EXTRACTING A BASIS WITH FIXED BLOCK INSIDE A MATRIX

PIERRE YOUSSEF

ABSTRACT. Given a matrix of rank and a block of columns inside , we consider the problem of extracting a block of columns which minimize the Hilbert-Schmidt norm of the inverse while preserving the block . This generalizes a previous result of Gluskin-Olevskii, and improves the estimates when given a "good" block .

1. INTRODUCTION

Let be an matrix, we see as an operator from to . We denote the operator norm of while the Hilbert-Schmidt norm of is given by .

Given , we denote the restriction of to the columns with indices in i.e. , where is the canonical coordinate projection.

Extracting a block of columns inside satisfying some properties reveals to be an important problem. Different properties of the restriction can be studied. In [2], the problem of extracting a block of columns which approximates the singular values of was investigated; it was proven that a number of columns proportional to is sufficient in order to achieve the approximation. Batson-Spielman-Srivastava developed a deterministic algorithm which selects these columns. In [6], for every fixed size , Kashin-Tzafriri considered selecting a block of columns which minimizes the operator norm among all restrictions of the same size; in [9], a randomized algorithm to achieve the selection was given while in [12], we were able to give a deterministic algorithm which slightly improves the result. In [3], Bourgain-Tzafriri considered selecting a block of columns which is well invertible i.e. whose smallest singular value is bounded away from zero; their result only dealt with square matrices and the extension to rectangular matrices was achieved by Spielman-Srivastava [8] who produced a deterministic algorithm to find the extraction. In [12], a generalization of this result was given by a similar algorithm. In [10], extracting a well conditioned block was considered i.e. a block which is well invertible and have a "good" operator norm; this result was improved in [11] with the advantage of producing a deterministic algorithm.

In this paper, we are interested in another column selection problem. Let be an matrix of rank , we want to extract an invertible matrix inside which minimizes the Hilbert-Schmidt norm of the inverse. The invertibility question is closely related to the restricted invertibility studied in [3] and [8]; the difference here is that we want to extract exactly rank columns while in the restricted invertibility principle, one is only allowed to extract a number of columns strictly less than . This problem was studied by Gluskin-Olevskii [4] who proved the following:

Theorem 1.1 (Gluskin-Olevskii). Let be an matrix of rank . Then there exists of size such that is invertible and

where . This problem was studied by Gluskin-Olevskii [4] who proved the following:

The proof of Gluskin-Olevskii is only existential. At first, we were interested in giving a deterministic algorithm achieving the extraction. This can be done by carefully removing the
"bad" columns of $U$. However, after doing so, it turned out that other algorithms were already established in \cite{1}. We include our algorithmic proof in the Appendix.

Our aim in this paper is to generalize Theorem 1.1 in different ways. The first generalization is considering the same problem with the constraint of keeping inside the extraction a block chosen at the beginning. This can be useful if we already know the existence of a "good" block inside $U$. In that case, we want to keep this block and try to complement it instead of doing the full extraction. This has the advantage of reducing the complexity of the algorithm and improve the estimate on the Hilbert-Schmidt norm of the inverse of the restriction.

The second generalization concerns arbitrary rank matrices; indeed, Theorem 1.1 states that given $UU^t = \sum_{i \in \mathbb{N}} u_i u_i^t$, there exists $\sigma \subset \{1, ..., m\}$ of size $n$ such that $\sum_{i \in \sigma} u_i u_i^t$ is invertible and

$$\text{Tr} \left( \sum_{i \in \sigma} u_i u_i^t \right)^{-1} \leq (m - n + 1) \cdot \text{Tr} \left( UU^t \right)^{-1}.$$

Whereas Theorem 1.1 only deals with rank one matrices, we consider the same problem when replacing the rank one matrices by arbitrary rank matrices.

