|\sum_{j=1}^m\boldsymbol{W}_{ij}\tilde{\boldsymbol{V}}_j\|$ in (4.4), where standard tools from the theory of measure concentration cannot be used as $\tilde{\boldsymbol{V}}$ depends on $\boldsymbol{W}$. Likewise, in certifying global optimality of a critical point, estimation of $\max_{1\le i\le m}\|\sum_{j=1}^m\boldsymbol{W}_{ij}\boldsymbol{O}_j\|$ in inequality (4.10) becomes a bottleneck. To solve this problem, some decoupling techniques in probability theory might be needed to disentangle the dependence structure. Another future direction is to use a more generic model than the additive Gaussian noise model, which would have a larger range of real-life applications.

Appendix A. Simulation study.

We conducted a simulation study to see how tight the conditions (3.6) and (3.7) are. Under the data generation model (MAXBET), we fixed $d=5$, $r=3$ and varied

TABLE 1
Frequency of satisfaction of conditions (3.6), (3.7) *and certificate* (2.4).

m	σ	(3.6)†	(3.7)	(2.4)†
10	0.01	100	TRUE	100
	0.10	10	FALSE	100
	1.00	0	FALSE	0
	1.50	0	FALSE	0
20	0.01	100	TRUE	100
	0.10	0	FALSE	100
	1.00	0	FALSE	21
	1.50	0	FALSE	0
30	0.01	100	TRUE	100
	0.10	0	FALSE	100
	1.00	0	FALSE	99
	1.50	0	FALSE	0

†Reported numbers are out of 100 replicates in each scenario.

the number of groups $m \in \{2, 5, 10\}$ and the noise level $\sigma \in \{0.01, 0.1, 1, 10\}$. The semiorthogonal matrices $\mathbf{\Theta}_1, \ldots, \mathbf{\Theta}_m$ were generated by taking the QR decomposition of random $d \times r$ matrices with i.i.d. standard normal entries. The upper triangular part including the diagonal of $\boldsymbol{W}$ was generated from i.i.d. normal with mean zero and variance σ^2. For each combination of m and σ, we generated 100 replicates and reported the number of replicates for which the proximal block ascent algorithm in [35] produced a critical point satisfying certificate (2.4) using the ten Berge initialization strategy ("tb" in [35]) in Table 1. In addition, we also counted the frequency of satisfying conditions (3.6) and (3.7) for Corollaries 3.10 and 3.11, respectively, and the certificate of global optimality of a critical point (2.4).

Table 1 shows that condition (3.6) is satisfied at small noise levels. Condition (3.7), which is fully determined by the combination of m and σ, is less frequently satisfied than (3.6). In case either condition (3.6) or (3.7) is satisfied, the certificate (2.4) is always satisfied as predicted by the theory. It is remarkable that certificate (2.4) is satisfied more frequently than condition (3.6) or (3.7), leaving room for improvement of these conditions.

REFERENCES

[1] E. Abbe, L. Massoulie, A. Montanari, A. Sly, and N. Srivastava, *Group Synchronization on Grids*, preprint, arXiv:1706.08561, 2017.
[2] M. Arie-Nachimson, S. Z. Kovalsky, I. Kemelmacher-Shlizerman, A. Singer, and R. Basri, *Global motion estimation from point matches*, in Proceedings of the 2012 Second International Conference on 3D Imaging, Modeling, Processing, Visualization, and Transmission, IEEE, 2012, pp. 81–88.
[3] F. Arrigoni, A. Fusiello, and B. Rossi, *Camera motion from group synchronization*, in Proceedings of the 2016 Fourth International Conference on 3D Vision (3DV), IEEE, 2016, pp. 546–555.
[4] A. S. Bandeira, N. Boumal, and A. Singer, *Tightness of the maximum likelihood semidefinite relaxation for angular synchronization*, Math. Program., 163 (2017), pp. 145–167.
[5] A. S. Bandeira, C. Kennedy, and A. Singer, *Approximating the little Grothendieck problem over the orthogonal and unitary groups*, Math. Program., 160 (2016), pp. 433–475.
[6] N. Boumal, *Nonconvex phase synchronization*, SIAM J. Optim., 26 (2016), pp. 2355–2377.
[7] N. Boumal, A. Singer, and P. A. Absil, *Robust estimation of rotations from relative measurements by maximum likelihood*, in Proceedings of the 52nd IEEE Conference on Decision and Control, 2013, pp. 1156–1161.

[8] N. Boumal, A. Singer, P.-A. Absil, and V. D. Blondel, *Cramér-Rao bounds for synchronization of rotations*, Inf. Inference, 3 (2013), pp. 1–39.

[9] K. N. Chaudhury, Y. Khoo, and A. Singer, *Global registration of multiple point clouds using semidefinite programming*, SIAM J. Optim., 25 (2015), pp. 468–501.

[10] Y. Chen and A. J. Goldsmith, *Information recovery from pairwise measurements*, in Proceedings of the 2014 IEEE International Symposium on Information Theory (ISIT), 2014, pp. 2012–2016.

