A Bounds for block sparse tensors

One of the main bounds to control is the spectral norm of the sparse perturbation tensor S. The success of the power iterations and the improvement in accuracy of recovery over iterative steps of RTD requires this bound.

Lemma 4 (Spectral norm bounds for block sparse tensors)**.** *Let $M \in \mathbb{R}^{n\times n\times n}$ satisfy the block sparsity assumption (S). Then*

$$\|M\|_2 = O(d^{1.5}\|M\|_\infty). \tag{10}$$

Proof. Let $\Psi \in \mathbb{R}^{n\times n\times n}$ be a tensor that encodes the sparsity of M i.e. $\Psi_{i,j,k} = 1$ iff $S^*_{i,j,k} \neq 0$ for all $i,j,k \in [n]$. We have that

$$\begin{aligned}
\|M\| &= \max_{u:\|u\|=1} \sum_{i,j,k} M_{i,j,k} u(i)u(j)u(k) \\
&= \max_{u:\|u\|=1} \sum_{i,j,k} M_{i,j,k}\Psi_{i,j,k} u(i)u(j)u(k) \\
&\le \max_{u:\|u\|=1} \sum_{i,j,k} |M_{i,j,k}\Psi_{i,j,k} u(i)u(j)u(k)| \\
&\le \|M\|_\infty \max_{u:\|u\|=1} \sum_{i,j,k} |\Psi_{i,j,k} u(i)u(j)u(k)| = \|M\|_\infty \|\Psi\|,
\end{aligned}$$

where the last inequality is from Perron Frobenius theorem for non-negative tensors [9]. Note that Ψ is non-negative by definition. Now we bound $\|\Psi\|$ on lines of [3, Lemma 4]. Recall that $\forall i \in [B]$, $j \in [n]$,

$$\Psi = \sum_{i=1}^{B} \psi_i \otimes \psi_i \otimes \psi_i, \quad \|\psi_i\|_0 \le d,\ \psi_i(j) = 0 \text{ or } 1.$$

By definition $\|\psi_i\|_2 = \sqrt{d}$. Define normalized vectors $\tilde{\psi_i} := \psi_i/\|\psi_i\|$. We have

$$\Psi = d^{1.5} \sum_{i=1}^{B} \tilde{\psi_i} \otimes \tilde{\psi_i} \otimes \tilde{\psi_i}$$

Define matrix $\tilde{\psi} := [\tilde{\psi_1}|\tilde{\psi_2}, \ldots \tilde{\psi_B}]$. Note that $\tilde{\psi}^\top\tilde{\psi} \in \mathbb{R}^{B\times B}$ is a matrix with unit diagonal entries and absolute values of off-diagonal entries bounded by η, by assumption. From Gershgorin Disk Theorem, every subset of L columns in $\tilde{\psi}$ has singular values within $1 \pm o(1)$, where $L < \frac{1}{\eta}$. Moreover, from Gershgorin Disk Theorem, $\|\tilde{\psi}\| < \sqrt{1 + B\eta}$.

For any unit vector u, let S be the set of L indices that are largest in $\tilde{\psi}^\top u$. By the argument above we know $\|(\tilde{\psi}_S)^\top u\| \le \|\tilde{\psi}_S\|\|u\| \le 1 + o(1)$. In particular, the smallest entry in $\tilde{\psi}_S^\top u$ is at most $2/\sqrt{L}$. By construction of S this implies for all i not in S, $|\tilde{\psi}_i^\top u|$ is at most $2/\sqrt{L}$. Now we can write the ℓ_3 norm of $\tilde{\psi}^\top u$ as

$$\begin{aligned}
\|\tilde{\psi}^\top u\|_3^3 &= \sum_{i\in S} |\tilde{\psi}_i^\top u|^3 + \sum_{i\notin S} |\tilde{\psi}_i^\top u|^3 \\
&\le \sum_{i\in S} |\tilde{\psi}_i^\top u|^2 + (2/\sqrt{L})^{3-2} \sum_{i\notin S} |\tilde{\psi}_i^\top u|^2 \\
&\le 1 + 2\sqrt{\eta}\|\tilde{\psi}\|^2 \le 1 + 2B\eta^{1.5}.
\end{aligned}$$

Here the first inequality uses that every entry outside S is small, and last inequality uses the bound argued on $\|(\tilde{\psi}_S)^\top u\|$, the spectral norm bound is assumed on A_{S^c}. Since $B = O(\eta^{-1.5})$, we have the result. ■

Another important bound required is ∞-norm of certain contractions of the (normalized) sparse tensor and its powers, which we denote by M below. We use a loose bound based on spectral norm and we require $\|M\| < 1/\sqrt{n}$. However, this constraint will also be needed for the power iterations to succeed and is not an additional requirement. Thus, the loose bound below will suffice for our results to hold.

Lemma 5 (Infinity norm bounds)**.** *Let $M \in \mathbb{R}^{n\times n\times n}$ satisfy the block sparsity assumption (S). Let u, v satisfy the assumption $\|u\|_\infty, \|v\|_\infty \le \frac{\mu}{n^{1/2}}$. Then, we have*

1. $\|M(u,v,I)\|_\infty \le \frac{\kappa\mu}{n^{1/2}}\|M\|_\infty$, *where* $\kappa := \frac{Bd^2\mu}{\sqrt{n}}$.

2. $\|[M(u,v,I)]^p\|_\infty \le \kappa\mu\|M\|_\infty\|M\|^{p-1}$ *for* $p > 1$.

3. $\sum_{p\ge 1}\|[M(u,I,I)]^p v\|_\infty \le \frac{\kappa\mu}{\sqrt{n}}\|M\|_\infty \cdot \frac{\|M\|}{1-\|M\|}$ *when* $\|M\| < 1/\sqrt{n}$.

Proof. We have from norm conversion

$$\|M(u,v,I)\|_\infty \le \|u\|_\infty \cdot \|v\|_\infty \max_j \|M(I,I,e_j)\|_1 \tag{11}$$

$$\le \frac{\mu^2}{n} \cdot Bd^2\|M\|_\infty, \tag{12}$$

where ℓ_1 norm (i.e. sum of absolute values of entries) of a slice $M(I,I,e_j)$ is Bd^2, since the number of non-zero entries in one block in a slice is d^2.

