Quantum Independent Component Analysis for Signal and Image Feature Extraction

Saleh Naghdi
Supervisors: Prof. Michelle Simmons, Dr. Casey Myers, Dr. Sam Gorman
z5403131

1 Introduction to Machine Learning for Feature Extraction

It was none other than a machine learning (ML) algorithm that, in 2020, finally cracked the famous problem of predicting the 3D structure of proteins [15]. Unsurprisingly, in the era of big data, machine learning has become of great utility for being able to leverage data to build statistical models that make accurate predictions [6]. Paralleling this growing investment in machine learning is the nascent field of quantum computing which can promise boosts in efficiency and performance of algorithms that are fundamentally inaccessible by classical computing [22]. The possibility for this *quantum advantage* has ushered in a great incentive to develop quantum analogues for otherwise classical machine learning algorithms, the process of which may colloquially be referred to as quantisation [27]. Efforts to quantize classical algorithms can be easily seen from the wide gamut of techniques that have already been quantized, spanning from generative learning with quantum general adversarial networks (qGAN) and quantum Boltzmann machines (QBM) [30, 31], to discriminative learning such as quantum K-nearest neighbours (qkNN) and quantum support vector machines (qSVM)[25, 8].

One crucial preliminary procedure in many ML algorithms is feature extraction, which consists of manipulating the raw attributes that describe a data point (e.g height, weight) to build more informative, derived quantities called features (e.g. Body Mass Index). While these features may be engineered according to domain expertise, there exists a suite of generic feature extraction techniques which extract from a raw dataset features possessing some useful generic property. For example, principle component analysis (PCA) [1] derives generic features which have maximal statistical variance; and in slow feature analysis (SFA) [28], features are chosen which have minimal concavity over time – a property known as "slowness". However, substantially outperforming [4] both PCA and SFA which consider only up to the second-order statistics of a dataset is the higher-order technique of independent component analysis (ICA) [13], which instead obtains features that are *statistically independent* from one another.

Since its formulation by Hyvarinen [13], ICA has become the epitomal tool for solving the ubiquitous cocktail party problem which describes the task of sifting out the independent conversations being had at a party provided only the noisy chatter of the room as input data. More formally, this refers to the procedure of separating a mixture of signals into its independent source components when provided only instances of the mixtures (Fig. 1). By the same token, ICA can identify noise signals and filter them out, which itself has significant use in, for example, artefact removal in brain EEG recordings [26], analysis of financial time-series data [23, 3], and image processing [4]. Its many recent extensions and refinements [14] are also a testament to the fact that any improvements made to either the discriminative power or efficiency of ICA are to be of tremendous relevance and utility to data scientists and specialists alike.

There is, however, a problem: the complexity of the best classical ICA trainer is quadratic in the dimension of the dataset [19]. Unlike qPCA and qSFA which boast a theoretical polylogarithmic speedup [1] [21, 16], due to its classical complexity ICA – which has no quantum counterpart yet – still suffers an inability to accommodate large and highly dimensional data sets as demanded by use cases such as image processing and signal extraction [19]. Such a powerful algorithm thus stands to benefit greatly from being quantized. Not only is there the possibility of attaining a quantum

[1] Classical PCA and SFA have polynomial complexity, whereas the quantized forms are polylogarithmic in the dataset size and dimension: $O(\mathrm{poly}(\log(n, d)))$.

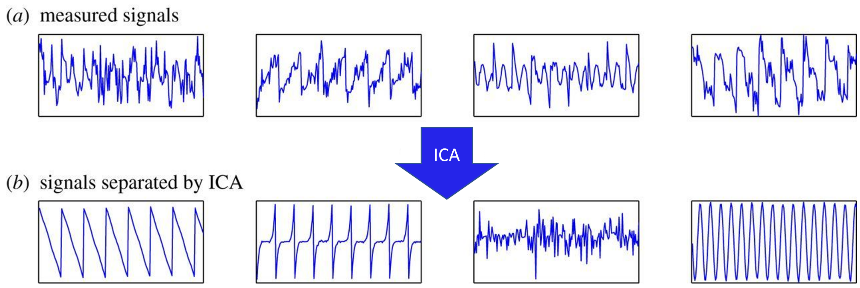

Figure 1: **Illustration of ICA** The ICA algorithm separating four mixture signals into the four statistically independent source components of which they are comprised.

speedup, but also that of a boosted accuracy [2] since the transition to the quantum regime affords the utilisation of phenomena like quantum entanglement and quantum superposition.

