Robust Principal Component Analysis: Exact Recovery of Corrupted Low-Rank Matrices via Convex Optimization

Supplementary material: proofs of main results

John Wright,[*] Yigang Peng, Yi Ma
Visual Computing Group
Microsoft Research Asia
{jowrig,v-yipe,mayi}@microsoft.com

Arvind Ganesh, Shankar Rao
Coordinated Science Laboratory
University of Illinois at Urbana-Champaign
{abalasu2,srrao}@uiuc.edu

This supplementary appendix provides additional details of the proofs of our results. Sections 1 and 2 prove our main result on Robust PCA, Theorem 1 of the paper. Section 3 explores the implications of this analysis for the related problem of low-rank matrix completion, and proves Theorem 2 of the paper.

For completeness, this supplement duplicates material appearing in the journal version of this paper [9]. The reader is encouraged to consult that work for a more thorough exposition of our main results.

1 Analysis Framework

In this section, we begin our analysis of the semidefinite program

$$\min_{A,E} \ \|A\|_* + \lambda\|E\|_1 \quad \text{subj} \quad A + E = D. \tag{1}$$

After introducing notation, we provide two sufficient conditions for a pair (A, E) to be the unique optimal solution to (1). The first condition, given in Lemma 1.1 below, is stated in terms of the existence of a dual vector W that certifies optimality of the pair (A, E). The existence (or non-existence) of such a dual vector is itself a random convex programming feasibility problem, and so the condition in this Lemma is not directly amenable to analysis. The next step, outlined in Lemma 1.3, is to show that if there exists some W_0 that does not violate the constraints of this feasibility problem too badly, then it can be refined to produce a W_∞ that satisfies the constraints and certifies optimality of (A, E). The probabilistic analysis in the following Section 2 will then show that, with high probability, such a W_0 exists.

1.1 Notation

For any $n \in \mathbb{Z}_+$, $[n] \doteq \{1 \dots n\}$. For $M \in \mathbb{R}^{m\times m}$, and $I, J \subset [m]$, $M_{I,J}$ will denote the submatrix of M consisting of those rows indexed by I and those columns indexed by J. We will use $\bullet$ as shorthand for the entire index set: $M_{I,\bullet}$ is the submatrix consisting of those rows indexed by I. M^* will denote the transpose of M. For matrices $P, Q \in \mathbb{R}^{m\times n}$, $\langle P, Q\rangle \doteq \operatorname{trace}[P^*Q]$ will denote the (Frobenius) inner product. The symbol $\mathtt{I}$ will denote the identity matrix or identity operator on matrices; the distinction will be clear from context.

$\|M\|_{p,q}$ will denote the operator norm of the matrix M, as a linear map between ℓ^p and ℓ^q. Important special cases are the spectral norm $\|M\|_{2,2}$, and the max row- and column norms,

$$\|M\|_{2,\infty} = \max_i \|M_{i,\bullet}\|_2 \quad \text{and} \quad \|M\|_{1,2} = \max_j \|M_{\bullet,j}\|_2, \tag{2}$$

and the max element norm

$$\|M\|_{1,\infty} = \max_{i,j} |M_{ij}|. \tag{3}$$

[*]For more information, see `http://perception.csl.illinois.edu/matrix-rank/home.html`. This work was partially supported by NSF IIS 08-49292, NSF ECCS 07-01676, and ONR N00014-09-1-0230.

We will also often reason about linear operators on matrices. If $\mathcal{L} : \mathbb{R}^{m\times m} \to \mathbb{R}^{m\times m}$ is a linear map, $\|\mathcal{L}\|_{F,F}$ will denote its operator norm with respect to the Frobenius norm on matrices:

$$\|\mathcal{L}\|_{F,F} \doteq \sup_{M\in\mathbb{R}^{m\times m}\setminus\{0\}} \frac{\|\mathcal{L}[M]\|_F}{\|M\|_F}. \tag{4}$$

If S is a subspace, π_S will denote the projection operator onto that subspace. For $U \in \mathbb{R}^{m\times r}$, π_U will denote the projection operator onto the range of U. Finally, if $I \subset [m]$, $\pi_I \doteq \sum_{i\in I} \boldsymbol{e}_i\boldsymbol{e}_i^*$ will denote the projection onto those coordinates indexed by I.

We will let

$$\Omega \doteq \{(i,j) \mid E_{0ij} \neq 0\} \tag{5}$$

denote the set of corrupted entries. By slight abuse of notation, we will identify Ω with subspace $\{M \mid M_{ij} = 0 \;\forall\, (i,j) \in \Omega^c\}$, so π_Ω will denote the projection onto the corrupted elements. We will let $\Sigma \in \mathbb{R}^{m\times m}$ be an iid Rademacher (Bernoulli ± 1) matrix that defines the signs of the corrupted entries:

$$\mathrm{sign}(E_0) = \pi_\Omega[\Sigma]. \tag{6}$$

Expressing $\mathrm{sign}(E_0)$ in this way allows us to exploit independence between Ω and Σ.

The symbol $\otimes$ will denote the Kronecker product between matrices. "vec" will denote the operator that vectorizes a matrix by stacking the columns. For $M \in \mathbb{R}^{m\times m}$, this operator stacks the entries M_{ij} in lexicographic order. For $\boldsymbol{x} = \mathrm{vec}[M] \in \mathbb{R}^{m\times m}$ we write $\boldsymbol{x}_\Omega$ as shorthand for $\boldsymbol{x}_{L(\Omega)}$, where $L(\Omega) = \{jm + i \mid (i,j) \in \Omega\} \subset [m^2]$, and as further shorthand, we will occasionally use $\mathrm{vec}_\Omega[M]$ to denote $[\mathrm{vec}[M]]_\Omega$.

In both our analysis and optimization algorithm, we will make frequent use of the soft thresholding operator, which we denote $\mathcal{S}_\gamma$:

$$\mathcal{S}_\gamma[x] = \begin{cases} 0, & |x| \leq \gamma, \\ \mathrm{sign}(x)(|x| - \gamma), & |x| > \gamma. \end{cases} \tag{7}$$

For matrices X, $\mathcal{S}_\gamma[X]$ will denote the matrix obtained by applying $\mathcal{S}_\gamma$ to each element of X_{ij}. Finally, we often find it convenient to use the notation

$$\mathcal{S}_\gamma^{\Omega^c}[X] \doteq \pi_{\Omega^\perp}[\mathcal{S}_\gamma[X]]. \tag{8}$$

That is, $\mathcal{S}_\gamma^{\Omega^c}$ applies the soft threshold to X but only retains those elements not in Ω.

Throughout, we will use the symbol $\mathbb{W}_r^m$ to refer to the Stiefel manifold of matrices $U \in \mathbb{R}^{m\times r}$ with orthonormal columns: $U^*U = \mathtt{I}_{r\times r}$. We will use $SO(r)$ to refer to the special orthogonal group of $r \times r$ matrices R such that $R^*R = RR^* = \mathtt{I}_{r\times r}$ and $\det(R) = 1$.

1.2 Optimality conditions for (A_0, E_0)

We begin with a simple sufficient condition for the pair (A_0, E_0) to be the unique optimal solution to (1). These conditions are stated in terms of a dual vector, the existence of which certifies optimality. Our analysis will then show that under the stated circumstances, such a dual certificate can be produced with high probability.

Lemma 1.1. *Let $(A_0, E_0) \in \mathbb{R}^{m\times m} \times \mathbb{R}^{m\times m}$, with*

$$\Omega \doteq \mathrm{supp}(E_0) \subseteq [m] \times [m]. \tag{9}$$

Let $A_0 = USV^*$ denote the compact singular value decomposition of A_0, and Θ denote the subspace generated by matrices with column space $\mathrm{range}(U)$ *or row space* $\mathrm{range}(V)$*:*

$$\Theta \doteq \{UM^* \mid M \in \mathbb{R}^{m\times r}\} + \{MV^* \mid M \in \mathbb{R}^{m\times r}\} \subset \mathbb{R}^{m\times m}. \tag{10}$$

Suppose that $\|\pi_\Omega\pi_\Theta\|_{F,F} < 1$ and there exists $W \in \mathbb{R}^{m\times m}$ such that

$$\left\{ \begin{array}{l} [UV^* + W]_{ij} \;=\; \lambda\,\mathrm{sign}(E_{0i,j}) \quad \forall\, i,j \in \Omega, \\ |[UV^* + W]_{ij}| \;<\; \lambda \quad \forall\, i,j \in \Omega^c, \\ U^*W \;=\; 0, \qquad WV \;=\; 0, \qquad \|W\|_{2,2} \;<\; 1 \end{array} \right\}. \tag{11}$$

Then the pair (A_0, E_0) is the unique optimal solution to (1)*.*

Proof. We show that if a Lagrange multiplier vector $Y \doteq UV^* + W$ satisfying this system exists, then any nonzero perturbation $A_0 \mapsto A_0 + \Delta$, $E_0 \mapsto E_0 - \Delta$ respecting the constraint $A + E = A_0 + E_0$ will increase the objective function. By convexity, we may restrict our interest to Δ satisfying $\|\Delta\|_{1,\infty} < \min_{(i,j)\in\Omega} |E_{0ij}|$. For such Δ,

$$\lambda\|E_0 - \Delta\|_1 - \lambda\|E_0\|_1 \;=\; -\lambda \sum_{(i,j)\in\Omega} \operatorname{sign}(E_{0ij})\Delta_{ij} + \lambda \sum_{(i,j)\notin\Omega} |\Delta_{ij}| \;\geq\; \langle Y, -\Delta\rangle.$$

with equality if and only if $\pi_{\Omega^\perp}[\Delta] = 0$ (since on this set $|Y_{ij}|$ is strictly smaller than λ). Similarly, since $Y = UV^* + W \in \partial\|\cdot\|_*\big|_{A_0}$,

$$\|A_0 + \Delta\|_* \geq \|A_0\|_* + \langle Y, \Delta\rangle. \tag{12}$$

Moreover, since $\|A_0\|_* = \langle UV^*, A_0\rangle = \langle Y, A_0\rangle$, $\|A_0\|_* + \langle Y, \Delta\rangle = \langle Y, A_0 + \Delta\rangle$. Thus, by Lemma 1.2 below, if equality holds in (12), then $\Delta \in \Theta$. Summing the two subgradient bounds, we have

$$\|A_0 + \Delta\|_* + \lambda\|E_0 - \Delta\|_1 \;\geq\; \|A_0\|_* + \lambda\|E_0\|_1,$$

with equality only if $\Delta \in \Omega \cap \Theta$. If $\|\pi_\Omega \pi_\Theta\|_{F,F} < 1$, then $\Omega \cap \Theta = \{0\}$ and so either

$$\|A_0 + \Delta\|_* + \lambda\|E_0 - \Delta\|_1 \;>\; \|A_0\|_* + \lambda\|E_0\|_1,$$

or $\Delta = 0$. □

Lemma 1.2. *Consider $P \in \mathbb{R}^{m\times m}$, with $\|P\|_{2,2} = 1$ and $\sigma_{min}(P) < 1$. Write the full singular value decomposition of P as $P = \begin{bmatrix} U_1 & U_2 \end{bmatrix} \begin{bmatrix} \mathtt{I} & 0 \\ 0 & \Sigma_2 \end{bmatrix} \begin{bmatrix} V_1 & V_2 \end{bmatrix}^*$ where $\|\Sigma_2\|_{2,2} < 1$. Let $Q \in \mathbb{R}^{m\times m}$ have reduced singular value decomposition $Q = U\Sigma V^*$. Then if $\langle P, Q\rangle = \|Q\|_*$, $U^*U_2 = 0$ and $V^*V_2 = 0$.*

Proof. Let r denote the rank of Q.

$$\langle P, Q\rangle = \left\langle \begin{bmatrix} \mathtt{I} & 0 \\ 0 & \Sigma_2 \end{bmatrix}, \begin{bmatrix} U_1^*U \\ U_2^*U \end{bmatrix} \Sigma \begin{bmatrix} V^*V_1 & V^*V_2 \end{bmatrix} \right\rangle.$$

Let $Y, Z \in \mathbb{R}^{m\times r}$ denote the matrices $Y \doteq \begin{bmatrix} U_1^*U \\ U_2^*U \end{bmatrix}$, $Z \doteq \begin{bmatrix} V_1^*V \\ V_2^*V \end{bmatrix}$, and notice that both Y and Z have orthonormal columns. Then, after the above rotation

$$\langle P, Q\rangle \;=\; \sum_{i=1}^m \sigma_i(P) \sum_{j=1}^r \sigma_j(Q) Z_{ij} Y_{ij} \;=\; \sum_{j=1}^r \sigma_j(Q) \sum_{i=1}^m \sigma_i(P) Z_{ij} Y_{ij}. \tag{13}$$

For all j, $\sum_{i=1}^m \sigma_i(P) Z_{ij} Y_{ij} \leq \left\| \left[\begin{smallmatrix} \mathtt{I} & 0 \\ 0 & \Sigma_2 \end{smallmatrix}\right] Z_{\bullet,j} \right\|_2 \|Y_{\bullet,j}\|_2 \leq 1$, with equality if and only if $\sigma_i(P) < 1 \implies Z_{ij} = 0$ and $Z_{\bullet,j} = Y_{\bullet,j}$. Since whenever $\langle P, Q\rangle = \|Q\|_*$ each of the $\sum_{i=1}^m \sigma_i(P) Z_{ij} Y_{ij}$ must be one, this implies that U_2^*U and V_2^*V are both zero. □

Thus, we can guarantee that (A_0, E_0) is optimal with high probability by asserting that a certain random convex feasibility problem is satisfiable with high probability. While there are potentially many W satisfying these constraints, most of them lack an explicit expression. As has proven fruitful in a variety of related problems, we will instead consider a putative dual vector W_0 that *does* have an explicit expression: the minimum Frobenius norm solution to the equality constraints in (11). We will see that this vector already satisfies the operator norm constraint with high probability. However, the box constraint due to $\operatorname{sign}(E_0)$ is likely to be violated. We then describe an iterative procedure that, with high probability, fixes the box constraint, while respecting the equality and operator norm constraints. The output of this iteration is the desired certifying dual vector.

