CS698C: Sketching and Sampling for Big Data Analysis

Affine and Subspace Embeddings *Exercise Set 7*

Note:

1. *All vector norms are ℓ_2 norms $\|\cdot\|_2$ unless explicitly specified.*
2. *Matrix norms are Frobenius norms $\|\cdot\|_F$. If matrix 2-norms are used, it will be explicitly specified.*

Review of Linear Regression. Let $A \in \mathbb{R}^{m\times n}$ and $b \in \mathbb{R}^m$. The problem of linear regression is $\min_x \|Ax - b\|_2$. We will assume that $A = U\Sigma V^T$ is the SVD of A, where, Σ is the $m \times n$ matrix with diagonal singular values entries given below. The pseudo inverse is defined as $A^- = V\Sigma^- U^T$, where, Σ^- is the $n \times m$ matrix with diagonal inverse singular values. The notation below assumes that the rank of A equals r.

$$\Sigma = \begin{bmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_r \\ & & \end{bmatrix}, \qquad \Sigma^- = \begin{bmatrix} \sigma_1^{-1} & & \\ & \ddots & \\ & & \sigma_r^{-1} \\ & & \end{bmatrix} .$$

The matrices Σ and Σ^- only have diagonal elements as shown above–all other entries are zeros. It is generally assumed that the left and right singular vectors of U and V are ordered such that

$$\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > 0 \ .$$

Since the rank of A is assumed to be r, let $U = \begin{bmatrix} u_1, & u_2, & \ldots, & u_m \end{bmatrix}$ and $V = \begin{bmatrix} v_1, & v_2, & \ldots, & v_n \end{bmatrix}$ be the orthogonal matrices corresponding to the left singular vectors and the right singular vectors. Since the rank of A is r, the singular vector matrices U and V are summarized as follows:

$$\begin{aligned} U &= \begin{bmatrix} U_r & U_r^\perp \end{bmatrix}, & U_r &= \begin{bmatrix} u_1, & \ldots, & u_r \end{bmatrix}, & U_r^\perp &= \begin{bmatrix} u_{r+1}, & \cdots & u_m \end{bmatrix} \ . \\ V &= \begin{bmatrix} V_r & V_r^\perp \end{bmatrix}, & V_r &= \begin{bmatrix} v_1, & \ldots, & v_r \end{bmatrix}, & V_r^\perp &= \begin{bmatrix} v_{r+1}, & \cdots & v_n \end{bmatrix} \ . \end{aligned}$$

Review of SVD.

1. Show that A has an equivalent SVD representation:
$$A = \sigma_1 u_1 v_1^T + \sigma_2 u_2 v_2^T + \cdots \sigma_r u_r v_r^T \ .$$
2. Show that the singular vector pairs satisfy the following properties
$$\begin{aligned} Av_j &= \sigma_j u_j, & j &= 1, 2, \ldots, r \ . \\ Av_j &= 0, & j &= r+1, \ldots, n \ . \\ u_j^T A &= \sigma_j v_j, & j &= 1, 2, \ldots, r \ . \\ u_j^T A &= 0, & j &= r+1, \ldots, m \ . \end{aligned}$$
Further, $v_{r+1}, \ldots, v_n$ forms an (arbitrary) orthonormal basis of the null space of A. Analogously, $u_{r+1}, \ldots, u_m$ forms an (arbitrary) orthonormal basis for the left nullspace of A.

3. *The relation between eigenvalues and singular values.*

   (a) Show that the eigenvalues of $A^T A = V\Sigma^T\Sigma V^T$ are the squares of the singular values of A. By the spectral theorem, conclude that the corresponding eigenvectors of $A^T A$ are the right singular vectors of A and form an orthogonal basis for $\mathbb{R}^n$. The first r eigenvectors correspond to the rowspace of A and the remaining r eigenvectors correspond to the nullspace of A.

   (b) Likewise show that the eigenvalues of $AA^T = U\Sigma\Sigma^T U^T$ are the squares of the singular values of A, and the corresponding eigenvectors are the left singular vectors of A thereby forming an orthogonal basis for $\mathbb{R}^m$. Analogously, argue regarding the column space of A and the left nullspace of A.

Review of Linear Regression.

1. From Euclidean geometry, a solution x^* attains the minimum value of $\min_x \|Ax - b\|_2 = Ax^* - b$ if and only if $Ax^* - b$ is orthogonal to the column space of A. Argue this with a figure.

2. From the above part, hence argue that any solution y that attains the minimum value of $\min_x \|Ax - b\|_2$ are solutions to the equation

$$A^T(Ax - b) = 0, \qquad \text{or,} \qquad A^T Ax = A^T b \ .$$

   Show vice-versa, namely that vectors y that do not satisfy the above equation do not attain the minimum value of $\min_x \|Ax - b\|$.

3. For the case when A has independent columns, show that the minimization problem $\min_x \|Ax - b\|$ has a unique solution given by

$$A^T(Ax - b) = 0, \qquad \text{which implies that } x_j^* = (A^T A)^{-1} A^T b \ .$$

   Hint: Derive the fact that $\text{rank}(A^T A) = \text{rank}(A) = n$ and hence invertible.

4. Hence show that when A has orthonormal columns,

$$x^* = A^T b \ .$$

5. Show that the column space projection matrix that projects any vector $b \in \mathbb{R}^m$ onto the column space of A is given by

$$U_r U_r^T = AA^- \ .$$

   The projected vector is given as $U_r U_r^T b = AA^- b$. Analogously, show that the row space projection matrix that projects any row vector c^T for $c \in \mathbb{R}^n$ onto the row space of A is given by

$$V_r V_r^T = A^- A \ .$$

   The projected vector is

$$(V_r V_r^T c)^T = c^T V_r V_r^T = c^T A^- A = (A^- Ac)^T \ .$$

6. Show that any $b \in \mathbb{R}^m$ can be written uniquely as a sum of orthogonal components comprised of projections onto the column space of A and on the left nullspace of A, namely,

$$b = AA^- b + (I - AA^-)b \ .$$

By Pythagoras' theorem,

$$\|b\|^2 = \|AA^- b\|^2 + \|(I - AA^-)b\|^2 \ .$$

7. Analogously, given any row vector c^T for $c \in \mathbb{R}^n$ can be written uniquely as the sum of its projections onto the row space of A and the nullspace of A, namely,

$$c^T = c^T A^- A + c^T (I - A^- A) \ .$$

This gives

$$\|c^T\|^2 = \|c\|^2 = \|c^T A^- A\|^2 + \|c^T (I - A^- A)\|^2 \ .$$

8. Extending the above from application of projection matrix AA^- to a vector $b \in \mathbb{R}^m$ to a matrix $B \in \mathbb{R}^{m \times p}$, show that the matrix B can be uniquely decomposed analogously as follows:

$$B = AA^- B + (I - AA^-)B$$

, where, each column b_j of B is written as the sum of orthogonal projections $AA^- b_j + (I - AA^-)b_j$, and therefore show that

$$\|B\|_F^2 = \|AA^- B\|_F^2 + \|(I - AA^-)B\|_F^2 \ .$$

9. Analogously, for any $C \in \mathbb{R}^{q \times m}$, show that

$$C = CA^- A + C(I - A^- A)$$

and

$$\|C\|_F^2 = \|CA^- A\|_F^2 + \|C(I - A^- A)\|_F^2 \ .$$

Review of affine embedding

A random matrix S drawn from a distribution $\mathcal{D}$ is said to be an (ϵ, δ) norm preserving affine embedding for a d-dimensional subspace defined by the column space of A, where, $A \in \mathbb{R}^{n \times d}$ and an $n \times p$ matrix $B \in \mathbb{R}^{n \times p}$ as the translation matrix, if for all $X \in \mathbb{R}^{n \times p}$

