QUANTUM LASSO: CLOSING THE GAP IN HIGH DIMENSIONS VIA ADAPTIVE FRANK–WOLFE

Anonymous authors
Paper under double-blind review

ABSTRACT

We study Lasso, a linear regression problem that minimizes the mean squared loss subject to the coefficient vector having ℓ_1-norm at most one. We use quantum queries to data entries on index superpositions. Chen & de Wolf (2023) (CdW) use approximate quantum minimum finding with additive error $O(\epsilon)$ for Frank–Wolfe direction selection. Each update costs $\widetilde{O}(\sqrt{d}/\epsilon)$ queries, leaving a gap to the high-dimensional lower bound. We propose a new algorithm, Adaptive Quantum Frank–Wolfe (AQFW), which returns an explicit feasible $O(1/\epsilon)$-sparse ϵ-optimal vector with constant confidence using $\widetilde{O}(\sqrt{d}\,\epsilon^{-3/2} + \epsilon^{-2})$ queries. For $d \gtrsim 1/\epsilon$ and $N \gtrsim \epsilon^{-2}\log d$, this improves CdW's bound by $\epsilon^{-1/2}$ and matches their lower bound up to logarithms. AQFW replaces ϵ-precision minimum finding with randomized threshold searches that test estimated loss-decrease slopes along signed coordinate directions (atoms). Smaller thresholds require more precise, costlier estimates, so we reduce Grover iteration caps, where each attempt costs $\widetilde{O}(\sqrt{d/\epsilon} + \epsilon^{-1})$ queries, cheaper than CdW's $\widetilde{O}(\sqrt{d}/\epsilon)$. Accepted step sizes are proportional to the threshold, ensuring sufficient loss decrease under accurate checks. A simplex lift represents θ by a distribution q over atoms. Mixing q with the uniform distribution reuses information and warm-starts searches while keeping every atom accessible. We propose a multiplicative potential that combines the objective suboptimality gap with an exponentiated Kullback–Leibler (KL) divergence from a fixed optimal atom distribution to the current one. Our analysis shows that $\widetilde{O}(\epsilon^{-1})$ searches suffice to reach ϵ-accuracy, matching the baseline's $O(\epsilon^{-1})$ searches up to logarithms. Finite-precision state preparation preserves the bound.

1 INTRODUCTION

The Lasso (Tibshirani, 1996), for a data matrix $X \in [-1,1]^{N\times d}$ and labels $y \in [-1,1]^N$, minimizes the empirical squared loss under an ℓ_1 constraint:

$$\min_{\|\theta\|_1 \le 1} L_{X,y}(\theta) := \frac{1}{N}\|X\theta - y\|_2^2. \tag{1}$$

We seek a feasible ϵ-optimal vector and count queries to entries of X or y (*coordinate query complexity*). Lasso promotes sparsity, which can improve interpretability in high-dimensional prediction problems. For example, genomics and document classification can involve hundreds of thousands of genetic markers or text features (Hastie et al., 2015).

Classical work has established nearly optimal coordinate-query bounds for Lasso in high dimensions. Coordinate queries capture the cost of input access (Cesa-Bianchi et al., 2011; Hazan & Koren, 2012). Reading the whole input costs $O(Nd)$. Classical exponentiated gradient on sampled entries uses $\widetilde{O}(d\epsilon^{-2})$ queries at constant confidence (Hazan & Koren, 2012), matching the lower bound for $d \gtrsim \epsilon^{-2}$ and $N \gtrsim \epsilon^{-2}\log(2d)$ up to logarithms (Chen & de Wolf, 2023; Appendix F.3).

Quantum search and amplitude estimation give quadratic query speedups (Grover, 1996; Brassard et al., 2002). Quantum algorithms also reduce the query complexity of convex optimization under different oracle models (van Apeldoorn et al., 2020; Chakrabarti et al., 2020). For Lasso, Frank–Wolfe repeatedly selects signed coordinate directions using the linearized loss. Quantum implementations accelerate selection using estimation and search (Du et al., 2022; Chen & de Wolf, 2023).

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

Table 1: Coordinate-query bounds for additive-error Lasso optimization at constant confidence, up to logarithmic factors. Details are in Section 2 and Appendix G.

Work	Access	Upper bound	Lower bound †
Hazan & Koren (2012)	classical	$\widetilde{O}(d\epsilon^{-2})$	$\widetilde{\Omega}(d\epsilon^{-2})$ for $d \gtrsim \epsilon^{-2}$
Du et al. (2022)	quantum	$\widetilde{O}(\sqrt{Nd}\,\epsilon^{-4})$	—
Chen & de Wolf (2023)	quantum	$\widetilde{O}(\sqrt{d}\,\epsilon^{-2})$	$\Omega(\sqrt{d}\,\epsilon^{-3/2})$ for $d \gtrsim \epsilon^{-1}$
He & Lui (2025) ‡	quantum	$\widetilde{O}(\sqrt{d}\,\epsilon^{-2})$ or $\widetilde{O}(d\epsilon^{-2})$	—
This work (Theorem 4.1)	quantum	$\widetilde{O}(\sqrt{d}\,\epsilon^{-3/2} + \epsilon^{-2})$	$\Omega(\sqrt{d}\,\epsilon^{-3/2})$ for $d \gtrsim \epsilon^{-1}$

† Lower bounds from Chen & de Wolf (2023) apply here for $N \gtrsim \epsilon^{-2}\log(2d)$ (Appendix F.3).
‡ Our coordinate-query adaptations of minimum finding and Jordan's quantum gradient estimation, respectively. See Appendix G.1 for conversions and Table 3 for other objectives.

Chen & de Wolf (2023) (CdW) use $O(\epsilon^{-1})$ updates, each selecting a direction by $O(\epsilon)$-accurate minimum finding using $\widetilde{O}(\sqrt{d}/\epsilon)$ coordinate queries. Their $\widetilde{O}(\sqrt{d}\,\epsilon^{-2})$ total leaves an $\epsilon^{-1/2}$ gap to their high-dimensional lower bound, up to logarithms (Table 1).

The $\epsilon^{-1/2}$ gap arises because CdW use an expensive $O(\epsilon)$-accurate direction search at every update. This precision is not needed at every update. Using coarser estimates, however, may yield less useful directions and require more iterations, offsetting the savings per search. The challenge is to keep both per-search query cost and total search attempts small.

We introduce Adaptive Quantum Frank–Wolfe (AQFW), summarizing our contributions as follows.

1. *Closing the gap.* Theorem 4.1 gives an algorithm returning an explicit feasible ϵ-optimal vector, sparsified by verified postprocessing to $O(1/\epsilon)$ nonzero entries, with probability $1-\delta$, using $\widetilde{O}(\sqrt{d}\,\epsilon^{-3/2} + \epsilon^{-2})$ coordinate queries up to polylogarithmic confidence factors. At constant confidence, this matches CdW's lower bound for $d \gtrsim 1/\epsilon$ and $N \gtrsim \epsilon^{-2}\log d$ (Corollary 4.2). The upper bound holds for all N. Optimality for $d \lesssim 1/\epsilon$ is not claimed.
2. *Adaptive search and progress analysis.* AQFW replaces reliable approximate quantum minimum finding with randomized threshold searches, the key to closing the gap. Smaller thresholds use fewer Grover iterations. Accepted step sizes are proportional to the threshold. A multiplicative potential tracks loss decrease and the KL divergence from a fixed optimal atom distribution to the current distribution. Its expected decrease bounds the total number of search calls.

Section 2 reviews quantum algorithms for Lasso and general optimization problems. Section 3 introduces the coordinate-query model, the lifted Lasso formulation, and basic quantum algorithms. Section 4 presents AQFW and its main guarantees. Section 5 analyzes convergence. Section 6 describes the quantum implementation and combines its guarantees. The appendices provide detailed proofs, resource bounds, population transfer, and literature comparisons.

2 Related Work

Lasso under different access models. Du et al. (2022) and Chen & de Wolf (2023) assume coherent entry access to X, y. He & Lui (2025) instead count coherent function-value queries returning objective values. Implementing these queries from entries incurs a precision-dependent cost. Table 1 reports our coordinate-query adaptations. However, the best of these entry-query bounds is $\widetilde{O}(\sqrt{d}\,\epsilon^{-2})$, leaving the high-dimensional accuracy gap closed by AQFW.

Other quantum regression guarantees. Li et al. (2019) accelerate sampled primal–dual optimization with multiplicative weights. Lim et al. (2024) quantize truncated gradients for sparse online learning, with regret bounds under data-norm and prediction-error assumptions. Doriguello et al. (2025) track the Lasso path via quantum least-angle regression. Their approximate version and quantum orthogonal matching pursuit (Bellante et al., 2025) use approximate minimum finding, with additional incoherence and conditioning parameters. Liu & Ji (2025) use leverage-score sampling for regression sparsification. Their printed Lasso guarantee is qualified in Appendix G.2. Table 3 lists their native complexities and output goals. However, these guarantees do not provide an explicit feasible ϵ-optimal solution for arbitrary bounded Lasso inputs at AQFW's query cost.

Broader quantum optimization. Finite-sum (Zhang et al., 2024) and stochastic optimization (Sidford & Zhang, 2023; Marsden et al., 2025) use gradient-vector oracles. Zeroth-order methods (van Apeldoorn et al., 2020; Chakrabarti et al., 2020; Kim et al., 2025) require precision-dependent function evaluations. Matrix-game solvers (Bouland et al., 2023; Gao et al., 2023) return classical strategies under different matrix-norm assumptions. However, simply substituting N, d, ϵ into these theorems does not yield a feasible Lasso solution with AQFW's data-entry query bound. The feasible domain, output guarantee, and oracle implementation must also match. Table 4 summarizes the complexities, assumptions, and output guarantees of these papers.

3 Preliminaries: the query model and the lifted problem

Write $[m] := \{1, \dots, m\}$, $\mathrm{supp}(z) := \{k : z_k \neq 0\}$, $x_+ := \max\{x, 0\}$, and $\mathbf{1}[E]$ for the indicator of E. $\widetilde{O}$ suppresses polylogarithmic factors except those in confidence parameters.

Query model. For $N, d \geq 1$, encode X, y entries as b_{in}-bit words via enc and use the oracles

$$O_X|i, j, w\rangle = |i, j, w \oplus \mathrm{enc}(X_{ij})\rangle, \qquad O_y|i, w\rangle = |i, w \oplus \mathrm{enc}(y_i)\rangle,$$

where $i \in [N]$, $j \in [d]$, and w is a b_{in}-bit register. The algorithmic statements use the encoded arrays X, y. Lemma F.1 bounds the effect of rounding. Following Chen & de Wolf (2023), each oracle call or inverse counts as one input query. Oracle realization and data loading are not charged.

Atoms. An *atom* is a signed coordinate vector $\pm e_j \in \mathbb{R}^d$, indexed by $a = (j, \pm) \in \mathcal{A}_d := [d] \times \{+, -\}$. Its corresponding simplex vertex is $e_a \in \mathbb{R}^n$, where $n = 2d$.

The lifted problem. Lift the unit ℓ_1 ball to the simplex over its $2d$ vertices $\pm e_j$ (Jaggi, 2014, Section 3.2). The signed design matrix, simplex, and decoding map are

$$Z_{\cdot,(j,\pm)} := \pm X_{\cdot j}, \qquad \Delta_n := \Big\{ q \in \mathbb{R}^n_+ : \textstyle\sum_{a \in \mathcal{A}_d} q_a = 1 \Big\}, \qquad (Bq)_j := q_{j,+} - q_{j,-},$$

Set $f(q) := N^{-1}\|Zq - y\|_2^2$. The decoding map satisfies $Be_{(j,\pm)} = \pm e_j$.