No one has quantized ICA before because the associated optimisation problem for finding statistically independent components is more general than those of PCA and SFA which instead reduce to a familiar eigenvalue problem. This has made it difficult to adapt ICA to be quantized by *block-encoding* methods employed by qSFA and qPCA [7]. However, these methods have attracted scrutiny for their reliance on the existence of a quantum RAM (qRAM) which is as yet only speculative [10, 27]. The challenge of quantizing ICA is thus twofold: finding a method that is not reliant on qRAM, or even if qRAM is assumed, discovering how to translate the problem of finding statistically independent components into a workable form able to be solved by the block-encoding methods used for qSFA and qPCA. This project aims to quantize ICA by devising approaches to solving each of those two goals independently: a qRAM-based and a kernel-based routine.

2 Theory of Classical Independent Component Analysis (ICA)

2.1 High-level Description

As seen in Figure 1, ICA recovers source signals (e.g. individual voices) from samples of combinations of them (e.g. inaudible chatter) [13]. In real life, source signals often follow a non-Gaussian distribution. ICA begins with a random linear combination of the mixtures, and iteratively changes the coefficients to make the term progressively more non-Gaussian. The point at which it's maximally non-Gaussian is when it is then assumed to be one of the source signals; this procedure is repeated for all remaining signals. In order to accelerate this iterative process, a preliminary statistical procedure called whitening is also performed on the samples of mixtures which involves decorrelating the mixtures from one another and standardising their means and variances.

2.2 Mathematical details of ICA [13]

Consider a room in which there are two speakers playing *independent* audio signals $s_1(t)$ and $s_2(t)$. If there are two microphones located at different points in the room, then each microphone respectively records some specific combination of the source signals $x_1(t)$ and $x_2(t)$. For this small example, the goal of ICA can then be proposed as such: given the two mixture signals $\mathbf{x}(t) = (x_1(t), x_2(t))$, obtain the two statistically independent source signals $\mathbf{s}(t) = (s_1(t), s_2(t))$. In general, out of

As a first assumption, these mixtures can be considered to be linear combinations of the source signals:

$$\begin{cases} x_1(t) = a_{11}s_1(t) + a_{12}s_2(t) \\ x_2(t) = a_{21}s_1(t) + a_{22}s_2(t) \end{cases}, \tag{1}$$

which in matrix form simplifies to

$$\mathbf{x} = \mathbf{A}\mathbf{s}\,, \tag{2}$$

where $[\mathbf{A}]_{ij} = a_{ij}$ and $\mathbf{A}$ is called the mixing matrix which changes from the basis of independent components to the mixed signal. Note we have suppressed the time dependence of the signals and henceforth will consider the signals to to be random variables[2]. This matrix form also captures the general case of N many source and mixed signals, where $\mathbf{x}, \mathbf{s} \in \mathbb{R}^n$. Solving the ICA problem now simplifies to obtaining an estimate $\tilde{\mathbf{W}} \approx \mathbf{A}^{-1}$ from which we can recover $\mathbf{s} = \mathbf{A}^{-1}\mathbf{x} \approx \tilde{\mathbf{W}}\mathbf{x}$, or equivalently, a single source component $s_i \approx \tilde{\mathbf{w}}_i^T\mathbf{x}$. What makes this problem uniquely ICA instead of PCA or SFA is the fact that the source signals s_i are to be statistically independent from one another.[3]

In the search for $\tilde{\mathbf{W}}$, we consider some trial matrix $\mathbf{W}$. Through the change of variables $\mathbf{z_i} = \mathbf{A}^T\mathbf{w_i}$, it follows that $y_i := \mathbf{w}_i^T\mathbf{x} = \mathbf{w}_i^T\mathbf{A}\mathbf{s} = \mathbf{z_i}^T\mathbf{s}$. Thus, y_i is a linear combination of the independent components $\mathbf{s}$ weighted by $\mathbf{z_i}$. By the central limit theorem, it is then the case that the random variable y_i – as a linear combination of independent components – is more Gaussian than any of the constituent components $\{s_i\}$. If the components $\{s_i\}$ are all non-Gaussian and identical random variables, then in fact, y_i would be the least Gaussian if and only if it was equal to only any one of the components s_i and not a linear combination of them. As a result, $y_i = \mathbf{w}_i^T\mathbf{x} = \tilde{\mathbf{w}}_i^T\mathbf{x} = s_i$ when y_i is maximally non-Gaussian. In relation to the analogy of the audio signals , this means that a linear combination y_i of the mixed signals will equate the source signal only when its distribution is the most non-Gaussian[4]. Additionally, given that statistical independence implies the signals are uncorrelated, the search space of $\mathbf{w}_i$ must be constrained to solutions that satisfy $\mathbb{E}[(\mathbf{w}_i^T\mathbf{x})(\mathbf{w}_j^T\mathbf{x})] = \mathbf{w}_i^T\mathbb{E}[\mathbf{x}\mathbf{x}^T]\mathbf{w}_i \overset{\text{must be}}{=} \delta_{ij}$. By letting $\mathbf{C} := \mathbb{E}[\mathbf{x}\mathbf{x}^T]$, the ICA algorithm can now be formulated as the following optimisation problem:

$$\begin{aligned} \underset{\mathbf{w}_i}{\text{maximize}} \quad & \text{non-Gaussanity of } \mathbf{w}_i^T\mathbf{x} \\ \text{subject to} \quad & \mathbf{w}_i^T\mathbf{C}\mathbf{w}_i = \delta_{ij} \end{aligned}$$

