A Bilinear Maximization Guarantees

Lemma 3.2. *For any real $m \times n$, rank-r matrix $\widetilde{\mathbf{A}}$ and arbitrary norm-bounded sets $\mathcal{X} \subset \mathbb{R}^{m\times k}$ and $\mathcal{Y} \subset \mathbb{R}^{n\times k}$, let*

$$(\widetilde{\mathbf{X}}_\star, \widetilde{\mathbf{Y}}_\star) \triangleq \underset{\mathbf{X}\in\mathcal{X},\mathbf{Y}\in\mathcal{Y}}{\arg\max} \operatorname{TR}\big(\mathbf{X}^\top \widetilde{\mathbf{A}}\mathbf{Y}\big).$$

If there exist operators $\mathrm{P}_\mathcal{X} : \mathbb{R}^{m\times k} \to \mathcal{X}$ such that

$$\mathrm{P}_\mathcal{X}(\mathbf{L}) = \underset{\mathbf{X}\in\mathcal{X}}{\arg\max} \operatorname{TR}\big(\mathbf{X}^\top \mathbf{L}\big)$$

and similarly, $\mathrm{P}_\mathcal{Y} : \mathbb{R}^{n\times k} \to \mathcal{Y}$ such that

$$\mathrm{P}_\mathcal{Y}(\mathbf{R}) = \underset{\mathbf{Y}\in\mathcal{Y}}{\arg\max} \operatorname{TR}\big(\mathbf{R}^\top \mathbf{Y}\big)$$

with running times $\mathrm{T}_\mathcal{X}$ and $\mathrm{T}_\mathcal{Y}$, respectively, then Algorithm 1 outputs $\widetilde{\mathbf{X}} \in \mathcal{X}$ and $\widetilde{\mathbf{Y}} \in \mathcal{Y}$ such that

$$\operatorname{TR}\big(\widetilde{\mathbf{X}}^\top \widetilde{\mathbf{A}}\widetilde{\mathbf{Y}}\big) \geq \operatorname{TR}\big(\widetilde{\mathbf{X}}_\star^\top \widetilde{\mathbf{A}}\widetilde{\mathbf{Y}}_\star\big) - 2\epsilon\sqrt{k}\cdot\|\widetilde{\mathbf{A}}\|_2\cdot\mu_\mathcal{X}\cdot\mu_\mathcal{Y},$$

where $\mu_\mathcal{X} \triangleq \max_{\mathbf{X}\in\mathcal{X}} \|\mathbf{X}\|_\mathrm{F}$ and $\mu_\mathcal{Y} \triangleq \max_{\mathbf{Y}\in\mathcal{Y}} \|\mathbf{Y}\|_\mathrm{F}$, in time $O\big(\big(2\sqrt{r}/\epsilon\big)^{r\cdot k}\cdot\big(\mathrm{T}_\mathcal{X}+\mathrm{T}_\mathcal{Y}+(m+n)r\big)\big)+\mathrm{T}_{\textsf{SVD}}(r)$.

Proof. In the sequel, $\widetilde{\mathbf{U}}$, $\widetilde{\mathbf{\Sigma}}$ and $\widetilde{\mathbf{V}}$ are used to denote the r-truncated singular value decomposition of $\widetilde{\mathbf{A}}$.

Without loss of generality, we assume that $\mu_\mathcal{X} = \mu_\mathcal{Y} = 1$ since the variables in $\mathcal{X}$ and $\mathcal{Y}$ can be normalized by $\mu_\mathcal{X}$ and $\mu_\mathcal{Y}$, respectively, while simultaneously scaling the singular values of $\widetilde{\mathbf{A}}$ by a factor of $\mu_\mathcal{X}\cdot\mu_\mathcal{Y}$. Then, $\|\mathbf{Y}\|_{\infty,2} \leq 1$, $\forall\, \mathbf{Y} \in \mathcal{Y}$, where $\|\mathbf{Y}\|_{\infty,2}$ denotes the maximum of the ℓ_2-norm of the columns of $\mathbf{Y}$.

Let $\widetilde{\mathbf{X}}_\star, \widetilde{\mathbf{Y}}_\star$ be the optimal pair on $\widetilde{\mathbf{A}}$, *i.e.*,

$$(\widetilde{\mathbf{X}}_\star, \widetilde{\mathbf{Y}}_\star) \triangleq \underset{\mathbf{X}\in\mathcal{X},\mathbf{Y}\in\mathcal{Y}}{\arg\max} \operatorname{TR}\big(\mathbf{X}^\top \widetilde{\mathbf{A}}\mathbf{Y}\big)$$

and define the $r \times k$ matrix $\widetilde{\mathbf{C}}_\star \triangleq \widetilde{\mathbf{V}}^\top \widetilde{\mathbf{Y}}_\star$. Note that

$$\begin{aligned}\|\widetilde{\mathbf{C}}_\star\|_{\infty,2} &= \|\widetilde{\mathbf{V}}^\top \widetilde{\mathbf{Y}}_\star\|_{\infty,2} \\ &= \max_{1\leq i\leq k} \|\widetilde{\mathbf{V}}^\top [\widetilde{\mathbf{Y}}_\star]_{:,i}\|_2 \leq 1, \end{aligned} \tag{14}$$

with the last inequality following from the facts that $\|\mathbf{Y}\|_{\infty,2} \leq 1\ \forall\ \mathbf{Y} \in \mathcal{Y}$ and the columns of $\widetilde{\mathbf{V}}$ are orthonormal. Alg. 1 iterates over the points in $(\mathbb{B}_2^{r-1})^{\otimes k}$. The latter is used to describe the set of $r \times k$ matrices whose columns have ℓ_2 norm at most equal to 1. At each point, the algorithm computes a candidate solution. By (14), the ϵ-net contains an $r \times k$ matrix $\mathbf{C}_\sharp$ such that

$$\|\mathbf{C}_\sharp - \widetilde{\mathbf{C}}_\star\|_{\infty,2} \leq \epsilon.$$