$$\mathrm{P}\Big[\text{ for all } X \in \mathbb{R}^{n \times p}, \|SAX - SB\|_F \in (1 \pm \epsilon)\|AX - B\|_F \Big] \geq 1 - \delta$$

where the probability is taken over S drawn from the distribution $\mathcal{D}$. In the following, the steps are reviewed and questions regarding inferences are posed. Denote B in column view as

$$B = \begin{bmatrix} b_1, & b_2, & \ldots, & b_p \end{bmatrix} \ .$$

1. Can we assume without loss of generality that A has orthonormal columns?

2. In all subsequent parts, let A have orthonormal columns. Let $X^* \in \mathbb{R}^{d\times p}$, where,

$$X^* = \begin{bmatrix} x_1^*, & x_2^*, & \ldots, & x_p^* \end{bmatrix}, \qquad\qquad x_j^* = A^- b_j, \quad j = 1, 2, \ldots, p$$

and so,

$$X^* = \begin{bmatrix} A^- b_1, A^- b_2, \ldots, A^- b_p \end{bmatrix} = A^- \begin{bmatrix} b_1, & b_2, & \ldots, b_p \end{bmatrix} = A^- B \ .$$

(a) Show that $\min_{x\in\mathbb{R}^d} \|Ax - b_j\|_2 = \|Ax_j^* - b_j\|_2$, for $j = 1, 2, \ldots, p$. In other words, show that x_j^* is an optimal solution to the linear regression problem: $\min_x \|Ax - b_j\|_2$.

(b) Hence, show that

$$\min_{X\in\mathbb{R}^{d\times p}} \|AX - B\|_F = \|AX^* - B\|_F \ .$$

(c) For any $j = 1, 2, \ldots, p$, show that $Ax_j^* - b_j$ is orthogonal to the column space of A. That is, $A^T(Ax_j^* - b_j) = 0$.

(d) Hence show the normal equation for linear regression,

$$\|Ax - b_j\|_2^2 = \|A(x - x_j^*)\|_2^2 + \|Ax_j^* - b_j\|_2^2 \ .$$

Hint: Expand $\|Ax - b_j\|_2^2 = \|A(x - x_j^*) + (Ax_j^* - b_j)\|_2^2 = \|A(x - x_j^*)\|_2^2 + \|Ax_j^* - b_j\|_2^2 + 2(A(x - x_j^*))^T(Ax_j^* - b_j)$, or argue by Pythagoras' theorem.

(e) Derive from the normal equation for linear regression the normal equation for generalized linear regression:

$$\|AX - B\|_F^2 = \|A(X - X^*)\|_F^2 + \|AX^* - B\|_F^2 \ .$$

(f) The following straightforward deductions are needed for matrix algebra with Frobenius norms. Let C and D be $m \times n$ matrices over reals. Show that,

$$\begin{aligned}\|C + D\|_F^2 &= \|C\|_F^2 + \|D\|_F^2 + 2\mathrm{tr}\ (C^T D) \\ &= \|C\|_F^2 + \|D\|_F^2 + 2\mathrm{tr}\ (CD^T) \ .\end{aligned}$$

(g) The following is a simple bound on trace of matrix products. Let C and D be real $m \times n$ matrices. Then,

$$\mathrm{tr}\ C^T D \le \|C\|_F \|D\|_F \ .$$

Hint: Use the famous Cauchy-Schwarz inequality, that for any t-dimensional vectors a and b, states that $\left|a^T b\right| \le \|a\|_2 \|b\|_2$.

3. As a first assumption, we assume that $S \sim \mathcal{D}$ is an $(\epsilon/5, \frac{\delta}{3})$ approximate norm preserving subspace embedding for the d-dimensional column space of A, that is,

$$\mathrm{P}\left[\text{ for all } x \in \mathbb{R}^n,\ \|SAx\| \in (1 \pm \epsilon/5)\|Ax\|\right] \ge 1 - \frac{\delta}{3} \qquad (A1)$$

for $\epsilon \le \frac{1}{2}$. where, $\epsilon < 1/4$.

(a) State and argue briefly the sufficient condition for the number of rows k in the distribution $\mathcal{D}$ over $k \times n$ matrices with i.i.d. normal distribution entries $N(0, 1/k)$.

(b) State and argue briefly the sufficient condition for the number of rows k in the class of random matrices over $\mathbb{R}^{k\times n}$ for the distribution $\mathcal{D}$ with Countsketch matrices.

4. (a) Denote $B^* = AX^* - B$. Then, show that

$$\begin{aligned}\|SAX - SB\|_F^2 &= \|SA(X - X)^*\|_F^2 + \|SB^*\|_F^2 + 2\mathrm{tr}(X - X^*)^T(SA)^T(SB^*)\\ &\leq \|SA(X - X)^*\|_F^2 + \|SB^*\|_F^2 + 2\|X - X^*\|_F\|(SA)^T SB^*\|_F \ .\end{aligned}$$

where the last step follows from the previous part.

(b) Make a second assumption that

$$\mathrm{P}\left[\|SB^*\|_F \in (1 \pm \epsilon/5)\|B\|_F\right] \leq \frac{\delta}{3} \qquad (A2) \ .$$

i. Assuming $\mathcal{D}$ is family of random $k \times n$ matrices over reals with i.i.d. normally distributed matrix entries $N(0, \frac{1}{k})$, what is the sufficient value of k to satisfy this assumption as a function of $\frac{\delta}{3}$? (*Hint:* Show that $k = O(\epsilon^{-2}\log(p/\frac{\delta}{3}))$ suffices).

ii. Assuming that $\mathcal{D}$ is a family of $k\times n$ CountSketch random matrices, find a sufficient value of k as a function of $\frac{\delta}{3}$. *Hint:* Show that $k = O(\epsilon^{-2}p/\frac{\delta}{3})$ suffices, for general $\frac{\delta}{3}$.

iii. *Can you show that if $\mathcal{D}$ is the family of $k \times n$ Countsketch random matrices, and $\frac{\delta}{3}$ is a constant (e.g., $\frac{\delta}{3} = 0.01$), then it suffices for k to be $k = O(\epsilon^{-2})$?

(c) Using assumptions $(A1)$ and $(A2)$ show that

$$\text{for all } X \in \mathbb{R}^{d\times p},\ \|SAX - SB\|_F^2 \in (1 \pm \epsilon)\|A(X - X^*)\|_F^2 + (1 \pm \epsilon)\|B^*\|_F^2 + 2\|X - X^*\|_F\|(SA)^T SB^*\|_F$$

holds with probability $1 - \frac{2\delta}{3}$ (using union bound to combine $A1$ and $A2$).

(d) Therefore show that it follows that, for all X,

$$\begin{aligned}&\big|\|SAX - SB\|_F^2 - \|AX - B\|_F^2\big|\\ &\qquad\leq \epsilon\left(\|A(X - X^*)\|_F^2 + \|B^*\|_F^2\right) + 2\|X - X^*\|_F\|(SA)^T SB^*\|_F \ .\end{aligned}$$

holds with probability $1 - 2\frac{\delta}{3}$.

5. Now the problem is to bound $\|(SA)^T SB^*\|_F$. Denote A in column view as $\begin{bmatrix}A_1, & A_2, & \cdots, & A_d\end{bmatrix}$.

