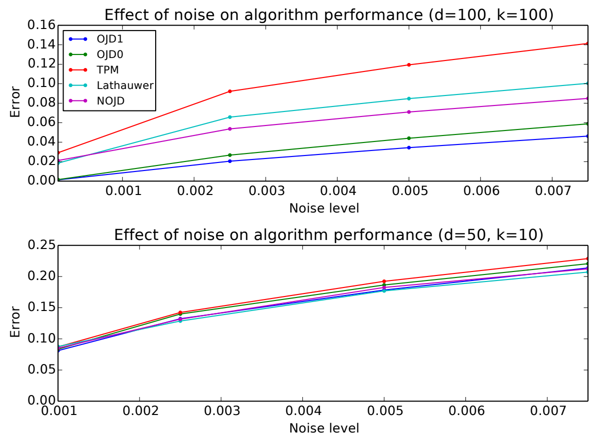

Figure 4: Algorithm performance in the orthogonal setting.

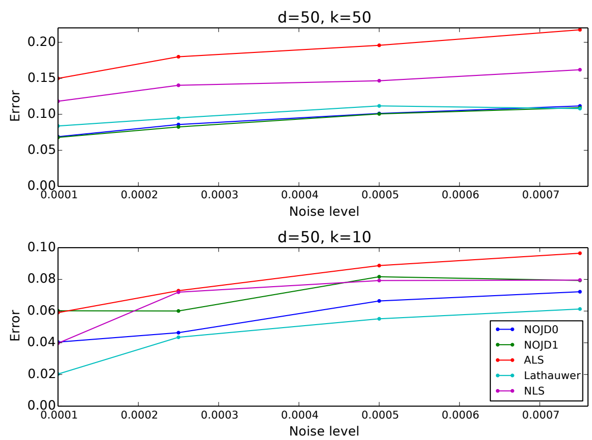

Figure 5: Algorithm performance in the non-orthogonal setting.

A Experiments

A.1 Synthetic experiments

Orthogonal tensors We start by generating random tensors $T = \sum_i \pi u_i^{\otimes 3} + \epsilon R$ with Gaussian entries in π, R and u_i distributed uniformly in the unit sphere $\mathcal{S}^{d-1}$. We let $d = 25, 50, 100$ and in each case consider two regimes: undercomplete tensors with $k = 0.2d$ and full rank tensors, $k = d$. We vary ϵ and report the average error $\|\tilde{u}_i - u_i\|_2$ across all eigenvectors u_i and across 50 trials. In the orthogonal setting, we compare our algorithms (**OJD0** uses random projections, **OJD1** is with plugin) with the tensor power method (**TPM**), alternating least squares (**ALS**), and with the method of de **Lauthauwer** [23]. Alternating least squares displayed very poor performance, and we omit it from our graphs. In the undercomplete case (Figure 4, right), all algorithms fare similarly and errors are within 10% of each other. Our method realizes its full potential in the full-rank setting, where OJD0 and OJD1 are up to three times more accurate than alternative methods ((Figure 4, left).

Non-orthogonal tensors In the non-orthogonal setting, we compare de **Lathauwer**, alternating least squares (**ALS**), non-linear least squares (**NLS**), and our non-orthogonal methods (**NOJD0** and **NOJD1**). We follow the same experimental setup as above and summarize our experiments in Figure 5. In the undercomplete setting, Lathauwer's algorithm has the highest accuracy, about a 10% more than our approach (Figure 5, right). In the full rank setting, there is little difference in performance between our method and Lathauwer's. In both settings, we consistently outperform the standard approaches, ALS and NLS, by 20-50% (Figure 5, left). Although we do

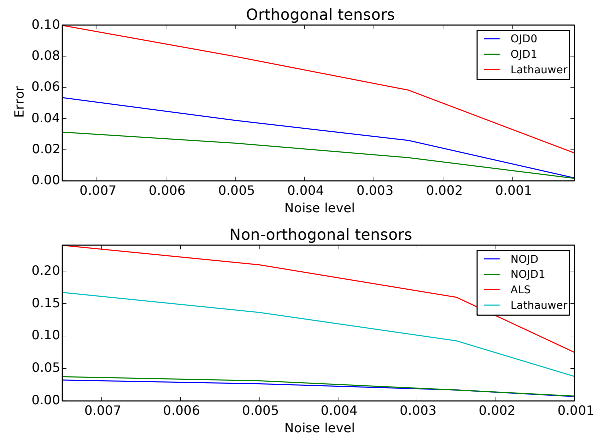

Figure 6: Algorithm performance on asymmetric tensors.

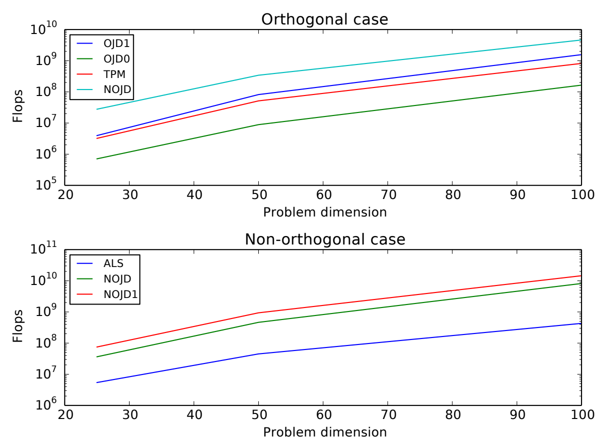

Figure 7: Number of flops performed by various algorithms.

not always outperform Lauthauwer's state-of-the-art method, NOJD0 and NOJD1 are faster and much simpler to implement.

Asymmetric tensors Lastly, we evaluate the extension of our algorithm to tensors of size $50 \times 50 \times 50$ having three distinct sets of asymmetric components (one in each mode). We find that performance is consistent with the symmetric setting, in both orthogonal and non-orthogonal regimes; our method outperforms is competitors by at least 25%, and in the non-orthogonal setting, it achieves an error reduction of up to 70% over Lathauer (Figure 6).

A.2 Algorithm running time

Figure 7 compares the running time in flops of the main algorithms.

We obtain the plots in Figure 7 by calculating flops as follows. The Jacobi method performs at each sweep $2dL(dk - \binom{k}{2})$ flops (where L is the number of matrices); the QRJ1 non-orthogonal diagonalization algorithm performs $4d^3L$ flops per sweep. The tensor power method performs a total of Lkd^3 flops (where L is the number of restarts), times the number of steps it takes to reach convergence for a given eigenvector. The flop count of Lathauwer's method is much higher than that of other method's: at one stage, it requires finding the SVD of a $d^4 \times k^2$ matrix. Consequently, we do not include it in our summary.

B Proofs for orthogonal tensor factorization

In this section we prove perturbation bounds for our algorithm in the setting of orthogonal tensors.

Recall that we observe $\widehat{T} = T + \epsilon R$ where $T = \sum_{i=1}^k \pi_i u_i^{\otimes 3}$ where π_i are factor weights, $u_i \in \mathbb{R}^d$ are orthogonal unit vectors and R is, without loss of generality, symmetric with $\|R\|_{\text{op}} = 1$. Our objective is to estimate π and (u_i). Algorithm 1 does so by simultaneously diagonalizing a number of projections of T; we make use of projections along random vectors and along approximate factors. In this section we will show why both schemes recover π_i and (u_i) with high probability.

Setup Let $\mathcal{M} = \{M_1, \ldots, M_L\}$ be the projections of T along vectors $w_1, \ldots, w_L$, and $\widehat{\mathcal{M}} = \{\widehat{M}_1, \ldots, \widehat{M}_L\}$ be the projections of $\widehat{T}$ along $w_1, \ldots, w_L$. We have that $M_l = \sum_{i=1}^d \pi_i (w_l^\top u_i) u_i \otimes u_i$ and that $\widehat{M}_l = M_l + \epsilon R_l$, where $R_l = R(I, I, w_l)$. Thus, M_l are a set of simultaneously diagonalizable matrices with factors U and factor weights $\lambda_{il} \triangleq \pi_i (w_l^\top u_i)$. From the discussion in Section 2, let $\bar{U}$ be a full-rank extension of U, with columns $u_1, u_2, \ldots u_d$.

Let $\tilde{\pi}$ and $\tilde{u}$ be a factorization of $\widehat{T}$ returned by Algorithm 1. From Lemma 1, we have that

$$\|\tilde{u}_j - u_j\|_2 \leq \epsilon \sqrt{\sum_{i=1}^d E_{ij}^2} + o(\epsilon), \tag{9}$$

for $j \in [k]$ where $E \in \mathbb{R}^{d \times k}$ has entries

$$E_{ij} = \begin{cases} 0 & \text{for } i = j \\ \frac{\sum_{l=1}^L (\lambda_{il} - \lambda_{jl}) u_j^\top R_l u_i}{\sum_{l=1}^L (\lambda_{il} - \lambda_{jl})^2} & \text{for } i \neq j. \end{cases} \tag{10}$$

For notational convenience, let $p_{ij} \triangleq (\pi_i u_i - \pi_j u_j)$ so that $\lambda_{il} - \lambda_{jl} = w_l^\top p_{ij}$. Let $r_{ij} \triangleq R(u_i, u_j, I)$ so that

$$u_j^\top R_l u_i = R(u_j, u_i, w_l) = R(u_i, u_j, I)^\top w_l = r_{ij}^\top w_l.$$

The expression for E_{ij} when $j \neq i$ simplifies to,

$$E_{ij} = \frac{\sum_{l=1}^L w_l^\top p_{ij} r_{ij}^\top w_l}{\sum_{l=1}^L w_l^\top p_{ij} p_{ij}^\top w_l}. \tag{11}$$

In the rest of this section, we will bound E_{ij} for different choices of $\{w_l\}_{l=1}^L$.