Lemma 3.1 (Exact lift)**.** *The map B sends Δ_n onto the unit ℓ_1 ball, and $Zq = X(Bq)$, $f(q) = L_{X,y}(Bq)$ for every $q \in \Delta_n$. Hence $\min_{q \in \Delta_n} f(q) = \min_{\|\theta\|_1 \leq 1} L_{X,y}(\theta)$.*

See Appendix A.1 for a proof.

Gradient and violations. With $M_Z := \frac{1}{N} Z^\top Z$ and $\|w\|_Z^2 := w^\top M_Z w = \frac{1}{N}\|Zw\|_2^2$,

$$\nabla f(q) = \frac{2}{N} Z^\top (Zq - y), \qquad \nabla^2 f(q) = 2M_Z.$$

The *violation* $v_a(q) := \langle \nabla f(q), q - e_a \rangle$ is the slope of loss decrease toward vertex e_a. Set $f^\star := \min_{z \in \Delta_n} f(z)$, fix a minimizer $q^\star$ for analysis, and write $q^\star(V) := \sum_{a \in V} q^\star_a$.

Bounded entries give, for $w \in \mathbb{R}^n$ and $q \in \Delta_n$,

$$\|w\|_Z \leq \|w\|_1, \qquad \|\nabla f(q)\|_\infty \leq 4, \qquad 0 \leq f(q) \leq 4, \qquad |v_a(q)| \leq 8, \tag{2}$$

The quadratic expansion is exact for $p', q \in \mathbb{R}^n$:

$$f(q) = f(p') + \langle \nabla f(p'), q - p' \rangle + \|q - p'\|_Z^2. \tag{3}$$

Two consequences drive the analysis (Lemma A.1). First, the suboptimality gap obeys

$$f(q) - f^\star \leq \sum_{a \in \mathcal{A}_d} q^\star_a v_a(q) \leq \max_{a \in \mathcal{A}_d} v_a(q), \tag{4}$$

The second balances the directional gain against curvature. For $0 \leq \gamma \leq 1$,

$$f\big((1-\gamma)q + \gamma e_a\big) \leq f(q) - \gamma v_a(q) + 4\gamma^2. \tag{5}$$

Frank–Wolfe. Frank–Wolfe (Frank & Wolfe, 1956; Jaggi, 2013) starts from $q_0 \in \Delta_n$ and updates

$$a_t \in \arg\max_{a \in \mathcal{A}_d} v_a(q_t), \qquad q_{t+1} := (1 - \gamma_t)\, q_t + \gamma_t\, e_{a_t}, \qquad \gamma_t := \frac{2}{t+2}. \tag{6}$$

Proposition 3.2 (Frank–Wolfe, Jaggi, 2013, Theorem 1, Chen & de Wolf, 2023, Theorems 8 and 9)**.** *If every a_t maximizes the violation up to additive error $2\gamma_t$, the iterates of Equation* (6) *satisfy $f(q_t) - f^\star \leq 24/(t+2)$ for all $t \geq 1$. Thus $O(1/\epsilon)$ steps reach accuracy ϵ.*

Proposition 3.3 (Quantum search, Boyer et al., 1998, Theorem 3)**.** *Let $g : [n] \to \{0,1\}$ mark an unknown number $m \geq 1$ of elements, accessed through $O_g|i\rangle|b\rangle = |i\rangle|b \oplus g(i)\rangle$. A marked element can be found with $O(\sqrt{n/m})$ expected queries to O_g. Classical sampling needs $\Theta(n/m)$.*

Proposition 3.4 (Amplitude estimation, informal version of Theorem A.2)**.** *If a k-query circuit outputs a bit equal to 1 with probability p, then p can be estimated to additive error ζ with probability $1 - \eta$ using $O(k\zeta^{-1}\log(1/\eta))$ queries, against $\Theta(k\zeta^{-2}\log(1/\eta))$ for classical sampling in the worst case. The estimate can also remain unmeasured for all indices in a superposition.*

4 Algorithm

4.1 Main Challenges and Algorithm Design

Each quantum minimum-finding search in Chen & de Wolf (2023) uses $\widetilde{O}(\sqrt{d}/\epsilon)$ coordinate queries to select an approximately best direction. However, this per-search cost is unnecessary for convergence. As far as we know, existing works (He & Lui, 2025; Braun et al., 2017; Kerdreux et al., 2018) do not directly resolve this coordinate-query issue. Simply capping search effort allows empty returns that leave the distribution unchanged in those iterations, without guaranteeing the original Frank–Wolfe convergence rate. The challenge is to lower per-search cost while keeping total attempts small. Therefore, we must use information from successful searches to retain sufficient average progress and design a suitable potential function to track this progress.

Balancing search cost across scales. Violations $v_a(q)$, the loss-decrease slopes, vary across atoms. Chen and de Wolf use $O(\epsilon)$-accurate estimates throughout, even when coarser estimates suffice. We sample a threshold r uniformly from $8, 4, \ldots, \epsilon/16$ and search for atoms with violations $v_a(q) \geq r/2$, using estimation precision proportional to r. Smaller thresholds require costlier estimates, so we use lower Grover iteration caps for smaller thresholds to balance the cost of different thresholds. Atoms with larger violations qualify at more thresholds, including higher ones with larger Grover caps, giving them more opportunities to be found. This encourages the selection of directions with larger violations, as in minimum finding. For $d \gtrsim \epsilon^{-1}$ at constant confidence, each call costs $\widetilde{O}(\sqrt{d/\epsilon})$, versus $\widetilde{O}(\sqrt{d}/\epsilon)$ for reliable selection (Proposition 6.3; Chen & de Wolf, 2023). Under accurate independent checks, returned atoms satisfy $v_a(q) \geq r/2$. The accepted step size is proportional to the threshold r, balancing linear gain against quadratic curvature and giving $\Omega(r^2)$ loss decrease (Equation (5)).

Reusing the current distribution. A uniform initial state can cause frequent failures under capped Grover search when few atoms pass the threshold. It also discards information from successful searches. We instead prepare the mixture $\sigma_a = \frac{1}{2}(q_a + \frac{1}{2d})$, where the current-distribution component q_a reuses information and the uniform component $1/(2d)$ keeps every atom accessible. An accepted step moves only partway toward its atom. If the threshold is below the atom's violation, the proportional step may leave it useful at later thresholds, while it is more likely to be selected again if it still qualifies, providing a warm start.

Accounting for both immediate and future progress. An accepted update changes both the loss and the distribution used by subsequent searches. A multiplicative potential tracks immediate loss decrease and the KL divergence from a fixed optimal atom distribution to the current distribution. This potential captures the progress of both designs and bounds the number of search calls by $\widetilde{O}(\epsilon^{-1})$ at constant confidence, matching CdW up to logarithmic factors.

4.2 Algorithm

Algorithm 1 starts at $u = (1/n, \ldots, 1/n)$, encoding $\theta = 0$. It makes at most $T = \widetilde{O}(\epsilon^{-1}\log(1/\delta))$ FindAtom calls (Theorem 5.3). Each sampled atom undergoes independent verification (a *post-check*) with failure-probability budget $\beta = \delta/(4T)$. Each call samples uniformly from

$$r_j := 8 \cdot 2^{-j}, \qquad j = 0, \ldots, L-1, \qquad L := \log_2(128/\epsilon) + 1, \tag{7}$$

and updates a returned atom by

$$\gamma := r_j/32, \qquad q \leftarrow (1-\gamma)q + \gamma e_a; \tag{8}$$

A rejection leaves q unchanged. The exact iterate uses the full update history and retains all n atoms in its support. Preparation stores only the latest K accepted updates (Section 6.1). The output

subroutine in Algorithm 3 samples and verifies a sparse vector with at most $16/\epsilon$ nonzero entries. The success tag means verification passed. Theorem 4.1 guarantees joint accuracy. Algorithm 1 uses the bounded uniform threshold sampler of Appendix D.2 and terminates with $(\perp, \text{failure})$ if sampling aborts.

Algorithm 1: $\text{Lasso}(\epsilon, \delta)$

Input: oracle access to X, y, accuracy $\epsilon = 2^{-k}$, failure probability δ
Output: either $(\perp, \text{failure})$ or $(\widehat{\theta}, \text{success})$

$q \leftarrow u$; choose T as in Theorem 5.3;
$\beta \leftarrow \delta/(4T)$;
for $t = 1, \ldots, T$ **do**
  sample threshold r uniformly from $\{8, 4, \ldots, \epsilon/16\}$;
  $a \leftarrow \text{FindAtom}(q, r, \epsilon, \beta)$; {Empty return leaves q unchanged}
  if $a \neq \perp$ **then**
    $\gamma \leftarrow r/32$; $q \leftarrow (1-\gamma)q + \gamma e_a$; {Step size proportional to r}
    append (a, γ) to the recent-update tree (Section 6.1);
return $\text{SparseOutput}(q, \epsilon, \delta)$ *(Algorithm 3)*;

4.3 Main Theorems

Theorem 4.1 (Main theorem). *Let $\epsilon = 2^{-k}$ for an integer $k \geq 3$, and $0 < \delta < 1/3$. In the coordinate-query model of Section 3, there is a quantum algorithm that returns either $(\perp, \text{failure})$ or $(\widehat{\theta}, \text{success})$, where every returned $\widehat{\theta}$ is an explicitly stored vector with $\|\widehat{\theta}\|_1 \leq 1$ and at most $\min\{d, 16/\epsilon\}$ nonzero entries, and*

$$\mathbb{P}\big[\text{success and } L_{X,y}(\widehat{\theta}) \leq \min_{\|\theta\|_1 \leq 1} L_{X,y}(\theta) + \epsilon\big] \geq 1 - \delta.$$

On every measurement branch the algorithm terminates and uses at most

$$Q_{X,y} = \widetilde{O}\Big(\sqrt{d}\,\epsilon^{-3/2}\log(1/\delta) + \epsilon^{-2}\log^2(1/\delta)\Big) \qquad \text{input queries.} \tag{9}$$

With the additional classically writable, coherently readable working-memory assumption of Appendix E, and suppressing confidence factors and polynomial word-length overheads, the algorithm uses $\widetilde{O}(\sqrt{d}\,\epsilon^{-3/2} + \epsilon^{-2})$ classical work (including the cost of generating quantum circuit instructions), $\widetilde{O}(\epsilon^{-1})$ classical memory words, and $\widetilde{O}(1)$ quantum workspace words. Memory excludes input storage. Quantum workspace excludes input-oracle and quantum random-access memory (QRAM) hardware. Resource bounds are in Appendix E.

Proof sketch. The FindAtom interface in Definition 4.3 requires every returned atom to have violation at least $r/2$ and imposes a joint bound on the probabilities of returning any atom and each atom whose violation reaches the threshold.

Assuming this interface, accurate measured post-checks, and uniform threshold sampling, Lemma 5.1 computes the potential decrease resulting from one accepted step. Combining this result with the search guarantee, Lemma 5.2 proves expected contraction averaged over thresholds and search outcomes, including empty returns. Theorem 5.3 gives a horizon $T = \widetilde{O}(\epsilon^{-1}\log(1/\delta))$ to reach $f(q_T) - f^\star < \epsilon/2$, with failure probability at most $\delta/4$.

