

# Quantum principal component analysis without eigenvector recovery

Yewei Yuan, Michele Minervini, Mark M. Wilde, and Nana Liu

## Abstract

Principal component analysis (PCA) typically requires eigenvector extraction and hard rank- $k$  projection, which can be costly and unstable in high-dimensional or quantum settings. For score-based applications like anomaly detection, we introduce a measurement-based soft PCA framework that replaces the hard projector with an entropy-regularized Fermi–Dirac filter, converging to the classical projector in the zero-temperature limit. For generated quantum covariance operators, a single calibrated circuit accesses all spectral filters without eigenvector recovery or classical centering. The same circuit yields soft scores and spectral profiles for new inputs. This framework achieves a dimension-independent sample complexity  $O(\eta^{-2})$  for fractional-rank or retained-variance scoring at accuracy  $\eta$ . **See arXiv article attached.**

## 1 Introduction

PCA is one of the most standard ways to reduce the effective dimension in high-dimensional data. Traditional methods for PCA rely on rigid eigendecomposition. These suffer from several key problems: (i) keeping only the top  $k$  eigenvectors imposes a hard spectral cut-off, so eigenmodes are either fully kept or completely discarded. Thus small perturbations in the data can change the selected subspace when the relevant eigengap is small; (ii) PCA requires the covariance matrix and kernel PCA requires the centered Gram matrix and its eigendecomposition, which can be costly in big data settings; (iii) even once the top eigenvectors are recovered, assigning a score to a new data point requires expensive projection onto the top eigenvectors.

However, many practical applications primarily require a principal-subspace score—quantifying how much input data lies in dominant covariance directions. Bypassing explicit eigenvector recovery avoids traditional PCA bottlenecks while directly serving downstream tasks:

- **Anomaly detection:** Flagging abnormal inputs that exhibit low projection scores. [1–3]
- **Classification:** Classifying test inputs by their projection distances to learned class subspaces. [4–6]
- **Scientific quality control:** Summarizing dominant physical patterns to monitor real-world processes, such as material synthesis. [7, 8]
- **Model-order diagnostics:** Evaluating cumulative explained variance to determine optimal data compression rates. [9, 10].

This distinction is particularly important for quantum data. Quantum PCA can process spectral information efficiently [11], and in many settings the covariance operator is available through state preparation or sampling rather than as a classical matrix. Existing approaches nevertheless typically aim to recover eigenvectors or implement hard spectral projectors. They also face an additional centering issue: for quantum-native data or nonlinear quantum feature maps, there may be no accessible classical feature vector to subtract and no explicit centered Gram matrix to build. Our goal is therefore to formulate PCA as a calibrated measurement task that directly scores new quantum feature states, without eigenvector recovery and without assuming classical centering.

## 2 Measurement-based PCA and quantum covariance

Let  $C \in \text{Herm}(\mathbb{C}^d)$  be a covariance operator. Classical rank- $k$  PCA can be written as the trace-constrained measurement optimization

$$\max_{0 \preceq M \preceq \mathbb{1}} \text{Tr}(CM) \quad \text{subject to} \quad \text{Tr}(M) = k. \quad (1)$$

The feasible variable  $M$  is already a POVM effect. When the  $k$ th eigengap is nonzero, the optimizer is the usual projector  $P_k = \sum_{j \leq k} |u_j\rangle\langle u_j|$ , and the score of a normalized input state  $\rho$  is  $\text{Tr}(P_k \rho)$ .

We soften the spectral cut by adding Fermi–Dirac entropy to the same trace-constrained problem. The unique optimizer is

$$M_{T,k}^* = M_{T,\mu_k^*} = (\mathbb{1} + e^{(\mu_k^* \mathbb{1} - C)/T})^{-1}, \quad \text{Tr}(M_{T,k}^*) = k. \quad (2)$$

If  $C = \sum_j \lambda_j |u_j\rangle\langle u_j|$ , then  $M_{T,\mu} = \sum_j (1 + e^{(\mu - \lambda_j)/T})^{-1} |u_j\rangle\langle u_j|$ . Thus binary PCA occupations are replaced by smooth logistic weights. Finite temperature regularizes the cut-off, while  $M_{T,k}^* \rightarrow P_k$  as  $T \rightarrow 0$  whenever  $\lambda_k > \lambda_{k+1}$ . The target object remains a measurement effect, so it can be estimated by repeated measurement rather than by eigenvector recovery.

For quantum feature states  $|\phi_i\rangle = |\phi(\mathbf{x}_i)\rangle$ , define the representation of covariance operator

$$C_\phi := \frac{1}{N} \sum_{i=1}^N (|\phi_i\rangle - |m\rangle)(\langle \phi_i| - \langle m|) = \bar{\rho}_\phi - |m\rangle\langle m|, \quad |m\rangle = \frac{1}{N} \sum_i |\phi_i\rangle. \quad (3)$$

This has the same nonzero spectrum as the centered kernel-PCA matrix  $K_c/N$ , but is accessed directly as an operator on the feature Hilbert space.

