SENSITIVITY OF LEVERAGE SCORES

ILSE C. F. IPSEN* AND THOMAS WENTWORTH†

Abstract. The sampling strategies in many randomized matrix algorithms are, either explicitly or implicitly, controlled by statistical quantities called leverage scores. We present four bounds for the sensitivity of leverage scores as well as an upper bound for the principal angles between two matrices. These bounds are expressed by considering two real matrices. Our bounds show that if the principal angles between $A$ and $B$ are small, then the leverage scores of $B$ are close to the leverage scores of $A$. Next, we show that the principal angles can be bounded above by the two-norm condition number of $A$, $\kappa(A)$ and $\|B - A\|_2$. Finally, we combine these bounds and derive bounds for the leverage scores of $B$ in terms of $\kappa(A)$ and $\|B - A\|_2/\|A\|_2$ and show that if $\|B - A\|_2/\|A\|_2$ and $\kappa(A)$ are small, then the leverage scores of $B$ are close to the leverage scores of $A$.

1. Introduction. We provide a brief overview of leverage scores and principal angles.

Leverage Scores. Statistical leverage scores were introduced in 1978 by Hoaglin and Welsch [8] to detect outliers when computing regression diagnostics, see also [3] [4]. To be specific, consider the least squares problem $\min_x \|Ax - b\|_2$, where $A$ is a real $m \times n$ matrix with $\text{rank}(A) = n$. The so-called hat matrix $H = A(A^TA)^{-1}A^T$ is the orthogonal projector onto $\text{range}(A)$, and determines the fit $\hat{b} = Hb$. The diagonal elements of the hat matrix are called the leverage scores of $A$,

$$\ell_j(A) \equiv H_{jj}, \quad 1 \leq j \leq m,$$

because $\ell_j(A)$ reflects the leverage of the $j$th point $b_j$ on the corresponding fit $\hat{b}_j$. To see this, suppose that $\ell_k(A) = 1$ for some $k$. Then $\hat{b}_k = b_k$. Because $b_k$ has maximal leverage, it completely determines the corresponding element of the fit. That is, $k$th canonical vector, $e_k$, is in the column space of $A$, and $b_k$ can be fitted completely without affecting fit of the other elements of $b$. In contrast, if $\ell_k(A) = 0$ then $\hat{b}_k$ has zero leverage on the fit $\hat{b}_k$ and $\hat{b}_k = 0$. That is, $e_k$ is perpendicular to the column space of $A$.

Leverage scores can be stably computed from a thin QR decomposition $A = QR$, where $Q$ is $m \times n$ with orthonormal columns, via $\ell_j(A) = \|e_j^TQ\|_2^2$. Leverage scores can also be expressed in terms of the left singular vectors of $A$ that are associated with the non-zero singular values. In fact, for any $n \times n$ orthogonal matrix $W$, $\|e_j^TQ\|_2 = \|e_j^TWQ\|_2$ which leads us to the following definition of the leverage scores of a matrix $A$.

Definition 1.1. Given a $m \times n$ real matrix $A$ with $m > n$ and full column rank, let $Q$ be any basis of orthonormal columns for the column space of $A$. Then, the leverage scores of $A$ are defined as

$$\ell_i(A) = \|e_i^TQ\|_2^2, \quad 1 \leq i \leq m.$$

Leverage scores are the basis for many sampling strategies in randomized matrix computations [10], including low rank approximations [5], CUR decompositions [6],

---

*Department of Mathematics, North Carolina State University, P.O. Box 8205, Raleigh, NC 27695-8205, USA, (ipsen@ncsu.edu, http://www4.ncsu.edu/~ipsen/)

†Department of Mathematics, North Carolina State University, P.O. Box 8205, Raleigh, NC 27695-8205, USA (thomas_wentworth@ncsu.edu)
Principal Angles. The principal angles, also called the canonical angles, between two matrices, \( A \) and \( B \), describe the minimal angles between their subspaces. They give a measure of the distance between the column space of \( A \) and the column space of \( B \). These angles can be described in terms of a singular value decomposition.