B.1 Plugin projections

In Section 4 we proposed using approximate factors $\tilde{u}_i$ as directions to project the tensor $\widehat{T}$ along. In this section, we show that doing so guarantees small errors in u_i.

We begin by bounding the terms E_{ij}.

Lemma 3 (E_{ij} with plug-in projections)**.** *Let $w_1, \ldots, w_k$ be unit-vectors approximations of the unit vectors $u_1, \ldots, u_k$: $\|w_l - u_l\|_2 \leq \gamma$ (so $L = k$), and let $\widehat{\mathcal{M}} = \{\widehat{M}_1, \ldots, \widehat{M}_L\}$ be constructed via projection of $\widehat{T}$ along $w_1, \ldots, w_L$. If the set of matrices $\widehat{\mathcal{M}}$ is simultaneously diagonalized, then to a first-order approximation,*

$$E_{ij} = \frac{p_{ij}^\top r_{ij}}{\|p_{ij}\|^2} + O(\gamma).$$

Proof. We have that

$$\begin{aligned} w_l^\top (p_{ij}) &= (u_l + (w_l - u_l))^\top (\pi_i u_i - \pi_j u_j) \\ &= \pi_i \delta_{il} - \pi_j \delta_{jl} + (w_l - u_l)^\top (\pi_i u_i - \pi_j u_j) \\ &\leq \pi_i \delta_{il} - \pi_j \delta_{jl} + \|w_l - u_l\|_2 \|\pi_i u_i - \pi_j u_j\|_2 \\ &= \pi_i \delta_{il} - \pi_j \delta_{jl} + O(\gamma), \end{aligned}$$

where $\delta_{ij} = 1$ if $i = j$ and 0 otherwise.

Thus,

$$\begin{aligned}
E_{ij} &= \frac{\sum_{l=1}^{L} w_l^\top p_{ij} r_{ij}^\top w_l}{\sum_{l=1}^{L} w_l^\top p_{ij} p_{ij}^\top w_l} \\
&= \frac{\sum_{l=1}^{L} \left(\pi_i \delta_{il} - \pi_j \delta_{jl} + O(\gamma)\right) r_{ij}^\top w_l}{\sum_{l=1}^{L} (\pi_i \delta_{il} - \pi_j \delta_{jl} + O(\gamma))^2} \\
&= \frac{\pi_i r_{ij}^\top w_i - \pi_j r_{ij}^\top w_j + O(\gamma)}{\pi_i^2 + \pi_j^2 + O(\gamma)} \\
&= \frac{\pi_i r_{ij}^\top u_i + \pi_i (w_i - u_i)^\top r_{ij} - \pi_j r_{ij}^\top u_j - \pi_j (w_j - u_j)^\top r_{ij} + O(\gamma)}{\pi_i^2 + \pi_j^2 + O(\gamma)}
\end{aligned}$$

Note that $(w_i - u_i)^\top r_{ij} = O(\gamma)$ and $(w_j - u_j)^\top r_{ij} = O(\gamma)$, and hence both can be included in the $O(\gamma)$ term.

$$E_{ij} = \frac{r_{ij}^\top (\pi_i u_i - \pi_j u_j) + O(\gamma)}{\pi_i^2 + \pi_j^2 + O(\gamma)}.$$

Finally, recall that $p_{ij} \triangleq (\pi_i u_i - \pi_j u_j)$ and that $\|p_{ij}\|^2 = \pi_i^2 + \pi_j^2$. Combining this with the observation that $\frac{1}{1-x} = 1 + x + o(x)$, we obtain

$$E_{ij} = \frac{p_{ij}^\top r_{ij}}{\|p_{ij}\|^2} + O(\gamma).$$

□

Next, we use these term-wise bounds to bound the error in u_i.

Theorem 5 (Tensor factorization with plugin projections)**.** *Let $w_1, \ldots, w_k$ be approximations of $u_1, \ldots, u_k$ such that $\|w_l - u_l\|_2 \le \gamma = O(\epsilon)$, and let $\widehat{\mathcal{M}} = \{\widehat{M}_1, \ldots, \widehat{M}_L\}$ be constructed via projection of $\widehat{T}$ along $w_1, \ldots, w_L$. Then, for $j \in [k]$,*

$$\|\tilde{u}_j - u_j\|_2 \le \left(\frac{2\sqrt{\|\pi\|_1 \pi_{\max}}}{\pi_i^2}\right) \epsilon + o(\epsilon).$$

Proof. From Equation 9, we have that,

$$\|\tilde{u}_j - u_j\|_2 \le \epsilon \sqrt{\sum_{j=1; j \ne i}^{d} E_{ij}^2},$$

for all $j \in [k]$. By Lemma 3, we get,

$$E_{ij} = \frac{p_{ij}^\top r_{ij}}{\|p_{ij}\|^2} + O(\epsilon),$$

and thus,

$$\|\tilde{u}_j - u_j\|_2 \le \epsilon \sqrt{\sum_{i=1; i \ne j}^{d} \left(\frac{p_{ij}^\top r_{ij}}{\|p_{ij}\|^2}\right)^2} + o(\epsilon).$$

Now, we must bound $\sum_{i=1;i\neq j}^{d} (p_{ij}^\top r_{ij})^2$. We expect this the projection to mostly preserve the norm of p_{ij} because r_{ij} are effectively random vectors. Using Lemma 10 with $\mu = 0$, we get that $\sum_{i=1;i\neq j}^{d} (p_{ij}^\top r_{ij})^2 \leq 4\|\pi\|_1 \pi_{\max}$. Finally, $\|p_{ij}\|_2^2 = \pi_i^2 + \pi_j^2 \geq \pi_j^2$.

$$\begin{aligned}\|\tilde{u}_j - u_j\|_2 &\leq \left(\frac{\sqrt{4\|\pi\|_1 \pi_{\max}}}{\pi_j^2}\right)\epsilon + o(\epsilon) \\ &\leq \left(\frac{2\sqrt{\|\pi\|_1 \pi_{\max}}}{\pi_j^2}\right)\epsilon + o(\epsilon).\end{aligned}$$

□

B.2 Random projections

Let us now consider the case when $\{w_l\}_{l=1}^{L}$ are random Gaussian vectors and present similar bounds.

Given Equation 11, we should expect E_{ij} to sharply, and now show that this is indeed the case.

Lemma 4 (Concentration of error E_{ij})**.** *Let $w_1, \ldots, w_L$ be i.i.d. random Gaussian vectors $w_l \sim \mathcal{N}(0, I)$, and let $\widehat{\mathcal{M}} = \{\widehat{M}_1, \ldots, \widehat{M}_L\}$ be constructed via projection of $\widehat{T}$ along $w_1, \ldots, w_L$. If the set of matrices $\widehat{\mathcal{M}}$ is simultaneously diagonalized, then the first-order error E_{ij} is sharply concentrated. If $L \geq 16\log(2\delta)$, then with probability at least $1-\delta$,*

$$E_{ij} \leq \frac{p_{ij}^\top r_{ij}}{\|p_{ij}\|_2^2} + \frac{10\log(2/\delta)}{\sqrt{L}} \frac{\|r_{ij}\|_2}{\|p_{ij}\|_2}.$$

Proof. The numerator and denominator of Equation 11 are both distributed as the sum of χ^2 variables; we show below that they respectively concentrate about $p_{ij}^\top r_{ij}$ and $\|p_{ij}\|_2^2$.

From Lemma 13, we have that the following hold independently with probability at least $1-\delta/2$,

$$\begin{aligned}\frac{1}{L}\sum_{l=1}^{L} w_l^\top p_{ij} r_{ij}^\top w_l &\leq p_{ij}^\top r_{ij} + \|p_{ij}\|\|r_{ij}\| \left(3\sqrt{\frac{\log(2/\delta)}{L}}\right) \\ \frac{1}{L}\sum_{l=1}^{L} w_l^\top p_{ij} p_{ij}^\top w_l &\geq \|p_{ij}\|^2 \left(1 - \frac{2\log(2/\delta)}{\sqrt{L}}\right)\end{aligned}$$

Applying a union bound on both these events, we get that with probability at least $1-\delta$,

$$\begin{aligned}E_{ij} &= \frac{\sum_{l=1}^{L} w_l^\top p_{ij} r_{ij}^\top w_l}{\sum_{m=1}^{L} \|w_m^\top p_{ij}\|_2^2} \\ &\leq \frac{p_{ij}^\top r_{ij} + \|p_{ij}\|_2 \|r_{ij}\|_2 \left(3\sqrt{\frac{\log(2/\delta)}{L}}\right)}{\|p_{ij}\|_2^2 \left(1 - \frac{2\log(2/\delta)}{\sqrt{L}}\right)}.\end{aligned}$$

Note that with the given condition on L, $\frac{2\log(2/\delta)}{\sqrt{L}} < \frac{1}{2}$. Using the property that when $x \leq \frac{1}{2}$, $\frac{1}{1-x} \leq 1 + 2x$, we have that

$$\frac{1}{1 - \frac{2\log(2/\delta)}{\sqrt{L}}} \leq 1 + \frac{4\log(2/\delta)}{\sqrt{L}}.$$

Consequently,

$$\begin{aligned}
E_{ij} &\leq \frac{1}{\|p_{ij}\|_2^2}\left(p_{ij}^\top r_{ij} + \|p_{ij}\|_2\|r_{ij}\|_2\left(3\sqrt{\frac{\log(2/\delta)}{L}}\right)\right)\left(1+\frac{4\log(2/\delta)}{\sqrt{L}}\right) \\
&\leq \frac{p_{ij}^\top r_{ij}}{\|p_{ij}\|_2^2}\left(1+\frac{4\log(2/\delta)}{\sqrt{L}}\right) + 6\frac{\|r_{ij}\|_2}{\|p_{ij}\|_2}\sqrt{\frac{\log(2/\delta)}{L}} \\
&\leq \frac{p_{ij}^\top r_{ij}}{\|p_{ij}\|_2^2} + \frac{10\log(2/\delta)}{\sqrt{L}}\frac{\|r_{ij}\|_2}{\|p_{ij}\|_2}.
\end{aligned}$$

□

With this term-wise bound, we can again proceed to bounding the error u_i.