Classically, this optimisation is based on an iterative routine called fastICA that grows as $O(d(d+1)mn)$, where d is the dimension of each of the n many training instances, and m is the number of iterations required until convergence of the optimisation. This makes ICA quadratic in the dimensionality of the dataset [19].

In practice, additional steps are taken to reduce the dimension of the search space of $\mathbf{w}_i$ in order to accelerate convergence. These include centering and whitening (see below) which are critical data preprocessing steps (Fig. 1) conducted for many ML algorithms in order to reduce optimisation times.

2.2.1 Centering and Whitening for Preprocessing Data

Centering and whitening help to constrain the mixing matrix $\mathbf{W}$ to be orthogonal so that its rows $\mathbf{w}_i$ are parameterised by fewer parameters, thus reducing the search space of $\mathbf{w}_i$ traversed in the optimisation problem.

Centering is the procedure of subtracting from $\mathbf{x}$ its own mean so as to make it a zero-mean variable centered at the origin:

$$\mathbf{x} \leftarrow \mathbf{x} - \mathbb{E}[\mathbf{x}]\,. \tag{3}$$

Whereas centering standardises the mean, whitening does so for the (co)variance of $\mathbf{x}$ which in particular involves a) normalising the variance of each mixture component x_i to unity and b) decorrelating cross terms:

$$\mathbb{E}[x_i x_j] = \delta_{ij} \tag{4}$$

$$\implies \mathbb{E}[\mathbf{x}\mathbf{x}^T] = \mathbf{I}. \tag{5}$$

This is achieved through the following linear transformation:

$$\mathbf{x} \leftarrow \sqrt{\mathbb{E}[\mathbf{x}\mathbf{x}^T]}\,\mathbf{x} \tag{6}$$

$$\text{or} \quad \mathbf{A} \leftarrow \sqrt{\mathbb{E}[\mathbf{x}\mathbf{x}^T]}\,\mathbf{A}\,. \tag{7}$$

[2]In this interpretation, $\{\mathbf{x}(t) \mid \forall t\}$ may be considered a sample set of the random variable $\mathbf{x}$. This generalises the domain of ICA away from time-series data to more general datasets sampled from a distribution, such as an image dataset

[3]As motivated by the fact that source signals are generally produced from entirely independent mechanisms (e.g. two speakers playing different audio source files)

[4]This rests on the assumption that each of the source signals is non-Gaussian which is true for most real-word source signals. In the case that the dataset *is* multivariate Gaussian, because it is uniquely defined by its covariance matrix then any second-order statistic feature extraction model such as PCA will exhaustively analyse the dataset

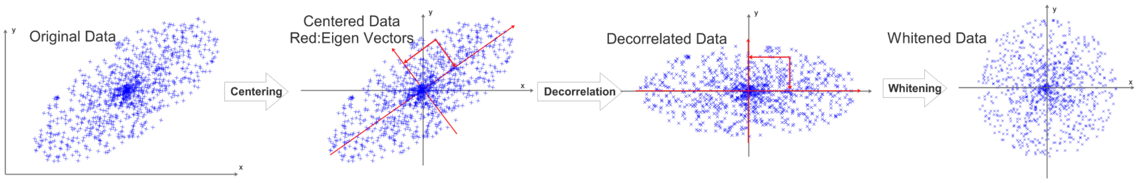

Figure 2: **Visualisation of whitening** the effect each operation involved in centering and whitening has on a given dataset. Observe that in the last two figures, the features are first decorrelated and only then standardised to unit variance. The circular shape of the final dataset suggests why whitening is also referred to as *sphering*.

For the purposes of ICA, centering and whitening have the benefit of making the mixing matrix $\mathbf{A}$ (and thus $\tilde{\mathbf{W}} \approx \mathbf{A}^{-1}$) orthogonal, since

$$\begin{aligned}\mathbb{E}[\mathbf{x}\mathbf{x}^T] &= \mathbb{E}[\mathbf{A}\mathbf{s}\mathbf{s}^T\mathbf{A}^T] && (8)\\ &= \mathbf{A}\mathbb{E}[\mathbf{s}\mathbf{s}^T]\mathbf{A}^T && (9)\\ &= \mathbf{A}\mathbf{A}^T && (10)\\ &= \mathbf{I}\,. && (11)\end{aligned}$$

Now the search space for $\tilde{\mathbf{W}}$ is significantly reduced as it takes $n(n-1)/2$ parameters to uniquely identify an orthogonal matrix as opposed to n^2 for an arbitrary matrix [13].