Let $Z = M(u,I,I) \in \mathbb{R}^{n\times n}$. Now, $\|M(u,I,I)^p v\|_\infty = \|Z^p v\|_\infty = \|Z^{p-1}a\|_\infty$ where $a = Zv$. Now,

$$\|Z^{p-1}a\|_\infty = \max_j |e_j^T Z^{p-1}a| \le \|Z^{p-1}\|_2\|a\|_2 \le \|Z\|_2^{p-1}\|a\|_2 \le \|M\|^{p-1}\sqrt{n}\|a\|_\infty \le \kappa\mu\|M\|_\infty\|M\|^{p-1}.$$

Hence, $\sum_{p\ge 1}\|[M(u,I,I)]^p v\|_\infty \le \kappa\mu\|M\|_\infty \cdot \frac{\|M\|_2}{1-\|M\|_2}$. ■

B Proof of Theorem 1

Lemma 6. *Let L^*, S^* be symmetric and satisfy the assumptions of Theorem 1 and let $S^{(t)}$ be the t^{th} iterate of the l^{th} stage of Algorithm 1. Let $\sigma_1^*, \dots, \sigma_r^*$ be the eigenvalues of L^*, such that $\sigma_1^* \ge \cdots \ge \sigma_r^* \ge 0$ and $\lambda_1, \cdots, \lambda_r$ be the eigenvalues of $T - S^{(t)}$ such that $\lambda_1 \ge \cdots \ge \lambda_r \ge 0$. Recall that $E^{(t)} := S^* - S^{(t)}$. Suppose further that*

1. $\|E^{(t)}\|_\infty \le \frac{8\mu^3 k}{n^{3/2}}\left(\sigma_{l+1}^* + \left(\frac{1}{2}\right)^{t-1}\sigma_l^*\right)$, *and*

2. $\operatorname{supp} E^{(t)} \subseteq \operatorname{supp} S^*$.

Then, for some constant $c \in [0,1)$, we have

$$(1-c)\left(\sigma_{l+1}^* + \left(\frac{1}{2}\right)^t \sigma_l^*\right) \le \left(\lambda_{l+1} + \left(\frac{1}{2}\right)^t \lambda_l\right) \le (1+c)\left(\sigma_{l+1}^* + \left(\frac{1}{2}\right)^t \sigma_l^*\right). \tag{13}$$

Proof. Note that $T - S^{(t)} = L^* + E^{(t)}$. Now,

$$\left|\lambda_{l+1} - \sigma_{l+1}^*\right| \le 8\left\|E^{(t)}\right\|_2 \le 8d^{3/2}\|E^{(t)}\|_\infty \le \frac{8\mu^3 r\gamma_t}{n^{3/2}}d^{3/2},$$

where $\gamma_t := \left(\sigma_{l+1}^* + \left(\frac{1}{2}\right)^{t-1}\sigma_l^*\right)$. That is, $\left|\lambda_{l+1} - \sigma_{l+1}^*\right| \le 8\mu^3 r\left(\frac{d}{n}\right)^{3/2}\gamma_t$. Similarly, $|\lambda_l - \sigma_l^*| \le 8\mu^3 r\left(\frac{d}{n}\right)^{3/2}\gamma_t$. So we have:

$$\begin{aligned}\left|\left(\lambda_{l+1} + \left(\frac{1}{2}\right)^t \lambda_l\right) - \left(\sigma_{l+1}^* + \left(\frac{1}{2}\right)^t \sigma_l^*\right)\right| &\le 8\mu^3 r\left(\frac{d}{n}\right)^{3/2}\gamma_t\left(1 + \left(\frac{1}{2}\right)^t\right) \\ &\le 16\mu^3 r\left(\frac{d}{n}\right)^{3/2}\gamma_t \\ &\le c\left(\sigma_{l+1}^* + \left(\frac{1}{2}\right)^t \sigma_l^*\right),\end{aligned}$$

where the last inequality follows from the bound $d \le \left(\frac{n}{c'\mu^3 k}\right)^{2/3}$ for some constant c'. ■

Lemma 7. *Assume the notation of Lemma 6. Also, let $L^{(t)}, S^{(t)}$ be the t^{th} iterates of r^{th} stage of Algorithm 1 and $L^{(t+1)}, S^{(t+1)}$ be the $(t+1)^{th}$ iterates of the same stage. Also, recall that $E^{(t)} := S^* - S^{(t)}$ and $E^{(t+1)} := S^* - S^{(t+1)}$.*

Suppose further that

1. $\|E^{(t)}\|_\infty \le \frac{8\mu^3 r}{n^{3/2}} \left(\sigma^*_{l+1} + \left(\frac{1}{2}\right)^{t-1} \sigma^*_l\right)$, *and*
2. $\operatorname{supp} E^{(t)} \subseteq \operatorname{supp} S^*$.
3. $\|E^{(t)}\|_2 < \frac{C\sigma^*_l}{\sqrt{n}}$, *where $C < 1/2$ is a sufficiently small constant.*

Then, we have:

$$\|L^{(t+1)} - L^*\|_\infty \le 2\frac{\mu^3 r}{n^{3/2}} \left(\sigma^*_{l+1} + \left(\frac{1}{2}\right)^t \sigma^*_l\right)$$

Proof. Let $L^{(t+1)} = \sum_{i=1}^l \lambda_i u_i^{(t+1)}$ be the eigen decomposition obtained using the tensor power method on $(T - S^{(t)})$ at the $(t+1)^{th}$ step of the l^{th} stage. Also, recall that $T - S^{(t)} = L^* + E^{(t)}$ where $L^* = \sum_{j=1}^r \sigma^*_j u_j^{\otimes 3}$. Define $E^{(t)} := S^* - S^{(t)}$. Define $E^i := E^{(t)}(u_i^{(t+1)}, I, I)$. Let $\|E^{(t)}\|_2 := \epsilon$.

Consider the eigenvalue equation $(T - S^{(t)})(u_i^{(t+1)}, u_i^{(t+1)}, I) = \lambda_i u_i^{(t+1)}$:

$$L^*(u_i^{(t+1)}, u_i^{(t+1)}, I) + E^{(t)}(u_i^{(t+1)}, u_i^{(t+1)}, I) = \lambda_i u_i^{(t+1)}$$

$$\sum_{j=1}^r \sigma^*_i \left\langle u_i^{(t+1)}, u_j \right\rangle^2 u_j + E^{(t)}(u_i^{(t+1)}, u_i^{(t+1)}, I) = \lambda_i u_i^{(t+1)}$$

$$[\lambda_i I - E^{(t)}(u_i^{(t+1)}, I, I)]u_i^{(t+1)} = \sum_{j=1}^r \sigma^*_i \left\langle u_i^{(t+1)}, u_j \right\rangle^2 u_j$$

$$u_i^{(t+1)} = \left[I + \sum_{p \ge 1} \left(\frac{E^i}{\lambda_i}\right)^p\right] \sum_{j=1}^r \frac{\sigma^*_i}{\lambda_i} \left\langle u_i^{(t+1)}, u_j \right\rangle^2 u_j$$

Now,

$$\begin{aligned}
\|L^{(t+1)} - L^*\|_\infty &\le \left\| \sum_{i \in [l]} \lambda_i (u_i^{(t+1)})^{\otimes 3} - \sum_{i \in [l]} \sigma^*_i {u_i}^{\otimes 3} \right\|_\infty + \left\| \sum_{i=l+1}^r \sigma^*_i {u_i}^{\otimes 3} \right\|_\infty \\
&\le \sum_{i \in [l]} \left\| \lambda_i (u_i^{(t+1)})^{\otimes 3} - \sigma^*_i {u_i}^{\otimes 3} \right\|_\infty + \sum_{i=l+1}^r \|\sigma^*_i {u_i}^{\otimes 3}\|_\infty
\end{aligned}$$

