

Complexity of Normalized Persistence Problems for Topological Data Analysis and Local Hamiltonians

Dominic Lowe¹, M. S. Kim¹, Roberto Bondesan², and Ryu Hayakawa³

¹*Blackett Laboratory, Imperial College London, SW7 2AZ, United Kingdom*

²*Department of Computing, Imperial College London, SW7 2AZ, United Kingdom*

³*Yukawa Institute for Theoretical Physics & The Hakubi Center, Kyoto University, Japan*

Abstract

Topological data analysis (TDA) is a machine learning technique that uses topology to extract patterns from data and has shown the potential to exhibit quantum advantage. A key concept in TDA is *persistent homology*, which measures the robustness of topological information at different lengthscales. In this paper, we introduce and study the problem of *normalized persistence*, a practically motivated and easily interpretable version of persistent homology that counts the fraction of holes that persist at different lengthscales. We prove that a variant of *normalized persistence* is DQC_1 -hard and contained in BQP , giving evidence of an exponential quantum speedup for TDA under the standard assumption that $\text{DQC}_1 \not\subseteq \text{BPP}$. These are the first DQC_1 -hardness results that are directly applicable to TDA instances. We also find a close connection between normalized persistence and the complexity of estimating spectral quantities in the *low-energy subspace* of local Hamiltonians. We study a family of such problems, including a low-energy normalized subtrace and spectral density. We show that these are DQC_1 -hard for $O(1)$ -local Hamiltonians, strengthening previous results that required log-local interactions. We also introduce a variant of DQC_1 with perfect completeness (SDQC_1) to characterize the hardness of problems normalized by an exact kernel. This includes normalized persistence for $O(1)$ -local Hamiltonians, which we show is SDQC_1 -hard.

Extended Abstract

Topological data analysis (TDA) uses tools from algebraic topology to study properties of data. It is a mature field in machine learning [DW22] with several applications [SL22, SFH⁺23], and has gained attention as a task that may exhibit quantum advantage. TDA takes data and forms a simplicial complex X , a topological space consisting of higher-dimensional triangles (simplices). One can then use algebraic topology to calculate the number of d -dimensional holes in the complex, revealing information about the manifold the data was drawn from. The number of d -dimensional holes in X is called the d 'th Betti number and denoted by β_d^X . The original quantum TDA algorithm [LGZ16] outputs the *normalized Betti number* $\beta_d^X/|X_d|$, where $|X_d|$ is the number of d -simplices in the complex. However, in practical TDA applications one is also interested in how many holes persist from one complex to another, when both are formed from the same dataset. This motivates the *persistent Betti number* $\beta_d^{1,2}$, the number of d -dimensional holes in a complex X_1 that are also present in a larger complex X_2 . Indeed [Hay22] developed a quantum algorithm for the *normalized persistent Betti number* $\beta_d^{1,2}/|X_{1,d}|$. Determining the computational complexity of these quantities has remained an open problem since these algorithms were proposed, with DQC_1 -hardness of the normalized Betti number being a central unresolved question [GCD22].

In this paper we consider an alternative normalized quantity: the *normalized persistence* $\frac{\beta_d^{1,2}}{\beta_d^1}$, the fraction of d -dimensional holes present in X_1 that persist into X_2 . The difference in normalization is practically motivated, since the quantity gives an easily interpretable measure of how robust the topological features are in the initial complex. Indeed, one of the key reasons for studying persistence in TDA is that holes which persist are less likely to be the result of noise in the data and more likely to be true features of the underlying manifold. In addition, since the normalized

persistence is independent of the simplex count $|X_d|$ it allows for reasonable comparison between the complexes on different datasets. The main goal of this paper is to study the computational complexity of normalized persistence and a family of related quantities.

Previous work [GCD22] has formalized the problem of approximating $\beta_d/|X_d|$, as ‘Approximate Betti Number Estimation’. Whilst the complexity of this is still open, Gyurik et al. establish a connection with *Low-Lying Spectral Density* (LLSD). This problem estimates the number of eigenvalues of a Hamiltonian below a given threshold, normalized by the full Hilbert space dimension, and is only known to be DQC_1 -hard for log-local Hamiltonians. In [CC24] Cade and Crichigno established DQC_1 -hardness of *normalized quasi-Betti number estimation*. This is estimation of the dimension of the low-energy subspace of a combinatorial Laplacian divided by the dimension of the chain space. However, their result only holds for log-local co-boundary operators and does not hold for clique complexes, which are the complexes arising in TDA. In related work, deciding whether a given hole persists from one complex to another was shown to be BQP_1 -hard and contained in BQP [GSK⁺26]. Although this provides evidence of exponential quantum advantage, it does not capture the complexity of learning the persistence of all holes in the initial complex. Moreover, it is phrased as a guided problem, in which a guiding state is supplied as part of the input; our problems instead stay close to the original setting of Lloyd, Garnerone, and Zanardi [LGZ16], estimating normalized quantities.