Theorem 6 (Tensor factorization with random projections)**.** *Let $w_1, \ldots, w_L$ be i.i.d. random Gaussian vectors, $w_l \sim \mathcal{N}(0, I)$, and let $\widehat{\mathcal{M}} = \{\widehat{M}_1, \ldots, \widehat{M}_L\}$ be constructed via projection of $\widehat{T}$ along $w_1, \ldots, w_L$. Furthermore, let $L \geq 16\log(2d(k-1)/\delta)^2$, then, with probability at least $1-\delta$,*

$$\|\tilde{u}_j - u_j\|_2 \leq \left(\frac{2\sqrt{2\|\pi\|_1\pi_{\max}}}{\pi_i^2}\right)\epsilon + \left(20\sqrt{2}\log(2d(k-1)/\delta)\frac{\sqrt{d/L}}{\pi_i}\right)\epsilon + o(\epsilon).$$

for all $j \in [k]$.

Proof. From Equation 9, we have that,

$$\|\tilde{u}_j - u_j\|_2 \leq \epsilon\sqrt{\sum_{i=1;i\neq j}^{d} E_{ij}^2} + o(\epsilon).$$

By Lemma 4, with probability at least $1-\delta/(d(k-1))$,

$$E_{ij} \leq \frac{|p_{ij}^\top r_{ij}|}{\|p_{ij}\|_2^2} + \frac{10\log(2d(k-1)/\delta)}{\sqrt{L}}\frac{\|r_{ij}\|_2}{\|p_{ij}\|_2}.$$

Applying a union bound over $(E_{ij})_{j\neq i}^d$, we have that with probability at least $1-\delta$,

$$\|\tilde{u}_j - u_j\|_2 \leq \epsilon\sqrt{\sum_{i=1;i\neq j}^{d} 2\left(\frac{p_{ij}^\top r_{ij}}{\|p_{ij}\|_2^2}\right)^2} + \epsilon\frac{10\log(2d(k-1)/\delta)}{\sqrt{L}}\sqrt{\sum_{i=1;i\neq j}^{d} 2\left(\frac{\|r_{ij}\|_2}{\|p_{ij}\|_2}\right)^2} + o(\epsilon),$$

for all $j \in [k]$. We have used the fact that for $a, b \geq 0$, $(a+b)^2 = a^2 + 2ab + b^2 \leq a^2 + (a^2+b^2) + b^2 = 2a^2 + 2b^2$ and $\sqrt{a+b} \leq \sqrt{a} + \sqrt{b}$.

Note that $\|p_{ij}\|_2 = \sqrt{\pi_i^2 + \pi_j^2} \geq |\pi_i|$. In Lemma 10, we show that $\sum_{i=1;i\neq j}^{d}(p_{ij}^\top r_{ij})^2 \leq 4\|\pi\|_1\pi_{\max}$. Furthermore, $\|r_{ij}\| \leq 1$ by the operator norm bound on R. Thus, we get,

$$\|\tilde{u}_j - u_j\|_2 \leq \left(\frac{2\sqrt{2\|\pi\|_1\pi_{\max}}}{\pi_i^2}\right)\epsilon + \left(20\sqrt{2}\log(2d(k-1)/\delta)\frac{\sqrt{d/L}}{\pi_i}\right)\epsilon + o(\epsilon).$$

□

C Proofs for non-orthogonal tensor factorization

In this section we extend our previous analysis to non-orthogonal tensor decomposition.

Setup As before, let $\mathcal{M} = \{M_1, \ldots, M_L\}$ be the projections of T along vectors $w_1, \ldots, w_L$, and $\widehat{\mathcal{M}} = \{\widehat{M}_1, \ldots, \widehat{M}_L\}$ be the projections of $\widehat{T}$ along $w_1, \ldots, w_L$. We have that $M_l = \sum_{i=1}^d \pi_i (w_l^\top u_i) u_i \otimes u_i$ and that $\widehat{M}_l = M_l + \epsilon R_l$, where $R_l = R(I, I, w_l)$. Thus, M_l are a set of simultaneously diagonalizable matrices with factors U and factor weights $\lambda_{il} \triangleq \pi_i(w_l^\top u_i)$. Let $\bar{U}$ be the full-rank extension of U with unit-norm columns $u_1, u_2, \ldots, u_d$. In this setting, however, the factor U is not orthogonal. Let $\bar{V} = \bar{U}^{-1}$, with rows $v_1, v_2, \ldots, v_d$. Note that we place our incoherence assumption on the columns of U and present results in terms of the 2-norm of $V^\top$. When U is incoherent, it can be shown that $\|V^\top\|_2 \leq 1 + O(\mu)$. Finally, note that in the orthogonal case, when $\mu = 0$, the rows (v_i) and columns (u_i) are identical, and no distinction between the two need be made.

Let $\tilde{\pi}$ and $\tilde{u}$ be a factorization of $\widehat{T}$ returned by Algorithm 1. From Lemma 2, we have that

$$\|\tilde{u}_j - u_j\|_2 = \epsilon \sqrt{\sum_{i=1}^d E_{ij}^2},$$

where the entries of $E \in \mathbb{R}^{d \times k}$ are bounded by Lemma 16:

$$|E_{ij}| \leq \frac{1}{1 - \rho_{ij}^2} \left(\frac{1}{\|\lambda_i\|_2^2} + \frac{1}{\|\lambda_j\|_2^2} \right) \left(\left| \sum_{l=1}^L v_i^\top R_l v_j \lambda_{jl} \right| + \left| \sum_{l=1}^L v_i^\top R_l v_j \lambda_{il} \right| \right), \tag{12}$$

where $\lambda_i \in \mathbb{R}^L$ is the vector of i-th factor values of M_l, i.e. λ_{il} is the i-th factor value of matrix M_l (i.e. $\lambda_{il} = (\Lambda_l)_{ii}$) and $\rho_{ij} = \frac{\lambda_i^\top \lambda_j}{\|\lambda_i\|_2 \|\lambda_j\|_2}$, the modulus of uniqueness, is a measure of the singularity of the problem.

When λ_{il} is generated by projections, $\lambda_{il} = \pi_i w_l^\top u_i$. Let $r_{ij} \triangleq R(v_i, v_j, I)$ so that

$$v_i^\top R_l v_j = R(v_i, v_j, w_l) = R(v_i, v_j, I)^\top w_l = r_{ij}^\top w_l.$$

Note that $\|r_{ij}\|_2 \leq \|v_i\|_2 \|v_j\|_2 \leq \|V^\top\|_2^2$.

Equation 12 then simplifies to,

$$|E_{ij}| \leq \frac{1}{1 - \rho_{ij}^2} \left(\frac{1}{\|\lambda_i\|_2^2} + \frac{1}{\|\lambda_j\|_2^2} \right) \left(|\pi_j| \left| \sum_{l=1}^L w_l^\top u_j r_{ij}^\top w_l \right| + |\pi_i| \left| \sum_{l=1}^L w_l^\top u_i r_{ij}^\top w_l \right| \right), \tag{13}$$

where $\|\lambda_i\|_2^2 = \pi_i^2 \sum_{l=1}^L w_l^\top u_i u_i^\top w_l$, and ρ_{ij} has the following expression,

$$\rho_{ij} = \frac{\lambda_i^\top \lambda_j}{\|\lambda_i\|_2 \|\lambda_j\|_2} = \frac{\sum_{l=1}^L w_l^\top u_i u_j^\top w_l}{\sqrt{(\sum_{l=1}^L w_l^\top u_i u_i^\top w_l)(\sum_{l=1}^L w_l^\top u_j u_j^\top w_l)}}. \tag{14}$$

Observe that the terms u_i interact with the factor weights λ_{il}, while the terms v_i interact only with the noise terms R_l.

In the rest of this section, we will bound E_{ij} and ρ_{ij} with different choices of $\{w_l\}_{l=1}^L$.

C.1 Plugin projections

We now assume we have plugin estimates (w_l) that are close to the inverse factors (v_l): $\|w_l - v_l\|_2 \leq O(\gamma)$ for $l \in [k]$. Then,

$$\begin{aligned} w_l^\top u_i &= (v_l + (w_l - v_l))^\top u_i \\ &= v_l^\top u_i + ||w_l - v_l||_2 \cdot \frac{(w_l - v_l)^\top u_i}{||w_l - v_l||_2} \\ &= v_l^\top u_i + O(\gamma). \end{aligned}$$

Recall that $V = U^{-1}$, so $v_l^\top u_i = \delta_{il}$.