We prepare the current-point state from the recent-update tree. Proposition 6.1 bounds the state-preparation error. Using this state, Proposition 6.2 bounds the violation-estimation errors in the coherent threshold test and the independent post-check. To implement the interface in Definition 4.3, we sample the Grover iteration count uniformly below a prescribed cap. Lemma A.4 shows that Grover iterations (amplitude amplification) preserve the relative sampling probabilities among atoms marked by the threshold test. Together with Proposition 6.2, this proves the interface for the accurate analysis process and a cost of $\widetilde{O}(\sqrt{d/\epsilon} + \epsilon^{-1}\log(2/\beta))$ input queries per FindAtom call (Proposition 6.3). Lemma 6.4 then transfers the convergence guarantee to the implemented process, adding at most $T\beta \leq \delta/4$ to the failure probability.

Combining Proposition 6.3, Lemma 6.4, and Theorem 5.3 with verified sparse output and the exact lift (Proposition D.1 and Lemma 3.1), we obtain an explicit feasible ϵ-optimal Lasso solution with failure probability at most δ. This yields the total input-query bound $\widetilde{O}(\sqrt{d}\,\epsilon^{-3/2}\log(1/\delta) +$

$\epsilon^{-2}\log^2(1/\delta)$) in Theorem 4.1. Proposition E.1 gives the auxiliary resource bounds. Corollary E.2 removes native working QRAM at additional gate cost. Budgets hold on every branch. Bounded loops ensure termination.

Corollary 4.2 (High-dimensional optimality)**.** *Fix a constant $0 < \delta < 1/20$, let $\epsilon = 2^{-k}$ for an integer $k \geq 3$, and set $c_3 := 2048$. There are universal constants $c_1, c_2 > 0$ such that, for $c_1/d < \epsilon < c_2$ and $N \geq c_3\epsilon^{-2}\log(2d/\delta)$, every quantum algorithm that solves the empirical problem of Theorem 4.1 with probability at least $1-\delta$ on every input makes $\Omega(\sqrt{d}\,\epsilon^{-3/2})$ coordinate queries, and Theorem 4.1 with failure-probability budget δ makes $\widetilde{O}(\sqrt{d}\,\epsilon^{-3/2})$ coordinate queries. In this regime the quantum coordinate query complexity of Lasso is therefore $\widetilde{\Theta}(\sqrt{d}\,\epsilon^{-3/2})$.*

The lower bound is that of Chen & de Wolf (2023), transferred to our model and the empirical problem in Proposition F.4. Our upper bound holds for every N. Its $\widetilde{O}(\epsilon^{-2})$ form for $d \lesssim 1/\epsilon$ is not claimed to be optimal. Theorem F.3 covers population risk.

4.4 The FindAtom interface

We analyze FindAtom through an auxiliary *accurate analysis process*, in which every measured post-check estimates $v_a(q)$ with total error at most $r/16$. This lets us analyze convergence separately from post-check estimation failures. Its construction and transfer to the physical implementation appear with Lemma 6.4 in Section 6.3.

Definition 4.3 (FindAtom interface)**.** All probabilities and guarantees below refer to this accurate analysis process. Fix the transcript, current point $q \in \Delta_n$, accuracy ϵ, a threshold r from Equation (7), and post-check failure-probability budget $0 < \beta < 1/2$. Let $A_{\mathrm{out}} \in \mathcal{A}_d \cup \{\perp\}$ be the output of $\mathrm{FindAtom}(q, r, \epsilon, \beta)$, and write $\sigma_a := (q_a + 1/n)/2$ and $c := \min\{r, 1/n\}$. The procedure satisfies the interface if both conditions hold:

(i) every returned atom a satisfies $v_a(q) \geq r/2$;
(ii) for every analysis parameter $H \geq \epsilon$ and every atom with $v_a(q) \geq r$,

$$r^2\mathbb{P}(A_{\mathrm{out}} \neq \perp) + \frac{Hc}{\sigma_a}\mathbb{P}(A_{\mathrm{out}} = a) \geq \frac{1}{16}\min\{r^2, Hnc\}, \tag{10}$$

with all probabilities conditional on the fixed transcript.

The probabilities $\mathbb{P}(A_{\mathrm{out}} \neq \perp)$ and $\mathbb{P}(A_{\mathrm{out}} = a)$ refer to returning any atom and the specified atom, respectively. The first term accounts for loss decrease. The second captures the contribution to reducing the KL divergence from a fixed optimal atom distribution to the current distribution. The KL divergence need not decrease at every step. Either term can compensate for the other. The algorithm knows neither H nor $q^\star$.

5 Convergence

Analysis conditions. Consider the update loop of Algorithm 1 with the parameters of Theorem 4.1, starting from $q_0 = u$. We analyze the accurate process with independent uniform threshold sampling without a trial cap and accurate post-checks. Rejections leave q unchanged. Let $\mathcal{F}$ denote the history before a search and assume Definition 4.3 conditional on $\mathcal{F}$ and the chosen threshold r. Section 6 realizes this interface and transfers the guarantee to the implemented run.

The potential. Fix an optimal point $q^\star$. Updates preserve positive coordinates, and accurate checks make the suboptimality gap nonincreasing. Hence $0 \leq f(q) - f^\star \leq f(u) - f^\star \leq 1$. We use

$$\mathcal{W}(q) := (f(q) - f^\star + \epsilon)e^{\rho D(q)}, \qquad D(q) := \sum_{a:\, q_a^\star > 0} q_a^\star \log\frac{q_a^\star}{q_a} \geq 0, \qquad \rho := 2^{-25}. \tag{11}$$

$D(q)$ is the Kullback–Leibler divergence from $q^\star$ to q. In the relative-decrease bound of Lemma 5.1, loss decrease and increased weight on an atom used by $q^\star$ contribute positively, while shrinking other coordinates incurs the negative term. The bound $\sigma_a\bar{\ell}_a \geq c/64$ links the KL gain to the per-atom probability term in Equation (10), allowing the product potential to capture progress that a loss-only analysis misses.

Lemma 5.1 (One accepted step, proof in Appendix B.2)**.** *Let $q \in \Delta_n$ have positive coordinates, $0 < r \le 8$, $\gamma := r/32$, and $q^+ := (1-\gamma)q + \gamma e_a$. Write*

$$H := f(q) - f^\star + \epsilon, \qquad \ell_a := \log\Big(1 + \frac{\gamma}{(1-\gamma)q_a}\Big), \qquad \bar{\ell}_a := \min\{\ell_a, 1\}.$$

$$D(q^+) - D(q) = -\log(1-\gamma) - q_a^\star \ell_a \le 2\gamma. \tag{12}$$

If $v_a(q) \ge r/2$, with $\sigma_a := (q_a + 1/n)/2$ and $c := \min\{r, 1/n\}$,

$$1 - \frac{\mathcal{W}(q^+)}{\mathcal{W}(q)} \ge \frac{3r^2}{512H} + \frac{\rho}{4} q_a^\star \bar{\ell}_a - \frac{\rho r}{8}, \qquad \sigma_a \bar{\ell}_a \ge \frac{c}{64}. \tag{13}$$

In the accurate analysis process, accepted updates satisfy $v_a(q) \ge r/2$ and obey Lemma 5.1, while rejections leave $\mathcal{W}$ unchanged. Although an accepted update decreases loss, it may increase $\mathcal{W}$. Using the probability guarantees in Definition 4.3, the next lemma proves expected potential decrease over thresholds and search outcomes when $f(q) - f^\star \ge \epsilon/2$.

Lemma 5.2 (Expected potential contraction per search)**.** *Under the analysis conditions above, fix the history $\mathcal{F}$ before sampling the next threshold, and let q^+ be the point after the search and any resulting update. If $f(q) - f^\star \ge \epsilon/2$, then*

$$\mathbb{E}[\mathcal{W}(q^+) \mid \mathcal{F}] \le \left(1 - \frac{2^{-45}\epsilon}{L}\right) \mathcal{W}(q). \tag{14}$$

$L = \log_2(128/\epsilon) + 1$ counts the number of different thresholds. The expectation is over the uniformly sampled threshold and all search outcomes, including empty returns.

Proof sketch (full version in Appendix B.2). Set $H = f(q) - f^\star + \epsilon$. Rejections leave $\mathcal{W}$ unchanged. For $r \ge 64\rho H$, the loss term in Equation (13) absorbs its negative term, and weighting Equation (10) by $q_a^\star$ gives

$$\mathbb{E}\left[1 - \frac{\mathcal{W}(q^+)}{\mathcal{W}(q)} \,\middle|\, \mathcal{F}, r\right] \ge \frac{\rho}{4096H} \min\{r^2, Hnc\} \sum_{a:\, v_a(q) \ge r} q_a^\star.$$

For $r \ge H/16$, the loss term in Equation (13) dominates its negative term, and Equation (4) ensures sufficient total expected progress. Intermediate thresholds contribute nonnegative progress. For $r \le 64\rho H$, the possible negative contributions sum to at most $16\rho^2 H$ in magnitude before averaging. Averaging over thresholds and using $H \le 3(f(q) - f^\star)$ and $\rho = 2^{-25}$ yields Equation (14).

Theorem 5.3 (Search budget for high-probability convergence)**.** *Under the analysis conditions above, choose $T := \lceil (2^{45} L/\epsilon)[\log(8/(\epsilon\delta)) + \rho \log n] \rceil = \widetilde{O}(\epsilon^{-1} \log(1/\delta))$. T counts FindAtom calls, including empty returns. The resulting point satisfies*

$$\mathbb{P}\big(f(q_T) - f^\star \ge \epsilon/2\big) \le \delta/4. \tag{15}$$

Proof sketch (full version in Appendix B.3). Let $\mathcal{F}_t$ be the history after t searches and their updates, and set $a := 2^{-45}\epsilon/L$. Uniform initialization gives $D(u) \le \log n$ and $\mathcal{W}(u) \le (1+\epsilon)n^\rho$. Define $Y_t := \mathbf{1}[f(q_t) - f^\star \ge \epsilon/2]\mathcal{W}(q_t)$. Since accurate updates never increase the gap, this quantity remains zero once the target is reached. Before then, Lemma 5.2 gives $\mathbb{E}[Y_{t+1} \mid \mathcal{F}_t] \le (1-a)Y_t$. Setting it to zero after reaching the target can only reduce it. Iterating yields $\mathbb{E}[Y_T] \le (1+\epsilon)n^\rho e^{-aT}$. On failure, $D(q_T) \ge 0$ implies $Y_T \ge 3\epsilon/2$, so Markov's inequality gives

$$\mathbb{P}(f(q_T) - f^\star \ge \epsilon/2) \le \frac{2(1+\epsilon)n^\rho}{3\epsilon} e^{-aT} \le \frac{2n^\rho}{\epsilon} e^{-aT} \le \frac{\delta}{4}$$

by the choice of T.

6 Implementing the search and composing the guarantees

6.1 Preparing the point and estimating violations

The stored point. The accepted steps $(a_i, \gamma_i)_{i=1}^U$ define the exact iterate:

$$q = \prod_{i \le U} (1-\gamma_i) u + \sum_{i \le U} \gamma_i \prod_{i < j \le U} (1-\gamma_j) e_{a_i}. \tag{16}$$

We use a recent-update tree for state preparation. To realize Equation (16), consider the following sampling procedure. Inspect updates from newest to oldest, returning a_i with probability γ_i at update i and continuing otherwise. If none is selected, sample a uniform atom. The tree implements this rule coherently. See Appendix C.1 for the detailed construction and proof.