The main result of this paper is the following

\textbf{Theorem 1.2.} \textit{Let $A, B, B_1, ..., B_m$ be $n \times n$ positive semidefinite matrices such that $A = B + \sum_{i \leq m} B_i$ is of rank $n$. Then for any $k \geq n - \lfloor \text{Tr} \left( A^{-1} \right) \rfloor$, there exists $\sigma_k$ of size $k$ such that

$$\text{Tr} \left( A_{\sigma_k}^{-1} \right) \leq \text{Tr} \left( A^{-1} \right) \cdot \left[ \frac{m - n + \text{Tr} \left( A^{-1} \right) + 1}{k - n + 1 + \text{Tr} \left( A^{-1} \right)} \right] - \frac{(m - k) \cdot \text{Tr} \left( A^{-2} \right)}{k - n + 1 + \text{Tr} \left( A^{-1} \right)},$$

where $A_{\sigma_k} = B + \sum_{i \in \sigma_k} B_i$.}

\textit{In particular, there exists $\sigma \subset \{1, ..., m\}$ with $|\sigma| = n - \lfloor \text{Tr} \left( A^{-1} \right) \rfloor$ such that

$$\text{Tr} \left( A_{\sigma}^{-1} \right) \leq \left( m - n + 1 + \text{Tr} \left( A^{-1} \right) \right) \cdot \text{Tr} \left( A^{-1} \right) - \text{Tr} \left( A^{-2} \right) \left( m - n + \text{Tr} \left( A^{-1} \right) \right),$$

where $A_{\sigma} = B + \sum_{i \in \sigma} B_i$.}

In section 2, we will discuss how this result generalizes Theorem 1.1 and state direct consequences of it. The proof of Theorem 1.2 will be given in section 3.

\section{2. Derived results}

Let us first derive the rank one case of Theorem 1.2. This case is interesting since the selection can be traced as a selection of columns inside the matrix. Preserving the matrix $B$ in Theorem 1.2 plays the role of preserving a block inside the matrix. Precisely, we have the following:

\textbf{Theorem 2.1.} \textit{Let $U$ be an $n \times m$ matrix of rank $n$ and $\nu \subset \{1, ..., m\}$. Denote by $V$ the block of columns inside $U$ corresponding to $\nu$ i.e. $V = U_{\nu}$. Let $A = UU^t$ and $B = VV^t$. For any $k \geq n - \lfloor \text{Tr} \left( A^{-1} \right) \rfloor$, there exists $\sigma_k' \subset \nu^c$ of size $k$ such that if $\sigma_k = \sigma_k' \cup \nu$ then $U_{\sigma_k}$ is of rank $n$ and

$$\text{Tr} \left( U_{\sigma_k} U_{\sigma_k}^t \right)^{-1} \leq \text{Tr} \left( A^{-1} \right) \cdot \left[ \frac{|\nu^c| - n + \text{Tr} \left( A^{-1} \right) + 1}{k - n + 1 + \text{Tr} \left( A^{-1} \right)} \right] - \frac{(|\nu^c| - k) \cdot \text{Tr} \left( A^{-2} \right)}{k - n + 1 + \text{Tr} \left( A^{-1} \right)}.$$}

\footnote{We would like to thank Nick Harvey for pointing out to us this reference.}
In particular, there exists $\sigma' \subset \nu^c$ with
\[ |\sigma'| = n - \left| \text{Tr} \left( A^{-1}B \right) \right|, \]
such that
\[ \text{Tr} \left( U_\sigma U_\sigma^t \right)^{-1} \leq \left( |\nu^c| - n + 1 + \text{Tr} \left( A^{-1}B \right) \right) \cdot \text{Tr} \left( A^{-1} \right) - \text{Tr} \left( A^{-2}B \right) \left( |\nu^c| - n + \text{Tr} \left( A^{-1}B \right) \right), \]
where $\sigma = \sigma' \cup \nu$.

**Proof.** Denote by $(u_i)_{i \leq m}$ the columns of $U$. Clearly, $UU^t = \sum_{i \leq m} u_i u_i^t$. For any $i \in \nu$, define $B_i = u_i u_i^t$. Then we have $A = B + \sum_{i \in \nu} B_i$. Theorem 2.1 follows by applying Theorem 1.2 to $A, B, (B_i)_{i \in \nu}$, with $m$ replaced by $|\nu^c|$. \( \square \)

Theorem 2.1 allows one to preserve the block $V$ while achieving the desired extraction. Let us illustrate how this can be useful. For this aim, consider the case where $UU^t = Id$.