For a fixed i, using $\lambda_i \le \sigma^*_i + \epsilon$ [2] and using Lemma 11, we obtain

$$\begin{aligned}
\left\| \lambda_i (u_i^{(t+1)})^{\otimes 3} - \sigma^*_i {u_i}^{\otimes 3} \right\|_\infty &\le \left\| (\sigma^*_i + \epsilon)(u_i^{(t+1)})^{\otimes 3} - \sigma^*_i {u_i}^{\otimes 3} \right\|_\infty \\
&\le \left\| \sigma^*_i (u_i^{(t+1)})^{\otimes 3} - \sigma^*_i {u_i}^{\otimes 3} \right\|_\infty + \epsilon \left\| (u_i^{(t+1)})^{\otimes 3} \right\|_\infty \\
&\le \sigma^*_i \left\| (u_i^{(t+1)})^{\otimes 3} - {u_i}^{\otimes 3} \right\|_\infty + \epsilon \left\| (u_i^{(t+1)})^{\otimes 3} \right\|_\infty \\
&\le \sigma^*_i [3\|u_i^{(t+1)} - u_i\|_\infty \|u_i\|^2_\infty + 3\|u_i^{(t+1)} - u_i\|^2_\infty \|u_i\|_\infty + \|u_i^{(t+1)} - u_i\|^3_\infty] \\
&\quad + \epsilon \|(u_i^{(t+1)})^{\otimes 3}\|_\infty \\
&\le 7\sigma^*_i \|u_i^{(t+1)} - u_i\|_\infty \|u_i\|^2_\infty + \epsilon \|(u_i^{(t+1)})\|^3_\infty
\end{aligned}$$

Now,

$$\begin{aligned}
\left\| u_i^{(t+1)} - u_i \right\|_\infty &= \left\| (\sum_{j=1}^{r} \frac{\sigma_i^*}{\lambda_i} \left\langle u_i^{(t+1)}, u_j \right\rangle^2 u_j - u_i) + \sum_{j=1, p\geq 1}^{r} \frac{\sigma_i^*}{\lambda_i} \left\langle u_i^{(t+1)}, u_j \right\rangle^2 (E^i)^p u_j \right\|_\infty \\
&\leq \left\| (1 - \frac{\sigma_i^*}{\lambda_i} \left\langle u_i^{(t+1)}, u_i \right\rangle^2) u_i \right\|_\infty + \left\| \sum_{j\neq i} \frac{\sigma_i^*}{\lambda_i} \left\langle u_i^{(t+1)}, u_j \right\rangle^2 u_j \right\|_\infty \\
&+ \left\| \sum_{p\geq 1} \frac{\sigma_i^*}{\lambda_i} \left\langle u_i^{(t+1)}, u_i \right\rangle^2 \left(\frac{E^i}{\lambda_i}\right)^p u_i \right\|_\infty + \left\| \sum_{p, j\neq i} \frac{\sigma_i^*}{\lambda_i} \left\langle u_i^{(t+1)}, u_j \right\rangle^2 \left(\frac{E^i}{\lambda_i}\right)^p u_j \right\|_\infty
\end{aligned}$$

For the first term, we have

$$\begin{aligned}
\left\| (1 - \frac{\sigma_i^*}{\lambda_i} \left\langle u_i^{(t+1)}, u_i \right\rangle^2) u_i \right\|_\infty &\leq \left(1 - \frac{\sigma_i^*}{\sigma_i^* + \epsilon} \left(1 - \left(\frac{\epsilon}{\sigma_i^*}\right)^2\right)\right) \|u_i\|_\infty \leq \left(1 - \left(1 - \frac{\epsilon}{\sigma_i^*}\right)\right) \frac{\mu}{n^{1/2}} \\
&\leq \frac{\mu}{\sigma_i^* n^{1/2}} \epsilon \leq \frac{C\mu\sigma_l^*}{\sigma_i^* n}
\end{aligned}$$

where we substitute for ϵ in the last step.

For the second term, we have

$$\left\| \sum_{j\neq i} \frac{\sigma_i^*}{\lambda_i} \left\langle u_i^{(t+1)}, u_j \right\rangle^2 u_j \right\|_\infty \leq \frac{\sigma_i^*}{\sigma_i^* - \epsilon} \left(\frac{\epsilon}{\sigma_i^*}\right)^2 \|u_i\|_\infty \leq 2 \left(\frac{\epsilon}{\sigma_i^*}\right)^2 \frac{\mu}{n^{1/2}},$$

which is a lower order term.

Next,

$$\begin{aligned}
\left\| \sum_{p\geq 1} \frac{\sigma_i^*}{\lambda_i} \left\langle u_i^{(t+1)}, u_i \right\rangle^2 \left(\frac{E^i}{\lambda_i}\right)^p u_i \right\|_\infty &\leq \left\| \sum_{p\geq 1} \frac{\sigma_i^*}{\lambda_i} \left(\frac{E^i}{\lambda_i}\right)^p u_i \right\|_\infty \leq \sum_{p\geq 1} \frac{\sigma_i^*}{\lambda_i} \left\| \left(\frac{E^i}{\lambda_i}\right)^p u_i \right\|_\infty \\
&\leq \frac{\sigma_i^*}{\lambda_i} \cdot \frac{\mu}{\sqrt{n}} \cdot \frac{\|E^{(t)}\|_2 \sqrt{n}/\lambda_i}{1 - \|E^{(t)}\|_2 \sqrt{n}/\lambda_i} \\
&\leq \frac{2}{(1-C)} \frac{\kappa_t \mu}{\lambda_i \sqrt{n}} \|E^{(t)}\|_\infty
\end{aligned}$$

from Lemma 5, and the assumption on spectral norm of $\|E^{(t)}\|_2$, where

$$\kappa_t := \frac{Bd^2\mu}{\sqrt{n}}.$$

For the remaining terms, we have

$$\left\| \sum_{p, j\neq i} \frac{\sigma_i^*}{\lambda_i} \left\langle u_i^{(t+1)}, u_j \right\rangle^2 \left(\frac{E^i}{\lambda_i}\right)^p u_j \right\|_\infty \leq \sum_{j\neq i} \frac{\sigma_i^*}{\lambda_i} \left\langle u_i^{(t+1)}, u_j \right\rangle^2 \left\| \sum_{p\geq 1} \left(\frac{E^i}{\lambda_i}\right)^p u_1 \right\|_\infty \leq \frac{\sigma_i^*}{\lambda_i} \left\| \sum_{p\geq 1} \left(\frac{E^i}{\lambda_i}\right)^p u_1 \right\|_\infty \left(\frac{\epsilon}{\sigma_i^*}\right)^2,$$

which is a lower order term.