Let $\mathbf{X}_\sharp, \mathbf{Y}_\sharp$ be the candidate pair computed at $\mathbf{C}_\sharp$ by the two step maximization, *i.e.*,

$$\mathbf{X}_\sharp \triangleq \underset{\mathbf{X}\in\mathcal{X}}{\arg\max} \operatorname{TR}\big(\mathbf{X}^\top \widetilde{\mathbf{U}}\widetilde{\mathbf{\Sigma}}\mathbf{C}_\sharp\big)$$

and

$$\mathbf{Y}_\sharp \triangleq \underset{\mathbf{Y}\in\mathcal{Y}}{\arg\max} \operatorname{TR}\big(\mathbf{X}_\sharp^\top \widetilde{\mathbf{A}}\mathbf{Y}\big). \tag{15}$$

We show that the objective values achieved by the candidate pair $\mathbf{X}_\sharp, \mathbf{Y}_\sharp$ satisfies the inequality of the lemma implying the desired result.

By the definition of $\widetilde{\mathbf{C}}_\star$ and the linearity of the trace,

$$\begin{aligned}&\operatorname{TR}\big(\widetilde{\mathbf{X}}_\star^\top \widetilde{\mathbf{A}}\widetilde{\mathbf{Y}}_\star\big) \\ &= \operatorname{TR}\big(\widetilde{\mathbf{X}}_\star^\top \widetilde{\mathbf{U}}\widetilde{\mathbf{\Sigma}}\widetilde{\mathbf{C}}_\star\big) \\ &= \operatorname{TR}\big(\widetilde{\mathbf{X}}_\star^\top \widetilde{\mathbf{U}}\widetilde{\mathbf{\Sigma}}\mathbf{C}_\sharp\big) + \operatorname{TR}\big(\widetilde{\mathbf{X}}_\star^\top \widetilde{\mathbf{U}}\widetilde{\mathbf{\Sigma}}\big(\widetilde{\mathbf{C}}_\star - \mathbf{C}_\sharp\big)\big) \\ &\leq \operatorname{TR}\big(\mathbf{X}_\sharp^\top \widetilde{\mathbf{U}}\widetilde{\mathbf{\Sigma}}\mathbf{C}_\sharp\big) + \operatorname{TR}\big(\widetilde{\mathbf{X}}_\star^\top \widetilde{\mathbf{U}}\widetilde{\mathbf{\Sigma}}\big(\widetilde{\mathbf{C}}_\star - \mathbf{C}_\sharp\big)\big). \end{aligned} \tag{16}$$

The inequality follows from the fact that (by definition (15)) $\mathbf{X}_\sharp$ maximizes the first term over all $\mathbf{X} \in \mathcal{X}$. We compute an upper bound on the right hand side of (16). Define

$$\widehat{\mathbf{Y}} \triangleq \underset{\mathbf{Y}\in\mathcal{Y}}{\arg\min} \|\widetilde{V}^\top \mathbf{Y} - \mathbf{C}_\sharp\|_{\infty,2}.$$

(We note that $\widehat{\mathbf{Y}}$ is used for the analysis and is never explicitly calculated.) Further, define the $r \times k$ matrix $\widehat{\mathbf{C}} \triangleq \widetilde{\mathbf{V}}^\top \widehat{\mathbf{Y}}$. By the linearity of the trace operator

$$\begin{aligned}&\operatorname{TR}\big(\mathbf{X}_\sharp^\top \widetilde{\mathbf{U}}\widetilde{\mathbf{\Sigma}}\mathbf{C}_\sharp\big) \\ &= \operatorname{TR}\big(\mathbf{X}_\sharp^\top \widetilde{\mathbf{U}}\widetilde{\mathbf{\Sigma}}\widehat{\mathbf{C}}\big) + \operatorname{TR}\big(\mathbf{X}_\sharp^\top \widetilde{\mathbf{U}}\widetilde{\mathbf{\Sigma}}\big(\mathbf{C}_\sharp - \widehat{\mathbf{C}}\big)\big) \\ &= \operatorname{TR}\big(\mathbf{X}_\sharp^\top \widetilde{\mathbf{U}}\widetilde{\mathbf{\Sigma}}\widetilde{\mathbf{V}}^\top \widehat{\mathbf{Y}}\big) + \operatorname{TR}\big(\mathbf{X}_\sharp^\top \widetilde{\mathbf{U}}\widetilde{\mathbf{\Sigma}}\big(\mathbf{C}_\sharp - \widehat{\mathbf{C}}\big)\big) \\ &\leq \operatorname{TR}\big(\mathbf{X}_\sharp^\top \widetilde{\mathbf{U}}\widetilde{\mathbf{\Sigma}}\widetilde{\mathbf{V}}^\top \mathbf{Y}_\sharp\big) + \operatorname{TR}\big(\mathbf{X}_\sharp^\top \widetilde{\mathbf{U}}\widetilde{\mathbf{\Sigma}}\big(\mathbf{C}_\sharp - \widehat{\mathbf{C}}\big)\big) \\ &= \operatorname{TR}\big(\mathbf{X}_\sharp^\top \widetilde{\mathbf{A}}\mathbf{Y}_\sharp\big) + \operatorname{TR}\big(\mathbf{X}_\sharp^\top \widetilde{\mathbf{U}}\widetilde{\mathbf{\Sigma}}\big(\mathbf{C}_\sharp - \widehat{\mathbf{C}}\big)\big). \end{aligned} \tag{17}$$