**Definition 1.2** (§12.4.3, [11]). Given \( m \times n \) real, full column rank matrices \( A \) and \( B \), let \( Q \) and \( \tilde{Q} \) be any bases of orthonormal columns for \( A \) and \( B \), respectively. Let \( Q^T \tilde{Q} = Y \Sigma Z^T \) be a SVD. Define

\[
\theta_j = \arccos(\Sigma_{j,j}) \in [0, \pi/2],
\]

for \( 1 \leq j \leq n \), to be the principal angles between \( A \) and \( B \), and define the columns of \( U = QY \) and \( V = QZ \) to be the principal vectors of \( A \) and \( B \), respectively. The principal angles are ordered such that \( 0 \leq \theta_1 \leq \ldots \leq \theta_n \leq \pi/2 \). The principal vectors have the special property that \( \cos(\theta_j) = (U)_{j,j}^T (V)_{j,j} \), and \( (U)_{j,j}^T (V)_{j,k} = 0 \) for \( 1 \leq k \leq n \) and \( j \neq k \).

We also use an alternative definition for \( \theta_n \) from [11].

**Definition 1.3** (§5.15, [11]). Given two matrices, \( A \) and \( B \), let \( P_A \) and \( P_B \) be orthogonal projectors onto the column space of \( A \) and \( B \), respectively, and define

\[
\theta_n(A, B) = \arcsin \left( \max \left\{ \max_{x, \|Ax\|_2 = 1} (\|Ax - P_BAx\|_2), \max_{x, \|Bx\|_2 = 1} (\|Bx - P_ABx\|_2) \right\} \right)
\]

be the maximal principal angle between the two subspaces.

2. Supplemental Results. Below we present two theorems and a lemma leading up to a leverage score perturbation bound. We consider two \( m \times n \) matrices of full column rank, \( A \) and \( B \), and bound \( \ell_i(B) \) in terms of \( \ell_i(A) \) and the principal angles between \( A \) and \( B \). Next, we bound the largest principal angle in terms of \( \|B - A\|_2 \) and \( \kappa(A) = \|A\|_2 \|A^T\|_2 \). Finally, we present a second upper bound on the largest principal angle. These results are used to obtain our main results in Section 3.

2.1. Leverage Score Perturbation in Terms of Principal Angles. Theorem 2.1 below uses principal angles and the triangle inequality to bound \( \ell_i(B) \) in terms of \( \ell_i(A) \).

**Theorem 2.1.** Let \( A \) and \( B \) be \( m \times n \) real, full column rank matrices, and let \( \theta_j \), for \( 1 \leq j \leq n \), be the principal angles as described in Definition 1.2. Then,

\[
\ell_i(B) \leq \left( \cos(\theta_1) \sqrt{\ell_i(A)} + \sin(\theta_n) \sqrt{1 - \ell_i(A)} \right)^2, \text{ for } 1 \leq i \leq m.
\]

In addition, for each \( 1 \leq i \leq m \), if \( \sqrt{\ell_i(A)} - \sin(\theta_n) \geq 0 \) and \( \cos(\theta_1) > 0 \), then

\[
\left( \frac{\sqrt{\ell_i(A)} - \sin(\theta_n)}{\cos(\theta_1)} \right)^2 \leq \ell_i(B).
\]

**Proof.** See Section A.1

Theorem 2.2 below uses principal angles to bound \( |\ell_i(A) - \ell_iB| \).
Theorem 2.2. Let $A$ and $B$ be $m \times n$ real, full column rank matrices, and let $\theta_j$, for $1 \leq j \leq n$, be the principal angles as described in Definition 1.2. Then,

$$|\ell_i(B) - \ell_i(A)| \leq \sin(\theta_n)^2 + 2\sqrt{\ell_i(A)}\sqrt{1 - \ell_i(A)}\sin(\theta_n).$$

Proof. See Section A.2. \qed

Theorem 2.1 & 2.2 show that the leverage scores of $A$ and $B$ are close if the principal angles between their column spaces are small. If $A$ and $B$ have the same column space, then all of the principal angles are zero and our bound confirms that their leverage scores are equal. On the other hand, if $A$ and $B$ are completely orthogonal to each other, then Theorem 2.1 confirms that $\ell_i(B) + \ell_i(A) \leq 1$.