Combining the above and recalling $\epsilon \ll \sigma_i^*$, $\forall i \in [l]$

$$\left\| u_i^{(t+1)} - u_i \right\|_\infty \leq \frac{8}{(1-C)} \frac{\kappa_t \mu}{\lambda_i \sqrt{n}} \|E^{(t)}\|_\infty.$$

Also, from Lemma 1

$$|\lambda_i - \sigma_i^*| \leq 8\|E^{(t)}\|_2 \leq 8\epsilon$$

Thus, from the above two equations, we obtain the bound for the parameters (eigenvectors and eigenvalues) of the low-rank tensor $\left\|u_i^{(t+1)} - u_i\right\|_\infty$ and $\|\lambda_i - \sigma_i^*\|_\infty$. We combine the individual parameter recovery bounds as:

$$\begin{aligned}\left\|\sum_{i\in[l]} \lambda_i (u_i^{(t+1)})^{\otimes 3} - \sum_{i\in[l]} \sigma_i^* u_i^{\otimes 3}\right\|_\infty &\le r[7\sigma_i^* \|u_i^{(t+1)} - u_i\|_\infty \|u_i\|_\infty^2 + \epsilon \|(u_i^{(t+1)})\|_\infty^3] \\ &\le \frac{224}{1-C}\frac{\kappa_t \mu^3 r}{n^{1.5}}\|E^{(t)}\|_\infty \end{aligned} \tag{14}$$

and the other term

$$\sum_{i=l+1}^{r} \|\sigma_i^* u_i^{\otimes 3}\|_\infty \le \sigma_{l+1}^* \frac{r\mu^3}{n^{1.5}}.$$

Combining bound in (14) with the above, we have

$$\|L^{(t+1)} - L^*\|_\infty \le \frac{r\mu^3}{n^{1.5}}\left(\frac{224}{1-C}\kappa_t \|E^{(t)}\|_\infty + \sigma_{l+1}^*\right) < \frac{1}{4}\|E^{(t)}\|_\infty.$$

where the last inequality comes from the fact that $\frac{r\mu^3}{n^{1.5}}\sigma_{l+1}^* \le \frac{\|E^{(t)}\|_\infty}{8}$ and the assumption that $C < 1/2$, and we can choose B and d s.t.

$$\frac{448 r\mu^3}{n^{1.5}}\kappa_t < \frac{1}{8}.$$

This is possible from assumption (S). ■

The following lemma bounds the support of $E^{(t+1)}$ and $\|E^{(t+1)}\|_\infty$, using an assumption on $\|L^{(t+1)} - L^*\|_\infty$.

Lemma 8. *Assume the notation of Lemma 7. Suppose*

$$\|L^{(t+1)} - L^*\|_\infty \le 2\frac{\mu^3 r}{n^{3/2}}\left(\sigma_{l+1}^* + \left(\frac{1}{2}\right)^{t-1}\sigma_l^*\right).$$

Then, we have:

1. $\operatorname{supp} E^{(t+1)} \subseteq \operatorname{supp} S^*$.
2. $\|E^{(t+1)}\|_\infty \le 7\frac{\mu^3 r}{n^{3/2}}\left(\sigma_{l+1}^* + \left(\frac{1}{2}\right)^t \sigma_l^*\right)$, *and*

Proof. We first prove the first conclusion. Recall that,

$$S^{(t+1)} = H_\zeta(T - L^{(t+1)}) = H_\zeta(L^* - L^{(t+1)} + S^*),$$

where $\zeta = 4\frac{\mu^3 r}{n^{3/2}}\left(\lambda_{l+1} + \left(\frac{1}{2}\right)^t \lambda_l\right)$ is as defined in Algorithm 1 and $\lambda_1, \cdots, \lambda_n$ are the eigenvalues of $T - S^{(t)}$ such that $\lambda_1 \ge \cdots \ge \lambda_n$.

If $S_{abc}^* = 0$ then $E_{ijk}^{(t+1)} = \mathbf{1}_{\left|L_{abc}^* - L_{abc}^{(t+1)}\right| > \zeta} \cdot (L_{abc}^* - L_{abc}^{(t+1)})$. The first part of the lemma now follows by using the assumption that $\|L^{(t+1)} - L^*\|_\infty \le 2\frac{\mu^3 r}{n^{3/2}}\left(\sigma_{l+1}^* + \left(\frac{1}{2}\right)^t \sigma_l^*\right) \overset{(\zeta_1)}{\le} 4\frac{\mu^3 r}{n^{3/2}}\left(\lambda_{l+1} + \left(\frac{1}{2}\right)^t \lambda_l\right) = \zeta$, where (ζ_1) follows from Lemma 6.

We now prove the second conclusion. We consider the following two cases:

1. $\left|T_{abc} - L_{abc}^{(t+1)}\right| > \zeta$: Here, $S_{abc}^{(t+1)} = S_{abc}^* + L_{abc}^* - L_{abc}^{(t+1)}$. Hence, $|S_{abc}^{(t+1)} - S_{abc}^*| \le |L_{abc}^* - L_{abc}^{(t+1)}| \le 2\frac{\mu^3 r}{n^{3/2}}\left(\sigma_{l+1}^* + \left(\frac{1}{2}\right)^t \sigma_l^*\right)$.
2. $\left|T_{abc} - L_{abc}^{(t+1)}\right| \le \zeta$: In this case, $S_{abc}^{(t+1)} = 0$ and $\left|S_{abc}^* + L_{abc}^* - L_{abc}^{(t+1)}\right| \le \zeta$. So we have, $\left|E_{abc}^{(t+1)}\right| = |S_{abc}^*| \le \zeta + \left|L_{abc}^* - L_{abc}^{(t+1)}\right| \le 7\frac{\mu^3 r}{n^{3/2}}\left(\sigma_{l+1}^* + \left(\frac{1}{2}\right)^t \sigma_l^*\right)$. The last inequality above follows from Lemma 6.

This proves the lemma. ■

Theorem 9. *Let L^*, S^* be symmetric and satisfy (L) and (S), and $\beta = 4\frac{\mu^3 r}{n^{3/2}}$. The outputs $\widehat{L}$ (and its parameters $\hat{u}_i$ and $\hat{\lambda}_i$) and $\widehat{S}$ of Algorithm 1 satisfy w.h.p.:*

$$\|\hat{u}_i - u_i\|_\infty \leq \frac{\delta}{\mu^2 r n^{1/2} \sigma^*_{\min}}, \quad |\hat{\lambda}_i - \sigma^*_i| \leq \delta, \quad \forall i \in [n],$$
$$\left\|\widehat{L} - L^*\right\|_F \leq \delta, \quad \|\widehat{S} - S^*\|_\infty \leq \frac{\delta}{n^{3/2}}, \text{ and } \quad \operatorname{supp} \widehat{S} \subseteq \operatorname{supp} S^*.$$

Proof. Recall that in the l^{th} stage, the update $L^{(t+1)}$ is given by: $L^{(t+1)} = P_l(T - S^{(t)})$ and $S^{(t+1)}$ is given by: $S^{(t+1)} = H_\zeta(T - L^{(t+1)})$. Also, recall that $E^{(t)} := S^* - S^{(t)}$ and $E^{(t+1)} := S^* - S^{(t+1)}$.