The inequality follows from the fact that (by definition (15)) $\mathbf{Y}_\sharp$ maximizes the first term over all $\mathbf{Y} \in \mathcal{Y}$. Combining (17) and (16), and rearranging the terms, we obtain

$$\begin{aligned}&\operatorname{TR}\big(\widetilde{\mathbf{X}}_\star^\top \widetilde{\mathbf{A}}\widetilde{\mathbf{Y}}_\star\big) - \operatorname{TR}\big(\mathbf{X}_\sharp^\top \widetilde{\mathbf{A}}\mathbf{Y}_\sharp\big) \\ &\leq \operatorname{TR}\big(\widetilde{\mathbf{X}}_\star^\top \widetilde{\mathbf{U}}\widetilde{\mathbf{\Sigma}}\big(\widetilde{\mathbf{C}}_\star - \mathbf{C}_\sharp\big)\big) + \operatorname{TR}\big(\mathbf{X}_\sharp^\top \widetilde{\mathbf{U}}\widetilde{\mathbf{\Sigma}}\big(\mathbf{C}_\sharp - \widehat{\mathbf{C}}\big)\big). \end{aligned} \tag{18}$$

By Lemma C.10,

$$\begin{aligned}&\big|\operatorname{TR}\big(\widetilde{\mathbf{X}}_\star^\top \widetilde{\mathbf{U}}\widetilde{\mathbf{\Sigma}}\big(\widetilde{\mathbf{C}}_\star - \mathbf{C}_\sharp\big)\big)\big| \\ &\qquad\leq \|\widetilde{\mathbf{X}}_\star^\top \widetilde{\mathbf{U}}\|_\mathrm{F}\cdot\|\widetilde{\mathbf{\Sigma}}\|_2\cdot\|\widetilde{\mathbf{C}}_\star - \mathbf{C}_\sharp\|_\mathrm{F} \\ &\qquad\leq \|\widetilde{\mathbf{X}}_\star\|_\mathrm{F}\cdot\sigma_1(\widetilde{\mathbf{A}})\cdot\sqrt{k}\cdot\epsilon \\ &\qquad\leq \max_{\mathbf{X}\in\mathcal{X}} \|\mathbf{X}\|_\mathrm{F}\cdot\sigma_1(\widetilde{\mathbf{A}})\cdot\sqrt{k}\cdot\epsilon \\ &\qquad\leq \sigma_1(\widetilde{\mathbf{A}})\cdot\sqrt{k}\cdot\epsilon. \end{aligned} \tag{19}$$

Similarly,

$$\begin{aligned}\left|\mathrm{Tr}\left(\mathbf{X}_\sharp^\top \widetilde{\mathbf{U}}\widetilde{\mathbf{\Sigma}}(\mathbf{C}_\sharp - \widehat{\mathbf{C}})\right)\right| &\le \|\mathbf{X}_\sharp \widetilde{\mathbf{U}}\|_\mathrm{F} \cdot \|\widetilde{\mathbf{\Sigma}}\|_2 \cdot \|\mathbf{C}_\sharp - \widehat{\mathbf{C}}\|_\mathrm{F} \\ &\le \max_{\mathbf{X}\in\mathcal{X}} \|\mathbf{X}\|_\mathrm{F} \cdot \sigma_1(\widetilde{\mathbf{A}}) \cdot \sqrt{k} \cdot \epsilon \\ &\le \sigma_1(\widetilde{\mathbf{A}}) \cdot \sqrt{k} \cdot \epsilon. \end{aligned} \tag{20}$$

The second inequality follows from the fact that by the definition of $\widehat{\mathbf{C}}$,

$$\begin{aligned}\|\widehat{\mathbf{C}} - \mathbf{C}_\sharp\|_{\infty,2} &= \|\widetilde{V}^\top \widehat{\mathbf{Y}} - \mathbf{C}_\sharp\|_{\infty,2} \le \|\widetilde{V}^\top \widetilde{\mathbf{Y}}_\star - \mathbf{C}_\sharp\|_{\infty,2} \\ &= \|\widetilde{\mathbf{C}}_\star - \mathbf{C}_\sharp\|_{\infty,2} \le \epsilon,\end{aligned}$$

which implies that

$$\|\widehat{\mathbf{C}} - \mathbf{C}_\sharp\|_\mathrm{F} \le \sqrt{k} \cdot \epsilon.$$

Continuing from (18) under (19) and (20),

$$\mathrm{Tr}\left(\mathbf{X}_\sharp^\top \widetilde{\mathbf{A}}\mathbf{Y}_\sharp\right) \ge \mathrm{Tr}\left(\widetilde{\mathbf{X}}_\star^\top \widetilde{\mathbf{A}}\widetilde{\mathbf{Y}}_\star\right) - 2 \cdot \epsilon \cdot \sqrt{k} \cdot \sigma_1(\widetilde{\mathbf{A}}).$$

Recalling that the singular values of $\widetilde{\mathbf{A}}$ have been scaled by a factor of $\mu_\mathcal{X} \cdot \mu_\mathcal{Y}$ yields the desired result.

