



Quantum algorithms for SVD-based data representation and analysis

Armando Bellante¹ · Alessandro Luongo^{2,3} · Stefano Zanero¹

Received: 2 February 2022 / Accepted: 7 June 2022
© The Author(s) 2022

Abstract

This paper narrows the gap between previous literature on quantum linear algebra and practical data analysis on a quantum computer, formalizing quantum procedures that speed-up the solution of eigenproblems for data representations in machine learning. The power and practical use of these subroutines is shown through new quantum algorithms, sublinear in the input matrix's size, for principal component analysis, correspondence analysis, and latent semantic analysis. We provide a theoretical analysis of the run-time and prove tight bounds on the randomized algorithms' error. We run experiments on multiple datasets, simulating PCA's dimensionality reduction for image classification with the novel routines. The results show that the run-time parameters that do not depend on the input's size are reasonable and that the error on the computed model is small, allowing for competitive classification performances.

Keywords Quantum computing · machine learning · Data analysis · Data representations · Singular value decomposition · Principal component analysis · Correspondence analysis · Latent semantic analysis

1 Introduction

Quantum computation is a computing paradigm that promises substantial speed-ups in a plethora of tasks that are computationally hard for classical computers. In 2009, Harrow, Hassidim, and Lloyd Harrow et al. (2009) presented quantum procedures to create a quantum state proportional to the solution of a linear system of equations $\mathbf{Ax} = \mathbf{b}$ in time logarithmic in the size of \mathbf{A} . This result has promoted further research on optimization, linear algebra, and machine learning problems, leading to faster quantum algorithms for linear regressions (Chakraborty et al. 2019), support vector machines (Rebentrost et al. 2014a), k-means (Kerenidis

et al. 2019a), and many others (Biamonte et al. 2017). Following this research line, in this work, we focus on quantum algorithms for singular value based data analysis and representation. When handling big data, it is crucial to learn effective representations that reduce the data's noise and help the learner perform better on the task. Many data representation methods for machine learning, such as principal component analysis (Partridge and Calvo 1997), correspondence analysis (Greenacre 1984), slow feature analysis (Kerenidis and Luongo 2020), or latent semantic analysis (Deerwester et al. 1990), heavily rely on singular value decomposition and are impractical to compute on classical computers for extensive datasets.

We have gathered and combined state-of-the-art quantum techniques to present a useful and easy-to-use framework for solving eigenvalue problems at large scale. While we focus on machine learning problems, these subroutines can be used for other problems that are classically solved via an SVD of a suitable matrix. More specifically, we formalize novel quantum procedures to compute classical estimates of the most relevant singular values, factor scores, factor score ratios of an $n \times m$ matrix in time poly-logarithmic in nm , and the most relevant singular vectors sub-linearly in nm . We show how to use these procedures to obtain a classical description of the models of three machine learning algorithms: principal component analysis, correspondence

✉ Armando Bellante
armando.bellante@polimi.it

Alessandro Luongo
ale@nus.edu.sg

Stefano Zanero
stefano.zanero@polimi.it

¹ DEIB, Politecnico di Milano, Via Ponzio, 34/5 - Building 20, 20133 Milan, Italy
² IRIF, Université Paris Diderot, Place Aurélie Nemours 8, 75205 Paris, France
³ CQT, National University of Singapore, Science Drive 2 3, 117543 Singapore, Singapore

analysis, and latent semantic analysis. We also discuss how to represent the data in the new feature space with a quantum computer. We provide a thorough theoretical analysis for all these algorithms bounding the run-time, the error, and the failure probability.

The remainder of the paper is organized as follows. Section 2 introduces our notation and discusses the relevant quantum preliminaries. Section 3 presents the novel quantum algorithms. In Section 4, we show applications of the algorithms to principal component analysis, correspondence analysis, and latent semantic analysis. Section 5 presents numerical experiments assessing the run-time parameters. Finally, we provide detailed information on the experiments and extensively discuss related work in quantum and classical literature in the appendix.

2 Quantum preliminaries and notation

2.1 Notation

Given a matrix \mathbf{A} , we write \mathbf{a}_i to denote its i^{th} row, \mathbf{a}_j for its j^{th} column, and a_{ij} for the element at row i , column j . We write its singular value decomposition as $\mathbf{A} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T$. \mathbf{U} and \mathbf{V} are orthogonal matrices, whose column vectors \mathbf{u}_i and \mathbf{v}_i are respectively the left and right singular vectors of \mathbf{A} . $\mathbf{\Sigma}$ is a diagonal matrix with positive, non-negative, entries σ_i : the singular values. The row/column size of $\mathbf{\Sigma}$ is the rank of \mathbf{A} and is denoted as r . We use λ_i to denote the i^{th} eigenvalue of the covariance matrix $\mathbf{A}^T\mathbf{A} = \mathbf{V}\mathbf{\Sigma}^2\mathbf{V}^T$, and $\lambda^{(i)} = \frac{\lambda_i}{\sum_j \lambda_j}$ to denote the relative magnitude of each eigenvalue. Using the notation of Hsu et al. (2019) for correspondence analysis, we refer to λ_i as factor scores and to $\lambda^{(i)}$ as factor score ratios. Note that $\lambda_i = \sigma_i^2$ and $\lambda^{(i)} = \frac{\sigma_i^2}{\sum_j \sigma_j^2}$. We denote the number of non-zero elements of a matrix/vector with $nnz()$. Given a scalar a , $|a|$ is its absolute value. The ℓ_∞ and ℓ_0 norm of a vector \mathbf{a} are defined as $\|\mathbf{a}\|_\infty = \max_i(\|a_i\|)$, $\|\mathbf{a}\|_0 = nnz(\mathbf{a})$. If the vector norm is not specified, we refer to the ℓ_2 norm. The Frobenius norm of a matrix is $\|\mathbf{A}\|_F = \sqrt{\sum_i \sigma_i^2}$, its spectral norm is $\|\mathbf{A}\| = \max_{\mathbf{x} \in \mathbb{R}^m} \frac{\|\mathbf{A}\mathbf{x}\|}{\|\mathbf{x}\|} = \sigma_{max}$, and finally $\|\mathbf{A}\|_\infty = \max_i(\|\mathbf{a}_i\|_1)$. A contingency table is a matrix that represents categorical variables in terms of the observed frequency counts. Finally, when stating the complexity of an algorithm, we use \tilde{O} instead of O to omit the poly-logarithmic terms on the size of the input data (i.e., $O(\text{polylog}(nm)) = \tilde{O}(1)$), on the error, and the failure probability.

2.2 Quantum preliminaries

We represent scalars as states of the computational basis of \mathcal{H}_n , where n is the number of bits required for binary encoding. The quantum state corresponding to a vector $\mathbf{v} \in \mathbb{R}^m$ is defined as a state-vector $|\mathbf{v}\rangle = \frac{1}{\|\mathbf{v}\|} \sum_j^m v_j |j\rangle$. Note that to build $|\mathbf{v}\rangle$ we need $\lceil \log m \rceil$ qubits.

Data access To access data in the form of state-vectors, we use the following definition of quantum access.

Definition 1 (Quantum access to a matrix) We have quantum access to a matrix $\mathbf{A} \in \mathbb{R}^{n \times m}$, if there exists a data structure that allows performing the mappings $|i\rangle|0\rangle \mapsto |i\rangle|\mathbf{a}_i\rangle = |i\rangle \frac{1}{\|\mathbf{a}_i\|} \sum_j^m a_{ij} |j\rangle$, for all i , and $|0\rangle \mapsto \frac{1}{\|\mathbf{A}\|_F} \sum_i^n \|\mathbf{a}_i\| |i\rangle$ in time $\tilde{O}(1)$.

By combining the two mappings we can create the state $|\mathbf{A}\rangle = \frac{1}{\|\mathbf{A}\|_F} \sum_i^n \sum_j^m a_{ij} |i\rangle |j\rangle$ in time $\tilde{O}(1)$.

Kerenidis and Prakash (2017, 2020a) have described one implementation of such quantum data access. Their implementation is based on a classical data structure such that the cost of updating/deleting/inserting one element of the matrix is poly-logarithmic in the number of its entries. In addition, their structure gives access to the Frobenius norm of the matrix and the norm of its rows in time $O(1)$. The cost of creating this data structure is $\tilde{O}(nnz(\mathbf{A}))$. This input model requires the existence of a QRAM (Giovannetti et al. 2008). While there has been some skepticism on the possibility of error-correcting such a complex device, recent results show that bucket-brigade QRAMs are highly resilient to generic noise (Hann et al. 2021).

Sometimes it is desirable to normalize the input matrix to have a spectral norm smaller than one. Kerenidis and Prakash (2020a) provide an efficient routine to estimate the spectral norm.

Theorem 1 (Spectral norm estimation (Kerenidis and Prakash 2020a)) *Let there be quantum access to the matrix $\mathbf{A} \in \mathbb{R}^{n \times m}$, and let $\epsilon > 0$ be a precision parameter. There exists a quantum algorithm that estimates $\|\mathbf{A}\|$ to additive error $\epsilon \|\mathbf{A}\|_F$ in time $\tilde{O}\left(\frac{\log(1/\epsilon) \|\mathbf{A}\|_F}{\epsilon \|\mathbf{A}\|}\right)$.*

If we have $\|\mathbf{a}\|$, we can create quantum access to $\mathbf{A}' = \frac{\mathbf{A}}{\|\mathbf{A}\|} = \mathbf{U} \frac{\mathbf{\Sigma}}{\sigma_{max}} \mathbf{V}^T$ in time $\tilde{O}(nnz(\mathbf{A}))$ by dividing each entry of the data structure. Once we have quantum access to a dataset, it is possible to apply a pipeline of quantum machine learning algorithms for data representation, analysis, clustering, and classification (Rebentrost et al. 2014b; Kerenidis et al. 2020a; Wang 2017; Allcock et al. 2020;

Kerenidis et al. 2019a; Kerenidis and Luongo 2020). Since the cost of each step of the pipeline should be evaluated independently, we consider $\tilde{O}(nnz(\mathbf{A}))$ to be a pre-processing cost and do not include it in our run-times.

We conclude this section by stating a useful claim that connects errors on classical vectors with errors on quantum states.

Claim 2 (Closeness of state-vectors (Kerenidis and Prakash 2020a)) *Let θ be the angle between vectors $\mathbf{x}, \bar{\mathbf{x}}$ and assume that $\theta < \pi/2$. Then, $\|\mathbf{x} - \bar{\mathbf{x}}\| \leq \epsilon$ implies $\|\|\mathbf{x}\rangle - \|\bar{\mathbf{x}}\rangle\| \leq \sqrt{2} \frac{\epsilon}{\|\mathbf{x}\|}$.*

Useful subroutines We state two relevant quantum linear algebra results: quantum singular value estimation (SVE) and quantum matrix-vector multiplication.

Theorem 3 (Singular value estimation (Kerenidis and Prakash 2020a)) *Let there be quantum access to $\mathbf{A} \in \mathbb{R}^{n \times m}$, with singular value decomposition $\mathbf{A} = \sum_i^r \sigma_i \mathbf{u}_i \mathbf{v}_i^T$ and $r = \min(n, m)$. Let $\epsilon > 0$ be a precision parameter. It is possible to perform the mapping $|b\rangle = \sum_i \alpha_i |\mathbf{v}\rangle_i \mapsto \sum_i \alpha_i |\mathbf{v}\rangle_i |\bar{\sigma}_i\rangle$, such that $|\frac{\sigma_i}{\mu(\mathbf{A})} - \bar{\sigma}_i| \leq \epsilon$ with probability at least $1 - 1/\text{poly}(m)$, in time $\tilde{O}(\frac{1}{\epsilon})$ where $\mu(\mathbf{A}) = \min_{p \in [0,1]} (\|\mathbf{A}\|_F, \sqrt{s_{2p}(\mathbf{A})s_{2(1-p)}(\mathbf{A}^T)})$ and $s_p(\mathbf{A}) = \max \|\mathbf{a}_{i \cdot}\|_p$. Similarly, we can have $|\sigma_i - \bar{\sigma}_i| \leq \epsilon$ in time $\tilde{O}(\frac{\mu(\mathbf{A})}{\epsilon})$.*

Unlike previous results with Hamiltonian simulations (Rebentrost et al. 2014b), this algorithm enables performing conditional rotations using the singular values of a matrix without any special requirement (e.g., sparsity, being square, Hermitian, etc.). By choosing the same matrix $|\mathbf{A}\rangle = \frac{1}{\|\mathbf{A}\|_F} \sum_i \sum_j^m a_{ij} |i\rangle |j\rangle = \frac{1}{\|\mathbf{A}\|_F} \sum_i \sigma_i |\mathbf{u}\rangle_i |\mathbf{v}\rangle_i$ as starting state $|b\rangle$, we obtain a superposition of all the singular values entangled with the respective left and right singular vectors $\frac{1}{\|\mathbf{A}\|_F} \sum_i \sigma_i |\mathbf{u}\rangle_i |\mathbf{v}\rangle_i |\bar{\sigma}_i\rangle$. In this case, the requirement $r = \min(n, m)$ is not needed anymore as $|b\rangle = |\mathbf{A}\rangle$ can be fully decomposed in terms of \mathbf{A} 's right singular vectors.

This algorithm uses phase estimation. In this work, we consider this algorithm to use a consistent version of phase estimation, so that the errors in the estimates of the singular values are consistent across multiple runs (Ta-Shma 2013; Kerenidis and Prakash 2020a).

Theorem 4 (Matrix-vector multiplication (Chakraborty et al. 2019) (Lemma 24, 25)) *Let there be quantum access to the matrix $\mathbf{A} \in \mathbb{R}^{n \times n}$, with $\sigma_{\max} \leq 1$, and to a vector $\mathbf{x} \in \mathbb{R}^n$. Let $\|\mathbf{Ax}\| \geq \gamma$. There exists a quantum algorithm that creates a state $|\mathbf{z}\rangle$ such that $\|\|\mathbf{z}\rangle - \|\mathbf{Ax}\rangle\| \leq \epsilon$ in time $\tilde{O}(\frac{1}{\gamma} \mu(\mathbf{A}) \log(1/\epsilon))$, with probability at least $1 - 1/\text{poly}(n)$.*

Increasing the run-time by a multiplicative factor $\tilde{O}(\frac{1}{\eta})$ one can retrieve an estimate of $\|\mathbf{Ax}\|$ to relative error η .

Data output Finally, to read out the quantum states, we state one version of amplitude amplification and estimation, and two *state-vector* tomographies.

Theorem 5 (Amplitude amplification and estimation (Brassard et al. 2002; Kerenidis et al. 2019c)) *Let there be a unitary that performs the mapping $U_x : |0\rangle \mapsto \sin(\theta)|\mathbf{x}, 0\rangle + \cos(\theta)|\mathbf{G}, 0^\perp\rangle$, where $|\mathbf{G}\rangle$ is a garbage state, in time $T(U_x)$. Then, $\sin(\theta)^2$ can be estimated to multiplicative error η in time $O(\frac{T(U_x)}{\eta \sin(\theta)})$ or to additive error η in time $O(\frac{T(U_x)}{\eta})$, and $|\mathbf{x}\rangle$ can be generated in expected time $O(\frac{T(U_x)}{\sin(\theta)})$.*

Theorem 6 (ℓ_2 state-vector tomography (Kerenidis and Prakash 2020b; Kerenidis et al. 2019c)) *Given a unitary mapping $U_x : |0\rangle \mapsto |\mathbf{x}\rangle$ in time $T(U_x)$ and $\delta > 0$, there is an algorithm that produces an estimate $\bar{\mathbf{x}} \in \mathbb{R}^m$ with $\|\bar{\mathbf{x}}\| = 1$ such that $\|\mathbf{x} - \bar{\mathbf{x}}\| \leq \delta$ with probability at least $1 - 1/\text{poly}(m)$ in time $O(T(U_x) \frac{m \log m}{\delta^2})$.*

Theorem 7 (ℓ_∞ state-vector tomography (Kerenidis et al. 2019b)) *Given access to a unitary mapping $U_x : |0\rangle \mapsto |\mathbf{x}\rangle$ and its controlled version in time $T(U_x)$, and $\delta > 0$, there is an algorithm that produces an estimate $\bar{\mathbf{x}} \in \mathbb{R}^m$ with $\|\bar{\mathbf{x}}\| = 1$ such that $\|\mathbf{x} - \bar{\mathbf{x}}\|_\infty \leq \delta$ with probability at least $1 - 1/\text{poly}(m)$ in time $O(T(U_x) \frac{\log m}{\delta^2})$.*

3 Novel quantum methods

Building from the previous section's techniques, we formalize a series of quantum algorithms that allow us to retrieve a classical description of the singular value decomposition of a matrix to which we have quantum access.

3.1 Estimating the quality of the representation

Algorithms such as principal component analysis and correspondence analysis are often used for visualization or dimensionality reduction purposes. These applications work better when a small subset of factor scores have high factor score ratios. We provide a fast procedure that allows verifying if this is the case: given efficient quantum access to a matrix $\mathbf{A} \in \mathbb{R}^{n \times m}$, it retrieves the most relevant singular values, factor scores, and factor score ratios in time poly-logarithmic

in the number of elements of \mathbf{A} , with no strict dependencies on its rank.

The main intuition behind this algorithm is that it is possible to create the state $\sum_i^r \sqrt{\lambda^{(i)}} |\mathbf{u}\rangle_i |\mathbf{v}\rangle_i |\bar{\sigma}_i\rangle$. The third register, when measured in the computational basis, outputs the estimate $\bar{\sigma}_i$ of a singular value with probability equal to its factor score ratio $\lambda^{(i)}$. This enables sampling the singular values of \mathbf{A} directly from the factor score ratios' distribution. When a matrix has a huge number of small singular values and only a few of them that are very big, the ones with the

greatest factor score ratios will appear many times during the measurements. In contrast, the negligible ones are not likely to be measured. This intuition has already appeared in literature (Gyurik et al. 2020; Cade and Montanaro 2018). Nevertheless, the analysis and the problem solved in these works are different from ours. In the context of data representation and analysis, this intuition has only been sketched for sparse or low rank square symmetric matrices by Lloyd et al. (2014), without a precise formalization. We thoroughly formalize it for any real matrix.

Algorithm 1 Quantum factor score ratio estimation.

Input: Quantum access to a matrix $\mathbf{A} \in \mathbb{R}^{n \times m}$. Two precision parameters $\gamma, \epsilon \in \mathbb{R}_{>0}$.