2.2. Upper Bounds for the Largest Principal Angle. In Theorem 2.3, we bound the maximal principal angle between $A$ and $B$ in terms of the relative difference between $A$ and $B$, $\frac{\|B-A\|_2}{\|A\|_2}$, and $\kappa(A)$. This bound is not the first of its kind. In [12, Theorem 3.1], the author proves a more general theorem that implies a slightly less tight version of our bound.

Theorem 2.3. Let $A$ and $B$ be $m \times n$ matrices with $m > n$. If $A$ has full column rank, and $\|B - A\|_2 < \|A\|_2^{-1}$, then

$$\theta_n \leq \arcsin\left(\frac{\|B - A\|_2 \|A\|_2}{1 - \|B - A\|_2 \|A\|_2}\right) = \arcsin\left(\frac{\kappa(A)\varepsilon}{1 - \kappa(A)\varepsilon}\right),$$

where

$$\varepsilon = \frac{\|B - A\|_2}{\|A\|_2}.$$

Proof. See Section A.3. \qed

Theorem 2.3 shows that the largest principal angle is small if $A$ is well conditioned and the relative difference between $A$ and $B$ is small.

In Lemma 2.4, we present a second upper bound on the largest principal angle between $A$ and $B$. We use this lemma to remove the condition $\cos(\theta_1) > 0$ from Theorem 2.1 for Corollary 3.1.

Lemma 2.4. Given $m \times n$, real, full column rank matrices $A$ and $B$, let $\theta_j$, for $1 \leq j \leq n$, be the principal angles of $A$ and $B$ as described in Definition 1.2. If $\|B - A\|_2 \leq \|A\|_2^{-1}$, then $\theta_1 \leq \theta_n < \pi/2$.

Proof. See Section A.4. \qed

3. Leverage Score Perturbation in terms of Matrix Perturbation. Here we combine our previous results to obtain a leverage score perturbation bounds. First, we give a two sided bound.

Corollary 3.1. Given $m \times n$ real matrices $A$ and $B$ with $m > n$ such that $A$ has full column rank and $\|B - A\|_2 < \|A\|_2^{-1}$. Then,

$$\ell_i(B) \leq \left(\ell_i(A) + \frac{\|A\|_2\|B - A\|_2}{1 - \|A\|_2\|B - A\|_2}\sqrt{1 - \ell_i(A)}\right)^2, \text{ for } 1 \leq i \leq m.$$
In addition, for each \(1 \leq i \leq m\), if 
\[
\sqrt{\ell_i(A)} - \frac{\|A^\dagger\|_2 \|B - A\|_2}{1 - \|A^\dagger\|_2 \|B - A\|_2} \geq 0,
\]
then
\[
\left(\sqrt{\ell_i(A)} - \frac{\|A^\dagger\|_2 \|B - A\|_2}{1 - \|A^\dagger\|_2 \|B - A\|_2}\right)^2 \leq \ell_i(B).
\]

Proof. See Section A.5. \(\square\)

This bound shows that the leverage scores of \(A\) and \(B\) are close if \(A\) is well conditioned and the relative difference between \(A\) and \(B\) is small.

While trying to examine Corollary 3.1 we found that the result is somewhat difficult to plot. Even with logarithmically scaled axes, the upper and lower bounds are indistinguishable since their difference is much smaller than their magnitude. To aid in examining our results, we reworked Theorem 2.1 and Corollary 3.1 as a bound on \(|\ell_i(A) - \ell_i(B)|\). While this bound is slightly less descriptive than Corollary 3.1, it has the advantage that one can actually see the results in a plot.

