Theorem (deviation bound) Same hypotheses as before, and \( p > 4 \). For all \( u, t > 1 \)

\[
\| (I - P_y) \| \leq \left( 1 + \frac{24}{p+1} \right) \sigma_{k+1} + t \frac{e^k + p}{p+1} \left( \frac{\epsilon}{2}\sigma_j^2 \right)^{1/6} \\
+ ut \frac{e^k + p}{p+1} \sigma_{k+1}
\]

except with probability at most \( 2e^{-t^2/2} \).

For example, set some parameters:

\[
\| (I - P_y) \| \leq \left( 1 + \frac{24}{p+1} \right) \log p \sigma_{k+1} \\
+ 3(e^k + p) \left( \frac{\epsilon}{2}\sigma_j^2 \right)^{1/6}
\]

\( \leq \) Frobenius truncation.

To prove this, we need various results regarding Gaussian matrices. Recall, previously we showed that

\[
\| (I - P_y) \| \leq \| \varepsilon_1 \|^2 + \| \varepsilon_2 \varepsilon_2^T \|^2
\]

in this analysis, \( \varepsilon_1, \varepsilon_2 \) are random, and the result comes from analyzing the 2nd term above.

| Example | \( k \) | \( p \) | Error | Prob
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>4</td>
<td>( 10^{0.01} )</td>
<td>0.01</td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>4</td>
<td>( 12^{0.01} )</td>
<td>0.01</td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>8</td>
<td>( 23^{0.01} )</td>
<td>10^{-7}</td>
<td></td>
</tr>
<tr>
<td>100</td>
<td>10</td>
<td>( 51^{0.01} )</td>
<td>10^{-10}</td>
<td></td>
</tr>
</tbody>
</table>
Probabilistic Error Bounds (worst case scenarios),

i.e. tail bounds.

Thin (deviation bound) Same hypothesis as previous theorem, but also assume \( p > 4 \). Then, for all

\[
\Pr\left(\| (I - P_{u}) A \|_{F} \geq \left( 1 + \sqrt{\frac{3k}{p+1}} \right) \left( \frac{1}{2} \sigma_{u} \right) \right) \\
\leq \frac{ut \frac{e^{k+1}}{p+1} \sigma_{u+1}}{2e^{p} + e^{-\frac{1}{2}}}
\]

In order to show this, we need some measure results for Gaussian matrix:

Let \( G \) be \( k \times (kp) \) Gaussian random matrix with

\[
\mathbb{E}[G_{ij}^{2}] = \frac{1}{p} \\
\mathbb{E}[G_{ij}^{4}] = \frac{3}{p^2}
\]

Let \( p > 4 \), and \( t > 1 \), then

\[
\Pr\left(\| G \|_{F} \geq \sqrt{\frac{3k}{p^2} + t} \right) \leq e^{-t'}
\]

\[
\Pr\left(\| G \|_{2} \geq \frac{e^{k+1}}{p+1} + t \right) \leq e^{-t'}
\]

Mutual relationship to singular value, etc.
Proof:

Define \( E_t \) as \( \{ \mathcal{S}_t : \| x_i^+ \|_2 \leq \frac{e^{\sqrt{k+p}}}{p+1} t \quad \text{and} \quad \| x_i^+ \|_F \leq \frac{3k}{p+1} t \} \)

Using the concentration of measure inequality,

\[
P(E_t^c) = P(\| x_i^+ \|_2 > \frac{e^{\sqrt{k+p}}}{p+1} t \quad \text{or} \quad \| x_i^+ \|_F > \frac{3k}{p+1} t ) \\
\leq P(\| x_i^+ \|_2 > \frac{e^{\sqrt{k+p}}}{p+1} t ) + P(\| x_i^+ \|_F > \frac{3k}{p+1} t ) \\
\leq 2t^{-\rho} + t^{-p}
\]

Next, consider \( h(x) = \| \Sigma_2 x \Sigma_i^+ \|_F \):

\[
|h(x) - h(y)| \leq \| \Sigma_2 x \Sigma_i^+ \|_F - \| \Sigma_2 y \Sigma_i^+ \|_F \\
\leq \| \Sigma_2 [(x-y)+y] \Sigma_i^+ \|_F - \| \Sigma_2 y \Sigma_i^+ \|_F \\
\leq \| \Sigma_2 (x-y) \Sigma_i^+ \|_F + \| \Sigma_2 y \Sigma_i^+ \|_F - \| \Sigma_2 y \Sigma_i^+ \|_F \\
\leq \| \Sigma_2 (x-y) \Sigma_i^+ \|_F \\
\leq \| \Sigma_2 \| \| x-y \|_F \| \Sigma_i^+ \|_F \\
\leq \| \Sigma_2 \| \| x-y \|_F \| \Sigma_i^+ \|_F \leq L 
\]