Why are the latest K records sufficient? Each accepted update scales earlier weights by $1-\gamma_i$. After K updates, the older history has weight at most $(1-\gamma_{\min})^K$, where $\gamma_{\min}$ is the minimum step size. This exponential decay lets us replace the older contribution by the same weight on the uniform point for preparation. We choose K to keep truncation error within budget ν (Equation (29)). The full history still defines the exact iterate. See Appendix C.2 for the detailed proof.

Proposition 6.1 (State preparation; proof in Appendix C.4)**.** *Measuring the atom register of $|\psi_q\rangle = \sum_a \sqrt{q_a}|a\rangle|\xi_a\rangle$ with normalized labels ξ_a gives q. The recent-update tree prepares, without input queries, a state $\widehat{U}_q|0\rangle$ with $\|\widehat{U}_q|0\rangle - |\psi_q\rangle\| \le \nu$ for some such labels. The representation uses $O(K) = \widetilde{O}(\epsilon^{-1})$ classical words. Initialization costs $O(K)$ word operations, and each accepted update costs $O(\log K)$, including cache maintenance. Each preparation uses $O(\log K)$ layers of tree-node reads and controlled rotations at superposed addresses, plus uniform-atom preparation.*

Estimating a violation. For I uniform on $[N]$ and independent $A_1, A_2 \sim q$,

$$v_a(q) = 2\mathbb{E}[(Z_{IA_1} - Z_{Ia})(Z_{IA_2} - y_I)]$$

(Proved in Lemma A.3). Independence of A_1, A_2 conditional on I turns the expectation into a product of row means. State preparation, four entry reads (eight oracle calls with uncomputation), and a qubit rotation encode $v_a(q)$ affinely in an ideal flag probability. Amplitude estimation acts on the implemented circuit. Finite-precision state preparation and gates bias its encoded mean (Proposition 6.2). The coherent *threshold-test circuit* V_r writes $F = 1$ when the estimate is at least $3r/4$, keeping all work registers. An independent measured estimate verifies the sampled atom.

Proposition 6.2 (Violation estimation)**.** *Under the preparations in Appendix C.4, the encoded mean differs from $v_a(q)$ by at most $r/32$. **(i) Coherent test.** Tolerance $r/16$ for this mean gives total error $< r/8$ with failure probability $\le \eta = 1/(64n)$. Flagging estimates $\ge 3r/4$ gives flag probabilities $\le \eta$ for $v_a(q) < 5r/8$ and $\ge 1-\eta$ for $v_a(q) \ge r$. Each use or inverse costs $O(r^{-1}\log n)$ queries. **(ii) Independent post-check.** Tolerance $r/32$ gives total error $\le r/16$ with conditional failure probability $\le \beta$. Accept estimates $\ge 9r/16$. Accurate checks accept $v_a(q) \ge 5r/8$ and reject $v_a(q) < r/2$. The cost is $O(r^{-1}\log(2/\beta))$ queries.*

6.2 Randomized amplitude amplification

The ideal initial search distribution is

$$\sigma_a := \tfrac{1}{2}(q_a + \tfrac{1}{n}), \qquad a \in \mathcal{A}_d. \tag{17}$$

The uniform half gives every atom probability at least $1/(2n)$. The current-distribution half favors atoms with larger q_a. Let W_r denote the finite circuit that prepares the mixture state and then applies the coherent threshold test V_r, which sets $F = 1$ when the estimated violation is at least $3r/4$. We retain the mixture label and work registers throughout Grover search so that $W_r^\dagger$ can reverse the preparation and threshold test. Let Π_F project onto $F = 1$. The reflection $I - 2\Pi_F$ flips the phase of flagged components without repeating the threshold test. Define the Grover iterate by

$$G := W_r(2|0\rangle\langle 0| - I)W_r^\dagger(I - 2\Pi_F).$$

The Grover budget balances the cost of a threshold test against the probability of useful progress. For Algorithm 2, define it with a local integer κ:

$$c := \min\{r, 1/n\}, \qquad R := 2^\kappa, \qquad \kappa := \left\lceil \log_2 \max\left\{1, \min\left\{r/\sqrt{\epsilon c}, \sqrt{n}\right\}\right\}\right\rceil. \tag{18}$$

Algorithm 2: $\mathrm{FindAtom}(q, r, \epsilon, \beta)$: randomized amplitude amplification for a passing atom

Input: current exact iterate q, threshold r, accuracy ϵ, post-check failure-probability budget β
Output: an atom passing an independent post-check, or $\perp$

prepare $W_r|0\rangle$ using the finite mixture and threshold test; {Warm start}
sample i uniformly from $\{0, \dots, R-1\}$; {Capped Grover search}
apply G^i to $W_r|0\rangle$ and measure (a, F);
if $F = 0$ **then return** $\perp$;
$\widetilde{v}_a \leftarrow$ independently estimate $v_a(q)$ with tolerance $r/32$ before bias and conditional failure probability at most β;
return a *if* $\widetilde{v}_a \ge 9r/16$, *else* $\perp$;

Proposition 6.3 (Guarantees of the FindAtom implementation)**.** *Fix the transcript before Algorithm 2, with $\epsilon = 2^{-k} \le 1/8$, $r = 8 \cdot 2^{-j} \in [\epsilon/16, 8]$, and post-check budget $0 < \beta < 1/2$.*

Use preparation tolerance $\nu = 2^{-20}\min\{1/n, \epsilon\}$, ensuring $2\nu \le 1/(4n) \le \sigma_a/2$ for every atom. In the accurate analysis process, returned atoms satisfy $v_a(q) \ge r/2$, and Equation (10) *holds for every $H \ge \epsilon$. Each call uses at most $2R-1$ applications of W_r or $W_r^\dagger$ and at most one post-check on every branch, costing $O((\sqrt{n/\epsilon} + \epsilon^{-1})\log n + \epsilon^{-1}\log(2/\beta))$ input queries.*

Proof sketch (full version in Appendix C.5). Let $\bar{\mu}$ be the initial flag probability, $x = R^2\bar{\mu}$, and fix a with $v_a(q) \ge r$. Finite preparation and Proposition 6.2 give $\widehat{\sigma}_a \ge \sigma_a/2$ and initial flag probability $\mu_a := \mathbb{P}(F = 1 \mid a) \ge 1 - \eta \ge 1/2$, hence $\bar{\mu} \ge 1/(8n)$ (Equation (38)). Flagged mass below $5r/8$ is at most $\eta = 1/(64n) \le \bar{\mu}/8$, so at least $7/8$ passes accurate post-checks. By Lemma A.4, randomized amplification gives flag probability $g \ge \min\{1, x\}/4$ and preserves the conditional atom distribution given $F = 1$. Thus $\mathbb{P}(A_{\mathrm{out}} \ne \perp) \ge (7/8)g \ge \min\{1, x\}/8$ and $\mathbb{P}(A_{\mathrm{out}} = a) = g\widehat{\sigma}_a\mu_a/\bar{\mu} \ge g\widehat{\sigma}_a/(2\bar{\mu}) \ge (R^2\sigma_a/16)\min\{1, x\}/x$. For $H \ge \epsilon$, the Grover budget gives $HcR^2 \ge \omega_H := \min\{r^2, Hnc\}$. If $x \le 1$, then $(Hc/\sigma_a)\mathbb{P}(A_{\mathrm{out}} = a) \ge HcR^2/16 \ge \omega_H/16$. If $x \ge 1$, then $r^2\mathbb{P}(A_{\mathrm{out}} \ne \perp) \ge r^2/8 \ge \omega_H/16$. Both terms are nonnegative, so either case proves Equation (10). Counting W_r calls and the post-check gives the cost.

6.3 Failure probabilities

Proposition 6.3 accounts for finite-circuit bias and coherent-test error $\eta = 1/(64n)$. Grover's fixed unitary and exact inverse need no extra preparation precision. We control adaptive post-check errors.

For each transcript h immediately before a post-check, including the current q, threshold r, and selected atom a, let P_h be the physical law of the measured record z, $\widetilde{v}_a(z)$ its estimate, and $\bar{v}_a(h)$ the finite violation circuit's encoded mean. Define the accurate records and replacement kernel by

$$\mathcal{G}_h := \{z : |\widetilde{v}_a(z) - \bar{v}_a(h)| \le r/32\}, \qquad \mathsf{K}_h(C) := \frac{P_h(C \cap \mathcal{G}_h)}{P_h(\mathcal{G}_h)}. \tag{19}$$

By Proposition 6.2 and Equation (36), $P_h(\mathcal{G}_h) \ge 1 - \beta > 0$ and $|\bar{v}_a(h) - v_a(q)| \le r/32$, so K_h is well defined and yields estimates with total error at most $r/16$. The auxiliary accurate process uses K_h at each post-check and the actual finite preparation and search circuits at its own transcript. It is used only for analysis. The physical algorithm neither computes $\bar{v}_a(h)$ nor tests membership in $\mathcal{G}_h$.

Lemma 6.4 (Adaptive failure transfer)**.** *Suppose the physical and accurate processes use identical initialization and threshold-sampling rules and make at most T measured post-checks. If $P_h(\mathcal{G}_h^c) \le \beta$ for every preceding transcript h and $0 < \beta \le \delta/(4T)$, their transcript and output distributions under the same randomized output subroutine differ by at most $T\beta \le \delta/4$ in total variation.*

Proof. At a common transcript h, couple the finite circuits and random controls identically. Draw $z \sim P_h$ for the physical post-check. The accurate run uses z if $z \in \mathcal{G}_h$ and an independent draw from K_h otherwise. Its marginal law is K_h, since for every set C,

$$P_h(C \cap \mathcal{G}_h) + P_h(\mathcal{G}_h^c)\mathsf{K}_h(C) = \mathsf{K}_h(C).$$

Each run follows its own conditional kernels after disagreement. Before then, each post-check disagrees with probability at most β. A union bound and the coupling inequality give transcript total variation at most $T\beta$, which the common randomized output subroutine cannot increase.

With failure budget $\delta/4$ each for convergence, measured post-checks, bounded threshold sampling, and verified sparse output, a union bound gives success probability at least $1 - \delta$ (Appendix D.2).

7 Conclusion, Limitation, and Discussion

Conclusion: AQFW reduces quantum Lasso query costs through adaptive threshold searches and reuse of the current distribution. A multiplicative potential bounds the number of searches despite empty returns. Our method closes the gap in the high-dimensional regime. **Limitation:** Our query bounds exclude data-loading costs, and constants are not optimized. We leave the regime $d \lesssim 1/\epsilon$ unoptimized, since the $\Omega(\sqrt{d}\,\epsilon^{-3/2})$ lower bound of Chen & de Wolf (2023) applies for $d \gtrsim 1/\epsilon$ under the conditions in Appendix F.3.

Discussion: Our Frank–Wolfe analysis relies on convexity and uniformly bounded directional curvature, while our violation estimators and sparse-output guarantees rely on squared loss. Extending AQFW to other objectives over convex hulls of finite atom sets is a natural direction, but requires separate convergence and query-complexity analyses. The potential function, update rule, and coherent threshold test may need to be adapted to the properties of the new objective.