**Corollary 3.2.** Given \(m \times n\) real matrices \(A\) and \(B\) with \(m > n\) such that \(A\) has full column rank and \(\|B - A\|_2 < \|A\|_2^{-1}\) and let
\[
\varepsilon = \frac{\|B - A\|_2}{\|A\|_2}.
\]

Then,
\[
|\ell_i(B) - \ell_i(A)| \leq 2\sqrt{\ell_i(A)} \sqrt{1 - \ell_i(A)} \frac{\|A^\dagger\|_2 \|B - A\|_2}{1 - \|A^\dagger\|_2 \|B - A\|_2} + \left(\frac{\|A^\dagger\|_2 \|B - A\|_2}{1 - \|A^\dagger\|_2 \|B - A\|_2}\right)^2, \text{ for } 1 \leq i \leq m.
\]

and if \(\ell_i(A) > 0\), then
\[
\frac{|\ell_i(B) - \ell_i(A)|}{\ell_i(A)} \leq 2\sqrt{\frac{1 - \ell_i(A)}{\ell_i(A)(1 - \kappa(A))\varepsilon}} + \left(\frac{\kappa(A)\varepsilon}{\ell_i(A)(1 - \kappa(A))\varepsilon}\right)^2, \text{ for } 1 \leq i \leq m.
\]

Proof. See Section A.6. \(\square\)

3.1. Leverage score bound for givens rotation. In this subsection, we bound the leverage scores of a matrix \(G_{i,j}A\) where \(G_{i,j}\) is a Givens matrix that rotates row \(i\) with row \(j\) by \(\phi\) radians.

**Theorem 3.3.** Given the \(m \times n\) real, full column rank matrix \(A\) with \(m > n\), let \(G_{i,j}\) be the \(m \times m\) Givens rotation matrix that rotates row \(i\) with row \(j\) by \(\phi\) radians. Then,
\[
\ell_i(G_{i,j}A) \leq \cos(\phi)^2 \ell_i(A) + \sin(\phi)^2 \ell_j(A) + \sqrt{\ell_i(A) \ell_j(A)}
\]
and
\[
\ell_i(G_{i,j}A) \geq \cos(\phi)^2 \ell_i(A) + \sin(\phi)^2 \ell_j(A) - \sqrt{\ell_i(U) \ell_j(U)}.
\]

Proof. See Section A.7. \(\square\)

This bound shows that the change in leverage score is small if the rotation angle of \(G_{i,j}\) is small.
Fig. 4.1. Here, $A$ has orthonormal columns and thus $\|A\|_2 = 1$ and $\kappa(A) = 1$ and $\|E\|_2 \approx 2.3 \times 10^{-15}$. The entries of $E$ have mean 0 and variance $10^{-16}$. On the left, we plot the absolute change in the leverage scores and the absolute bound from Corollary 3.2 and on the right we plot the relative change in the leverage scores against the relative bound from Corollary 3.2.

4. Experiments. In this section, we examine the tightness and behavior of Corollary 3.2 with a few carefully constructed examples. In all of the examples, we set $m = 500$ and $n = 3$ and we use [9, Algorithm 6.1] to construct the unperturbed matrix $A$ with orthonormal columns. We use this algorithm because it allows us to create a matrix with a wide distribution of leverage scores. It is important to note that since Corollary 3.2 depends on $\kappa(A)\|A - B\|_2/\|A\|_2$, we would not gain any insight on the performance of the bound by considering matrices with different condition numbers. It is for this reason that we only consider matrices $A$ with orthonormal columns.

In Figures 4.1, 4.2 and 4.3 the horizontal coordinate axis represents the leverage scores of $A$ (the unperturbed leverage scores) and the vertical coordinate axis represents either the absolute or relative magnitude of the change in leverage scores. The dots represent the absolute difference between the leverage scores of $A$ and $B$, $|\ell_i(B) - \ell_i(A)|$, and the black line represents the bound from Corollary 3.2.

Example 1. Here we present two plots to show how well Corollary 3.2 predicts the behavior of the leverage scores of $B = A + E$ where $E$ is a matrix whose entries are independent realizations of a normal random variable.
In Figures 4.1 and 4.2 we see that Corollary 3.2 accurately bounds the absolute difference between the leverage scores of $A$ and $B$. In particular, we see the following behaviors.

- **Small leverage scores have a larger relative perturbation.** When $\|A - B\|_2$ is small, as in Figure 4.1, the relative bound on the leverage score perturbation is $O\left( \frac{1 - \ell_i(A)}{\ell_i(A)} \right)$, which is larger for small $\ell_i(A)$. On the other hand, when $\|A - B\|_2$ is large, as in Figure 4.2, the relative perturbation bound on the small leverage scores is $O\left( \frac{\kappa(A) \varepsilon}{\ell_i(A)(1 - \kappa(A) \varepsilon)} \right)^2$, which is again larger for small $\ell_i(A)$.