Thus, \( h \) is Lipschitz with \( L = \| \Sigma_2 \| \| \Sigma_i^+ \|_F \).
Then
\[ E(h(x_2) | x_1) = E(h(x_2 x_1^+ | x_1) \leq \| x_2 \| F \| x_1^+ \| F \| x_1^+ \| \]

Using some known results and conditioning on \( E_t \), we have that
\[ P(\| x_2 x_1^+ \| > E(\| x_2 x_1^+ \|) + LU \mid E_t) \leq e^{-u \frac{1}{2}} \]

But
\[ P(\| x_2 x_1^+ \| > \| x_2 \| F \| x_1^+ \| F + \| x_2 \| x_1^+ \| u \| E_t) \leq e^{-u \frac{1}{2}} \]

So
\[ P(\| x_2 x_1^+ \| > \| x_2 \| F \| x_1^+ \| F + \| x_2 \| x_1^+ \| u \| E_t) \leq e^{-u \frac{1}{2}} \]

But under \( E_t \), we have bounds on \( \| x_2 \| F \) and \( \| x_1^+ \| F \):
\[ P(\| x_2 x_1^+ \| > \| x_2 \| F \frac{e^{\frac{k}{p+1}} + \| x_2 \| F \frac{e^{k+\frac{p}{p+1}}} + \| x_2 \| e^{k+\frac{p}{p+1}} u \mid E_t) \leq e^{-u \frac{1}{2}} \]

Likewise,
\[ P(\| x_2 x_1^+ \| > \| x_2 \| F x_1^+ \mid E_t) \leq e^{-u \frac{1}{2}} \]

\[ P(E_t^c) \leq 2t^p \]

\[ \Rightarrow P(\| x_2 x_1^+ \| > x) \leq e^{-u \frac{1}{2} + 2t^p} \]

And note that if \( \| (I-P) \| \) \( \| x_2 \| F \) \( \| x_1^+ \| F \), then
\[ \| x_2 x_1^+ \| \leq \| x_2 \| F + \| x_1^+ \| F \]
Application: The SVD + analysis

Let \( A \in \mathbb{R}^{m \times n} \).

Choose \( k = p + k \), and compute a rank-\( k \) approximate SVD via the following algorithm:

1. Draw \( \mathbf{Q} \in \mathbb{R}^{m \times k} \).

2. Compute \( Y = \mathbf{A} \mathbf{Q} \) and then \( Y = QR \).

   Then \( A \approx QQ^\top A \); see previous analysis for error.

3. Set \( B = Q^\top A \in k \times n \).

4. Compute SVD factorization directly (Golub-Kahan, etc.).

   \[
   B = \hat{U} \hat{\Sigma} \hat{V}^\top
   \]

5. Set \( \hat{\Sigma} = \Sigma(1:k, 1:k) \)

   \[
   \hat{U} = Q \hat{U}(1:k, 1:k)
   \]

   \[
   \hat{V} = \hat{V}(1:k)
   \]

6. Then \( A \approx U \hat{\Sigma} V^\top \).

Note: Step 4 is a full SVD, which is then truncated. Not obvious how the terms behave, dependent on quality of \( A \), of course.
Analysis

Let $A \in \mathbb{R}^{m \times m}$ with singular values $\sigma_1, \sigma_2, \ldots$, and let $Y \in \mathbb{R}^{m \times l}$ with $l > k$. (In our case $Y = A\Omega = QR$.) Let $\hat{A}(k)$ be a best rank-$k$ approximation of $PA$ in spectral norm (i.e., $Q^TA$ in our case). Then

$$
\| A - \hat{A}(k) \| \leq \sigma_{k+1} + \| (I - P)A \| .
$$

Note in our case, $\hat{A}(k) = Q\tilde{U}(i,k) \tilde{S}(i,k) \tilde{V}(i,k)^T$

$$
= \text{SVD of } Q^TA .
$$

Proof: By the triangle inequality

$$
\| A - \hat{A}(k) \| \leq \| A - P_yA \| + \| P_yA - \hat{A}(k) \| .
$$

(we've detailed the first term)

To analyze this, let $\hat{A}(k)$ be best rank-$k$ approx of $A$.