The runtime of Alg. 1 follows from the cost per iteration and the cardinality of the ϵ-net. Matrix multiplications can exploit the truncated singular value decomposition of $\widetilde{\mathbf{A}}$ which is performed only once. ☐

Lemma A.6. *For any $\mathbf{A}, \widetilde{\mathbf{A}} \in \mathbb{R}^{m\times n}$, and norm-boudned sets $\mathcal{X} \subseteq \mathbb{R}^{m\times k}$ and $\mathcal{Y} \subseteq \mathbb{R}^{n\times k}$, let*

$$(\mathbf{X}_\star, \mathbf{Y}_\star) \triangleq \underset{\mathbf{X}\in\mathcal{X},\mathbf{Y}\in\mathcal{Y}}{\arg\max}\ \mathrm{Tr}\left(\mathbf{X}^\top \mathbf{A}\mathbf{Y}\right),$$

and

$$(\widetilde{\mathbf{X}}_\star, \widetilde{\mathbf{Y}}_\star) \triangleq \underset{\mathbf{X}\in\mathcal{X},\mathbf{Y}\in\mathcal{Y}}{\arg\max}\ \mathrm{Tr}\left(\mathbf{X}^\top \widetilde{\mathbf{A}}\mathbf{Y}\right).$$

For any $(\widetilde{\mathbf{X}}, \widetilde{\mathbf{Y}}) \in \mathcal{X} \times \mathcal{Y}$ such that

$$\mathrm{Tr}\left(\widetilde{\mathbf{X}}^\top \widetilde{\mathbf{A}}\widetilde{\mathbf{Y}}\right) \ge \gamma \cdot \mathrm{Tr}\left(\widetilde{\mathbf{X}}_\star^\top \widetilde{\mathbf{A}}\widetilde{\mathbf{Y}}_\star\right) - C$$

for some $0 < \gamma \le 1$, we have

$$\begin{aligned}\mathrm{Tr}\left(\widetilde{\mathbf{X}}^\top \mathbf{A}\widetilde{\mathbf{Y}}\right) &\ge \gamma \cdot \mathrm{Tr}\left(\mathbf{X}_\star^\top \mathbf{A}\mathbf{Y}_\star\right) - C \\ &\quad - 2 \cdot \|\mathbf{A} - \widetilde{\mathbf{A}}\|_2 \cdot \mu_\mathcal{X} \cdot \mu_\mathcal{Y}.\end{aligned}$$

where $\mu_\mathcal{X} \triangleq \max_{\mathbf{X}\in\mathcal{X}} \|\mathbf{X}\|_\mathrm{F}$ and $\mu_\mathcal{Y} \triangleq \max_{\mathbf{Y}\in\mathcal{Y}} \|\mathbf{Y}\|_\mathrm{F}$.

Proof. By the optimality of $\widetilde{\mathbf{X}}_\star, \widetilde{\mathbf{Y}}_\star$ for $\widetilde{\mathbf{A}}$, we have

$$\mathrm{Tr}\left(\widetilde{\mathbf{X}}_\star^\top \widetilde{\mathbf{A}}\widetilde{\mathbf{Y}}_\star\right) \ge \mathrm{Tr}\left(\mathbf{X}_\star^\top \widetilde{\mathbf{A}}\mathbf{Y}_\star\right).$$

In turn, for any $(\widetilde{\mathbf{X}}, \widetilde{\mathbf{Y}}) \in \mathcal{X} \times \mathcal{Y}$ such that

$$\mathrm{Tr}\left(\widetilde{\mathbf{X}}^\top \widetilde{\mathbf{A}}\widetilde{\mathbf{Y}}\right) \ge \gamma \cdot \mathrm{Tr}\left(\widetilde{\mathbf{X}}_\star^\top \widetilde{\mathbf{A}}\widetilde{\mathbf{Y}}_\star\right) - C$$

for some $0 < \gamma < 1$ (if such pairs exist), we have

$$\mathrm{Tr}\left(\widetilde{\mathbf{X}}^\top \widetilde{\mathbf{A}}\widetilde{\mathbf{Y}}\right) \ge \gamma \cdot \mathrm{Tr}\left(\mathbf{X}_\star^\top \widetilde{\mathbf{A}}\mathbf{Y}_\star\right) - C. \tag{21}$$