- **The absolute perturbation bound for Small leverage scores if flat when $\|A - B\|_2$ is large.** When $\|A - B\|_2$ is large, as in Figure 4.2, the $(\frac{\|A\|_2\|B - A\|_2}{1 - \|A\|_2\|B - A\|_2})^2$ term in Corollary 3.2 becomes dominate. Thus, the absolute perturbation bound for small leverage scores below a certain threshold is flat as it does not significantly depend on $\ell_i(A)$. The intuition for this is that even a zero leverage score can be increased by a certain amount for a given $\|A - B\|_2$ and this amount becomes significant for small leverage scores.

- **Large leverage scores have a small relative perturbation even when $\varepsilon$ is large.** From Figure 4.2 we can see that even when $\varepsilon \approx 2.3 \times 10^{-3}$, the relative perturbation of the large leverage scores remains small. This is because, for $\ell_i(A) \approx 1$, the relative perturbation bound is $O(\frac{\kappa(A) \varepsilon}{1 - \kappa(A) \varepsilon})$.

At first glance, Corollary 3.2 does not look particularly tight as there is a large gap between the experimental data and the bound. This is because the elements of the perturbation, $E$, were sampled from a normal random variable and thus its affects are spread out among all of the leverage scores.

**Example 2.** By carefully constructing $E$, it is possible to focus its affects on a particular leverage score. In Figure 4.3 we have constructed $E$ in a way that attempts...
Fig. 4.3. Here, $A$ has orthonormal columns and thus $\|A\|_2 = 1$ and $\kappa(A) = 1$ and $\|E\|_2 \approx 10^{-4}$.

The $\times$ represents $\ell_{100}(A) - \ell_{100}(B)$ and is $\sim 3 \times 10^{-9}$ below the bound implied by Corollary 3.2 and on the right we plot the relative change in the leverage scores against the relative bound from Corollary 3.2.

Figure 4.3 shows that the bound from Corollary 3.2 can be very tight for particular leverage scores.

5. Conclusion. We have proven multiple perturbation bounds for leverage scores as well as an upper bound for the principal angles between two matrices. Theorem 2.1 shows that if the principal angles between $A$ and $B$ are small, then the leverage scores of $A$ and $B$ are close. Theorem 2.3 shows that if the two-norm relative difference between $A$ and $B$ is small and if the two-norm condition number of $A$ is small, then the principal angles between $A$ and $B$ are small. Finally, we combine Theorem 2.1 and Theorem 2.3 to obtain Corollary 3.1 which shows that the leverage scores of $A$ and $B$ are close if the condition number of $A$ and the two-norm relative difference between $A$ and $B$ are small.

REFERENCES


The algorithm used to construct this example is similar to a gradient descent method where we examine $\partial \ell_i(G_{i,j}A)/\partial\|G_{i,j}A\|_2$ and will be included in a future version of this paper.
Appendix A. Proofs.

A.1. Proof for Theorem 2.1. Let $U$ and $V$ be as defined in Definition 1.2 and define the $m \times m$ orthogonal matrix $C = \begin{bmatrix} U & U_\perp \end{bmatrix}$. Then $C^T V = \begin{bmatrix} \Sigma \end{bmatrix}$, where $D = U_\perp^T V$, hence $V = C \begin{bmatrix} \Sigma \\ D \end{bmatrix} = U \Sigma + U_\perp D$. Since the leverage scores of $B$ do not depend on the choice of basis (Definition 1.1),

$$\ell_i(B) = \| e_i^T V \|_2^2 = \| e_i^T (U \Sigma + U_\perp D) \|_2^2.$$

The triangle inequality gives us

$$\ell_i(B) \leq \left( \| e_i^T U \|_2 \| \Sigma \|_2 + \| e_i^T U_\perp \|_2 \| D \|_2 \right)^2 = \left( \sqrt{\ell_i(A)} \cos(\theta_i) + \| e_i^T U_\perp \|_2 \| D \|_2 \right)^2.$$