ARTIFICIAL INTELLIGENCE (AI) USE STATEMENT

We used Claude (Anthropic) and ChatGPT/Codex (OpenAI) as supporting tools in the preparation of this work. The underlying mathematical ideas, proof strategies, and overall approach were developed by the authors. These tools were used to assist in carrying out mathematical derivations and constructing proofs based on the authors' ideas and approaches, as well as to assist with auxiliary statements, literature comparisons, and the translation, polishing, restructuring, and editing of text originally drafted by the authors, together with pseudocode, tables, and figures. All AI-assisted mathematical derivations and proofs were carefully checked and verified by the authors, and all literature-related content and citations were independently reviewed by the authors. This is a theoretical work. No datasets or experimental results were generated using these tools. The authors take full responsibility for the final mathematical claims, proofs, citations, and other submitted content.

REPRODUCIBILITY STATEMENT

Algorithms 1 to 3 specify the main algorithm, atom search, and sparse-output verification, with all parameters. Complete convergence proofs are in Appendix B. Appendix C gives state preparation and finite-circuit error bounds, Appendix D proves the explicit sparse-output guarantee, and Appendix E gives resource bounds on every measurement branch. Statistical and lower-bound transfers are in Appendix F. Section 2 and Appendix G detail the baseline reductions and differences in access models. No experiments are reported.

REFERENCES

Armando Bellante, Stefano Vanerio, and Stefano Zanero. Quantum sparse recovery and quantum orthogonal matching pursuit, 2025. URL https://arxiv.org/abs/2510.06925v1. arXiv:2510.06925.

Charles H. Bennett. Logical reversibility of computation. *IBM Journal of Research and Development*, 17(6):525–532, 1973. doi: 10.1147/rd.176.0525.

Adam Bouland, Yosheb M. Getachew, Yujia Jin, Aaron Sidford, and Kevin Tian. Quantum speedups for zero-sum games via improved dynamic Gibbs sampling. In Andreas Krause, Emma Brunskill, Kyunghyun Cho, Barbara Engelhardt, Sivan Sabato, and Jonathan Scarlett (eds.), *Proceedings of the 40th International Conference on Machine Learning*, volume 202 of *Proceedings of Machine Learning Research*, pp. 2932–2952. PMLR, 2023. URL https://proceedings.mlr.press/v202/bouland23a.html.

Michel Boyer, Gilles Brassard, Peter Høyer, and Alain Tapp. Tight bounds on quantum searching. *Fortschritte der Physik*, 46(4-5):493–505, 1998. doi: 10.1002/(SICI)1521-3978(199806)46:4/5<493::AID-PROP493>3.0.CO;2-P.

Gilles Brassard, Peter Høyer, Michele Mosca, and Alain Tapp. Quantum amplitude amplification and estimation. In Samuel J. Lomonaco, Jr. and Howard E. Brandt (eds.), *Quantum Computation and Information*, volume 305 of *Contemporary Mathematics*, pp. 53–74. American Mathematical Society, 2002. doi: 10.1090/conm/305/05215.

Gábor Braun, Sebastian Pokutta, and Daniel Zink. Lazifying conditional gradient algorithms. In Doina Precup and Yee Whye Teh (eds.), *Proceedings of the 34th International Conference on Machine Learning*, volume 70 of *Proceedings of Machine Learning Research*, pp. 566–575. PMLR, 2017. URL https://proceedings.mlr.press/v70/braun17a.html.

Nicolò Cesa-Bianchi, Shai Shalev-Shwartz, and Ohad Shamir. Efficient learning with partially observed attributes. *Journal of Machine Learning Research*, 12(87):2857–2878, 2011. URL https://jmlr.org/papers/v12/cesa-bianchi11a.html.

Shouvanik Chakrabarti, Andrew M. Childs, Tongyang Li, and Xiaodi Wu. Quantum algorithms and lower bounds for convex optimization. *Quantum*, 4:221, 2020. doi: 10.22331/q-2020-01-13-221.

Yanlin Chen and Ronald de Wolf. Quantum algorithms and lower bounds for linear regression with norm constraints. In Kousha Etessami, Uriel Feige, and Gabriele Puppis (eds.), *50th International Colloquium on Automata, Languages, and Programming (ICALP 2023)*, volume 261

of *Leibniz International Proceedings in Informatics (LIPIcs)*, pp. 38:1–38:21. Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2023. doi: 10.4230/LIPIcs.ICALP.2023.38. Full version: arXiv:2110.13086v2.

Christopher M. Dawson and Michael A. Nielsen. The Solovay–Kitaev algorithm. *Quantum Information and Computation*, 6(1):81–95, 2006. doi: 10.26421/QIC6.1-6. arXiv:quant-ph/0505030.

Joao F. Doriguello, Debbie Lim, Chi Seng Pun, Patrick Rebentrost, and Tushar Vaidya. Quantum algorithms for the pathwise Lasso. *Quantum*, 9:1674, 2025. doi: 10.22331/q-2025-03-25-1674.

Yuxuan Du, Min-Hsiu Hsieh, Tongliang Liu, Shan You, and Dacheng Tao. Quantum differentially private sparse regression learning. *IEEE Transactions on Information Theory*, 68(8):5217–5233, 2022. doi: 10.1109/TIT.2022.3164726.

Marguerite Frank and Philip Wolfe. An algorithm for quadratic programming. *Naval Research Logistics Quarterly*, 3(1-2):95–110, 1956. doi: 10.1002/nav.3800030109.

Minbo Gao, Zhengfeng Ji, Tongyang Li, and Qisheng Wang. Logarithmic-regret quantum learning algorithms for zero-sum games. In A. Oh, T. Naumann, A. Globerson, K. Saenko, M. Hardt, and S. Levine (eds.), *Advances in Neural Information Processing Systems*, volume 36, pp. 31177–31203. Curran Associates, Inc., 2023. doi: 10.52202/075280-1359. URL https://proceedings.neurips.cc/paper_files/paper/2023/file/637df18481a6aa74238bd2cafff94cb9-Paper-Conference.pdf.

Lov K. Grover. A fast quantum mechanical algorithm for database search. In *Proceedings of the Twenty-Eighth Annual ACM Symposium on Theory of Computing*, STOC '96, pp. 212–219. ACM, 1996. doi: 10.1145/237814.237866.

Lov K. Grover and Terry Rudolph. Creating superpositions that correspond to efficiently integrable probability distributions, 2002. URL https://arxiv.org/abs/quant-ph/0208112v1. arXiv:quant-ph/0208112.

Trevor Hastie, Robert Tibshirani, and Martin Wainwright. *Statistical Learning with Sparsity: The Lasso and Generalizations*. Chapman and Hall/CRC, 2015. doi: 10.1201/b18401.

Elad Hazan and Tomer Koren. Linear regression with limited observation. In John Langford and Joelle Pineau (eds.), *Proceedings of the 29th International Conference on Machine Learning*, pp. 807–814. Omnipress, 2012. URL https://icml.cc/2012/papers/433.pdf.

Jianhao He and John C. S. Lui. Quantum algorithms for projection-free sparse convex optimization, 2025. URL https://arxiv.org/abs/2507.08543v1. arXiv:2507.08543.

Martin Jaggi. Revisiting Frank-Wolfe: Projection-free sparse convex optimization. In Sanjoy Dasgupta and David McAllester (eds.), *Proceedings of the 30th International Conference on Machine Learning*, volume 28 of *Proceedings of Machine Learning Research*, pp. 427–435. PMLR, 2013. URL https://proceedings.mlr.press/v28/jaggi13.html.

Martin Jaggi. An equivalence between the Lasso and support vector machines. In Johan A. K. Suykens, Marco Signoretto, and Andreas Argyriou (eds.), *Regularization, Optimization, Kernels, and Support Vector Machines*, chapter 1, pp. 19–44. Chapman and Hall/CRC, 2014. doi: 10.1201/b17558-4. arXiv:1303.1152.

Stephen P. Jordan. Fast quantum algorithm for numerical gradient estimation. *Physical Review Letters*, 95(5):050501, 2005. doi: 10.1103/PhysRevLett.95.050501.

Ishani Karmarkar, Liam O'Carroll, and Aaron Sidford. Solving matrix games with near-optimal matvec complexity. In *Proceedings of the 58th Annual ACM Symposium on Theory of Computing*, STOC '26, pp. 1728–1738. ACM, 2026. doi: 10.1145/3798129.3800880. Full version: arXiv:2601.02347v3.

Thomas Kerdreux, Fabian Pedregosa, and Alexandre d'Aspremont. Frank-Wolfe with subsampling oracle. In Jennifer Dy and Andreas Krause (eds.), *Proceedings of the 35th International Conference on Machine Learning*, volume 80 of *Proceedings of Machine Learning Research*, pp. 2591–2600. PMLR, 2018. URL https://proceedings.mlr.press/v80/kerdreux18a.html.

Junhyung Lyle Kim, Brandon Augustino, Dylan Herman, Enrico Fontana, Jacob Watkins, Marco Pistoia, and Shouvanik Chakrabarti. Fast zeroth-order convex optimization with quantum gradient methods. In D. Belgrave, C. Zhang, H. Lin, R. Pascanu, P. Koniusz, M. Ghassemi, and N. Chen (eds.), *Advances in Neural Information Processing Systems*, volume 38, pp. 84556–84604. Curran Associates, Inc., 2025. doi: 10.52202/085713-2834. URL https://proceedings.neurips.cc/paper_files/paper/2025/file/79f3fb79a8fbc57a63ae7126c8193508-Paper-Conference.pdf.

Tongyang Li, Shouvanik Chakrabarti, and Xiaodi Wu. Sublinear quantum algorithms for training linear and kernel-based classifiers. In Kamalika Chaudhuri and Ruslan Salakhutdinov (eds.), *Proceedings of the 36th International Conference on Machine Learning*, volume 97 of *Proceedings of Machine Learning Research*, pp. 3815–3824. PMLR, 2019. URL https://proceedings.mlr.press/v97/li19b.html.

Debbie Lim, Yixian Qiu, Patrick Rebentrost, and Qisheng Wang. Quantum algorithm for sparse online learning with truncated gradient descent, 2024. URL https://arxiv.org/abs/2411.03925v1. arXiv:2411.03925.

Chengchang Liu, Zongqi Wan, Jialin Zhang, Xiaoming Sun, and John C. S. Lui. Quantum speedups for minimax optimization and beyond. In D. Belgrave, C. Zhang, H. Lin, R. Pascanu, P. Koniusz, M. Ghassemi, and N. Chen (eds.), *Advances in Neural Information Processing Systems*, volume 38, pp. 119100–119128. Curran Associates, Inc., 2025. doi: 10.52202/085713-3974. URL https://proceedings.neurips.cc/paper_files/paper/2025/file/acb3e20075b0a2dfa3565f06681578e5-Paper-Conference.pdf.

Chenghua Liu and Zhengfeng Ji. Accelerating regression tasks with quantum algorithms, 2025. URL https://arxiv.org/abs/2509.24757v1. arXiv:2509.24757.

Annie Marsden, Liam O'Carroll, Aaron Sidford, and Chenyi Zhang. Isotropic noise in stochastic and quantum convex optimization. In D. Belgrave, C. Zhang, H. Lin, R. Pascanu, P. Koniusz, M. Ghassemi, and N. Chen (eds.), *Advances in Neural Information Processing Systems*, volume 38, pp. 111482–111523. Curran Associates, Inc., 2025. doi: 10.52202/085713-3722. URL https://proceedings.neurips.cc/paper_files/paper/2025/file/a1c0bb652b64f99e26b5bcd60ed2c68f-Paper-Conference.pdf.