Output: An estimate of the factor score ratios $\lambda^{(i)} > \gamma$. An estimate of the corresponding singular values and factor scores.

- 1: **for** $N \sim O(\frac{1}{\gamma^2})$ **times do**
 - 2: Prepare the state $\frac{1}{\|\mathbf{A}\|_F} \sum_i^n \sum_j^m a_{ij} |i\rangle |j\rangle$.
 - 3: Apply SVE to get $\frac{1}{\sqrt{\sum_j^m \sigma_j^2}} \sum_i^r \sigma_i |\mathbf{u}_i\rangle |\mathbf{v}_i\rangle |\bar{\sigma}_i\rangle$.
 - 4: Measure the last register and store it in a set data structure.
 - 5: **end for**
 - 6: For each $\bar{\sigma}_i$ measured, output $\bar{\sigma}_i$, its factor score $\bar{\lambda}_i = \bar{\sigma}_i^2$, and its factor score ratio $\bar{\lambda}^{(i)} = \frac{\bar{\lambda}_i}{\|\mathbf{A}\|_F^2}$.
-

Theorem 8 (Quantum factor score ratio estimation) *Let there be quantum access to a matrix $\mathbf{A} \in \mathbb{R}^{n \times m}$, with singular value decomposition $\mathbf{A} = \sum_i \sigma_i \mathbf{u}_i \mathbf{v}_i^T$. Let γ, ϵ be precision parameters. There exists a quantum algorithm that runs in time $\tilde{O}\left(\frac{1}{\gamma^2} \frac{\mu(\mathbf{A})}{\epsilon}\right)$ and estimates:*

- all the factor score ratios $\lambda^{(i)} > \gamma$, with probability at least $1 - 1/\text{poly}(r)$, such that $|\lambda^{(i)} - \bar{\lambda}^{(i)}| \leq 2\epsilon \frac{\sigma_i}{\|\mathbf{A}\|_F^2}$, with probability at least $1 - 1/\text{poly}(n)$;
- the corresponding singular values σ_i , such that $|\sigma_i - \bar{\sigma}_i| \leq \epsilon$ with probability at least $1 - 1/\text{poly}(n)$;
- the corresponding factor scores λ_i , such that $|\lambda_i - \bar{\lambda}_i| \leq 2\epsilon \sqrt{\lambda_i}$ with probability at least $1 - 1/\text{poly}(n)$.

The proof consists in bounding the run-time, the error, and the probability of failure of Algorithm 1.

Proof By the definition of quantum access, the cost of step 2 is $\tilde{O}(1)$. The singular value estimation in step 3 can be performed using Theorem 3 in time $\tilde{O}\left(\frac{\mu(\mathbf{A})}{\tau}\right)$, such that

$\|\sigma_i - \bar{\sigma}_i\| \leq \epsilon$ with probability at least $1 - 1/\text{poly}(n)$. A measurement of the third register at step 4 can output any $\bar{\sigma}_i$ with probability $\lambda^{(i)} = \frac{\sigma_i^2}{\sum_j \sigma_j^2}$.

Theorem 7 guarantees that with $O(1/\gamma^2)$ measurements we can get estimates $|\lambda^{(i)} - \bar{\lambda}^{(i)}| \leq \gamma$. In particular, Kerenidis et al. (2019b) estimate that $N = 36 \log(r)/\gamma^2$ measures should suffice for our goal.

Alternatively, we could consider the measurement process as performing r Bernoulli trials: one for each $\bar{\lambda}^{(i)}$, so that if we measure $\bar{\sigma}_i$ it is a success for the i^{th} Bernoulli trial and a failure for all the others. Given a confidence level z , it is possible to use the Wald confidence interval to determine a value for N such that $|\lambda^{(i)} - \frac{\zeta_{\bar{\sigma}_i}}{N}| \leq \gamma$ with confidence level z , where $\zeta_{\bar{\sigma}_i}$ is the number of times that $\bar{\sigma}_i$ has appeared in the measurements. In this case, it suffice to choose $N = \frac{z^2}{4\gamma^2}$ [Schuld and Petruccione (2018), Section 5.1.3]. Having $|\lambda^{(i)} - \bar{\lambda}^{(i)}| \leq \gamma$ means measuring all the $\bar{\sigma}_i$ whose factor score ratio is greater than λ .

We now proceed with the error analysis. We can compute $\bar{\lambda}_i = \bar{\sigma}_i^2$.

$$|\lambda_i - \bar{\sigma}_i^2| \leq |\lambda_i - (\sigma_i \pm \epsilon)^2| = |\pm 2\epsilon\sigma_i + \epsilon^2| \leq 2\epsilon\sigma_i + \epsilon^2. \tag{1}$$

If we keep the error analysis at the first order and consider that $\sigma_i = \sqrt{\lambda_i}$, we can conclude the bound as $|\lambda_i - \bar{\sigma}_i^2| \leq 2\epsilon\sqrt{\lambda_i}$. Similarly, we can compute $\bar{\lambda}^{(i)} = \frac{\bar{\sigma}_i^2}{\|A\|_F^2}$.

$$|\lambda^{(i)} - \bar{\lambda}^{(i)}| = \frac{|\lambda_i - \bar{\lambda}_i|}{\|A\|_F^2} \leq 2\epsilon \frac{\sigma_i}{\|A\|_F^2}. \tag{2}$$

The parameter γ is the one that controls how big a factor score ratio should be for the singular value/factor score to be measured. If we choose γ bigger than the least factor scores

ratio of interest, the estimate for the smaller ones is likely to be 0, as $|\lambda^{(i)} - 0| \leq \gamma$ would be a plausible estimation.

Often in data representations, the cumulative sum of the factor score ratios is a measure of the quality of the representation. By slightly modifying Algorithm 1 to use Theorem 7, it is possible to estimate this sum such that $|\sum_i^k \lambda^{(i)} - \sum_i^k \bar{\lambda}^{(i)}| \leq k\epsilon$ with probability $1 - 1/\text{poly}(r)$. However, a slight variation of Algorithm IV.3 for spectral norm estimation in Kerenidis and Prakash (2020a) provides a more accurate estimation in less time, given a threshold θ for the smallest singular value to retain.

Algorithm 2 Quantum check on the factor score ratios' sum.

Input: Quantum access to a matrix $A \in \mathbb{R}^{n \times m}$. A threshold parameter θ . Two precision parameters $\epsilon, \eta \in \mathbb{R}_{>0}$, such that the greatest singular value smaller than θ is not more than ϵ distant to it.

Output: An estimate \bar{p} of the factor score ratios' sum $p = \sum_{i:\sigma_i \geq \theta} \lambda^{(i)}$.

- 1: Prepare the state $\frac{1}{\|A\|_F} \sum_i^n \sum_j^m a_{ij} |i\rangle |j\rangle$.
- 2: Apply SVE with precision ϵ to get $\frac{1}{\sqrt{\sum_j^r \sigma_j^2}} \sum_i^r \sigma_i |\mathbf{u}_i\rangle |\mathbf{v}_i\rangle |\bar{\sigma}_i\rangle$.
- 3: Append a quantum register $|0\rangle$ to the state and set it to $|1\rangle$ if $\bar{\sigma}_i < \theta$.
- 4: Uncompute the SVE

$$\frac{1}{\sqrt{\sum_j^r \sigma_j^2}} \sum_{i:\bar{\sigma}_i \geq \theta} \sigma_i |\mathbf{u}_i\rangle |\mathbf{v}_i\rangle |0\rangle + \frac{1}{\sqrt{\sum_j^r \sigma_j^2}} \sum_{i:\bar{\sigma}_i < \theta} \sigma_i |\mathbf{u}_i\rangle |\mathbf{v}_i\rangle |1\rangle$$

- 5: Perform amplitude estimation with precision η on the last register being $|0\rangle$, to estimate $p = \frac{\sum_{i:\bar{\sigma}_i \geq \theta} \sigma_i^2}{\sum_j^r \sigma_j^2} = \sum_{i:\bar{\sigma}_i \geq \theta} \lambda^{(i)}$.
-

Theorem 9 (Quantum check on the factor score ratios' sum) *Let there be quantum access to a matrix $A \in \mathbb{R}^{n \times m}$, with singular value decomposition $A = \sum_i \sigma_i \mathbf{u}_i \mathbf{v}_i^T$. Let η, ϵ be precision parameters, and θ be a threshold for the smallest singular value to consider. There exists a quantum algorithm that estimates $p = \sum_{i:\bar{\sigma}_i \geq \theta} \lambda^{(i)}$, where $|\sigma_i - \bar{\sigma}_i| \leq \epsilon$, to relative error η in time $\tilde{O}\left(\frac{\mu(A)}{\epsilon} \frac{1}{\eta\sqrt{p}}\right)$.*

Proof As discussed in the previous proof, the cost of preparing the state at step 2 is $\tilde{O}\left(\frac{\mu(A)}{\epsilon}\right)$. The complexity of step 3 is $\tilde{O}(1)$, as it is an arithmetic operation that only depends on the encoding of $|\bar{\sigma}_i\rangle$. Step 4 consists in uncomputing step 2 and has its same cost. Finally, the cost of amplitude

estimation, with relative precision η , on the last register being $|0\rangle$ is equal to $O\left(T(U_4) \frac{1}{\eta\sqrt{p}}\right)$, where $p = \frac{\sum_{i:\bar{\sigma}_i \geq \theta} \sigma_i^2}{\sum_j^r \sigma_j^2}$ is the probability of measuring $|0\rangle$ (Theorem 5). The overall complexity is proven: $\tilde{O}\left(\frac{\mu(A)}{\epsilon} \frac{1}{\eta\sqrt{p}}\right)$.

Since the sum of factor score ratios p is a measure of the representation quality, in problems such as PCA, CA, and LSA, this is usually a constant number bigger than 0 (i.e., often in practice, $p \in [0.3, 1]$). This makes the term \sqrt{p} negligible in most of the practical applications. Moreover, we further modify Algorithm IV.3 to perform a binary search of θ given the desired sum of factor score ratios.

Algorithm 3 Quantum binary search for the singular value threshold.

Input: Quantum access to a matrix $\mathbf{A} \in \mathbb{R}^{n \times m}$. The desired amount of factor score ratios sum $p \in [0, 1]$. Two precision parameters $\epsilon, \eta \in \mathbb{R}_{>0}$.

Output: A threshold θ such that $|p - \sum_{i:\bar{\sigma}_i \geq \theta} \lambda^{(i)}| \leq \eta$, where $|\bar{\sigma}_i - \sigma_i| \leq \epsilon$, or -1 if no such θ exists.

- 1: Let $l = 0$ and $u = 1$ be upper and lower bounds for the binary search.
- 2: If $|1 - p| \leq \eta$, then return $\theta = 0$.
- 3: If $|0 - p| \leq \eta$, then return $\theta = \mu(\mathbf{A})$.
- 4: Initialize $\tau = (l + u)/2$.
- 5: **for** $O(\log(\frac{\mu(\mathbf{A})}{\epsilon}))$ **times do**
- 6: Prepare the state $|\mathbf{A}\rangle$ and apply SVE to get $\frac{1}{\sqrt{\sum_j^r \sigma_j^2}} \sum_i^r \sigma_i |\mathbf{u}_i\rangle |\mathbf{v}_i\rangle |\hat{\sigma}_i\rangle$,
so that $|\hat{\sigma}_i - \frac{\sigma_i}{\mu(\mathbf{A})}| \leq \frac{\epsilon}{\mu(\mathbf{A})}$.
- 7: Append a quantum register $|0\rangle$ to the state and set it to $|1\rangle$ if $\hat{\sigma}_i < \tau$.
- 8: Uncompute the SVE

$$\frac{1}{\sqrt{\sum_j^r \sigma_j^2}} \sum_{i:\hat{\sigma}_i \geq \tau} \sigma_i |\mathbf{u}_i\rangle |\mathbf{v}_i\rangle |0\rangle + \frac{1}{\sqrt{\sum_j^r \sigma_j^2}} \sum_{i:\hat{\sigma}_i < \tau} \sigma_i |\mathbf{u}_i\rangle |\mathbf{v}_i\rangle |1\rangle.$$

- 9: Perform amplitude estimation with precision η on the last register being $|0\rangle$, to estimate $p_\tau = \frac{\sum_{i:\hat{\sigma}_i \geq \tau} \sigma_i^2}{\sum_j^r \sigma_j^2} = \sum_{i:\hat{\sigma}_i \geq \tau} \lambda^{(i)}$, such that $|\bar{p}_\tau - p_\tau| \leq \eta/2$.
- 10: If $|\bar{p}_\tau - p| \leq \eta/2$, then return $\theta = \tau \mu(\mathbf{A})$.
- 11: If $\bar{p}_\tau < p$, then set $u = \tau$ and set $l = \tau$ otherwise.
- 12: Update $\tau = (u + l)/2$.
- 13: **end for**
- 14: Return -1 .

Theorem 10 (Quantum binary search for the singular value threshold) *Let there be quantum access to a matrix $\mathbf{A} \in \mathbb{R}^{n \times m}$. Let η, ϵ be precision parameters, and θ be a threshold for the smallest singular value to consider. Let $p \in [0, 1]$ be the factor score ratios sum to retain. There exists a quantum algorithm that runs in time $\tilde{O}\left(\frac{\mu(\mathbf{A}) \log(\mu(\mathbf{A})/\epsilon)}{\epsilon \eta}\right)$ and outputs an estimate θ such that $|p - \sum_{i:\bar{\sigma}_i \geq \theta} \lambda^{(i)}| \leq \eta$, where $|\bar{\sigma}_i - \sigma_i| \leq \epsilon$, or detects whether such θ does not exist.*

The proof consists in proving the correctness and the run-time of Algorithm 3.

Proof The algorithm searches for θ using τ as an estimate between 0 and 1. The search is performed using $\text{sign}(p_\tau - p)$ as an oracle that tells us whether to update the lower or upper bound for τ .

The algorithm terminates when $|\bar{p}_\tau - p| \leq \eta/2$ or when it is not possible to update τ anymore (i.e., there are not enough qubits to express the next τ). In this last case, there is no θ that satisfies the requisites and the algorithm returns -1 .

In the first case, instead, we need to guarantee that $|p - \sum_{i:\bar{\sigma}_i \geq \theta} \lambda^{(i)}| = |p - p_\tau| \leq \eta$. Since we run amplitude estimation with additive error $\eta/2$ we have $|\bar{p}_\tau - p_\tau| \leq \eta/2$, and we require $|\bar{p}_\tau - p| \leq \eta/2$ to stop. This two conditions entail

$$|p - p_\tau| \leq |p - \bar{p}_\tau| + \eta/2 \leq \eta. \quad (3)$$

If we want θ to be comparable with the singular values of \mathbf{A} and use τ for the binary search, we have to use Theorem 3 with error $\frac{\mu(\mathbf{A})}{\epsilon}$, meaning that Step 6 can be done in time $O(\mu(\mathbf{A})/\epsilon)$. The total cost of the inner loop has to be evaluated at the end of Step 9, which runs in time $O(\frac{\mu(\mathbf{A})}{\epsilon \eta})$.

The maximum number of updates of τ is bounded by the number of qubits that we use to store the singular values $\hat{\sigma}_i$. This is given by the logarithm of the error used in Step 6, and is $O\left(\log\left(\frac{\mu(\mathbf{A})}{\epsilon}\right)\right)$.

The run-time of this algorithm is bounded by $\tilde{O}\left(\frac{\mu(\mathbf{A}) \log(\mu(\mathbf{A})/\epsilon)}{\epsilon \eta}\right)$.

Using the quantum counting algorithms of Brassard et al. (2002) after step 3 of Algorithm 2, it is possible to count the number of singular values retained by a certain threshold θ .

Corollary 11 (Quantum reduced rank estimation) *Let there be quantum access to a matrix $\mathbf{A} \in \mathbb{R}^{n \times m}$, with singular value decomposition $\mathbf{A} = \sum_i^r \sigma_i \mathbf{u}_i \mathbf{v}_i^T$ and rank r . Let ϵ be a precision parameter, and θ be a threshold for the smallest singular value to consider. There exists a quantum algorithm that estimates the exact number k of singular values such that $\bar{\sigma}_i \geq \theta$, where $|\sigma_i - \bar{\sigma}_i| \leq \epsilon$, in time $\tilde{O}\left(\frac{\mu(\mathbf{A})}{\epsilon} \sqrt{(k+1)(r-k+1)}\right)$ with probability at least $\frac{2}{3}$.*

Similarly, given a parameter η , it is possible to produce an estimate \bar{k} such that $|\bar{k} - k| \leq \eta k$ in time $\tilde{O}\left(\frac{\mu(\mathbf{A})}{\epsilon \eta} \sqrt{\frac{r}{k}}\right)$ with probability at least $\frac{2}{3}$.

Estimating the number of singular values retained by θ is helpful. When the singular values are dense around θ , this Corollary, together with Theorem 9, can help the analyst evaluate trade-offs between big p and small k . On the one hand, the bigger p is, the more information on the dataset one can retain. On the other hand, the bigger k is, the slower will the algorithms in the next section be.

3.2 Extracting the SVD representation

After introducing the procedures to test for the most relevant singular values, factor scores and factor score ratios of \mathbf{A} ,

we present a routine to extract the corresponding right/left singular vectors. The inputs of this algorithm, other than the matrix, are a parameter δ for the precision of the singular vectors, a parameter ϵ for the precision of the singular value estimation, and a threshold θ to discard the non interesting singular values/vectors. The output guarantees a unit estimate $\bar{\mathbf{x}}_i$ of each singular vector such that $\|\mathbf{x}_i - \bar{\mathbf{x}}_i\| \leq \delta$, ensuring that the estimate has a similar orientation to the original vector. Additionally, this subroutine can provide an estimation of the singular values greater than θ , to absolute error ϵ .