It will be useful to keep track of $\|\lambda_i\|_2^2$,

$$\begin{aligned}\|\lambda_i\|_2^2 &= \sum_{l=1}^{L} \pi_i^2 (w_l^\top u_i)^2 \\ &= \pi_i^2 \sum_{l=1}^{k} (v_l^\top u_i + O(\gamma))^2 \\ &= \pi_i^2 + O(\gamma). \end{aligned} \tag{15}$$

Lemma 5 (Modulus of uniqueness for plugin projections)**.** *Let $w_1, \dots, w_k$ be approximations of $v_1, \dots, v_k$: $\|w_l - v_l\|_2 \leq O(\gamma)$ for $l \in [k]$, and let $\widehat{\mathcal{M}} = \{\widehat{M}_1, \dots, \widehat{M}_L\}$ be constructed via projection of $\widehat{T}$ along $w_1, \dots, w_L$. Then, for $i \neq j$,*

$$\rho_{ij}^2 \leq O(\gamma),$$

Proof. Let us first bound the numerator of Equation 14.

$$\begin{aligned}(\lambda_i^\top \lambda_j)^2 &= \pi_i^2 \pi_j^2 \left(\sum_{l=1}^{L} w_l^\top u_i u_j^\top w_l \right)^2 \\ &= \pi_i^2 \pi_j^2 \left(\sum_{l=1}^{L} v_l^\top u_i u_j^\top v_l + O(\gamma) \right)^2 \\ &= \pi_i^2 \pi_j^2 \delta_{ij} + O(\gamma) \\ &= O(\gamma).\end{aligned}$$

Using Equation 15, we get that

$$\begin{aligned}\rho_{ij}^2 &= \frac{O(\gamma)}{(1+O(\gamma))(1+O(\gamma))} \\ &= O(\gamma).\end{aligned}$$

where in the last line we used the fact that $\frac{1}{1-x} = 1 + x + o(x)$. □

Lemma 6 (Bound on E_{ij} for non-orthogonal plugin projections)**.** *Let $w_1, \dots, w_k$ be approximations of $v_1, \dots, v_k$: $\|w_l - v_l\|_2 \leq O(\gamma)$ for $l \in [k]$, and let $\widehat{\mathcal{M}} = \{\widehat{M}_1, \dots, \widehat{M}_L\}$ be constructed via projection of $\widehat{T}$ along $w_1, \dots, w_L$.*

$$|E_{ij}| \leq \left(\frac{1}{\pi_i^2} + \frac{1}{\pi_j^2} \right) \|V^\top\|_2 \; p_{ij}^\top r_{ij} + O(\gamma),$$

where $p_{ij} \triangleq |\pi_i| \frac{v_i}{\|v_i\|_2} + |\pi_j| \frac{v_j}{\|v_j\|_2}$.

Proof. Let us bound each term within our expression for E_{ij} (Equation (13)).

$$\begin{aligned}\sum_{l=1}^{k} w_l^\top u_j r_{ij}^\top w_l &= \sum_{l=1}^{k} v_l^\top u_j r_{ij}^\top v_l + O(\gamma) \\ &\leq r_{ij}^\top v_j + O(\gamma).\end{aligned}$$

Similarly,

$$\sum_{l=1}^{k} w_l^\top u_i r_{ij}^\top w_l \leq r_{ij}^\top v_i + O(\gamma),$$

From Equation (15), we have

$$\|\lambda_i\|_2^2 = |\pi_i|^2 + O(\gamma)$$
$$\|\lambda_j\|_2^2 = |\pi_j|^2 + O(\gamma).$$

From Lemma 5 we have that

$$\begin{aligned} \rho_{ij}^2 &\le O(\gamma) \\ \frac{1}{1-\rho_{ij}^2} &\le \frac{1}{1-O(\gamma)} + O(\gamma) \\ &\le 1 + O(\gamma). \end{aligned}$$

Finally,

$$\begin{aligned} |E_{ij}| &\le \left(\frac{1}{\pi_i^2} + \frac{1}{\pi_j^2}\right)\left((|\pi_i|v_i + |\pi_j|v_j)^\top r_{ij}\right) + O(\gamma) \\ &\le \left(\frac{1}{\pi_i^2} + \frac{1}{\pi_j^2}\right)\|V^\top\|_2 \; p_{ij}^\top r_{ij} + O(\gamma). \end{aligned}$$

□

Note that the error terms depend not on u_i but rather v_i. This is because the projections (w_l) are chosen to be close to the v_i. Now, let us bound the error in u_i.

Theorem 7 (Non-orthogonal tensor factorization with plug-in projections)**.** *Let $w_1, \ldots, w_k$ be approximations of $v_1, \ldots, v_k$: $\|w_l - v_l\|_2 \le O(\epsilon)$ for $l \in [k]$ and let $\widehat{\mathcal{M}} = \{\widehat{M}_1, \ldots, \widehat{M}_L\}$ be constructed via projection of $\widehat{T}$ along $w_1, \ldots, w_L$. Then, for all $j \in [k]$,*

$$\|\tilde{u}_j - u_j\|_2 \le 8\epsilon \; \frac{\sqrt{\|\pi\|_1 \pi_{\max}}}{\pi_{\min}^2} \|V^\top\|_2^3 + o(\epsilon).$$

Proof. From Lemma 15 we have that

$$\|\tilde{u}_j - u_j\|_2 \le \epsilon \sqrt{\sum_{i=1}^d E_{ij}^2} + o(\epsilon),$$

for $j \in [k]$, where E_{ij} is bounded in Lemma 6 as follows:

$$\begin{aligned} |E_{ij}| &\le \left(\frac{1}{\pi_i^2} + \frac{1}{\pi_j^2}\right)\|V^\top\|_2 \; p_{ij}^\top r_{ij} + O(\epsilon) \\ &\le \frac{2}{\pi_{\min}^2}\|V^\top\|_2 \; p_{ij}^\top r_{ij} + O(\epsilon). \end{aligned}$$

Consequently,

$$\begin{aligned} \|\tilde{u}_j - u_j\|_2 &\le \epsilon \sqrt{\sum_{i \ne j}^d E_{ij}^2} \\ &\le \frac{2\epsilon}{\pi_{\min}^2} \sqrt{\sum_{i \ne j}^d \left(\|V^\top\|_2 \; p_{ij}^\top r_{ij} + O(\epsilon)\right)^2} + o(\epsilon) \\ &\le \frac{4\epsilon}{\pi_{\min}^2} \left(\sqrt{\sum_{i \ne j}^d \left(\|V^\top\|_2 \; p_{ij}^\top r_{ij}\right)^2 +} \right) + o(\epsilon), \end{aligned}$$

where we have used the fact that $(a+b)^2 \leq 2(a^2+b^2)$ and that $\sqrt{a+b} \leq \sqrt{a}+\sqrt{b}$.

From Lemma 10 we have, $p_{ij}^\top r_{ij} \leq 4\|\pi\|_1 \pi_{\max} \|V^\top\|_2^4$,

$$\begin{aligned}
\|\tilde{u}_j - u_j\|_2 &\leq \frac{4\epsilon}{\pi_{\min}^2}\left(\sqrt{4\|\pi\|_1\pi_{\max}\|V^\top\|_2^6}\right) + o(\epsilon) \\
&\leq 8\epsilon\, \frac{\sqrt{\|\pi\|_1\pi_{\max}}}{\pi_{\min}^2}\|V^\top\|_2^3 + o(\epsilon).
\end{aligned}$$

□

C.2 Random projections

We now study the case where the random projections, (w_l), are drawn from a standard Gaussian distribution. First let us show that the modulus of uniqueness ρ_{ij} sharply concentrates around $u_i^\top u_j$.

Lemma 7 (Modulus of Uniqueness with random projections)**.** *Let $w_1, \cdots w_L \in \mathbb{R}^d$ be entries drawn i.i.d. from the standard Normal distribution. Let $L > 16\log(3/\delta)^2$ Then, with probability at least $1-\delta$,*

$$\rho_{ij} \leq u_i^\top u_j + \frac{10\log(3/\delta)}{\sqrt{L}}.$$

Proof. Observe from Equation 14 that the numerator and the denominator of ρ_{ij} are essentially distributed as a χ^2 distribution (Lemma 13). Thus, with probability at least $1-\delta/3$ each, the following hold,

$$\begin{aligned}
\frac{1}{L}\sum_{l=1}^{L} w_l^\top u_i u_j^\top w_l &\leq u_i^\top u_j + \|u_i\|_2\|u_j\|_2\left(3\sqrt{\frac{\log(3/\delta)}{L}}\right) \\
\frac{1}{L}\sum_{l=1}^{L}(w_l^\top u_i)^2 &\geq \|u_i\|_2\left(1-\frac{2\log(3/\delta)}{\sqrt{L}}\right) \\
\frac{1}{L}\sum_{l=1}^{L}(w_l^\top u_j)^2 &\geq \|u_j\|_2\left(1-\frac{2\log(3/\delta)}{\sqrt{L}}\right).
\end{aligned}$$

Noting that $\|u_i\|_2 = \|u_j\|_2 = 1$ and applying a union bound on the above three events, we get that with probability at least $1-\delta$,

$$\rho_{ij} \leq \frac{u_i^\top u_j + 3\sqrt{\frac{\log(3/\delta)}{L}}}{1-\frac{2\log(3/\delta)}{\sqrt{L}}}.$$

Under the conditions on L, $\frac{2\log(3/\delta)}{\sqrt{L}} \leq \frac{1}{2}$. Applying the property that when $x < \frac{1}{2}$, $\frac{1}{1-x} \leq 1+2x$,

$$\frac{1}{1-\frac{2\log(3/\delta)}{\sqrt{L}}} \leq 1 + \frac{4\log(3/\delta)}{\sqrt{L}} < 2.$$

Finally,

$$\begin{aligned}
\rho_{ij} &\leq \left(u_i^\top u_j + 3\sqrt{\frac{\log(3/\delta)}{L}}\right)\left(1+\frac{4\log(3/\delta)}{\sqrt{L}}\right) \\
&\leq u_i^\top u_j\left(1+\frac{4\log(3/\delta)}{\sqrt{L}}\right) + 3\sqrt{\frac{\log(3/\delta)}{L}} \times 2 \\
&\leq u_i^\top u_j + \frac{10\log(3/\delta)}{\sqrt{L}}.
\end{aligned}$$

□

Let's now bound the inverse modulus of uniqueness.