1.3 Iterative construction of the dual certificate

In constructing the dual certificate, we will use the fact that the violations of the inequality constraints, viewed as a matrix, often have sparse rows and sparse columns. To formalize this, fix a small constant $c \in [0, 1]$, and define

$$\Psi_c \doteq \left\{ M \in \mathbb{R}^{m\times m} \mid \|M_{\bullet,j}\|_0 \leq cm \,\forall\, j,\; \|M_{i,\bullet}\|_0 \leq cm \,\forall\, i,\; \pi_\Omega[M] = 0 \right\}. \tag{14}$$

We will show that such matrices are near the nullspace of the equality constraints in (11). The equality constraints in (11) can be expressed as

$$\pi_\Omega[W] = \pi_\Omega[\lambda \operatorname{sign}(E_0) - UV^*] \quad \text{and} \quad \pi_\Theta[W] = 0.$$

We will let Γ denote the orthogonal complement of the nullspace of these constraints:

$$\Gamma \doteq \Theta + \Omega. \tag{15}$$

Let ξ_c denote the operator norm of π_Γ, with respect to the Frobenius norm on $\mathbb{R}^{m\times m}$, restricted to Ψ_c:

$$\xi_c \doteq \sup_{M\in\Psi_c\setminus\{0\}} \frac{\|\pi_\Gamma M\|_F}{\|M\|_F} \in [0,1]. \tag{16}$$

In Section 2.3, we establish a useful probabilistic upper bound for this quantity. Before investigating these properties, we show that if ξ_c and W_0 are well-controlled, we can find a dual vector certifying optimality of (A_0, E_0).

Lemma 1.3 (Iterative construction of a dual certificate)**.** *Suppose for some $c \in (0,1)$ and $\varepsilon > 0$, there exists W_0 satisfying $\pi_\Theta[W_0] = 0$ and $\pi_\Omega[W_0] = \pi_\Omega[\lambda \operatorname{sign}(E_0) - UV^*]$, such that*

$$\|UV^* + W_0\|_{1,2} + \frac{1}{1-\xi_c}\left\|\mathcal{S}^{\Omega^c}_{\lambda-\varepsilon}(UV^* + W_0)\right\|_F \le (\lambda-\varepsilon)\sqrt{cm}, \tag{17}$$

$$\|UV^* + W_0\|_{2,\infty} + \frac{1}{1-\xi_c}\left\|\mathcal{S}^{\Omega^c}_{\lambda-\varepsilon}(UV^* + W_0)\right\|_F \le (\lambda-\varepsilon)\sqrt{cm}, \tag{18}$$

and

$$\|W_0\|_{2,2} + \frac{1}{1-\xi_c}\|\mathcal{S}^{\Omega^c}_{\lambda-\varepsilon}(UV^* + W_0)\|_F < 1. \tag{19}$$

Then there exists W_∞ satisfying the system (11).

Proof. We construct a convergent sequence $W_0, W_1, \ldots$ whose limit satisfies (11). For each k, set

$$W_k = W_{k-1} - \pi_{\Gamma^\perp}\mathcal{S}^{\Omega^c}_{\lambda-\varepsilon}\left(UV^* + W_{k-1}\right). \tag{20}$$

Notice that $\pi_\Theta[W_k] = \pi_\Theta[W_{k-1}]$ and $\pi_\Omega[W_k] = \pi_\Omega[W_{k-1}]$, so for all k, $\pi_\Theta[W_k] = 0$ and $\pi_\Omega[W_k] = \pi_\Omega[\lambda \operatorname{sign}(E_0) - UV^*]$. We will further show by induction that the sequence (W_k) satisfies the following properties:

$$\|UV^* + W_k\|_{1,2} \le \|UV^* + W_0\|_{1,2} + \left\|\mathcal{S}^{\Omega^c}_{\lambda-\varepsilon}(UV^* + W_0)\right\|_F \sum_{i=0}^{k-1}\xi_c^i, \tag{21}$$

$$\|UV^* + W_k\|_{2,\infty} \le \|UV^* + W_0\|_{2,\infty} + \left\|\mathcal{S}^{\Omega^c}_{\lambda-\varepsilon}(UV^* + W_0)\right\|_F \sum_{i=0}^{k-1}\xi_c^i, \tag{22}$$

$$\max_i \left\|\left[\mathcal{S}^{\Omega^c}_{\lambda-\varepsilon}\left(UV^* + W_{k-1}\right)\right]_{i,\bullet}\right\|_0 \le cm, \tag{23}$$

$$\max_j \left\|\left[\mathcal{S}^{\Omega^c}_{\lambda-\varepsilon}\left(UV^* + W_{k-1}\right)\right]_{\bullet,j}\right\|_0 \le cm, \tag{24}$$

$$\|\mathcal{S}^{\Omega^c}_{\lambda-\varepsilon}(UV^* + W_k)\|_F \le \xi_c^k\|\mathcal{S}^{\Omega^c}_{\lambda-\varepsilon}(UV^* + W_0)\|_F. \tag{25}$$

For $k = 0$, (21), (22) and (25) hold trivially. The sparsity assertion (23) follows from the assumptions of the lemma: for all i,

$$\left\|\left[\mathcal{S}^{\Omega^c}_{\lambda-\varepsilon}(UV^* + W_0)\right]_{i,\bullet}\right\|_0 \le \frac{\|[UV^* + W_0]_{i,\bullet}\|_2^2}{(\lambda-\varepsilon)^2} \le \frac{\|UV^* + W_0\|_{1,2}^2}{(\lambda-\varepsilon)^2} \le cm.$$

The exact same reasoning applied to the columns gives (24). Now, suppose the statements (21)-(25) hold for $W_0 \ldots W_{k-1}$. Then

$$\begin{aligned}
\|UV^* + W_k\|_{1,2} &\le \|UV^* + W_{k-1}\|_{1,2} + \left\|\mathcal{S}^{\Omega^c}_{\lambda-\varepsilon}(UV^* + W_{k-1})\right\|_F \\
&\le \|UV^* + W_{k-1}\|_{1,2} + \left\|\mathcal{S}^{\Omega^c}_{\lambda-\varepsilon}(UV^* + W_0)\right\|_F \xi_c^{k-1} \\
&\le \|UV^* + W_0\|_{1,2} + \left\|\mathcal{S}^{\Omega^c}_{\lambda-\varepsilon}(UV^* + W_0)\right\|_F \sum_{i=0}^{k-1}\xi_c^i,
\end{aligned}$$

establishing (21) for k. The same reasoning applied to $\|\cdot\|_{2,\infty}$ establishes (22) for k. Bounding the summation in (21) by $\frac{1}{1-\xi_c}$ and applying assumption (17) of the lemma gives that $\|UV^* + W_k\|_{1,2}^2 \leq (\lambda - \varepsilon)^2 cm$. This implies (23): the number of entries of each column of $UV^* + W_k$ that exceed $\lambda - \varepsilon$ in absolute value is at most cm. The same chain of reasoning establishes that (24): the rows of $\mathcal{S}_{\lambda-\varepsilon}^{\Omega^c}(UV^* + W_k)$ are also cm-sparse.

Finally, notice that

$$\begin{aligned}\|\mathcal{S}_{\lambda-\varepsilon}^{\Omega^c}(UV^* + W_k)\|_F &= \left\|\mathcal{S}_{\lambda-\varepsilon}^{\Omega^c}\left(UV^* + W_{k-1} - \pi_{\Gamma^\perp}\mathcal{S}_{\lambda-\varepsilon}^{\Omega^c}(UV^* + W_{k-1})\right)\right\|_F \\ &= \left\| \mathcal{S}_{\lambda-\varepsilon}^{\Omega^c}\Big(UV^* + W_{k-1} - \mathcal{S}_{\lambda-\varepsilon}^{\Omega^c}(UV^* + W_{k-1}) + \pi_\Gamma \mathcal{S}_{\lambda-\varepsilon}^{\Omega^c}(UV^* + W_{k-1})\Big) \right\|_F.\end{aligned}$$

Since the entries of $UV^* + W_{k-1} - \mathcal{S}_{\lambda-\varepsilon}^{\Omega^c}(UV^* + W_{k-1})$ have magnitude $\leq \lambda - \varepsilon$, $\pi_\Gamma \mathcal{S}_{\lambda-\varepsilon}^{\Omega^c}(UV^* + W_{k-1})$ dominates $\mathcal{S}_{\lambda-\varepsilon}^{\Omega^c}(UV^* + W_k)$ elementwise in absolute value. Hence, $\|\mathcal{S}_{\lambda-\varepsilon}^{\Omega^c}(UV^* + W_k)\|_F \leq \|\pi_\Gamma \mathcal{S}_{\lambda-\varepsilon}^{\Omega^c}(UV^* + W_{k-1})\|_F \leq \xi_c \|\mathcal{S}_{\lambda-\varepsilon}^{\Omega^c}(UV^* + W_{k-1})\|_F$, where we have used that $\mathcal{S}_{\lambda-\varepsilon}^{\Omega^c}(UV^* + W_{k-1}) \in \Psi_c$. Thus each of the three statements holds for all k, and $\|\mathcal{S}_{\lambda-\varepsilon}^{\Omega^c}(UV^* + W_k)\|_F$ decreases geometrically. For all k sufficiently large, $\|UV^* + W_k\|_{1,\infty} \leq \lambda$.

Meanwhile, notice that

$$\begin{aligned}\|W_k\|_{2,2} &\leq \|W_{k-1}\|_{2,2} + \|\pi_{\Gamma^\perp}\mathcal{S}_{\lambda-\varepsilon}^{\Omega^c}(UV^* + W_{k-1})\|_{2,2} \\ &\leq \|W_{k-1}\|_{2,2} + \|\mathcal{S}_{\lambda-\varepsilon}^{\Omega^c}(UV^* + W_{k-1})\|_F \\ &\leq \|W_{k-1}\|_{2,2} + \xi_c^{k-1}\|\mathcal{S}_{\lambda-\varepsilon}^{\Omega^c}(UV^* + W_0)\|_F.\end{aligned}$$

By induction, it is easy to show that

$$\|W_k\|_{2,2} \leq \|W_0\|_{2,2} + \|\mathcal{S}_{\lambda-\varepsilon}^{\Omega^c}(UV^* + W_0)\|_F \sum_{i=0}^{k-1} \xi_c^i.$$

By the second assumption of the lemma, this quantity is bounded by 1 for all k. □

2 Probabilistic Analysis of the Initial Dual Vector

In this section, we analyze the minimum Frobenius norm solution to the equality constraints in the optimality condition (11). We show that with high probability, this initial dual vector satisfies the operator norm constraint, and that the violations of the box constraint are small enough that the iteration described in (20) will succeed in producing a dual certificate. The analysis is organized as follows: in Section 2.1, we introduce several tools and preliminary results. Section 2.2 then shows that with overwhelming probability the operator norm of the initial dual vector is bounded by a constant that can be made arbitrarily small by assuming the error probability ρ_s and rank r are both low enough. Section 2.3 then shows that the restricted operator norm ξ_c is also bounded by a small constant, establishing that all row- and column- sparse matrices are nearly orthogonal to the subspace spanned by the equality constraints in (11). Section 2.4 analyzes the violations of the box constraint. Finally, Section 2.5 combines these results with the results of Section 1 to prove our main result, Theorem 1.

2.1 Tools and preliminaries

In this section, we will often need to refer to the following subspaces

$$\begin{aligned}\Xi_U &\doteq \{UM^* \mid M \in \mathbb{R}^{r\times m}\}, \\ \Xi_V &\doteq \{MV^* \mid M \in \mathbb{R}^{m\times r}\}, \\ \Xi_{UV} &\doteq \{UMV^* \mid M \in \mathbb{R}^{r\times r}\}.\end{aligned}$$

Notice that $\pi_\Theta = \pi_{\Xi_U} + \pi_{\Xi_V} - \pi_{\Xi_{UV}}$ and that $\pi_{\Xi_{UV}} = \pi_{\Xi_U}\pi_{\Xi_V} = \pi_{\Xi_V}\pi_{\Xi_U}$.

In the Bernoulli support model, the number of errors in each row and column concentrates about $\rho_s m$. For each $j \in [m]$, let

$$I_j \doteq \{i \mid (i,j) \in \Omega\}, \tag{26}$$

i.e., I_j contains the indices of the errors in the j-th column. Similarly, for $i \in [m]$, set

$$J_i \doteq \{j \mid (i,j) \in \Omega\}. \tag{27}$$

In terms of these quantities, for each $\eta > 0$, define the event

$$\mathcal{E}_\Omega(\eta) \ : \ \max_j |I_j| < (1+\eta)\rho_s \text{ and } \max_i |J_i| < (1+\eta)\rho_s.$$

Much of our analysis hinges on the operator norm of $\pi_\Omega \pi_\Theta$. We will see that on the above event $\mathcal{E}_\Omega(\eta)$, this quantity can be controlled by bounding norms of submatrices of the singular vectors U and V. This results in a number of bounds involving the following function of the rank and error probability:

$$\tau(r/m, \rho_s) \ \doteq \ 2\frac{\sqrt{r/m} + \sqrt{\rho_s}}{1 - \sqrt{r/m}}. \tag{28}$$

Where τ is used as shorthand, it should be understood as $\tau(r/m, \rho_s)$. We will repeatedly refer to the following good events:

$$\begin{aligned} \mathcal{E}_{\Omega U} &\ : \ \|\pi_\Omega \pi_{\Xi_U}\|_{F,F} \le \tau\left(r/m, \rho_s\right), \\ \mathcal{E}_{\Omega V} &\ : \ \|\pi_\Omega \pi_{\Xi_V}\|_{F,F} \le \tau\left(r/m, \rho_s\right), \\ \mathcal{E}_{\Omega \Theta} &\ : \ \|\pi_\Omega \pi_\Theta\|_{F,F} \le 2\tau\left(r/m, \rho_s\right). \end{aligned}$$

In establishing that these events are overwhelmingly likely, the following result on singular values of Gaussian matrices will prove useful:

Fact 2.1. *Let $M \in \mathbb{R}^{m\times n}$, $m > n$ and suppose that the elements of M are independent identically distributed $\mathcal{N}(0,1)$ random variables. Then*

$$\begin{aligned} \mathbb{P}\left[\sigma_{max}(M) \ge \sqrt{m} + \sqrt{n} + t\right] &\le \exp\left(-\frac{t^2}{2}\right), \\ \mathbb{P}\left[\sigma_{min}(M) \le \sqrt{m} - \sqrt{n} - t\right] &\le \exp\left(-\frac{t^2}{2}\right). \end{aligned}$$

This result is widely used in the literature, with various estimates of the error exponent. The form stated here can be obtained by via the bounds $\mathbb{E}\left[\sigma_{max}(M)\right] \le \sqrt{m}+\sqrt{n}$ and $\mathbb{E}\left[\sigma_{min}(M)\right] \ge \sqrt{m}-\sqrt{n}$, in conjunction with [6] Proposition 2.18, Equation (2.35), and the observation that the singular values are 1-Lipschitz.