Theorem 12 (Top-k singular vectors extraction) *Let there be efficient quantum access to a matrix $\mathbf{A} \in \mathbb{R}^{n \times m}$, with singular value decomposition $\mathbf{A} = \sum_i^r \sigma_i \mathbf{u}_i \mathbf{v}_i^T$. Let $\delta > 0$ be a precision parameter for the singular vectors, $\epsilon > 0$ a precision parameter for the singular values, and $\theta > 0$ be a threshold such that \mathbf{A} has k singular values greater than θ . Define $p = \frac{\sum_{i:\bar{\sigma}_i \geq \theta} \sigma_i^2}{\sum_j^r \sigma_j^2}$. There exist quantum algorithms that estimate:*

- The top k left singular vectors \mathbf{u}_i of \mathbf{A} with unit vectors $\bar{\mathbf{u}}_i$ such that $\|\mathbf{u}_i - \bar{\mathbf{u}}_i\|_2 \leq \delta$ with probability at least $1 - 1/\text{poly}(n)$, in time $\tilde{O}\left(\frac{\|\mathbf{A}\|}{\theta} \frac{1}{\sqrt{p}} \frac{\mu(\mathbf{A})}{\epsilon} \frac{kn}{\delta^2}\right)$;

Algorithm 4 Quantum top-k singular vectors extraction.

Input: Quantum access to a matrix $\mathbf{A} \in \mathbb{R}^{n \times m}$. A threshold θ that captures the top-k singular values. Two precision parameters $\delta, \epsilon \in \mathbb{R}_{>0}$.

Input: The top-k singular vectors such that $\|\mathbf{u}_i - \bar{\mathbf{u}}_i\| \leq \delta$ and $\|\mathbf{v}_i - \bar{\mathbf{v}}_i\| \leq \delta$. Optionally the top-k singular values, such that $|\sigma_i - \bar{\sigma}_i| \leq \epsilon$.

- 1: Prepare the state $\frac{1}{\|\mathbf{A}\|_F} \sum_i^n \sum_j^m a_{ij} |i\rangle |j\rangle$.
 - 2: Apply SVE to get $\frac{1}{\sqrt{\sum_j^r \sigma_j^2}} \sum_i^r \sigma_i |\mathbf{u}_i\rangle |\mathbf{v}_i\rangle |\bar{\sigma}_i\rangle$, where $|\sigma_i - \bar{\sigma}_i| \leq \epsilon$.
 - 3: Append a quantum register $|0\rangle$ to the state and set it to $|1\rangle$ if $|\bar{\sigma}_i\rangle < \theta$.
 - 4: Perform amplitude amplification for $|0\rangle$, to get the state $\frac{1}{\sqrt{\sum_j^k \sigma_j^2}} \sum_i^k \sigma_i |\mathbf{u}_i\rangle |\mathbf{v}_i\rangle |\bar{\sigma}_i\rangle$.
 - 5: Append a second ancillary register $|0\rangle$ and perform the controlled rotation $\frac{C}{\|\mathbf{A}^{(k)}\|_F} \sum_i^k \frac{\sigma_i}{\bar{\sigma}_i} |\mathbf{u}_i\rangle |\mathbf{v}_i\rangle |\bar{\sigma}_i\rangle |0\rangle + \frac{1}{\|\mathbf{A}^{(k)}\|_F} \sum_i^k \sqrt{1 - \frac{C^2}{\bar{\sigma}_i^2}} |\mathbf{u}_i\rangle |\mathbf{v}_i\rangle |\bar{\sigma}_i\rangle |1\rangle$ where C is a normalization constant.
 - 6: Perform again amplitude amplification for $|0\rangle$ to get the uniform superposition $\frac{1}{\sqrt{k}} \sum_i^k |\mathbf{u}_i\rangle |\mathbf{v}_i\rangle |\bar{\sigma}_i\rangle$.
 - 7: Measure the last register and, according to the measured $|\bar{\sigma}_i\rangle$, apply state-vector tomography on $|\mathbf{u}_i\rangle$ for the i^{th} left singular vector or on $|\mathbf{v}_i\rangle$ for the right one.
 - 8: Repeat 1-7 until the tomography requirements are met.
 - 9: Output the k singular vectors \mathbf{u}_i or \mathbf{v}_i and, optionally, the singular values $\bar{\sigma}_i$.
-

- The top k right singular vectors \mathbf{v}_i of \mathbf{A} with unit vectors $\bar{\mathbf{v}}_i$ such that $\|\mathbf{v}_i - \bar{\mathbf{v}}_i\|_2 \leq \delta$ with probability at least $1 - 1/\text{poly}(m)$, in time $\tilde{O}\left(\frac{\|\mathbf{A}\|}{\theta} \frac{1}{\sqrt{p}} \frac{\mu(\mathbf{A})}{\epsilon} \frac{km}{\delta^2}\right)$.
- The top k singular values σ_i , factor scores λ_i , and factor score ratios $\lambda^{(i)}$ of \mathbf{A} to precision ϵ , $2\epsilon\sqrt{\lambda_i}$, and $\epsilon \frac{\sigma_i}{\|\mathbf{A}\|_F^2}$ respectively, with probability at least $1 - 1/\text{poly}(m)$, in time $\tilde{O}\left(\frac{\|\mathbf{A}\|}{\theta} \frac{1}{\sqrt{p}} \frac{\mu(\mathbf{A})k}{\epsilon}\right)$ or during any of the two procedures above.

The proof consists in proving the time complexity and the error of Algorithm 4.

Proof Like in the previous proofs, the cost of preparing the state at step 4, is $\tilde{O}\left(\frac{1}{\sqrt{p}} \frac{\mu(\mathbf{A})}{\epsilon}\right)$, where $\tilde{O}\left(\frac{\mu(\mathbf{A})}{\epsilon}\right)$ is the cost of singular value estimation and $\tilde{O}\left(\frac{1}{\sqrt{p}}\right)$ is the one of amplitude amplification. Step 5 is a conditional rotation and similarly to step 3 it has a negligible cost. The next step is to analyze the amplitude amplification at 6. The constant C is a normalization factor in the order of $\tilde{O}(1/\kappa(\mathbf{A}^{(k)}))$ where $\kappa(\mathbf{A}^{(k)}) = \frac{\sigma_{\max}}{\sigma_{\min}}$ is the condition number of the low-rank matrix $\mathbf{A}^{(k)}$. Since for construction $\sigma_{\min} \geq \theta$, we can bound the condition number $\kappa(\mathbf{A}^{(k)}) \leq \frac{\|\mathbf{A}\|}{\theta}$. From the famous work of Harrow, Hassidim and Lloyd Harrow et al. (2009) we know that applying amplitude amplification on the state above, with the the third register being $|0\rangle$, would cost $T(U_6) \sim \tilde{O}(\kappa(\mathbf{A}^{(k)})T(U_5)) \sim \tilde{O}\left(\frac{\|\mathbf{A}\|}{\theta} \frac{1}{\sqrt{p}} \frac{\mu(\mathbf{A})}{\epsilon}\right)$.

This last amplitude amplification leaves the registers in the state

$$\frac{1}{\sqrt{\sum_i^k \frac{\sigma_i^2}{\bar{\sigma}_i^2}}} \sum_i^k \frac{\sigma_i}{\bar{\sigma}_i} |\mathbf{u}\rangle_i |\mathbf{v}\rangle_i |\bar{\sigma}_i\rangle \sim \frac{1}{\sqrt{k}} \sum_i^k |\mathbf{u}\rangle_i |\mathbf{v}\rangle_i |\bar{\sigma}_i\rangle \tag{4}$$

where $\bar{\sigma}_i \in [\sigma_i - \epsilon, \sigma_i + \epsilon]$ and $\frac{\sigma_i}{\bar{\sigma}_i} \rightarrow 1$ for $\epsilon \rightarrow 0$.

When measuring the last register of state 6 in the computational basis, we measure $|\bar{\sigma}_i\rangle$ and the first two registers collapse in the state $|\mathbf{u}\rangle_i |\mathbf{v}\rangle_i$. It is possible to perform vector-state tomography on $|\mathbf{u}\rangle_i |\mathbf{v}\rangle_i$, using Theorem 6 on the first register to retrieve $\bar{\mathbf{u}}_i$, or on the second one to retrieve $\bar{\mathbf{v}}_i$. The costs are $O\left(\frac{n \log n}{\delta^2}\right)$ and $O\left(\frac{m \log m}{\delta^2}\right)$, respectively. Using a coupon collector’s argument (Erdős and Rényi 1961), if the k states $|\bar{\sigma}_i\rangle$ are uniformly distributed, to get all the k possible couples $|\mathbf{u}\rangle_i |\mathbf{v}\rangle_i$ at least once, we would need $k \log k$ measurements on average. This proves that it is possible to estimate all the singular values, factor scores and factor score ratios with the guarantees of Theorem 3 in time $\tilde{O}\left(\frac{\|\mathbf{A}\|}{\theta} \frac{1}{\sqrt{p}} \frac{\mu(\mathbf{A})k}{\epsilon}\right)$.

To perform tomography on each state-vector, one should satisfy the coupon collector the same number of times as the measurements needed by the tomography procedure. The costs of the tomography for all the vectors $\{\bar{\mathbf{u}}_i\}_i^k$ and $\{\bar{\mathbf{v}}_i\}_i^k$ are $O\left(T(U_{(6)}) \frac{k \log k \cdot n \log n}{\delta^2}\right)$, and $O\left(T(U_{(6)}) \frac{k \log k \cdot m \log m}{\delta^2}\right)$. Therefore, the following complexities are proven: $\tilde{O}\left(\frac{\|\mathbf{A}\|}{\theta} \frac{1}{\sqrt{p}} \frac{\mu(\mathbf{A})}{\epsilon} \frac{km}{\delta^2}\right), \tilde{O}\left(\frac{\|\mathbf{A}\|}{\theta} \frac{1}{\sqrt{p}} \frac{\mu(\mathbf{A})}{\epsilon} \frac{km}{\delta^2}\right)$.

In the appendix, Section 6.3., we provide experiments that show that the coupon collector’s argument of Eq. 4 is accurate for practical ϵ . Besides $1/\sqrt{p}$ being negligible, it is interesting to note that the parameter θ can be computed using: 1. the procedures of Theorems 8, 9; 2. the binary search of Theorem 10; 3. the available literature on the type of data stored in the input matrix \mathbf{A} .

About the latter, the original paper of latent semantic indexing (Deerwester et al. 1990) states that the first $k = 100$ singular values are enough for a good representation. We believe that, in the same way, fixed thresholds θ can be defined for different machine learning applications. The experiments of Kerenidis and Luongo (2020) on the run-time parameters of the polynomial expansions of the MNIST dataset support this expectation: even though in qSFA they keep the k smallest singular values and refer to θ as the biggest singular value to retain, this value does not vary much when the dimensionality of their dataset grows. In our experiments, we observe that different datasets for image classification have similar θ s. For completeness, we also state a different version of Theorem 12, with ℓ_∞ guarantees on the vectors.

Corollary 13 (Fast top-k singular vectors extraction) *The run-times of 12 can be improved to $\tilde{O}\left(\frac{\|\mathbf{A}\|}{\theta} \frac{1}{\sqrt{p}} \frac{\mu(\mathbf{A})}{\epsilon} \frac{k}{\delta^2}\right)$ with estimation guarantees on the ℓ_∞ norms.*

Proof The proof consists in using ℓ_∞ tomography (Theorem 7) at step 7 of Algorithm 4.

Note that, given a vector with d non-zero entries, performing ℓ_∞ tomography with error $\frac{\delta}{\sqrt{d}}$ provides the same guarantees of ℓ_2 tomography with error δ . This implies that the extraction of the singular vectors with ℓ_2 guarantees can be faster if we can make assumptions on their sparseness: $\tilde{O}\left(\frac{\|\mathbf{A}\|}{\theta} \frac{1}{\sqrt{p}} \frac{\mu(\mathbf{A})}{\epsilon} \frac{kd}{\delta^2}\right)$.

4 Applications to machine learning

The new quantum procedures can be used for principal component analysis, correspondence analysis, and latent semantic analysis. Besides extracting the orthogonal factors and measuring their importance, we provide a procedure to

represent the data in PCA’s reduced feature space on a quantum computer. In a similar way, it is possible to compute the representations of CA and LSA.

4.1 Principal component analysis

Principal component analysis is a widely-used multivariate statistical method for continuous variables with applications in machine learning. Its uses range from outlier detection to dimensionality reduction and data visualization. Given a matrix $\mathbf{A} \in \mathbb{R}^{n \times m}$ storing information about n data points with m coordinates, its *principal components* are the set of orthogonal vectors along which the variance of the data points is maximized. The goal of PCA is to compute the principal components with the amount of variance they capture and rotate the data points to express their coordinates along the principal components. It is possible to represent the data using only the k coordinates that express the most variance for dimensionality reduction.

Model The model of PCA is closely related to the singular value decomposition of the data matrix \mathbf{A} , shifted to row mean 0. The model consists of the principal components and the amount of variance they explain. The principal components coincide with the right singular vectors \mathbf{v}_i , the factor scores $\lambda_i = \sigma_i^2$ represent the amount of variance along each of them, and the factor score ratios $\lambda^{(i)} = \frac{\lambda_i}{\sum_j \lambda_j}$ express the percentage of retained variance. For datasets with 0 mean, the transformation consists in a rotation along the principal components: $\mathbf{Y} = \mathbf{A}\mathbf{V} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T\mathbf{V} = \mathbf{U}\mathbf{\Sigma} \in \mathbb{R}^{n \times m}$. When performing dimensionality reduction, it suffice to use the top k singular values and vectors.

Using the procedures from Section 3 it is possible to extract the model for principal component analysis. In particular, Theorems 8, 9, and 10 allow to retrieve information on the factor scores and on the factor score ratios, while Theorem 12 allows extracting the principal components. The run-time of the model extraction is the sum of the run-times of the theorems: $\tilde{O}\left(\left(\frac{1}{\gamma^2} + \frac{km}{\theta\delta^2}\right)\frac{\mu(\mathbf{A})}{\epsilon}\right)$. The model comes with the following guarantees: $|\sigma_i - \bar{\sigma}_i| \leq \frac{\epsilon}{2}$; $|\lambda_i - \bar{\lambda}_i| \leq \epsilon\sqrt{\lambda_i}$; $|\lambda^{(i)} - \bar{\lambda}^{(i)}| \leq \epsilon\frac{\sigma_i}{\|\mathbf{A}\|_F}$; $\|\mathbf{v}_i - \bar{\mathbf{v}}_i\| \leq \delta$ for $i \in \{0, \dots, k - 1\}$. This run-time is generally smaller than the number of elements of the input data matrix, providing polynomial speed-ups on the best classical routines for non-sparse matrices. In writing the time complexity of the routines, we have omitted the term $\frac{1}{\sqrt{p}}$ because usually p is chosen to be a number greater than 0.5 (generally in the order of 0.8/0.9).

When performing dimensionality reduction, the goal is to obtain the matrix $\mathbf{Y} = \mathbf{U}\mathbf{\Sigma} \in \mathbb{R}^{n \times k}$, where $\mathbf{U} \in \mathbb{R}^{n \times k}$ and $\mathbf{\Sigma} \in \mathbb{R}^{k \times k}$ are composed respectively of the top k left singular vectors and singular values. In Lemma 14, we provide a

theoretical error bound for \mathbf{Y} , using the estimated entries of \mathbf{U} and $\mathbf{\Sigma}$. For sake of completeness, the error bound is also stated for $\mathbf{V}\mathbf{\Sigma}$. These bounds stand regardless of how the singular values and vectors are extracted and hold when the multiplication is done with a classical computer.

Lemma 14 (Accuracy of $\bar{\mathbf{U}}\bar{\mathbf{\Sigma}}$ and $\bar{\mathbf{V}}\bar{\mathbf{\Sigma}}$) *Let $\mathbf{A} \in \mathbb{R}^{n \times m}$ be a matrix. Given some approximate procedures to retrieve estimates $\bar{\sigma}_i$ of the singular values σ_i such that $|\sigma_i - \bar{\sigma}_i| \leq \epsilon$ and unit estimates $\bar{\mathbf{u}}_i$ of the left singular vectors \mathbf{u}_i such that $\|\bar{\mathbf{u}}_i - \mathbf{u}_i\|_2 \leq \delta$, the error on $\mathbf{U}\mathbf{\Sigma}$ can be bounded as $\|\mathbf{U}\mathbf{\Sigma} - \bar{\mathbf{U}}\bar{\mathbf{\Sigma}}\|_F \leq \sqrt{\sum_j^k (\epsilon + \delta\sigma_j)^2}$. Similarly, $\|\mathbf{V}\mathbf{\Sigma} - \bar{\mathbf{V}}\bar{\mathbf{\Sigma}}\|_F \leq \sqrt{\sum_j^k (\epsilon + \delta\sigma_j)^2}$. Both are bounded by $\sqrt{k}(\epsilon + \delta\|\mathbf{a}\|)$.*

We prove this result for $\|\mathbf{U}\mathbf{\Sigma} - \bar{\mathbf{U}}\bar{\mathbf{\Sigma}}\|_F$. The proof for $\|\mathbf{V}\mathbf{\Sigma} - \bar{\mathbf{V}}\bar{\mathbf{\Sigma}}\|_F$ is analogous.

Proof We first bound the error on the columns:

$$\|\bar{\sigma}_i\bar{\mathbf{u}}_i - \sigma_i\mathbf{u}_i\| \leq \|(\sigma_i \pm \epsilon)\bar{\mathbf{u}}_i - \sigma_i\mathbf{u}_i\| = \|\sigma_i(\bar{\mathbf{u}}_i - \mathbf{u}_i) \pm \epsilon\bar{\mathbf{u}}_i\| \tag{5}$$

Because of the triangular inequality, $\|\sigma_i(\bar{\mathbf{u}}_i - \mathbf{u}_i) \pm \epsilon\bar{\mathbf{u}}_i\| \leq \sigma_i\|\bar{\mathbf{u}}_i - \mathbf{u}_i\| + \epsilon\|\bar{\mathbf{u}}_i\|$. Also by hypothesis, $\|(\bar{\mathbf{u}}_i - \mathbf{u}_i)\| \leq \delta$ and $\|\mathbf{u}_i\| = 1$. Thus, $\sigma_i\|\bar{\mathbf{u}}_i - \mathbf{u}_i\| + \epsilon\|\bar{\mathbf{u}}_i\| \leq \sigma_i\delta + \epsilon$. Since $f(x) = \sqrt{x}$ is an increasing monotone function, it is possible to prove:

$$\begin{aligned} \|\bar{\mathbf{U}}\bar{\mathbf{\Sigma}} - \mathbf{U}\mathbf{\Sigma}\|_F &= \sqrt{\sum_i^n \sum_j^k \|\bar{\sigma}_j\bar{\mathbf{u}}_{ij} - \sigma_j\mathbf{u}_{ij}\|^2} = \sqrt{\sum_j^k (\|\sigma_j\bar{\mathbf{u}}_j - \sigma_j\mathbf{u}_j\|)^2} \\ &\leq \sqrt{\sum_j^k (\epsilon + \delta\sigma_j)^2} \leq \sqrt{k(\epsilon + \delta\sigma_{max})^2} \leq \sqrt{k}(\epsilon + \delta\|\mathbf{A}\|) \end{aligned} \tag{6}$$

Using matrix-multiplication from Theorem 4, we can have algorithms to produce quantum states proportional to the data representation in the new feature space. Having access to $\mathbf{V}^{(k)} \in \mathbb{R}^{m \times k}$, these routines create the new data points in almost constant time and are helpful when chained to other quantum machine learning algorithms that need to be executed multiple times.