**Corollary 2.2.** Let $U$ be an $n \times m$ matrix such that $UU^t = Id$. Suppose that $U$ contains $r$ columns of norm 1 for some $r < n$. Then there exists $\sigma \subset \{1, \ldots, m\}$ of size $n$ such that $U_\sigma$ is invertible and
\[ \|U_\sigma^{-1}\|_{\text{HS}}^2 \leq (m - n) \cdot (n - r) + n. \]

**Proof.** Suppose that $UU^t = Id$ and that $U$ contains $r$ columns of norm 1. Let $\nu \subset \{1, \ldots, m\}$ be the set of indices of the norm one columns and denote $V = U_{\nu}$. Clearly, $V$ is an $n \times r$ matrix of rank $r$ since $r = \|V\|_{\text{HS}}^2 \leq \|V\|^2 \cdot \text{rank}(V)$ and $\|V\| \leq 1$. By preserving $V$, we already have $r$ linearly independent columns and therefore we need just to complement them by $n - r$ linearly independent columns.

Following the notations of Theorem 2.1, we have $A = Id$ and $B = V V^t$. Applying Theorem 2.1, one can find $\sigma'$ with
\[ |\sigma'| = n - \left| \text{Tr} \left( A^{-1}B \right) \right| = n - r, \]
such that if $\sigma = \sigma' \cup \nu$, then $|\sigma| = n$ and
\[ \|U_\sigma^{-1}\|_{\text{HS}}^2 \leq \left( |\nu^c| - n + 1 + \text{Tr} \left( A^{-1}B \right) \right) \cdot \text{Tr} \left( A^{-1} \right) - \text{Tr} \left( A^{-2}B \right) \left( |\nu^c| - n + \text{Tr} \left( A^{-1}B \right) \right) \]
\[ = (m - r - n + 1 + r) \cdot n - r \cdot (m - r - n + r) \]
\[ = (m - n) \cdot (n - r) + n. \]

\( \square \)

**Remark 2.3.** When we don’t look for columns of norm 1 inside $U$ (i.e. the case where $r = 0$), the estimate would be the same as in Theorem 1.1 while if $r \neq 0$, the estimate of Corollary 2.2 improves the one of Theorem 1.1 by $r \cdot (m - n)$. Moreover, one can view Corollary 2.2 as an extraction of a basis inside $U$ with a "good" estimate on the Hilbert-Schmidt of the inverse while preserving a "nice" block.

To have a better understanding of the generalization of Theorem 1.1 to the case of arbitrary rank matrices, let us take $B = 0$ in Theorem 1.2 to get the following corollary:

**Corollary 2.4.** Let $A, B_1, \ldots, B_m$ be $n \times n$ positive semidefinite matrices such that $A = \sum_{i \leq m} B_i$ is of rank $n$. Then for any $k \geq n$, there exists $\sigma \subset \{1, \ldots, m\}$ of size $k$ such that $\sum_{i \in \sigma} B_i$ is invertible and
\[ \text{Tr} \left( \sum_{i \in \sigma} B_i \right)^{-1} \leq \frac{m - n + 1}{k - n + 1} \text{Tr} \left( A^{-1} \right). \]
3. Proof of Theorem 1.2

The proof is based on an iteration of the following lemma:

**Lemma 3.1.** Let $A_p, B, B_1, \ldots, B_p$ be $n \times n$ positive semidefinite matrices such that $A_p = B + \sum_{i \leq p} B_i$ is of rank $n$. Then there exists $j \in \{1, \ldots, p\}$ such that $A_p - B_j$ is of rank $n$ satisfying $(A_p - B_j)^{-1} \succeq A_p^{-1}$ and

$$
\text{Tr} (A_p - B_j)^{-1} \leq \text{Tr} (A_p^{-1}) \cdot \left[ \frac{p - n + \text{Tr} (A_p^{-1} B) + 1}{p - n + \text{Tr} (A_p^{-1} B)} \right] - \frac{\text{Tr} (A_p^{-2} B)}{p - n + \text{Tr} (A_p^{-1} B)}.
$$