We prove the lemma by induction on both l and t. For the base case ($l = 1$ and $t = -1$), we first note that the first inequality on $\|L^{(0)} - L^*\|_\infty$ is trivially satisfied. Due to the thresholding step (step 3 in Algorithm 1) and the incoherence assumption on L^*, we have:

$$\|E^{(0)}\|_\infty \leq \frac{8\mu^3 r}{n^{3/2}} \left(\sigma^*_2 + 2\sigma^*_1\right), \text{ and}$$
$$\operatorname{supp} E^{(0)} \subseteq \operatorname{supp} S^*.$$

So the base case of induction is satisfied.

We first do the inductive step over t (for a fixed r). By inductive hypothesis we assume that: a) $\|E^{(t)}\|_\infty \leq \frac{8\mu^3 r}{n^{3/2}} \left(\sigma^*_{l+1} + \left(\frac{1}{2}\right)^{t-1} \sigma^*_l\right)$, b) $\operatorname{supp} E^{(t)} \subseteq \operatorname{supp} S^*$. Then by Lemma 7, we have:

$$\|L^{(t+1)} - L^*\|_\infty \leq \frac{2\mu^3 r}{n^{3/2}} \left(\sigma^*_{l+1} + \left(\frac{1}{2}\right)^t \sigma^*_l\right).$$

Lemma 8 now tells us that

1. $\|E^{(t+1)}\|_\infty \leq \frac{8\mu^3 r}{n^{3/2}} \left(\sigma^*_{l+1} + \left(\frac{1}{2}\right)^t \sigma^*_l\right)$, and
2. $\operatorname{supp} E^{(t+1)} \subseteq \operatorname{supp} S^*$.

This finishes the induction over t. Note that we show a stronger bound than necessary on $\|E^{(t+1)}\|_\infty$.

We now do the induction over l. Suppose the hypothesis holds for stage l. Let T denote the number of iterations in each stage. We first obtain a lower bound on T. Since

$$\left\|T - S^{(0)}\right\|_2 \geq \|L^*\|_2 - \left\|E^{(0)}\right\|_2 \geq \sigma^*_1 - d^{3/2}\|E^{(0)}\|_\infty \geq \frac{3}{4}\sigma^*_1,$$

we see that $T \geq 10 \log\left(3\mu^3 r \sigma^*_1/\delta\right)$. So, at the end of stage r, we have:

1. $\|E^{(T)}\|_\infty \leq \frac{7\mu^3 r}{n^{3/2}} \left(\sigma^*_{l+1} + \left(\frac{1}{2}\right)^T \sigma^*_l\right) \leq \frac{7\mu^3 r \sigma^*_{l+1}}{n^{3/2}} + \frac{\delta}{10n}$, and
2. $\operatorname{supp} E^{(T)} \subseteq \operatorname{supp} S^*$.

Recall, $\left|\sigma_{r+1}\left(T - S^{(T)}\right) - \sigma^*_{r+1}\right| \leq \left\|E^{(T)}\right\|_2 \leq \frac{d}{n}\left(\mu^3 r \left|\sigma^*_{r+1}\right| + \delta\right)$. We will now consider two cases:

1. **Algorithm 1 terminates:** This means that $\beta\sigma_{r+1}\left(T - S^{(T)}\right) < \frac{\delta}{2n^{3/2}}$ which then implies that $\sigma^*_{r+1} < \frac{\delta}{6\mu^3 r}$. So we have:

$$\|\widehat{L} - L^*\|_\infty = \|L^{(T)} - L^*\|_\infty \leq \frac{2\mu^3 r}{n^{3/2}} \left(\sigma^*_{r+1} + \left(\frac{1}{2}\right)^T \sigma^*_r\right) \leq \frac{\delta}{5n^{3/2}}.$$

   This proves the statement about $\widehat{L}$ and its parameters (eigenvalues and eigenvectors). A similar argument proves the claim on $\|\widehat{S} - S^*\|_\infty$. The claim on $\operatorname{supp} \widehat{S}$ follows since $\operatorname{supp} E^{(T)} \subseteq \operatorname{supp} S^*$.

2. **Algorithm 1 continues to stage $(r+1)$:** This means that $\beta\sigma_{r+1}\left(L^{(T)}\right) \geq \frac{\delta}{2n^{3/2}}$ which then implies that $\sigma^*_{r+1} > \frac{\delta}{8\mu^3 r}$. So we have:

$$\begin{aligned}\|E^{(T)}\|_\infty &\leq \frac{8\mu^3 r}{n^{3/2}}\left(\sigma^*_{r+1} + \left(\frac{1}{2}\right)^T \sigma^*_r\right)\\ &\leq \frac{8\mu^3 r}{n^{3/2}}\left(\sigma^*_{l+1} + \frac{\delta}{10\mu^3 r n^{3/2}}\right)\\ &\leq \frac{8\mu^3 r}{n^{3/2}}\left(\sigma^*_{l+1} + \frac{8\sigma^*_{l+1}}{10n}\right)\\ &\leq \frac{8\mu^3 r}{n^{3/2}}\left(\sigma^*_{l+2} + 2\sigma^*_{l+1}\right).\end{aligned}$$

Similarly for $\|L^{(T)} - L^*\|_\infty$.

This finishes the proof. ■

B.1 Short proof of Corollary 1

The state of art guarantees for robust matrix PCA requires that the overall sparsity along any row or column of the input matrix be $D = O(\frac{n}{r\mu^2})$ (when the input matrix is $\mathbb{R}^{n\times n}$).

Under (S), the total sparsity along any row or column of M_i is given by $D := dB$. Now, Theorem 1 holds when the sparsity condition in (5) is satisfied. That is, RTD succeeds when

$$D = O(d\cdot B) = O\left(\min\left(\frac{n^{4/3}}{r^{1/3}\mu^2}, \frac{n^{2/3}}{r^{2/3}\mu^2}(\frac{n}{r})^{1/3}\right)\right) = O\left(\frac{n}{r\mu^2}\right).$$

Hence, RTD can handle larger amount of corruption than the matrix methods and the gain becomes more significant for smaller η.

B.2 Some auxiliary lemmas

We recall Theorem 5.1 from [2]. Let $\epsilon = 8\|E^{(t)}\|_2$ where $E^{(t)} := S^* - S^{(t)}$.

Lemma 10. *Let $L^{(t+1)} = \sum_{i=1}^k \lambda_i u_i^{(t+1)}$ be the eigen decomposition obtained using Algorithm 1 on $(T - S^{(t)})$. Then,*

1. *If $\|u_i^{(t+1)} - u_i\|_2 \leq \frac{\epsilon}{\sigma^*_{\min}}$, then $\mathrm{dist}(u_i^{(t+1)}, u_i) \leq \frac{\epsilon}{\sigma^*_{\min}}$.*
2. $\sum_{j\neq i}\left\langle u_i^{(t+1)}, u_j\right\rangle^2 \leq \left(\frac{\epsilon}{\sigma^*_{\min}}\right)^2$.
3. $\|u_i^{(t+1)}\|_\infty \leq \frac{\mu}{n^{1/2}} + \frac{\epsilon}{\sigma^*_{\min}}$.
4. $|\sigma^*_i| - \epsilon \leq |\lambda_i| \leq |\sigma^*_i| + \epsilon$.