(a) Show that $\|(SA)^T SB^*\|_F^2 = \sum_{i=1}^d \sum_{j=1}^p ((SA_i)^T(SB_j^*))^2$.

(b) Consider the problem of using union bound (Boole's inequality) to ensure that for each i, j in the ranges as above,

$$\begin{aligned}\mathrm{P}\Big[\text{ for all } i \in \{1, \dots, d\}, j \in \{1, \dots, p\},\ \ &\big|(SA_i)^T(SB_j^*) - A_i^T B_j^*\big| \leq \epsilon'\|A_i\|\|B_j^*\|\Big]\\ &\geq 1 - \frac{\delta}{3} \qquad (A3)\end{aligned}$$

Reviewing the discussion earlier, state and argue the conditions on the number of rows k of the distribution $\mathcal{D}$ over random matrices over $\mathbb{R}^{k\times n}$ so that the above property holds for each $i \in \{1, 2, \dots, d\}$ and $j \in \{1, 2, \dots, p\}$, jointly with probability $1 - \frac{\delta}{3}$.

- For normally distributed random matrices as discussed, show that for $(A3)$ to hold it suffices to have $k = O(\epsilon'^{-2} \log(dp/\delta))$.
- For Countsketch matrices, show that for $(A3)$ to hold, it suffices to have $k = O(\epsilon'^{-2} pd\delta^{-1})$ suffices.

(c) Since, $A^T B^* = 0_{d\times p}$, show that if the property mentioned in the part above holds, then,

$$\|(SA)^T SB^*\|_F = \epsilon' \|A\|_F \|B^*\|_F \ .$$

Hint: Since $A^T B^* = 0_{d\times p}$, the summation over the *LHS* simplifies to

$$\sum_{i,j} \left|(SA_i)^T(SB_j^*) - A_i^T B_j^*\right|^2 = \sum_{i,j}((SA_i)^T(SB_j^*)^2 = \|(SA)^T SB^*\|_F^2 \ .$$

The summation of the *RHS* simplifies to

$$\sum_{i,j} \epsilon'^2 \|A_i\|_2^2 \|B_j^*\|_2^2 = \epsilon'^2 \|A\|_F^2 \|B^*\|_F^2 \ .$$

This gives

$$\|(SA)^T SB^*\|_F^2 = \|(SA)^T SB^* - A^T B^*\|_F^2 = \sum_{i,j} \epsilon'^2 \|A_i\|_2^2 \|B_j^*\|_2^2 = \epsilon'^2 \|A\|_F^2 \|B^*\|_F^2$$

where, the second step above follows from approximate preservation of all inner-products $A_i^T B_j^*$, $i \in \{1, \dots, d\}, j \in \{1, \dots, p\}$.

6. Substituting the above part (c) in part (4) above, show that this implies the following.

$$\text{for all } X \in \mathbb{R}^{d\times p},\ \left|\|SAX - SB\|_F^2 - \|AX - B\|_F^2\right| \\ \leq \epsilon(\|A(X-X)^*\|_F^2 + \|B^*\|_F^2 + 2\epsilon' \|(X-X)^*\| \|A\|_F \|B\|_F$$

holds with probability $1-\delta$, by using union bound for the events $A1, A2$ and $A3$.

7. Since A has orthonormal columns, show that $\|A\|_F = \sqrt{d}$. Also show that $\|X - X^*\| = \|A(X - X^*)\|_F$. Hence, show that letting $\epsilon' = \frac{\epsilon}{\sqrt{d}}$ in the event $(A3)$,

$$\text{for all } X \in \mathbb{R}^{d\times p},\ \left|\|SAX - SB\|_F^2 - \|AX - B\|_F^2\right| \\ \leq \epsilon(\|A(X-X)^*\|_F^2 + \|B^*\|_F^2 + 2\epsilon \|A(X-X)^*\| \|B^*\|_F$$

holds with probability $1 - 3\frac{\delta}{3}$.

(a) Show that for random normal sketching matrices it suffices to have the number of rows $k = O\left(\frac{d}{\epsilon^2}\log\frac{1}{\delta} + \frac{1}{\epsilon^2}\log\frac{p}{\delta} + \frac{d}{\epsilon^2}\log\frac{pd}{\delta}\right)$.

(b) For random Countsketch family of matrices show that it suffices to have the number of rows to be $k = O\left(\frac{d^4}{\epsilon^2\delta} + \frac{p}{\epsilon^2\delta} + \frac{d^2 p}{\epsilon^2\delta}\right)$.

(c) Using $AM \geq GM$, it follows that $\|A(X - X^*)\|_F \|B^*\|_F \leq \frac{1}{2}\left(\|A(X - X^*)\|_F^2 + \|B^*\| + F^2\right)$. Hence, show that the following statement holds with probability $1-\delta$. For all $X \in \mathbb{R}^{d\times p}$,

$$\begin{aligned}\left|\|SAX - SB\|_F^2 - \|AX - B\|_F^2\right| &\leq 2\epsilon\left(\|A(X - X^*)\|_F^2 + \|B^*\|_F^2\right) \\ &= 2\epsilon\|AX - B\|_F^2\end{aligned}$$

with probability $1-\delta$.

A partially alternate proof procedure

Suppose we deviate from the statement in part (4) above and proceed alternatively as follows.

$$\left|\|SAX - SB\|_F^2 - \|AX - B\|_F^2\right| \le \epsilon \left(\|A(X - X^*)\|_F^2 + \|B^*\|_F^2\right) + 2\left|\mathrm{tr}(SA(X - X^*))^T SB^*\right| \ .$$

We would like to preserve the following property $(A4)$.

$$\forall x \in \mathbb{R}^d,\ \text{for } j = 1, 2 \ldots, p, \left|(SAx)^T (SB_j^*)\right| \le \epsilon \|Ax\| \|B_j^*\| \qquad (A4_\epsilon)$$

with probability $1 - \frac{\delta}{3}$. The following questions develop this approach.

1. Let M be a $\gamma = \frac{1}{2}$-net in the unit sphere $T = \{Ax \mid x \in \mathbb{R}^d, \|x\|_2 = 1\}$. Recall that for any vector $y \in T$,

$$y = y_0 + \alpha_1 y_1 + \alpha_2 y_2 + \cdots$$

   where, $y_0, y_1, \ldots \in M$ and $|\alpha_j| \le \gamma^j$, for each $j = 1, 2, \ldots$. Recall that

$$|M| \le \frac{(1 + \frac{\gamma}{2})^d}{\left(\frac{\gamma}{2}\right)^d} = \left(1 + \frac{2}{\gamma}\right)^d \ .$$

2. Define the following event E_ϵ.

$$\forall y \in M, j \in \{1, \ldots, d\}, \quad \left|(Sy)^T (SB_j^*)\right| \le \epsilon \|B_j^*\| \ .$$

   Show that if the event $E_{\epsilon/2}$ holds then event $A4_\epsilon$ holds.

   Suppose E_ϵ holds. Let $y \in T$. Then assuming E holds show that

$$\left|(Sy)^T (SB_j^*)\right| \le 2\epsilon \|B_j^*\| \ .$$

   Hint:

$$(Sy)^T (SB_j^*) = \left(\sum_i S(\alpha_i y_i\right)^T SB_j^* = \sum_i \alpha_i (Sy_i)^T (SB_j^*) \ .$$

   Taking absolute values on both sides,

$$\begin{aligned}
\left|(Sy)^T (SB_j^*)\right| &\le \sum_i |\alpha_i| \left|(Sy_i)^T (SB_j^*)\right| \\
&= \sum_i |\alpha_i| \, \epsilon \|y_i\| \|B_j^*\| \\
&= \sum_i |\alpha_i| \, \epsilon \|B_j^*\| \\
&\le \epsilon \|B_j^*\| \sum_i \gamma^i \\
&\le \frac{\epsilon}{1 - \gamma} \|B_j^*\| \\
&\le 2\epsilon \|B_j^*\|
\end{aligned}$$

   assuming $\gamma = 1/2$.

