Deterministic column subset selection for single-cell RNA-Seq: Supplementary Material

Shannon R. McCurdy\textsuperscript{1*}, Vasilis Ntranos\textsuperscript{2}, Lior Pachter\textsuperscript{3}

\textsuperscript{1} California Institute for Quantitative Biosciences, University of California Berkeley, Berkeley, California, United States of America
\textsuperscript{2} Department of Electrical Engineering and Computer Sciences, University of California Berkeley, Berkeley, California, United States of America
\textsuperscript{3} Division of Biology and Biological Engineering, Department of Computing and Mathematical Sciences, California Institute of Technology, Pasadena, California, United States of America

* Corresponding author
E-mail: smccurdy@berkeley.edu (SRM)

A Additional figures

\textbf{Figure A.} Average spectral clustering ARI for nine clusters for DCSS, count, variance, and index of dispersion thresholding on the data matrix from the mouse cortex scRNA-Seq experiment \cite{1} and the clustering workflow of \cite{2}. We vary the error tolerance $\epsilon$ with $k = 5$ for DCSS. Increasing the error tolerance decreases the agreement between clusters.

\textbf{Figure B.} Average spectral clustering ARI for nine clusters for DCSS, count, variance, and index of dispersion thresholding on the data matrix from the mouse cortex scRNA-Seq experiment \cite{1} and the clustering workflow of \cite{2}. We vary the dimension $k$ with fixed error tolerance $\epsilon = 0.1$ for DCSS. Increasing the dimension increases the agreement between clusters.
The **singular value decomposition** (SVD) of any complex matrix \( A \) is \( A = U \Sigma V^\dagger \), where \( U \) and \( V \) are square unitary matrices (\( U^\dagger U = UU^\dagger = I \), \( V^\dagger V = VV^\dagger = I \)), \( \Sigma \) is a rectangular diagonal matrix with real non-negative non-increasingly ordered entries. \( U^\dagger \) is the complex conjugate and transpose of \( U \), and \( I \) is the identity matrix. The diagonal elements of \( \Sigma \) are called the **singular values**, and they are the positive square roots of the eigenvalues of both \( AA^\dagger \) and \( A^\dagger A \), which have eigenvectors \( U \) and \( V \), respectively. \( U \) and \( V \) are the left and right **singular vectors** of \( A \).

Defining \( U_k \) as the first \( k \) columns of \( U \) and analogously for \( V \), and \( \Sigma_k \) the square diagonal matrix with the first \( k \) entries of \( \Sigma \), then \( A_k = U_k \Sigma_k V_k^\dagger \) is the rank-\( k \) SVD approximation to \( A \), and \( T_k = AV_k = U_k \Sigma_k \) is a rank-\( k \) SVD truncation of \( A \). Furthermore, we refer to matrix with only the last \( n - k \) columns of \( U \), \( V \) and last \( n - k \) entries in \( \Sigma \) as \( U \_k \), \( V \_k \), and \( \Sigma \_k \).

The Moore-Penrose pseudo inverse of a rank \( r \) matrix \( A \) is given by \( A^+ = V \Sigma_r^{-1} U^\dagger \). The Frobenius norm \( ||A||_F \) of a matrix \( A \) is given by \( ||A||_F = \sqrt{\text{tr}(AA^\dagger)} \). Recall that the trace has a cyclic property. The spectral norm \( ||A||_2 \) of a matrix \( A \) is given by the largest singular value of \( A \).

The Eckart-Young-Mirsky theorem [4] states that, for \( A = U \Sigma V^\dagger \) the SVD of \( A \), and \( B \) any complex matrix with compatible dimension to \( A \) and rank \( \leq k \),

\[
A_k = \arg\min\_{\text{rank}(B) \leq k} ||A - B||_F
\]

min \( \text{rank}(B) \leq k \)

\[
||A - B||_F = \sqrt{\text{tr}(\Sigma \_k \Sigma \_k^T)}.
\]  

(S1)

The minimizer \( A_k \) is unique if and only if \( \sigma_{k+1} \neq \sigma_k \), where \( \sigma_i \) are the respective non-increasingly ordered singular values in \( \Sigma \).