Since $C$ is an orthogonal matrix,

$$1 = \| e_i^T C \|_2^2 = \| e_i^T U \|_2^2 + \| e_i^T U_\perp \|_2^2 = \ell_i(A) + \| e_i^T U_\perp \|_2^2$$

implies

$$\ell_i(B) \leq \left( \sqrt{\ell_i(A)} \cos(\theta_i) + \sqrt{1 - \ell_i(A)} \| D \|_2 \right)^2.$$

Also, since $C^T V$ has orthonormal columns, we have $\Sigma^2 + D^T D = I$, and it follows that

$$\| D \|_2 = \sqrt{\| I - \Sigma^2 \|_2} = \sqrt{1 - \cos(\theta_n)^2} = \sin(\theta_n).$$

Finally,

$$\ell_i(B) \leq \left( \cos(\theta_1) \sqrt{\ell_i(A)} + \sin(\theta_n) \sqrt{1 - \ell_i(A)} \right)^2.$$

Reversing the roles of $A$ and $B$ gives

$$\ell_i(A) \leq \left( \cos(\theta_1) \sqrt{\ell_i(B)} + \sin(\theta_n) \sqrt{1 - \ell_i(B)} \right)^2 \leq \left( \cos(\theta_1) \sqrt{\ell_i(B)} + \sin(\theta_n) \right)^2.$$

Rearranging the terms gives

$$\sqrt{\ell_i(B)} \geq \frac{\sqrt{\ell_i(A)} - \sin(\theta_n)}{\cos(\theta_1)}.$$
In addition, for $1 \leq i \leq m$, if $\sqrt{\ell_i(A)} - \sin(\theta_i) \geq 0$,

$$\ell_i(B) \geq \left( \frac{\sqrt{\ell_i(A)} - \sin(\theta_i)}{\cos(\theta_1)} \right)^2.$$

**A.2. Proof for Theorem 2.2** Let $U$ and $V$ be as defined in Definition 1.2 and define the $m \times m$ orthogonal matrix $C = [U \quad U_\perp]$. Then $C^T V = \begin{bmatrix} \Sigma \\ D \end{bmatrix}$, where $D = U_\perp V$, hence $V = C \begin{bmatrix} \Sigma \\ D \end{bmatrix} = U \Sigma + U_\perp D$. Since the leverage scores of $B$ do not depend on the choice of basis (Definition 1.1),

$$\ell(B) = \|e_i^T V\|_2^2 = \|e_i^T (U \Sigma + U_\perp D)\|_2^2.$$

We begin with

$$|\ell_i(B) - \ell_i(A)| = \|\|e_i^T (U \Sigma + U_\perp D)\|_2^2 - \|e_i^T U\|_2^2|.$$ 

The triangle inequality and rearranging terms gives us

$$|\ell_i(B) - \ell_i(A)| \leq (\|e_i^T U\|_2^2 \|\Sigma\|_2 + \|e_i^T U_\perp\|_2 \|D\|_2)^2 - \|e_i^T U\|_2^2$$

$$\leq \|e_i^T U\|_2^2 (\|\Sigma\|_2^2 - 1 + 2 \|e_i^T U\|_2 \|e_i^T U_\perp\|_2 \|\Sigma\|_2 \|D\|_2 + \|e_i^T U_\perp\|_2 \|D\|_2^2)$$

$$\leq |\ell_i(A)(1 - \cos(\theta_n)^2) + 2 \sqrt{\ell_i(A)} \|e_i^T U\|_2 \|e_i^T U_\perp\|_2 \|\Sigma\|_2 \|D\|_2^2 + \|e_i^T U_\perp\|_2 \|D\|_2^2|$$

Since $C$ is an orthogonal matrix,

$$1 = \|e_i^T C\|_2^2 = \|e_i^T U\|_2^2 + \|e_i^T U_\perp\|_2^2 = \ell_i(A) + \|e_i^T U_\perp\|_2^2$$

implies

$$|\ell_i(B) - \ell_i(A)| \leq \ell_i(A)(1 - \cos(\theta_n)^2) + 2 \sqrt{\ell_i(A)} \sqrt{1 - \ell_i(A)} \cos(\theta_1) \|\Sigma\|_2 \|D\|_2^2 + (1 - \ell_i(A)) \|D\|_2^2.$$ 