Mehryar Mohri, Afshin Rostamizadeh, and Ameet Talwalkar. *Foundations of Machine Learning*. MIT Press, second edition, 2018. URL https://mitpress.mit.edu/9780262039406/foundations-of-machine-learning/.

Ashley Montanaro. Quantum speedup of Monte Carlo methods. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 471(2181):20150301, 2015. doi: 10.1098/rspa.2015.0301.

Aaron Sidford and Chenyi Zhang. Quantum speedups for stochastic optimization. In A. Oh, T. Naumann, A. Globerson, K. Saenko, M. Hardt, and S. Levine (eds.), *Advances in Neural Information Processing Systems*, volume 36, pp. 35300–35330. Curran Associates, Inc., 2023. doi: 10.52202/075280-1534. URL https://proceedings.neurips.cc/paper_files/paper/2023/file/6ed9931d6e1fb6a85efa1b2c014a47e1-Paper-Conference.pdf.

Robert Tibshirani. Regression shrinkage and selection via the Lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, 58(1):267–288, 1996. doi: 10.1111/j.2517-6161.1996.tb02080.x.

Tommaso Toffoli. Reversible computing. In Jaco de Bakker and Jan van Leeuwen (eds.), *Automata, Languages and Programming*, volume 85 of *Lecture Notes in Computer Science*, pp. 632–644. Springer, 1980. doi: 10.1007/3-540-10003-2_104.

Joran van Apeldoorn, András Gilyén, Sander Gribling, and Ronald de Wolf. Convex optimization using quantum oracles. *Quantum*, 4:220, 2020. doi: 10.22331/q-2020-01-13-220.

Yexin Zhang, Chenyi Zhang, Cong Fang, Liwei Wang, and Tongyang Li. Quantum algorithms and lower bounds for finite-sum optimization. In Ruslan Salakhutdinov, Zico Kolter, Katherine Heller,

Adrian Weller, Nuria Oliver, Jonathan Scarlett, and Felix Berkenkamp (eds.), *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, pp. 60244–60270. PMLR, 2024. URL https://proceedings.mlr.press/v235/zhang24bz.html.

A Proofs for Section 3 and the quantum primitives

A.1 Geometry of the lifted problem

Lemma 3.1 (Exact lift). *Let $N, d \geq 1$, $X \in [-1,1]^{N\times d}$, and $y \in [-1,1]^N$. Set $\mathcal{A}_d = [d]\times\{+,-\}$, $n = 2d$, and let e_a be the atom basis vectors. Define the signed design matrix, decoding map, simplex, and objectives by*

$$Z_{\cdot,(j,\pm)} = \pm X_{\cdot j}, \qquad (Bq)_j = q_{j,+} - q_{j,-}, \qquad \Delta_n = \{q \in \mathbb{R}^n_+ : \textstyle\sum_a q_a = 1\},$$
$$f(q) = \frac{1}{N}\|Zq - y\|_2^2, \qquad L_{X,y}(\theta) = \frac{1}{N}\|X\theta - y\|_2^2.$$

The following properties hold.

(i) The map B sends Δ_n onto the unit ℓ_1 ball.
(ii) For every $q \in \Delta_n$, $Zq = X(Bq)$ and $f(q) = L_{X,y}(Bq)$.
(iii) The minima satisfy $\min_{q\in\Delta_n} f(q) = \min_{\|\theta\|_1\leq 1} L_{X,y}(\theta)$.

Proof. **Proof of (i).** For $q \in \Delta_n$, $\|Bq\|_1 \leq \sum_j (q_{j,+} + q_{j,-}) = 1$. Conversely, write $\theta = \theta^+ - \theta^-$ for a vector with $\|\theta\|_1 \leq 1$. Set $q_{j,\pm} = \theta_j^\pm$ and add $(1-\|\theta\|_1)/2$ to both $q_{1,+}$ and $q_{1,-}$. The resulting $q \in \Delta_n$ satisfies $Bq = \theta$.

Proof of (ii). By definition, $(Zq)_i = \sum_j X_{ij}(q_{j,+} - q_{j,-}) = (X(Bq))_i$. Thus the objectives agree.

Proof of (iii). Surjectivity in (i) and the objective identity in (ii) give equality of the minima. □

Lemma A.1 (Geometry, suboptimality gap, and Lipschitz bounds)**.** *For the lifted objective, let $M_Z = Z^\top Z/N$ and $\|w\|_Z^2 = w^\top M_Z w$. The following properties hold.*

(i) The basic bounds are
$$\|w\|_Z \leq \|w\|_1, \quad \|\nabla f(q)\|_\infty \leq 4, \quad 0 \leq f(q) \leq 4, \quad |v_a(q)| \leq 8 \qquad (q \in \Delta_n), \tag{2}$$
(ii) For $q, z \in \mathbb{R}^n$, the quadratic identity is
$$f(q) = f(z) + \langle \nabla f(z), q - z\rangle + \|q - z\|_Z^2 \qquad (q, z \in \mathbb{R}^n). \tag{3}$$
(iii) The minimum is attained, and every minimizer $q^\star$ satisfies
$$f(q) - f^\star \leq \sum_a q_a^\star v_a(q) \leq \max_a v_a(q). \tag{4}$$
(iv) For $q \in \Delta_n$ and $0 \leq \gamma \leq 1$,
$$f((1-\gamma)q + \gamma e_a) \leq f(q) - \gamma v_a(q) + 4\gamma^2. \tag{5}$$
(v) For $q, z \in \Delta_n$, the bound needed for output sampling is
$$|f(q) - f(z)| \leq 4\|q - z\|_1. \tag{20}$$

Proof. **Proof of (i).** The inequalities $|(Zw)_i| \leq \|w\|_1$, $|(Zq)_i - y_i| \leq 2$, and $\|q - e_a\|_1 \leq 2$ give the basic bounds.

Proof of (ii). Expanding the squared norm gives the quadratic identity.

Proof of (iii). Compactness of Δ_n and continuity of f give a minimizer. Expanding at q yields
$$f(q) - f^\star = \sum_a q_a^\star v_a(q) - \|q^\star - q\|_Z^2.$$
The sum is an average, proving both gap inequalities.

Proof of (iv). Taking displacement $\gamma(e_a - q)$ instead bounds the quadratic remainder by $4\gamma^2$.

Proof of (v). Integrate ∇f along the segment from z to q, which stays in Δ_n, to obtain Equation (20). □

A.2 Quantum primitives

Finite-circuit convention. We use an inverse-closed finite universal gate set with exact reversible Boolean gates and controlled versions of its one-qubit generators. Finite-word arithmetic and

lookups are reversible (Bennett, 1973; Toffoli, 1980). Constant-qubit gates can be synthesized to operator error τ with $\mathrm{polylog}(1/\tau)$ overhead (Dawson & Nielsen, 2006). For a sequence of M unitary operations, the bound

$$\|\widehat{V}_M \cdots \widehat{V}_1 - V_M \cdots V_1\| \le \sum_{\ell=1}^{M} \|\widehat{V}_\ell - V_\ell\|$$

controls the accumulated approximation error. An error allowance τ per operation gives total error at most $M\tau$. To achieve total synthesis error ϵ_{syn}, take $\tau = \epsilon_{\mathrm{syn}}/M$, giving $\mathrm{polylog}(M/\epsilon_{\mathrm{syn}})$ synthesis overhead per constant-qubit operation. Controlled rotations are approximated on the target, preserving control registers exactly. Amplitude estimation and Grover search need a preparation circuit and its inverse. We therefore keep classical tables and synthesized gates fixed during each quantum subroutine. Each fixed finite preparation circuit defines a unitary. Reversing its gate sequence and taking adjoints gives its adjoint. Zero-state reflections include the registers initialized by that preparation and exclude read-only memory and external controlling labels.

Theorem A.2 (Amplitude estimation and bounded means)**.** *Let a fixed unitary U set its flag to one with probability p. For every integer $M \ge 1$, amplitude estimation uses at most M calls each to U and $U^\dagger$ and returns $\widetilde{p}$ with*

$$|\widetilde{p} - p| \le \frac{2\pi\sqrt{p(1-p)}}{M} + \frac{\pi^2}{M^2} \quad \textit{with probability at least } 8/\pi^2.$$

Consequently, a finite bounded random variable $|Y| \le A$, with input-query-free sample preparation (no calls to O_X, O_y) and $O(1)$-query reversible evaluation, admits a mean estimate of accuracy $\zeta > 0$ and failure probability $0 < \eta < 1/2$ using

$$O\big((A/\zeta)\log(2/\eta)\big) \tag{21}$$

queries and preparation calls when $A > \zeta$. Zero suffices otherwise. The estimate may be measured or retained coherently, preserving a controlling atom label and all workspace. Finite output rounding and finite auxiliary gates preserve the complexity bound in Equation (21)*. For a fixed finite sample circuit with flag probability p, the target is the encoded mean $2Ap - A$. We account separately for its deterministic bias from the desired mean.*

Proof. **Mean estimation.** The amplitude-estimation bound is Brassard et al. (2002, Theorem 12). For bounded means, apply Montanaro (2015, Theorem 2.3) to $(Y + A)/(2A)$ and rescale. Flag-probability accuracy $\zeta/(2A)$ gives mean accuracy ζ and Equation (21). If $A \le \zeta$, return zero.

Finite implementation. For a fixed finite sample circuit, this argument targets its actual flag probability. Controlled Grover powers use $U, U^\dagger$ and controlled reflections. When the control is zero, the two sample-circuit calls cancel. Run each trial at accuracy $\zeta/2$ and round its output within $\zeta/2$. Synthesize its auxiliary gates to total operator error at most $1/32$. The telescoping bound requires only logarithmic precision in the trial length, and its success probability remains at least $8/\pi^2 - 1/32 > 3/4$.

Confidence and coherent output. A median of $O(\log(2/\eta))$ independent trials then gives the stated failure probability. For coherent output, defer measurement and compute the median reversibly, retaining all trial workspaces. Higher confidence requires more trials at the same synthesis accuracy. This controls estimation failure for the implemented circuit's actual mean. The remaining deterministic bias relative to the desired mean is bounded in Proposition 6.2. □

The next lemma estimates violations and loss from coordinate queries. Violation estimates support threshold tests and independent atom checks. Loss estimates verify the final sparse output.

Lemma A.3 (Sample circuits and estimators)**.** *Let I be uniform on $[N]$ and let A_1, A_2 be independent samples from $q \in \Delta_n$, independent of I. The following properties hold.*

(i) The violation and loss satisfy

$$v_a(q) = 2\mathbb{E}[(Z_{IA_1} - Z_{Ia})(Z_{IA_2} - y_I)], \tag{22}$$

$$f(q) = \mathbb{E}[(Z_{IA_1} - y_I)(Z_{IA_2} - y_I)]. \tag{23}$$

(ii) Given input-query-free preparations of purifications of these laws, the corresponding sample circuits have flag probabilities $(v_a(q) + 8)/16$ and $(f(q) + 4)/8$, using at most eight input queries per use or inverse use. The first circuit is block diagonal in a controlling atom a.

(iii) Under the preparation assumptions in (ii), either quantity is estimable to accuracy ζ with $O(\zeta^{-1}\log(2/\eta))$ queries. With finite preparations the target and bias are interpreted as in Theorem A.2.