A square complex matrix \( F \) is **Hermitian** if \( F = F^\dagger \). Symmetric positive semi-definite (S.P.S.D) matrices are Hermitian matrices. The set of \( n \times n \) Hermitian matrices is a real linear space. As such, it is possible to define a *partial ordering* (also called a Loewner partial ordering, denoted by \( \preceq \)) on the real linear space. One matrix is “greater” than another if their difference lies in the closed convex cone of S.P.S.D. matrices. Let \( F, G \) be Hermitian and the same size, and \( x \) a complex vector of compatible dimension. Then,

\[ F \preceq G \iff x^\dagger F x \leq x^\dagger G x \quad \forall x \neq 0. \]  

(S2)

A few simple consequences of the Loewner partial ordering are as follows. If \( F \) is Hermitian and S.P.S.D., then \( 0 \preceq F \), where \( 0 \) is the zero matrix.

If \( F \) is Hermitian with smallest and largest eigenvalues \( \lambda_{\text{min}}(F), \lambda_{\text{max}}(F) \), respectively, then,

\[ \lambda_{\text{min}}(F) I \preceq F \preceq \lambda_{\text{max}}(F) I. \]  

(S3)

Let \( F, G \) be Hermitian and the same size, and let \( H \) be any complex rectangular matrix of compatible dimension. The **conjugation rule** is,

\[ \text{If } F \preceq G, \text{ then } HFH^\dagger \preceq HGH^\dagger. \]  

(S4)

In addition, let \( \lambda_i(F) \) and \( \lambda_i(G) \) be the non-decreasingly ordered eigenvalues of \( F, G \). Then,

\[ \text{If } F \preceq G, \text{ then } \forall i, \lambda_i(F) \leq \lambda_i(G). \]  

(S5)
Since the trace of a matrix $F$ is the sum of its eigenvalues, $\text{tr} F = \sum_i \lambda_i(F)$, and the Loewner ordering implies the ordering of eigenvalues (Eq S5), the Loewner ordering also implies the ordering of their sum,

$$\text{If } F \preceq G, \text{ then } \text{tr} F \leq \text{tr} G. \quad (S6)$$

Let $F_1, G_1, F_2, G_2$ be Hermitian and the same size. Then if $F_1 \preceq G_1$ and $F_2 \preceq G_2$, then

$$F_1 + F_2 \preceq G_1 + G_2. \quad (S7)$$

As a simple consequence of Eq S2 consider the real matrices $FF^T$ and $GG^T$, and the vector $x$ which has a one in row $i$ and a minus one in row $j$, and zeros elsewhere. The Euclidean distance between rows $i,j$ with respect to $G$ is $d_{i,j}(G)$:

$$d_{i,j}(G) = x^T GG^T x. \quad (S8)$$

Thus, if $FF^T \preceq GG^T$, by Eq S2 with appropriate vectors, $d_{i,j}(F) \leq d_{i,j}(G) \forall i,j$. Furthermore, let $F$ be Hermitian and dimension $n$, $U_k$ be a semi-orthogonal rectangular matrix $(U_k^H U_k = I)$ of compatible dimension $n \times k$, $1 \leq k \leq n$, and $\lambda_i(M)$ refer to the non-decreasingly ordered eigenvalues of a matrix $M$. Then the upper bound of the Poincaré separation theorem states,

$$\lambda_i(U_k^H Fu_k) \leq \lambda_{n-k+i}(F) \quad i = 1, \ldots, k. \quad (S9)$$

We will also use the von Neumann trace inequality. Let $F, G$ be complex matrices of compatible dimension and minimum dimension $n$. Let $\sigma_i(F), \sigma_i(G)$ be the respective non-increasingly ordered singular values. Then

$$\text{Re}(\text{tr} FG^\dagger) \leq \sum_{i=1}^n \sigma_i(F)\sigma_i(G). \quad (S10)$$

### C Proof of Eq 2

Eq 2 is a generalization of Lemma 2 in [5]. The proof is as follows. The minimum norm solution to the least-squares minimization problem $\min_x \|A_k x - a_i\|_2^2$ is,

$$\hat{x} = \hat{A}_k^+ a_i = V_k \Sigma_k^{-1} U_k^\dagger a_i. \quad (S11)$$

And, by definition,

$$\|\hat{x}\|_2^2 = a_i^T U_k \Sigma_k^{-1} V_k \Sigma_k^{-1} U_k^\dagger a_i = a_i^T U_k \Sigma_k^{-2} U_k^\dagger a_i = \tau_i(A_k). \quad (S12)$$

### D Proof of Eq 9

The upper bound (Eq 9) in Theorem 1 follows from the fact that $0 \preceq I - SS^T$ and the conjugation rule (Eq S4),

$$0 \preceq \text{A}(I - SS^T)A^T = AA^T - CC^T. \quad (S13)$$

This upper bound is true for any column selection of $A$. A second application of the conjugation rule gives the upper bound in Eq 9.