Lemma 2.2. *Fix any $\eta \in (0, 1/16)$. Consider (U, V, Ω) drawn from the random orthogonal model of rank $r < m$ with error probability $\rho_s > 0$. Then there exists $t^\star(r/m, \rho_s) > 0$ such that*

$$\mathbb{P}_{U,V,\Omega}\left[\mathcal{E}_\Omega(\eta) \cap \mathcal{E}_{\Omega U} \cap \mathcal{E}_{\Omega V} \cap \mathcal{E}_{\Omega\Theta}\right] \ge 1 - 4m \exp\left(-\frac{m t^{\star 2}}{2}\right) - 4m \exp\left(-\frac{\eta^2 \rho_s^2 m}{2}\right). \tag{29}$$

In particular, if for all m larger than some $m_0 \in \mathbb{Z}$, $r/m \le \rho_r < 1$, then

$$\mathbb{P}_{U,V,\Omega}\left[\mathcal{E}_\Omega(\eta) \cap \mathcal{E}_{\Omega U} \cap \mathcal{E}_{\Omega V} \cap \mathcal{E}_{\Omega\Theta}\right] \ge 1 - \exp\left(-Cm + O(\log(m))\right). \tag{30}$$

Proof. Each of the random variables $|I_j|$ is a sum of m independent $\mathrm{Bernoulli}(\rho_s)$ random variables $X_{1,j}, X_{2,j}, \dots X_{m,j}$. The partial sum $\sum_{i=1}^k X_k - \rho_s k$ is a Martingale whose differences are all bounded by 1. So by Azuma's inequality, we have $\mathbb{P}\left[|I_j| - \rho_s m > t\right] < \exp\left(-\frac{t^2}{2m}\right)$. The same calculation clearly holds for the J_i, and so setting $t = \eta\rho_s m$,

$$\mathbb{P}\left[\mathcal{E}_\Omega(\eta)^c\right] \ \le \ 2m\mathbb{P}\left[|I_j| > \rho_s(1+\eta)m\right] \ \le \ 2m \exp\left(-\frac{\eta^2\rho_s^2 m}{2}\right). \tag{31}$$

So, with high probability each row and column of E_0 is $(1+\eta)\rho_s m$-sparse. We next show that such sparse vectors are nearly orthogonal to the random subspace $\mathrm{range}(U)$. The matrix U is uniformly distributed on $\mathbb{W}_r^m$, and can be realized by orthogonalizing a Gaussian matrix. Let $Z \in \mathbb{R}^{m\times r}$ be an iid $\mathcal{N}(0, \frac{1}{m})$ matrix; then U is equal in distribution to $Z(Z^*Z)^{-1/2}$. For each I_j,

$$\left\| Z_{I_j,\bullet}(Z^*Z)^{-1/2} \right\|_{2,2} \le \frac{\|Z_{I_j,\bullet}\|_{2,2}}{\sigma_{min}(Z)},$$

where $\sigma_{min}(Z)$ denotes the r-th singular value of the $m \times r$ matrix Z. Now, for any $t > 0$

$$\mathbb{P}\left[\sigma_{min}(Z) < 1 - \sqrt{r/m} - t\right] < \exp\left(-\frac{t^2 m}{2}\right). \tag{32}$$

Meanwhile, on the event $\mathcal{E}_\Omega(\eta)$,

$$\mathbb{P}_{Z|\Omega}\left[\|Z_{I_j,\bullet}\|_{2,2} > \sqrt{r/m} + \sqrt{|I_j|/m} + t\right] < \exp\left(-\frac{t^2 m}{2}\right). \tag{33}$$

Hence,

$$\begin{aligned}
&\mathbb{P}\left[\|U_{I_j,\bullet}\|_{2,2} > \frac{\sqrt{r/m} + \sqrt{1+\eta}\sqrt{\rho_s} + t}{1 - \sqrt{r/m} - t}\right] \\
\leq \quad & \mathbb{P}_{U|I_j}\left[\|U_{I_j,\bullet}\|_{2,2} > \frac{\sqrt{r/m} + \sqrt{1+\eta}\sqrt{\rho_s} + t}{1 - \sqrt{r/m} - t} \mid |I_j| < (1+\eta)\rho_s m\right] + \mathbb{P}\left[|I_j| \geq (1+\eta)\rho_s m\right] \\
< \quad & 2\exp\left(-\frac{mt^2}{2}\right) + \exp\left(-\frac{\eta^2 \rho_s^2 m}{2}\right).
\end{aligned}$$

Set $t^\star \leq \min\left(\frac{1}{3} - \frac{1}{3}\sqrt{\frac{r}{m}}, \sqrt{\eta\rho_s}\right)$. By the assumption of the lemma, $\sqrt{1+\eta} + \sqrt{\eta} < 4/3$, and

$$\begin{aligned}
\frac{\sqrt{r/m} + \sqrt{1+\eta}\sqrt{\rho_s} + t^\star}{1 - \sqrt{r/m} - t^\star} \quad & \leq \quad \frac{\sqrt{r/m} + \left(\sqrt{1+\eta} + \sqrt{\eta}\right)\sqrt{\rho_s}}{\frac{2}{3}\left(1 - \sqrt{r/m}\right)} \\
& \leq \quad \frac{\sqrt{r/m} + \frac{4}{3}\sqrt{\rho_s}}{\frac{2}{3}\left(1 - \sqrt{r/m}\right)} \ \leq \ \tau(r/m, \rho_s).
\end{aligned}$$

So, (applying a symmetric argument to the V_{J_i}), then, for the event $\mathcal{E}_1$ defined below,

$$\mathcal{E}_1 \ : \ \max\left(\max_j \|U_{I_j,\bullet}\|_{2,2}, \max_i \|V_{J_i,\bullet}\|_{2,2}\right) \leq \tau(r/m, \rho_s), \tag{34}$$

$$\mathbb{P}_{U,V,\Omega}\left[\mathcal{E}_1 \cap \mathcal{E}_\Omega(\eta)\right] > 1 - 4m\exp\left(-\frac{mt^{\star 2}}{2}\right) - 4m\exp\left(-\frac{\eta^2\rho_s^2 m}{2}\right). \tag{35}$$

If for all $m > m_0$, $r/m \leq \rho_r < 1$, then t^* is bounded away from zero. We next show that $\mathcal{E}_1$ implies $\mathcal{E}_{\Omega U}$, $\mathcal{E}_{\Omega V}$, and $\mathcal{E}_{\Omega\Theta}$. We can express $\pi_\Omega[M]$ in terms of its action on the columns of M: $\pi_\Omega[M] = \sum_j \pi_{I_j} M_{\bullet,j}\boldsymbol{e}_j^*$. So,

$$\pi_{\Xi_U}\pi_\Omega[M] = \pi_U \sum_j \pi_{I_j} M_{\bullet,j}\boldsymbol{e}_j^*, \tag{36}$$

$$\begin{aligned}
\text{and} \quad \|\pi_{\Xi_U}\pi_\Omega[M]\|_F^2 \quad & = \quad \sum_j \|\pi_U \pi_{I_j} M_{\bullet,j}\|_2^2 \ = \ \sum_j \|U^* \pi_{I_j} M_{\bullet,j}\|_2^2 \\
& \leq \quad \sum_j \|U_{I_j,\bullet}\|_{2,2}^2 \|M_{\bullet,j}\|_2^2 \ \leq \ \left(\max_j \|U_{I_j,\bullet}\|_{2,2}\right)^2 \|M\|_F^2,
\end{aligned}$$

and so on $\mathcal{E}_1$, $\|\pi_\Omega\pi_{\Xi_U}\|_{F,F} = \|\pi_{\Xi_U}\pi_\Omega\|_{F,F} \leq \tau(r/m, \rho_s)$; and so $\mathcal{E}_1 \implies \mathcal{E}_{\Omega U}$. A symmetric argument establishes that $\mathcal{E}_1 \implies \mathcal{E}_{\Omega V}$. Now, notice that since

$$\pi_\Omega\pi_\Theta[M] = \pi_\Omega\pi_{\Xi_U}[M] + \pi_\Omega\left[\pi_{U^\perp} M \pi_V\right], \tag{37}$$

if we choose a basis $B \in \mathbb{R}^{m-r}$ for the orthogonal complement of $\mathrm{range}(U)$ and define $\Xi_{U^\perp V} \doteq \{BQV^* \mid Q \in \mathbb{R}^{m-r\times r}\} \subset \mathbb{R}^{m\times m}$, then

$$\|\pi_\Omega\pi_\Theta\|_{F,F} \leq \|\pi_\Omega\pi_{\Xi_U}\|_{F,F} + \left\|\pi_\Omega\pi_{\Xi_{U^\perp V}}\right\|_{F,F}. \tag{38}$$

Since

$$\begin{aligned}
\left\|\pi_\Omega\pi_{\Xi_{U^\perp V}}\right\|_{F,F} \quad & = \quad \sup_{M\in\Xi_{U^\perp V}\setminus\{0\}} \|\pi_\Omega[M]\|_F \\
& \leq \quad \sup_{M\in\Xi_V\setminus\{0\}} \|\pi_\Omega[M]\|_F \ = \ \|\pi_\Omega\pi_{\Xi_V}\|_{F,F},
\end{aligned}$$

$\mathcal{E}_1 \implies \mathcal{E}_{\Omega\Theta}$ and the proof is complete. $\square$

We will need to understand concentration of Lipschitz functions of matrices that are uniformly distributed (according to the Haar measure) on two manifolds of interest: the Stiefel manifold $\mathbb{W}_r^m$ and the group of $r \times r$ orthogonal matrices with determinant one, $SO(r)$. This is governed by the *concentration function* on these spaces:

Definition 2.3. *[6] Let (X, d) be a metric space. For a given measure μ on X, the concentration function is defined as*

$$\alpha_{X,d,\mu}(t) \doteq \sup\left\{1 - \mu(A_t) \mid A \subset X, \mu(A) \geq \tfrac{1}{2}\right\}, \tag{39}$$

where $A_t = \{x \mid d(x, A) < t\}$ is a t-neighborhood of A.

The concentration functions for $\mathbb{W}_r^m$ and $SO(r)$ are well known:

Fact 2.4 ([7] Theorems 6.5.1 and 6.7.1)**.** *For $r < m$ the manifold $\mathbb{W}_r^m$, with distance $d(X, Y) \doteq \|X - Y\|_F$, the Haar measure μ has concentration function*

$$\alpha_{\mathbb{W},d,\mu}(t) \leq \sqrt{\frac{\pi}{8}} \exp\left(-\frac{mt^2}{8}\right). \tag{40}$$

Similarly, on $SO(r)$ with $\delta(X, Y) \doteq \|X - Y\|_F$, and ν the Haar measure,

$$\alpha_{SO(r),\delta,\nu}(t) \leq \sqrt{\frac{\pi}{8}} \exp\left(-\frac{rt^2}{8}\right). \tag{41}$$

Propositions 1.3 and 1.8 of [6] then imply that Lipschitz functions on these spaces concentrate about their medians and expectations:

Corollary 2.5. *Suppose $r < m$, and let $f : \mathbb{R}^{m \times r} \to \mathbb{R}$ with Lipschitz norm*

$$\|f\|_{lip} \doteq \sup_{X \neq Y} \frac{|f(X) - f(Y)|}{\|X - Y\|_F}. \tag{42}$$

Then if U is distributed according to the Haar measure on $\mathbb{W}_r^m$, and $\mathrm{median}(f)$ *denotes any median,*

$$\mathbb{P}\left[f(U) \geq \mathrm{median}(f) + t\right] \leq \exp\left(-\frac{mt^2}{8\|f\|_{lip}^2}\right). \tag{43}$$

Similarly, if $g : \mathbb{R}^{r \times r} \to \mathbb{R}^m$ with Lipschitz norm

$$\|g\|_{lip} \doteq \sup_{X \neq Y} \frac{|g(X) - g(Y)|}{\|X - Y\|_F}. \tag{44}$$

Then if R is distributed according to the Haar measure on $SO(r)$, g satisfies the following tail bound:

$$\mathbb{P}\left[|g(R) - \mathbb{E}[g(R)]| \geq \|g\|_{lip}\sqrt{\tfrac{8\pi}{r}} + t\right] \quad \leq \quad 2\,\exp\left(-\frac{rt^2}{8\|g\|_{lip}^2}\right). \tag{45}$$

Proof. The concentration result (43) is a restatement of [6] Proposition 1.3. For (45), notice first that [6] Proposition 1.3 implies that

$$\mathbb{P}\left[|g(R) - \mathrm{median}(g)| \geq t\right] \leq 2\exp\left(-\frac{rt^2}{8\|g\|_{lip}^2}\right). \tag{46}$$

If we set

$$\bar{a} \;\doteq\; \int_0^\infty 2\exp\left(-\frac{rt^2}{8\|g\|_{lip}^2}\right) dt \;=\; \|g\|_{lip}\sqrt{\frac{8\pi}{r}}, \tag{47}$$

then [6] Proposition 1.8 gives

$$\mathbb{P}\left[|g(R) - \mathbb{E}[g]| \geq \bar{a} + t\right] \leq 2\exp\left(-\frac{rt^2}{8\|g\|_{lip}^2}\right). \tag{48}$$

□

2.2 Bounding the operator norm

In this section, we begin our analysis of the minimum Frobenius norm solution to the equality constraints of the feasibility problem (11). We show that with overwhelming probability this matrix also has operator norm bounded by a small constant. An important byproduct of this analysis is a simple proof for that, with the same models studied here, the easier problem of *matrix completion* – filling missing entries in a low-rank matrix – can be efficiently and exactly solved by convex optimization, even for cases when the rank of the matrix grows in proportion to the dimensionality. Section 3 further elucidates connections to that problem.

2.2.1 General approach

We begin by showing that with high probability the minimum Frobenius norm solution is unique, and giving an explicit representation for the pseudoinverse operator in that case. This operator, denoted $\mathcal{H}^\dagger$, is applied to the matrix $\lambda\,\mathrm{sign}[E_0] - UV^*$ to give the initial dual vector. We separately bound the norm of the two terms induced by $\|\mathcal{H}^\dagger \mathrm{sign}[E_0]\|_{2,2}$ and $\|\mathcal{H}^\dagger[UV^*]\|_{2,2}$, respectively. Both arguments follow in a fairly straightforward manner by reducing to a net and then applying concentration inequalities. Throughout this section, N will denote a $\frac{1}{2}$-net for $\mathbb{S}^{m-1}$. By [6] Lemma 3.18, there is such a net with size at most $\exp(4m)$. Moving from $\|A\|_{2,2} = \sup_{\boldsymbol{x},\boldsymbol{y}\in\mathbb{S}^{m-1}} \boldsymbol{x}^* A \boldsymbol{y}$ to our product of nets loses at most a constant factor in the estimate: $\|A\|_{2,2} \le 4\,\sup_{\boldsymbol{x},\boldsymbol{y}\in N} \boldsymbol{x}^* A\boldsymbol{y}$ (see e.g., [8] Proposition 2.6). We will argue that for our A of interest, $f(A) = \boldsymbol{x}^* A \boldsymbol{y}$ concentrates, and union bound over all $\exp(8m)$ pairs in $N \times N$.