**Proof.** Our aim is to find $C$ among the $(B_i)_{i \leq p}$ such that $A_p - C$ is still invertible and then have a control on $\text{Tr} (A_p - C)^{-1}$. We would like to use the Sherman-Morrison-Woodbury which states that

$$
(A_p - C)^{-1} = A_p^{-1} + A_p^{-1} C \left( Id - C A_p^{-1} C \right)^{-1} C A_p^{-1}.
$$

For this formula to hold, we should ensure choosing $C$ such that $Id - C A_p^{-1} C$ is invertible. Since

$$
C A_p^{-1} C \preceq \text{Tr} (A_p^{-1} C) \cdot Id,
$$

then it would be sufficient to have $1 - \text{Tr} (A_p^{-1} C) > 0$ in order to use (1). Since $\text{Tr} (A_p^{-2} C)$ is positive then we may search for $C$ satisfying

$$
\text{Tr} (A_p^{-2} C) \leq \alpha \left( 1 - \text{Tr} (A_p^{-1} C) \right),
$$

where $\alpha = \frac{\text{Tr} (A_p^{-1}) - \text{Tr} (A_p^{-2} B)}{p - n + \text{Tr} (A_p^{-1} B)}$.

In order to guarantee the existence of $C$ among the $(B_i)_{i \leq p}$ that satisfies (2), it is sufficient to prove that (2) holds when taking the sum over all $(B_i)_{i \leq p}$. Therefore, we need to prove that

$$
\sum_{i \leq p} \text{Tr} (A_p^{-2} B_i) \leq \alpha \left( p - \sum_{i \leq p} \text{Tr} (A_p^{-1} B_i) \right).
$$

Now since the trace is linear and $\sum_{i \leq p} B_i = A_p - B$ then (3) is equivalent to the following

$$
\text{Tr} (A_p^{-1}) - \text{Tr} (A_p^{-2} B) \leq \alpha \left( p - n + \text{Tr} (A_p^{-1} B) \right),
$$

which is true by the choice of $\alpha$. Therefore, there exists $j \leq p$ such that $C = B_j$ satisfies (2).

Now, we may use the Sherman-Morrison-Woodbury formula (1) and write

$$
(A_p - B_j)^{-1} = A_p^{-1} + A_p^{-1} B_j \left( Id - B_j A_p^{-1} B_j \right)^{-1} B_j A_p^{-1} \preceq A_p^{-1} + \frac{A_p^{-1} B_j A_p^{-1}}{1 - \text{Tr} (A_p^{-1} B_j)}.
$$

Now since $B_j$ satisfies (2), then

$$
A_p^{-1} \preceq (A_p - B_j)^{-1} \preceq A_p^{-1} + \alpha \frac{A_p^{-1} B_j A_p^{-1}}{\text{Tr} (A_p^{-2} B_j)}.
$$

Taking the trace in (4), we get

$$
\text{Tr} (A_p - B_j)^{-1} \leq \text{Tr} (A_p^{-1}) + \alpha.
$$

Replacing $\alpha$ with its value, we finish the proof of the lemma.
Proof of Theorem 7.2 \  The proof will be based on an iteration of the previous lemma.

We will construct the set $\sigma$ step by step, starting with $\sigma_0 = \{1, \ldots, m\}$ and at each time taking away the "bad" indices. Denote by $A_0 = A = B + \sum_{i \in \sigma_0} B_i$. Apply Lemma 3.1 to find $j_0 \in \sigma_0$ such that $A_1 := A_0 - B_{j_0}$ is of rank $n$ satisfying $A_1^{-1} \succeq A_0^{-1}$ and

\[
(5) \quad \text{Tr} \left( A_1^{-1} \right) \leq \text{Tr} \left( A_0^{-1} \right) \cdot \left[ \frac{|\sigma_0| - n + \text{Tr}(A_0^{-1}B) + 1}{|\sigma_0| - n + \text{Tr}(A_0^{-1}B)} \right] - \frac{\text{Tr}(A_0^{-1}B)}{|\sigma_0| - n + \text{Tr}(A_0^{-1}B)}.
\]