3. The question pertains to how many rows in the family of random matrices are needed to preserve $A1, A2$ and E. Show that if the distribution over random matrices is iid normal, then, it suffices to have $m = O\left(\frac{d+\log p}{\epsilon^2} \log \frac{1}{\delta}\right)$.

   Hint. Consider the classical random normally distributed family of random matrices. Then the following inequality holds for a fixed $y \in T$,

$$\left|(Sy)^T(SB_j^*)\right| \leq \epsilon \|B_j^*\|, \qquad \text{Event } E_{y,j} \ .$$

   provided the following three conditions hold, denoted G_y^1, G_j^2 and $G_{y,j}^3$.

$$\begin{aligned} &1. \|Sy\|_2 \in (1 \pm \frac{\epsilon}{3})\|y\|_2, && y \in M \ . \\ &2. \|SB_j^*\|_2 \in (1 \pm \frac{\epsilon}{3})\|B_j^*\|_2, && j = 1, 2, \ldots, p \ . \\ &3. \|S(y + B_j^*)\|_2 \in (1 \pm \frac{\epsilon}{3})\|y + B_j^*\|_2, && y \in M, j = 1, 2, \ldots, p \ . \end{aligned}$$

   By replacing ϵ by $\epsilon/3$ in Assumptions $A1$ and $A2$ satisfies G_y^1 and G_y^2 respectively. Condition $G_{y,j}^3$ has to be satisfied for all $y \in M$ and $j \in \{1, \ldots, p\}$. Under random iid normal distribution entries, this requires J-L application for all $y \in T, j \in \{1, \ldots, p\}$. There gives $|T|\, p = 5^d p$ applications of J-L Lemma. Preserving each of them with probability $\frac{\delta}{3 \cdot 5^d \cdot p}$ would satisfy $E_{\epsilon/2}$ with probability $1 - \delta/3$. Since, $A1$ requires less than 5^{2d} applications of JL lemma, $A2$ requires p applications and E requires $5^d p$ applications, using union bound,

$$\mathrm{P}\left[A1 \cap A2 \cap E\right] \geq 1 - (5^{2d} + p + 5^d p) \exp\left\{-\frac{\epsilon^2 m}{8}\right\} \geq 1 - \delta$$

   provided,

$$m \geq O\left(\frac{d}{\epsilon^2} \log \frac{1}{\delta} + \frac{1}{\epsilon^2} \log \frac{p}{\delta}\right) \ .$$

4. Assume that event $E_{\epsilon/2}$ holds. Show that

$$\left|\mathrm{tr}\ (SA(X - X^*))^T SB^*\right| \leq (\epsilon/2)\left[\|A(X - X^*)\|_F^2 + \|B^*\|_F^2\right] = (\epsilon/2)\|AX - B\|_F^2 \ .$$

   Hint:

$$\begin{aligned} \left|\mathrm{tr}\ (SA(X - X^*))^T SB^*\right| &= \left|\sum_{j=1}^{p} ((SA(X_j - X_j^*))^T SB_j^*\right| \\ &\leq \sum_{j=1}^{p} \left|((SA(X_j - X_j^*))^T SB_j^*\right| \\ &\leq \sum_j \epsilon \|A(X_j - X_j^*)\| \|B_j^*\|, && \text{since } E_{\epsilon/2} \text{ holds.} \\ &\leq \sum_j (\epsilon/2)\|A(X_j - X_j^*)\|^2 + \|B_j^*\|^2 && \text{AM} \geq \text{GM} \\ &= (\epsilon/2)\left[\|A(X - X^*)\|_F^2 + \|B^*\|_F^2\right] \\ &= (\epsilon/2)\|AX - B\|_F^2 \ . \end{aligned}$$

5. *Putting it together.* Assuming that events $A1_{\epsilon/3}, A2_{\epsilon/3}$ and $A4_\epsilon$ holds then show that, for any $X \in \mathbb{R}^{d\times p}$,

$$\left| \|SAX - SB\|_F^2 - \|AX - B\|_F^2 \right| \leq 2\epsilon \|AX - B\|_F^2 \ .$$

Moreover, for iid random normally distributed entries, it suffices for the number of rows in the random embedding matrix to be $O\left(\frac{d}{\epsilon^2}\log\frac{1}{\delta} + \frac{1}{\epsilon^2}\log\frac{p}{\delta}\right)$. *Note that this matches the bound on the number of rows found in the original proof procedure while avoiding obtaining the term* $\epsilon^{-2}d\log(d/\delta)$.

Approximate norm preserving affine embedding for $\min_X \|XB - C\|_F$

This briefly considers the minimization problem

$$\min_{X\in\mathbb{R}^{p\times d}} \|XB - C\|$$

where, $B \in \mathbb{R}^{d\times n}$ dimensional matrix and $C \in \mathbb{R}^{p\times n}$. We will generally assume that n is large and comparatively d is much smaller. An obvious way to solve this problem is to solve its equivalent transpose problem, namely,

$$\min_X \|B^T X^T - C^T\| \ .$$

The other approach is to solve it directly; both methods are virtually the same. We consider the direct approach. For any matrix A let A'_i denote the ith row vector of A.

The affine subspace embedding norm preservation property under the subspace embedding $R \sim \mathcal{D}$ is that

$$\mathrm{P}\left[\forall X \in \mathbb{R}^{d\times n},\ \|XBR - CR\|_F \in (1\pm\epsilon)\|XB - C\|_F\right] \geq 1 - \delta \ .$$

Writing $X^* = CB^-$, we have for each row index $i = 1, 2, \ldots, d$, that $X_i^{*\prime}B - C'_i$ is orthogonal to the row space of B, that is,

$$C'_i = X_i^{*\prime}B + C'_i - X_i^{*\prime}B = C'_iB^-B + C'_i(I - B^-B) \ .$$

If the SVD of B is $U\Sigma V^T$, then, $B^- = V\Sigma^- U^T$ and $BB^- = V_rV_r^T$, assuming that $\mathrm{rank}(B) = r$. Thus, $I - BB^- = I - V_rV_r^T$ is the projection matrix of a row vector onto the nullspace of B and $BB^- = V_rV_r^T$ is the projection matrix onto the row space of B. Therefore,

$$\|C'_i\|^2 = \|C'_iB^-B\|^2 + \|C'_i(I - B^-B)\|^2 = \|C'_iB^-B\|^2 + \|C'_i - X_i^{*\prime}B\|^2 \ .$$

Accordingly,

$$\|C\|_F^2 = \|CB^-B\|_F^2 + \|C(I - B^-B)\|_F^2 \ .$$

For any X therefore,

$$\|XB - C\|_F^2 = \|(X - X^*)B + X^*B - C\|_F^2 = \|(X - X^*)B\|_F^2 + \|X^*B - C\|_F^2 \ .$$

The matrix of the first term $(X - X^*)B$ has row space that is a subspace of the row space of B. The matrix of the second term $X^*B - C = CB^-B - C = -C(I - B^-B)$ has row space that is orthogonal to the row space of B. Further arguments are similar as above. Let R be an $n \times m$

matrix that acts as an embedding of the row space of its argument, in this case, XB or C. We can proceed along lines as before,

$$\|XBR - CR\|_F^2 = \|(X - X^*)BR + (X^*B - C)R\|_F^2$$
$$= \|(X - X^*)BR\|_F^2 + \|(X^*B - C)R\|_F^2 + 2\text{tr } (X - X^*)BR((X^*B - C)R)^T$$

Denoting $X^*B - C$ by C^*,

$$\|XBR - CR\|_F^2 = \|(X - X^*)BR\|_F^2 + \|C^*\|_F^2 + 2\text{tr } (X - X^*)BR(C^*R)^T \ .$$

Assuming as before that R is an approximate norm preserving subspace embedding for the row space of B (assumption $A1_{\epsilon/3}$), and $\|C^*\|_F$ is preserved to within $1 \pm \epsilon$ factors (assumption $A2_\epsilon$), it remains to obtain bounds on the cross term tr $(X - X^*)BR(C^*R)^T$.