Proof. 1. Let $z \perp u$ and $\|z\|_2 = 1$.

$$\begin{aligned}u_i^{(t+1)} &= \left\langle u_i^{(t+1)}, u_i\right\rangle u_i + \mathrm{dist}(u_i^{(t+1)}, u_i) z\\ \|u_i^{(t+1)} - u_i\|_2^2 &= (\left\langle u_i^{(t+1)}, u_i\right\rangle - 1)^2\|u_i\|_2^2 + \mathrm{dist}(u_i^{(t+1)}, u_i)\|z\|_2^2 + 0\\ &\geq (\mathrm{dist}(u_i^{(t+1)}, u_i))^2\end{aligned}$$

Then using Theorem 5.1 from [2], we obtain the result. Next, since $\langle u_i^{(t+1)}, u_i\rangle^2 + \mathrm{dist}(u_i^{(t+1)}, u_i)^2 = 1$, we have $\langle u_i^{(t+1)}, u_i\rangle^2 \geq 1 - \left(\frac{\epsilon}{\sigma^*_{\min}}\right)^2$.

2. Note that

$$u_i^{(t+1)} = \sum_{j=1}^{k} \left\langle u_i^{(t+1)}, u_j \right\rangle u_j + \mathrm{dist}(u_i^{(t+1)}, U)z$$

where $z \perp U$ such that $\|z\|_2 = 1$. Using $\|u_i^{(t+1)}\|_2 = 1$ and the Pythagoras theorem, we get

$$1 - \left\langle u_i^{(t+1)}, u_i \right\rangle^2 = \sum_{j\neq i} \left\langle u_i^{(t+1)}, u_j \right\rangle^2 + \mathrm{dist}(u_i^{(t+1)}, U)^2.1 \geq \sum_{j\neq i} \left\langle u_i^{(t+1)}, u_j \right\rangle^2$$

Using part 1 of Lemma 10, we get $\sum_{j\neq i} \left\langle u_i^{(t+1)}, u_j \right\rangle^2 \leq \left(\frac{\epsilon}{\sigma^*_{\min}}\right)^2$.

3. We have

$$u_i^{(t+1)} = \left\langle u_i^{(t+1)}, u_i \right\rangle u_i + \mathrm{dist}(u_i^{(t+1)}, u_i)z$$
$$\|u_i^{(t+1)}\|_\infty \leq | \left\langle u_i^{(t+1)}, u_i \right\rangle | \|u_i\|_\infty + | \,\mathrm{dist}(u_i^{(t+1)}, u_i)| \|z\|_\infty \leq 1.\frac{\mu}{n^{1/2}} + \frac{\epsilon}{\sigma^*_{\min}}$$

4. This follows from Theorem 5.1 from [2], i.e., $\forall i,\ ||\lambda_i| - |\sigma_i^*|| \leq \epsilon$.

■

Lemma 11. *Let $a = b+\epsilon.\overrightarrow{1}$ where a, b are any 2 vectors and $\epsilon > 0$. Then, $\|a^{\otimes 3} - b^{\otimes 3}\|_\infty \leq \|a-b\|_\infty.\|b\|_\infty^2 + O(\epsilon^2)$.*

Proof. We have

$$\|a^{\otimes 3} - b^{\otimes 3}\|_\infty = \|(b + \epsilon\overrightarrow{1})^{\otimes 3} - b^{\otimes 3}\|_\infty$$

Let (i, j, k) be the maximum element. Therefore,

$$\begin{aligned}\|(b + \epsilon\overrightarrow{1})^{\otimes 3} - b^{\otimes 3}\|_\infty &= (b_i + \epsilon)(b_j + \epsilon)(b_k + \epsilon) - b_i b_j b_k \\ &= \epsilon(b_i b_j + b_j b_k + b_k b_i) + \epsilon^2(b_i + b_j + b_k) + \epsilon^3\end{aligned}$$

With $b_i \leq c\ \forall i$ for some $c > 0$ and $\epsilon = \|a - b\|_\infty$, we have $\|a^{\otimes 3} - b^{\otimes 3}\|_\infty \leq 3\epsilon c^2 + O(\epsilon^2)$ ■

C Symmetric embedding of an asymmetric tensor

We use the symmetric embedding $sym(L)$ of a tensor L as defined in Section 2.3 of [24]. We focus on third order tensors which have low CP-rank. We have three properties to derive that is relevant to us:

1. *Symmetry:* From Lemma 2.2 of [24] we see that $sym(L)$ for any tensor is symmetric.

2. *CP-Rank:* From Equation 6.5 of [24] we see that CP-rank($sym(L)$) $\leq$ 6.CP-rank(L). Since this is a constant, we see that the symmetric embedding is also a low-rank tensor.

3. *Incoherece:* Theorem 4.7 of [24] says that if u_1, u_2 and u_3 are unit modal singular vectors of T, then the vector $\tilde{u} = 3^{-1/2}[u_1; u_2; u_3]$ is a unit eigenvector of $sym(T)$. Without loss of generality, assume that T is of size $n_1 \times n_2 \times n_3$ with $n_1 \leq n_2 \leq n_3$. In this case, we have

$$\|\tilde{u}\|_\infty \leq \frac{\mu}{(3n_1)^{1/2}} \tag{15}$$

and

$$\|\tilde{u}\|_\infty \leq \frac{\tilde{\mu}}{(n_1 + n_2 + n_3)^{1/2}} \tag{16}$$

for $\tilde{\mu} = c\mu$ for some constant c to be calculated. Equating the right hand sides of Equations (15) and (16), we obtain $c = [(n_1 + n_2 + n_3)/(3n_1)]^{1/2}$. When $\Theta(n_1) = \Theta(n_2) = \Theta(n_3)$, we see that the eigenvectors $\tilde{u}$ of $sym(T)$ as specified above have the incoherence-preserving property.

D Proof of Theorem 2

Let $\widetilde{L}$ be a symmetric tensor which is a perturbed version of an orthogonal tensor L^*, $\widetilde{L} = L^* + E \in \mathbb{R}^{n\times n\times n}$, $L^* = \sum_{i\in[r]} \sigma_i^* u_i^{\otimes 3}$, where $\sigma_1^* \geq \sigma_2^* \ldots \sigma_r^* > 0$ and $\{u_1, u_2, \ldots, u_r\}$ form an orthonormal basis.

The analysis proceeds iteratively. First, we prove convergence to eigenpair of $\widetilde{L}$, which is close to top eigenpair (σ_1^*, u_1) of L^*. We then argue that the same holds on the deflated tensor, when the perturbation E satisfies (8). from This finishes the proof of Theorem 2.

To prove convergence for the first stage, i.e. convergence to eigenpair of $\widetilde{L}$, which is close to top eigenpair (σ_1^*, u_1) of L^*, we analyze two phases of the shifted power iteration. In the first phase, we prove that with N_1 initializations and N_2 power iterations, we get close to true top eigenpair of L^*, i.e. (σ_1^*, u_1). After this, in the second phase, we prove convergence to an eigenpair of $\widetilde{L}$.

The proof of the second phase is outlined in the main text. Here, we now provide proof for the first phase.