2.2.2 Representation and uniqueness of W_0.

Lemma 2.6 (Representation of the pseudoinverse)**.** *Suppose that we have $\|\pi_\Theta \pi_\Omega\|_{F,F} < 1$. Then the operator $\mathtt{I} - \pi_\Omega\pi_\Theta\pi_\Omega$ is invertible, and for any M the optimization problem*

$$\min \|W\|_F \quad \text{subj} \quad \pi_\Theta[W] = 0, \;\; \pi_\Omega[W] = \pi_\Omega[M] \tag{49}$$

has unique solution

$$\hat{W} \;=\; \pi_{\Theta^\perp}\pi_\Omega \left(\mathtt{I} - \pi_\Omega\pi_\Theta\pi_\Omega\right)^{-1} \pi_\Omega[M] \;=\; \pi_{\Theta^\perp}\pi_\Omega \sum_{k=0}^{\infty} (\pi_\Omega\pi_\Theta\pi_\Omega)^k \pi_\Omega[M].$$

Proof. Choose matrices $U^\perp, V^\perp \in \mathbb{R}^{m\times m-r}$ whose columns form orthonormal bases for the orthogonal complement of the ranges of U and V, respectively. Let Q denote the matrix

$$Q \doteq \begin{bmatrix} {V^\perp}^* \otimes U^* \\ V^* \otimes U^* \\ V^* \otimes {U^\perp}^* \end{bmatrix}.$$

Notice that the rows of Q are orthonormal, and that they form a basis for the subspace of vectors $\{\mathrm{vec}[M] \mid M \in \Theta\}$. The equality constraint in (49) can therefore be expressed as

$$\begin{bmatrix} Q \\ \mathtt{I}_{\Omega,\bullet} \end{bmatrix} \mathrm{vec}(W) = \begin{bmatrix} \mathbf{0} \\ \mathrm{vec}_\Omega[M] \end{bmatrix}, \tag{50}$$

Here, $\mathtt{I}$ is the $m^2 \times m^2$ identity matrix, and $\mathtt{I}_{\Omega,\bullet}$ is the submatrix of $\mathtt{I}$ consisting of those rows indexed by Ω, taken in lexicographic order. The minimum Frobenius norm solution W is simply the minimum ℓ^2 norm solution to this system of equations. Define the matrix

$$P \doteq Q\,\mathtt{I}_{\bullet,\Omega}.$$

Notice that for any matrix M,

$$Q^* P\,\mathtt{I}_{\Omega,\bullet}\mathrm{vec}[M] = \mathrm{vec}\left[\pi_\Theta\pi_\Omega M\right].$$

Since Q and $\mathtt{I}_{\Omega,\bullet}$ each have orthonormal rows, $\|\pi_\Theta\pi_\Omega\|_{F,F} = \|Q^* P \mathtt{I}_{\Omega,\bullet}\|_{2,2} = \|P\|_{2,2}$. So, by the assumption of the lemma $\|P\|_{2,2} < 1$, and the matrix $\begin{bmatrix} \mathtt{I} & P \\ P^* & \mathtt{I} \end{bmatrix}$ is nonsingular. We therefore have the following explicit expression for the unique minimum ℓ^2-norm solution to (50):

$$\mathrm{vec}[\hat{W}] = \begin{bmatrix} Q^* & \mathtt{I}_{\bullet,\Omega} \end{bmatrix} \begin{bmatrix} \mathtt{I} & P \\ P^* & \mathtt{I} \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{0} \\ \mathrm{vec}_\Omega[M] \end{bmatrix}. \tag{51}$$

Applying the Schur complement formula (which is justified since $\mathtt{I} \succ P^*P$), the above is equal to

$$\begin{aligned}
&\left[\begin{array}{cc} Q^* & \mathtt{I}_{\bullet,\Omega} \end{array}\right] \left[\begin{array}{c} -P \\ \mathtt{I} \end{array}\right] (\mathtt{I} - P^*P)^{-1} \mathrm{vec}_\Omega[M] \\
&= \quad (\mathtt{I} - Q^*Q)\mathtt{I}_{\bullet,\Omega}(\mathtt{I} - P^*P)^{-1}\mathtt{I}_{\Omega,\bullet}\mathrm{vec}[M] \\
&= \quad (\mathtt{I} - Q^*Q)\mathtt{I}_{\bullet,\Omega}\sum_{k=0}^{\infty}(\mathtt{I}_{\Omega,\bullet}Q^*Q\mathtt{I}_{\bullet,\Omega})^k\mathtt{I}_{\Omega,\bullet}\mathrm{vec}[M] \\
&= \quad \mathrm{vec}\left[\pi_{\Theta^\perp}\pi_\Omega\sum_{k=0}^{\infty}(\pi_\Omega\pi_\Theta\pi_\Omega)^k\pi_\Omega[M]\right],
\end{aligned}$$

yielding the representation in the statement of the lemma. □

2.2.3 Effect of the singular vectors

We next analyze the part of the initial dual vector W_0 induced by UV^*. From the previous lemma we have the expression

$$\mathcal{H}^\dagger[\cdot] = \pi_{\Theta^\perp}\pi_\Omega\sum_{k=0}^{\infty}(\pi_\Omega\pi_\Theta\pi_\Omega)^k\pi_\Omega[\cdot], \tag{52}$$

whenever $\|\pi_\Omega\pi_\Theta\|_{F,F} < 1$. Analysis of $\mathcal{H}^\dagger[UV^*]$ is complicated by the fact that UV^* and Θ are dependent random variables. Notice, however, that π_Θ depends only on the *subspace spanned* by the columns of U, not on any choice of basis for that subspace. UV^*, on the other hand, does depend on the choice of basis. We will use this fact to decouple $\mathcal{H}^\dagger$ and UV^*, and show that the ℓ^2 operator norm of $\mathcal{H}^\dagger[UV^*]$ is bounded below a small constant with overwhelming probability.

Lemma 2.7. *Let (U, V, Ω) be sampled from the random orthogonal model of rank r with error probability ρ_s. Suppose that r and ρ_s satisfy*

$$\tau(r/m, \rho_s) < \tfrac{1}{4}. \tag{53}$$

Then with probability at least $1 - \exp(-Cm + O(\log m))$, the solution W_0^{UV} to the optimization problem

$$\min \|W\|_F \quad \mathrm{subj} \quad U^*W = 0,\ WV = 0,\ \pi_\Omega[W] = -\pi_\Omega[UV^*] \tag{54}$$

is unique, and satisfies

$$\left\|W_0^{UV}\right\|_{2,2} \le 64\,\tau\left(\frac{r}{m}, \rho_s\right)\left(1 + \frac{8}{3}\sqrt{\frac{m}{m-r-1}}\right) + \frac{8}{3}\sqrt{\frac{8\pi}{m-r-1}}. \tag{55}$$

Proof. First consider the event

$$\mathcal{E}_1 \ : \ \|\pi_\Omega[UV^*]\|_{2,2} \le 64\tau.$$

Fix $\boldsymbol{x}, \boldsymbol{y} \in N$ and write

$$\boldsymbol{x}^*\pi_\Omega[UV^*]\boldsymbol{y} = \langle U^*\pi_\Omega[\boldsymbol{x}\boldsymbol{y}^*], V^*\rangle\,.$$

On the event $\mathcal{E}_{\Omega U}$, $\|U^*\pi_\Omega[\boldsymbol{x}\boldsymbol{y}^*]\|_F \le \tau$. So, as a function of V, $f(V) \doteq \boldsymbol{x}^*\pi_\Omega[UV^*]\boldsymbol{y}$ is τ-Lipschitz. Since the distribution of V is invariant under the orthogonal transformation $V \mapsto -V$, $f(V)$ is a symmetric random variable and 0 is a median. Hence, on the event $\mathcal{E}_{\Omega U}$ the tail bound (43) implies that

$$\mathbb{P}_{V|U,\Omega}\left[f(V) > t\right] < \exp\left(-\frac{mt^2}{8\tau^2}\right). \tag{56}$$

Set $t = 16\tau$. A union bound over the $\le \exp(8m)$ elements of $N \times N$ shows that on $\mathcal{E}_{\Omega U}$,

$$\begin{aligned}
\mathbb{P}_{V|U,\Omega}\left[\,\|\pi_\Omega[UV^*]\|_{2,2} > 64\tau\right] \quad &\le \quad \mathbb{P}_{V|U,\Omega}\left[\sup_{\boldsymbol{x},\boldsymbol{y}\in N\times N}\boldsymbol{x}^*\pi_\Omega[UV^*]\boldsymbol{y} \ > \ 16\tau\right] \\
&\le \quad \exp\left(-24m\right),
\end{aligned}$$

and hence we can conclude that

$$\begin{aligned}
\mathbb{P}_{U,V,\Omega}\left[\mathcal{E}_1\right] \quad &\ge \quad 1 - \exp\left(-24m\right) - \mathbb{P}[\mathcal{E}_{\Omega U}^c] \\
&\ge \quad 1 - \exp\left(-Cm + O(\log m)\right).
\end{aligned}$$

Now, consider the combined event $\mathcal{E}_1 \cap \mathcal{E}_{\Omega\Theta}$. On this event, the representation in Lemma 2.6 is valid and

$$W_0^{UV} = -\pi_{\Theta^\perp}\pi_\Omega \sum_{k=0}^{\infty} (\pi_\Omega\pi_\Theta\pi_\Omega)^k \pi_\Omega[UV^*]. \tag{57}$$

For any M, $\pi_{\Theta^\perp}[M] = \pi_{U^\perp} M \pi_{V^\perp}$, $\|\pi_{\Theta^\perp}[M]\|_{2,2} \leq \|M\|_{2,2}$, so

$$\begin{aligned} \|W_0^{UV}\|_{2,2} &\leq \left\| \pi_\Omega \sum_{k=0}^{\infty} (\pi_\Omega\pi_\Theta\pi_\Omega)^k \pi_\Omega[UV^*] \right\|_{2,2} \\ &\leq \|\pi_\Omega\,[UV^*]\|_{2,2} + \left\| \sum_{k=1}^{\infty} (\pi_\Omega\pi_\Theta\pi_\Omega)^k [UV^*] \right\|_{2,2}. \end{aligned} \tag{58}$$

We have already addressed the first term. To bound the second, we expand our probability space as follows. Consider $\tilde{U}$ and $\tilde{V}$ distributed according to the Haar measure on $\mathbb{W}^m_{m-1}$. Identify U and V with the first r columns of $\tilde{U}$ and $\tilde{V}$, respectively, and write $\hat{U}$ and $\hat{V}$ for the remaining $m-1-r$ columns. Notice that U and V are indeed distributed according to the random orthogonal model of rank r, and that $\hat{U}$ and $\hat{V}$ follow the random orthogonal model of rank $m-r-1$ (although, of course, U and $\hat{U}$ now dependent random variables). Write

$$\begin{aligned} \left\| \sum_{k=1}^{\infty} (\pi_\Omega\pi_\Theta\pi_\Omega)^k \,[UV^*] \right\|_{2,2} &= \left\| \sum_{k=1}^{\infty} (\pi_\Omega\pi_\Theta\pi_\Omega)^k \left[\tilde{U}\tilde{V}^* - \hat{U}\hat{V}^*\right] \right\|_{2,2} \\ &\leq \left\| \sum_{k=1}^{\infty} (\pi_\Omega\pi_\Theta\pi_\Omega)^k \left[\tilde{U}\tilde{V}^*\right] \right\|_{2,2} + \left\| \sum_{k=1}^{\infty} (\pi_\Omega\pi_\Theta\pi_\Omega)^k \left[\hat{U}\hat{V}^*\right] \right\|_{2,2}. \end{aligned}$$

We next show that with overwhelming probability each of these terms is well-controlled. Notice that if $R \in SO\,(m-r-1)$ is any orthogonal matrix, then the joint distribution $\tilde{U}$, U, and $\hat{U}$ is invariant under the map

$$\tilde{U} \mapsto \tilde{U} \begin{bmatrix} \mathtt{I} & 0 \\ 0 & R \end{bmatrix}. \tag{59}$$

This follows from the right orthogonal invariance of the Haar measure on $\mathbb{W}^m_{m-1}$ (see e.g., [5] Section 1.4.3). Since this map preserves U and V, it also preserves Θ. Hence, the term of interest, $\left\|\sum_{k=1}^{\infty} (\pi_\Omega\pi_\Theta\pi_\Omega)^k \left[\hat{U}\hat{V}^*\right]\right\|_{2,2}$ is equal in distribution to $\left\|\sum_{k=1}^{\infty} (\pi_\Omega\pi_\Theta\pi_\Omega)^k \left[\hat{U}R\hat{V}^*\right]\right\|_{2,2}$. The orthogonal matrix R is independent of all of the other terms in this expression. This independence allows us to estimate the norm by first bounding the operator norm of the map $\sum_{k=1}^{\infty} (\pi_\Omega\pi_\Theta\pi_\Omega)^k$ and then applying measure concentration on $SO\,(m-r-1)$. For fixed $\boldsymbol{x}, \boldsymbol{y} \in N$, consider the quantity

$$\begin{aligned} \boldsymbol{x}^* \sum_{k=1}^{\infty} (\pi_\Omega\pi_\Theta\pi_\Omega)^k \left[\hat{U}R\hat{V}^*\right] \boldsymbol{y} &= \left\langle \hat{U}^* \left(\sum_{k=1}^{\infty} (\pi_\Omega\pi_\Theta\pi_\Omega)^k [\boldsymbol{x}\boldsymbol{y}^*] \right) \hat{V}, R \right\rangle \\ &\doteq \langle M, R \rangle\,. \end{aligned}$$

This is a $\|M\|_F$-Lipschitz function of R. Since for any $\boldsymbol{w} \in \mathbb{S}^{m-1}$, $R\boldsymbol{w}$ is uniformly distributed on $\mathbb{S}^{m-1}$, $\mathbb{E}[\boldsymbol{e}_i^* R\boldsymbol{w}] = 0$ for all i. So, $\mathbb{E}[R] = 0$ and $\mathbb{E}_{R|M}\,\langle M, R\rangle = 0$. On the event $\mathcal{E}_{\Omega\Theta}$,

$$\|M\|_F \;\leq\; \sum_{k=1}^{\infty} (\pi_\Omega\pi_\Theta\pi_\Omega)^k [\boldsymbol{x}\boldsymbol{y}^*] \;\leq\; \frac{\|\pi_\Omega\pi_\Theta\pi_\Omega\|_{F,F}}{1 - \|\pi_\Omega\pi_\Theta\pi_\Omega\|_{F,F}} \|\boldsymbol{x}\boldsymbol{y}^*\|_F \;\leq\; \frac{4\tau^2}{1-4\tau^2} \;\leq\; \frac{4\tau}{3}. \tag{60}$$