Corollary 15 (Quantum PCA: vector dimensionality reduction) *Let ξ be a precision parameter. Let there be efficient quantum access to the top k right singular vectors $\bar{\mathbf{V}}^{(k)} \in \mathbb{R}^{m \times k}$ of a matrix $\mathbf{A} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T \in \mathbb{R}^{n \times m}$, such that $\|\mathbf{V}^{(k)} - \bar{\mathbf{V}}^{(k)}\| \leq \frac{\xi}{\sqrt{2}}$. Given quantum access to a row \mathbf{a}_i of \mathbf{A} , the quantum state $|\bar{\mathbf{y}}_i\rangle = \frac{1}{\|\bar{\mathbf{y}}_i\|} \sum_k \bar{y}_k |i\rangle$, proportional to its projection onto the PCA space, can be created in time $\tilde{O}\left(\mu(\mathbf{V}^{(k)})\frac{\|\mathbf{a}_i\|}{\|\bar{\mathbf{y}}_i\|}\right)$ with probability at least $1 - 1/\text{poly}(m)$ and precision $\| |\mathbf{y}_i\rangle - |\bar{\mathbf{y}}_i\rangle \| \leq \frac{\|\mathbf{a}_i\|}{\|\bar{\mathbf{y}}_i\|} \xi$. An estimate of $\|\bar{\mathbf{y}}_i\|$, to relative error η , can be computed in $\tilde{O}(1/\eta)$.*

Proof Here with \mathbf{V} we denote $\mathbf{V}^{(k)} \in \mathbb{R}^{m \times k}$. Given a vector \mathbf{a}_i , its projection onto the k -dimensional PCA space of \mathbf{A} is $\mathbf{y}_i^T = \mathbf{a}_i^T \mathbf{V}$, or equivalently $\mathbf{y}_i = \mathbf{V}^T \mathbf{a}_i$. Note that $\|\mathbf{y}_i\| = \|\mathbf{V}^T \mathbf{a}_i\|$.

It is possible to use Theorem 4 to multiply the quantum state $|\mathbf{a}_i\rangle$ by \mathbf{V}^T , appropriately padded with 0s to be a square $\mathbb{R}^{m \times m}$ matrix. In this way, we can create an approximation $|\bar{\mathbf{y}}_i\rangle$ of the state $|\mathbf{y}_i\rangle = |\mathbf{v}\rangle^T \mathbf{a}_i$ in time $\tilde{O}\left(\frac{\mu(\mathbf{V}^T) \log(1/\epsilon)}{\gamma}\right)$ with probability $1 - 1/\text{poly}(m)$, such that $\| |\mathbf{y}_i\rangle - |\bar{\mathbf{y}}_i\rangle \| \leq \epsilon$. Since \mathbf{V}^T has rows with unit ℓ_2 norm, we can prepare efficient quantum access to it by creating access to its rows [Kerenidis and Prakash (2020a), Theorem IV.1]. Having $\gamma = \|\mathbf{V}^T \mathbf{a}_i\| / \|\mathbf{a}_i\|$, we get a run-time of $\tilde{O}\left(\mu(\mathbf{V}) \frac{\|\mathbf{a}_i\|}{\|\mathbf{y}_i\|} \log(1/\epsilon)\right)$. The term $\log(1/\epsilon)$ can be considered negligible. We conclude that the state $|\mathbf{y}_i\rangle$ can be created in time $\tilde{O}\left(\mu(\mathbf{V}) \frac{\|\mathbf{a}_i\|}{\|\mathbf{y}_i\|}\right)$ with probability $1 - 1/\text{poly}(m)$ and that its norm can be estimated to relative error η in time $\tilde{O}\left(\mu(\mathbf{V}) \frac{\|\mathbf{a}_i\|}{\|\mathbf{y}_i\|} \frac{1}{\eta}\right)$.

For what concerns the error, we start by bounding $\|\mathbf{y}_i - \bar{\mathbf{y}}_i\|$ and then use Claim 2 to bound the error on the quantum states. Assume to have estimates $\bar{\mathbf{v}}_i$ of the columns of \mathbf{V} such that $\|\mathbf{v}_i - \bar{\mathbf{v}}_i\| \leq \delta$.

$$\|\mathbf{V} - \bar{\mathbf{V}}\|_F = \sqrt{\sum_i^n \sum_j^k (v_{ij} - \bar{v}_{ij})^2} \leq \sqrt{k} \delta \tag{7}$$

Considering that $\|\mathbf{y}_i - \bar{\mathbf{y}}_i\| = \|\mathbf{a}_i^T \mathbf{V}^{(k)} - \mathbf{a}_i^T \bar{\mathbf{V}}^{(k)}\| \leq \|\mathbf{a}_i\| \sqrt{k} \delta$, we can use Claim 2 to state

$$\| |\mathbf{y}_i\rangle - |\bar{\mathbf{y}}_i\rangle \| \leq \frac{\|\mathbf{a}_i\|}{\|\mathbf{y}_i\|} \sqrt{2k} \delta = \frac{\|\mathbf{a}_i\|}{\|\mathbf{y}_i\|} \xi. \tag{8}$$

Setting $\delta = \frac{\xi}{\sqrt{2k}}$ leads to the requirement $\|\mathbf{V} - \bar{\mathbf{V}}\|_F \leq \frac{\xi}{\sqrt{2}}$.

This result also holds when \mathbf{a}_i is a previously unseen data point, not necessarily stored in \mathbf{A} . Note that from the row orthogonality of $\mathbf{V}^{(k)}$ it follows that $\mu(\mathbf{V}^{(k)}) \leq \|\mathbf{V}^{(k)}\|_F = \sqrt{k}$. Furthermore, $\frac{\|\mathbf{y}_i\|}{\|\mathbf{a}_i\|}$ is expected to be close to 1, as it is the percentage of support of \mathbf{a}_i on the new feature space spanned by $\mathbf{V}^{(k)}$. We formalize this better using Definition 2 below.

Definition 2 (PCA-representable data) A set of n data points described by m coordinates, represented through a matrix $\mathbf{A} = \sum_i^r \sigma_i \mathbf{u}_i \mathbf{v}_i^T \in \mathbb{R}^{n \times m}$ is said to be PCA-representable if there exists $p \in [\frac{1}{2}, 1]$, $\epsilon \in [0, 1/2]$, $\beta \in [p - \epsilon, p + \epsilon]$, $\alpha \in [0, 1]$ such that:

- $\exists k \in O(1)$ such that $\frac{\sum_i^k \sigma_i^2}{\sum_i^m \sigma_i^2} = p$
- for at least αn points \mathbf{a}_i it holds $\frac{\|\mathbf{y}_i\|}{\|\mathbf{a}_i\|} \geq \beta$, where $\|\mathbf{y}_i\| = \sqrt{\sum_i^k |\langle \mathbf{a}_i | \mathbf{v}_i \rangle|^2} \|\mathbf{a}_i\|$.

Claim 16 (Quantum PCA on PCA-representable datasets) Let \mathbf{a}_i be a row of $\mathbf{A} \in \mathbb{R}^{n \times d}$. Then, for $p \in [1/2, 1]$, the runtime of Corollary 15 is $\mu(\mathbf{V}) \frac{\|\mathbf{a}_i\|}{\|\bar{\mathbf{y}}_i\|} = \mu(\mathbf{V}) \frac{1}{\beta} = O(\mu(\mathbf{V}))$ with probability greater than α .

It is known that, in practical machine learning datasets, α is a number fairly close to one. We have tested the value of α for the MNIST, Fashion MNIST and CIFAR-10 datasets, finding values over 0.85 for any $p \in (0, 1]$.

The next corollary shows how to perform dimensionality reduction on the whole matrix, enabling quantum access to the data matrix in the reduced feature space.

Corollary 17 (Quantum PCA: matrix dimensionality reduction) Let ξ be a precision parameter and p be the amount of variance retained after the dimensionality reduction. Let there be efficient quantum access to $\mathbf{A} = \mathbf{U} \mathbf{\Sigma} \mathbf{V}^T \in \mathbb{R}^{n \times m}$ and to its top k right singular vectors $\bar{\mathbf{V}}^{(k)} \in \mathbb{R}^{m \times k}$, such that $\|\mathbf{V}^{(k)} - \bar{\mathbf{V}}^{(k)}\| \leq \frac{\xi \sqrt{p}}{\sqrt{2}}$. There exists a quantum algorithm that, with probability at least $1 - 1/\text{poly}(m)$, creates the state $|\bar{\mathbf{Y}}\rangle = \frac{1}{\|\mathbf{Y}\|_F} \sum_i^n \|\mathbf{y}_i\| |i\rangle |\mathbf{y}_i\rangle$, proportional to the projection of \mathbf{A} in the PCA subspace, with error $\| |\mathbf{Y}\rangle - |\bar{\mathbf{Y}}\rangle \| \leq \xi$ in time $\tilde{O}(\mu(\mathbf{V})/\sqrt{p})$. An estimate of $\|\bar{\mathbf{Y}}\|_F$, to relative error η , can be computed in $\tilde{O}\left(\frac{\mu(\mathbf{V})}{\sqrt{p\eta}}\right)$.

Proof Here with \mathbf{V} we denote $\mathbf{V}^{(k)} \in \mathbb{R}^{m \times k}$. Using the same reasoning as the proof above and giving a closer look at the proof of Theorem 4 (Lemma 24 (Chakraborty et al. 2019)), we see that it is possible to create the state $|0\rangle \left(\frac{\bar{\mathbf{V}}^T}{\mu(\mathbf{V})} |\mathbf{a}_i\rangle\right) + |0_\perp\rangle$ in time $\tilde{O}(1)$ and that the term $\frac{\mu(\mathbf{V})}{\gamma}$ is introduced to boost the probability of getting the right state. Indeed, if we apply Theorem 4 without the amplitude amplification step to the superposition of the rows of \mathbf{A} , we obtain the following mapping in time $\tilde{O}(1)$:

$$|\mathbf{A}\rangle = \frac{1}{\|\mathbf{A}\|_F} \sum_i^n \|\mathbf{a}_i\| |i\rangle |\mathbf{a}_i\rangle \mapsto \frac{1}{\|\mathbf{A}\|_F \mu(\mathbf{V})} \sum_i^n (\|\mathbf{y}_i\| |0\rangle |i\rangle |\mathbf{y}_i\rangle + \|\mathbf{y}_{i,\perp}\| |0_\perp\rangle), \tag{9}$$

where $\|\mathbf{y}_{i,\perp}\|$ are normalization factors. Keeping in mind that $\|\mathbf{A}\|_F = \sqrt{\sum_i^r \sigma_i^2}$ and $\|\mathbf{Y}\|_F = \sqrt{\sum_i^n \|\mathbf{y}_i\|^2} = \sqrt{\sum_i^k \sigma_i^2}$, we see that the amount of explained variance is $p = \frac{\sum_i^k \sigma_i^2}{\sum_i^r \sigma_i^2} = \left(\frac{\|\mathbf{Y}\|_F}{\|\mathbf{A}\|_F}\right)^2$. The probability of obtaining $|\mathbf{Y}\rangle = \frac{1}{\|\mathbf{Y}\|_F} \sum_i^n \|\mathbf{y}_i\| |i\rangle |\mathbf{y}_i\rangle$ is $\frac{p}{\mu(\mathbf{V})^2} = \frac{\|\mathbf{Y}\|_F^2}{\|\mathbf{A}\|_F^2 \mu(\mathbf{V})^2} = \frac{\sum_i^n \|\mathbf{y}_i\|^2}{\|\mathbf{A}\|_F^2 \mu(\mathbf{V})^2}$. We conclude that, using $\tilde{O}(\mu(\mathbf{V})/\sqrt{p})$ rounds of amplitude amplification, we obtain $|\mathbf{Y}\rangle$ with probability $1 - 1/\text{poly}(m)$ (Theorem 5).

For the error, consider that $\|\mathbf{Y} - \bar{\mathbf{Y}}\| = \|\mathbf{A} \mathbf{V}^{(k)} - \mathbf{A} \bar{\mathbf{V}}^{(k)}\| \leq \|\mathbf{A}\| \sqrt{k} \delta$, so we can use Claim 2 to state

$$\| |\mathbf{Y}\rangle - |\bar{\mathbf{Y}}\rangle \| \leq \frac{\|\mathbf{A}\|_F}{\|\mathbf{Y}\|_F} \sqrt{2k} \delta = \xi. \tag{10}$$

We can set $\delta = \frac{\xi}{\sqrt{2k}} \frac{\|\mathbf{Y}\|_F}{\|\mathbf{A}\|_F} = \frac{\xi \sqrt{p}}{\sqrt{2k}}$, so we require $\|\mathbf{V} - \bar{\mathbf{V}}\|_F \leq \frac{\xi \sqrt{p}}{\sqrt{2}}$.

The error requirements of the two corollaries propagate to the run-time of the model extraction in the following way.

Corollary 18 (Quantum PCA: fitting time) *Let ϵ be a precision parameter and $p = \frac{\sum_{i:\bar{\sigma}_i \geq 0} \sigma_i^2}{\sum_j \sigma_j^2}$ the amount of variance to retain, where $|\sigma_i - \bar{\sigma}_i| \leq \epsilon$. Given efficient quantum access to a matrix $\mathbf{A} \in \mathbb{R}^{n \times m}$, the run-time to extract $\mathbf{V}^{(k)} \in \mathbb{R}^{m \times k}$ for corollaries 15, 17 is $\tilde{O}\left(\frac{\mu(\mathbf{A})k^2m}{\theta \epsilon \xi^2}\right)$.*

Proof The procedure to train the model consists in using Theorem 10 or 8 to extract the threshold θ , given the amount of variance to retain p , and to leverage Theorem 12 to extract the k right singular vectors that compose $\mathbf{V} \in \mathbb{R}^{m \times k}$. The run-time of Theorem 10 and 8 are smaller than the one of Theorem 12, so we can focus on the last one. To have $\|\mathbf{V} - \bar{\mathbf{V}}\|_F \leq \frac{\xi \sqrt{p}}{\sqrt{2}}$ we need $\|\mathbf{v}_i - \bar{\mathbf{v}}_i\| \leq \frac{\xi \sqrt{p}}{\sqrt{2k}}$. Substituting $\delta = \frac{\xi \sqrt{p}}{\sqrt{2k}}$ in the run-time of Theorem 12, we get $\tilde{O}\left(\frac{\mu(\mathbf{A})k^2m}{p^{3/2} \theta \epsilon \xi^2}\right)$. If we consider that p to be a reasonable number (e.g., at least greater than 0.05), we can consider it a constant factor that is independent from the input's size. The asymptotic run-time is proven to be $\tilde{O}\left(\frac{\mu(\mathbf{A})k^2m}{\theta \epsilon \xi^2}\right)$.

When training the model for Corollary 17, the run-time has a dependency on $1/p^{3/2}$. However, this term is constant and independent from the size of the input dataset. With this additional $1/p^{3/2}$ cost, the error of Corollary 15 drops to ξ for every row of the matrix and generally decreases in case of new data points.

Using the same framework and proof techniques, it is possible to produce similar results for the representations of CA and LSA.

Remark: Note that [Yu et al. (2019), Theorem 1] propose a lower bound for a quantity similar to our α . However, their result seems to be a loose bound: using their notation and setting $\eta = 1, \theta = 1$ they bound this quantity with 0, while a tight bound should give 1.

4.2 Correspondence analysis

Correspondence analysis is a multivariate statistical tool from the family of *factor analysis* methods. It is used to explore relationships among categorical variables. Given two random variables, X and Y , with possible outcomes in $\{x_1, \dots, x_n\}$ and $\{y_1, \dots, y_m\}$, the model of Correspondence Analysis enables

representing the outcomes as vectors in two related Euclidean spaces. These vectors can be used for data visualization, exploration, and other unsupervised machine learning tasks.

Model Given a contingency table for X and Y (see Section 2), it is possible to compute the matrix $\mathbf{A} = \mathbf{D}_X^{-1/2} (\hat{\mathbf{P}}_{X,Y} - \hat{\mathbf{p}}_X \hat{\mathbf{p}}_Y^T) \mathbf{D}_Y^{-1/2} \in \mathbb{R}^{n \times m}$, where $\hat{\mathbf{P}}_{X,Y} \in \mathbb{R}^{n \times m}$ is the estimated matrix of joint probabilities, $\hat{\mathbf{p}}_X \in \mathbb{R}^n$ and $\hat{\mathbf{p}}_Y \in \mathbb{R}^m$ are the vectors of marginal probabilities, and $\mathbf{D}_X^{-1/2} = \text{diag}(\hat{\mathbf{p}}_X)$, $\mathbf{D}_Y^{-1/2} = \text{diag}(\hat{\mathbf{p}}_Y)$. The computation of \mathbf{A} requires linear time in the non-zero entries of the contingency table. The singular value decomposition of \mathbf{A} is strictly related to the model of correspondence analysis (Greenacre 1984; Hsu et al. 2019). The new coordinates of X 's outcomes are given by the rows of $\mathbf{D}_X^{-1/2} \mathbf{U} \in \mathbb{R}^{n \times k}$, while the ones of Y by the rows of $\mathbf{D}_Y^{-1/2} \mathbf{V} \in \mathbb{R}^{m \times k}$. Like in PCA, it is possible to choose only a subset of the orthogonal factors as coordinates for the representation. Factor scores and factor score ratios measure of how much ‘‘correspondence’’ is captured by the respective orthogonal factor, giving an estimate of the quality of the representation.