(iv) For finite implementation and $0 < \tau \le 1/2$, uniform row ($s = N$) and auxiliary atom ($s = n$) states can be prepared without input queries to state error τ using $O(\log(s/\tau))$-bit arithmetic and $\mathrm{polylog}(s/\tau)$ gates, via interval counts with empty branches excluded exactly.

Proof. **Proof of (i).** On the sample indices and labels $S = (I, A_1, A_2, T)$, the input-query-free preparation is

$$P|0\rangle_S = |\Psi_q\rangle := N^{-1/2}\sum_{i,b,c}\sqrt{q_b q_c}|i,b,c\rangle|\tau_{ibc}\rangle, \qquad \|\tau_{ibc}\| = 1.$$

For fixed a, define the sample values and diagonal observables (with identity action on T):

$$h_v(i,b,c) := (Z_{ib} - Z_{ia})(Z_{ic} - y_i), \qquad h_f(i,b,c) := (Z_{ib} - y_i)(Z_{ic} - y_i),$$
$$D_t := \textstyle\sum_{i,b,c} h_t(i,b,c)|i,b,c\rangle\langle i,b,c| \otimes I_T, \qquad t \in \{v, f\}.$$

Independence and $\sum_b q_b = 1$ give

$$\begin{aligned}\langle\Psi_q|D_v|\Psi_q\rangle &= N^{-1}\textstyle\sum_{i,b,c} q_b q_c h_v(i,b,c)\\ &= N^{-1}\textstyle\sum_i((Zq)_i - Z_{ia})((Zq)_i - y_i) = v_a(q)/2,\\ \langle\Psi_q|D_f|\Psi_q\rangle &= N^{-1}\textstyle\sum_{i,b,c} q_b q_c h_f(i,b,c) = N^{-1}\sum_i((Zq)_i - y_i)^2 = f(q).\end{aligned}$$

Proof of (ii). Let $E_{t,a}$ be an arithmetic unitary that preserves label S and computes h_t into H using four/three input queries for v/f. Write B for its workspace (including H) and F for the flag qubit. Here, $|h\rangle_H$ denotes the computational basis state encoding the sample value $h = h_t(i,b,c)$. Since $|h_t| \le 4$, define the controlled rotation (extended to a unitary on H, F)

$$\mathsf{R}|h\rangle_H|0\rangle_F = |h\rangle_H\big(\sqrt{(4-h)/8}|0\rangle_F + \sqrt{(4+h)/8}|1\rangle_F\big).$$

With $\Pi_F = I_{S,B} \otimes |1\rangle\langle 1|_F$, the sample unitary and flag probabilities are

$$U_{t,a} := E_{t,a}^\dagger \mathsf{R} E_{t,a}(P \otimes I_{B,F}), \tag{24}$$
$$p_t := \langle 0|U_{t,a}^\dagger \Pi_F U_{t,a}|0\rangle = \langle\Psi_q|(D_t + 4I)/8|\Psi_q\rangle.$$

Thus $p_v = (v_a(q) + 8)/16$ and $p_f = (f(q) + 4)/8$. Uncomputing with $E_{t,a}^\dagger$ resets B, giving eight/six queries per use or inverse use. The extension $U_t = \sum_a |a\rangle\langle a| \otimes U_{t,a}$ preserves a. The inner zero-state reflection excludes this label.

Proof of (iii). Apply Theorem A.2, with finite bias treated in Proposition 6.2.

Proof of (iv). The stated bounds for uniform index preparation follow by applying the recursive construction of Grover & Rudolph (2002, Eqs. (2)–(6)) and the gate synthesis of Dawson & Nielsen (2006, Theorem 1) under the finite-circuit convention in Appendix A.2. Empty branches are skipped, and deterministic branches use exact Boolean gates to exclude invalid indices. □

The next lemma separates two effects of Grover iterations: the probability of reaching the good subspace and the conditional state within that subspace. The latter determines which atom is sampled when the flag is one.

Lemma A.4 (Grover iterations with a random count)**.** *Let W be a unitary and Π a projector onto a good subspace, with good probability $\mu := \|\Pi W|0\rangle\|^2$ and Grover angle $\phi := \arcsin\sqrt{\mu} \in [0, \pi/2]$. Let $\mathsf{Grov} := -W(I - 2|0\rangle\langle 0|)W^\dagger(I - 2\Pi)$.*

(i) For integer $j \ge 0$, $\mathsf{Grov}^j W|0\rangle$ has good probability $\sin^2((2j+1)\phi)$ and, when nonzero, normalized good component $\pm\Pi W|0\rangle/\|\Pi W|0\rangle\|$, independently of j up to the sign of $\sin((2j+1)\phi)$.

(ii) For integer $R \ge 1$ and j uniform on $\{0, \dots, R-1\}$, the average good probability is

$$s_R(\mu) = \frac{1}{R}\sum_{j=0}^{R-1}\sin^2\big((2j+1)\phi\big) \;\ge\; \frac{1}{4}\min\{1, R^2\mu\}. \tag{25}$$

Proof. **Proof of (i).** The statement, including the common normalized good direction, follows from Brassard et al. (2002, Eq. (8)).

Proof of (ii). Outside the immediate cases $R = 1$ and $\mu \in \{0, 1\}$, Boyer et al. (1998, Lemma 2) gives

$$s_R(\mu) = \frac{1}{2} - \frac{\sin(4R\phi)}{4R\sin(2\phi)}.$$

We prove the lower bound in (ii) in three cases, using $\sin x \geq 2x/\pi$ for $0 \leq x \leq \pi/2$, which follows from concavity of the sine function on this interval.

Case 1: $R\phi \leq \pi/4$**.** Every angle $(2j+1)\phi$, $0 \leq j < R$, lies in $[0, \pi/2]$. Applying the sine bound to each term gives

$$s_R(\mu) \geq \frac{4\phi^2}{\pi^2 R} \sum_{j=0}^{R-1} (2j+1)^2 = \frac{4(4R^2-1)}{3\pi^2}\phi^2$$
$$\geq \frac{4}{\pi^2} R^2\mu \geq \frac{1}{4} R^2\mu.$$

The equality uses $\sum_{j=0}^{R-1}(2j+1)^2 = R(4R^2-1)/3$. The last line uses $4R^2 - 1 \geq 3R^2$, $\phi^2 \geq \sin^2\phi = \mu$, and $4/\pi^2 > 1/4$.

Case 2: $R\phi > \pi/4$ **and** $\phi \leq \pi/4$**.** Here $2\phi \in [0, \pi/2]$, so the same sine bound gives

$$R\sin(2\phi) \geq \frac{4R\phi}{\pi} > 1.$$

Thus the $s_R(\mu) \geq 1/4$ conclusion of Boyer et al. (1998, Lemma 2) applies.

Case 3: $\phi > \pi/4$**.** Set $\psi := \pi/2 - \phi \in (0, \pi/4)$. Since R is an integer, $\sin(4R\phi) = -\sin(4R\psi)$. Substituting into the average-success formula gives

$$s_R(\mu) = \frac{1}{2} + \frac{\sin(4R\psi)}{4R\sin(2\psi)}.$$

If $R\psi \leq \pi/4$, then $4R\psi \in [0, \pi]$ and the added term is nonnegative, so $s_R(\mu) \geq 1/2$. Otherwise, $R\sin(2\psi) \geq 4R\psi/\pi > 1$. Using $\sin(4R\psi) \geq -1$ then yields

$$s_R(\mu) \geq \frac{1}{2} - \frac{1}{4R\sin(2\psi)} \geq \frac{1}{4}.$$

Case 1 gives $s_R(\mu) \geq R^2\mu/4$, and Cases 2–3 give $s_R(\mu) \geq 1/4$. Together they prove $s_R(\mu) \geq \frac{1}{4}\min\{1, R^2\mu\}$. □

B Proofs for Section 5: convergence

B.1 The accurate process and deterministic updates

The next lemma keeps the KL potential well defined and bounds its initial value. In the accurate process, it also bounds the loss decrease of each accepted update and ensures loss monotonicity, as needed for the convergence analysis.

Lemma B.1 (Feasibility, initial gap, and accurate updates)**.** *Fix $\epsilon = 2^{-k} \leq 1/8$, $k \geq 3$, $0 < \delta < 1/3$, $L := \log_2(128/\epsilon) + 1$, and an optimal point $q^\star \in \Delta_n$. Start from $q_0 = u = (1/n, \ldots, 1/n)$. Before each search independently sample $r_j = 8 \cdot 2^{-j}$ uniformly over $j = 0, \ldots, L-1$. Every returned atom a is followed by $q^+ = (1-\gamma)q + \gamma e_a$, $\gamma = r/32$. A rejection leaves q unchanged. In the accurate process, every returned atom satisfies $v_a(q) \geq r/2$, and, conditional on the current history $\mathcal{F}$ and threshold r, the output probabilities satisfy, for every $H \geq \epsilon$ and every atom with $v_a(q) \geq r$,*

$$r^2\mathbb{P}(A_{\text{out}} \neq \perp \mid \mathcal{F}, r) + \frac{Hc}{\sigma_a}\mathbb{P}(A_{\text{out}} = a \mid \mathcal{F}, r) \geq \frac{1}{16}\min\{r^2, Hnc\},$$
$$\sigma_a := \frac{q_a + 1/n}{2}, \quad c := \min\{r, 1/n\}. \tag{10}$$

For analysis define

$$\rho := 2^{-25},$$
$$D(q) := \sum_{a:\, q_a^\star > 0} q_a^\star \log\frac{q_a^\star}{q_a}, \qquad \mathcal{W}(q) := (f(q) - f^\star + \epsilon)e^{\rho D(q)}. \tag{11}$$

Then the following properties hold.

(i) Every iterate has positive coordinates and lies in Δ_n, even if some checks err. Hence $D(q_t)$ and $\mathcal{W}(q_t)$ are finite.
(ii) The initial point satisfies $f(u) - f^\star \le 1$ and $D(u) \le \log n$.
(iii) For an accurate update,

$$f(q) - f(q^+) \ge \frac{3r^2}{256}. \tag{26}$$

(iv) The suboptimality gap is nonincreasing and satisfies $0 \le f(q_t) - f^\star \le 1$ throughout the accurate process.

Proof. **Proof of (i).** The smallest threshold is $r_{L-1} = \epsilon/16$ and the largest is 8, so $0 < \gamma \le 1/4$. Starting from the positive point $q_0 = u$, each update preserves positivity since $q_b^+ \ge (1-\gamma)q_b \ge 3q_b/4 > 0$, and the coordinates still sum to one. These properties hold for any selected atom, independently of whether its check is accurate. A rejection changes nothing. Positivity makes $D(q_t)$ and $\mathcal{W}(q_t)$ finite at every finite time. Nonnegativity of D is the standard KL divergence inequality.