Hence, on $\mathcal{E}_{\Omega\Theta}$ by (45)

$$\mathbb{P}_{R|M}\left[\langle M, R\rangle > \gamma(m) + t\right] \;<\; 2\,\exp\left(-\frac{9(m-r-1)t^2}{128 \times \tau^2}\right), \tag{61}$$

where $\gamma(m) \;\doteq\; \frac{4\tau\left(\frac{r}{m},\rho_s\right)^2}{1-4\tau\left(\frac{r}{m},\rho_s\right)^2}\sqrt{\frac{8\pi}{m-r-1}} \;\leq\; \frac{1}{3}\sqrt{\frac{8\pi}{m-r-1}}$. For compactness of notation, set $\beta(m,r) \;\doteq\; \sqrt{\frac{m-r-1}{m}}$. Then, setting $t = \frac{64\tau}{3\beta}$, union bounding over the $\leq \exp(8m)$ pairs in $N \times N$, and then moving from $N \times N$ to $\mathbb{S}^{m-1} \times \mathbb{S}^{m-1}$ (losing at most a factor of 4) gives

$$\mathbb{P}_{R|\tilde{U},V,\Omega}\left[\left\|\sum_{k=1}^{\infty}(\pi_\Omega\pi_\Theta\pi_\Omega)^k\left[\hat{U}R\hat{V}^*\right]\right\|_{2,2} > \frac{256\,\tau}{3\beta(m,r)} + 4\gamma(m)\right] \;<\; 2\,\exp\left(-24m\right). \tag{62}$$

And so,

$$\begin{aligned}
&\mathbb{P}_{\tilde{U},V,\Omega}\left[\left\|\sum_{k=1}^{\infty}(\pi_\Omega\pi_\Theta\pi_\Omega)^k\left[\hat{U}\hat{V}^*\right]\right\|_{2,2} > \frac{256\,\tau}{3\beta(m,r)} + 4\gamma(m)\right] \\
&= \;\; \mathbb{P}_{R|\tilde{U},V,\Omega}\left[\left\|\sum_{k=1}^{\infty}(\pi_\Omega\pi_\Theta\pi_\Omega)^k\left[\hat{U}R\hat{V}^*\right]\right\|_{2,2} > \frac{256\,\tau}{3\beta(m,r)} + 4\gamma(m)\right] \\
&\leq \;\; 2\,\exp\left(-24m\right) + \mathbb{P}\left[\mathcal{E}_{\Omega\Theta}^c\right] \;\leq\; \exp\left(-Cm + O(\log m)\right).
\end{aligned}$$

An identical argument, this time randomizing over $\tilde{R} \doteq \begin{bmatrix} \mathtt{I} & 0 \\ 0 & R \end{bmatrix}$ shows that

$$\mathbb{P}_{\tilde{U},V,\Omega}\left[\left\|\sum_{k=1}^{\infty}(\pi_\Omega\pi_\Theta\pi_\Omega)^k\left[\tilde{U}\tilde{V}^*\right]\right\|_{2,2} > \frac{256\,\tau}{3\beta(m,r)} + 4\gamma(m)\right] \;\leq\; \exp\left(-Cm + O(\log m)\right). \tag{63}$$

Combining terms yields the bound. □

2.2.4 Effect of the error signs

We next consider the effect of the error signs. As in the previous proposition, we handle the $k = 0$ and $k = 1 \ldots \infty$ parts of the representation in Lemma 2.6 separately.

Lemma 2.8. *There exists a function $\phi(\rho_s)$ satisfying $\lim_{\rho_s \searrow 0} \phi(\rho_s) = 0$, such that if E_0 is distributed according to the Bernoulli sign and support model with error probability ρ_s.*

$$\mathbb{P}_{\Omega,\Sigma}\left[\|\mathrm{sign}(E_0)\|_{2,2} \geq \phi(\rho_s)\sqrt{m}\right] \;\leq\; \exp\left(-Cm\right). \tag{64}$$

Proof. We first provide a bound on the moment-generating-function for the iid random variables $Y_{ij} \doteq \mathrm{sign}(E_{0ij})$ of the form $\mathbb{E}[\exp(tY)] \leq \exp(\alpha t^2)$. The moments of Y are

$$\mathbb{E}[Y^k] = \begin{cases} 1 & k = 0, \\ 0 & k \text{ odd}, \\ \rho_s & k > 0,\ k \text{ even}, \end{cases}$$

so $\mathbb{E}[\exp(tY)] = 1 + \sum_{k=1}^{\infty} \frac{\rho_s t^{2k}}{(2k)!}$. Since $\exp(\alpha t^2) = 1 + \sum_{k=1}^{\infty} \frac{\alpha^k t^{2k}}{k!}$, it suffices to choose α such that

$$\rho_s \leq \frac{(2k)!}{k!}\alpha^k \quad \forall\ k = 1, 2, \ldots.$$

Since $\frac{(2k)!}{k!} \geq k^k$, $\alpha \geq \max_{k=1,2,\ldots} \frac{1}{k}\rho_s^{\frac{1}{k}}$ suffices. Consider the function $\psi : [1,\infty) \times [0,1] \to \mathbb{R}$ defined by $\psi(x,y) = \frac{1}{x}y^{\frac{1}{x}}$. $\psi(1,y) = y$, and for all y $\lim_{x\to\infty}\psi(x,y) = 0$. The only stationary point occurs at $x^\star = \log(1/y)$, and hence for $y > 0$ its maximum on $[1,\infty)$ is $\psi(x^\star(y),y) = \max\left(y, \frac{1}{\log(y^{-1})} y^{\frac{1}{\log(y^{-1})}}\right)$. Notice that $\lim_{y\searrow 0}\psi(x^\star(y),y) = 0$, and that

$$\mathbb{E}[\exp(tY)] \leq \exp(\psi(x^\star(\rho_s),\rho_s)t^2). \tag{65}$$

Now, for any fixed pair $\boldsymbol{x},\boldsymbol{y} \in N$, let $Z \doteq \boldsymbol{x}^*\ \mathrm{sign}(E_0)\ \boldsymbol{y}$. Then

$$\begin{aligned}
\mathbb{E}[\exp(tZ)] &= \prod_{ij}\mathbb{E}[\exp(tx_iy_jY_{ij})] \\
&\leq \prod_{ij}\exp(\psi(x^\star(\rho_s),\rho_s)t^2(x_iy_j)^2) = \exp(\psi(x^\star(\rho_s),\rho_s)t^2).
\end{aligned}$$

Applying a Chernoff bound (and optimizing the exponent) gives

$$\mathbb{P}\left[\boldsymbol{x}^* \operatorname{sign}(E_0)\boldsymbol{y} > t\right] < \exp\left(-\frac{t^2}{4\psi(x^\star(\rho_s),\rho_s)}\right). \tag{66}$$

Union bounding (and recognizing that moving from $N \times N$ to $\mathbb{S}^{m-1}$ loses at most a factor of 4) gives

$$\mathbb{P}\left[\|\operatorname{sign}(E_0)\|_{2,2} \geq 4t\sqrt{m}\right] \quad \leq \quad \exp\left(8m - m\frac{t^2}{4\psi(x^\star(\rho_s),\rho_s)}\right). \tag{67}$$

If we set, e.g., $t(\rho_s) = 8\psi^{1/2}(x^\star(\rho_s),\rho_s)$, the probability of failure will be bounded by $\exp(-8m)$. Further setting $\phi(\rho_s) = 4t(\rho_s)$ gives the statement of the lemma. $\square$

The above lemma goes part of the way to controlling the component of the initial dual vector induced by the errors, $\mathcal{H}^\dagger[\lambda \operatorname{sign}(E_0)]$. A straightforward Martingale argument, detailed in the following lemma, completes the proof.

Lemma 2.9. *Consider (U,V,Ω,Σ) drawn from the random orthogonal model of rank $r < m$ with Bernoulli error probability ρ_s and random signs, and suppose that r and ρ_s satisfy*

$$\tau\left(r/m,\rho_s\right) \;\leq\; \tfrac{1}{4}. \tag{68}$$

Then with probability at least $1-\exp(-Cm+O(\log m))$ in (U,V,Ω,Σ), the solution W_0^E to the optimization problem

$$\min \|W\|_F \quad \text{subj} \quad U^*W = 0,\; WV = 0,\; \pi_\Omega[W] = -\pi_\Omega[\lambda \operatorname{sign}(E_0)] \tag{69}$$

is unique, and satisfies

$$\left\|W_0^E\right\|_{2,2} \;\leq\; \phi(\rho_s) + \frac{128\tau(r/m,\rho_s)}{3}, \tag{70}$$

where $\phi(\cdot)$ is the function defined in Lemma 2.8.

Proof. On event $\mathcal{E}_{\Omega\Theta}$, the minimizer is unique, and can be expressed as

$$W_0^E = \pi_{\Theta^\perp}\pi_\Omega \sum_{k=0}^{\infty}(\pi_\Omega\pi_\Theta\pi_\Omega)^k \pi_\Omega[\lambda \operatorname{sign}(E_0)].$$

Since $\pi_{\Theta^\perp}$ is a contraction in the $(2,2)$ norm,

$$\begin{aligned}\|W_0^E\|_{2,2} \;&\leq\; \left\|\pi_\Omega \sum_{k=0}^{\infty}(\pi_\Omega\pi_\Theta\pi_\Omega)^k \pi_\Omega[\lambda \operatorname{sign}(E_0)]\right\|_{2,2} \\ &\leq\; \|\pi_\Omega[\lambda \operatorname{sign}[E_0]]\|_{2,2} + \left\|\sum_{k=1}^{\infty}(\pi_\Omega\pi_\Theta\pi_\Omega)^k \pi_\Omega[\lambda \operatorname{sign}(E_0)]\right\|_{2,2}.\end{aligned}$$

Lemma 2.8 controls the first term below $\phi(\rho_s)$ with overwhelming probability. For the second term, it is convenient to treat $\operatorname{sign}(E_0)$ as the projection of an $m \times m$ iid Rademacher matrix Σ onto Ω: $\operatorname{sign}(E_0) = \pi_\Omega[\Sigma]$. Fix $\boldsymbol{x},\boldsymbol{y} \in N$, and notice that

$$\begin{aligned}\boldsymbol{x}^* \sum_{k=1}^{\infty}(\pi_\Omega\pi_\Theta\pi_\Omega)^k \pi_\Omega[\lambda \operatorname{sign}(E_0)]\boldsymbol{y} \;&=\; \left\langle \lambda \sum_{k=1}^{\infty}(\pi_\Omega\pi_\Theta\pi_\Omega)^k[\boldsymbol{x}\boldsymbol{y}^*], \Sigma \right\rangle. \\ &\doteq\; \langle M, \Sigma\rangle.\end{aligned}$$

On the event $\mathcal{E}_{\Omega\Theta}$, $M \in \mathbb{R}^{m\times m}$ has Frobenius norm at most $\frac{1}{\sqrt{m}}\frac{4\tau^2}{1-4\tau^2} \;\leq\; \frac{4\tau}{3\sqrt{m}}$. Order the indices (i,j) $1 \leq i,j \leq m$ arbitrarily, and consider the Martingale Z defined by $Z_0 = 0$, $Z_k = Z_{k-1} + M_{i_k,j_k}\Sigma_{i_k,j_k}$. The k-th Martingale difference is bounded by $|M_{i_k,j_k}|$, and the overall squared ℓ^2-norm of the differences is bounded by $\|M\|_F^2$. Hence, by Azuma's inequality,

$$\mathbb{P}\left[\langle M,\Sigma\rangle > t\right] \;<\; \exp\left(-\frac{9mt^2}{32\tau^2}\right). \tag{71}$$

Set $t = \frac{32\tau}{3}$. A union bound over the $\leq \exp(8m)$ elements in $N \times N$ shows that on the event $\mathcal{E}_{\Omega\Theta}$,

$$\mathbb{P}_{\Sigma|U,V,\Omega}\left[\left\|\sum_{k=1}^{\infty}(\pi_\Omega\pi_\Theta\pi_\Omega)^k\pi_\Omega[\lambda\,\mathrm{sign}(E_0)]\right\|_{2,2} > \frac{128\tau}{3}\right] \leq \exp\left(-24m\right). \tag{72}$$

Hence, $\mathbb{P}_{U,V,\Omega,\Sigma}\left[\left\|\sum_{k=1}^{\infty}(\pi_\Omega\pi_\Theta\pi_\Omega)^k\pi_\Omega[\lambda\,\mathrm{sign}(E_0)]\right\|_{2,2} > \frac{128\tau}{3}\right]$ is bounded by

$$\exp\left(-24m\right) + \mathbb{P}[\mathcal{E}_{\Omega\Theta}^c] \;\leq\; \exp\left(-Cm + O(\log m)\right).$$

□

2.3 Controlled feedback for sparse matrices

We next bound the operator norm of the projection π_Γ, restricted to row- and column-sparse matrices Ψ_c.

Lemma 2.10 (Representation of iterates)**.** *Let Θ, Γ, Ω be defined as above. Suppose that $\|\pi_\Omega\pi_\Theta\|_{F,F} < 1$. Then for all $M \in \Omega^\perp$,*

$$\pi_\Gamma[M] \;=\; \pi_{\Omega^\perp}\pi_\Theta\sum_{k=0}^{\infty}(\pi_\Theta\pi_\Omega\pi_\Theta)^k\pi_\Theta[M]. \tag{73}$$

Proof. For $M \in \Omega^\perp$,

$$\begin{aligned}
\mathrm{vec}[\pi_\Gamma M] &= [Q^* \mid \mathtt{I}_{\bullet,\Omega}]\begin{bmatrix}\mathtt{I} & P\\ P^* & \mathtt{I}\end{bmatrix}^{-1}\begin{bmatrix}Q\\ \mathtt{I}_{\Omega,\bullet}\end{bmatrix}\mathrm{vec}[M]\\
&= [Q^* \mid \mathtt{I}_{\bullet,\Omega}]\begin{bmatrix}\mathtt{I} & P\\ P^* & \mathtt{I}\end{bmatrix}^{-1}\begin{bmatrix}Q\,\mathrm{vec}[M]\\ \mathbf{0}\end{bmatrix}.
\end{aligned}$$

Under the condition of the lemma, $\|P\|_{2,2} < 1$ so the above inverse is indeed well-defined, and can be expressed via the Schur complement formula:

$$\begin{aligned}
\mathrm{vec}[\pi_\Gamma M] &= Q^*(\mathtt{I} - PP^*)^{-1}Q\mathrm{vec}[M] - \mathtt{I}_{\bullet,\Omega}P^*(\mathtt{I} - PP^*)^{-1}Q\mathrm{vec}[M]\\
&= \pi_{\Omega^\perp}Q^*(\mathtt{I} - PP^*)^{-1}Q\mathrm{vec}[M]\\
&= \pi_{\Omega^\perp}Q^*\sum_{k=0}^{\infty}(PP^*)^kQ\mathrm{vec}[M].
\end{aligned}$$

Recognizing that

$$Q^*\sum_{k=0}^{\infty}(PP^*)^kQ\mathrm{vec}[M] = \mathrm{vec}\left[\pi_\Theta\sum_{k=0}^{\infty}(\pi_\Theta\pi_\Omega\pi_\Theta)^k\pi_\Theta[M]\right]$$

completes the proof. □

Lemma 2.11. *Fix any $c \in (0,1)$. Consider (U,V,Ω) drawn from the random orthogonal model of rank $r < m$, with Bernoulli error probability ρ_s. Further suppose that r and ρ_s satisfy*

$$\tau\left(\tfrac{r}{m}, \rho_s\right) \;\leq\; \frac{1}{4}. \tag{74}$$

Then with probability at least $1 - \exp\left(-Cm + O(\log m)\right)$,

$$\xi_c \;\leq\; \frac{32}{9}\left(\sqrt{\tfrac{r}{m}} + \sqrt{c} + 2\sqrt{H(c)}\right), \tag{75}$$

where $H(c)$ is the base-e binary entropy function.