1. Following the original approach, show that

$$\text{tr } (X - X^*)BR(C^*R)^T \leq \|X - X^*\|_F \|BR(C^*R)^T\| \ .$$

2. Without loss of generality, assume that B has orthonormal rows. Now show that

$$\|BR(C^*R)^T\|_F^2 \leq \frac{\epsilon^2}{d^2} \|B\|_F^2 \|C^*\|_F^2$$

provided, R preserves each of the inner products of row i of B with row i' of C^*, namely,

$$\forall i, i' \in \{1, 2, \ldots, d\}, \quad \left|B_i' R (C_{i'}^{*\prime})^T\right| \leq \frac{\epsilon}{\sqrt{d}} \|B_i'\| \|C^{*\prime}\|$$

with probability $1 - \frac{\delta}{3}$.

3. Give an upper bound for the sufficient number of rows of the random matrix distribution for the conditions above to holds. Show that this is identical to the expressions obtained in the previous section.

Generalized affine embedding

We are given matrices $A \in \mathbb{R}^{n \times d}, B \in \mathbb{R}^{e \times p}, C \in \mathbb{R}^{n \times p}$ and the variable matrix $X \in \mathbb{R}^{d \times e}$. The minimization problem is the following.

$$\min_{X \in \mathbb{R}^{d \times p}} \|AXB - C\|_F \ .$$

Typically, n and p are large, $n \gg d$ and $p \gg e$. Dimensionality reduction with respect to both n and p is considered.

The following notation is used.

$Y^* = A^- C$.

$C^* = AA^- C - C = -(I - AA^-)C,$ col. space of $\mathbb{C}^*$ is perpendicular to col. space of C.

$X^* = A^- C B^-$.

$B^* = AX^*B - AY^* = AA^- CB^- B - AA^- C = -AA^- C(I - B^- B)$.

1. Let $X^* = A^- C B^-$ and $Y^* = A^- C$. Denote $C^* = AY^* - C = AA^- C - C$ (as usual) and let $B^* = AX^* B - AY^* = AA^- C B^- B - AA^- C$. Show that

$$\|AXB - C\|_F^2 = \|A(X - X^*)B\|_F^2 + \|A(X^* B - Y^*)\|_F^2 + \|AY^* - C\|_F^2$$

Soln.

$$\begin{aligned}
\|AXB - C\|_F^2 &= \|AXB - AY^* + AY^* - C\|_F^2 \\
&= \|AXB - AY^*\|_F^2 + \|AY^* - C\|_F^2 \\
&= \|AXB - AX^*B + AX^*B - AY^*\|_F^2 + \|AY^* - C\|_F^2 \\
&= \|AXB - AX^*B\|_F^2 + \|AX^*B - AY^*\|_F^2 + \|AY^*C - C\|_F^2 \\
&= \|A(X - X^*)B\|_F^2 + \|B^*\|_F^2 + \|C^*\|_F^2 \ .
\end{aligned}$$

We consider applying a dimensionality reduction embedding matrices S, R where S is applied from the left to A and R is applied from the right. For simplicity, let $S \sim \mathcal{D}_1$ and $R \sim \mathcal{D}_2$, where, $\mathcal{D}_1$ is a distribution over $k \times n$ matrices and $\mathcal{D}_2$ is a distribution over $p \times l$ matrices. The following events and their probabilities are assumed.

1. EA^{Subembd}: Under the distribution $\mathcal{D}_1$, the column space of A is approximately preserved, that is,

$$\mathrm{P}\left[\forall x \in \mathbb{R}^d,\ \|SAx\| \in (1 \pm \epsilon/20)\|Ax\|\right] \geq 1 - \frac{\delta}{20} \ . \tag{1}$$

2. EB^{Subembd}: $\mathcal{D}_2$ preserves the row space of B approximately, that is,

$$\mathrm{P}\left[\forall y^T \in \mathbb{R}^e,\ \|y^T BR\| \in (1 \pm \epsilon/20)\|y^T B\|\right] \geq 1 - \delta/20 \ . \tag{2}$$

3. EB^{Affembed}: $\mathcal{D}_2$ is an affine embedding, namely,

$$\mathrm{P}\left[\forall Y \in\ \mathbb{R}^{n \times e}, \|YBR - AY^*R\|_F \in (1 \pm \epsilon)\|YB - AY^*\|_F\right] \geq 1 - \delta/20 \ . \tag{3}$$

where, $AY^* = AA^- C$.

4. $E^{\text{Frob}}_{C^*R}$: $(\mathcal{D}_2, \mathcal{D}_1)$ satisfy the following property:

$$\mathrm{P}_{S|R}\left[\|SC^*R\|_F \in (1 \pm \epsilon/20)\|C^*R\|_F\right] \geq 1 - \delta/20 \ . \tag{4}$$

5. $E^{\text{Frob}}_{C^*}$:

$$\mathrm{P}\left[\|C^*R\|_F \in (1 \pm \epsilon/20)\|C^*\|_F\right] \geq 1 - \delta/20 \ . \tag{5}$$

6. $E^{\text{Frob}}_{B^*R}$: $(\mathcal{D}_2, \mathcal{D}_1)$ satisfy the following property:

$$\mathrm{P}_{S|R}\left[\|SB^*R\|_F \in (1 \pm \epsilon/20)\|B^*R\|_F\right] \geq 1 - \delta/20 \ . \tag{6}$$

7. $E^{\text{Frob}}_{B^*}$:

$$\mathrm{P}\left[\|B^*R\|_F \in (1 \pm \epsilon/20)\|B^*\|_F\right] \geq 1 - \delta/20 \ . \tag{7}$$

8. $E_A^{\text{subsp}\cdot\text{ip}\cdot\text{set}}$:

$$\mathrm{P}_{S|R}\Big\{\forall Y \in \mathbb{R}^d \times p, \quad \left|(S(AY - AY^*)R)_j^T S(AY^*R - CR)_j - ((AXB - AY^*)R_j)^T(AY^* - C)R_j\right| \\ \leq (\epsilon/20)\|(AXB - AY^*)R_j\|\|(AY^* - C)R_j\|\} \geq 1 - \delta/20 \ . \quad (8)$$

9. E_A^{ipsimple}:

$$\mathrm{P}_{S|R}\left[\forall\ j = 1\ldots d, j' = 1\ldots l\ \ \left|(SA_j)^T(SC^*R_{j'})\right| \leq \frac{\epsilon}{20\sqrt{d}}\|A_j\|\|C^*R_{j'}\|\right] \geq 1 - \delta/20 \ . \quad (9)$$