D.1 Analysis of first phase of shifted power iteration

In this section, we prove that the output of shifted power method is close to original eigenpairs of the (unperturbed) orthogonal tensor, i.e. Theorem 2 holds, except for the property that the output corresponds to the eigenpairs of the perturbed tensor. We adapt the proof of tensor power iteration from [2] but here, since we consider the shifted power method, we need to modify it. We adopt the notation of [2] in this section.

Recall the update rule used in the shifted power method. Let $\theta_t = \sum_{i=1}^k \theta_{i,t} v_i \in \mathbb{R}^k$ be the unit vector at time t. Then

$$\theta_{t+1} = \sum_{i=1}^k \theta_{i,t+1} v_i := (\tilde{T}(I, \theta_t, \theta_t) + \alpha\theta_t)/\|(\tilde{T}(I, \theta_t, \theta_t) + \alpha\theta_t)\|.$$

In this subsection, we assume that $\tilde{T}$ has the form

$$\tilde{T} = \sum_{i=1}^k \tilde{\lambda}_i v_i^{\otimes 3} + \tilde{E} \tag{17}$$

where $\{v_1, v_2, \ldots, v_k\}$ is an orthonormal basis, and, without loss of generality,

$$\tilde{\lambda}_1 |\theta_{1,t}| = \max_{i\in[k]} \tilde{\lambda}_i |\theta_{i,t}| > 0.$$

Also, define

$$\tilde{\lambda}_{\min} := \min\{\tilde{\lambda}_i : i \in [k],\ \tilde{\lambda}_i > 0\}, \quad \tilde{\lambda}_{\max} := \max\{\tilde{\lambda}_i : i \in [k]\}.$$

We assume the error $\tilde{E}$ is a symmetric tensor such that, for some constant $p > 1$,

$$\|\tilde{E}(I, u, u)\| \leq \tilde{\epsilon}, \quad \forall u \in S^{k-1}; \tag{18}$$

$$\|\tilde{E}(I, u, u)\| \leq \tilde{\epsilon}/p, \quad \forall u \in S^{k-1} \text{ s.t. } (u^\top v_1)^2 \geq 1 - (3\tilde{\epsilon}/\tilde{\lambda}_1)^2. \tag{19}$$

In the next two propositions (Propositions D.1 and D.2) and Lemmas D.1, we analyze the power method iterations using $\tilde{T}$ at some arbitrary iterate θ_t using only the property (18) of $\tilde{E}$. But throughout, the quantity $\tilde{\epsilon}$ can be replaced by $\tilde{\epsilon}/p$ if θ_t satisfies $(\theta_t^\top v_1)^2 \geq 1 - (3\tilde{\epsilon}/\tilde{\lambda}_1)^2$ as per property (19).

Define

$$\begin{aligned} R_\tau &:= \left(\frac{\theta_{1,\tau}^2}{1-\theta_{1,\tau}^2}\right)^{1/2}, \qquad r_{i,\tau} := \frac{\tilde{\lambda}_1 \theta_{1,\tau}}{\tilde{\lambda}_i |\theta_{i,\tau}|}, \\ \gamma_\tau &:= 1 - \frac{1}{\min_{i\neq 1} |r_{i,\tau}|}, \quad \delta_\tau := \frac{\tilde{\epsilon}}{\tilde{\lambda}_1 \theta_{1,\tau}^2}, \quad \kappa := \frac{\tilde{\lambda}_{\max}}{\tilde{\lambda}_1} \end{aligned} \tag{20}$$

for $\tau \in \{t, t+1\}$.

Proposition D.1.

$$\min_{i\neq 1}|r_{i,t}| \geq \frac{R_t}{\kappa}, \qquad \gamma_t \geq 1-\frac{\kappa}{R_t}, \qquad \theta_{1,t}^2 = \frac{R_t^2}{1+R_t^2}.$$

Proposition D.2.

$$r_{i,t+1} \geq r_{i,t}^2 \cdot \frac{1-\delta_t+\frac{\alpha}{\tilde{\lambda}_1\theta_{1,t}}}{1+\kappa\delta_t r_{i,t}^2+\frac{\alpha}{\tilde{\lambda}_i\theta_{i,t}}}, \quad i\in[k], \tag{21}$$

$$R_{t+1} \geq= R_t \cdot \frac{1-\delta_t+\frac{\alpha}{\tilde{\lambda}_1|\theta_{1,t}|}}{1-\gamma_t+\left(\delta_t+\frac{\alpha(1-\theta_{1,t})^{1/2}}{\tilde{\lambda}_1\theta_{1,t}^2}\right)R_t} \geq \frac{1-\delta_t+\frac{\alpha}{\tilde{\lambda}_1|\theta_{1,t}|}}{\frac{\kappa}{R_t^2}+\delta_t+\frac{\alpha(1-\theta_{1,t})^{1/2}}{\tilde{\lambda}_1\theta_{1,t}^2}}. \tag{22}$$

Proof. Let $\breve{\theta}_{t+1} := \tilde{T}(I,\theta_t,\theta_t)+\alpha\theta_t$, so $\theta_{t+1} = \breve{\theta}_{t+1}/\|\breve{\theta}_{t+1}\|$. Since $\breve{\theta}_{i,t+1} = \tilde{T}(v_i,\theta_t,\theta_t) = T(v_i,\theta_t,\theta_t)+\alpha\theta_t + E(v_i,\theta_t,\theta_t)$, we have

$$\breve{\theta}_{i,t+1} = \tilde{\lambda}_i\theta_{i,t}^2 + E(v_i,\theta_t,\theta_t)+\alpha\theta_t^\top v_i, \quad i\in[k].$$

By definition, we have $\theta_{i,t} = \theta_t^\top v_i$. Using the triangle inequality and the fact $\|E(v_i,\theta_t,\theta_t)\| \leq \tilde{\epsilon}$, we have

$$\breve{\theta}_{i,t+1} \geq \tilde{\lambda}_i\theta_{i,t}^2 - \tilde{\epsilon}+\alpha\theta_{i,t} \geq |\theta_{i,t}|\cdot\left(\tilde{\lambda}_i|\theta_{i,t}| - \tilde{\epsilon}/|\theta_{i,t}|+\alpha\right) \tag{23}$$

and

$$|\breve{\theta}_{i,t+1}| \leq |\tilde{\lambda}_i\theta_{i,t}^2| + \tilde{\epsilon}+\alpha\theta_{i,t} \leq |\theta_{i,t}|\cdot\left(\tilde{\lambda}_i|\theta_{i,t}| + \tilde{\epsilon}/|\theta_{i,t}|+\alpha\right) \tag{24}$$

for all $i\in[k]$. Combining (23) and (24) gives

$$r_{i,t+1} = \frac{\tilde{\lambda}_1\theta_{1,t+1}}{\tilde{\lambda}_i|\theta_{i,t+1}|} = \frac{\tilde{\lambda}_1\breve{\theta}_{1,t+1}}{\tilde{\lambda}_i|\breve{\theta}_{i,t+1}|} \geq r_{i,t}^2 \cdot \frac{1-\delta_t+\frac{\alpha}{\tilde{\lambda}_1\theta_{1,t}}}{1+\frac{\tilde{\epsilon}}{\tilde{\lambda}_i\theta_{i,t}^2}+\frac{\alpha}{\tilde{\lambda}_i\theta_{i,t}}} = r_{i,t}^2 \cdot \frac{1-\delta_t+\frac{\alpha}{\tilde{\lambda}_1\theta_{1,t}}}{1+(\tilde{\lambda}_i/\tilde{\lambda}_1)\delta_t r_{i,t}^2+\frac{\alpha}{\tilde{\lambda}_i\theta_{i,t}}} \geq r_{i,t}^2 \cdot \frac{1-\delta_t+\frac{\alpha}{\tilde{\lambda}_1\theta_{1,t}}}{1+\kappa\delta_t r_{i,t}^2+\frac{\alpha}{\tilde{\lambda}_i\theta_{i,t}}}.$$