Also, since $C^T V$ has orthonormal columns, we have $\Sigma^2 + D^T D = I$, and it follows that

$$\|D\|_2 = \sqrt{\|I - \Sigma^2\|_2} = \sqrt{1 - \cos(\theta_n)^2} = \sin(\theta_n).$$

This gives,

$$|\ell_i(B) - \ell_i(A)| \leq \ell_i(A)(1 - \cos(\theta_n)^2) + 2 \sqrt{\ell_i(A)} \sqrt{1 - \ell_i(A)} \cos(\theta_1) \sin(\theta_n) + (1 - \ell_i(A)) \sin(\theta_n)^2.$$ 

Finally, since $\sin(\theta_n)^2 + \cos(\theta_n)^2 = 1$ and $\cos(\theta_1) < 1$ we have,

$$|\ell_i(B) - \ell_i(A)| \leq \sin(\theta_n)^2 + 2 \sqrt{\ell_i(A)} \sqrt{1 - \ell_i(A)} \sin(\theta_n).$$

**A.3. Proof for Theorem 2.3** Start with Definition 1.3 for $\theta_n$. Let $x_1$ be a vector with $\|Ax_1\|_2 = 1$ that maximizes $\max_{x_1, \|Ax_1\|_2 = 1} (\|Ax - P_B Ax\|_2)$. Then, $P_B B x_1 = B x_1$ implies that

$$\|Ax_1 - P_B Ax_1\|_2 = \|Ax_1 - P_B (Ax_1 + B x_1 - B x_1)\|_2$$

$$= \|-(B - A)x_1 + P_B (B - A)x_1\|_2$$

$$= \|(I - P_B)(B - A)x_1\|_2.$$
Since \( P_B \) is an orthogonal projector, \( I - P_B \) is also an orthogonal projector, and thus \( \|I - P_B\|_2 = 1 \). It then follows that
\[
\|(I - P_B)(B - A)x_1\|_2 \leq \|B - A\|_2 \|x_1\|_2.
\]
Since \( A \) has a left inverse, it also follows that
\[
\|B - A\|_2 \|x_1\|_2 = \|B - A\|_2 \|A^\dagger A x_1\|_2 \leq \|B - A\|_2 \|A^\dagger\|_2 \|Ax_1\|_2 = \|B - A\|_2 \|A^\dagger\|_2.
\]
Thus,
\[
\|Ax_1 - P_B Ax_1\|_2 \leq \|B - A\|_2 \|A^\dagger\|_2. \tag{A.1}
\]
Let \( x_2 \) be a vector with \( \|Bx_2\|_2 = 1 \) that maximizes \( \max_{x_2} \|Bx - PAx\|_2 \). Since \( A \) has full column rank, and \( \|B - A\|_2 < \|A^\dagger\|^{-1}_2 \), it follows from the well conditioning of singular values that
\[
\|\sigma_n(B) - \sigma_n(A)\| \leq \|B - A\|_2 < \|A^\dagger\|^{-1}_2 = \sigma_n(A),
\]
and thus \( B \) has full column rank. Similarly to Equation (A.1) we have
\[
\|Bx_2 - P_A B x_2\|_2 \leq \|B - A\|_2 \|B^\dagger\|_2 = \frac{\|B - A\|_2}{\sigma_n(B)}.
\]
Again, from the well conditioning of singular values, it follows that \( \sigma_n(B) \geq \sigma_n(A) - \|B - A\|_2 \). The assumption \( \sigma_n(A) - \|B - A\|_2 > 0 \) then gives
\[
\|Bx_2 - P_A B x_2\|_2 \leq \frac{\|B - A\|_2}{\|A^\dagger\|^{-1}_2 - \|B - A\|_2} = \frac{\|B - A\|_2 \|A^\dagger\|_2}{\|B - A\|_2 \|A^\dagger\|_2}.
\]
Since \( \|B - A\|_2 < \|A^\dagger\|^{-1}_2 \), we know that
\[
\frac{\|B - A\|_2 \|A^\dagger\|_2}{\|B - A\|_2 \|A^\dagger\|_2} \geq \|B - A\|_2 \|A^\dagger\|_2,
\]
which finally gives us
\[
\theta_n \leq \arcsin\left(\frac{\|B - A\|_2 \|A^\dagger\|_2}{\|B - A\|_2 \|A^\dagger\|_2}\right) = \arcsin\left(\frac{\kappa(A)\varepsilon}{1 - \kappa(A)\varepsilon}\right).
\]

**A.4. Proof for Lemma 2.4** We prove the contrapositive; if \( \theta_n = \pi/2 \), then \( \|B - A\|_2 > \sigma_n(A) \).