3 Past Work on Quantum Feature Extraction

3.1 The assumption of qRAM

Algorithms like qSFA only claim an exponential speedup on classical datasets insofar as the input model of a quantum RAM (qRAM) exists [16]. qRAM is a hypothetical memory system analogous to classical RAM that can efficiently load classical data into a quantum superposition state [10]. A complication of formulating qICA so that it relies on qRAM is that its proposed polylogarithmic complexity remains contingent on the existence of this input model, which is as yet only speculative.

Moreover, Ewin Tang recently provided compelling evidence that by assuming a classical input model that was equivalent to qRAM, it may be possible to classically gain as much of a computational speedup for many of these algorithm [27]. As a result, Tang formulated a series of quantum-inspired classical algorithms which effectively *dequantize* the earlier works by Kerenedis who pioneered qSFA and quantum recommendation systems [27]. Given the currently dubious footing of the literature on qRAM, this project will consider both qRAM-based and qRAM-free approaches to quantizing ICA.

3.2 Works dependent on qRAM

SFA is a closely related feature extraction technique to ICA which has already been quantized [28]. Rather than statistical independence, SFA finds signal components $s_i = \tilde{\mathbf{w}}_i^T\mathbf{x}$ which are instead the most slowly varying over time:

$$\begin{aligned}&\underset{\mathbf{w}_i}{\text{maximize}} && \text{Slowness of } \mathbf{w}_i^T\mathbf{x}\\ &\text{subject to} && \mathbf{w}_i^T\mathbf{C}\mathbf{w}_i = \delta_{ij}\,.\end{aligned}$$

Slowness can be quantified as the time-average of the squared second derivative of the trial signal, which has the following convenient representation:

$$\begin{aligned}\mathbb{E}[\dot{y}_i^2] &= \mathbb{E}[(\mathbf{w}_i^T\dot{\mathbf{x}})(\mathbf{w}_i^T\dot{\mathbf{x}})] && (12)\\ &= \mathbf{w}_i^T\mathbb{E}[\dot{\mathbf{x}}\dot{\mathbf{x}}^T]\mathbf{w}_i && (13)\\ &= \mathbf{w}_i^T\mathbf{B}\mathbf{w}_i && (14)\\ &= \frac{\mathbf{w}_i^T\mathbf{B}\mathbf{w}_i}{\mathbf{w}_i^T\mathbf{C}\mathbf{w}_i} && (15)\end{aligned}$$

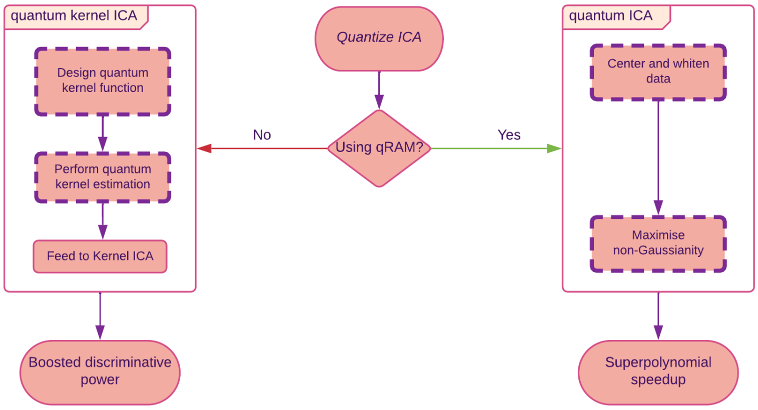

Figure 3: **Project outline flowchart** The two main approaches that will be considered for quantizing ICA are determining a quantum advantage for performance (left) and efficiency (right). The boxes with dashed borders depict steps that would require to be run on a quantum computer. The bottom rounded boxes are the expected outcomes.

where $\mathbf{B} := \mathbb{E}[\dot{\mathbf{x}}\dot{\mathbf{x}}^T]$, and the denominator in the last line is simply unity by virtue of the constraint of the signals being uncorrelated. The solution $\tilde{\mathbf{w}}_i$ for which Eq 15 is minimised (and slowness maximised) also solves the generalised eigenvalue problem

$$\mathbf{B}\mathbf{w}_i = \lambda_i \mathbf{C}\mathbf{w}_i \tag{16}$$

which after whitening the data set ($\mathbf{C} = \mathbf{I}$) reduces to the standard eigenvalue problem , effectively rendering the original problem Eq 15 to one that is based on tractable linear algebra techniques. For the whitening stage (Eq. (5)), quantum SFA [16] employs quantum methods for efficiently multiplying matrices and computing matrix powers – both of which benefit from running in polylogarithmic time. These methods belong to a more general suite of quantum linear algebraic algorithms which rely on representing arbitrary matrices in block-encoding unitary form by using a quantum accessible data structure and quantum RAM (qRAM) [7]. These methods can also be employed for quantizing ICA which considers whitening as a preliminary step.