Similarly to what we have already discussed, it is possible to extract the model for CA by creating quantum access to the matrix \mathbf{A} and using Theorems 8, 9, and 12 to extract the orthogonal factors, the factor scores and the factor score ratios in time $\tilde{O}\left(\left(\frac{1}{\gamma^2} + \frac{k(n+m)}{\theta \delta^2}\right) \frac{\mu(\mathbf{A})}{\epsilon}\right)$. We provide a theoretical bound for the data representations in Lemma 19.

Lemma 19 (Accuracy of $\mathbf{D}_X^{-1/2} \mathbf{U}$ and $\mathbf{D}_Y^{-1/2} \mathbf{V}$) *Let $\mathbf{A} \in \mathbb{R}^{n \times m}$ be a matrix. Given some approximate procedures to retrieve unit estimates $\bar{\mathbf{u}}_i$ of the left singular vectors \mathbf{u}_i such that $\|\bar{\mathbf{u}}_i - \mathbf{u}_i\| \leq \delta$, the error on $\mathbf{D}_X^{-1/2} \mathbf{U}$ can be bounded as $\|\mathbf{D}_X^{-1/2} \mathbf{U} - \mathbf{D}_X^{-1/2} \bar{\mathbf{U}}\|_F \leq \|\mathbf{D}_X^{-1/2}\|_F \sqrt{k} \delta$. Similarly, $\|\mathbf{D}_Y^{-1/2} \mathbf{V} - \mathbf{D}_Y^{-1/2} \bar{\mathbf{V}}\|_F \leq \|\mathbf{D}_Y^{-1/2}\|_F \sqrt{k} \delta$.*

Proof It suffices to note that $\|\mathbf{D}_X^{-1/2} \bar{\mathbf{U}} - \mathbf{D}_X^{-1/2} \mathbf{U}\|_F \leq \|\mathbf{D}_X^{-1/2}\|_F \|\bar{\mathbf{U}} - \mathbf{U}\|_F \leq \|\mathbf{D}_X^{-1/2}\|_F \sqrt{k} \delta$. Similar conclusions can be drawn for $\|\mathbf{D}_Y^{-1/2} \bar{\mathbf{V}} - \mathbf{D}_Y^{-1/2} \mathbf{V}\|_F$.

4.3 Latent semantic analysis

Latent semantic analysis is a data representation method used to represent words and text documents as vectors in Euclidean spaces. Using these vector spaces, it is possible to compare terms, documents, and terms and documents. LSA spaces automatically model synonymy and polysemy (Deerwester et al. 1990), and their applications in machine learning range from topic modeling to document clustering and retrieval.

Model The input of LSA is a contingency table of n words and m documents $\mathbf{A} \in \mathbb{R}^{n \times m}$. Inner products of rows are a measure of words similarity, and can be computed at once as

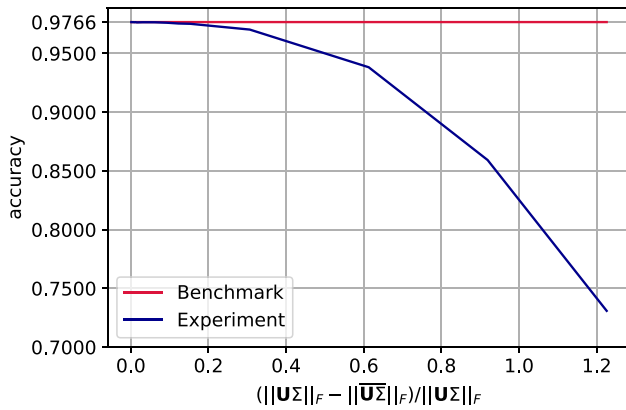


Fig. 1 Accuracy of 10-fold cross-validation using K-Nearest-Neighbors, with 7 neighbors, on the MNIST dataset after PCA’s dimensionality reduction (0.8580% of variance retained). The *benchmark* accuracy was computed with an exact PCA. The *experiment* line shows how the classification accuracy decreases as error is introduced in the Frobenius norm of the representation

$\mathbf{A}\mathbf{A}^T = \mathbf{U}\Sigma^2\mathbf{U}^T$. Inner products of columns $\mathbf{A}^T\mathbf{A} = \mathbf{V}\Sigma^2\mathbf{V}^T$ are a measure of documents similarity, and the a_{ij} entry of $\mathbf{A} = \mathbf{U}\Sigma\mathbf{V}^T$ is a measure of similarity between word i and document j . We can use SVD to express words and documents in new spaces where we can compare them with respect to this similarity measure. In particular, we can compute: 1. a representation for word comparisons $\mathbf{U}\Sigma \in \mathbb{R}^{n \times k}$; 2. a representation for document comparisons $\mathbf{V}\Sigma \in \mathbb{R}^{m \times k}$; 3. two representations for word and document comparisons $\mathbf{U}\Sigma^{1/2} \in \mathbb{R}^{n \times k}$ and $\mathbf{V}\Sigma^{1/2} \in \mathbb{R}^{m \times k}$.

When using LSA for document indexing, like in a search engine, we need to represent the query as a vector in the document space. In this case, instead of increasing \mathbf{A} ’s size and recomputing the document space, the new vector can be expressed as $\mathbf{v}_q^T = \mathbf{x}_q^T \mathbf{U}\Sigma^{-1}$, where $\mathbf{x}_q \in \mathbb{R}^n$ is obtained using the same criteria used to store a document in \mathbf{A} . The representation of the query can then be used to compare the query to the other documents in the document representation space. Finally, factor score ratios play an important role in LSA too. For instance, the columns of \mathbf{V} can be seen as latent topics of the corpus. The importance of each topic is proportional to the corresponding factor score ratio. This paragraph only stresses how computing the SVD of \mathbf{A} is connected to LSA. For a better introduction to LSA and indexing, we invite the reader to consult the original paper (Deerwester et al. 1990).

Even in this case, the cost of extracting the orthogonal factors and the factor scores is bounded by $\tilde{O}\left(\left(\frac{1}{\gamma^2} + \frac{k(n+m)}{\theta\delta^2}\right)\frac{\mu(\mathbf{A})}{\epsilon}\right)$. In some applications, the data analyst might use a fixed number of singular values and vectors, regardless of the factor score ratios. In Deerwester et al.

Table 1 Summary of the run-time parameters. The parameters that depend on k have been computed using the estimated k

Parameter	MNIST	F-MNIST	CIFAR-10
$\mu(\mathbf{A}) = \ \mathbf{a}\ _F$	3.2032	1.8551	1.8540
Estimated k	62	45	55
Exact k	59	43	55
Estimated p	0.8510	0.8510	0.8510
Exact p	0.8580	0.8543	0.8514
Thrs. ϵ	0.0030	0.0009	0.0006
θ	0.1564	0.0776	0.0746
δ	0.1124	0.0106	0.0340

(1990), $k = 100$ is found to be a good number for document indexing. Similarly, we believe that if we scale the singular values by the spectral norm, it is possible to empirically determine a threshold θ to use in practice. Determining such threshold would reduce the complexity of model computation to the one of Theorem 12: $\tilde{O}\left(\frac{k(n+m)}{\theta\delta^2} \frac{\mu(\mathbf{A})}{\epsilon}\right)$.

For what concerns the error bounds, we already know that it is possible to retrieve an approximation $\overline{\mathbf{U}\Sigma}$ and $\overline{\mathbf{V}\Sigma}$ with precision $\sqrt{k}(\epsilon + \delta\|\mathbf{a}\|)$ (Lemma 14), where δ is the precision on the singular vectors and ϵ the precision on the singular values. To provide bounds on the estimations of $\mathbf{U}\Sigma^{1/2}$, $\mathbf{V}\Sigma^{1/2}$, and $\mathbf{U}\Sigma^{-1}$ we introduce Lemmas 20 and Lemma 21.

Lemma 20 (Accuracy of $\overline{\mathbf{U}\Sigma}^{1/2}$ and $\overline{\mathbf{V}\Sigma}^{1/2}$) *Let $\mathbf{A} \in \mathbb{R}^{n \times m}$ be a matrix. Given some approximate procedures to retrieve estimates $\overline{\sigma}_i$ of the singular values σ_i such that $|\overline{\sigma}_i - \sigma_i| \leq \epsilon$ and unitary estimates $\overline{\mathbf{u}}_i$ of the left singular vectors \mathbf{u}_i such that $\|\overline{\mathbf{u}}_i - \mathbf{u}_i\| \leq \delta$, the error on $\mathbf{U}\Sigma^{1/2}$ can be bounded as*

$$\|\mathbf{U}\Sigma^{1/2} - \overline{\mathbf{U}\Sigma}^{1/2}\|_F \leq \sqrt{\sum_j^k \left(\delta\sqrt{\sigma_j} + \frac{\epsilon}{2\sqrt{\theta}}\right)^2}.$$

Similarly, $\|\mathbf{V}\Sigma^{1/2} - \overline{\mathbf{V}\Sigma}^{1/2}\|_F \leq \sqrt{\sum_j^k \left(\delta\sqrt{\sigma_j} + \frac{\epsilon}{2\sqrt{\theta}}\right)^2}$. Both are bounded by $\sqrt{k}\left(\delta\sqrt{\|\mathbf{A}\|} + \frac{\epsilon}{2\sqrt{\theta}}\right)$

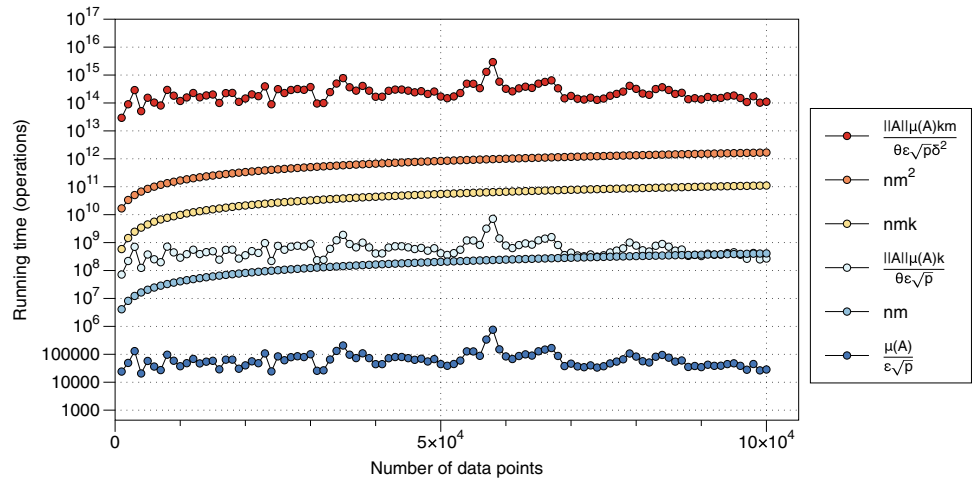
We prove this result for $\|\overline{\mathbf{U}\Sigma}^{1/2} - \mathbf{U}\Sigma^{1/2}\|_F$.

Proof We start by bounding $|\sqrt{\overline{\sigma}_i} - \sqrt{\sigma_i}|$. Let’s define $\epsilon = \gamma\sigma_i$ as a relative error:

$$\begin{aligned} |\sqrt{\sigma_i + \epsilon} - \sqrt{\sigma_i}| &= |\sqrt{\sigma_i + \gamma\sigma_i} - \sqrt{\sigma_i}| = |\sqrt{\sigma_i}(\sqrt{1 + \gamma} - 1)| \\ &= \sqrt{\sigma_i} \left| \frac{(\sqrt{1 + \gamma} - 1)(\sqrt{1 + \gamma} + 1)}{\sqrt{1 + \gamma} + 1} \right| \\ &= \sqrt{\sigma_i} \left| \frac{\gamma + 1 - 1}{\sqrt{1 + \gamma} + 1} \right| \leq \sqrt{\sigma_i} \frac{\gamma}{2}. \end{aligned} \tag{11}$$

By definition $\gamma = \frac{\epsilon}{\sigma_i}$ and we know that $\sigma_{\min} \geq \theta$:

Fig. 2 Run-time comparison on Imagenet as the number of data points increases. The plots have been computed setting $\delta = 0.1$ and $p = 0.85$ and are logarithmic w.r.t. the y axis



$$|\sqrt{\bar{\sigma}_i} - \sqrt{\sigma_i}| \leq \frac{\sqrt{\sigma_i} \epsilon}{\sigma_i 2} = \frac{\epsilon}{2\sqrt{\sigma_i}} \leq \frac{\epsilon}{2\sqrt{\theta}}. \tag{12}$$

Using the bound on the square roots, we can bound the columns of $\overline{\mathbf{U}\Sigma}^{1/2}$:

$$\begin{aligned} \left\| \sqrt{\bar{\sigma}_i} \bar{\mathbf{u}}_i - \sqrt{\sigma_i} \mathbf{u}_i \right\| &\leq \left\| \left(\sqrt{\sigma_i} + \frac{\epsilon}{2\sqrt{\theta}} \right) \bar{\mathbf{u}}_i - \sqrt{\sigma_i} \mathbf{u}_i \right\| = \\ \left\| \sqrt{\sigma_i} (\bar{\mathbf{u}}_i - \mathbf{u}_i) + \frac{\epsilon}{2\sqrt{\theta}} \bar{\mathbf{u}}_i \right\| &\leq \sqrt{\sigma_i} \delta + \frac{\epsilon}{2\sqrt{\theta}} \end{aligned} \tag{13}$$

From the error bound on the columns we derive the bound on the matrices:

$$\begin{aligned} \left\| \overline{\mathbf{U}\Sigma}^{1/2} - \mathbf{U}\Sigma^{1/2} \right\|_F &= \sqrt{\sum_j^k \left(\left\| \sqrt{\bar{\sigma}_j} \bar{\mathbf{u}}_j - \sqrt{\sigma_j} \mathbf{u}_j \right\|^2 \right)} \\ &\leq \sqrt{\sum_j^k \left(\delta \sqrt{\sigma_j} + \frac{\epsilon}{2\sqrt{\theta}} \right)^2} \leq \sqrt{k} \left(\delta \sqrt{\|\mathbf{A}\|} + \frac{\epsilon}{2\sqrt{\theta}} \right). \end{aligned} \tag{14}$$

Lemma 21 (Accuracy of $\overline{\mathbf{U}\Sigma}^{-1}$ and $\overline{\mathbf{V}\Sigma}^{-1}$) *Let $\mathbf{A} \in \mathbb{R}^{n \times m}$ be a matrix. Given some approximate procedures to retrieve estimates $\bar{\sigma}_i$ of the singular values σ_i such that $|\bar{\sigma}_i - \sigma_i| \leq \epsilon$ and unitary estimates $\bar{\mathbf{u}}_i$ of the left singular vectors \mathbf{u}_i such that $\|\bar{\mathbf{u}}_i - \mathbf{u}_i\| \leq \delta$, the error on $\mathbf{U}\Sigma^{-1}$ can be bounded as $\|\mathbf{U}\Sigma^{-1} - \overline{\mathbf{U}\Sigma}^{-1}\|_F \leq \sqrt{k} \left(\frac{\delta}{\theta} + \frac{\epsilon}{\theta^2 - \theta\epsilon} \right)$. Similarly, $\|\mathbf{V}\Sigma^{-1} - \overline{\mathbf{V}\Sigma}^{-1}\|_F \leq \sqrt{k} \left(\frac{\delta}{\theta} + \frac{\epsilon}{\theta^2 - \theta\epsilon} \right)$.*

We prove this result for $\|\overline{\mathbf{U}\Sigma}^{-1} - \mathbf{U}\Sigma^{-1}\|_F$.

Proof We start by bounding $|\frac{1}{\bar{\sigma}_i} - \frac{1}{\sigma_i}|$. Knowing that $\sigma_{\min} \geq \theta$ and $\epsilon < \theta$:

$$\left| \frac{1}{\bar{\sigma}_i} - \frac{1}{\sigma_i} \right| \leq \left| \frac{1}{\sigma_i - \epsilon} - \frac{1}{\sigma_i} \right| \leq \frac{\epsilon}{\theta^2 - \theta\epsilon}. \tag{15}$$

From the bound on the inverses, we can obtain the bound on the columns of $\overline{\mathbf{U}\Sigma}^{-1}$:

$$\left\| \frac{1}{\bar{\sigma}_i} \bar{\mathbf{u}}_i - \frac{1}{\sigma_i} \mathbf{u}_i \right\| \leq \left\| \left(\frac{1}{\sigma_i \pm \frac{\epsilon}{\theta^2 - \theta\epsilon}} \right) \bar{\mathbf{u}}_i - \frac{1}{\sigma_i} \mathbf{u}_i \right\| \leq \frac{1}{\sigma_i} \delta + \frac{\epsilon}{\theta^2 - \theta\epsilon} \leq \frac{\delta}{\theta} + \frac{\epsilon}{\theta^2 - \theta\epsilon}. \tag{16}$$

To complete the proof, we compute the bound on the matrices:

$$\left\| \overline{\mathbf{U}\Sigma}^{-1} - \mathbf{U}\Sigma^{-1} \right\|_F = \sqrt{\sum_j^k \left(\left\| \frac{1}{\bar{\sigma}_j} \bar{\mathbf{u}}_j - \frac{1}{\sigma_j} \mathbf{u}_j \right\|^2 \right)} \leq \sqrt{k} \left(\frac{\delta}{\theta} + \frac{\epsilon}{\theta^2 - \theta\epsilon} \right) \tag{17}$$

5 Experiments

All of our experiments are numerical and can be carried out on classical computers.¹ We have analysed the distribution of the factor score ratios in the MNIST, Fashion MNIST, CIFAR-10, Tiny Imagenet and Research Papers datasets. They decrease exponentially fast (figures in the appendix), confirming the low rank nature of the data. Focusing on MNIST, Fashion-MNIST, and CIFAR-10, we have simulated PCA's dimensionality reduction for image classification. The datasets have been shifted to row mean 0 and normalized so that $\sigma_{\max} = 1$. We have simulated Algorithm 1 by sampling $1/\gamma^2 = 1000$ times from the state $\sum_i^r \lambda_i |\mathbf{u}\rangle_i |\mathbf{v}\rangle_i |\bar{\sigma}_i\rangle$ to search the first k principal components that account for a factor score ratios sum $p = 0.85$. The simulation occurs by sampling with replacement from the discrete probability distribution given by the λ_i . We then estimated the measured λ_i

¹ The code of the experiments is available at <https://github.com/ikigai/qadra>.

using the Wald estimator (see the proof of Theorem 8) and searched for the most important k .² In all cases, sampling the singular values has been enough to decide how many to keep. However, as p increases, the gap between the factor score ratios decreases and the quality of the estimation of k or θ decreases. As discussed in Section 3.1, it is possible to detect this problem using Theorem 9 and solve it with a binary search for θ (Theorem 10). We have tested the quality of the representation by observing the accuracy of 10-fold cross-validation k -nearest neighbors with $k = 7$ as we introduce error in the representation's Frobenius norm (see Fig. 1). To introduce the error, we have added truncated Gaussian noise to each element of $\mathbf{U}\boldsymbol{\Sigma}$ to have $\|\mathbf{U}\boldsymbol{\Sigma} - \overline{\mathbf{U}\boldsymbol{\Sigma}}\| \leq \xi = \sqrt{k}(\epsilon + \delta)$ (Lemma 14). The parameter δ has been estimated using the bound above, choosing the error so that the accuracy drops no more than 0.01 and fixing ϵ to a number that allows for correct thresholding. Table 1 summarizes the run-time parameters. The results show that Theorems 8, 9, 10 are already advantageous on small datasets, while Theorem 12 requires bigger datasets to express its speed-up. We have also simulated the creation of the state at step 6 of Algorithm 4 to test the average number of measurements needed to collect all the singular values as ϵ increases. The analysis has confirmed the run-time's expectations. To end with, we have tested the value of α (Definition 2, Claim 16) for the MNIST dataset, fixing $\epsilon = 0$ and trying $p \in \{0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9\}$. We have observed that $\alpha = 0.97 \pm 0.03$, confirming that the run-time of Corollary 15 can be assumed $\tilde{O}(\mu(\mathbf{V})^{(k)})$ for the majority of the data points of a PCA-representable dataset.