Moreover, by the triangle inequality and Hölder's inequality,

$$\begin{aligned}
\left(\sum_{i=2}^n [\breve{\theta}_{i,t+1}]^2\right)^{1/2} &= \left(\sum_{i=2}^n \left(\tilde{\lambda}_i\theta_{i,t}^2 + E(v_i,\theta_t,\theta_t)+\alpha\theta_{i,t}\right)^2\right)^{1/2} \\
&\leq \left(\sum_{i=2}^n \tilde{\lambda}_i^2\theta_{i,t}^4\right)^{1/2} + \left(\sum_{i=2}^n E(v_i,\theta_t,\theta_t)^2\right)^{1/2} + \left(\sum_{i=2}^k \alpha^2\theta_{i,t}^2\right)^{1/2} \\
&\leq \max_{i\neq 1}\tilde{\lambda}_i|\theta_{i,t}|\left(\sum_{i=2}^n \theta_{i,t}^2\right)^{1/2} + \tilde{\epsilon} + \left(\alpha^2\sum_{i=2}^k \theta_{i,t}^2\right)^{1/2} \\
&= (1-\theta_{1,t}^2)^{1/2}\cdot\left(\max_{i\neq 1}\tilde{\lambda}_i|\theta_{i,t}| + \tilde{\epsilon}/(1-\theta_{1,t}^2)^{1/2}+\alpha\right).
\end{aligned} \tag{25}$$

Combining (23) and (25) gives

$$\frac{|\theta_{1,t+1}|}{(1-\theta_{1,t+1}^2)^{1/2}} = \frac{|\breve{\theta}_{1,t+1}|}{\left(\sum_{i=2}^n [\breve{\theta}_{i,t+1}]^2\right)^{1/2}} \geq \frac{|\theta_{1,t}|}{(1-\theta_{1,t}^2)^{1/2}} \cdot \frac{\tilde{\lambda}_1|\theta_{1,t}| - \tilde{\epsilon}/|\theta_{1,t}|+\alpha}{\max_{i\neq 1}\tilde{\lambda}_i|\theta_{i,t}| + \tilde{\epsilon}/(1-\theta_{1,t}^2)^{1/2}+\alpha}.$$

In terms of R_{t+1}, R_t, γ_t, and δ_t, this reads

$$\begin{aligned}
R_{t+1} &\geq \frac{1-\delta_t+\frac{\alpha}{\tilde{\lambda}_1|\theta_{1,t}|}}{(1-\gamma_t)\left(\frac{1-\theta_{1,t}^2}{\theta_{1,t}^2}\right)^{1/2}+\delta_t+\frac{\alpha(1-\theta_{1,t})^{1/2}}{\tilde{\lambda}_1\theta_{1,t}^2}} = R_t \cdot \frac{1-\delta_t+\frac{\alpha}{\tilde{\lambda}_1|\theta_{1,t}|}}{1-\gamma_t+\left(\delta_t+\frac{\alpha(1-\theta_{1,t}^2)^{1/2}}{\tilde{\lambda}_1\theta_{1,t}^2}\right)R_t} \\
&= \frac{1-\delta_t+\frac{\alpha}{\tilde{\lambda}_1|\theta_{1,t}|}}{\frac{1-\gamma_t}{R_t}+\left(\delta_t+\frac{\alpha(1-\theta_{1,t}^2)^{1/2}}{\tilde{\lambda}_1\theta_{1,t}^2}\right)} \geq \frac{1-\delta_t+\frac{\alpha}{\tilde{\lambda}_1|\theta_{1,t}|}}{\frac{\kappa}{R_t^2}+\delta_t+\frac{\alpha(1-\theta_{1,t}^2)^{1/2}}{\tilde{\lambda}_1\theta_{1,t}^2}}
\end{aligned}$$

where the last inequality follows from Proposition D.1. ∎

Lemma D.1. *Fix any $\rho > 1$. Assume*

$$0 \leq \delta_t < \min\left\{\frac{1}{2(1+2\kappa\rho^2)}, \ \frac{1-1/\rho}{1+\kappa\rho}\right\}$$

and $\gamma_t > 2(1+2\kappa\rho^2)\delta_t$.

1. *If* $r_{i,t}^2 \leq 2\rho^2$, *then* $r_{i,t+1} \geq |r_{i,t}|\left(1+\frac{\gamma_t}{2}\right)$.

Proof. By (21) from Proposition D.2,

$$r_{i,t+1} \geq r_{i,t}^2 \cdot \frac{1-\delta_t+\frac{\alpha}{\tilde{\lambda}_1\theta_{1,t}}}{1+\kappa\delta_t r_{i,t}^2+\frac{\alpha}{\tilde{\lambda}_i\theta_{i,t}}} \geq |r_{i,t}| \cdot \frac{1}{1-\gamma_t} \cdot \frac{1-\delta_t+\frac{\alpha}{\tilde{\lambda}_1\theta_{1,t}}}{1+2\kappa\rho^2\delta_t+\frac{\alpha}{\tilde{\lambda}_i\theta_{i,t}}} \geq |r_{i,t}|\left(1+\frac{\gamma_t}{2}\right)$$

where the last inequality is seen as follows: Let

$$\xi = 2.\frac{1-\delta_t+\frac{\alpha}{\tilde{\lambda}_1\theta_{1,t}}}{1+2\kappa\rho^2\delta_t+\frac{\alpha}{\tilde{\lambda}_i\theta_{i,t}}}$$

Then, we have $\gamma_t^2+\gamma_t-2+\xi \geq 0$. The positive root is $\frac{-1+(9-4\xi)^{1/2}}{2}$. Since $\gamma_t \geq 0$, we have $(9-4\xi)^{1/2} \geq 1$, so we assume $\xi \leq 2$ for the inequality to hold, i.e., $\frac{\alpha}{\tilde{\lambda}_1\theta_{1,t}} - \frac{\alpha}{\tilde{\lambda}_i\theta_{i,t}} \leq (1+2\kappa\rho^2)\delta_t$. ∎

The rest of the proof is along the similar lines of [2], except that we use SVD initialization instead of random initialization. The proof of SVD initialization is given in [3].