Let $\sigma_1 = \sigma_0 \setminus \{j_0\}$, then $A_1 = B + \sum_{i \in \sigma_1} B_i$ and $|\sigma_1| = m - 1$. Now apply Lemma 3.1 again in order to find $j_1 \in \sigma_1$ such that $A_2 := A_1 - B_{j_1}$ is of rank $n$ satisfying $A_2^{-1} \succeq A_1^{-1}$ and

\[
(6) \quad \text{Tr} \left( A_2^{-1} \right) \leq \text{Tr} \left( A_1^{-1} \right) \cdot \left[ \frac{|\sigma_1| - n + \text{Tr}(A_1^{-1}B) + 1}{|\sigma_1| - n + \text{Tr}(A_1^{-1}B)} \right] - \frac{\text{Tr}(A_1^{-1}B)}{|\sigma_1| - n + \text{Tr}(A_1^{-1}B)}.
\]

Let $\sigma_2 = \sigma_1 \setminus \{j_1\}$, then $A_2 = B + \sum_{i \in \sigma_2} B_i$ and $|\sigma_2| = m - 2$. Combining (5) and (6), we have

\[
(7) \quad \text{Tr} \left( A_2^{-1} \right) \leq \text{Tr} \left( A_1^{-1} \right) \cdot \left[ \frac{|\sigma_1| - n + \text{Tr}(A_1^{-1}B) + 1}{|\sigma_1| - n + \text{Tr}(A_1^{-1}B)} \right] - \frac{2\text{Tr}(A_1^{-1}B)}{|\sigma_1| - n + \text{Tr}(A_1^{-1}B)}.
\]

Suppose that we constructed $\sigma_p$ of size $m - p$ such that $A_p = B + \sum_{i \in \sigma_p} B_i$ satisfies $A_p^{-1} \succeq A_0^{-1}$ and

\[
(8) \quad \text{Tr} \left( A_p^{-1} \right) \leq \text{Tr} \left( A_0^{-1} \right) \cdot \left[ \frac{|\sigma_p| - n + \text{Tr}(A_0^{-1}B) + 1}{|\sigma_p| - n + \text{Tr}(A_0^{-1}B)} \right] - \frac{p\text{Tr}(A_0^{-1}B)}{|\sigma_p| - n + \text{Tr}(A_0^{-1}B)}.
\]

Applying Lemma 3.1 we find $j_p \in \sigma_p$ such that $A_{p+1} := A_p - B_{j_p}$ is of rank $n$ satisfying $A_{p+1}^{-1} \succeq A_p^{-1}$ and

\[
(9) \quad \text{Tr} \left( A_{p+1}^{-1} \right) \leq \text{Tr} \left( A_p^{-1} \right) \cdot \left[ \frac{|\sigma_p| - n + \text{Tr}(A_p^{-1}B) + 1}{|\sigma_p| - n + \text{Tr}(A_p^{-1}B)} \right] - \frac{\text{Tr}(A_p^{-1}B)}{|\sigma_p| - n + \text{Tr}(A_p^{-1}B)}.
\]

Let $\sigma_{p+1} = \sigma_p \setminus \{j_p\}$, then $A_{p+1} = B + \sum_{i \in \sigma_{p+1}} B_i$ and $|\sigma_{p+1}| = m - p - 1$. Combining (8) and (9), we have

\[
(10) \quad \text{Tr} \left( A_{p+1}^{-1} \right) \leq \text{Tr} \left( A_p^{-1} \right) \cdot \left[ \frac{|\sigma_p| - n + \text{Tr}(A_p^{-1}B) + 1}{|\sigma_p| - n + \text{Tr}(A_p^{-1}B)} \right] - \frac{(p+1)\text{Tr}(A_p^{-1}B)}{|\sigma_p| - n + \text{Tr}(A_p^{-1}B)}.
\]