Lemma 8 (Bounding inverse modulus of uniqueness)**.** *Let $w_1, \cdots w_L \in \mathbb{R}^d$ be entries drawn i.i.d. from the standard Normal distribution. Assume incoherence μ for that the (u_i): $u_i^\top u_j \le \mu$ for $i \neq j$. Let $L_0 \triangleq \left(\frac{50}{(1-\mu^2)}\right)^2$ Let $L \ge L_0 \log(3/\delta)^2$. Then, with probability at least $1-\delta$,*

$$\frac{1}{1-\rho_{ij}^2} \le \frac{1}{1-(u_i^\top u_j)^2}\left(1 + \sqrt{\frac{L_0}{L}}\log(3/\delta)\right).$$

Proof. From Lemma 7, we have that with probability at least $1-\delta$,

$$\rho_{ij} \le u_i^\top u_j + \frac{10\log(3/\delta)}{\sqrt{L}}.$$

Then,

$$\rho_{ij}^2 \le (u_i^\top u_j)^2 + 2u_i^\top u_j\left(\frac{10\log(3/\delta)}{\sqrt{L}}\right) + \left(\frac{10\log(3/\delta)}{\sqrt{L}}\right)^2.$$

Given the assumptions on L, we have that $L \ge L_0 \log(3/\delta)^2 \ge 50\log(3/\delta)^2$ and thus $\frac{10\log(3/\delta)}{\sqrt{L}} \le \frac{1}{2}$:

$$\begin{aligned}
\rho_{ij}^2 &\le (u_i^\top u_j)^2 + 2\left(\frac{10\log(3/\delta)}{\sqrt{L}}\right) + \frac{1}{2}\frac{10\log(3/\delta)}{\sqrt{L}} \\
&= (u_i^\top u_j)^2 + \frac{25\log(3/\delta)}{\sqrt{L}}.
\end{aligned}$$

Now, we bound $\frac{1}{1-\rho_{ij}^2}$,

$$\begin{aligned}
\frac{1}{1-\rho_{ij}^2} &\le \frac{1}{1-(u_i^\top u_j)^2 - \frac{25\log(3/\delta)}{\sqrt{L}}} \\
&\le \frac{1}{1-(u_i^\top u_j)^2}\frac{1}{1-\frac{25\log(3/\delta)}{(1-(u_i^\top u_j)^2)\sqrt{L}}} \\
&\le \frac{1}{1-(u_i^\top u_j)^2}\frac{1}{1-\frac{25\log(3/\delta)}{(1-\mu^2)\sqrt{L}}} \\
&\le \frac{1}{1-(u_i^\top u_j)^2}\frac{1}{1-\frac{1}{2}\log(3/\delta)\sqrt{\frac{L_0}{L}}}.
\end{aligned}$$

Again, given assumptions on L, $\frac{1}{2}\log(3/\delta)\sqrt{\frac{L_0}{L}} \le \frac{1}{2}$. Using the identity that if $x < \frac{1}{2}$, $\frac{1}{1-x} \le 1 + 2x$,

$$\frac{1}{1-\rho_{ij}^2} \le \frac{1}{1-(u_i^\top u_j)^2}\left(1 + \log(3/\delta)\sqrt{\frac{L_0}{L}}\right).$$

□

We are now ready to bound the termwise entries of E.

Lemma 9 (Concentration of E_{ij})**.** *Let $w_1, \ldots, w_L$ be i.i.d. random Gaussian vectors $w_l \sim \mathcal{N}(0, I)$, and let $\widehat{\mathcal{M}} = \{\widehat{M}_1, \ldots, \widehat{M}_L\}$ be constructed via projection of $\widehat{T}$ along $w_1, \ldots, w_L$. Assume incoherence μ for that the (u_i): $u_i^\top u_j \le \mu$ for $i \neq j$. Furthermore, let $L \ge L_0 \log(15/\delta)^2$. Then, with probability at least $1-\delta$,*

$$|E_{ij}| \le \left(\frac{1}{\pi_i^2} + \frac{1}{\pi_j^2}\right)\left(\frac{\bar{p}_{ij}^\top r_{ij}}{1-(u_i^\top u_j)^2} + \frac{\bar{\pi}_{ij}\|r_{ij}\|_2}{1-(u_i^\top u_j)^2}\frac{(20+\sqrt{L_0})\log(15/\delta)}{\sqrt{L}}\right),$$

where $\bar{p}_{ij} \triangleq |\pi_i|u_i + |\pi_j|u_j$ and $\bar{\pi}_{ij} \triangleq |\pi_i| + |\pi_j|$.

Proof. Each term in Equation 13 concentrates sharply about its mean value. We bound each in turn.

First, consider $\|\lambda_i\|_2^2/L = \frac{1}{L}|\pi_i|^2 \sum_{l=1}^{L}(w_l^\top u_i)^2$. With probability at least $1-\delta/5$ each, the following hold,

$$\frac{1}{L}\|\lambda_i\|_2^2 \geq \pi_i^2 \|u_i\|_2^2 \left(1 - \frac{2\log(5/\delta)}{\sqrt{L}}\right)$$
$$\frac{1}{L}\|\lambda_j\|_2^2 \geq \pi_j^2 \|u_j\|_2^2 \left(1 - \frac{2\log(5/\delta)}{\sqrt{L}}\right).$$

Thus, using the fact that $\|u_i\|_2^2 = 1$,

$$L\left(\frac{1}{\|\lambda_i\|_2^2} + \frac{1}{\|\lambda_j\|_2^2}\right) \leq \frac{\frac{1}{\pi_i^2} + \frac{1}{\pi_j^2}}{1 - \frac{2\log(5/\delta)}{\sqrt{L}}}.$$

Given our assumption on L, it follows that $\frac{2\log(5/\delta)}{\sqrt{L}} \leq \frac{1}{2}$. Thus we can use the fact that $\frac{1}{1-x} \leq 1+2x$ when $x \leq \frac{1}{2}$ to obtain the following bound:

$$L\left(\frac{1}{\|\lambda_i\|_2^2} + \frac{1}{\|\lambda_j\|_2^2}\right) \leq \left(\frac{1}{\pi_i^2} + \frac{1}{\pi_j^2}\right)\left(1 + \frac{4\log(5/\delta)}{\sqrt{L}}\right).$$

Next, we bound $\frac{1}{L}\sum_{l=1}^{L} w_l^\top u_i r_{ij}^\top w_l$ and $\frac{1}{L}\sum_{l=1}^{L} w_l^\top u_j r_{ij}^\top w_l$. From Lemma 13, we have with probability at least $1-\delta/5$ each,

$$\frac{1}{L}\sum_{l=1}^{L} w_l^\top u_j r_{ij}^\top w_l \leq r_{ij}^\top u_j + \|r_{ij}\|_2 \|u_j\|_2 \left(3\sqrt{\frac{\log(5/\delta)}{L}}\right)$$
$$\frac{1}{L}\sum_{l=1}^{L} w_l^\top u_i r_{ij}^\top w_l \leq r_{ij}^\top u_i + \|r_{ij}\|_2 \|u_i\|_2 \left(3\sqrt{\frac{\log(5/\delta)}{L}}\right).$$

Note that by definition, $\|u_i\|_2 = 1$.