However, it is the fact that the original optimisation problem reduces to solving the eigenvalue problem that truly inspires the quantization of SFA into qSFA. Based on an earlier work in building a quantum recommendation system, Kerenedis projects the whitened dataset onto the eigenspace spanned by $\{\tilde{\mathbf{w}}_i\}$ by using an efficient quantum singular value estimation algorithm, relying on rudimentary techniques such as quantum phase estimation [17].

Indeed, it has been shown that a special case of ICA called second-order ICA is equivalent to SFA, but the highly desirable search for a general quantum ICA still remains due to the fact that maximising non-Gaussianity in full generality is not as easy as the generalised eigenvalue problem in terms of being solved by block-encoding methods [5].

3.3 Works independent of qRAM

Recently, it was shown that as an alternative to qRAM, classical data points can instead have their attributes directly encoded as the parameters of a parameterised quantum circuit called an *ansatz* [12]. This is known as a *feature map* as it casts data points into a larger *feature space*. Moreover, using a procedure called *quantum kernel estimation* , quantum circuits are able to efficiently evaluate a subset of *kernel functions* (functions which calculate inner products of data points in a larger feature space) that are classically intractable to be computed, these are called *quantum kernel functions* [20].

Kernel ICA is a non-linear algorithm based on classical ICA which aims to unmix observation signals $\mathbf{x}_i$ that aren't simply linear combinations of the source signals $\mathbf{s}_i$ [2]. It takes as input a kernel function which introduces non-linearities in how the source and observations signals are related.

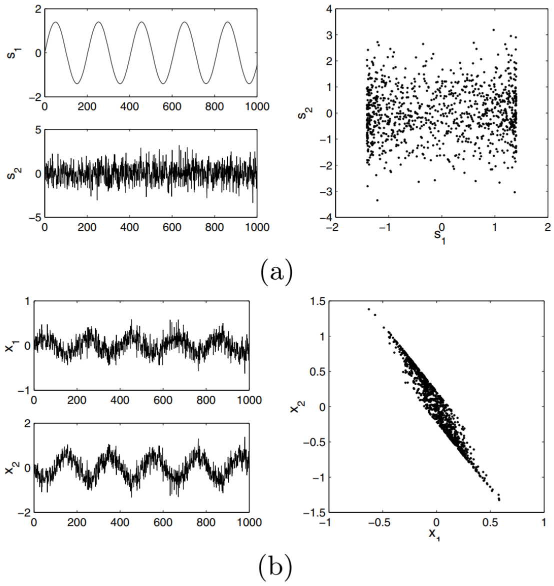

Figure 4: **Time-series dataset for qRAM-based benchmarking**: The two observed signals x_1 and x_2 in the dataset are randomly chosen linear combinations of a sinusoidal signal s_1 and Gaussian signals s_2.

4 Aims

The overall aims of this project can be summarised as the following:

1. Derive a qRAM-based quantum analogue to classical ICA and simulate this on an artificial signal dataset to determine a) how well the source signals are recovered and b) how the execution time scales with the dataset size, as compared to the best classical algorithm (fastICA). A quantum advantage obtained from qICA would allow the classical ICA algorithm to be executed superpolynomially faster[5], allowing for classification tasks to be done on highly dimensional image datasets which currently face classical limits [19].

2. Alternatively, Derive a qRAM-free quantum analogue to classical ICA by designing and evaluating a quantum kernel function tailored to the MNIST datatset of handwritten digits using *quantum kernel estimation*, which is then fed to kernel ICA. This is expected to give a boost in performance on non-linearly separable signals compared to the performance of a) non-kernel qICA and b) classical kernel ICA. This would demonstrate kernel methods being used for the first time on a practical (e.g. image) dataset as opposed to the contrived one in the source paper, thus highlighting the utility of kernel-based quantum machine learning.

3. Time permitting, optimally map qICA to the native gate set of the silicon devices available at SQC for implementation. This would contribute to demonstrating the first ever quantum algorithm on the silicon systems as a proof-of-concept. SQC is uniquely positioned to implement in the future on this hardware, given its atomically precise in-house manufacturing capability [18].