For the lower bound (Eq 9), consider the quantity

$$Y = \Sigma_k^{-1} U_k^\dagger A(I - SS^T)A^T U_k \Sigma_k^{-1} = V_k^\dagger (I - SS^T) V_k. \quad (S14)$$

By the conjugation rule (Eq

January 10, 2019
a small, positive, non-zero number \( \gamma \)

**Eq 1. Construct \( C \)**

If \( A \)

Let \( \text{Theorem 1.} \)

The following theorem pertains to a new spectral bound for the square \( C \) by **Eq S6** and the cyclic property of the trace. Similarly, **Eq S13** implies

\[
 \text{rank-} C \leq \sum_{k=1}^{m} \text{tr} U_k^TCC^TU_k \leq (1-\epsilon) \text{rank } A \leq U_k^TCC^TU_k, \quad (S14)
\]

providing the lower bound of **Eq 9**.

For **Eq 10** the lower bound of **Eq 9** implies,

\[
(1-\epsilon) \text{tr } A_k^{T} \leq \text{tr } C \leq \text{tr } AA^T. \quad (S15)
\]

by **Eq S6** and the cyclic property of the trace. Similarly, **Eq S13** implies \( \text{tr } C \leq \text{tr } AA^T \). Since \( U_k \) is semi-orthogonal \( (U_k^TU_k = I) \), by **Eq S9** every ordered eigenvalue of \( U_k^TCC^TU_k \) is smaller than its counterpart ordered eigenvalue of \( AA^T \).

Since the trace is the sum of eigenvalues, this implies **Eq 10**

\[
(1-\epsilon) \text{tr } A_k^{T} \leq \text{tr } U_k^TCC^TU_k \leq \text{tr } C \leq \text{tr } AA^T. \quad (S16)
\]

Note that if \( A \) is full rank and \( k = \text{rank}(A) = n \), then **Eq 9** becomes,

\[
(1-\epsilon) \text{AA}^T \leq AA^T. \quad (S17)
\]

### E Proof of **Eq 11** for random sampling

The following theorem pertains to a new spectral bound for the square \( C \) selected by rank-\( k \) subspace leverage scores and the random sampling procedure from [6].

**Theorem 1.** Let \( A \in \mathbb{R}^{n \times d} \) be a matrix of at least rank \( k \) and \( \tau_j(A_k) \) be defined as in **Eq 1** Construct \( C \) by sampling \( t \) columns of \( A \), reweighted to \( 1/\sqrt{p_i}a_i \), with probability \( p_i = (\tau_j(A_k) + \gamma 1(\tau_j(A_k) = 0))/\left(\sum_{i=1}^{d} p_i\right) \), where \( 1() \) is the indicator function and \( \gamma \) is a small, positive, non-zero number \( \gamma = \min_{\tau_j(A_k) > 0} (\tau_j(A_k)) \). Let

\[
m = \sum_{i=1}^{d} 1(\tau_j(A_k) = 0), \quad \sum_{i=1}^{d} p_i = k + m\gamma. \quad \text{If the number of selected columns} \quad t \geq \frac{k}{\epsilon^2(k + m\gamma)} (1 + \frac{1}{\epsilon}) \ln \left(\frac{16k}{\delta}\right), \quad \text{then with probability} \quad 1 - \delta, \quad \text{the matrix} \quad C \quad \text{satisfies:}
\]

\[
(1-\epsilon)A_kA_k^T \leq U_k^TCC^TU_k \leq (1+\epsilon)A_kA_k^T. \quad (S18)
\]

If \( A \) is full rank and \( k = \text{rank}(A) = n \), then **Eq S18** becomes,

\[
(1-\epsilon) \text{AA}^T \leq AA^T. \quad (S19)
\]

The proof of **Theorem 1** is similar in structure to **Theorem 3 in [7]**. **Theorem 3 in [7]** pertains to a different type of leverage score.