Results for TDA. We formalize the task of estimating normalized persistence $\frac{\beta_d^{1,2}}{\beta_d^1}$ to inverse polynomial additive precision under inverse-polynomial spectral gap promise for combinatorial Laplacians as *Normalized Harmonic Persistence*. As an initial step, we introduce a low-energy relaxation, *Normalized Quasi-Harmonic Persistence*. Here we denote the clique complex of a graph G_k by $\text{Cl}(G_k)$ and the corresponding d ’th combinatorial Laplacian by $\Delta_{k,d}$; recall that $\beta_d = \dim \ker(\Delta_d)$. Then an informal definition of *Normalized Quasi-Harmonic Persistence* is: given input dimension d and two weighted graphs $G_1 \subseteq G_2$, output an estimate χ such that

$$\frac{\dim \text{Im } \mathcal{P}_{2,b} |_{\text{Im } \mathcal{P}_{1,\eta}}}{\tilde{\beta}_{d,\eta}^1} - \epsilon \leq \chi \leq \frac{\dim \text{Im } \mathcal{P}_{2,b+\xi} |_{\text{Im } \mathcal{P}_{1,\eta}}}{\tilde{\beta}_{d,\eta}^1} + \epsilon,$$

where $\mathcal{P}_{k,b}$ denotes the orthogonal projector onto the eigenvectors of $\Delta_{k,d}$ with energy at most b and $\tilde{\beta}_{d,\eta}^1 = \text{rank } \mathcal{P}_{1,\eta}$ is the d ’th quasi-Betti number of $\text{Cl}(G_1)$. We also informally describe the numerator of this quantity as $\tilde{\beta}_{d,\eta}^{1,2}$, the *quasi-persistent Betti number*, in Table 1. All threshold and error parameters above are taken to be $\Omega(\frac{1}{\text{poly}(n)})$ where n is the number of nodes on G_1 and G_2 . When b, η, ξ are sufficiently small, and we have suitable promises on the spectral gaps of $\Delta_{1,d}$ and $\Delta_{2,d}$ then the output χ will be within ϵ of the *normalized persistence*. In this paper we prove the following theorem.

Theorem (Main result for TDA). *Normalized Quasi-Harmonic Persistence is DQC_1 -hard and contained in BQP , under a suitable assumption. Moreover, the hardness remains with this assumption.*

In order to study the original problem (*Normalized Harmonic Persistence*), we introduce a new complexity class. This is motivated by the fact that previous hardness results in TDA have usually required complexity classes with perfect completeness, such as BQP_1 or QMA_1 . Thus, we introduce a variant of DQC_1 with perfect completeness, which we call *Subspace DQC_1* (SDQC_1). Informally, this class can be seen as DQC_1 with the additional constraint that there exists an unknown subspace of the maximally mixed register, such that any state in that subspace is accepted with probability 1. This complexity class captures the complexity of TDA and local Hamiltonian problems that are normalized by an exact kernel, such as β_d^1 . We prove SDQC_1 -hardness for several such problems,

Problem	Quantity	Hardness	Containment
<i>Prior works</i>			
Normalized Subtrace	$2^{-n} \sum_{0 \leq \lambda_i \leq \eta} \lambda_i$	DQC ₁ [Bra08]	DQC ₁ [CC24]
Low-lying Spectral Density (LLSD)	$2^{-n} \sum_{0 \leq \lambda_i \leq b} 1$	DQC ₁ [GCD22]	DQC ₁ [GCD22]
<i>This work</i>			
Low-energy Norm. Subtrace (LENS)	$\sum_{0 \leq \lambda_i \leq \eta} \lambda_i / \dim \mathcal{S}_{\leq \eta}$	DQC ₁	BQP
Low-energy Spectral Density (LESD)	$\dim \mathcal{S}_{\leq b} / \dim \mathcal{S}_{\leq \eta}$	DQC ₁	BQP
Low-energy Kernel Density (LEKD)	$\dim \ker H / \dim \mathcal{S}_{\leq \eta}$	SDQC ₁	BQP
Norm. Quasi-Persistence (NQP)	$\dim \text{Im}(\mathcal{P}_{2,b} _{\ker H_1}) / \dim \ker H_1$	DQC ₁	BQP
Normalized Persistence (NP)	$\dim \text{Im}(\mathcal{P}_{\ker H_2} _{\ker H_1}) / \dim \ker H_1$	SDQC ₁	BQP
Norm. Quasi-Harmonic Persistence (NQHP)	$\tilde{\beta}_{d,\eta}^{1,2} / \beta_{d,\eta}^1$	DQC ₁	BQP
Normalized Harmonic Persistence (NHP)	$\beta_d^{1,2} / \beta_d^1$	Conj.: SDQC ₁	BQP
<i>Pure-state specializations</i>			
Guided Local Hamiltonian	ground-energy promise with guiding state	BQP [GLG22]	BQP
Low-energy Overlap	$\ \text{Proj}_{\mathcal{S}_{\leq \eta}} \psi\rangle\ ^2$	BQP [GSK+26]	BQP
Kernel Overlap	$\ \text{Proj}_{\ker H} \psi\rangle\ ^2$	BQP ₁ [GSK+26]	BQP
Harmonic Persistence	whether a supplied hole persists	BQP ₁ [GSK+26]	BQP