Using Lemma 8, we have that with probability at least $1-\delta/5$,

$$\frac{1}{1-\rho_{ij}^2} \leq \frac{1}{1-(u_i^\top u_j)^2}\left(1 + \sqrt{\frac{L_0}{L}}\log(15/\delta)\right).$$

Putting it all together, we get that with probability at least $1-\delta$,

$$|E_{ij}| \leq \frac{1}{1-(u_i^\top u_j)^2}\left(1 + \sqrt{\frac{L_0}{L}}\log(15/\delta)\right)\left(\frac{1}{\pi_i^2} + \frac{1}{\pi_j^2}\right)\left(1 + \frac{4\log(5/\delta)}{\sqrt{L}}\right)$$
$$\left(|\pi_i| r_{ij}^\top u_i + |\pi_j| r_{ij}^\top u_j + (|\pi_i| + |\pi_j|)\|r_{ij}\|_2 \left(3\sqrt{\frac{\log(5/\delta)}{L}}\right)\right).$$

Let us define $\bar{p}_{ij} \triangleq |\pi_i| u_i + |\pi_j| u_j$ and $\bar{\pi}_{ij} \triangleq |\pi_i| + |\pi_j|$:

$$|E_{ij}| \leq \frac{1}{1-(u_i^\top u_j)^2}\left(1 + \sqrt{\frac{L_0}{L}}\log(15/\delta)\right)\left(\frac{1}{\pi_i^2} + \frac{1}{\pi_j^2}\right)\left(1 + \frac{4\log(5/\delta)}{\sqrt{L}}\right)$$
$$\left(\bar{p}_{ij}^\top r_{ij} + \bar{\pi}_{ij}\|r_{ij}\|_2 \left(3\sqrt{\frac{\log(5/\delta)}{L}}\right)\right).$$

Given that $L \geq L_0 \log(15/\delta)^2$, we have that $\sqrt{\frac{L_0}{L}}\log(15/\delta) \leq 1$ and $\frac{4\log(5/\delta)}{\sqrt{L}} \leq 1$, thus

$$\left(1 + \sqrt{\frac{L_0}{L}}\log(15/\delta)\right)\left(1 + \frac{4\log(5/\delta)}{\sqrt{L}}\right) \leq 2 \times 2$$
$$\leq 4.$$

Finally, note that $|\pi_i|r_{ij}^\top u_i + |\pi_j|r_{ij}^\top u_j \le (|\pi_i| + |\pi_j|)\|r_{ij}\|_2$, giving us,

$$\begin{aligned}
|E_{ij}| &\le \left(\frac{1}{\pi_i^2} + \frac{1}{\pi_j^2}\right) \frac{(\bar{p}_{ij}^\top r_{ij})}{1 - (u_i^\top u_j)^2} \\
&\quad + \left(\frac{1}{\pi_i^2} + \frac{1}{\pi_j^2}\right) \frac{\bar{\pi}_{ij}\|r_{ij}\|_2}{1 - (u_i^\top u_j)^2} \left(\sqrt{\frac{L_0}{L}} \log(15/\delta) + 2\frac{4\log(5/\delta)}{\sqrt{L}} + 4\left(3\sqrt{\frac{\log(5/\delta)}{L}}\right)\right) \\
&\le \left(\frac{1}{\pi_i^2} + \frac{1}{\pi_j^2}\right) \left(\frac{\bar{p}_{ij}^\top r_{ij}}{1 - (u_i^\top u_j)^2} + \frac{\bar{\pi}_{ij}\|r_{ij}\|_2}{1 - (u_i^\top u_j)^2} \frac{(20 + \sqrt{L_0})\log(15/\delta)}{\sqrt{L}}\right).
\end{aligned}$$

□

Finally, we bound the error in estimating u_j.

Theorem 8 (Non-orthogonal tensor factorization with random projections)**.** *Let $w_1, \dots, w_L$ be i.i.d. random Gaussian vectors, $w_l \sim \mathcal{N}(0, I)$, and let $\widehat{\mathcal{M}} = \{\widehat{M}_1, \dots, \widehat{M}_L\}$ be constructed via projection of $\widehat{T}$ along $w_1, \dots, w_L$. Assume incoherence μ for both (u_i) and (v_i): $u_i^\top u_j \le \mu$ and $v_i^\top v_j \le \mu$ for $i \ne j$. Let $L_0 \triangleq \left(\frac{50}{1-\mu^2}\right)^2$. Let $L \ge L_0 \log(15d(k-1)/\delta)^2$. Then, with probability at least $1 - \delta$ and for ϵ small enough,*

$$\|\tilde{u}_j - u_j\|_2 \le \frac{8\epsilon}{1 - \mu^2} \frac{\sqrt{\|\pi\|_1 \pi_{\max}}}{\pi_{\min}^2} \|V^\top\|_2^2 \left(1 + C(\delta)\sqrt{d}\right),$$

where $C(\delta) \triangleq \frac{20+\sqrt{L_0}}{\sqrt{L}} \log(15(d(k-1))/\delta)$.

Proof. From Lemma 15 we have that

$$\|\tilde{u}_j - u_j\|_2 \le \epsilon \sqrt{\sum_{i=1}^{d} E_{ij}^2} + o(\epsilon),$$

for $j \in [k]$.

Using Lemma 9, we have that with probability at least $1 - \delta/(d(k-1))$,

$$\begin{aligned}
|E_{ij}| &\le \left(\frac{1}{\pi_i^2} + \frac{1}{\pi_j^2}\right) \left(\frac{\bar{p}_{ij}^\top r_{ij}}{1 - (u_i^\top u_j)^2} + \frac{\bar{\pi}_{ij}\|r_{ij}\|_2}{1 - (u_i^\top u_j)^2} \frac{(20 + \sqrt{L_0})\log(15(d(k-1))/\delta)}{\sqrt{L}}\right) \\
&\le \left(\frac{2}{\pi_{\min}^2}\right) \left(\frac{1}{1 - \mu^2}\right) \left(\bar{p}_{ij}^\top r_{ij} + 2|\pi_{\min}|\|V^\top\|_2^2 \frac{(20 + \sqrt{L_0})\log(15(d(k-1))/\delta)}{\sqrt{L}}\right) \\
&\le \left(\frac{2}{\pi_{\min}^2}\right) \left(\frac{1}{1 - \mu^2}\right) \left(\bar{p}_{ij}^\top r_{ij} + 2|\pi_{\min}|\|V^\top\|_2^2 C(\delta)\right),
\end{aligned}$$

where we have defined $C(\delta) \triangleq \frac{20+\sqrt{L_0}}{\sqrt{L}} \log(15(d(k-1))/\delta)$ and are using the fact that $u_i^\top u_j \le \mu$ and $\bar{\pi}_{ij} = |\pi_i| + |\pi_j| \le 2|\pi_{\max}|$.

Applying a union bound on all the entries of E_{ij}, we arrive at the following bound for all j.

$$\begin{aligned}
\|\tilde{u}_j - u_j\|_2 &\leq \epsilon \sqrt{\sum_{i \neq j} E_{ij}^2} \\
&\leq \frac{2\epsilon}{\pi_{\min}^2 (1-\mu^2)} \sqrt{\left(\sum_{i \neq j}^{d} \bar{p}_{ij}^\top r_{ij} + 2\pi_{\max} \|V^\top\|_2^2 C(\delta) \right)^2} \\
&\leq \frac{4\epsilon}{\pi_{\min}^2 (1-\mu^2)} \left(\sqrt{\sum_{i \neq j}^{d} (\bar{p}_{ij}^\top r_{ij})^2} + 2\pi_{\max} \|V^\top\|_2^2 C(\delta) \sqrt{\sum_{i \neq j}^{d} 1} \right) \\
&\leq \frac{4\epsilon}{\pi_{\min}^2 (1-\mu^2)} \left(\sqrt{\sum_{i \neq j}^{d} (\bar{p}_{ij}^\top r_{ij})^2} + 2\pi_{\max} \|V^\top\|_2^2 C(\delta) \sqrt{d} \right).
\end{aligned}$$

where we use the fact that $(a+b)^2 \leq 2(a^2+b^2)$.

By Lemma 10 we also have, $\sum_{j \neq i}^{d} (\bar{p}_{ij}^\top r_{ij})^2 \leq 4\|\pi\|_1 \pi_{\max} \|V^\top\|_2^4$. Finally, note that $\pi_{\max} \leq \sqrt{\pi_{\max}\|\pi\|_1}$:

$$\begin{aligned}
\|\tilde{u}_j - u_j\|_2 &\leq \frac{4\epsilon}{\pi_{\min}^2 (1-\mu^2)} \left(\sqrt{4\|\pi\|_1 \pi_{\max}} \|V^\top\|_2^2 + 2\pi_{\max} \|V^\top\|_2^2 C(\delta) \sqrt{d} \right) \\
&\leq \frac{8\epsilon}{1-\mu^2} \frac{\sqrt{\|\pi\|_1 \pi_{\max}}}{\pi_{\min}^2} \|V^\top\|_2^2 \left(1 + C(\delta)\sqrt{d} \right).
\end{aligned}$$

□

D Proofs of auxiliary lemmas

In this section, we prove some auxiliary results that appear as intermediate steps in the main lemmas above.

Lemma 10 (Bounding $p_{ij}^\top r_{ij}$)**.** *Let $p_{ij} \triangleq \pi_i u_i - \pi_j u_j \in \mathbb{R}^d$ and $r_{ij} \triangleq R(v_i, v_j, I) \in \mathbb{R}^d$, where R is a tensor with unit operator norm and where $(u_i) \in \mathbb{R}^d$ are unit vectors and $(v_i) \in \mathbb{R}^{d'}$ form the columns of the matrix V with bounded 2 norm. Then,*

$$\sum_{i \neq j}^{d} (p_{ij}^\top r_{ij})^2 \leq 4\pi_{\max} \|\pi\|_1 \|V\|_2^4.$$

Proof. Firstly, note that it is trivial to bound the sum as follows,

$$\begin{aligned}
\sum_{i \neq j}^{d} (p_{ij}^\top r_{ij})^2 &\leq \sum_{i \neq j}^{d} \|p_{ij}\|_2^2 \|r_{ij}\|_2^2 \\
&\leq 4(d-1)\pi_{\max}^2 \|V\|_2^4,
\end{aligned}$$

using the properties that $p_{ij} \triangleq \pi_i u_i - \pi_j u_j$ and that R has unit operator norm and thus $\|p_{ij}\|_2 \leq 2\pi_{\max}$ and $\|r_{ij}\|_2 = \|R(v_i, v_j, I)\|_2 \leq \|V\|_2^2$.