Consider the quantity \( Y = \Sigma_k^{-1}U_k^T(CC^T - AA^T)U_k\Sigma_k^{-1} \). Note the sign change from Section Proof of **Eq 9**. This can be rewritten as,

\[
Y = \sum_{j=1}^{t} \Sigma_k^{-1}U_k^{-1}(e_je_j^T - \frac{1}{\gamma}AA^T)U_k\Sigma_k^{-1} \quad \forall j, (X_j), = \frac{1}{\gamma}\Sigma_k^{-1}U_k^{-1}(\frac{1}{\epsilon}a_i\epsilon^T - AA^T)U_k\Sigma_k^{-1} \quad \text{with categorical probability} \quad p_i.
\]

(S20)
If \( \| Y \|_2 \leq \epsilon \), then \(-\epsilon I \leq Y \leq \epsilon I\), and Eq. S18 follows from this and the definition of \( Y \). Thus, the proof of Eq. S18 relies on showing that \( \| Y \|_2 \leq \epsilon \). We use an intrinsic dimension matrix Bernstein inequality ([8], Theorem 7.3.1), specialized to Hermitian matrices, to show that \( \| Y \|_2 \) is small with high probability. The Bernstein inequality requires that, for a finite sequence \( Y = \sum_{j=1}^{t} X_j \) of random Hermitian matrices \( X_j \) of the same size,

1. \( \forall j, \mathbb{E}(X_j) = 0 \),
2. \( \forall j, \| X_j \|_2 \leq L \),
3. and that \( \sum_j \mathbb{E}(X_j X_j^T) \leq V \).

Then, for \( \epsilon \geq \sqrt{\| Y \|_2^2 + L/3} \),

\[
P(\| Y \|_2 \geq \epsilon) \leq 8 \frac{\epsilon^2}{\| Y \|_2} \exp \left( -\frac{\epsilon^2/2}{\| Y \|_2} \right).
\]

Requirement 1 is satisfied because,

\[
\mathbb{E}(X_j) = \sum_{i=1}^{d} p_i(X_j)_i = \frac{1}{t} \sum_{j=1}^{d} \sum_{i=1}^{d} a_i a_i^T (U_k^T \Sigma_k^{-1} U_k) \Sigma_k^{-1} = 0.
\]

Eq. S23 follows from the fact that for all \( y \in \mathbb{R}^k \),

\[
y^T U_k \Sigma_k^{-1} U_k^T a_i a_i^T U_k \Sigma_k^{-1} U_k^T y = \text{tr} \left( (U_k \Sigma_k^{-1} U_k^T a_i a_i^T U_k \Sigma_k^{-1} U_k^T) (U_k \Sigma_k^{-1} U_k^T a_i a_i^T U_k \Sigma_k^{-1} U_k^T) \right) \leq \tau_i(A_k) y^T y.
\]

where the inequality comes from the Von Neumann trace inequality (Eq. S10) applied to the product of two rank 1 matrices. Using Eq. S23 in the definition of \( X \), gives,

\[
X_j = \frac{1}{t p_i} \sum_{k} \sum_{j=1}^{d} \sum_{i=1}^{d} a_i a_i^T (U_k^T \Sigma_k^{-1} U_k) \Sigma_k^{-1} - \frac{1}{t} \frac{1}{t p_i} \tau_i(A_k) I - \frac{1}{t} \frac{1}{t p_i} \frac{(k+m\gamma)}{\tau_i(A_k)} I - \frac{1}{t} \frac{1}{t p_i} \frac{(k+m\gamma)}{\tau_i(A_k)} I - \frac{1}{t} I
\]

and \( \| X_j \|_2 \leq L = \frac{k+m\gamma}{\tau_i(A_k)} \) follows immediately.