Proof. From the above representation, whenever $\|\pi_\Omega \pi_\Theta\|_{F,F} < 1$,

$$\begin{aligned} \xi_c &= \sup_{\substack{M \in \Psi_c \\ \|M\|_F \le 1}} \left\| \pi_{\Omega^\perp} \pi_\Theta \sum_{k=0}^\infty (\pi_\Theta \pi_\Omega \pi_\Theta)^k \pi_\Theta [M] \right\|_F \\ &\le \left\| \sum_{k=0}^\infty (\pi_\Theta \pi_\Omega \pi_\Theta)^k \right\|_{F,F} \sup_{\substack{M \in \Psi_c \\ \|M\|_F \le 1}} \|\pi_\Theta[M]\|_F \\ &\le \frac{1}{1 - \|\pi_\Theta \pi_\Omega \pi_\Theta\|_{F,F}} \sup_{\substack{M \in \Psi_c \\ \|M\|_F \le 1}} \|\pi_\Theta[M]\|_F . \end{aligned}$$

Now,

$$\begin{aligned} \sup_{\substack{M \in \Psi_c \\ \|M\|_F \le 1}} \|\pi_\Theta[M]\|_F &\le \sup_{\substack{M \in \Psi_c \\ \|M\|_F \le 1}} \|\pi_U M\|_F + \|M \pi_V\|_F \\ &\le \sup_{|I| \le cm} \|U_{I,\bullet}\|_{2,2} + \sup_{|J| \le cm} \|V_{J,\bullet}\|_{2,2}. \end{aligned}$$

Identify U with the orthogonalization $Z(Z^*Z)^{-1/2}$ of an $m \times r$ iid $\mathcal{N}(0, \frac{1}{m})$ matrix Z. Then for any given $I \subset [m]$ of size cm, $\|U_{I,\bullet}\|_{2,2} \le \frac{\|Z_{I,\bullet}\|_{2,2}}{\sigma_{min}(Z)}$. Now,

$$\mathbb{P}\left[\sigma_{min}(Z) > 1 - \sqrt{r/m} - t_1\right] < \exp\left(-\frac{mt_1^2}{2}\right). \tag{76}$$

Meanwhile, for each I of size cm, again

$$\mathbb{P}\left[\|Z_{I,\bullet}\|_{2,2} > \sqrt{r/m} + \sqrt{c} + t_2\right] < \exp\left(-\frac{mt_2^2}{2}\right). \tag{77}$$

There are at most $\exp(mH(c) + O(\log m))$ such subsets I, where $H(\cdot)$ denotes the base-e binary entropy function, so

$$\mathbb{P}\left[\max_{I \in \binom{[m]}{cm}} \|Z_{I,\bullet}\|_{2,2} > \sqrt{r/m} + \sqrt{c} + t_2\right] < \exp\left(-m(t_2^2/2 - H(c)) + O(\log m)\right). \tag{78}$$

Choosing $t_1 = \frac{1}{2} - \frac{1}{2}\sqrt{\frac{r}{m}}$, and set $t_2 = 2\sqrt{H(c)}$ and combining bounds gives

$$\xi_c \le 2 \frac{\sqrt{r/m} + \sqrt{c} + 2\sqrt{H(c)}}{(1 - 4\tau(r/m, \rho_s)^2)(1 - \sqrt{r/m})}. \tag{79}$$

Noticing that $\sqrt{r/m} < \tau$ and applying the bound $\tau < 1/4$ to the terms in the denominator completes the proof. □

2.4 Controlling the initial violations

In this section, we analyze the initial dual vector W_0 given by the minimum Frobenius norm solution to the equality constraints of the optimality condition (11), and show that for any constant β, the probability that $\|\mathcal{S}_{\frac{5}{6\sqrt{m}}}[UV^* + W_0]\|_F > 3\beta$ approaches zero. We will treat the parts of $UV^* + W_0 = UV^* + \mathcal{H}^\dagger[-UV^*] + \mathcal{H}^\dagger[\lambda \operatorname{sign}(E_0)]$ separately, and use the following simple lemma to combine the bounds:

Lemma 2.12. *For all matrices A, B, and $\alpha \in (0,1)$,*

$$\|\mathcal{S}_\gamma^{\Omega^c}[A+B]\|_F \le \|\mathcal{S}_{\alpha\gamma}^{\Omega^c}[A]\|_F + \|\mathcal{S}_{(1-\alpha)\gamma}^{\Omega^c}[B]\|_F. \tag{80}$$

Proof. Notice that for scalars x, $|\mathcal{S}_\gamma[x]|$ is a convex nonnegative function. Hence, for matrices $X \in \mathbb{R}^{m\times m}$, $\|\mathcal{S}_\gamma^{\Omega^c}[X]\|_F$ is again convex (see, e.g, [1] Example 3.14), and so

$$\begin{aligned}
\|\mathcal{S}_\gamma^{\Omega^c}[A+B]\|_F &= \left\|\mathcal{S}_\gamma^{\Omega^c}\left[\alpha\tfrac{A}{\alpha} + (1-\alpha)\tfrac{B}{1-\alpha}\right]\right\|_F \\
&\leq \alpha\left\|\mathcal{S}_\gamma^{\Omega^c}\left[\tfrac{A}{\alpha}\right]\right\|_F + (1-\alpha)\left\|\mathcal{S}_\gamma^{\Omega^c}\left[\tfrac{B}{1-\alpha}\right]\right\|_F \\
&= \|\mathcal{S}_{\alpha\gamma}^{\Omega^c}[A]\|_F + \|\mathcal{S}_{(1-\alpha)\gamma}^{\Omega^c}[B]\|_F.
\end{aligned}$$

□

The strategy, then, is to bound the expectation of $\|S_{\lambda/3}^{\Omega^c}[\cdot]\|_F$ for each of the three terms, using the following lemma. It will turn out that for any prespecified p, we can choose C_0 such if $r < C_0\frac{m}{\log m}$, the expected Frobenius norm of the violations is $O(m^{-p})$; an application of the Markov inequality then bounds the probability that the value deviates above any fixed constant β.

Lemma 2.13. *Let X be a symmetric random variable satisfying the (subgaussian) tail bound*

$$\mathbb{P}\left[X \geq t\right] \leq \exp(-Ct^2).$$

Let $Y \doteq \mathcal{S}_\gamma(X)$. Then

$$\mathbb{E}\left[Y^2\right] \;\leq\; \frac{2}{C}\exp\left(-C\gamma^2\right). \tag{81}$$

Proof. Since X is symmetric, Y is also symmetric, so

$$\begin{aligned}
\mathbb{E}[Y^2] &= 4\int_0^\infty t\,\mathbb{P}[Y\geq t]dt \;\leq\; 4\int_0^\infty t\exp(-C(t+\gamma)^2)dt \\
&\leq 4\int_\gamma^\infty (s-\gamma)\exp(-Cs^2)ds \;\leq\; 4\int_\gamma^\infty s\exp(-Cs^2)ds.
\end{aligned}$$

□

Lemma 2.14. *Let X be a random variable satisfying a tail bound of the form*

$$\mathbb{P}\left[|X| \geq a+t\right] \leq C_1\exp(-C_2t^2).$$

Suppose $\gamma > a$ and let $Y \doteq \mathcal{S}_\gamma(X)$. Then

$$\mathbb{E}\left[Y^2\right] \;\leq\; \frac{C_1}{C_2}\exp\left(-C_2(\gamma-a)^2\right). \tag{82}$$

Proof.

$$\begin{aligned}
\mathbb{E}[Y^2] &= 2\int_0^\infty t\,\mathbb{P}[|Y|\geq t]dt \;=\; 2\int_0^\infty t\mathbb{P}[|X|\geq \gamma+t]dt \\
&= 2\int_\gamma^\infty (s-\gamma)\mathbb{P}[|X|\geq s]ds \;\leq\; 2C_1\int_\gamma^\infty (s-\gamma)\exp\left(-C_2(s-a)^2\right)ds \\
&= 2C_1\int_{\gamma-a}^\infty (q+a-\gamma)\exp(-C_2q^2)dq \;\leq\; \frac{C_1}{C_2}\exp\left(-C_2(\gamma-a)^2\right).
\end{aligned}$$

□

In the previous section, we were interested in controlling the quadratic products $\boldsymbol{x}^*W_0\boldsymbol{y}$ involving the initial dual vector W_0 and arbitrary unit vectors. Here, because the soft thresholding operator S_γ acts elementwise, in this section, we require tighter control over $\boldsymbol{e}_i^*W_0\boldsymbol{e}_j$. Because there are only m^2 pairs $(\boldsymbol{e}_i, \boldsymbol{e}_j)$, much tighter control can be established, as is formalized in the following lemma:

Lemma 2.15. *Consider U, V drawn from the random orthogonal model of rank r. For any $\kappa > 0$, define the events*

$$\mathcal{E}_U(\kappa) \quad : \quad \max_i \|U_{i,\bullet}\|_2 \ \leq \sqrt{\frac{1}{\kappa \log m}},$$
$$\mathcal{E}_V(\kappa) \quad : \quad \max_i \|V_{i,\bullet}\|_2 \ \leq \sqrt{\frac{1}{\kappa \log m}}.$$

Then if $r < \frac{C_0 m}{\log(m)}$ for some $C_0 < \frac{1}{4\kappa}$,

$$\mathbb{P}_{U,V}\left[\mathcal{E}_U(\kappa) \cap \mathcal{E}_V(\kappa)\right] \geq 1 - 2m \exp\left(-\frac{(\kappa^{-1/2} - 2\sqrt{C_0})^2 m}{8 \log m}\right). \tag{83}$$

On these good events,

$$\max_{i,j} \left\|\pi_\Theta[\boldsymbol{e}_i \boldsymbol{e}_j^*]\right\|_F \ \leq \ \frac{2}{\sqrt{\kappa \log m}}. \tag{84}$$

Proof. Notice that $f(U) \doteq \|U_{i,\bullet}\|_2$ is a 1-Lipschitz function of U. Since $\sum_i \|U_{i,\bullet}\|_2^2 = \|U\|_F^2 = r$, by symmetry $\mathbb{E}[\|U_{i,\bullet}\|_2^2] = \frac{r}{m}$. By the Markov inequality, f has a median no larger than $2\mathbb{E}[f] \leq 2\sqrt{\mathbb{E}[f^2]} = 2\sqrt{\frac{r}{m}}$. Invoking Lipschitz concentration on $\mathbb{W}_r^m$ and union bounding over the m choices of i,

$$\mathbb{P}\left[\max_i \|U_{i,\bullet}\|_2 > \sqrt{\frac{1}{\kappa \log m}}\right] \ \leq \ m \ \exp\left(-\frac{m}{8}\left(\sqrt{\frac{1}{\kappa \log m}} - 2\sqrt{\frac{r}{m}}\right)^2\right). \tag{85}$$

An identical calculation applies to $\mathcal{E}_V(\kappa)$. Summing the two probabilities of failure completes the proof. □

Lemma 2.16. *Let (U, V) be distributed according to the random orthogonal model of rank $r < m$. For any fixed $\Omega \subseteq [m] \times [m]$ and $\kappa > 0$, on the good event $\mathcal{E}_U(\kappa)$*

$$\mathbb{E}_{V|U}\left\|\mathcal{S}^{\Omega^c}_{\frac{1}{3\sqrt{m}}}\left[UV^*\right]\right\|_F \ \leq \ \frac{4}{\sqrt{\kappa \log m}} m^{\left(\frac{1}{2} - \frac{\kappa}{144}\right)}. \tag{86}$$

Proof. Fix U and consider $[UV^*]_{i,j} = U_{i,\bullet}V_{j,\bullet}^*$ as a $\|U_{i,\bullet}\|_2$-Lipschitz function of V. On $\mathcal{E}_U(\kappa)$, the Lipschitz constant is bounded by $(\kappa \log m)^{-1/2}$, and

$$\mathbb{P}_{V|U}\left[[UV^*]_{ij} > t\right] < \exp\left(-\frac{\kappa t^2 m \log m}{8}\right). \tag{87}$$

By Lemma 2.13, on $\mathcal{E}_U$,

$$\mathbb{E}_{V|U}\left[\left(\mathcal{S}_{\frac{1}{3\sqrt{m}}}\left[[UV^*]_{ij}\right]\right)^2\right] \leq \frac{16}{\kappa m \log m} \exp\left(-\frac{\kappa}{72}\log(m)\right). \tag{88}$$

Hence, summing over $\leq m^2$ pairs $(i,j) \in \Omega^c$

$$\mathbb{E}_{V|U}\left[\left\|\mathcal{S}^{\Omega^c}_{\frac{1}{3\sqrt{m}}}\left[UV^*\right]\right\|_F^2\right] \leq \frac{16}{\kappa \log m} m^{\left(1 - \frac{\kappa}{72}\right)}. \tag{89}$$

Applying the Cauchy-Schwarz inequality completes the proof. □

Lemma 2.17. *Fix any $\beta > 0$, $\kappa > 0$. Consider (U, V, Ω) drawn from the random orthogonal model of rank r, with Bernoulli error probability ρ_s, and suppose that r, ρ_s satisfy*

$$r \ < \ C_0 \frac{m}{\log m}, \qquad C_0 < \frac{1}{4\kappa}, \qquad \tau\left(\tfrac{r}{m}, \rho_s\right) \ < \ \frac{1}{4}. \tag{90}$$

Then there exists m_0, C_1, $C_2 > 0$ such that for all $m > m_0$,

$$\mathbb{P}_{U,V,\Omega}\left[\left\|\mathcal{S}^{\Omega^c}_{\frac{1}{3\sqrt{m}}}\left[\mathcal{H}^\dagger\left[UV^*\right]\right]\right\|_F > \beta\right] \ \leq \ \frac{C_1 m^{\frac{1}{2} - \frac{\kappa}{2048}}}{\beta\sqrt{\kappa \log m}} \ + \ \mathbb{P}_{U,V,\Omega}\left[\mathcal{E}_U(\kappa)^c \cup \mathcal{E}_V(\kappa)^c \cup \mathcal{E}_{\Omega\Theta}^c\right].$$