$$
\Rightarrow \| P_yA - \hat{A}(k) \| \leq \| P_yA - \hat{A}(k) \| ,
$$

since $\hat{A}(k)$ was best approx to $P_yA$.

$$
\Rightarrow \| P_yA - \hat{A}(k) \| \leq \| (I - P_y)(A - \hat{A}(k)) \| \leq \| A - \hat{A}(k) \| = \sigma_{k+1} .
$$
The Power Method

As has been mentioned several times, the performance of the standard randomized range finder scheme can suffer significantly if the singular values of \( A \) decay slowly. The fix is to "basically" apply the algorithm combined with a few iterations of subspace iteration:

Idea: Apply algorithm to \( B = \beta \Theta (A A^*)^q A \)
with \( q \) a small integer.

Note \( B = (U \Sigma V^*)^q U \Sigma V^* \)
\[ = (U \Sigma^2 V^*)^q U \Sigma V^* \]
\[ = U \Sigma^{2q+1} V^* \]

More rapidly decaying singular values for \( A \)

A stable implementation

1. Choose \( k + p \) / draw \( \Omega \in \mathbb{R}^{n \times k} \)
2. Set \( Y_0 = \Omega \), compute \( Y_0 = Q_0 R_0 \)
3. For \( j = 1, \ldots, q \)
   - Set \( Y_j = A^* Q_{j-1} \), compute \( Y_j = Q_j \hat{R}_j \)
   - Set \( Y_j = A \tilde{Q}_{j-1} \), compute \( Y_j = Q_j \tilde{R}_j \)
4. Set \( Q = Q_q \)
Deterministic Error Bound

Let $A \in \mathbb{R}^{m \times n}$, $\Omega \in \mathbb{R}^{n \times d}$, $q > 0$, $B = (AA^*)^q A$, and $Z = B \Omega$. Then
\[ \| (I-P_z) A \| \leq \| (I-P_z) B \| \frac{1}{\sqrt{2^{q+1}}} \]

Proof:
\[ \| (I-P_z) A \| \leq \| (I-P_z) (AA^*)^q A \| \frac{1}{\sqrt{2^{q+1}}} = \| (I-P_z) B \| \frac{1}{\sqrt{2^{q+1}}} \]

Because (see Prop 8.6 for details.)

On its face it looks bad because $\| \cdot \|$ might possibly grow (e.g. $\sqrt{0.1} = 1$). But put in context of earlier result:

If $Y = AZ$, then previously we showed:
\[ \| (I-P_z) A \| \leq (1 + \| \Omega \|_2^2 \| \Omega \|_2) \frac{1}{\sqrt{2^{q+1}}} \frac{1}{\sqrt{s_{k+1}}} \]

But if $B = (AA^*)^q A$, then $s_{k+1} \rightarrow \infty$ and
\[ \| (I-P_z) A \| \leq \| (I-P_z) B \| \frac{1}{\sqrt{2^{q+1}}} \leq \left[ \frac{\sqrt{2^{q+1}}}{s_{k+1}} + \| \Omega \|_2^2 \| \Omega \|_2 \| \frac{1}{\sqrt{2^{q+1}}} \right] \frac{1}{\sqrt{s_{k+1}}} \]
\[ = \left( 1 + \| \Omega \|_2^2 \| \Omega \|_2 \right) \frac{1}{\sqrt{2^{q+1}}} \frac{1}{\sqrt{s_{k+1}}} \]
\[ \rightarrow 1 \text{ exponentially fast as } q \rightarrow \infty. \]
Average Case Behavior for Power Scheme.

Corollary Let \( B = (A A^*)^q A \) and \( Z = A \Omega, \quad \Omega \in \mathbb{R}^{n \times k} \)
standard Gaussian test matrix. Then

\[
\mathbb{E}(\| (I - P_2) A \|) \leq \left[ (1 + \frac{k}{p+1}) S_{w1}^{2q+1} + \frac{e^{k+p}}{p} \left( \sum_{i=1}^{2} S_{i}^{2q+1} \right)^{1/2} \right]^{1/2q+1}
\]

skip proof...

A more illuminating form:

\[
\mathbb{E}(\| (I - P_2) A \|) \leq \left( 1 + \frac{k}{p+1} + \frac{e^{k+p}}{p} \cdot \min(m,n) - k \right) \frac{1}{\sqrt{k+1}} \to 1 \text{ as } q \to \infty.
\]

c.i.e. \( q \text{ as } q \to \log(\min(m,n)) \)

Applications Hierarchical Matrix Compression