To show that requirement 3 is satisfied, we compute directly,

\[
\begin{align*}
\mathbb{E}(Y^2) &= t \mathbb{E}(X_j X_j^T) \\
 &= t \sum_{i=1}^{d} p_i \left( \frac{1}{t} \sum_{k} \sum_{j=1}^{d} \sum_{i=1}^{d} a_i a_i^T (U_k^T \Sigma_k^{-1} U_k) \Sigma_k^{-1} \right) \\
 &= t \sum_{i=1}^{d} p_i \left( \frac{1}{t} \sum_{k} \sum_{j=1}^{d} \sum_{i=1}^{d} a_i a_i^T (U_k^T \Sigma_k^{-1} U_k) \Sigma_k^{-1} \left( \frac{1}{t p_i} \tau_i(A_k) I - \frac{1}{t} \frac{1}{t p_i} \frac{(k+m\gamma)}{\tau_i(A_k)} I - \frac{1}{t} I \right) \right) \\
 &\leq t \sum_{i=1}^{d} p_i \left( \frac{1}{t p_i} \sum_{k} \sum_{j=1}^{d} \sum_{i=1}^{d} a_i a_i^T (U_k^T \Sigma_k^{-1} U_k) \Sigma_k^{-1} \right) - \frac{1}{t} I
\end{align*}
\]
\[
\sum_{i=1}^{d} \left( \Sigma_k^{-1} U_k^T a_i a_i^T U_k \Sigma_k^{-1} \right) = \frac{k+m\gamma}{t} \mathbf{I} = \mathbf{V}.
\] (S25)

It follows immediately that \( ||\mathbf{V}||_2 = \frac{k+m\gamma}{t} \) and \( \text{tr} \mathbf{V} = \frac{k(k+m\gamma)}{t} \).

Then, for \( \epsilon \geq \sqrt{\frac{k+m\gamma}{t} + \frac{k+m\gamma}{3t}} \),

\[
P(||\mathbf{Y}||_2 \geq \epsilon) \leq 8k \exp \left( -\frac{\epsilon^2 / 2}{(k+m\gamma)(\epsilon/3+1)} \right) \leq \frac{1}{2} \delta.
\] (S26)

Solving for \( t \) as a function of \( \epsilon, \delta, \) and \( \gamma \) gives,

\[
t \geq \frac{2}{\epsilon^2} (k+m\gamma) \left( 1 + \frac{1}{3} \epsilon \right) \ln \left( \frac{16k}{\delta} \right).
\] (S27)

Eq (S18) also holds for \( \mathbf{C} \) selected by the DCSS algorithm, as a consequence of Eq 9. Thus DCSS selects fewer columns with the same accuracy for power-law decay for Eq (S18) when \( |\Theta| < t \).

References

1. Zeisel A, Muñoz-Manchado AB, Codeluppi S, Lönnberg P, Manno GL, Juréus A, et al. Cell types in the mouse cortex and hippocampus revealed by single-cell RNA-seq. Science. 2015;347(6226):1138–1142. doi:10.1126/science.aaa1934.

2. Ntranos V, Kamath GM, Zhang JM, Pachter L, Tse DN. Fast and accurate single-cell RNA-seq analysis by clustering of transcript-compatibility counts. Genome Biology. 2016;17(1):1–14. doi:10.1186/s13059-016-0970-8.

3. Horn RA, Johnson CR. Matrix analysis. 2nd ed. New York: Cambridge University Press; 2013.

4. Eckart C, Young G. The approximation of one matrix by another of lower rank. Psychometrika. 1936;1:211–218. doi:10.1007/BF02288367.

5. Cohen MB, Lee YT, Musco C, Musco C, Peng R, Sidford A. Uniform Sampling for Matrix Approximation. In: Proceedings of the 2015 Conference on Innovations in Theoretical Computer Science, ITCS ’15. New York, NY, USA: ACM; 2015. p. 181–190. Available from: http://dl.acm.org/citation.cfm?id=3039686.3039801.

6. Drineas P, Mahoney MW, Muthukrishnan S. Relative-Error CUR Matrix Decompositions. SIAM J Matrix Anal Appl. 2008;30(2):844–881. doi:10.1137/07070471X.

7. Cohen MB, Musco C, Musco C. Input Sparsity Time Low-rank Approximation via Ridge Leverage Score Sampling. In: Proceedings of the Twenty-Eighth Annual ACM-SIAM Symposium on Discrete Algorithms. SODA ’17. Philadelphia, PA, USA: Society for Industrial and Applied Mathematics; 2017. p. 1758–1777. Available from: http://dl.acm.org/citation.cfm?id=3039686.3039801.

8. Tropp JA. An Introduction to Matrix Concentration Inequalities. Found Trends Mach Learn. 2015;8(1-2):1–230. doi:10.1561/2200000048.