Proof. We use a similar splitting trick to that in Lemma 2.7. Again let $\tilde{U}$ and $\tilde{V}$ be uniformly distributed on $\mathbb{W}_{m-1}^{m}$. Identify U and V with their first r columns, and let $\hat{U}$ and $\hat{V}$ denote the remaining $m-r-1$ columns. Write

$$\begin{aligned}\left\|\mathcal{S}^{\Omega^c}_{\frac{1}{3\sqrt{m}}}\left[\mathcal{H}^\dagger[UV^*]\right]\right\|_F &= \left\|\mathcal{S}^{\Omega^c}_{\frac{1}{3\sqrt{m}}}\left[\mathcal{H}^\dagger\left[\tilde{U}\tilde{V}^* - \hat{U}\hat{V}^*\right]\right]\right\|_F \\ &\leq \left\|\mathcal{S}^{\Omega^c}_{\frac{1}{6\sqrt{m}}}\left[\mathcal{H}^\dagger\left[\tilde{U}\tilde{V}^*\right]\right]\right\|_F + \left\|\mathcal{S}^{\Omega^c}_{\frac{1}{6\sqrt{m}}}\left[\mathcal{H}^\dagger\left[\hat{U}\hat{V}^*\right]\right]\right\|_F.\end{aligned}$$

We first address the second term. On the event $\mathcal{E}_{\Omega\Theta}$,

$$\begin{aligned}\left\|\mathcal{S}^{\Omega^c}_{\frac{1}{6\sqrt{m}}}\left[\mathcal{H}^\dagger\left[\hat{U}\hat{V}^*\right]\right]\right\|_F &= \left\|\mathcal{S}^{\Omega^c}_{\frac{1}{6\sqrt{m}}}\left[\pi_{\Theta^\perp}\pi_\Omega\sum_{k=0}^{\infty}(\pi_\Omega\pi_\Theta\pi_\Omega)^k\pi_\Omega\left[\hat{U}\hat{V}^*\right]\right]\right\|_F \\ &= \left\|\mathcal{S}^{\Omega^c}_{\frac{1}{6\sqrt{m}}}\left[\pi_{\Omega^\perp}\pi_{\Theta^\perp}\pi_\Omega\sum_{k=0}^{\infty}(\pi_\Omega\pi_\Theta\pi_\Omega)^k\pi_\Omega\left[\hat{U}\hat{V}^*\right]\right]\right\|_F \\ &= \left\|\mathcal{S}^{\Omega^c}_{\frac{1}{6\sqrt{m}}}\left[\pi_\Theta\pi_\Omega\sum_{k=0}^{\infty}(\pi_\Omega\pi_\Theta\pi_\Omega)^k\pi_\Omega\left[\hat{U}\hat{V}^*\right]\right]\right\|_F.\end{aligned}$$

Here, we have used that $\pi_{\Omega^\perp}\pi_{\Theta^\perp}\pi_\Omega = \pi_{\Omega^\perp}(\mathtt{I}-\pi_\Theta)\pi_\Omega = -\pi_{\Omega^\perp}\pi_\Theta\pi_\Omega$. Now, since for any $R \in SO(m-r-1)$, the joint distribution of $(\tilde{U},\tilde{V},\Omega)$ is invariant under the map

$$\tilde{U} \mapsto \tilde{U}\begin{bmatrix}\mathtt{I} & 0\\ 0 & R\end{bmatrix}. \tag{91}$$

Moreover, this map preserves U and V, and therefore $\mathcal{H}^\dagger$. Hence, if R is distributed according to the invariant measure on $SO(m-r-1)$, $\left\|\mathcal{S}^{\Omega^c}_{\frac{1}{6\sqrt{m}}}\left[\mathcal{H}^\dagger[\hat{U}\hat{V}^*]\right]\right\|_F$ is equal in distribution to $\left\|\mathcal{S}^{\Omega^c}_{\frac{1}{6\sqrt{m}}}\left[\pi_\Theta\pi_\Omega\sum_{k=0}^{\infty}(\pi_\Omega\pi_\Theta\pi_\Omega)^k\pi_\Omega[\hat{U}R\hat{V}^*]\right]\right\|_F$. Consider the i,j element of this matrix,

$$\begin{aligned}\boldsymbol{e}_i^*\pi_\Theta\pi_\Omega\sum_{k=0}^{\infty}(\pi_\Omega\pi_\Theta\pi_\Omega)^k\pi_\Omega[\hat{U}R\hat{V}^*]\boldsymbol{e}_j &= \left\langle \hat{U}^*\left(\sum_{k=0}^{\infty}(\pi_\Omega\pi_\Theta\pi_\Omega)^k\pi_\Omega\pi_\Theta[\boldsymbol{e}_i\boldsymbol{e}_j^*]\right)\hat{V}, R\right\rangle \\ &\doteq \langle M, R\rangle.\end{aligned}$$

This is a $\|M\|_F$-Lipschitz function of R. On $\mathcal{E}_U(\kappa)\cap\mathcal{E}_V(\kappa)\cap\mathcal{E}_{\Omega\Theta}$,

$$\|M\|_F \leq \frac{\|\pi_\Omega\pi_\Theta\|_{F,F}}{1-\|\pi_\Omega\pi_\Theta\pi_\Omega\|_{F,F}}\|\pi_\Theta[\boldsymbol{e}_i\boldsymbol{e}_j^*]\|_F \leq \frac{4\tau}{1-4\tau^2}\frac{1}{\sqrt{\kappa\log(m)}} \leq \frac{4}{3\sqrt{\kappa\log m}}.$$

Let $\delta \doteq \sqrt{\frac{m-r-1}{m}}$. Since $\mathbb{E}_R\left[\langle M,R\rangle\right] = 0$, the tail bound (45) implies that

$$\mathbb{P}_{R|M}\left[\langle M,R\rangle > q(m)+t\right] < 2\exp\left(-\frac{9\,\kappa\delta^2 m\log(m)t^2}{32}\right), \tag{92}$$

where

$$q(m) = \frac{4}{3\delta}\sqrt{\frac{8\pi}{\kappa m\log m}}. \tag{93}$$

By Lemma 2.14, $\mathbb{E}_{R|\tilde{U},\tilde{V},\Omega}\left[\mathcal{S}_{\frac{1}{6\sqrt{m}}}[\langle M,R\rangle]^2\right]$

$$\leq \frac{64}{9\kappa\delta^2 m\log(m)}\exp\left(-\frac{\kappa\delta^2\log(m)}{32}\left(\frac{1}{2}-\frac{4}{\delta}\sqrt{\frac{8\pi}{\kappa\log(m)}}\right)^2\right). \tag{94}$$

For compactness, let

$$\zeta(r/m,\rho_s,m) \doteq 1-\frac{\kappa\delta^2}{32}\left(\frac{1}{2}-\frac{4}{\delta}\sqrt{\frac{8\pi}{\kappa\log(m)}}\right)^2. \tag{95}$$

Then summing over the $\leq m^2$ elements in Ω^c gives that on the event $\mathcal{E}_U(\kappa) \cap \mathcal{E}_V(\kappa) \cap \mathcal{E}_{\Omega\Theta}$,

$$\mathbb{E}_{R|\tilde{U},\tilde{V},\Omega}\left[\left\|\mathcal{S}^{\Omega^c}_{\frac{1}{6\sqrt{m}}}\left[\mathcal{H}^\dagger\left[\hat{U}R\hat{V}^*\right]\right]\right\|_F^2\right] \leq \frac{64}{9\kappa\delta^2\log m} m^{\zeta}. \tag{96}$$

By the Markov inequality, on $\mathcal{E}_U(\kappa) \cap \mathcal{E}_V(\kappa) \cap \mathcal{E}_{\Omega\Theta}$,

$$\mathbb{P}_{R|\tilde{U},\tilde{V},\Omega}\left[\left\|\mathcal{S}^{\Omega^c}_{\frac{1}{6\sqrt{m}}}\left[\mathcal{H}^\dagger\left[\hat{U}R\hat{V}^*\right]\right]\right\|_F > \frac{\beta}{2}\right] < \frac{16 m^{\zeta/2}}{3\beta\delta\sqrt{\kappa\log(m)}}. \tag{97}$$

An identical calculation shows that if we set $\tilde{R} = \begin{bmatrix} \mathtt{I} & 0 \\ 0 & R \end{bmatrix} \in SO(m-1)$, on $\mathcal{E}_U(\kappa) \cap \mathcal{E}_V(\kappa) \cap \mathcal{E}_{\Omega\Theta}$

$$\mathbb{P}_{R|\tilde{U},\tilde{V},\Omega}\left[\left\|\mathcal{S}^{\Omega^c}_{\frac{1}{6\sqrt{m}}}\left[\mathcal{H}^\dagger\left[\tilde{U}\tilde{R}\tilde{V}^*\right]\right]\right\|_F > \frac{\beta}{2}\right] < \frac{16 m^{\zeta/2}}{3\beta\delta\sqrt{\kappa\log(m)}}. \tag{98}$$

So,

$$\mathbb{P}_{U,V,\Omega}\left[\left\|\mathcal{S}^{\Omega^c}_{\frac{1}{3\sqrt{m}}}\left[\mathcal{H}^\dagger\left[UV^*\right]\right]\right\|_F > \beta\right] \;<\; \frac{32 m^{\zeta/2}}{3\beta\delta\sqrt{\kappa\log(m)}} + \mathbb{P}[(\mathcal{E}_U(\kappa)\cap\mathcal{E}_V(\kappa))^c] + \mathbb{P}[\mathcal{E}^c_{\Omega\Theta}].$$

Under the assumptions of the Lemma, $\delta = \sqrt{\frac{m-r-1}{m}} \geq \sqrt{1 - \frac{C_0}{\log m} - m^{-1}}$. Hence, $\exists m_\delta > 0$ such that for all $m > m_\delta$, $\delta(m,r) > \frac{1}{\sqrt{2}}$ and so $\zeta \leq 1 - \frac{\kappa}{64}\left(\frac{1}{2} - 16\sqrt{\frac{\pi}{\kappa\log m}}\right)^2$. Furthermore, for any κ, $\exists m_\kappa$ such that for $m > m_\kappa$, $16\sqrt{\frac{8\pi}{\kappa\log m}} < \frac{1}{4}$. For $m > \max(m_\delta, m_\kappa)$, $\zeta < 1 - \frac{\kappa}{1024}$. For such sufficiently large m, the multiplier $\frac{32}{3\delta} \leq \frac{32\sqrt{2}}{3}$. Choosing this value for C_1 and setting $m_0 = \max(m_\delta, m_\kappa)$ gives the statement of the lemma. $\square$

Lemma 2.18 (Box violations induced by error)**.** *Fix any $\alpha \in (0,1)$. Let (U,V,Ω,Σ) be distributed according to the random orthogonal model of rank r and with error probability ρ_s, with r and ρ_s satisfying*

$$\tau(r/m, \rho_s) \;<\; 1/4. \tag{99}$$

Then on the good event $\mathcal{E}_U(\kappa) \cap \mathcal{E}_V(\kappa) \cap \mathcal{E}_{\Omega\Theta}$,

$$\mathbb{E}_{\Sigma|U,V,\Omega}\left\|\mathcal{S}^{\Omega^c}_{\frac{\alpha}{\sqrt{m}}}\left[\mathcal{H}^\dagger\left[\lambda\,\mathrm{sign}(E_0)\right]\right]\right\|_F \;\leq\; \frac{2}{\sqrt{\kappa\log(m)}} m^{\frac{1}{2}-\frac{\alpha\kappa}{4}}. \tag{100}$$

Proof. We can exploit independence of Ω and Σ by writing $\mathrm{sign}(E_0) = \pi_\Omega[\Sigma]$. From the representation in the previous lemma,

$$\left\|\mathcal{S}^{\Omega^c}_{\frac{\alpha}{\sqrt{m}}}\left[\mathcal{H}^\dagger\left[\lambda\,\mathrm{sign}(E_0)\right]\right]\right\|_F = \left\|S_{\frac{\alpha}{\sqrt{m}}}\left[\sum_{k=1}^{\infty}(\pi_\Theta\pi_\Omega)^k\left[\lambda\,\Sigma\right]\right]\right\|_F.$$

The i,j element of the matrix of interest is

$$\boldsymbol{e}_i^*\sum_{k=1}^{\infty}(\pi_\Theta\pi_\Omega)^k\left[\lambda\,\Sigma\right]\boldsymbol{e}_j \quad=\quad \left\langle \lambda\sum_{k=1}^{\infty}(\pi_\Omega\pi_\Theta)^k\pi_\Theta[\boldsymbol{e}_i\boldsymbol{e}_j^*], \Sigma\right\rangle \;\doteq\; \langle M, \Sigma\rangle.$$

On $\mathcal{E}_U(\kappa) \cap \mathcal{E}_V(\kappa)$, $\|\pi_\Theta[\boldsymbol{e}_i\boldsymbol{e}_j^*]\|_F \;\leq\; \sqrt{\frac{1}{\kappa\log m}}$. On $\mathcal{E}_{\Omega\Theta}$, $\left\|\sum_{k=1}^{\infty}(\pi_\Theta\pi_\Omega)^k\right\|_{F,F} \;\leq\; \frac{2\tau}{1-2\tau} \;\leq\; 1$, and so $\|M\|_F \leq \frac{1}{\sqrt{\kappa m\log m}}$. The same Martingale argument as in Lemma 2.9 shows that

$$\mathbb{P}_\Sigma\left[\langle M,\Sigma\rangle > t\right] < \exp\left(-\frac{t^2}{2\|M\|_F^2}\right) = \exp\left(-\frac{\kappa m\log(m)t^2}{2}\right). \tag{101}$$

Then by Lemma 2.13,

$$\mathbb{E}_{\Sigma|U,V,\Omega}\left[\left(\mathcal{S}_{\frac{\alpha}{\sqrt{m}}}\left[\langle M,\Sigma\rangle\right]\right)^2\right] \leq \frac{2}{\kappa m\log(m)}\exp\left(-\frac{\alpha\kappa\log(m)}{2}\right). \tag{102}$$

Summing over the $\leq m^2$ elements in Ω^c to bound $\mathbb{E}[\|\cdot\|_F^2]$ and then applying the Cauchy-Schwarz inequality establishes the result. $\square$

The three lemmas in this section combine to yield the following bound on the Frobenius norm of the violations.