By the linearity of the trace operator,

$$\begin{aligned}&\mathrm{Tr}\left(\widetilde{\mathbf{X}}^\top \widetilde{\mathbf{A}}\widetilde{\mathbf{Y}}\right) \\ &\quad= \mathrm{Tr}\left(\widetilde{\mathbf{X}}^\top \mathbf{A}\widetilde{\mathbf{Y}}\right) - \mathrm{Tr}\left(\widetilde{\mathbf{X}}^\top (\mathbf{A} - \widetilde{\mathbf{A}})\widetilde{\mathbf{Y}}\right) \\ &\quad\le \mathrm{Tr}\left(\widetilde{\mathbf{X}}^\top \mathbf{A}\widetilde{\mathbf{Y}}\right) + \left|\mathrm{Tr}\left(\widetilde{\mathbf{X}}^\top (\mathbf{A} - \widetilde{\mathbf{A}})\widetilde{\mathbf{Y}}\right)\right|.\end{aligned} \tag{22}$$

By Lemma C.10,

$$\begin{aligned}&\left|\mathrm{Tr}\left(\widetilde{\mathbf{X}}^\top (\mathbf{A} - \widetilde{\mathbf{A}})\widetilde{\mathbf{Y}}\right)\right| \\ &\le \|\widetilde{\mathbf{X}}\|_\mathrm{F} \cdot \|\widetilde{\mathbf{Y}}\|_\mathrm{F} \cdot \|\mathbf{A} - \widetilde{\mathbf{A}}\|_2 \\ &\le \|\mathbf{A} - \widetilde{\mathbf{A}}\|_2 \cdot \max_{\mathbf{X}\in\mathcal{X}} \|\mathbf{X}\|_\mathrm{F} \cdot \max_{\mathbf{Y}\in\mathcal{Y}} \|\mathbf{Y}\|_\mathrm{F} \triangleq R.\end{aligned} \tag{23}$$

Continuing from (22),

$$\mathrm{Tr}\left(\widetilde{\mathbf{X}}^\top \widetilde{\mathbf{A}}\widetilde{\mathbf{Y}}\right) \le \mathrm{Tr}\left(\widetilde{\mathbf{X}}^\top \mathbf{A}\widetilde{\mathbf{Y}}\right) + R. \tag{24}$$

Similarly,

$$\begin{aligned}&\mathrm{Tr}\left(\mathbf{X}_\star^\top \widetilde{\mathbf{A}}\mathbf{Y}_\star\right) \\ &\quad= \mathrm{Tr}\left(\mathbf{X}_\star^\top \mathbf{A}\mathbf{Y}_\star\right) - \mathrm{Tr}\left(\mathbf{X}_\star^\top (\mathbf{A} - \widetilde{\mathbf{A}})\mathbf{Y}_\star\right) \\ &\quad\ge \mathrm{Tr}\left(\mathbf{X}_\star^\top \mathbf{A}\mathbf{Y}_\star\right) - \left|\mathrm{Tr}\left(\mathbf{X}_\star^\top (\mathbf{A} - \widetilde{\mathbf{A}})\mathbf{Y}_\star\right)\right| \\ &\quad\ge \mathrm{Tr}\left(\mathbf{X}_\star^\top \mathbf{A}\mathbf{Y}_\star\right) - R.\end{aligned} \tag{25}$$

Combining the above, we have

$$\begin{aligned}\mathrm{Tr}\left(\widetilde{\mathbf{X}}^\top \mathbf{A}\widetilde{\mathbf{Y}}\right) &\ge \mathrm{Tr}\left(\widetilde{\mathbf{X}}^\top \widetilde{\mathbf{A}}\widetilde{\mathbf{Y}}\right) - R \\ &\ge \gamma \cdot \mathrm{Tr}\left(\mathbf{X}_\star^\top \widetilde{\mathbf{A}}\mathbf{Y}_\star\right) - R - C \\ &\ge \gamma \cdot \left(\mathrm{Tr}\left(\mathbf{X}_\star^\top \mathbf{A}\mathbf{Y}_\star\right) - R\right) - R - C \\ &= \gamma \cdot \mathrm{Tr}\left(\mathbf{X}_\star^\top \mathbf{A}\mathbf{Y}_\star\right) - (1+\gamma) \cdot R - C \\ &\ge \gamma \cdot \mathrm{Tr}\left(\mathbf{X}_\star^\top \mathbf{A}\mathbf{Y}_\star\right) - 2 \cdot R - C,\end{aligned}$$

where the first inequality follows from (24) the second from (21), the third from (25), and the last from the fact that $R \ge 0$. This concludes the proof. ☐

Lemma 3.3. *For any $\mathbf{A} \in \mathbb{R}^{m\times n}$, let*

$$(\mathbf{X}_\star, \mathbf{Y}_\star) \triangleq \underset{\mathbf{X}\in\mathcal{X},\mathbf{Y}\in\mathcal{Y}}{\arg\max}\ \mathrm{Tr}\left(\mathbf{X}^\top \mathbf{A}\mathbf{Y}\right),$$

where $\mathcal{X} \subseteq \mathbb{R}^{m\times k}$ and $\mathcal{Y} \subseteq \mathbb{R}^{n\times k}$ are sets satisfying the conditions of Lemma 3.2. Let $\widetilde{\mathbf{A}}$ be a rank-r approximation of $\mathbf{A}$, and $\widetilde{\mathbf{X}} \in \mathcal{X}$, $\widetilde{\mathbf{Y}} \in \mathcal{Y}$ be the output of Alg. 1 with input $\widetilde{\mathbf{A}}$ and accuracy ϵ. Then,