5 Methods for quantizing ICA

5.1 Two Datasets for Benchmarking

To determine if the qICA models can claim any boost in efficiency or accuracy, their performance will have to be benchmarked with reference to two datasets: MNIST Database of Handwritten Digits [9], and an artificially-constructed time-series signal dataset, samples from which are displayed in Figure 1.

[5]provided the existence of qRAM

5.1.1 Signal Time-Series [29]

This artificially constructed time series consists of two mixed signals obtained from two freqeuently occurring source signals: a sinusoidal waveform and a Gaussian white signal. The dimensionality of the dataset is increased by creating mixtures of more than just two source signals. The performance of the qRAM-based qICA will be based on how well the source signals are able to be recovered from the datasets, and how the execution time scales with dimensionality.

5.1.2 MNIST Dataset of Handwritten Digits [9]

This common database consists of 28x28 pixel images of handwritten digits in greyscale labeled with respect to their corresponding digit. It is comprised of 60,000 training and 10,000 test instances, and used to train classifiers to predict the number corresponding to a handwritten digit. Each image from this dataset can be vectorised into a 784-dimensional vector with pixel values as entries.

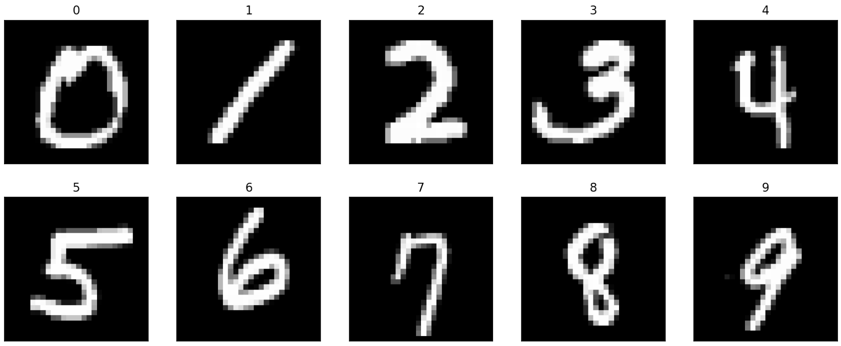

Figure 5: **A Sample of digits from the MNIST dataset of Handwritten Digits.**

5.2 Quantizing with qRAM

ICA consists of two steps: whitening and optimisation. Whitening can be quantized just as in the SFA algorithm by means of qRAM-based block-encoding methods of matrix multiplication [7]. For the second step, we will follow Reference [29] to reformulate the problem of maximising non-Gaussianity into one of carrying out non-linear PCA which involves the following steps:

1. Nonlinearly transform the data points $\mathbf{x}$ to $z = \|\mathbf{x}\|\mathbf{x}$ using qRAM-based quantum matrix multiplication [7].

2. Perform qPCA on the transformed dataset and the orthogonal matrix received as output U will be an approximation of the desired mixing matrix $\tilde{\mathbf{W}}$, thus solving the optimisation problem.

To evaluate whether this method correctly recovers the independent components, a simulation of the algorithm will be run on the mixed time-series dataset (see above). Then, the average align-

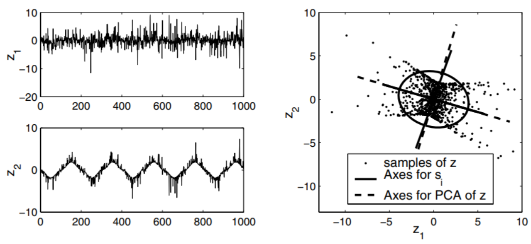

Figure 6: **Comparison of independent component axes obtained by qPCA and classical ICA**: (right) The qRAM-based method in Section 5.2 will be considered successful if the independent component axes obtained by peforming qPCA on the transformed mixed signals $z_i = \|\mathbf{x}_i\|\mathbf{x}_i$ aligns with the classical axes. The individual transformed mixed signals $\mathbf{z}_i$ are shown on the left.

ment[6] of the row (basis) vectors of **U** and those obtained by classical ICA will be calculated (as depicted in Figure 6) [29]. A high alignment will mean that our quantum algorithm is successfully obtaining the independent components which upon being graphed should correspond to sinusoidal and Gaussian waveforms. After validating the method above, the resultant polylogarithmic complexity of the algorithm will be checked by plotting execution time of finding **U** against increasing dimensionality (i.e. number of source signals) of the time-series dataset.

5.3 Quantizing without qRAM

Three steps will be taken to obtain a quantized kernel ICA:

1. Design a quantum advantageous kernel function[7] by using as reference the underlying group structure of the MNIST dataset, as was done in Reference [11] albeit for a contrived dataset.