The two terms in (3) are available coherently. A data-loading unitary prepares  $|\Omega\rangle = N^{-1/2} \sum_i |i\rangle |\phi_i\rangle$ , whose feature marginal is  $\bar{\rho}_\phi$ . Applying the inverse uniform-index preparation gives

$$(U_N^\dagger \otimes \mathbb{1})|\Omega\rangle = |0\rangle|m\rangle + |\perp\rangle, \quad (4)$$

where the index part of  $|\perp\rangle$  is orthogonal to  $|0\rangle$ . This branch supplies  $|m\rangle\langle m|$  directly. The same branch centers a test input by an LCU step: from  $|\phi(\mathbf{x})\rangle$  and (4), postselection prepares an unnormalized vector proportional to  $|z_{\mathbf{x}}\rangle = |\phi(\mathbf{x})\rangle - |m\rangle$ . Thus training and inference use the same coherent centering.

## 3 Training-free calibration

The soft filter can be implemented by a thermal measurement. Prepare a continuous-variable control state with logistic position density  $g_{T_1}(q)$ , apply the fixed data-side evolution

$$U_C = e^{-i\hat{p} \otimes C_\phi / T_2}, \quad T = T_1 T_2, \quad (5)$$

and measure the control position  $q$ . Accepting the outcome when  $q > \beta$  implements the effect

$$\widetilde{M}_{T,0,\beta} = (\mathbb{1} + e^{(\beta T_2 \mathbb{1} - C_\phi)/T})^{-1}. \quad (6)$$

The chemical potential is therefore  $\mu = \beta T_2$ , controlled by the classical threshold rather than by a rank-dependent circuit. The same  $U_C$  is used for every target rank and every inference query. In the access model induced by (3),  $U_C$  can be simulated by alternating short swap steps for  $\bar{\rho}_\phi$  with selector-swap steps for  $|m\rangle\langle m|$ .

Rank calibration uses the maximally mixed state  $\tau_{\text{mm}} = \mathbb{1}/d$ . Then

$$F_C(\beta_k^*) = 1 - \Pr(q \leq \beta \mid \tau_{\text{mm}}) = \text{Tr}(\widetilde{M}_{T,0,\beta} \tau_{\text{mm}}) = d^{-1} \text{Tr}(M_{T,\beta T_2}). \quad (7)$$

Thus the optimal trace- $k$  threshold is the empirical quantile satisfying  $F_C(\beta_k^*) = k/d$ , and  $M_{T,k}^* = \widetilde{M}_{T,0,\beta_k^*}$ . One sampled position distribution calibrates all ranks and requires no hybrid optimization.

Retained-variance calibration is analogous. If  $V_C = \text{Tr}(C_\phi) > 0$ , prepare the covariance probe  $\rho_C = C_\phi/V_C$  by uniformly sampling a training feature state, coherently centering it, and conditioning on success. With  $G_C(\beta) = \Pr(q > \beta \mid \rho_C)$ , a target retained covariance fraction  $\vartheta$  is selected by  $G_C(\beta_\vartheta^*) = \vartheta$ . No classical estimate of  $V_C$  is required.

## 4 Inference outputs and cost

After calibration, apply the same measurement to a centered test state  $|\tilde{z}_{\mathbf{x}}\rangle = |z_{\mathbf{x}}\rangle/\sqrt{\nu_{\mathbf{x}}}$ . For a rank or retained-variance label  $\ell \in \{K, \vartheta\}$ , with calibrated threshold  $\beta_\ell^*$ , define  $M_{T,\ell}^* = M_{T,\beta_\ell^* T_2}$ . The soft score and residual anomaly score are

$$\bar{s}_{T,\ell}(\mathbf{x}) = \text{Tr}(M_{T,\ell}^* |\tilde{z}_{\mathbf{x}}\rangle\langle\tilde{z}_{\mathbf{x}}|), \quad 1 - \bar{s}_{T,\ell}(\mathbf{x}). \quad (8)$$

Using the unpostselected LCU-centered input instead gives unnormalized energies directly from joint success-and-position probabilities.

Rank-indexed thresholds also define a soft spectral profile. Set  $M_{T,0}^* = 0$ ,  $M_{T,d}^* = \mathbb{1}$ , and  $D_{T,k} = M_{T,k}^* - M_{T,k-1}^*$ . Then  $\{D_{T,k}\}_{k=1}^d$  is a POVM; at low temperature it concentrates on the ordinary principal eigenspaces. The measured window probabilities  $\pi_{T,k}(\mathbf{x}) = \text{Tr}(D_{T,k} |\tilde{z}_{\mathbf{x}}\rangle\langle\tilde{z}_{\mathbf{x}}|)$  form a normalized soft spectral energy profile, and the cumulative rank- $K$  score is  $\sum_{k \leq K} \pi_{T,k}(\mathbf{x})$ .