Corollary 2.19 (Control of box violations)**.** *Fix any $\beta > 0$, $p > 0$. There exist constants $C_0(p) > 0$, $\rho_s^\star > 0$, m_0 with the following property: if $m > m_0$ and (U, V, Ω, Σ) are distributed according to the random orthogonal model of rank*

$$r \; < \; C_0(p) \frac{m}{\log m}, \tag{103}$$

with Bernoulli error probability $\rho_s \le \rho_s^\star$ and random signs, then

$$\mathbb{P}_{U,V,\Omega,\Sigma}\left[\left\| \mathcal{S}^{\Omega^c}_{\frac{5}{6\sqrt{m}}} \left[UV^* + W_0 \right] \right\|_F \; \le \; 3\beta \right] \quad \ge \quad 1 - \frac{C}{\beta} m^{-p}.$$

Proof. By Lemma 2.12,

$$\begin{aligned} & \left\| \mathcal{S}^{\Omega^c}_{\frac{5}{6\sqrt{m}}} [UV^* + W_0] \right\|_F \\ \le \quad & \left\| \mathcal{S}^{\Omega^c}_{\frac{1}{3\sqrt{m}}} [UV^*] \right\|_F + \left\| \mathcal{S}^{\Omega^c}_{\frac{1}{3\sqrt{m}}} [\mathcal{H}^\dagger \left[UV^* \right]] \right\|_F + \left\| \mathcal{S}^{\Omega^c}_{\frac{1}{6\sqrt{m}}} \left[\mathcal{H}^\dagger [\lambda \operatorname{sign}(E_0)] \right] \right\|_F . \end{aligned}$$

We use the three previous lemmas to estimate each of these terms. Set $\kappa = 2048p + 1024$, and $C_0 = \frac{1}{16\kappa}$. From Lemma 2.16 and the Markov inequality, for any Ω, on $\mathcal{E}_U$,

$$\mathbb{P}_{V|U}\left[\left\| \mathcal{S}^{\Omega^c}_{\frac{1}{3\sqrt{m}}} [UV^*] \right\|_F \ge \beta \right] \; \le \; \frac{4}{\beta\sqrt{\kappa \log(m)}} m^{1/2 - \kappa/144},$$

and so

$$\mathbb{P}_{U,V,\Omega,\Sigma}\left[\left\| \mathcal{S}^{\Omega^c}_{\frac{1}{3\sqrt{m}}} [UV^*] \right\|_F \ge \beta \right] \le \frac{4}{\beta\sqrt{\kappa \log(m)}} m^{1/2 - \kappa/144} + \mathbb{P}[\mathcal{E}_U(\kappa)^c] \; = \; o\left(m^{-p}\right).$$

From Lemma 2.18, $\mathbb{P}_{U,V,\Omega,\Sigma}\left[\left\| \mathcal{S}^{\Omega^c}_{\frac{1}{6\sqrt{m}}} \left[\mathcal{H}^\dagger [\lambda \operatorname{sign}(E_0)] \right] \right\|_F \ge \beta \right]$

$$\le \quad \frac{2}{\sqrt{\kappa \log(m)}} m^{1/2 - \kappa/24} + \mathbb{P}\left[(\mathcal{E}_U(\kappa) \cap \mathcal{E}_V(\kappa))^c \right] + \mathbb{P}\left[\mathcal{E}^c_{\Omega\Theta} \right] \; = \; o\left(m^{-p}\right).$$

Finally, by Lemma 2.17 $\mathbb{P}_{U,V,\Omega,\Sigma}\left[\left\| \mathcal{S}^{\Omega^c}_{\frac{1}{3\sqrt{m}}} \left[\mathcal{H}^\dagger [UV^*] \right] \right\|_F \ge \beta \right]$

$$\begin{aligned} \le \quad & \frac{C_1}{\beta\sqrt{\kappa \log(m)}} m^{1/2 - \kappa/2048} + \mathbb{P}\left[(\mathcal{E}_U(\kappa) \cap \mathcal{E}_V(\kappa))^c \right] + \mathbb{P}\left[\mathcal{E}^c_{\Omega\Theta} \right] \\ = \quad & \frac{C_1 m^{-p}}{\beta\sqrt{\kappa \log(m)}} \; + o\left(m^{-p}\right). \end{aligned}$$

Summing the three failure probabilities completes the proof. ∎

2.5 Pulling it all together

We close by fitting the probabilistic lemmas in the previous three sections together to give a proof of our main result, Theorem 1.

Proof. We show that C_0 and m_0 can be selected such that with high probability the conditions of Lemma 1.3 are satisfied. Set $\varepsilon = \frac{1}{6\sqrt{m}}$. Fix m_0 large enough and $c^\star > 0$ small enough that on the good event from Lemma 2.11, as long as $C_0 \le 1$,

$$\xi_{c^\star} \;\; \le \;\; \frac{32}{9} \left(\sqrt{\frac{1}{\log m_0}} + \sqrt{c^\star} + 2\sqrt{H(c^\star)} \right) \;\; \le \;\; \tfrac{1}{2}. \tag{104}$$

We then must show that

$$\|UV^* + W_0\|_{1,2} + 2 \left\| \mathcal{S}^{\Omega^c}_{\frac{5}{6\sqrt{m}}} [UV^* + W_0] \right\|_F \;\; \le \;\; \frac{5}{6}\sqrt{c^\star}. \tag{105}$$

Notice that

$$\|UV^* + W_0\|_{1,2} \;\leq\; \|UV^*\|_{1,2} + \|W_0\|_{1,2} \;\leq\; \max_i \|V_{i,\bullet}\|_2 + \|W_0\|_{2,2}. \tag{106}$$

From Lemmas 2.7 and 2.9, for any choice of C_0, $\|W_0\|_{2,2}$ is with overwhelming probability bounded by a linear function of τ. For any fixed C_0, $\lim_{m\to\infty} \tau(r/m, \rho_s) = 2\sqrt{\rho_s}$. Hence, by choosing m_0 large and ρ_s small, we can bound

$$\|W_0\|_{2,2} \;\leq\; \frac{5}{18}\sqrt{c^\star} \tag{107}$$

with overwhelming probability. Meanwhile, on the overwhelmingly likely good event $\mathcal{E}_V$,

$$\|V_{i,\bullet}\|_2 \;\leq\; \frac{5}{18}\sqrt{c^\star} \tag{108}$$

for m sufficiently large. Finally, fix $\beta = \frac{5}{36}\sqrt{c^\star}$ and choose C_0 small enough and m_0 large enough that the conditions of Corollary 2.19 are satisfied. Then, with probability at least $1 - Cm^{-p}$, (17) is satisfied. The same calculations apply to (18).

Finally, on these good events,

$$\|W_0\|_{2,2} + \frac{1}{1-\xi_c}\left\| \mathcal{S}^{\Omega^c}_{\frac{5}{6\sqrt{m}}}[UV^* + W_0]\right\|_F \leq \frac{5}{18}\sqrt{c^\star} + 2 \times \frac{5}{36}\sqrt{c^\star} < 1, \tag{109}$$

so (19) holds. By Lemma 1.3, then, there exists a certifying dual vector, and the proof is complete. □

3 Implications on Low-Rank Matrix Completion

Our result has strong implications for the low-rank matrix completion problem studied in [3, 4, 2]. In matrix completion, the goal is to recover a low rank matrix A_0 from an observation consisting of a known subset $\Upsilon \doteq [m] \times [m] \setminus \Omega$ of its entries.[1] [3] and [4] studied the following convex programming heuristic for the matrix completion problem

$$A_0 \;=\; \arg\min \|A\|_* \quad \text{subj} \quad A(i,j) = A_0(i,j) \;\forall\; (i,j) \in \Upsilon. \tag{110}$$

Duality considerations in [3] give the following characterization of A_0 that can be recovered by solving (110):

Lemma 3.1 ([3], Lemma 3.1). *Let $A_0 \in \mathbb{R}^{m\times m}$ be a rank-r matrix with reduced singular value decomposition USV^*. As above, let*

$$\Theta \doteq \operatorname{span}\left(\{UM^* \mid M \in \mathbb{R}^{m\times r}\} \cup \{MV^* \mid M \in \mathbb{R}^{m\times r}\}\right) \subset \mathbb{R}^{m\times m}. \tag{111}$$

Suppose that[2]

$$\Theta \cap \{M \in \mathbb{R}^{m\times m} \mid \pi_\Upsilon[M] = 0\} = \{0\} \tag{112}$$

and that there exists a matrix Y such that

$$\{\; \pi_\Theta[Y] \;=\; UV^*, \quad \pi_{\Upsilon^\perp}[Y] \;=\; 0, \quad \|\pi_{\Theta^\perp}[Y]\|_{2,2} < 1 \;\}. \tag{113}$$

Then A_0 is the unique solution to the semidefinite program

$$\min \;\|A\|_* \quad \text{subj} \quad \pi_\Upsilon[A] = \pi_\Upsilon[A_0]. \tag{114}$$

Candes and collaborators then consider the minimum Frobenius solution Y_0 to the system of equations (113) and show that if the number of observations $|\Upsilon|$ is large enough, $\|Y_0\|_{2,2}$ is bounded below one with high probability. Recall that in Section 2, we analyzed the operator norm of the minimum Frobenius norm solution to a similar system of equations. In fact, we will see that Y_0 is exactly equal to $UV^* + W_0^{UV}$, where W_0^{UV} is the part of the initial (RPCA) dual vector that was induced by the singular vectors of A_0. Using Lemma 2.7, the operator norm of W_0^{UV} can be bounded below one with overwhelming probability, even in a proportional growth setting. This yields Theorem 2, which we formally prove below:

[1] In [3], Ω denotes *this* set.

[2] The first condition is equivalent to assuming that the sampling operator π_Υ is *injective* when restricted to the subspace Θ. We thank Yaniv Plan of Caltech for pointing out that this condition was omitted from an early version of this work.

Proof. Notice that (113) is feasible if and only if the system

$$\{ \; \pi_\Theta[W] \; = \; 0, \quad \pi_{\Upsilon^\perp}[W] \; = \; \pi_{\Upsilon^\perp}[-UV^*], \quad \|W\|_{2,2} < 1 \; \} \tag{115}$$

is feasible in $W \doteq Y - UV^*$. Here, $\Upsilon^\perp$ is the set of *missing* elements from the matrix to be completed; we can identify this set with the set of corrupted entries Ω in the low-rank recovery problem that is the main focus of this paper. This is again a random subset of $[m] \times [m]$ in which the inclusion of each pair (i, j) is an independent $\mathrm{Bernoulli}(\rho_s)$ random variable. Hence, if we can show that the matrix W_0^{UV} defined in Lemma 2.7 as the minimum Frobenius norm solution to the equality constraints $\pi_\Theta[W] = 0, \pi_\Omega[W] \; = \; \pi_\Omega[-UV]$ has operator norm bounded below 1, we will have further established that A_0 uniquely solves (110), and hence can be efficiently recovered by nuclear norm minimization. Lemma 2.7 immediately implies Theorem 2 as follows.

First, notice that $\max(\rho_r, \rho_s) < \frac{1}{289} \implies \tau(\rho_r, \rho_s) < \frac{1}{4}$. Under this condition, with probability at least $1 - \exp\left(Cm + O(\log m)\right)$, the minimizer W_0^{UV} is uniquely defined and satisfies

$$\|W_0^{UV}\|_{2,2} \le 64\tau(\rho_r, \rho_s) \left(1 + \tfrac{8}{3} \sqrt{\frac{1}{1 - \rho_r - m^{-1}}} \right) + \tfrac{8}{3} \sqrt{\frac{2\pi}{(1 - \rho_r)m - 1}}. \tag{116}$$

Choose m_0 large enough that the second term is $< \frac{1}{2}$ and $1 + \frac{8}{3}\sqrt{\frac{1}{1-\rho_r-m^{-1}}} \le 4$. Then on the above good event, $\|W_0^{UV}\|_{2,2} < 256\tau(\rho_r, \rho_s) + \frac{1}{2}$. For ρ_r, ρ_s sufficiently small, this is bounded below one: for example, $\max(\rho_r, \rho_s) < \frac{1}{2049^2}$ suffices.[3]

Under this condition, a dual vector Y satisfying (113) exists with high probability. By Lemma 2.2, the injectivity condition (113) is also satisfied with overwhelming probability, and A_0 is indeed the unique optimal solution to the nuclear norm minimization problem (114). □

Thus, matrices A_0 distributed according to the random orthogonal model of rank as large as $r = Cm$ can be recovered from incomplete subsets Υ of size $\le (1 - \rho_s)m^2$. This is the first result to suggest that matrix completion should succeed in such a *proportional growth* scenario. The previous best result for A_0 distributed according to the random orthogonal model, due to [4], showed that

$$|\Upsilon| = Cm\,r \;\; \log^8(m) \tag{117}$$

observations suffice. When $r \ge \frac{m}{C\log^8(m)}$, this result becomes empty, since in this case the prescribed number of measurements exceeds m^2.

The fact that our result holds in proportional growth may seem surprising in light of discussions in [4] – as discussed there, if all one assumes is that the singular spaces of A_0 are *incoherent* with the standard basis, then $|\Upsilon| = O(mr\log m)$ measurements are necessary. It is important to note, though, that the singular spaces of matrices A_0 distributed according to the random orthogonal model satisfy rich regularity properties in addition to incoherence. In particular, the submatrix norm considerations used in bounding $\|\pi_\Theta\pi_\Omega\|_{F,F}$ in our proof of Theorem 1 can be viewed as a kind of restricted isometry property for the singular vectors U and V. Furthermore, our proofs make heavy use of the independence of U and V in the random orthogonal model. Independence and orthogonal invariance allow us to introduce an auxiliary randomization R over the choice of basis for U, decoupling $\mathcal{H}^\dagger$ (which only depends on the subspace spanned by U and not on any choice of basis for the space) from UV^*, which very clearly does depend on the choice of basis. Finally, one should also note that our Theorem 2 only supersedes (117) for relatively large rank r. For smaller (say fixed) rank r, (117) remains the strongest available result for matrix completion in the random orthogonal model.

References

[1] S. Boyd and L. Vandenberghe. *Convex Optimization*. Cambridge University Press, Cambridge, United Kingdom, 2004.

[2] E. Candes and Y. Plan. Matrix completion with noise. *Proceedings of the IEEE, to appear*.

[3] These estimates are clearly pessimistic: larger constant fractions ρ_r are possible. However, the argument given here already suffices to establish that matrix completion succeeds in a new regime of matrices whose rank is a constant fraction of the dimensionality.

[3] E. Candes and B. Recht. Exact matrix completion via convex optimzation. *Foundations of Computational Mathematics, to appear*.

[4] E. Candes and T. Tao. The power of convex relaxation: Near-optimal matrix completion. *IEEE Transactions on Information Theory, submitted for publication*.

[5] Y. Chikuse. *Statistics on Special Manifolds*. Springer-Verlag, New York, New York, 2003.

[6] M. Ledoux. *The Concentration of Measure Phenomenon, Mathematical Surveys and Monographs 89*. American Mathematical Society, Providence, Rhode Island, 2001.

[7] V. Milman and G. Schechtman. *Asymptotic Theory of Finite Dimensional Normed Spaces*. Springer-Verlag, New York, New York, 1986.

[8] R. Vershynin. Spectral norms of products of random and deterministic matrices. preprint, `http://arxiv.org/abs/0812.2432`, 2008.

[9] J. Wright, A. Ganesh, S. Rao, and Y. Ma. Robust principal component analysis: Exact recovery of corrupted low-rank matrices via convex optimization. *Journal of the ACM, submitted for publication*.