10. $E_B^{\text{subsp}\cdot\text{ip}\cdot\text{set}}$:

$$\mathrm{P}_{R|S}\left[\forall\ y \in \mathbb{R}^e, i = 1, 2, \ldots, k, \quad \left|y^TBR(SB^*)_i'R^T - y^TB(SB^*)_i\right| \leq (\epsilon/20)\|y^TB\|_2\|(SB^*)_i'\|_2\right] \geq 1 - \delta/20 \ . \quad (10)$$

11. $E_B^{ipsimple}$.

$$\mathrm{P}_{R|S}\left[\left|B_i'(SAB)_{i'} - B_i'R((SAB)_{i'}R)^T\right| \leq \frac{\epsilon}{20\sqrt{e}}\|B_i'\|\|(SAB)_{i'}\|\right] \geq 1 - \delta/20 \qquad (11)$$

12. $E_{C^*}^{\text{Frob}}$: $\mathcal{D}_2$ preserves the Frobenius norm of C^* approximately, that is,

$$\mathrm{P}\left[\|C^*R\|_F \in (1 \pm \epsilon/30)\|C^*\|_F\right] \geq 1 - \frac{\delta}{20} \ . \qquad (12)$$

13. $E_A^{JLunion}$:,

$$\mathrm{P}\left[\|SAB^*\|_F \in (1 \pm \epsilon/20)\|AB^*\|_F\right] \geq 1 - \delta/20 \ . \qquad (13)$$

This event is contained in the event E_A^{Subembd}.

For all analysis, and all further questions, we assume that $\epsilon \leq \frac{1}{4}$.

1. Show that

$$\|SAXBR - SCR\|_F^2 \\ = \|SA(X - X^*)BR\|_F^2 + \|SA(X^*B - Y^*R)\|_F^2 + \|SAY^*R - SCR\|_F^2 \\ +2\mathrm{tr}\ SA(X - X^*)BR(SA(X^*B - Y^*)R)^T + 2\mathrm{tr}\ (S(AXB - AY^*)R)^T(SAY^*R - SCR) \ . \quad (14)$$

Soln.

$$
\begin{aligned}
&\|SAXBR - SCR\|_F^2 \\
&= \|S(AXB - AY^*)R + SAY^*R - SCR\|_F^2 \\
&= \|S(AXB - AY^*)R\|_F^2 + \|SAY^*R - SCR\|_F^2 + 2\text{tr } (S(AXB - AY^*)R)^T(SAY^*R - SCR) \\
&= \|SA(X - X^*)BR + SA(X^*B - Y^*)R\|_F^2 + \|SAY^*R - SCR\|_F^2 \\
&\quad + 2\text{tr } (S(AXB - AY^*)R)^T(SAY^*R - SCR) \\
&= \|SA(X - X^*)BR\|_F^2 + \|SA(X^*B - Y^*R)\|_F^2 + 2\text{tr } SA(X - X^*)BR(SA(X^*B - Y^*)R)^T \\
&\quad + \|SAY^*R - SCR\|_F^2 + 2\text{tr } (S(AXB - AY^*)R)^T(SAY^*R - SCR) \\
&= \|SA(X - X^*)BR\|_F^2 + \|SA(X^*B - Y^*R)\|_F^2 + \|SAY^*R - SCR\|_F^2 \\
&\quad + 2\text{tr } SA(X - X^*)BR(SA(X^*B - Y^*)R)^T + 2\text{tr } (S(AXB - AY^*)R)^T(SAY^*R - SCR)
\end{aligned}
$$

2. Assuming events E_A^{Subembd} and E_B^{Subembd}, show that

$$\|SA(X - X^*)BR\|_F^2 \in (1 \pm \epsilon)\|A(X - X^*)B\|_F^2 \ .$$

Soln.

$$
\begin{aligned}
\|SA(X - X^*)BR\|_F^2 &\in (1 \pm \epsilon/20)^2\|A(X - X^*)BR\|_F^2, \quad \text{assuming } E_A^{\text{Subembd}} \text{ holds} \\
&\in (1 \pm \epsilon/20)^4\|A(X - X^*)B\|_F^2, \quad \text{assuming } E_B^{\text{Subembd}} \text{ holds} \\
&\in (1 \pm \epsilon)\|A(X - X^*)B\|_F^2 \ .
\end{aligned}
$$

3. Show that under assumption $E_A^{\text{subsp}\cdot\text{ip}\cdot\text{set}}$ and assumption $E_{C^*}^{\text{Frob}}$,

$$\forall X \in \mathbb{R}^{d\times e}, \ \left|\text{tr } (S(AXB - AY^*)R)^T(SAY^*R - SCR)\right| \le (\epsilon/10)\|(AXB - AY^*)\|_F\|C^*\|_F \ .$$

Soln.

$$\text{tr } (S(AXB - AY^*)R)^T(SAY^*R - SCR) = \sum_{j=1}^{l}(S(AXB - AY^*)R)_j^T S(AY^*R - CR)_j$$

Consider the inner products of the jth columns of $((AXB - AY^*)R)$ with $(AY^*R - CR)$. Assuming the event $E_A^{\text{subsp}\cdot\text{ip}\cdot\text{set}}$ given by Eqn (10),

$$
\begin{aligned}
&\left|(S(AXB - AY^*)R)_j^T S(AY^*R - CR)_j - ((AXB - AY^*)R_j)^T(AY^* - C)R_j\right| \\
&\qquad \le \epsilon\|(AXB - AY^*)R_j\|\|(AY^* - C)R_j\| \ . \quad (15)
\end{aligned}
$$

As discussed earlier, each column q of $AY^* - C$ is $AY_q^* - C_q = AA^-C_q - C_q = -(I - AA^-)C_q$ and is orthogonal to the column space of A. Therefore, $AY^* - C = -(I - AA^-)C$ and has column space that is orthogonal to the column space of A. Thus,

$$((AXB - AY^*)R_j)^T(AY^* - C)R_j = R_j^T(AXB - AY^*)^T(AY^* - C)R_j = R_j^T 0_{l\times l}R = 0 \ .$$

Therefore,

$$
\begin{aligned}
&\left|\text{tr } (S(AXB - AY^*)R)^T(SAY^*R - SCR)\right| \\
&\le \sum_{j=1}^{l} \left|(S(AXB - AY^*)R)_j^T S(AY^*R - CR)_j\right|, \quad \text{by triangle inequality over reals} \\
&\le \sum_{j=1}^{l} \left|(S(AXB - AY^*)R)_j^T S(AY^*R - CR)_j - ((AXB - AY^*)R_j)^T(AY^* - C)R_j\right| \\
&\quad + \sum_{j=1}^{l} \left|((AXB - AY^*)R_j)^T(AY^* - C)R_j\right|, \quad \text{triangle inequality} \\
&\le \sum_{j=1}^{l} \epsilon\|(AXB - AY^*)R_j\|\|(AY^* - C)R_j\| \\
&\le (\epsilon/20)\left(\sum_{j=1}^{l}\|(AXB - AY^*)R_j\|^2\right)^{1/2}\left(\sum_{j=1}^{l}\|(AY^* - C)R_j\|^2\right)^{1/2}, \quad \text{Cauchy-Schwarz inequality} \\
&= (\epsilon/20)\|(AXB - AY^*)R\|_F\|C^*R\|_F, \quad \text{since, } AY^* - C = C^* \ . \qquad (16)
\end{aligned}
$$