We can continue this procedure as long as $|\sigma_p| \geq n - \text{Tr}(A_0^{-1}B) + 1$. Therefore, we have proved by induction that for any $l \leq m - n + \left\lfloor \text{Tr}(A_0^{-1}B) \right\rfloor$, there exists $\sigma_l$ of size $m - l$ such that

\[
\text{Tr} \left( A_{\sigma_l}^{-1} \right) \leq \text{Tr} \left( A_0^{-1} \right) \cdot \left[ \frac{m - n + \text{Tr}(A_0^{-1}B) + 1}{m - n + 1 + \text{Tr}(A_0^{-1}B) - l} \right] - \frac{l \cdot \text{Tr}(A_0^{-1}B)}{m - n + 1 + \text{Tr}(A_0^{-1}B) - l}.
\]

where $A_{\sigma_l} = B + \sum_{i \in \sigma_l} B_i$.

\[\square\]
Acknowledgement: The author would like to thank Christos Boutsidis for his valuable comments regarding this manuscript.

REFERENCES

[1] H. Avron and C. Boutsidis. Faster Subset Selection for Matrices and Applications. SIAM J. Matrix Anal. Appl., 34(4):1464–1499, 2013.
[2] J. D. Batson, D. A. Spielman, and N. Srivastava. Twice-Ramanujan sparsifiers. In STOC’09—Proceedings of the 2009 ACM International Symposium on Theory of Computing, pages 255–262. ACM, New York, 2009.
[3] J. Bourgain and L. Tzafriri. Invertibility of “large” submatrices with applications to the geometry of Banach spaces and harmonic analysis. Israel J. Math., 57(2):137–224, 1987.
[4] E. Gluskin and A. Olevskii. Invertibility of sub-matrices and the octahedron width theorem. Israel J. Math., 186:61–68, 2011.
[5] N. Harvey and N. Olver. Pipage rounding, pessimistic estimators and matrix concentration. Available at arXiv:1307.2274.
[6] B. Kashin and L. Tzafriri. Some remarks on the restriction of operators to coordinate subspaces. preprint.
[7] A. Marcus, D. Spielman, and N. Srivasatava. Interlacing families ii: Mixed characteristic polynomials and the kadosin-singer problem. Available at arXiv:1306.3969v3.
[8] D. A. Spielman and N. Srivastava. An elementary proof of the restricted invertibility theorem. Israel J. Math., 190:83–91, 2012.
[9] J. A. Tropp. Column subset selection, matrix factorization, and eigenvalue optimization. In Proceedings of the Twentieth Annual ACM-SIAM Symposium on Discrete Algorithms, pages 978–986, Philadelphia, PA, 2009. SIAM.
[10] R. Vershynin. John’s decompositions: selecting a large part. Israel J. Math., 122:253–277, 2001.
[11] P. Youssef. A note on column subset selection. International Mathematics Research Notices, 2013.
[12] P. Youssef. Restricted invertibility and the banach-mazur distance to the cube. Mathematika, FirstView:1–18, 9 2013.

4. APPENDIX

We will present here an algorithmic proof of Theorem 1.1. Our proof is inspired by the tools developed in [2] but the procedure will be quite opposite. Write $UU^t = \sum_{i \leq m} u_i u_i^t$, where $(u_i)_{i \leq m}$ are the columns of $U$. The aim is to construct $A_\sigma = \sum_{i \in \sigma} u_i u_i^t$ which satisfies the conclusion needed. In [2], the construction is done step by step starting from $A_0 = 0$ and studying the evolution of the eigenvalues when adding a suitable rank one matrix. Our construction will also be done step by step, starting however with $A_0 = A$ and studying the evolution of the Hilbert-Schmidt norm of the inverse when subtracting a suitable rank one matrix. Precisely, at each step we will remove from $U$ the "bad" columns until we have $n$ remaining "good" linearly independent columns.

The proof is an iteration of the following Lemma:

**Lemma 4.1.** Let $U$ be an $n \times m$ matrix of rank $n$. There exists $\sigma \subset \{1, \ldots, m\}$ of size $m - 1$ such that $U_\sigma$ is of rank $n$ and

$$
\text{Tr} \left( U_\sigma U_\sigma^t \right)^{-1} \leq \frac{m - n + 1}{m - n} \text{Tr} \left( UU^t \right)^{-1}.
$$