2. Compute the kernel matrix efficiently according to the *quantum kernel estimation method* [12].

3. Feed the kernel matrix into the classical kernel ICA. The optimisation problem based on this method is actually based on the generalised eigenvalue problem [2], for which the methods introduced in Section 3 could be utilised.

A challenge associated with this is the task of computing a kernel that's useful for a practical dataset, since the only former demonstration of quantum supremacy by a kernel method was a proof-of-concept work on a contrived dataset. The main workaround to this will be identifying the symmetry properties of signal processing datasets to then apply the generic methods of covariant quantum kernels as mentioned in Reference [11].

The MNIST dataset will be used to evaluate this kernel-based approach by comparing the independent components generated by the quantum kernel function with those from a classical kernel function. In particular, the independent components derived from each approach will be fed into a K-nearest neighbours classifier [24] and the accuracy of correctly classified digits over increasing resolution of the images (and thus dimensionality) will be recorded. A similar evaluation procedure was employed when the qSFA was benchmarked, the findings of which can be seen in Figure 7 [16].

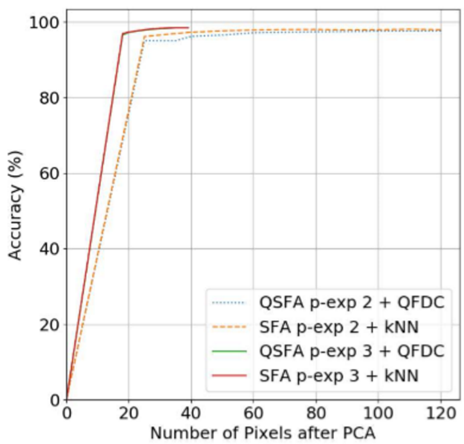

Figure 7: **Accuracy of qSFA vs SFA over increasing image resolution of the MNIST dataset** Note the two sets of lines refer to different kernels that were used as feature maps, where p-exp 2 a second degree polynomial expansion, and p-exp 3 a third degree one. In the qSFA paper, kNN was also quantized to yield Quantum Frobenius Distance Classifier (QFBC) which was used instead.

[6] measured by, for example, the cosine similarity.

[7] A quantum advantageous kernel function is a kernel function that is a) only efficiently computable on a quantum computer; and b) able to solve a classically intractable problem.

6 Summary of project outline

Be it for ECG data, facial recognition or image compression, ICA plays a critical role in signal analysis. However, throttled by its polynomial complexity, classical ICA cannot be liberally applied to highly dimensional image data sets, thus motivating this project's aim to formulate a quantum ICA to determine a potential superpolynomial speedup. Based on past works, this will likely necessitate the assumption of quantum RAM.

As an alternate, qRAM-free approach, we will also adopt quantum feature maps to quantize Kernel ICA which aims to make traditional ICA more discriminative of source signals by introducing non-linearities with the help of a kernel function. With recent findings that there exists an entire class of useful kernel functions exclusively implementable on quantum computers, this project will also aim to design one such quantum kernel function tailored to signal datasets which will be fed as input to kernel ICA. In this way, the switch to a quantum regime has the potential to afford not only a boost in speed, but also accuracy.

References

[1] Hervé Abdi and Lynne J Williams. Principal component analysis. *Wiley interdisciplinary reviews: computational statistics*, 2(4):433–459, 2010.

[2] Francis R Bach and Michael I Jordan. Kernel independent component analysis. *Journal of machine learning research*, 3(Jul):1–48, 2002.

[3] Andrew D Back and Andreas S Weigend. A first application of independent component analysis to extracting structure from stock returns. *International journal of neural systems*, 8(04):473–484, 1997.

[4] Marian Stewart Bartlett, Javier R Movellan, and Terrence J Sejnowski. Face recognition by independent component analysis. *IEEE Transactions on neural networks*, 13(6):1450–1464, 2002.

[5] Tobias Blaschke, Pietro Berkes, and Laurenz Wiskott. What is the relation between slow feature analysis and independent component analysis? *Neural computation*, 18(10):2495–2508, 2006.

[6] Alan Brnabic and Lisa M Hess. Systematic literature review of machine learning methods used in the analysis of real-world data for patient-provider decision making. *BMC medical informatics and decision making*, 21(1):1–19, 2021.

[7] Shantanav Chakraborty, András Gilyén, and Stacey Jeffery. The power of block-encoded matrix powers: improved regression techniques via faster hamiltonian simulaction. *arXiv preprint arXiv:1804.01973*, 2018.

[8] Yijie Dang, Nan Jiang, Hao Hu, Zhuoxiao Ji, and Wenyin Zhang. Image classification based on quantum k-nearest-neighbor algorithm. *Quantum Information Processing*, 17(9):1–18, 2018.