The sampling cost depends on the requested output. A uniform all-rank trace guarantee carries a  $d^2$  factor through the empirical CDF. For a fixed rank the binomial bound is sharper. For normalized rank fractions or retained-variance profiles to additive accuracy  $\eta$ , the sampling contribution is  $O(\eta^{-2} \log(1/\delta_{\text{fail}}))$ , independent of  $d$ ; covariance-probe attempts acquire the centered-branch overhead. Each sample uses the same simulated evolution (5).

## 5 Advantages over Existing Approaches

Our PCA framework provides several distinct operational advantages for downstream scoring tasks:

- **Scoring Complexity:** Classical Krylov methods require  $\widetilde{O}(NdK)$  to extract eigenvectors plus  $O(dK)$  for each test score. QPE-based qPCA requires  $\widetilde{O}(R^2\kappa^2/\epsilon^3)$  copies and repeated phase estimation per query. Our protocol bypasses reconstruction, evaluating the soft score via a single calibrated measurement with dimension-independent sample complexity  $O(\eta^{-2})$  for fractional-rank or retained-variance tasks.
- **Multi-Rank Calibration:** QSVT-based PCA requires  $\widetilde{O}(K\kappa/\Delta_K)$  queries and must recompile the singular-value polynomial for each target rank  $K$ . Variational qPCA requires separate optimization loops per rank. Conversely, our protocol extracts all rank-dependent filters  $\{M_{T,k}^*\}_{k=1}^{d-1}$  from a single fixed evolution  $U_C$  and one classical CDF, requiring no circuit recompilation.
- **Operator-Level Kernel:** By representing the covariance matrix directly as a linear combination of quantum states, the protocol mitigates the  $\Theta(N^2)$  quantum measurement overhead required to construct a full empirical Gram matrix in standard quantum kernel workflows.
- **Coherent Centering:** The framework incorporates data centering into the quantum evolution via a subnormalized mean branch. This avoids explicit classical mean-vector subtraction or Gram matrix centering, facilitating PCA on quantum-native data and nonlinear feature maps where classical mean estimation is strictly prohibitive.

## References

- [1] J. E. Jackson and G. S. Mudholkar, “Control procedures for residuals associated with principal component analysis,” *Technometrics*, vol. 21, no. 3, pp. 341–349, 1979.
- [2] A. Lakhina, M. Crovella, and C. Diot, “Diagnosing network-wide traffic anomalies,” in *Proceedings of ACM SIGCOMM 2004*, pp. 219–230, 2004.
- [3] Q. Ding and E. D. Kolaczyk, “A compressed PCA subspace method for anomaly detection in high-dimensional data,” *IEEE Transactions on Information Theory*, vol. 59, no. 11, pp. 7419–7433, 2013.
- [4] S. Watanabe and N. Pakvasa, “Subspace method of pattern recognition,” in *Proceedings of the 1st International Joint Conference on Pattern Recognition*, pp. 25–32, 1973.
- [5] M. Turk and A. Pentland, “Eigenfaces for recognition,” *Journal of Cognitive Neuroscience*, vol. 3, no. 1, pp. 71–86, 1991.
- [6] K. Fukui, “Subspace methods,” in *Computer Vision: A Reference Guide* (K. Ikeuchi, ed.), Springer, 2020.
- [7] M. Kano, S. Hasebe, I. Hashimoto, and H. Ohno, “A new multivariate statistical process monitoring method using principal component analysis,” *Computers & Chemical Engineering*, vol. 25, no. 7–8, pp. 1103–1113, 2001.
- [8] E. Marin, D. Redolfi Bristol, A. Rondinella, A. Lanzutti, and P. Riello, “Statistical approaches to Raman imaging: Principal component score mapping,” *Analytical Methods*, vol. 16, pp. 2707–2720, 2024.
- [9] T. P. Minka, “Automatic choice of dimensionality for PCA,” in *Advances in Neural Information Processing Systems 13* (T. K. Leen, T. G. Dietterich, and V. Tresp, eds.), pp. 598–604, MIT Press, 2000.
- [10] R. Perry, S. Panigrahi, J. Bien, and D. Witten, “Inference on the proportion of variance explained in principal component analysis,” *Journal of the American Statistical Association*, 2025.
- [11] S. Lloyd, M. Mohseni, and P. Rebentrost, “Quantum principal component analysis,” *Nature Physics*, vol. 10, no. 9, pp. 631–633, 2014.