**Proof.** Denote $A = UU^t = \sum_{i \leq m} u_i u_i^t$, where $(u_i)_{i \leq m}$ are the columns of $U$. We are searching for vector $v$ chosen among the columns of $U$ such that $A - vv^t$ is still invertible and has a control on the Hilbert-Schmidt norm of its inverse. We would like to use the Sherman-Morrison formula which states that if $v^t A^{-1}v \neq 1$ then

$$
\text{Tr} \left( A - vv^t \right) = \text{Tr} \left( A^{-1} \right) + \frac{v^t A^{-2}v}{1 - v^t A^{-1}v}.
$$

For that, we will search for $v$ such that $v^t A^{-1}v < 1$. Since $v^t A^{-2}v$ is positive, then it is sufficient to search for $v$ satisfying

$$
v^t A^{-2}v \leq \alpha \cdot \left( 1 - v^t A^{-1}v \right),
$$

for some $\alpha > 0$.
where \( \alpha = \frac{\text{Tr}(A^{-1})}{m-n} \). To guarantee that such \( v \) exists, we may prove that (12) holds when taking the sum over all columns of \( U \)

\[
\sum_{i \leq m} u_i^t A^{-2} u_i \leq \alpha \cdot \left( m - \sum_{i \leq m} u_i^t A^{-1} u_i \right).
\]

This is equivalent to the following

\[
\text{Tr} \left( A^{-2} \sum_{i \leq m} u_i^t u_i^t \right) \leq \alpha \cdot \left( m - \text{Tr} \left( A^{-1} \sum_{i \leq m} u_i^t u_i^t \right) \right).
\]

Since \( A = \sum_{i \leq m} u_i^t u_i^t \), it is then reduced to prove that

\[
\text{Tr} \left( A^{-1} \right) \leq \alpha \cdot (m - n),
\]

which is true by the choice of \( \alpha \). Therefore we have found \( j \in \{1, \ldots, m\} \) such that \( u_j \) satisfies (12). We may now use (11) to get

\[
\text{Tr} \left( A - u_j u_j^t \right) = \text{Tr} \left( A^{-1} \right) + \frac{u_j^t A^{-2} u_j}{1 - u_j^t A^{-1} u_j} \leq \frac{m - n + 1}{m - n} \text{Tr} \left( A^{-1} \right).
\]

The Lemma follows by taking \( \sigma = \{1, \ldots, m\} \setminus \{j\} \).

\[\Box\]

**Remark 4.2.** In [1], Avron-Boutsidis gave a different proof of Lemma 4.1. Their proof produces an algorithm based on volume sampling. We refer the interested reader to [1] for more details.

**Proof of Theorem 1.1.** Start with \( A_0 = A = U U^t \) and apply Lemma 4.1 to find \( \sigma_1 \) of size \( m - 1 \) such that \( A_1 = U_{\sigma_1} U_{\sigma_1}^t \) is of rank \( n \) and satisfies

\[
\text{Tr} \left( A_1^{-1} \right) \leq \frac{m - n + 1}{m - n} \text{Tr} \left( A_0^{-1} \right).
\]

Now apply Lemma 4.1 again with \( U_{\sigma_1} \) to find \( \sigma_2 \) of size \( m - 2 \) such that \( A_2 = U_{\sigma_2} U_{\sigma_2}^t \) is of rank \( n \) and satisfies

\[
\text{Tr} \left( A_2^{-1} \right) \leq \frac{m - n + 1}{m - n - 1} \text{Tr} \left( A_1^{-1} \right) \leq \frac{m - n + 1}{m - n - 1} \text{Tr} \left( A_0^{-1} \right).
\]

If we continue this procedure, after \( k \) steps we find \( \sigma_k \) of size \( m - k \) such that \( A_k = U_{\sigma_k} U_{\sigma_k}^t \) is of rank \( n \) and satisfies

\[
\text{Tr} \left( A_k^{-1} \right) \leq \frac{m - n + 1}{m - n - k + 1} \text{Tr} \left( A_0^{-1} \right).
\]

This holds for any \( k \leq m - n \). After \( k = m - n \) steps, Theorem 1.1 is proved.

\[\Box\]

**Remark 4.3.** One can check that the running time of the algorithm is \( O(mn^2(m - n)) \). Slightly faster algorithms were developed in [1].