Further, from the assumption E_B^{Affembed}, we have that

$$\forall W \in \mathbb{R}^{n\times p}, \|(WB - AY^*)R\|_F \in (1 \pm \epsilon/20))\|WB - AY^*\|_F \ .$$

Therefore conditional on the event E_B^{Affembed}, it follows that

$$\forall X \in \mathbb{R}^{d\times e},\ \|(AXB - AY^*)R\|)F \le (1 + \epsilon/20)\|AXB - AY^*\|_F \ .$$

(Can the event E_B^{Affembed} be more restricted to consider only the embeddings satisfying the above condition?) From Assumption $E_{C^*}^{\text{Frob}}$, we have that $\|C^*R\|_F \in (1 \pm \epsilon/20)\|C^*\|_F$. Conditional on event$E^{\text{Frob}}$ and applying it to Eqn (16), we have,

$$
\begin{aligned}
\left|\text{tr } (S(AXB - AY^*)R)^T(SAY^*R - SCR)\right| &\le (\epsilon/20)(1 + \epsilon/20)\|(AXB - AY^*)\|_F\|C^*\|_F \\
&\le (\epsilon/10)\|(AXB - AY^*)\|_F\|C^*\|_F \ .
\end{aligned}
$$

4. We consider an alternative way to bound tr $(S(AXB-AY^*C)R)^T(S(AY^*-C)R) = \text{tr } (S(AXB-AY^*C)R)^T SC^*R$. Show that under assumptions $E_{C^*}^{\text{Frob}}$ and E_A^{ipsimple},

$$\left|\text{tr } ((AXB - AY^*C)R)^T SC^*R\right| \le (\epsilon/10)\|(XB - Y^*C)\|_F\|C^*\|_F \ .$$

Soln.

$$
\begin{aligned}
&\text{tr } (S(AXB - AY^*C)R)^T SC^*R \\
&= \text{tr } (SA(XBR - Y^*CR))^T SC^*R \\
&= \text{tr } (XBR - Y^*CR)^T(SA)^T SC^*R \\
&\le \|XBR - Y^*CR\|_F\|(SA)^T SC^*R\|_F, \quad \text{tr } C^TD \le \|C\|_F\|D\|_F.
\end{aligned}
$$

We now assume that the event E_A^{ipsimple} and $E_{C^*}^{\text{Frob}}$ holds. Since, $A^T C^* = 0$, following earlier analysis,

$$\begin{aligned}
&\|(SA)^T SC^* R\|_F^2 \\
&= \|(SA)^T SC^* R - A^T C^* R\|_F^2 \\
&= \sum_{j=1}^{d} \sum_{j'=1}^{l} \left|(SA_j)^T S(C^* R_{j'}) - A_j^T C^* R_{j'}\right|^2 \\
&\le \sum_{j=1}^{d} \sum_{j'=1}^{l} \frac{\epsilon^2}{(10\sqrt{d})^2} \|A_j\|_2^2 \|C^* R_{j'}\|_2^2, \qquad \text{assuming event } E_A^{\text{ipsimple}} \text{ holds.} \\
&= \left(\frac{\epsilon}{20\sqrt{d}}\right)^2 \|A\|_F^2 \|C^* R\|_F^2 \\
&= \frac{\epsilon^2}{(20)^2} \|C^* R\|_F^2, \qquad \text{since, } \|A\|_F^2 = d \\
&\le \frac{\epsilon^2}{20^2} (1 + \epsilon/20)^2 \|C^*\|_F^2, \qquad \text{assuming event } E_{C^*}^{\text{Frob}} \text{ holds.} \\
&\le \frac{\epsilon^2}{16^2} \|C^*\|_F^2
\end{aligned}$$

and therefore

$$\left|\text{tr } ((AXB - AY^*C)R)^T SC^* R\right| \le \frac{\epsilon}{16} \|(XB - Y^*C)R\|_F \|C^* R\|_F \; .$$

5. We now consider the first trace term in Eqn (14). Show that by assuming the events $E_B^{\text{subsp·ip·set}}$, E_A^{Subembd} and $E_B^{\text{subsp·ip·set}}$, we have.

$$\left|\text{tr } SA(X - X^*)BR(SA(X^*B - Y^*)R)^T\right| \;\le\; (\epsilon/10)\|A(X - X^*)B\|_F \|A(X^*B - Y^*)\|_F \; . \tag{17}$$

Soln.

$$\begin{aligned}
&\text{tr } SA(X - X^*)BR(SA(X^*B - Y^*)R)^T \\
&= \sum_{i=1}^{k} ((SA(X - X^*)B)'_i R((SA(X^*B - Y^*))'_i R)^T
\end{aligned}$$

We will assume the event $E_B^{\text{subsp·ip·set}}$ which states that for all $y \in \mathbb{R}^e$ and $i = 1, 2, \ldots, k$,

$$\forall \; y \in \mathbb{R}^e, i = 1, 2, \ldots, k, \quad \left|y^T BR(SB^*)'_i R^T - y^T B(SB^*)_i\right| \le (\epsilon/20)\|y^T B\|_2 \|(SB^*)'_i\|_2 \; .$$

We note that

$$X^*B - Y^* = Y^*B^-B - Y^* = -(I - B^-B)$$

has row space that is orthogonal to the rowspace of B. Hence, for any matrix $Y \in \mathbb{R}^{k \times p}$ and $T \in R^{k \times e}$,

$$(YB)'_i((T(X^*B - Y^*))'_i)^T = (YB)'_i(T'_i(I - B^-B))^T = 0$$

and therefore,

$$YB(T(X^*B - Y^*))^T = 0 \ .$$

Using the event $E_B^{\text{subsp·ip·set}}$, and assuming $Y = SA(X - X^*)$ and $T = SA$ (can this be improved?), we have,

$$\begin{aligned}
&\left|\text{tr } SA(X - X^*)BR(SA(X^*B - Y^*)R)^T\right| && (18)\\
&= \left|\text{tr } SA(X - X^*)BR(SA(X^*B - Y^*)R)^T - \text{tr } SA(X - X^*)B(SA(X^*B - Y^*)\right| \\
&\leq \sum_{i=1}^{l} \left|((SA(X - X^*)B)_i)R(SA(X^*B - Y^*)_iR)^T - (SA(X - X^*)B_i(SA(X^*B - Y^*)_i\right| \\
&\leq \sum_{i=1}^{l} (\epsilon/20)\|(SA(X - X^*)B)_i\|\|(SA(X^*B - Y^*))_i\|, \quad \text{assuming event } E_B^{\text{subsp·ip·set}} \\
&\leq (\epsilon/20)\|(SA(X - X^*)B\|_F\|SA(X^*B - Y^*)\|_F \ . && (19)
\end{aligned}$$

Further assuming colspace embedding event E_A^{Subembd} and preservation of Frobenius norm event $E_{B^*}^{\text{Frob}}$, these yield that

$$\|SA(X - X^*)B\|_F\|SA(X^*B - Y^*)\|_F \in (1 \pm \epsilon/20)^2\|A(X - X^*)B\|_F\|A(X^*B - Y^*)\|_F \ .$$

Substituting in Eqn (19), we get

$$\begin{aligned}
&\left|\text{tr } SA(X - X^*)BR(SA(X^*B - Y^*)R)^T\right| \\
&\leq (\epsilon/20)(1 + \epsilon/20)^2\|A(X - X^*)B\|_F\|A(X^*B - Y^*)\|_F \\
&\leq (\epsilon/10)\|A(X - X^*)B\|_F\|A(X^*B - Y^*)\|_F \ .
\end{aligned}$$