If \( \theta_n = \pi/2 \), then \( \Sigma_{n,n} = \cos(\theta_n) = 0 \). This implies that \( Q^T \tilde{Q} \) has a zero singular value and thus, \( A^T B \) also has a zero singular value. Thus, there exists a vector \( x \), where \( \|x\|_2 = 1 \), such that \( x^T B^T A x = 0 \). Additionally, since \( A \) and \( B \) are full column rank, \( Ax \) and \( Bx \) are non-zero. Hence, Pythagoras implies the strict inequality, \( \|B - A\|_2 \geq \|(B - A)x\|_2 \). Using \( \|B - A\|_2 \geq \|(B - A)x\|_2 \) and \( \|Ax\|_2 \geq \sigma_n(A) \) gives us our desired result,
\[
\|B - A\|_2 > \sigma_n(A).
\]

**A.5. Proof for Corollary 3.1** Start with Theorem 2.1. We begin by removing the condition that \( \cos(\theta_1) > 0 \) since Lemma 2.4 ensures that this is always true when \( \|Ax\|_2 \geq \sigma_n(A) \). Finally, Theorem 2.3 allows us to substitute \( \arcsin\left(\frac{\|A^\dagger\|_2 \|B - A\|_2}{\|B - A\|_2 \|A^\dagger\|_2}\right) \) for \( \theta_n \).
A.6. Proof for Corollary 3.2: Start with Theorem 2.2 and use Theorem 2.3 to substitute \( \arcsin\left(\frac{\kappa(A)\varepsilon}{1-\kappa(A)\varepsilon}\right) \) for \( \theta_n \).

A.7. Proof for Theorem 3.3: Let \( U \) be a basis of orthonormal columns for \( A \). Then, \( G_{i,j}U \) is a basis for \( G_{i,j}A \) and

\[
G_{i,j}U = \cos(\phi)e_i^T U + \sin(\phi)e_j^T U.
\]

From here we can compute the \( i \)-th leverage score of \( G_{i,j}U \),

\[
\ell_i(G_{i,j}U) = \left\| \cos(\phi)e_i^T U + \sin(\phi)e_j^T U \right\|^2 = \cos^2(\phi)\|e_i^T U\|^2 + \sin^2(\phi)\|e_j^T U\|^2 + 2\cos(\phi)\sin(\phi)e_i^T UU^T e_j.
\]

Using the double angle formula for \( \sin \) and recognizing that the row norms are leverage scores yields,

\[
\ell_i(G_{i,j}U) = \cos^2(\phi)\ell_i(U) + \sin^2(\phi)\ell_j(U) + \sin(2\phi)e_i^T UU^T e_j.
\]

Using the Cauchy-Schwarz inequality gives us our upper bound,

\[
\ell_i(G_{i,j}U) \leq \cos^2(\phi)\ell_i(U) + \sin^2(\phi)\ell_j(U) + |\sin(2\phi)| |e_i^T UU^T e_j| \\
= \cos^2(\phi)\ell_i(U) + \sin^2(\phi)\ell_j(U) + |\sin(2\phi)| \sqrt{\ell_i(U)\ell_j(U)}.
\]

A similar process gives us our lower bound,

\[
\ell_i(G_{i,j}U) \geq \cos^2(\phi)\ell_i(U) + \sin^2(\phi)\ell_j(U) - |\sin(2\phi)| \|e_i^T UU^T e_j\| \\
= \cos^2(\phi)\ell_i(U) + \sin^2(\phi)\ell_j(U) - |\sin(2\phi)| \sqrt{\ell_i(U)\ell_j(U)}.
\]