$$\begin{aligned}&\mathrm{Tr}\left(\mathbf{X}_\star^\top \mathbf{A}\mathbf{Y}_\star\right) - \mathrm{Tr}\left(\widetilde{\mathbf{X}}^\top \mathbf{A}\widetilde{\mathbf{Y}}\right) \\ &\quad\le 2 \cdot \left(\epsilon\sqrt{k} \cdot \|\widetilde{\mathbf{A}}\|_2 + \|\mathbf{A} - \widetilde{\mathbf{A}}\|_2\right) \cdot \mu_\mathcal{X} \cdot \mu_\mathcal{Y},\end{aligned}$$

where $\mu_\mathcal{X} \triangleq \max_{\mathbf{X}\in\mathcal{X}} \|\mathbf{X}\|_\mathrm{F}$ and $\mu_\mathcal{Y} \triangleq \max_{\mathbf{Y}\in\mathcal{Y}} \|\mathbf{Y}\|_\mathrm{F}$.

Proof. The proof follows the approximation guarantees of Alg. 1 in Lemma 3.2 and Lemma A.6. ☐

B Correctness of Algorithm 2

In the sequel, we use $\|\mathbf{X}\|_{\infty,1}$ to denote the maximum of the ℓ_1 norm of the rows of $\mathbf{X}$. When $\mathbf{X} \in \{0,1\}^{d\times k}$, the constraint $\|\mathbf{X}\|_{\infty,1} = 1$ effectively implies that each row of $\mathbf{X}$ has exactly one nonzero entry.

Lemma 4.4. *Let* $\mathcal{X} \triangleq \{\mathbf{X} \in \{0,1\}^{d\times k} : \|\mathbf{X}\|_{\infty,1} = 1\}$. *For any $d \times k$ real matrix $\mathbf{L}$, Algorithm 2 outputs*

$$\widetilde{\mathbf{X}} = \arg\max_{\mathbf{X}\in\mathcal{X}} \mathrm{TR}\big(\mathbf{X}^\top \mathbf{L}\big),$$

in time $O(k \cdot d)$

Proof. By construction, each row of $\mathbf{X}$ has exactly one nonzero entry. Let $j_i \in [k]$ denote the index of the nonzero entry in the ith row of $\mathbf{X}$. For any $\mathbf{X} \in \mathcal{X}$,

$$\begin{aligned}\mathrm{TR}\big(\mathbf{X}^\top \mathbf{L}\big) &= \sum_{j=1}^{k} \mathbf{x}_j^\top \mathbf{l}_j = \sum_{j=1}^{k} \sum_{i\in\mathrm{supp}(\mathbf{x}_j)} 1 \cdot L_{ij} \\ &= \sum_{i=1}^{d} L_{ij_i} \le \sum_{i=1}^{d} \max_{j\in[k]} L_{ij}. \qquad (26)\end{aligned}$$

Algorithm 2 achieves equality in (26) due to the choice of j_i in line 3. Finally, the running time follows immediately from the $O(k)$ time required to determine the maximum entry of each of the d rows of $\mathbf{L}$. □

C Auxiliary Lemmas

Lemma C.7. *Let $a_1, \ldots, a_n$ and $b_1, \ldots, b_n$ be $2n$ real numbers and let p and q be two numbers such that $1/p + 1/q = 1$ and $p > 1$. We have*

$$\left|\sum_{i=1}^{n} a_i b_i\right| \le \left(\sum_{i=1}^{n} |a_i|^p\right)^{1/p} \cdot \left(\sum_{i=1}^{n} |b_i|^q\right)^{1/q}.$$

Lemma C.8. *For any* $\mathbf{A}, \mathbf{B} \in \mathbb{R}^{n\times k}$,

$$\big|\langle \mathbf{A}, \mathbf{B}\rangle\big| \triangleq \big|\mathrm{TR}\big(\mathbf{A}^\top \mathbf{B}\big)\big| \le \|\mathbf{A}\|_{\mathrm{F}} \|\mathbf{B}\|_{\mathrm{F}}.$$

Proof. Treating $\mathbf{A}$ and $\mathbf{B}$ as vectors, the lemma follows immediately from Lemma C.7 for $p = q = 2$. □

Lemma C.9. *For any two real matrices $\mathbf{A}$ and $\mathbf{B}$ of appropriate dimensions,*

$$\|\mathbf{A}\mathbf{B}\|_{\mathrm{F}} \le \min\big\{\|\mathbf{A}\|_2 \|\mathbf{B}\|_{\mathrm{F}},\ \|\mathbf{A}\|_{\mathrm{F}} \|\mathbf{B}\|_2\big\}.$$

Proof. Let $\mathbf{b}_i$ denote the ith column of $\mathbf{B}$. Then,

$$\begin{aligned}\|\mathbf{A}\mathbf{B}\|_{\mathrm{F}}^2 &= \sum_i \|\mathbf{A}\mathbf{b}_i\|_2^2 \le \sum_i \|\mathbf{A}\|_2^2 \|\mathbf{b}_i\|_2^2 \\ &= \|\mathbf{A}\|_2^2 \sum_i \|\mathbf{b}_i\|_2^2 = \|\mathbf{A}\|_2^2 \|\mathbf{B}\|_{\mathrm{F}}^2.\end{aligned}$$