6. This gives an alternative approach to the previous part. Here we assume that B has orthonormal rows. Assuming events E_B^{Subembd}, $E_{B^*}^{frobsimp}$ and E_A^{Subembd}, show that

$$\left|\text{tr } SA(X - X^*)BR(SA(X^*B - Y^*)R)^T\right| \leq (\epsilon/10)\|A(X - X)^*\|_F\|AB^*\|_F \ .$$

Soln. .

$$\begin{aligned}
&\left|\text{tr } SA(X - X^*)BR(SA(X^*B - Y^*)R)^T\right| \\
&= \left|(SA(X - X^*)BRR^T(X^*B - Y^*)^TA^TS^T\right| \\
&= \|SA(X - X^*)\|_F\|BR(SAB^*R)^T\|_F \ . && (20)
\end{aligned}$$

We now assume the event $EBIP_2$. It states that

$$\left|B_i'(SAB)_{i'} - B_i'R((SAB^*)_{i'}R)^T\right| \leq \frac{\epsilon}{20\sqrt{e}}\|B_i'\|\|(SAB^*)_{i'}\| \ .$$

Noting that the rowspaces of B and B^* are orthogonal to each other, we have,

$$
\begin{aligned}
&\|BR(SAB^*R)^T\|_F^2 \\
&= \|BR(SAB^*R)^T - B(SAB^*)^T\|_F^2 \\
&= \sum_{i=1}^{e}\sum_{i'=1}^{l} \left|(B_iR)^T((SAB^*)_{i'})R)^T - B_i^T(SAB^*)_{i'}^T\right|^2 \\
&\le \sum_{i=1}^{e}\sum_{i'=1}^{l} (\epsilon/(20\sqrt{e})^2)\|B_i\|^2\|(SAB^*)_{i'}\|^2 \\
&= \left(\frac{\epsilon}{20\sqrt{e}}\right)^2 \|B\|_F^2\|SAB^*\|_F^2 \\
&= \frac{\epsilon^2}{(20)^2}\|SAB^*\|_F^2
\end{aligned}
$$

since, $\|B\|_F^2 = d$.

We therefore have that

$$\|BR(SAB^*R)^T\|_F \le (\epsilon/20)\|SAB^*\|_F \ .$$

Assuming event E_A^{JLunion}, namely, $\|SAB^*\|_F \in (1 \pm \epsilon/20)\|AB^*\|_F$, we have, Subtituting this and the above inequality in Eqn (20), we obtain

$$\left|\text{tr } SA(X - X^*)BR(SA(X^*B - Y^*)R)^T\right| \le (\epsilon/20)(1 + \epsilon/20)\|SA(X - X^*)\|_F\|B\|_F\|AB^*\|_F \ .$$

Using a further step of assuming E_A^{Subembd}, we have $\|SA(X - X^*)\|_F \in (1 \pm \epsilon/20)\|A(X - X^*)\|_F$. This gives,

$$
\begin{aligned}
&\left|\text{tr } SA(X - X^*)BR(SA(X^*B - Y^*)R)^T\right| \\
&\le (\epsilon/20)(1 + \epsilon/20)^2\|A(X - X)^*\|_F\|AB^*\|_F \\
&\le (\epsilon/10)\|A(X - X)^*\|_F\|AB^*\|_F \ .
\end{aligned}
$$

7. Assuming that the events $E_A^{\text{Subembd}}, E_B^{\text{Subembd}}, E_A^{\text{subsp}\cdot\text{ip}\cdot\text{set}}, E_{C^*}^{\text{Frob}}, E_B^{\text{subsp}\cdot\text{ip}\cdot\text{set}}, E_{B^*}^{frobsimp}$, show that

$$\|SAXBR - SCR\|_F \in (1 \pm \epsilon/4)\|AXB - C\|_F \ .$$

Soln.

$$
\begin{aligned}
&\|SAXBR - SCR\|_F^2 \\
&\|SA(X - X^*)BR\|_F^2 + \|SA(X^*BR - Y^*R)\|_F^2 + \|SAY^*R - SCR\|_F^2 \\
&\qquad + 2\text{tr } SA(X - X^*)BR(SA(X^*B - Y^*)R)^T + 2\text{tr } (S(AXB - AY^*)R)^T(SAY^*R - SCR) \ . \\
&= (1 \pm \epsilon/10)\|A(X - X^*)B\|_F^2 + (1 \pm \epsilon/10)\|AB^*\|_F^2 + (1 \pm \epsilon/10)\|C^*\|_F^2 \\
&\quad + 2(\epsilon/16)\|AXB - AY^*\|_F\|C^*\|_F + 2(\epsilon/16)\|A(X - X^*)B\|_F\|AB^*\|_F \\
&= (1 \pm \epsilon/10)\left[\|A(X - X^*)B\|_F^2 + \|AB^*\|_F^2 + \|C^*\|_F^2\right] \\
&\quad + (\epsilon/8)(1/2)\left[\|AXB - AY^*\|_F^2 + \|C^*\|_F^2 + \|A(X - X^*)B\|_F^2 + \|AB^*\|_F^2\right] \qquad \text{since, } AM \ge GM \ .
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
&\left|\|SAXBR - SCR\|_F^2 - \|AXB - C\|_F^2\right| \\
&\leq (\epsilon/10)\left[\|A(X - X^*)B\|_F^2 + \|AB^*\|_F^2 + \|C^*\|_F^2\right] \\
&\quad (\epsilon/16)\left[\|AXB - AY^*\|_F^2 + \|C^*\|_F^2 + \|A(X - X^*)B\|_F^2 + \|AB^*\|_F^2\right] \\
&= (\epsilon/10 + \epsilon/16)\|AXB - C\|_F^2 + (\epsilon/16)\|AXB - C\|_F^2 \\
&\leq (\epsilon/4)\|AXB - C\|_F^2
\end{aligned}
$$

8. Show that under the assumption of random matrices $i.i.d$ random normal entries $S \sim \mathcal{D}_1$ with $S_{ij} \sim N(0, \frac{1}{k})$ and $R_{i',j'} \sim N(0, 1/l)$, then, the events E_A^{Subembd}, E_B^{Subembd}, $E_A^{\text{subsp·ip·set}}$, $E_{C^*}^{\text{Frob}}$, $E_B^{\text{subsp·ip·set}}$, $E_{B^*}^{frobsimp}$ hold provided,

$$
\begin{aligned}
&1. \quad k \geq O\left(\frac{d}{\epsilon^2}\log\frac{1}{\delta} + \frac{1}{\epsilon^2}\log\frac{p}{\delta}\right), \text{ and} \\
&2. \quad l \geq O\left(\frac{e}{\epsilon^2}\log\frac{1}{\delta} + \frac{e}{\epsilon^2}\log\frac{n}{\delta}\right) \quad .
\end{aligned}
$$

This gives a sufficient condition for the dimensionality reduction solution to be computed as

$$
\hat{X} = (SA)^- SCR(SB)^-
$$

namely, following standard calculation,

$$
\begin{aligned}
(1 - \epsilon)\|A\hat{X}B - C\|_F \leq \|SA\hat{X}BR - SCR\|_F \leq \|SAX^*BR - SCR\|_F& \\
\leq (1 + \epsilon/4)\|AX^*B - C\|_F&
\end{aligned}
$$

or, that

$$
\|A\hat{X}B - C\|_F \leq \left(\frac{1 + \epsilon/4}{1 - \epsilon/4}\right)\|AX^*B - C\| \leq (1 + \epsilon/2)\|AX^*B - C\|_F \quad .
$$