We point out that more experiments on the run-time parameters have been extensively discussed in other works that rely on the same parameters (Kerenidis and Luongo 2020; Kerenidis et al. 2020b). These works study the scaling of the parameters as the dataset size increases, both in features and samples, and conclude that the parameters of interest are almost constant. In addition to the existing experiments, we have studied the trend of the run-time parameters on the Tiny Imagenet dataset as the number of samples scales. While the spectral norm increases, the other run-time parameters become constant after a certain number of samples. Figure 2 shows that the algorithms discussed in Section 3.1 are already of practical use for small datasets, while the singular vector extraction routines of Section 3.2 require larger datasets to be convenient over their classical counterparts. We refer the interested reader to the appendix for more details about the experiments.

² Note that, in practice, one could also estimate the factor score ratios as $\bar{\lambda}_i = \frac{\sigma_i}{\|\mathbf{A}\|_F}$. This method should require less measurements: a bound on the necessary number of measurements can be obtained via the coupon collector's problem with non-uniform probabilities.

6 Conclusions

In this paper, we formulate many eigenvalue problems in machine learning within a useful framework, filling the gap left open by previous literature with new algorithms. Our new procedures fill the gap by estimating the quality of a representation and extracting a classical description of the top- k singular values and vectors. We have shown how to use the new tools to extract the information needed by SVD-based data representation algorithms, computing theoretical error bounds for three machine learning applications. Besides identifying the proper quantum tools and formalizing the novel quantum algorithms, the main technical difficulty was analyzing how the error propagates to bound the algorithms' run-time properly.

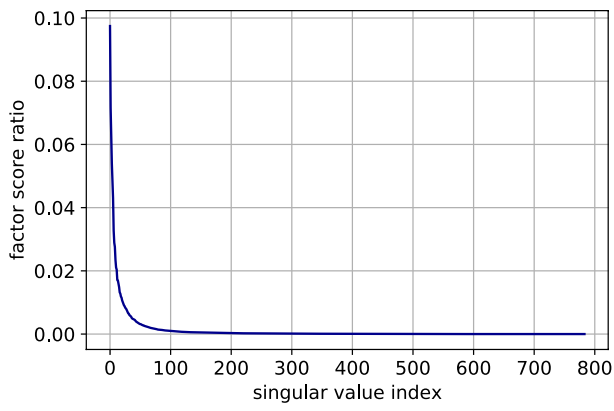
We do not expect run-time improvements that exceed poly-logarithmic factors or constant factors, using similar techniques. For non-zero singular values and dense singular vectors, the run-time of the extraction can not be smaller than kz , as one needs to read vectors of size kz . The δ^2 parameter is a tight bound for the ℓ_2 norm of the vectors, as it is a result of Chernoff's bound. The parameter ϵ is a tight error bound from phase estimation, which is necessary to distinguish the singular vectors. θ is the condition number of the low-rank approximation of the matrix, and it is necessary to amplify the amplitudes of the smallest singular values.

As future work, we deem it interesting to explore quantum algorithms for incremental SVD or for datasets whose points are available as a data streaming. It might be possible to reduce the overhead due to tomography and achieve greater speed-ups in these settings. It also remains an open question whether there are particular applications and dataset distributions for which the singular vector extraction algorithms offer a practical advantage over their classical counterparts. Finally, an appropriate resource estimation that takes into consideration different quantum hardware architectures, noise models, and error correction codes is out of the scope of this paper and is left for future work.

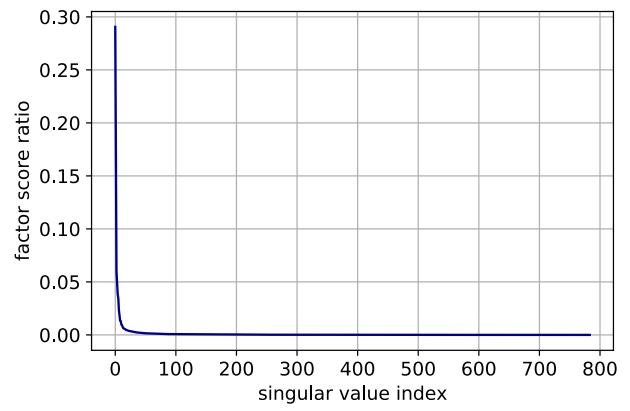
Appendix A: Experiments

A.1 Factor score ratios distribution in real data

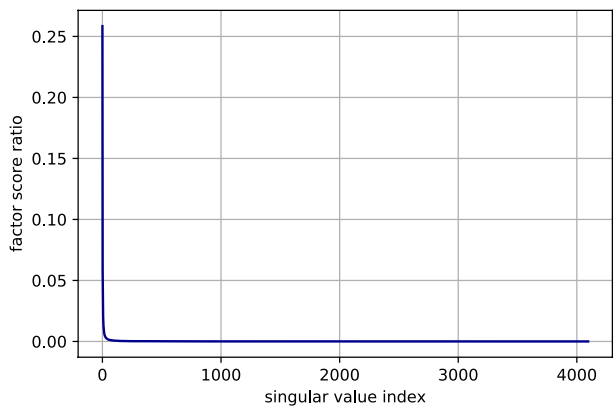
Throughout the work, we often claim that real datasets for machine learning are low-rank and that the distribution of their singular values is so that a few of them are much bigger than the rest. To verify this fact, we have selected five datasets for machine learning and investigated the distribution of the factor score ratios $\frac{\sigma_i^2}{\sum_j \sigma_j^2}$ in all of them (see Fig. 3). We briefly describe the datasets and our pre-processing steps.



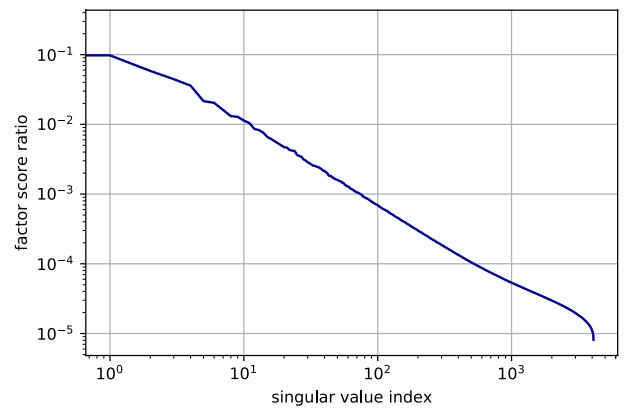
(a) MNIST.



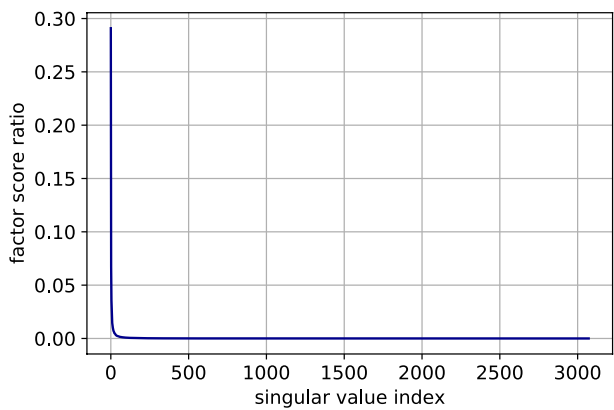
(b) Fashion MNIST.



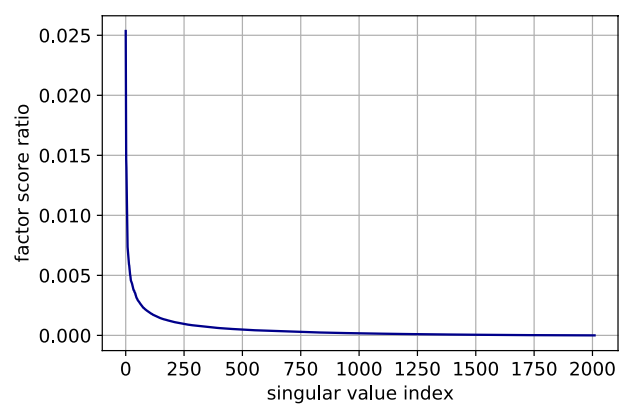
(c) Tiny Imagenet.



(d) Tiny Imagenet log-log scale.

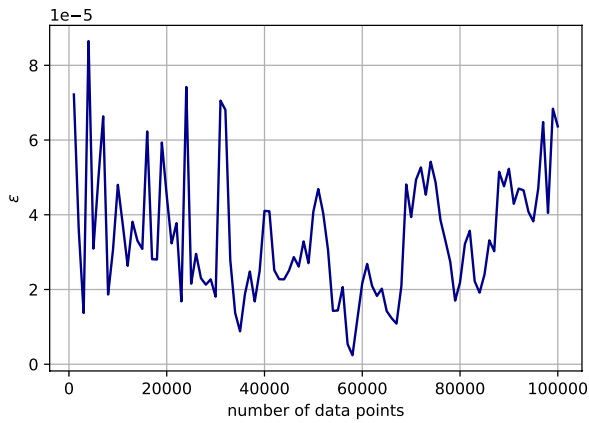
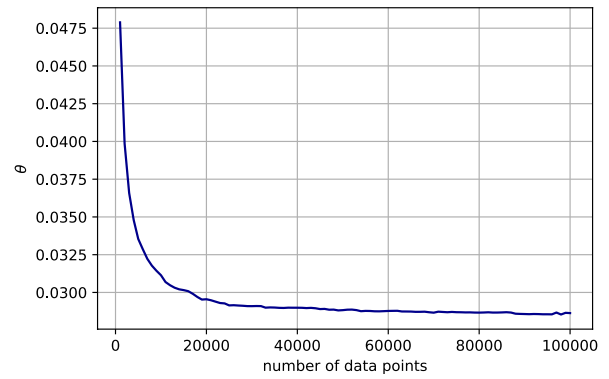
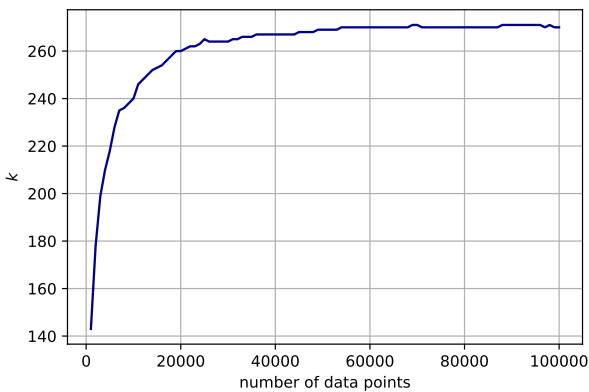
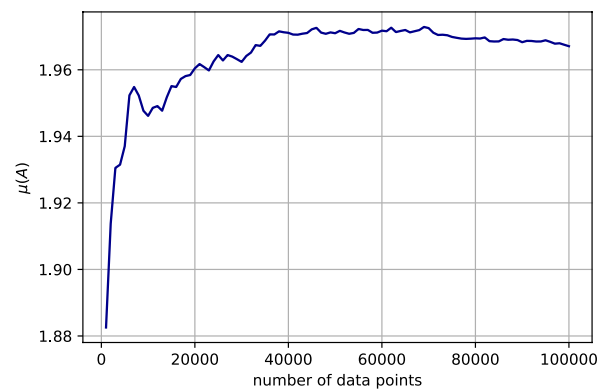
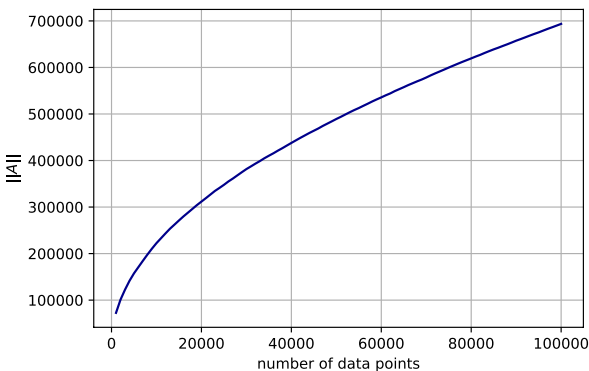


(e) CIFAR-10.



(f) Research Papers.

Fig. 3 Factor score ratios distributions in machine learning datasets

(a) Threshold ϵ .(b) θ .(c) Number of principal components k .(d) $\mu(\mathbf{A}) = \|\mathbf{A}\|_F$.

(e) Spectral norm.

Fig. 4 Tiny Imagenet run-time parameters

MNIST MNIST (LeCun et al. 1998) is probably the most used dataset in image classification. It is a collection of 70000 images of $28 \times 28 = 784$ pixels. Each image is a black and white hand-written digit between 0 and 9 and it is paired with

a label that specifies the digit. Since the images are black and white, they are represented as arrays of 784 values that encode the lightness of each pixel. The dataset, excluding the labels, can be encoded in a matrix of size 70000×784 .

Table 2 Results of the estimation of the number of principal components to retain. The parameter k is the number of components needed to retain at least $p = 0.85$ of the total variance. The parameter p is computed with respect to the estimated k

Parameter	MNIST	F-MNIST	CIFAR-10
Estimated k	62	45	55
Exact k	59	43	55
Estimated p	0.8510	0.8510	0.8510
Exact p	0.8580	0.8543	0.8514
γ	0.0316	0.0316	0.0316

Fashion MNIST Fashion MNIST (Xiao et al. 2017) is a recent dataset for benchmarking in image classification. Like the MNIST, it is a collection of 70000 images composed of $28 \times 28 = 784$ pixels. Each image represents a black and white fashion item among {T-shirt/top, Trouser, Pullover, Dress, Coat, Sandal, Shirt, Sneaker, Bag, Ankle boot}. Each image is paired with a label that specifies the item represented in the image. Since the images are black and white, they are represented as arrays of 784 values that encode the lightness of each pixel. The dataset, excluding the labels, can be encoded in a matrix of size 70000×784 .

CIFAR-10 CIFAR-10 (Krizhevsky and et al 2009) is another widely used dataset for benchmarking image classification. It contains 60000 colored images of 32×32 pixels, with the values for each of the 3 RGB colors. Each image represents an object among {airplane, automobile, bird, cat, deer, dog, frog, horse, ship, truck} and is paired with the appropriate label. We use all the images, reshaping them to unroll the three channels in a single vector. The resulting size of the dataset is 60000×3072 .

Tiny Imagenet Tiny Imagenet (Le and Yang 2015) is a subset of Imagenet, a large dataset for image classification. It is a collection of 100000 colored images of 64×64 pixels. Tiny Imagenet contains images of 200 object classes. Each class is composed of 500 images. We process the dataset to have only black and white images. Though the size is considerably less than the one of Imagenet, its complexity is higher than CIFAR-10's. The dataset, excluding the labels, can be encoded in a matrix of size 100000×4096 .

Research Paper Research Paper (Harun-Ur-Rashid 2018) is a dataset for text classification, available on Kaggle. It contains 2507 titles of papers together with the labels of the venue where they have been published. The labels are {WWW, INFOCOM, ISCAS, SIGGRAPH, VLDB}. We pre-process the titles to compute a contingency table of $papers \times words$: the value of the $i^{th} - j^{th}$ cell is the number of times that the j^{th} word is contained in the i^{th} title. We remove the English stop-words, the words that appear in

only one document, and those that appear in more than half the documents. The result is a contingency table of size 2507×2010 .

A.2 Run-time parameters

We have computed the run-time parameters on the Tiny Imagenet dataset, maintaining the number of features steady (i.e., 4096 black and white pixels) and observing how the parameters scale as we consider an increasing number of data points. The results are shown in Fig. 4. In these plots, epsilon is half the gap between the least singular value to retain and the one below, leading to correct thresholding, while theta is computed as the least singular value to retain. Although we would fine-tune θ and ϵ better in practice, the trend and the order of magnitudes of these parameters would remain like our plots. We have computed the best $\mu(\mathbf{A})$ over a finite set of $p \in [0, 1]$, and for any number of data points, the Frobenius norm was the most convenient. Finally, in this experiment, we did not estimate δ . This is because δ can only be estimated with respect to a specific classification task. We did not run classification on this dataset for practical computational reasons. However, the following sections contain more run-time parameters for image classification datasets on smaller datasets, including estimates for δ .