Table 1: Overview of the problems and results studied in this paper. Except for the promise-problem rows, the task is to estimate the displayed quantity to inverse-polynomial additive precision. Here $\mathcal{S}_{\leq a}$ denotes the subspace spanned by eigenvectors with eigenvalue at most a , $\mathcal{P}_{2,b}$ and $\mathcal{P}_{\ker H_2}$ are the projectors onto the low-energy subspace of H_2 and the kernel of H_2 , and $\tilde{\beta}_{d,\eta}^{1,2}$ is the quasi-persistent Betti number. The containment results are under an approximate state-preparation assumption; moreover, the hard instances constructed in our reductions satisfy this assumption.

however we leave the hardness of Normalized Harmonic Persistence as a conjecture. We further prove that Normalized Harmonic Persistence is contained in BQP under a suitable assumption.

The conjectured hardness of Normalized Harmonic Persistence would follow from an efficient construction in which the Laplacians $\Delta_{1,d}$ and $\Delta_{2,d}$ simulate the exact kernel structure of a particular class of Hamiltonians H_1 and perturbed Hamiltonians $H_2 = H_1 + H_{12}$. The obstruction is that the Hamiltonian simulation results used for the low-energy reductions preserve spectra only approximately, and hence do not generally preserve kernel dimensions.

Results for Local Hamiltonians. The same viewpoint gives a natural generalization to low-energy problems for local Hamiltonians: normalized persistence becomes a question about estimating spectral or overlap quantities inside a relevant low-energy subspace. This setting is practically motivated, since in many physically relevant scenarios the state of interest lies in the low-energy subspace of a local Hamiltonian rather than the full Hilbert space. It is also algorithmically natural: recent Hamiltonian simulation results show that restricting to this subspace can make the simulation error depend on an effective low-energy norm rather than the full operator norm [SS21, GZL24]. On the complexity side, our Hamiltonian results strengthen previous log-local hardness results by establishing DQC₁-hardness for $O(1)$ -local Hamiltonian problems. A comprehensive summary of our results can be seen in Table 1. We remark that the problems studied in this work naturally recover previously studied problems in pure-state specializations such as the guided Local Hamiltonian and harmonic persistence.

Summary and implications. These results give complexity-theoretic evidence for a quantum advantage in TDA based on DQC₁-hardness, closer to the original setting of normalized Betti number estimation. We have introduced a new TDA problem family, centered on the fraction of existing topological features that persist, and connect it to low-energy subspace problems for local Hamiltonians. This viewpoint suggests a broader program of studying normalized spectral and persistence quantities, their state-preparation assumptions, and the boundary between dequantizable and DQC₁-hard regimes.

References

- [Bra08] Fernando GSL Brandão. Entanglement theory and the quantum simulation of many-body physics. *arXiv preprint arXiv:0810.0026*, 2008.
- [CC24] Chris Cade and P Marcos Crichigno. Complexity of supersymmetric systems and the cohomology problem. *Quantum*, 8:1325, 2024.
- [DW22] Tamal Krishna Dey and Yusu Wang. *Computational topology for data analysis*. Cambridge University Press, 2022.
- [GCD22] Casper Gyurik, Chris Cade, and Vedran Dunjko. Towards quantum advantage via topological data analysis. *Quantum*, 6:855, 2022.
- [GLG22] Sevag Gharibian and François Le Gall. Dequantizing the quantum singular value transformation: hardness and applications to quantum chemistry and the quantum pcg conjecture. In *Proceedings of the 54th annual ACM SIGACT symposium on theory of computing*, pages 19–32, 2022.
- [GSK⁺26] Casper Gyurik, Alexander Schmidhuber, Robbie King, Vedran Dunjko, and Ryu Hayakawa. Provable quantum speedups for computing persistence in topological data analysis. *PRX Quantum*, 7:020361, 2026.
- [GZL24] Weiyuan Gong, Shuo Zhou, and Tongyang Li. Complexity of digital quantum simulation in the low-energy subspace: Applications and a lower bound. *Quantum*, 8:1409, 2024.
- [Hay22] Ryu Hayakawa. Quantum algorithm for persistent Betti numbers and topological data analysis. *Quantum*, 6:873, 2022.
- [LGZ16] Seth Lloyd, Silvano Garnerone, and Paolo Zanardi. Quantum algorithms for topological and geometric analysis of data. *Nature communications*, 7(1):10138, 2016.
- [SFH⁺23] Yashbir Singh, Colleen M. Farrelly, Quincy A. Hathaway, Tim Leiner, Jaidip Jagtap, Gunnar E. Carlsson, and Bradley J. Erickson. Topological data analysis in medical imaging: current state of the art. *Insights into Imaging*, 14(1):58, April 2023.
- [SL22] Yara Skaf and Reinhard Laubenbacher. Topological data analysis in biomedicine: A review. *Journal of Biomedical Informatics*, 130:104082, June 2022.
- [ŞS21] Burak Şahinoğlu and Rolando D Somma. Hamiltonian simulation in the low-energy subspace. *npj Quantum Information*, 7(1):119, 2021.