However, we would like a tighter bound with a lower-order dependence on k. To do so, let us expand p_{ij},

$$\begin{aligned}
\sum_{i \neq j}^{d} (p_{ij}^\top r_{ij})^2 &= \sum_{i \neq j}^{d} ((\pi_i u_i - \pi_j u_j)^\top r_{ij})^2 \\
&= \sum_{i \neq j}^{d} (\pi_i R(v_i, v_j, u_i) - \pi_j R(v_i, v_j, u_j))^2 \\
&= \sum_{i \neq j}^{d} \pi_j^2 R(v_i, v_j, u_j)^2 + \sum_{i \neq j}^{d} \pi_i^2 R(v_i, v_j, u_i)^2 - \sum_{i \neq j}^{d} 2\pi_i \pi_j R(v_i, v_j, u_i) R(v_i, v_j, u_j).
\end{aligned}$$

Using the assumption that R has unit norm, the latter two terms can be bounded by $\|\pi\|_2^2\|V\|_2^4$ and $2\,\pi_j\|\pi\|_1\|V\|_2^4$ respectively.

We now focus on the first term, $\pi_j^2\sum_{i\neq j}^d R(v_i,v_j,u_j)^2$. Note that $R(v_i,v_j,u_j) = R(I,v_j,u_j)^\top v_i = \tilde{r}_j^\top v_i$, where $\tilde{r}_j \triangleq R(I,v_j,u_j)$ and $\|\tilde{r}_j\|_2 \le \|V\|_2$ by the operator norm condition on R.

$$\begin{aligned}\sum_{i=1}^d (\tilde{r}_j^\top v_i)^2 &= \|V\tilde{r}_j\|_2^2 \\ &\le \|V\|_2^2\|\tilde{r}_j\|_2^2 \\ &= \|V\|_2^4\end{aligned}$$

Put together, we get that,

$$\sum_{i\neq j}^d (p_{ij}^\top r_{ij})^2 \le \pi_j^2\|V\|_2^4 + \|\pi\|_2^2\|V\|_2^4 + 2\pi_i\|\pi\|_1\|V\|_2^4.$$

Finally, $\pi_i^2 \le \pi_{\max}\|\pi\|_1$ and, by Hölder's inequality, $\|\pi\|_2^2 \le \pi_{\max}\|\pi\|_1$, giving us,

$$\sum_{i\neq j}^d (p_{ij}^\top r_{ij})^2 \le 4\pi_{\max}\|\pi\|_1\|V\|_2^4.$$

□

E Concentration Inequalities

In this section, we present several concentration results that are key to our results. The χ^2 tail bounds presented in Laurent and Massart [34] play a key role and are reproduced below.

Lemma 11 (χ_k^2 tail inequality)**.** *Let $q \sim \chi_k^2$ be distributed as a chi-squared variable with k degrees of freedom. Then, for any $t > 0$,*

$$\begin{aligned}\mathbb{P}(q - k > 2\sqrt{kt} + 2t) &\le e^{-t} \\ \mathbb{P}(k - q > 2\sqrt{kt}) &\le e^{-t}.\end{aligned}$$

Alternatively, we have that with probability at least $1-\delta$,

$$q \ge k\left(1 - \frac{2\log(1/\delta)}{\sqrt{k}}\right). \tag{16}$$

and similarly, with probability at least $1-\delta$,

$$q \le k\left(1 + 2\sqrt{\frac{\log(1/\delta)}{k}} + \frac{2\log(1/\delta)}{k}\right). \tag{17}$$

Proof. See Laurent and Massart [34, Lemma 1]. □

Lemma 12 (Gaussian quadratic forms)**.** *Let $x \sim \mathcal{N}(0, I) \in \mathbb{R}^d$ be a random Gaussian vector. If A is symmetric, $x^\top Ax$ is distributed as the sum of d independent χ^2 variables, $\sum_{i=1}^d \lambda_i(A)\chi_1^2$, where λ_i are the eigenvalues of A.*

Proof. Let $A = \sum_{i=1}^d \lambda_i u_i u_i^\top$ be the eigendecomposition of A. Then, $x^\top Ax = \sum_{i=1}^d \lambda_i\|u_i^\top x_i\|^2$. However, $u_i \top x_i$ is distributed as independent χ_1^2 random variables. Thus, $x^\top Ax = \sum_{i=1}^d \lambda_i\chi_1^2$. □

Lemma 13 (Gaussian products)**.** *Let $x_i \sim \mathcal{N}(0, I) \in \mathbb{R}^d$ for $i = 1, \ldots, L$ be random Gaussian vectors. Let $L \ge 4\log(1/\delta)$. Then,*

1. $\sum_{i=1}^{L}(x_i^\top a)^2$ *where* $a \in \mathbb{R}^d$ *is distributed as* $\|a\|_2^2\chi_L^2$. *Consequently, with probability at least* $1-\delta$,

$$\frac{1}{L}\sum_{i=1}^{L}(x_i^\top a)^2 \leq \|a\|_2^2\left(1+2\sqrt{\frac{\log(1/\delta)}{L}}+\frac{2\log(1/\delta)}{L}\right)$$
$$\leq \|a\|_2^2\left(1+3\sqrt{\frac{\log(1/\delta)}{L}}\right)$$
$$\frac{1}{L}\sum_{i=1}^{L}(x_i^\top a)^2 \geq \|a\|_2^2\left(1-\frac{2\log(1/\delta)}{\sqrt{L}}\right).$$

2. $\sum_{i=1}^{L} x_i^\top ab^\top x_i$ $a, b \in \mathbb{R}^d$ *and* $a \neq b$ *is sharply concentrated around* $a^\top b$*: with probability at least* $1-\delta$,

$$\frac{1}{L}\sum_{i=1}^{L} x_i^\top ab^\top x_i \leq a^\top b + \|a\|_2\|b\|_2\left(2\sqrt{\frac{\log(1/\delta)}{L}}+\frac{2\log(1/\delta)}{L}\right)$$
$$\leq a^\top b + \|a\|_2\|b\|_2\left(3\sqrt{\frac{\log(1/\delta)}{L}}\right).$$

Proof. The first part follows directly from Lemma 12 and the χ^2 tail bound, Lemma 11.

For the second part, let $A = \frac{ab^\top + ba^\top}{2}$. Note that $x_i^\top ab^\top x_i = x_i^\top A x_i$. Then, by Lemma 12, $x_i^\top A x_i = \lambda_1\chi_1^2 + \lambda_2\chi_1^2$, where λ_1 and λ_2 are the eigenvalues of A. Furthermore, because $A = \frac{ab^\top + ba^\top}{2}$, one of λ_1 or λ_2 is negative, and the other is positive. Without loss of generality, let $\lambda_1 > 0 > \lambda_2$.

Applying the χ^2 tail bound, Lemma 11, we get that with probability at least $1-\delta$,

$$\lambda_1\chi_1^2 \leq \lambda_1(1+2\sqrt{\frac{\log(2/\delta)}{L}}+2\frac{\log(2/\delta)}{L})$$
$$|\lambda_2|\chi_1^2 \geq |\lambda_2|(1-\frac{2\log(2/\delta)}{\sqrt{L}}).$$

Applying a union bound, we get,

$$\frac{1}{L}\sum_{i=1}^{L} x_i^\top ab^\top x_i \leq \lambda_1(1+2\sqrt{\frac{\log(2/\delta)}{L}}+2\frac{\log(2/\delta)}{L}) + \lambda_2(1-\frac{2\log(2/\delta)}{\sqrt{L}})$$
$$\leq (\lambda_1+\lambda_2) + |\lambda_1|\left(2\sqrt{\frac{\log(2/\delta)}{L}}+\frac{2\log(2/\delta)}{L}\right) + |\lambda_2|\frac{2\log(2/\delta)}{\sqrt{L}}$$
$$\leq (\lambda_1+\lambda_2) + (|\lambda_1|+|\lambda_2|)\left(2\sqrt{\frac{\log(2/\delta)}{L}}+\frac{2\log(2/\delta)}{L}\right).$$

Observe that $\lambda_1+\lambda_2 = \mathrm{tr}(A) = a^\top b$. Similarly, $|\lambda_1|+|\lambda_2| = \|A\|_* = 2(\frac{1}{2}\|a\|_2\|b\|_2)$. Thus, we finally have that with probability at least $1-\delta$,

$$\frac{1}{L}\sum_{i=1}^{L} x_i^\top ab^\top x_i \leq a^\top b + \|a\|_2\|b\|_2\left(2\sqrt{\frac{\log(2/\delta)}{L}}+\frac{2\log(2/\delta)}{L}\right).$$

□

F Perturbation bounds for joint diagonalization

In this section, we present minor extensions to the perturbation bounds of Cardoso [28] and Afsari [24] so that they apply in the low-rank setting.

Notation Let $M_l = U\Lambda_l U^\top + \epsilon R_l$ for $l = 1, 2, \ldots, L$ be a set of $d \times d$ matrices to be jointly diagonalized. $\Lambda_l \in \mathbb{R}^{k\times k}$ is a diagonal matrix, $R_l \in \mathbb{R}^{d\times d}$ is an arbitrary unit operator norm matrix and ϵ is a scalar. In the orthogonal setting, $U \in \mathbb{R}^{d\times k}$ is orthogonal, while in the non-orthogonal setting $U \in \mathbb{R}^{d\times k}$ is an arbitrary matrix with unit operator norm. Let $\lambda_{il} \triangleq (\Lambda_l)_i$ be the i-th factor weight of matrix M_l. Finally, we say that a set of matrices $\{M_1, \cdots, M_L\}$, $M_l = \sum_{i=1}^d \lambda_{il} u_i v_i^T$ has joint rank k if $\left|\{i \mid \sum_{l=1}^L |\lambda_{il}| > 0\}\right| = k$.