[9] Li Deng. The mnist database of handwritten digit images for machine learning research. *IEEE Signal Processing Magazine*, 29(6):141–142, 2012.

[10] Vittorio Giovannetti, Seth Lloyd, and Lorenzo Maccone. Quantum random access memory. *Physical review letters*, 100(16):160501, 2008.

[11] Jennifer R Glick, Tanvi P Gujarati, Antonio D Corcoles, Youngseok Kim, Abhinav Kandala, Jay M Gambetta, and Kristan Temme. Covariant quantum kernels for data with group structure. *arXiv preprint arXiv:2105.03406*, 2021.

[12] Vojtěch Havlíček, Antonio D Córcoles, Kristan Temme, Aram W Harrow, Abhinav Kandala, Jerry M Chow, and Jay M Gambetta. Supervised learning with quantum-enhanced feature spaces. *Nature*, 567(7747):209–212, 2019.

[13] Aapo Hyvärinen. Survey on independent component analysis. 1999.

[14] Aapo Hyvärinen. Independent component analysis: recent advances. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 371(1984):20110534, 2013.

[15] John Jumper, Richard Evans, Alexander Pritzel, Tim Green, Michael Figurnov, Olaf Ronneberger, Kathryn Tunyasuvunakool, Russ Bates, Augustin Žídek, Anna Potapenko, et al. Highly accurate protein structure prediction with alphafold. *Nature*, 596(7873):583–589, 2021.

[16] Iordanis Kerenidis and Alessandro Luongo. Quantum classification of the mnist dataset via slow feature analysis. *arXiv preprint arXiv:1805.08837*, 2018.

[17] Iordanis Kerenidis and Anupam Prakash. Quantum recommendation systems. *arXiv preprint arXiv:1603.08675*, 2016.

[18] Ludwik Kranz, Samuel Keith Gorman, Brandur Thorgrimsson, Yu He, Daniel Keith, Joris Gerhard Keizer, and Michelle Yvonne Simmons. Exploiting a single-crystal environment to minimize the charge noise on qubits in silicon. *Advanced Materials*, 32(40):2003361, 2020.

[19] Valero Laparra, Gustavo Camps-Valls, and Jesús Malo. Iterative gaussianization: from ica to random rotations. *IEEE transactions on neural networks*, 22(4):537–549, 2011.

[20] Yunchao Liu, Srinivasan Arunachalam, and Kristan Temme. A rigorous and robust quantum speed-up in supervised machine learning. *Nature Physics*, 17(9):1013–1017, 2021.

[21] Seth Lloyd, Masoud Mohseni, and Patrick Rebentrost. Quantum principal component analysis. *Nature Physics*, 10(9):631–633, 2014.

[22] Khadijeh Najafi, Susanne F Yelin, and Xun Gao. The development of quantum machine learning. 2022.

[23] Erkki Oja, Kimmo Kiviluoto, and Simona Malaroiu. Independent component analysis for financial time series. In *Proceedings of the IEEE 2000 Adaptive Systems for Signal Processing, Communications, and Control Symposium (Cat. No. 00EX373)*, pages 111–116. IEEE, 2000.

[24] Leif E Peterson. K-nearest neighbor. *Scholarpedia*, 4(2):1883, 2009.

[25] Patrick Rebentrost, Masoud Mohseni, and Seth Lloyd. Quantum support vector machine for big data classification. *Physical review letters*, 113(13):130503, 2014.

[26] Lisha Sun, Ying Liu, and Patch J Beadle. Independent component analysis of eeg signals. In *Proceedings of 2005 IEEE International Workshop on VLSI Design and Video Technology, 2005.*, pages 219–222. IEEE, 2005.

[27] Ewin Tang. A quantum-inspired classical algorithm for recommendation systems. In *Proceedings of the 51st Annual ACM SIGACT Symposium on Theory of Computing*, pages 217–228, 2019.

[28] Laurenz Wiskott and Terrence J Sejnowski. Slow feature analysis: Unsupervised learning of invariances. *Neural computation*, 14(4):715–770, 2002.

[29] Kun Zhang and Lai-Wan Chan. Ica by pca approach: relating higher-order statistics to second-order moments. In *International Conference on Independent Component Analysis and Signal Separation*, pages 311–318. Springer, 2006.

[30] Christa Zoufal, Aurélien Lucchi, and Stefan Woerner. Quantum generative adversarial networks for learning and loading random distributions. *npj Quantum Information*, 5(1):1–9, 2019.

[31] Christa Zoufal, Aurélien Lucchi, and Stefan Woerner. Variational quantum boltzmann machines. *Quantum Machine Intelligence*, 3(1):1–15, 2021.