[11] M. Cucuringu, *Synchronization over Z2 and community detection in signed multiplex networks with constraints*, J. Complex Netw., 3 (2015), pp. 469–506.

[12] M. Cucuringu, A. Singer, and D. Cowburn, *Eigenvector synchronization, graph rigidity and the molecule problem*, Inf. Inference, 1 (2012), pp. 21–67.

[13] K. R. Davidson and S. J. Szarek, *Local operator theory, random matrices and Banach spaces*, in Handbook of the Geometry of Banach Spaces, Vol. 1, North-Holland, Amsterdam, 2001, pp. 317–366, https://doi.org/10.1016/S1874-5849(01)80010-3.

[14] Y. Fei and Y. Chen, *Achieving the Bayes Error Rate in Synchronization and Block Models by SDP, Robustly*, preprint, arXiv:1904.09635, 2019.

[15] S. Foucart and H. Rauhut, *A Mathematical Introduction to Compressive Sensing*, Birkhauser, New York, 2013.

[16] T. Gao, J. Brodzki, and S. Mukherjee, *The geometry of synchronization problems and learning group actions*, Discrete Comput. Geom., 65 (2021), pp. 150–211.

[17] J. C. Gower, *Generalized Procrustes analysis*, Psychometrika, 40 (1975), pp. 33–51.

[18] M. Hanafi and H. A. Kiers, *Analysis of k sets of data, with differential emphasis on agreement between and within sets*, Comput. Statist. Data Anal., 51 (2006), pp. 1491–1508.

[19] Y. Khoo and A. Kapoor, *Non-iterative rigid 2D/3D point-set registration using semidefinite programming*, IEEE Trans. Image Process., 25 (2016), pp. 2956–2970.

[20] K. Lange, *MM Optimization Algorithms*, SIAM, Philadelphia, 2016.

[21] S. Ling, *Solving Orthogonal Group Synchronization via Convex and Low-rank Optimization: Tightness and Landscape Analysis*, preprint, arxiv:2006.00902, 2020, https://arxiv.org/abs/2006.00902.

[22] S. Ling, *Improved Performance Guarantees for Orthogonal Group Synchronization via Generalized Power Method*, preprint, arXiv:2012.00470, 2021, https://arxiv.org/abs/2012.00470.

[23] S. Ling, R. Xu, and A. S. Bandeira, *On the landscape of synchronization networks: A perspective from nonconvex optimization*, SIAM J. Optim., 29 (2019), pp. 1879–1907.

[24] H. Liu, M.-C. Yue, and A. Man-Cho So, *On the estimation performance and convergence rate of the generalized power method for phase synchronization*, SIAM J. Optim., 27 (2017), pp. 2426–2446.

[25] H. Liu, M.-C. Yue, and A. M.-C. So, *A Unified Approach to Synchronization Problems Over Subgroups of the Orthogonal Group*, preprint, arxiv:2009.07514, 2020, https://arxiv.org/abs/2009.07514.

[26] O. Özyesil, N. Sharon, and A. Singer, *Synchronization over Cartan motion groups via contraction*, SIAM J. Appl. Algebra Geom., 2 (2018), pp. 207–241.

[27] A. Perry, A. S. Wein, A. S. Bandeira, and A. Moitra, *Message-passing algorithms for synchronization problems over compact groups*, Comm. Pure Appl. Math., 71 (2018), pp. 2275–2322.

[28] A. Singer, *Angular synchronization by eigenvectors and semidefinite programming*, Appl. Comput. Harmon. Anal., 30 (2011), pp. 20–36.

[29] A. Singer and Y. Shkolnisky, *Three-dimensional structure determination from common lines in cryo-EM by eigenvectors and semidefinite programming*, SIAM J. Imaging Sci., 4 (2011), pp. 543–572.

[30] J. M. F. Ten Berge, *Orthogonal Procrustes rotation for two or more matrices*, Psychometrika, 42 (1977), pp. 267–276.

[31] J. Thunberg, F. Bernard, and J. Goncalves, *Distributed methods for synchronization of orthogonal matrices over graphs*, Automatica, 80 (2017), pp. 243–252.

[32] R. Tron, X. Zhou, and K. Daniilidis, *A survey on rotation optimization in structure from motion*, in Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), 2016, pp. 1032–1040.

[33] L. Wang and A. Singer, *Exact and stable recovery of rotations for robust synchronization*, Inf. Inference, 2 (2013), pp. 145–193.

[34] L. Wang, A. Singer, and Z. Wen, *Orientation determination of cryo-EM images using least unsquared deviations*, SIAM J. Imaging Sci., 6 (2013), pp. 2450–2483.

[35] J.-H. Won, H. Zhou, and K. Lange, *Orthogonal trace-sum maximization: Applications, local algorithms, and global optimality*, SIAM J. Matrix Anal. Appl., 42 (2021), pp. 859–882, https://doi.org/10.1137/20M1363388.

[36] T. Zhang and A. Singer, *Disentangling orthogonal matrices*, Linear Algebra Appl., 524 (2017), pp. 159–181.

[37] Y. Zhong and N. Boumal, *Near-optimal bounds for phase synchronization*, SIAM J. Optim., 28 (2018), pp. 989–1016.