Lemma 14 (Cardoso [28])**.** *Let $M_l = U\Lambda_l U^\top + \epsilon R_l$, $l \in [L]$, be matrices with common factors $U \in \mathbb{R}^{d\times k}$ and diagonal $\Lambda_l \in \mathbb{R}^{k\times k}$. Let $\bar{U} \in \mathbb{R}^{d\times d}$ be a full-rank extension of U with columns $u_1, u_2, \ldots, u_d$ and let $\tilde{U} \in \mathbb{R}^{d\times d}$ be the orthogonal minimizer of the joint diagonalization objective $F(\cdot)$. Then, for all u_j, $j \in [k]$, there exists a column $\tilde{u}_j$ of $\tilde{U}$ such that*

$$\|\tilde{u}_j - u_j\|_2 \leq \epsilon\sqrt{\sum_{i=1}^{d} E_{ij}^2} + o(\epsilon), \tag{18}$$

where $E \in \mathbb{R}^{d\times k}$ is

$$E_{ij} \triangleq \frac{\sum_{l=1}^{L}(\lambda_{il} - \lambda_{jl}) u_j^\top R_l u_i}{\sum_{l=1}^{L}(\lambda_{il} - \lambda_{jl})^2} \tag{19}$$

when $i \neq j$ and $i \leq k$ or $j \leq k$. We define $E_{ij} = 0$ when $i = j$ and $\lambda_{il} = 0$ when $i > k$.

Proof. See Cardoso [28, Proposition 1]. Note that in the low rank setting, the entries of E_{ij} (Cardoso [28, Equation 15]) where $i, j > k$ are not defined, however, these terms only effect the last $d - k$ columns of $\tilde{U}$. The bounds for vectors $u_1, ..., u_k$ only depend on E_{ij} where $i \in [d]$ and $j \in [k]$, and these are derived in the low-rank setting in the same way as they are derived in the full-rank proof of Cardoso [28]. $\square$

We now present the corresponding perturbation bounds in Afsari [24] to the low rank setting.

Lemma 15 (Afsari [24])**.** *Let $M_l = U\Lambda_l U^\top + \epsilon R_l$, $l \in [L]$, be matrices with common factors $U \in \mathbb{R}^{d\times k}$ and diagonal $\Lambda_l \in \mathbb{R}^{k\times k}$. Let $\bar{U} \in \mathbb{R}^{d\times d}$ be a full-rank extension of U with columns $u_1, u_2, \ldots, u_d$ and let $\bar{V} = \bar{U}^{-1}$, with rows $v_1, v_2, \ldots, v_d$. Let $\tilde{V} \in \mathbb{R}^{d\times d}$ be the minimizer of the joint diagonalization objective $F(\cdot)$ and let $\tilde{U} = \tilde{V}^{-1}$.*

Then, for all u_j, $j \in [k]$, there exists a column $\tilde{u}_j$ of $\tilde{U}$ such that

$$\|\tilde{u}_j - u_j\|_2 \leq \epsilon\sqrt{\sum_{i=1}^{d} E_{ij}^2} + o(\epsilon), \tag{20}$$

where the entries of $E \in \mathbb{R}^{d\times k}$ satisfy the equation

$$\begin{bmatrix} E_{ij} \\ E_{ji} \end{bmatrix} = \frac{-1}{\gamma_{ij}(1 - \rho_{ij}^2)} \begin{bmatrix} \eta_{ij} & -\rho_{ij} \\ -\rho_{ij} & \eta_{ij}^{-1} \end{bmatrix} \begin{bmatrix} T_{ij} \\ T_{ji} \end{bmatrix}.$$

when $i \neq j$ and either $i \leq k$ or $j \leq k$. When $i = j$, $E_{ij} = 0$. The matrix T has zero on-diagonal elements, and is defined as

$$T_{ij} = \sum_l v_i^\top R_l v_j \lambda_{jl}, \qquad \text{for } 1 \leq j \neq i \leq d$$

and the other parameters are

$$\gamma_{ij} = \|\lambda_i\|_2 \|\lambda_j\|_2, \qquad \eta_{ij} = \frac{\|\lambda_i\|_2}{\|\lambda_j\|_2}, \qquad \rho_{ij} = \frac{\lambda_i^\top \lambda_j}{\|\lambda_j\|_2 \|\lambda_i\|_2}, \qquad (\lambda_i)_k = \lambda_{ik}.$$

We define $\lambda_{il} = 0$ when $i > k$.

Proof. In Afsari [24, Theorem 3] it is shown that $\tilde{V} = (I + \epsilon E)V + o(\epsilon)$, where E_{ij} is defined for $i, j \in [d]$ (Afsari [24, Equation 36]). Then,

$$\begin{aligned}\tilde{U} &= \tilde{U}(I + \epsilon E)^{-1} + o(\epsilon) \\ &= \tilde{U}(I - \epsilon E) + o(\epsilon).\end{aligned}$$

Note that, once again, in the low rank setting, the entries of E_{ij} when $i, j > k$ are not characterized by Afsari's results; however, these terms only effect the last $d - k$ columns of $\tilde{U}$. □

Lemma 16. *Let $M_l = U\Lambda_l U^\top + \epsilon R_l$, $l \in [L]$, be matrices with common factors $U \in \mathbb{R}^{d\times k}$ and diagonal $\Lambda_l \in \mathbb{R}^{k\times k}$. Let $\bar{U} \in \mathbb{R}^{d\times d}$ be a full-rank extension of U with columns $u_1, u_2, \ldots, u_d$ and let $\bar{V} = \bar{U}^{-1}$, with rows $v_1, v_2, \ldots, v_d$. Let $\tilde{V} \in \mathbb{R}^{d\times d}$ be the minimizer of the joint diagonalization objective $F(\cdot)$ and let $\tilde{U} = \tilde{V}^{-1}$.*

Then, for all u_j, $j \in [k]$, there exists a column $\tilde{u}_j$ of $\tilde{U}$ such that

$$\|\tilde{u}_j - u_j\|_2 \leq \epsilon\sqrt{\sum_{i=1}^{d} E_{ij}^2} + o(\epsilon), \tag{21}$$

where the entries of $E \in \mathbb{R}^{d\times k}$ are bounded by

$$|E_{ij}| \leq \frac{1}{1-\rho_{ij}^2}\left(\frac{1}{\|\lambda_i\|_2^2} + \frac{1}{\|\lambda_j\|_2^2}\right)\left(\left|\sum_{l=1}^{L} v_i^\top R_l v_j \lambda_{jl}\right| + \left|\sum_{l=1}^{L} v_i^\top R_l v_j \lambda_{il}\right|\right),$$

when $i \neq j$ and $E_{ij} = 0$ when $i = j$ and $\lambda_{il} = 0$ when $i > k$. Here $\lambda_i = (\lambda_{i1}, \lambda_{i2}, ..., \lambda_{iL}) \in \mathbb{R}^L$ and $\rho_{ij} = \frac{\lambda_i^\top \lambda_j}{\|\lambda_i\|_2\|\lambda_j\|_2}$ is the modulus of uniqueness, a measure of how ill-conditioned the problem is.

Proof. From Lemma 15, we have that

$$\left\|\begin{bmatrix}E_{ij}\\E_{ji}\end{bmatrix}\right\| \leq \frac{\eta_{ij}+\eta_{ji}}{\gamma_{ji}(1-\rho_{ij}^2)}\left\|\begin{bmatrix}T_{ij}\\T_{ji}\end{bmatrix}\right\|,$$

where

$$\gamma_{ij} = \|\lambda_i\|_2\|\lambda_j\|_2, \qquad \eta_{ij} = \frac{\|\lambda_i\|_2}{\|\lambda_j\|_2}, \qquad \rho_{ij} = \frac{\lambda_i^\top\lambda_j}{\|\lambda_j\|_2\|\lambda_i\|_2},$$

and the matrix T is defined to be zero on the diagonal and for $i \neq j$ defined as

$$T_{ij} = \sum_{l=1}^{L} v_i^\top R_l v_j \lambda_{jl}, \qquad \text{for } 1 \leq j \neq i \leq d$$

Taking $\|\cdot\|$ to be the l_1-norm in the above expression, we have that

$$|E_{ij}| \leq |E_{ij}| + |E_{ji}| \leq \frac{\eta_{ij}+\eta_{ji}}{\gamma_{ji}(1-\rho_{ij}^2)}\left(|T_{ij}| + |T_{ji}|\right).$$

Since

$$\frac{\eta_{ij}+\eta_{ji}}{\gamma_{ji}} = \frac{\|\lambda_i\|_2^2 + \|\lambda_j\|_2^2}{\|\lambda_i\|_2^2\|\lambda_j\|_2^2} = \frac{1}{\|\lambda_i\|_2^2} + \frac{1}{\|\lambda_j\|_2^2}$$

and

$$T_{ij} = \sum_{l=1}^{L} v_i^\top R_l v_j \lambda_{jl},$$

the claim follows. □