Similarly, using the previous inequality,

$$\|\mathbf{A}\mathbf{B}\|_{\mathrm{F}}^2 = \|\mathbf{B}^\top \mathbf{A}^\top\|_{\mathrm{F}}^2 \le \|\mathbf{B}^\top\|_2^2 \|\mathbf{A}^\top\|_{\mathrm{F}}^2 = \|\mathbf{B}\|_2^2 \|\mathbf{A}\|_{\mathrm{F}}^2.$$

The desired result follows combining the two upper bounds. □

Lemma C.10. *For any real $m \times k$ matrix $\mathbf{X}$, $m \times n$ matrix $\mathbf{A}$, and $n \times k$ matrix $\mathbf{Y}$,*

$$\big|\mathrm{TR}\big(\mathbf{X}^\top \mathbf{A}\mathbf{Y}\big)\big| \le \|\mathbf{X}\|_{\mathrm{F}} \cdot \|\mathbf{A}\|_2 \cdot \|\mathbf{Y}\|_{\mathrm{F}}.$$

Proof. We have

$$\big|\mathrm{TR}\big(\mathbf{X}^\top \mathbf{A}\mathbf{Y}\big)\big| \le \|\mathbf{X}\|_{\mathrm{F}} \cdot \|\mathbf{A}\mathbf{Y}\|_{\mathrm{F}} \le \|\mathbf{X}\|_{\mathrm{F}} \cdot \|\mathbf{A}\|_2 \cdot \|\mathbf{Y}\|_{\mathrm{F}},$$

with the first inequality following from Lemma C.8 on $|\langle \mathbf{X},\, \mathbf{A}\mathbf{Y}\rangle|$ and the second from Lemma C.9. □

Lemma C.11. *For any real $m \times n$ matrix $\mathbf{A}$, and pair of $m \times k$ matrix $\mathbf{X}$ and $n \times k$ matrix $\mathbf{Y}$ such that $\mathbf{X}^\top \mathbf{X} = \mathbf{I}_k$ and $\mathbf{Y}^\top \mathbf{Y} = \mathbf{I}_k$ with $k \le \min\{m,\ n\}$, the following holds:*

$$\big|\mathrm{TR}\big(\mathbf{X}^\top \mathbf{A}\mathbf{Y}\big)\big| \le \sqrt{k} \cdot \Big(\sum_{i=1}^{k} \sigma_i^2(\mathbf{A})\Big)^{1/2}.$$

Proof. By Lemma C.8,

$$\begin{aligned}|\langle \mathbf{X},\, \mathbf{A}\mathbf{Y}\rangle| &= \big|\mathrm{TR}\big(\mathbf{X}^\top \mathbf{A}\mathbf{Y}\big)\big| \\ &\le \|\mathbf{X}\|_{\mathrm{F}} \cdot \|\mathbf{A}\mathbf{Y}\|_{\mathrm{F}} = \sqrt{k} \cdot \|\mathbf{A}\mathbf{Y}\|_{\mathrm{F}}.\end{aligned}$$

where the last inequality follows from the fact that $\|\mathbf{X}\|_{\mathrm{F}}^2 = \mathrm{TR}\big(\mathbf{X}^\top \mathbf{X}\big) = \mathrm{TR}\big(\mathbf{I}_k\big) = k$. Further, for any $\mathbf{Y}$ such that $\mathbf{Y}^T \mathbf{Y} = \mathbf{I}_k$,

$$\|\mathbf{A}\mathbf{Y}\|_{\mathrm{F}}^2 \le \max_{\substack{\widehat{\mathbf{Y}}\in\mathbb{R}^{n\times k} \\ \widehat{\mathbf{Y}}^\top \widehat{\mathbf{Y}} = \mathbf{I}_k}} \|\mathbf{A}\widehat{\mathbf{Y}}\|_{\mathrm{F}}^2 = \sum_{i=1}^{k} \sigma_i^2(\mathbf{A}). \qquad (27)$$

Combining the two inequalities, the result follows. □

Lemma C.12. *For any real $m \times n$ matrix $\mathbf{A}$, and any $k \le \min\{m,\ n\}$,*

$$\max_{\substack{\mathbf{Y}\in\mathbb{R}^{n\times k} \\ \mathbf{Y}^\top \mathbf{Y} = \mathbf{I}_k}} \|\mathbf{A}\mathbf{Y}\|_{\mathrm{F}} = \left(\sum_{i=1}^{k} \sigma_i^2(\mathbf{A})\right)^{1/2}.$$

The above equality is realized when the k columns of $\mathbf{Y}$ coincide with the k leading right singular vectors of $\mathbf{A}$.