From the plots, we can see that the spectral norm increases with the number of data points and that the thresholding epsilon is independent of this quantity. All the other parameters asymptotically approach a constant after introducing a certain number of data points. Our intuition suggests that the number of data points after which the parameters are constant depends on the number of classes in the dataset. Indeed, this quantity should be related to the amount of information that a new data point adds to the dataset. The reader might find it weird that the Frobenius norm, in

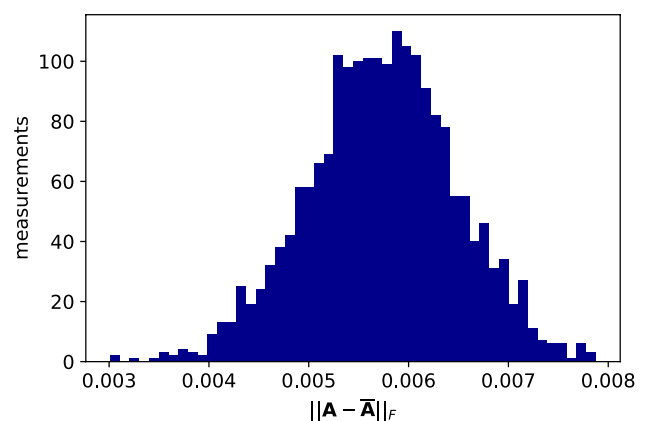


Fig. 5 Introducing some error in the Frobenius norm of a matrix \mathbf{A} . The error was introduced such that $\|\mathbf{A} - \bar{\mathbf{A}}\|_F \leq 0.01$. The figure shows the distribution of the error over 2000 measurements

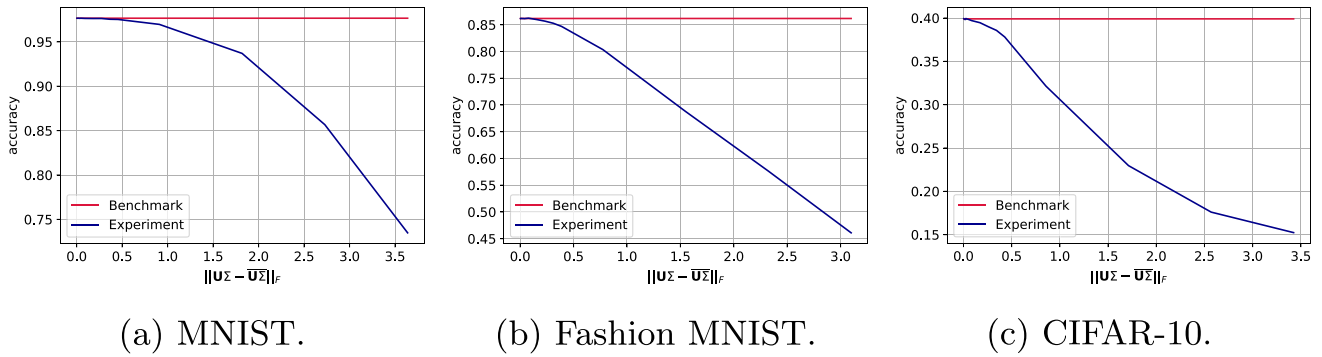


Fig. 6 Classification accuracy of 7-Nearest Neighbor on three machine learning datasets after PCA's dimensionality reduction. The drop in accuracy is plotted with respect to the *bound* on the Frobenius norm of the difference between the exact data representation and its approximation

Fig. 4, slightly decreases towards the end. However, this trend is justified by the fact that we compute these parameters after the dataset is divided by the spectral norm, and this parameter continues to increase (Fig. 4). The fact that $\mu(\mathbf{A})$ is a positive homogeneous function makes it so that scaling by the spectral norm does not improve the overall run-time. If we did not divide the dataset by the spectral norm, we would have seen the effect of its trend in ϵ , θ , and $\mu(\mathbf{A})$. The decrease of μ after the normalization corresponds to a decrease of ϵ and θ , making the overall run-time remain the same.

We have used this data to generate the run-time plots in the main text (Fig. 2). In that figure, we can see that the algorithms of Section 3.1 are already convenient on datasets of this size. In contrast, the ones for singular vector extraction of Section 3.2 require datasets of greater size to show their potential.

A.3 Image classification with quantum PCA

To provide the reader with a clearer view of our new algorithms and their use in machine learning, we provide experiments on quantum PCA for image classification. We perform PCA on the three datasets for image classification (MNIST,

Fashion MNIST, and CIFAR 10) and classify them with a K-Nearest Neighbors model. First, we simulate the extraction of the singular values and the percentage of variance explained by the principal components (top k factor score ratios' sum) using the procedure from Theorem 8. Then, we study the error of the model extraction, using Lemma 14, by introducing errors on the Frobenius norm of the representation to see how this affects the accuracy.

Estimating the number of principal components We shift MNIST, Fashion MNIST, and CIFAR-10 to row mean 0 and divide them by their spectral norm. We directly simulate Theorem 8 to decide the number of principal components needed to retain 0.85 of the total variance. For each dataset, we classically compute the singular values with an exact classical algorithm and simulate the quantum state $\frac{1}{\sqrt{\sum_i \sigma_i^2}} \sum_i \sigma_i |\sigma_i\rangle$ to emulate the measurement process of Algorithm 1. After initializing the random object with the correct probabilities, we measure it $\frac{1}{\gamma^2} = 1000$ times and estimate the factor score ratios with a frequentist approach (i.e., dividing the number of measurements of each outcome by the total number of measurements). Measuring 1000 times guarantees us an error of at most $\gamma = 0.03$ on each factor score ratio. In practice, the error is much smaller. To

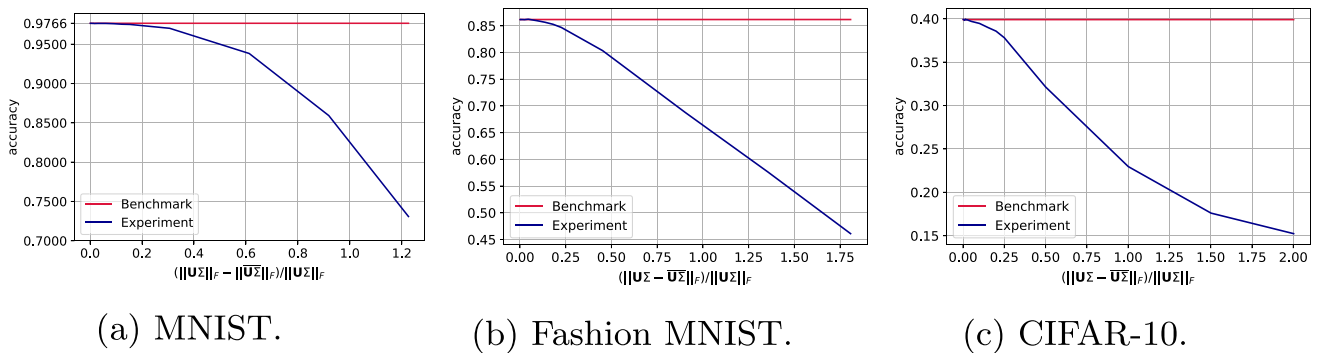


Fig. 7 Classification accuracy of 7-Nearest Neighbor on three machine learning datasets after PCA's dimensionality reduction. The drop in accuracy is plotted with respect to the *effective* Frobenius norm of the difference between the exact data representation and its approximation

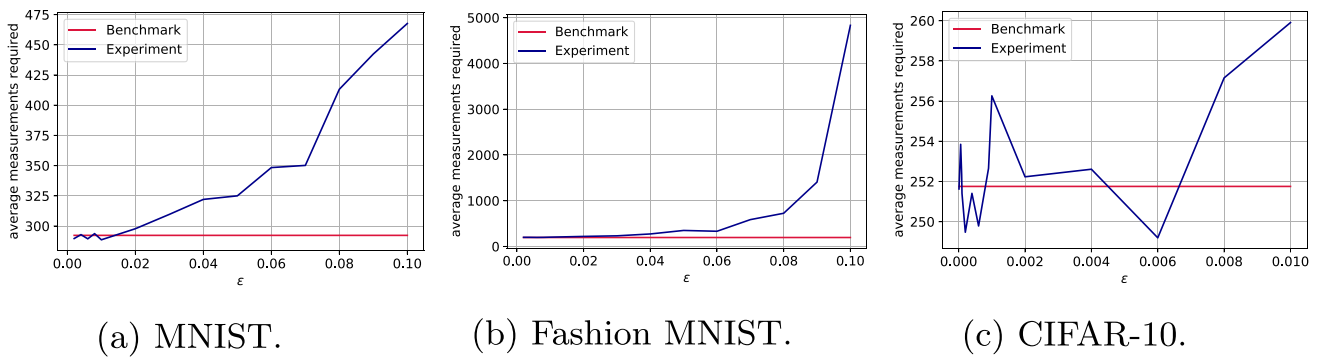


Fig. 8 Number of measurements needed to obtain all the k singular values from the quantum state $\frac{1}{\sqrt{\sum_i \sigma_i^2}} \sum_i \frac{\sigma_i}{\sigma_i} |\bar{\sigma}_i\rangle$, where $\|\sigma_i - \bar{\sigma}_i\| \leq \epsilon$, as ϵ increases. The benchmark line is $k \log_{2.4}(k)$

determine the number of principal components to retain, we sum the factor score ratios until the percentage of explained variance becomes more significant than 0.85. We report the results of these experiments in Table 2. We obtained good results for all the datasets, estimating no more than three extra principal components than needed.

We could further refine the number of principal components using Theorems 9, 10. When we increase the percentage of variance to retain, the factor score ratios become smaller and the estimation worsens. When the factor score ratios become too small to perform efficient sampling, it is possible to establish the threshold θ for the smaller singular value to retain using Theorems 9 and 10. Suppose one is interested in refining the exact number k of principal components, rather than θ . In that case, it is possible to obtain it using a combination of the Theorems 9, 10 and the quantum counting algorithm in time that scales with the square root of k (Theorem 11) to find a good trade-off. Once one sets the number of principal components, the next step is to use Theorem 12 to extract the top singular vectors. To do so, we can retrieve the threshold θ from the previous step by checking the gap between the last singular value to retain and the first to exclude.

Studying the error in the data representation We continue the experiment by checking how much error in the data representation a classifier can tolerate. We compute the exact PCA representation for the three datasets and the 10-fold Cross-validation error using k-Nearest Neighbors with 7 neighbors. For each dataset, we introduce errors in the representation and check how the accuracy decreases. To simulate the error, we perturb the exact representation by adding truncated Gaussian error (zero mean and unit variance, truncated on the interval $[-\frac{\epsilon}{\sqrt{nm}}, \frac{\epsilon}{\sqrt{nm}}]$) to each matrix entry. The graph in Fig. 5 shows the distribution of the simulated error on 2000 approximation of a matrix \mathbf{A} , such that

$\|\mathbf{A} - \bar{\mathbf{A}}\| \leq 0.1$. The distribution is still Gaussian, centered almost at half the bound.

The results show a reasonable tolerance of the errors; we report them in two sets of figures. Figure 6 shows the drop of accuracy in classification as the error bound increases. Figure 7 shows the accuracy trend against the approximation's

Analyzing the run-time parameters As discussed in Section 4, the model extraction's run-time is $\tilde{O}\left(\left(\frac{1}{\gamma^2} + \frac{kz}{\theta\sqrt{p}\delta^2}\right)\frac{\mu(\mathbf{A})}{\epsilon}\right)$, where $\mathbf{A} \in \mathbb{R}^{n \times m}$ is PCA's input matrix, $\mu(\mathbf{A})$ is a parameter bounded by $\min(\|\mathbf{A}\|_F, \|\mathbf{A}\|_\infty)$, k is the number of principal components retained, θ is the value of the last singular value retained, γ is the precision to estimate the factor score ratios, ϵ bounds the absolute error on the estimation of the singular values, δ bounds the ℓ_2 norm of the distance between the singular vectors and their approximation, and z is either n, m depending on whether we extract the left singular vectors, to compute the classical representation, or the right ones, to retrieve the model and allow for further quantum/classical computation. This run-time can be further lowered using Theorem 10 if we are not interested in the factor score ratios. This paragraph aims to show how to determine the run-time parameters for a specific dataset. We enrich the parameters of Table 2 with the ones in Table 3, and we discuss how to compute them. From the previous paragraphs, it should be clear how to determine k, θ, γ , and p , and it is worth noticing again that $1/\sqrt{p} \simeq 1$.

Table 3 Run-time parameters

Parameter	MNIST	F-MNIST	CIFAR-10
$\mu(\mathbf{A})$	3.2032	1.8551	1.8540
Thrs. ϵ	0.0030	0.0009	0.0006
θ	0.1564	0.0776	0.0746
δ	0.1124	0.0106	0.0340

We have computed $\mu(\mathbf{A})$ over a finite set of values $p \in [0, 1]$ and have seen that $\|\mathbf{A}\|_F$ is the best $\mu(\mathbf{A})$ (this is true for CIFAR-10, Fashion MNIST, Tiny Imagenet, and Research Papers as well). To compute the parameter ϵ one should consider the epsilon that allows for a correct singular value thresholding. We refer to this as the thresholding ϵ and set it as the difference between the last retained singular value and the first that is excluded. For the sake of completeness, we have run experiments to check how the Coupon Collector's problem changes as ϵ increases. Recall that in the proof of Theorem 12, we use $\frac{1}{\sqrt{\sum_i \frac{\sigma_i^2}{\sigma_i^2}}} \sum_i \frac{\sigma_i}{\sigma_i} |\mathbf{u}\rangle_i |\mathbf{v}\rangle_i |\bar{\sigma}_i\rangle \sim \frac{1}{\sqrt{k}} \sum_i |\mathbf{u}\rangle_i |\mathbf{v}\rangle_i |\bar{\sigma}_i\rangle$ to

say that the number of measurements needed to observe all the singular values is $O(k \log(k))$, and this is true only if ϵ is small enough to let the singular values distribute uniformly. We observe that the thresholding ϵ always satisfies the Coupon Collector's argument, and we have plotted the results of our tests in Fig. 8.

Furthermore, we have computed δ by using the fact that $\|\mathbf{A} - \bar{\mathbf{A}}\| \leq \sqrt{k}(\epsilon + \delta)$ (Lemma 14). We have computed an estimate for δ by inverting the equation and considering the thresholding ϵ . In particular, we have fixed $\|\mathbf{A} - \bar{\mathbf{A}}\|$ to the biggest value in our experiments so that the accuracy doesn't drop more than 1%.

These results show that Theorem 8, 9, and 10 can already provide speed-ups on datasets as small as the MNIST. Even though their speed-up is not exponential, they still run sub-linearly on the number matrix entries even though all the entries are taken into account during the computation, offering a polynomial speed-up with respect to their traditional classical counterparts. On the other hand, Theorem 12 requires bigger datasets. These algorithms are expected to show their full speed-up on big low-rank datasets that maintain a good distribution of singular values. As a final remark, the parameters have similar orders of magnitude.

Appendix B: Related works

One of the first papers that faced the problem of performing the eigendecomposition of a matrix with a quantum computer is the well-known Lloyd et al. (2014), which leveraged the intuition that density matrices are covariance matrices whose trace has been normalized. In this work, the authors assume to have quantum access to a matrix in the form of a density matrix and develop a method for fast density matrix exponentiation that enables preparing the eigendecomposition of the input matrix in time logarithmic on its dimensions. However, this algorithm requires the input matrix to be square, symmetric, and sparse or low-rank. More recently, the works of Kerenidis et al. on recommendation systems (Kerenidis and Prakash 2017) and least-squares (Kerenidis

and Prakash 2020a) have used a different definition of quantum access to a matrix (the one used throughout this work) and defined the task of singular value estimation. Their singular value decomposition scales better with respect to the error parameters, eliminates the dependency on the condition number, and does not have requirements on the input matrix. Several recent works, such as Lin et al. (2019); Reberstrost et al. (2018); Gu et al. (2019), have improved or extended the quantum singular value decomposition techniques. Almost none of them have provided a formal analysis of an algorithm that ensures classical access to the singular vectors, values and the amount of variance explained by each. There have also been attempts at creating near-term quantum algorithms for singular value decomposition. These works propose quantum circuits for singular value decomposition of quantum states on noisy intermediate-scale quantum (NISQ) devices using variational circuits (Bravo-Prieto et al. 2020; Wang et al. 2020c). However, the complexity of such methods is unclear, and recent works have questioned the efficacy of the speed-ups of variational quantum algorithms due to (entanglement and noise-induced) barren plateaus in the optimization landscape (Wang et al. 2020a; Marrero et al. 2020).

In classical computer science, most diffused implementations of PCA, CA, and LSA available (Pedregosa et al. 2011) relies on ARPACK (Lehoucq et al. 1998) or similar packages, which implement improvements of the Lanczos method, like the Implicitly restarted Arnoldi method (IRAM) (Sorensen 1997), an improvement upon the simple Arnoldi iteration, which dates back to 1951 (a more general case of Lanczos algorithm, which works only for Hermitian matrices). The run-time of these algorithms is bounded by $O(nmk \frac{\ln(m/\epsilon)}{\sqrt{\epsilon}})$, where ϵ is an approximation error related to the relative spectral gap between eigenvalues (Saad 1992).

The realization of quantum procedures that provide exponential speed-ups in linear algebra tasks has given inspiration for the realization of classical quantum-inspired algorithms that try to achieve the same run-time as their quantum counterparts. The process of transforming a quantum algorithm into a classical algorithm with a similar speed-up is usually referred to as "dequantization". In our case, the comparison with dequantized algorithms is often not easy, as they solve problems that are different from ours. Most of these works are based on a famous algorithm by Frieze, Kannan, and Vempala, which computes a low-rank approximation of a matrix in time that is sub-linear in the number of entries (see Frieze et al. 2004; Chia et al. 2020; Arrazola et al. 2020). Such algorithms promise exponential speed-ups over the traditional SVD algorithm for low-rank matrices. However, the high polynomial dependency of the run-times on the condition number, the rank, and the estimation error makes them advantageous only for matrices of extremely

large dimensions, with low ranks and small condition numbers. The research described in Arrazola et al. (2020) observed that the dependencies like $O\left(\frac{\|\mathbf{A}\|_F^6}{\epsilon^6}\right)$ are far from being tight in real implementations, but still order of magnitudes slower than the best classical algorithms.

Concomitantly to our work, a new important result (Chepurko et al. 2020) was able to lower the complexity of these dequantizations by better leveraging all the previous literature of classical algorithms in randomized linear algebra and re-framing them into a more complete mathematical framework. Indeed, previous sample-based dequantizations were just doing a form of leverage score sampling. These new algorithms seem to be tighter than previous results and offer a better comparison with quantum algorithms, solving problems related to ours. While we believe that it is not possible to have classical algorithms with run-times comparable to the ones of Theorems 8, 9, 10 (see the relationships between LLSD, SUES, and DQC1 in Cade and Montanaro (2018)) and Corollaries 15 and 17, we have found that the work of Chepurko et al. (2020) may question the practical advantage of our Theorem 12 over a classical counterpart. At first sight, their Theorem 33 might seem relevant for this work, as it provides a set of linearly independent rows of the input matrix. We stress that this problem is not related to finding the singular vectors provided by SVD, which are linearly independent and orthonormal. Moreover, even after further orthonormalization processing (e.g., Gram-Schmidt), the computed row basis wouldn't necessarily be the one provided by SVD. This is why we cannot compare the run-time of this procedure to our Theorem 12. On the other hand, Theorem 37 is more similar to our Theorem 12 but still aims to solve a different problem. While ours provides estimates $\|\mathbf{v}_i - \tilde{\mathbf{v}}_i\| \leq \epsilon, \forall i \in [k]$ (which we recall are also relative-error estimates, as $\|\mathbf{v}_i\| = 1$), their Theorem 37 provides a rank- k projector matrix $\mathbf{Q}^{(k)}$, with orthonormal columns, such that $\|\mathbf{A} - \mathbf{A}\mathbf{Q}^{(k)}\mathbf{Q}^{(k)T}\|_F^2 \leq (1 + \epsilon')\|\mathbf{A} - \mathbf{A}^{(k)}\|_F^2$ in time $\tilde{O}(n\text{nz}(\mathbf{A}) + \frac{k^{v-1}m}{\epsilon'} + \frac{k^{1.01}m}{\epsilon'^2})$. While it is easy to see that $\mathbf{Q} \rightarrow \mathbf{V}$ as $\epsilon \rightarrow 0$, it is not easy to see how $\|\mathbf{Q} - \mathbf{V}\|_F$ varies as ϵ varies and that becomes even less clear if we are interested in the error on a specific singular vector. If the run-time of this algorithm is shown to be better than its quantum equivalent, it would still be great to include it in our framework instead of Theorem 12 and continue to take advantage of the speed-ups of the other quantum procedures. One downside of using the dequantized subroutines would be that, in general, the $\tilde{O}(n\text{nz}(\mathbf{A}))$ data pre-processing step is different from the one required to provide efficient quantum access. Even though it can be possible that a classical algorithm could extract the singular vectors with a run-time comparable to the quantum one, using it would require paying additional costs both in time and space. Those costs arise from the need for an ad hoc data structure that would not

be adequate to provide competitive speed-ups with respect to the other available quantum machine learning and data analysis algorithms. We believe that both the classical and quantum versions of singular vectors extraction may be used in the future, depending on the computational capabilities available to the interested data analysts.

B.1 Principal component analysis

Probably no other algorithm in ML has been studied as much as PCA, so the literature around this algorithm is vast (Halko et al. 2011; Jolliffe and Cadima 2016). To mention an improvement upon the standard Lanczos method for PCA (Wang et al. 2020b), the authors used more Lanczos iterations to improve the numerical stability of PCA, by obtaining a better description of the Krylov subspace (i.e., more iterations help obtain a more orthonormal base). As mentioned, the problem of PCA has been studied previously within the model of quantum computation. Lin et al. (2019); He et al. (2020) focus on a circuit implementation of qPCA, whose run-time has been superseded by more recent techniques used in this paper. The work of Yu et al. (2019) faces the problem of performing PCA for dimensionality reduction on quantum states achieving an exponential advantage over the best known classical algorithms. However, their algorithm is somewhat impractical, due to the overall error dependence, which can be of $\tilde{O}(\epsilon^{-5})$. Furthermore they use old Hamiltonian simulation techniques, superseded by the techniques that we use in our paper. To our knowledge, there are no works that provide a theoretical analysis of the run-time for the procedure needed to *select* the number of singular vectors needed to retain enough variance, obtain a classical description of the model, and map new data points in the new feature space with theoretical guarantees on the run-time (which we believe cannot be improved, as in this work we show that the run-time for this mapping is almost constant).

B.2 Correspondence analysis

While correspondence analysis has been really popular in the past, so much that entire books have been written about it (Clausen 1998; Greenacre 2017), it seems to have become out of fashion in the last decades, probably overshadowed by the wave of results in deep learning. The novel formulation of Hsu et al. (2019) gives a new perspective of CA. The authors connects correspondence analysis to the principal inertia components theory, making it relevant also in tasks that concern privacy in machine learning (Wang et al. 2019). As said before, similarly and independently from us, Koide-Majima and Majima (2021) have extended the dequantized subroutines to perform canonical correspondence analysis. This algorithm is not expected to beat the performance of our quantum algorithm, let alone the performance of the best classical algorithm for CA.

B.3 Latent semantic indexing

LSA was first introduced in Deerwester et al. (1990), which spurred a flurry of applications (Landauer et al. 2013). Some notable works are streaming and/or distributed algorithms for incremental (LSA Řehřůřek 2011; Cavanagh et al. 2009; Zhang et al. 2017). While these work might offer inspiration for new quantum algorithms, their distributed nature make it an unfair comparison with a single-QPU quantum algorithm. LSA with neural networks has also been explored in the past years (Yu et al. 2008), albeit without guarantees on the runtime or the approximation error. During the preparation of this manuscript we discovered a previous work on quantum LSA, which pointed at the similarities between quantum states and LSA, albeit without offering any practical algorithm (González and Caicedo 2011).

Acknowledgements A.B. and S.Z. thank Prof. Ferruccio Resta and Prof. Donatella Sciuto for their support. A.L. has been supported by QuantERA ERA-NET Cofund in Quantum Technologies implemented within the European Union's Horizon 2020 Programme (QuantAlgo project), the ANRT, and Singapore's National Research Foundation, the Prime Minister's Office, Singapore, the Ministry of Education, Singapore under the Research Centres of Excellence program under research grant R 710-000-012-135.

Funding Open access funding provided by Politecnico di Milano within the CRUI-CARE Agreement.

Declarations

Conflict of interest The authors have no competing interests to declare.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Allcock J, Hsieh CY, Kerenidis I et al (2020) Quantum algorithms for feedforward neural networks. *ACM Transactions on Quantum Computing* 1(1):1–24
- Arrazola JM, Delgado A, Bardhan BR et al (2020) Quantum-inspired algorithms in practice. *Quantum* 4:307. <https://doi.org/10.22331/q-2020-08-13-307>
- Biamonte J, Wittek P, Pancotti N et al (2017) Quantum machine learning. *Nature* 549(7671):195–202. <https://doi.org/10.1038/nature23474>
- Brassard G, Hoyer P, Mosca M et al (2002) Quantum amplitude amplification and estimation. *Contemporary Mathematics* 305:53–74. <https://doi.org/10.1090/conm/305/052152>
- Bravo-Prieto C, García-Martín D, Latorre JI (2020) Quantum singular value decomposer. *Physical Review A* 101(6):062,310. <https://doi.org/10.1103/PhysRevA.101.062310>
- Cade C, Montanaro A (2018) The Quantum Complexity of Computing Schatten p -norms. In: 13th Conference on the Theory of Quantum Computation, Communication and Cryptography. 2018
- Cavanagh JM, Potok TE, Cui X (2009) Parallel latent semantic analysis using a graphics processing unit. In: Proceedings of the 11th Annual Conference Companion on Genetic and Evolutionary Computation Conference: Late Breaking Papers, pp 2505–2510 <https://doi.org/10.1145/1570256.1570352>
- Chakraborty S, Gilyén A, Jeffery S (2019) The Power of Block-Encoded Matrix Powers: Improved Regression Techniques via Faster Hamiltonian Simulation. In: 46th International Colloquium on Automata, Languages, and Programming (ICALP 2019), Leibniz International Proceedings in Informatics (LIPIcs), vol 132. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik, pp 33:1–33:14 <https://doi.org/10.4230/LIPIcs.ICALP.2019.33>
- Chepurko N, Clarkson KL, Horesh L, et al (2020) Quantum-inspired algorithms from randomized numerical linear algebra. arXiv preprint [arXiv:2011.04125](https://arxiv.org/abs/2011.04125)
- Chia NH, Gilyén A, Li T, et al (2020) Sampling-based sublinear low-rank matrix arithmetic framework for dequantizing quantum machine learning. In: Proceedings of the 52nd Annual ACM SIGACT Symposium on Theory of Computing, pp 387–400 <https://doi.org/10.1145/3357713.3384314>
- Clausen SE (1998) Applied correspondence analysis: An introduction, vol 121. Sage. <https://doi.org/10.4135/9781412983426>
- Deerwester S, Dumais ST, Furnas GW et al (1990) Indexing by latent semantic analysis. *Journal of the American society for information science* 41(6):391–407. [https://doi.org/10.1002/\(SICI\)1097-4571\(199009\)41:6<3c391::AID-ASII>3e3.0.CO;2-9](https://doi.org/10.1002/(SICI)1097-4571(199009)41:6<3c391::AID-ASII>3e3.0.CO;2-9)
- Erdős P, Rényi A (1961) On a classical problem of probability theory. *Magyar Tud Akad Mat Kutató Int Közl* 6(1–2):215–220
- Frieze A, Kannan R, Vempala S (2004) Fast monte-carlo algorithms for finding low-rank approximations. *Journal of the ACM (JACM)* 51(6):1025–1041. <https://doi.org/10.1145/1039488.1039494>
- Giovannetti V, Lloyd S, Maccone L (2008) Quantum random access memory. *Physical review letters* 100(16):160,501
- González FA, Caicedo JC (2011) Quantum latent semantic analysis. In: Amati G, Crestani F (eds) *Advances in Information Retrieval Theory*. Springer Berlin Heidelberg, Berlin, Heidelberg, pp 52–63 https://doi.org/10.1007/978-3-642-23318-0_7
- Greenacre M (2017) Correspondence analysis in practice. CRC Press. <https://doi.org/10.1201/9781315369983>
- Greenacre MJ (1984) Theory and applications of correspondence analysis. London (UK) Academic Press
- Gu L, Wang X, Zhang G (2019) Quantum higher order singular value decomposition. In: 2019 IEEE International Conference on Systems, Man and Cybernetics (SMC), IEEE, pp 1166–1171 <https://doi.org/10.1109/SMC.2019.8914525>
- Gyurik C, Cade C, Dunjko V (2020) Towards quantum advantage for topological data analysis. arXiv preprint [arXiv:2005.02607](https://arxiv.org/abs/2005.02607)
- Halko N, Martinsson PG, Shkolnisky Y et al (2011) An algorithm for the principal component analysis of large data sets. *SIAM Journal on Scientific computing* 33(5):2580–2594. [https://doi.org/10.1016/S0169-7439\(01\)00130-7](https://doi.org/10.1016/S0169-7439(01)00130-7)
- Hann CT, Lee G, Girvin S, et al (2021) Resilience of quantum random access memory to generic noise. *PRX Quantum* 2(2):020,311

- Harrow AW, Hassidim A, Lloyd S (2009) Quantum algorithm for linear systems of equations. *Physical review letters* 103(15):150,502. <https://doi.org/10.1103/PhysRevLett.103.150502>
- Harun-Ur-Rashid (2018) Research paper dataset. <https://www.kaggle.com/harunshimanto/research-paper>
- He C, Li J, Liu W (2020) An exact quantum principal component analysis algorithm based on quantum singular value threshold. arXiv preprint [arXiv:2010.00831](https://arxiv.org/abs/2010.00831)
- Hsu H, Salamati S, Calmon FP (2019) Correspondence analysis using neural networks. In: *The 22nd International Conference on Artificial Intelligence and Statistics*, pp 2671–2680
- Jolliffe IT, Cadima J (2016) Principal component analysis: a review and recent developments. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 374(2065):20150,202. <https://doi.org/10.1098/rsta.2015.0202>
- Kerenidis I, Luongo A (2020) Classification of the mnist data set with quantum slow feature analysis. *Physical Review A* 101(6):062,327. <https://doi.org/10.1103/PhysRevA.101.062327>
- Kerenidis I, Prakash A (2017) Quantum recommendation systems. In: *8th Innovations in Theoretical Computer Science Conference (ITCS 2017)*, Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik <https://doi.org/10.4230/LIPIcs.ITCS.2017.49>
- Kerenidis I, Prakash A (2020a) Quantum gradient descent for linear systems and least squares. *Physical Review A* 101(2):022,316. <https://doi.org/10.1103/PhysRevA.101.022316>
- Kerenidis I, Prakash A (2020) A quantum interior point method for lps and sdps. *ACM Transactions on Quantum Computing* 1(1):1–32. <https://doi.org/10.1145/3406306>
- Kerenidis I, Landman J, Luongo A, et al (2019a) q-means: A quantum algorithm for unsupervised machine learning. In: *Advances in neural information processing systems*, pp 4134–4144
- Kerenidis I, Landman J, Prakash A (2019b) Quantum algorithms for deep convolutional neural networks. In: *International conference on learning representations*
- Kerenidis I, Prakash A, Szilágyi D (2019c) Quantum algorithms for portfolio optimization. In: *Proceedings of the 1st ACM conference on advances in financial technologies*, pp 147–155 <https://doi.org/10.1145/3318041.3355465>
- Kerenidis I, Luongo A, Prakash A (2020a) Quantum expectation-maximization for gaussian mixture models. In: *International conference on machine learning*, PMLR, pp 5187–5197
- Kerenidis I, Luongo A, Prakash A (2020b) Quantum expectation-maximization for gaussian mixture models. In: *International conference on machine learning*, PMLR, pp 5187–5197
- Koide-Majima N, Majima K (2021) Quantum-inspired canonical correlation analysis for exponentially large dimensional data. *Neural Networks* 135:55–67. <https://doi.org/10.1016/j.neunet.2020.11.019>
- Krizhevsky A, et al (2009) Learning multiple layers of features from tiny images
- Landauer TK, McNamara DS, Dennis S et al (2013) *Handbook of latent semantic analysis*. Psychology Press. <https://doi.org/10.4324/9780203936399>
- Le Y, Yang X (2015) Tiny imagenet visual recognition challenge. *CS 231N* 7(7):3
- LeCun Y, Bottou L, Bengio Y et al (1998) Gradient-based learning applied to document recognition. *Proceedings of the IEEE* 86(11):2278–2324. <https://doi.org/10.1109/5.726791>
- Lehoucq RB, Sorensen DC, Yang C (1998) ARPACK users' guide: solution of large-scale eigenvalue problems with implicitly restarted Arnoldi methods. *SIAM* DOI 10(1137/1):9780898719628
- Lin J, Bao WS, Zhang S et al (2019) An improved quantum principal component analysis algorithm based on the quantum singular threshold method. *Physics Letters A* 383(24):2862–2868. <https://doi.org/10.1016/j.physleta.2019.06.026>
- Lloyd S, Mohseni M, Rebentrost P (2014) Quantum principal component analysis. *Nature Physics* 10(9):631–633. <https://doi.org/10.1038/nphys3029>
- Marrero CO, Kieferová M, Wiebe N (2020) Entanglement induced barren plateaus. arXiv preprint [arXiv:2010.15968](https://arxiv.org/abs/2010.15968)
- Partridge M, Calvo R (1997) Fast dimensionality reduction and simple pca. *Intelligent data analysis* 2(3):292–298. <https://doi.org/10.3233/IDA-1998-2304>
- Pedregosa F, Varoquaux G, Gramfort A et al (2011) Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research* 12:2825–2830
- Rebentrost P, Mohseni M, Lloyd S (2014a) Quantum support vector machine for big data classification. *Physical review letters* 113(13):130,503. <https://doi.org/10.1103/PhysRevLett.113.130503>
- Rebentrost P, Mohseni M, Lloyd S (2014b) Quantum support vector machine for big data classification. *Physical review letters* 113(13):130,503
- Rebentrost P, Steffens A, Marvian I, et al (2018) Quantum singular-value decomposition of nonsparse low-rank matrices. *Physical review A* 97(1):012,327. <https://doi.org/10.1103/PhysRevA.97.012327>
- Saad Y (1992) *Numerical methods for large eigenvalue problems*. Manchester University Press DOI 10(1137/1):9781611970739
- Schuld M, Petruccione F (2018) *Supervised Learning with Quantum Computers*. Springer. <https://doi.org/10.1007/978-3-319-96424-9>
- Sorensen DC (1997) Implicitly restarted arnoldi/lanczos methods for large scale eigenvalue calculations. In: *Parallel Numerical Algorithms*. Springer, p 119–165 https://doi.org/10.1007/978-94-011-5412-3_5
- Ta-Shma A (2013) Inverting well conditioned matrices in quantum log-space. In: *Proceedings of the forty-fifth annual ACM symposium on Theory of computing*, pp 881–890
- Wang G (2017) Quantum algorithm for linear regression. *Physical review A* 96(1):012,335
- Wang H, Vo L, Calmon FP et al (2019) Privacy with estimation guarantees. *IEEE Transactions on Information Theory* 65(12):8025–8042. <https://doi.org/10.1109/TIT.2019.2934414>
- Wang S, Fontana E, Cerezo M, et al (2020a) Noise-induced barren plateaus in variational quantum algorithms. *Bulletin of the American Physical Society*
- Wang X, Chen B, Sheng J, et al (2020b) An improved lanczos algorithm for principal component analysis. In: *Proceedings of 2020 the 6th International Conference on Computing and Data Engineering*, pp 70–74 <https://doi.org/10.1145/3379247.3379250>
- Wang X, Song Z, Wang Y (2020c) Variational quantum singular value decomposition. arXiv pp arXiv–2006
- Xiao H, Rasul K, Vollgraf R (2017) Fashion-mnist: a novel image dataset for benchmarking machine learning algorithms. arXiv preprint [csLG/170807747](https://arxiv.org/abs/1708.07747)
- Yu B, Zb Xu, Li Ch (2008) Latent semantic analysis for text categorization using neural network. *Knowledge-Based Systems* 21(8):900–904. <https://doi.org/10.1016/j.knosys.2008.03.045>
- Yu CH, Gao F, Lin S et al (2019) Quantum data compression by principal component analysis. *Quantum Information Processing* 18(8):249. <https://doi.org/10.1007/s11128-019-2364-9>
- Zhang M, Li P, Wang W (2017) An index-based algorithm for fast on-line query processing of latent semantic analysis. *PLoS One* 12(5):e0177,523. <https://doi.org/10.1371/journal.pone.0177523>
- Řehřůfek, R (2011) Subspace tracking for latent semantic analysis. In: *European Conference on Information Retrieval*, Springer, pp 289–300 https://doi.org/10.1007/978-3-642-20161-5_29