Proof. Let $\mathbf{U}\mathbf{\Sigma}\mathbf{V}^\top$ be the singular value decomposition of $\mathbf{A}$, with $\Sigma_{jj} = \sigma_j$ being the jth largest singular value of $\mathbf{A}$, $j = 1, \ldots, d$, where $d \triangleq \min\{m, n\}$. Due to the invariance of the Frobenius norm under unitary multiplication,

$$\|\mathbf{A}\mathbf{Y}\|_{\mathrm{F}}^2 = \|\mathbf{U}\mathbf{\Sigma}\mathbf{V}^\top\mathbf{Y}\|_{\mathrm{F}}^2 = \|\mathbf{\Sigma}\mathbf{V}^\top\mathbf{Y}\|_{\mathrm{F}}^2. \tag{28}$$

Continuing from (28),

$$\begin{aligned}\|\mathbf{\Sigma}\mathbf{V}^\top\mathbf{Y}\|_{\mathrm{F}}^2 &= \mathrm{T}\mathrm{R}\big(\mathbf{Y}^\top\mathbf{V}\mathbf{\Sigma}^2\mathbf{V}^\top\mathbf{Y}\big) \\ &= \sum_{i=1}^{k} \mathbf{y}_i^\top \left(\sum_{j=1}^{d} \sigma_j^2 \cdot \mathbf{v}_j\mathbf{v}_j^\top\right) \mathbf{y}_i \\ &= \sum_{j=1}^{d} \sigma_j^2 \cdot \sum_{i=1}^{k} \left(\mathbf{v}_j^\top \mathbf{y}_i\right)^2.\end{aligned}$$

Let $z_j \triangleq \sum_{i=1}^{k} \big(\mathbf{v}_j^\top\mathbf{y}_i\big)^2$, $j = 1, \ldots, d$. Note that each individual z_j satisfies

$$0 \leq z_j \triangleq \sum_{i=1}^{k} \big(\mathbf{v}_j^\top\mathbf{y}_i\big)^2 \leq \|\mathbf{v}_j\|^2 = 1,$$

where the last inequality follows from the fact that the columns of $\mathbf{Y}$ are orthonormal. Further,

$$\begin{aligned}\sum_{j=1}^{d} z_j &= \sum_{j=1}^{d}\sum_{i=1}^{k}\big(\mathbf{v}_j^\top\mathbf{y}_i\big)^2 = \sum_{i=1}^{k}\sum_{j=1}^{d}\big(\mathbf{v}_j^\top\mathbf{y}_i\big)^2 \\ &= \sum_{i=1}^{k} \|\mathbf{y}_i\|^2 = k.\end{aligned}$$

Combining the above, we conclude that

$$\|\mathbf{A}\mathbf{Y}\|_{\mathrm{F}}^2 = \sum_{j=1}^{d} \sigma_j^2 \cdot z_j \leq \sigma_1^2 + \ldots + \sigma_k^2. \tag{29}$$

Finally, it is straightforward to verify that if $\mathbf{y}_i = \mathbf{v}_i$, $i = 1, \ldots, k$, then (29) holds with equality. □

Lemma C.13. *For any real $m \times n$ matrix $\mathbf{A}$, let $\sigma_i(\mathbf{A})$ be the ith largest singular value. For any $r, k \leq \min\{m, n\}$,*

$$\sum_{i=r+1}^{r+k} \sigma_i(\mathbf{A}) \leq \frac{k}{\sqrt{r+k}} \|\mathbf{A}\|_{\mathrm{F}}.$$

Proof. By the Cauchy-Schwartz inequality,

$$\begin{aligned}\sum_{i=r+1}^{r+k} \sigma_i(\mathbf{A}) &= \sum_{i=r+1}^{r+k} |\sigma_i(\mathbf{A})| \leq \left(\sum_{i=r+1}^{r+k} \sigma_i^2(\mathbf{A})\right)^{1/2} \|\mathbf{1}_k\|_2 \\ &= \sqrt{k} \cdot \left(\sum_{i=r+1}^{r+k} \sigma_i^2(\mathbf{A})\right)^{1/2}.\end{aligned}$$

Note that $\sigma_{r+1}(\mathbf{A}), \ldots, \sigma_{r+k}(\mathbf{A})$ are the k smallest among the $r + k$ largest singular values. Hence,

$$\begin{aligned}\sum_{i=r+1}^{r+k} \sigma_i^2(\mathbf{A}) &\leq \frac{k}{r+k} \sum_{i=1}^{r+k} \sigma_i^2(\mathbf{A}) \leq \frac{k}{r+k} \sum_{i=1}^{l} \sigma_i^2(\mathbf{A}) \\ &= \frac{k}{r+k} \|\mathbf{A}\|_{\mathrm{F}}^2.\end{aligned}$$

Combining the two inequalities, the desired result follows. □

Corollary 1. *For any real $m \times n$ matrix $\mathbf{A}$, the rth largest singular value $\sigma_{\mathrm{r}}(\mathbf{A})$ satisfies $\sigma_{\mathrm{r}}(\mathbf{A}) \leq \|\mathbf{A}\|_{\mathrm{F}}/\sqrt{r}$.*

Proof. It follows immediately from Lemma C.13. □

First, we define the $\|\cdot\|_{\infty,2}$ norm of a matrix as the l_2 norm of the column with the maximum l_2 norm, *i.e.*, for an $r \times k$ matrix $\mathbf{C}$

$$\|\mathbf{C}\|_{\infty,2} = \max_{1 \leq i \leq k} \|\mathbf{c}_i\|_2.$$

Note that

$$\begin{aligned}\|\mathbf{C}\|_{\mathrm{F}}^2 &= \sum_{i=1}^{k} \|\mathbf{c}_i\|_2^2 \leq k \cdot \max_{1 \leq i \leq k} \|\mathbf{c}_i\|_2^2 \\ &= k \cdot \left(\max_{1 \leq i \leq k} \|\mathbf{c}_i\|_2\right)^2 = k \cdot \|\mathbf{C}\|_{\infty,2}.\end{aligned} \tag{30}$$