Proof of (ii). The signed columns cancel at u, so $f(u) - f^\star \le f(u) = \|y\|_2^2/N \le 1$. Also

$$D(u) = \log n + \sum_{a:\, q_a^\star > 0} q_a^\star \log q_a^\star \le \log n.$$

Proof of (iii). For an accurate update, the quadratic descent bound Equation (5) gives

$$f(q) - f(q^+) \ge \gamma v_a(q) - 4\gamma^2 \ge \frac{\gamma r}{2} - 4\gamma^2 = \frac{3r^2}{256}.$$

Proof of (iv). By (iii), accurate accepted steps decrease f, while rejections leave it unchanged. The gap therefore never increases and is at most one by (ii). Its nonnegativity follows from optimality of $q^\star$. □

B.2 Relative potential change and expected contraction

Lemma 5.1 (One accepted step). *Let $q \in \Delta_n$ have positive coordinates, $0 < r \le 8$, $\gamma := r/32$, and $q^+ := (1-\gamma)q + \gamma e_a$. Write*

$$H := f(q) - f^\star + \epsilon, \qquad \ell_a := \log\Big(1 + \frac{\gamma}{(1-\gamma)q_a}\Big), \qquad \bar{\ell}_a := \min\{\ell_a, 1\}.$$

Then the following statements hold.

(i) The KL divergence satisfies

$$D(q^+) - D(q) = -\log(1-\gamma) - q_a^\star \ell_a \le 2\gamma. \tag{12}$$

(ii) If $v_a(q) \ge r/2$, then, with $\sigma_a := (q_a + 1/n)/2$ and $c := \min\{r, 1/n\}$,

$$\begin{aligned} 1 - \frac{\mathcal{W}(q^+)}{\mathcal{W}(q)} &\ge \frac{3r^2}{512H} + \frac{\rho}{4} q_a^\star \bar{\ell}_a - \frac{\rho r}{8}, \\ \sigma_a \bar{\ell}_a &\ge \frac{c}{64}. \end{aligned} \tag{13}$$

Proof. **Proof of (i).** For $b \ne a$, $q_b^+ = (1-\gamma)q_b$, whereas $q_a^+ = (1-\gamma)q_a(1 + \gamma/((1-\gamma)q_a))$. Substitution into D proves the exact KL difference. Since $-\log(1-\gamma) \le 2\gamma$ for $\gamma \le 1/4$ and $\ell_a \ge 0$, the increase bound holds for every atom.

Proof of (ii). For an accurate update, Equation (26) gives $f(q) - f(q^+) \ge 3r^2/256$. Also $0 \le f(q) - f(q^+) \le f(q) - f^\star < H$, since $f(q^+) - f^\star \ge 0$. The exact relative change is

$$\frac{\mathcal{W}(q^+)}{\mathcal{W}(q)} = \left(1 - \frac{f(q) - f(q^+)}{H}\right)(1-\gamma)^{-\rho} e^{-\rho q_a^\star \ell_a}. \tag{27}$$

We bound its last two factors. First, $0 \le -\rho\log(1-\gamma) \le 2\rho\gamma \le \rho/2 < 1$, so $e^x \le 1 + 2x$ for $0 \le x \le 1$ gives $(1-\gamma)^{-\rho} \le 1 + 4\rho\gamma$. Second, $0 \le \rho q_a^\star \bar{\ell}_a \le \rho < 1$ and $e^{-x} \le 1 - x/2$ for $0 \le x \le 1$, whence

$$e^{-\rho q_a^\star \ell_a} \le e^{-\rho q_a^\star \bar{\ell}_a} \le 1 - \frac{\rho}{2} q_a^\star \bar{\ell}_a.$$

For $x := (f(q) - f(q^+))/H \in [0,1]$ and $y := \rho q_a^\star \bar{\ell}_a/2 \in [0,1]$, Equation (27) and $x + y - xy \geq (x+y)/2$ give

$$1 - \frac{\mathcal{W}(q^+)}{\mathcal{W}(q)} \geq x + y - xy - 4\rho\gamma(1-x)(1-y)$$
$$\geq \frac{x+y}{2} - 4\rho\gamma \geq \frac{3r^2}{512H} + \frac{\rho}{4} q_a^\star \bar{\ell}_a - \frac{\rho r}{8}.$$

The clipping ensures that the exponential bound is valid even when q_a is very small and ℓ_a is large.

Finally, $\log(1+t) \geq t/(1+t)$ gives

$$\ell_a \geq \frac{\gamma}{(1-\gamma)q_a + \gamma} \geq \frac{\gamma}{q_a + \gamma}.$$

The last expression lies in $[0,1]$, so the same lower bound holds for $\bar{\ell}_a$. Therefore

$$\sigma_a \bar{\ell}_a \geq \frac{1}{2} \frac{(q_a + 1/n)\gamma}{q_a + \gamma} \geq \frac{1}{2} \min\{\gamma, 1/n\} \geq \frac{1}{64} \min\{r, 1/n\} = \frac{c}{64},$$

which completes the proof. □

Lemma 5.2 (Expected potential contraction per search). *Under the analysis conditions above, fix the history $\mathcal{F}$ before sampling the next threshold, and let q^+ be the point after the search and any resulting update. If $f(q) - f^\star \geq \epsilon/2$, then*

$$\mathbb{E}[\mathcal{W}(q^+) \mid \mathcal{F}] \leq \left(1 - \frac{2^{-45}\epsilon}{L}\right) \mathcal{W}(q). \tag{14}$$

Here $L = \log_2(128/\epsilon) + 1$ counts the number of different thresholds. The expectation is over the uniformly sampled threshold and all search outcomes, including empty returns.

Proof. Conditioning on $\mathcal{F}$ fixes q, its positive potential, and $H = f(q) - f^\star + \epsilon$. By Lemma B.1, $\epsilon \leq H \leq 1 + \epsilon$ and, under the stated gap condition, $H \leq 3\big(f(q) - f^\star\big)$. On rejection, $q^+ = q$, so the relative potential decrease is zero.

We first derive a bound for a fixed threshold $r \geq 64\rho H$. Its negative term in Equation (13) obeys $\rho r/8 \leq r^2/(512H)$. Thus every returned atom satisfies

$$1 - \frac{\mathcal{W}(q^+)}{\mathcal{W}(q)} \geq \frac{r^2}{256H} + \frac{\rho}{4} q_a^\star \bar{\ell}_a \geq 0.$$

Write $c := \min\{r, 1/n\}$ and $\mathcal{V}_r := \{a : v_a(q) \geq r\}$. Here $q^\star(\mathcal{V}_r) = \sum_{a \in \mathcal{V}_r} q_a^\star$ is the total weight that the fixed optimal distribution $q^\star$ assigns to atoms whose current violation is at least r. Averaging over outputs and keeping divergence terms only from $\mathcal{V}_r$ gives

$$\mathbb{E}\left[1 - \frac{\mathcal{W}(q^+)}{\mathcal{W}(q)} \,\middle|\, \mathcal{F}, r\right] \geq \frac{r^2 \mathbb{P}(A_{\text{out}} \neq \perp \mid \mathcal{F}, r)}{256H} + \frac{\rho c}{256} \sum_{a \in \mathcal{V}_r} q_a^\star \frac{\mathbb{P}(A_{\text{out}} = a \mid \mathcal{F}, r)}{\sigma_a}$$
$$\geq \frac{\rho}{256H} \sum_{a \in \mathcal{V}_r} q_a^\star \left(r^2 \mathbb{P}(A_{\text{out}} \neq \perp \mid \mathcal{F}, r) + \frac{Hc}{\sigma_a} \mathbb{P}(A_{\text{out}} = a \mid \mathcal{F}, r) \right)$$
$$\geq \frac{\rho}{4096H} \min\{r^2, Hnc\}\, q^\star(\mathcal{V}_r).$$

The first inequality uses $\sigma_a \bar{\ell}_a \geq c/64$ from Lemma 5.1. The second uses $\rho q^\star(\mathcal{V}_r) \leq 1$. The last is exactly Equation (10) with this H. This expectation includes empty returns. The first term accounts for loss decrease whenever an atom is returned. The second accounts for KL gains from qualifying atoms.

We now average over the L uniformly sampled thresholds. Large thresholds provide the required contraction, intermediate thresholds contribute a nonnegative amount, and the total possible increase at small thresholds is bounded by a geometric tail. We treat these three ranges separately.

Case 1: $r_j \geq H/16$. These thresholds satisfy $r_j \geq 64\rho H$, since $64\rho \leq 1/16$. For $r = r_j$, the minimum in the fixed-threshold bound satisfies

$$\min\{r_j^2, Hn \min\{r_j, 1/n\}\} = \min\{r_j^2, H, Hnr_j\} \geq Hr_j/16,$$

using respectively $r_j \geq H/16$, $r_j \leq 8$, and $n \geq 2$ to bound its three terms. Hence

$$\mathbb{E}\left[1 - \frac{\mathcal{W}(q^+)}{\mathcal{W}(q)} \,\middle|\, \mathcal{F}, r_j\right] \geq \frac{\rho r_j}{2^{16}} q^\star(\mathcal{V}_{r_j}).$$

If $v_a(q) > H/8$, the largest threshold not exceeding $v_a(q)$ exists because $H \geq \epsilon$ and $v_a(q) \leq 8$ by Equation (2). Successive thresholds differ by a factor of two, so this threshold is greater than $v_a(q)/2 > H/16$. If $v_a(q) \leq H/8$, the right-hand side below is zero. Thus, for every atom a,

$$\sum_{j:\, H/16 \leq r_j \leq v_a(q)} r_j \geq \frac{1}{2}\big(v_a(q) - H/8\big)_+.$$

Multiplying by $q_a^\star$ and summing over atoms gives

$$\begin{aligned}\sum_{j:\, r_j \geq H/16} r_j q^\star(\mathcal{V}_{r_j}) &\geq \frac{1}{2}\sum_a q_a^\star \big(v_a(q) - H/8\big)_+ \\ &\geq \frac{1}{2}\big(f(q) - f^\star - H/8\big) \geq \frac{f(q) - f^\star}{4}.\end{aligned}$$

The second inequality uses the suboptimality gap bound Equation (4) and $\sum_a q_a^\star = 1$. The last uses $H \leq 3(f(q) - f^\star)$. Consequently this case contributes at least

$$\begin{aligned}\frac{1}{L}\sum_{j:\, r_j \geq H/16} \mathbb{E}\left[1 - \frac{\mathcal{W}(q^+)}{\mathcal{W}(q)} \,\middle|\, \mathcal{F}, r_j\right] &\geq \frac{\rho}{2^{16} L} \sum_{j:\, r_j \geq H/16} r_j q^\star(\mathcal{V}_{r_j}) \\ &\geq \frac{\rho(f(q) - f^\star)}{2^{18} L}.\end{aligned}$$

Case 2: $64\rho H \leq r_j < H/16$**.** Since $r_j \geq 64\rho H$, we have $\rho r_j/8 \leq r_j^2/(512H)$. Thus Equation (13) gives, for every returned atom a,

$$\begin{aligned}1 - \frac{\mathcal{W}(q^+)}{\mathcal{W}(q)} &\geq \frac{3r_j^2}{512H} + \frac{\rho}{4} q_a^\star \bar{\ell}_a - \frac{\rho r_j}{8} \\ &\geq \frac{r_j^2}{256H} + \frac{\rho}{4} q_a^\star \bar{\ell}_a \geq 0.\end{aligned}$$

The last inequality uses $q_a^\star, \bar{\ell}_a \geq 0$. On rejection, $q^+ = q$. Hence the expected relative potential decrease is nonnegative for each threshold in this range, as is its contribution to the average.

Case 3: $r_j < 64\rho H$**.** By Equation (13), $1 - \mathcal{W}(q^+)/\mathcal{W}(q) \geq -\rho r_j/8$ on acceptance. This bound also holds on rejection, when $q^+ = q$. If this range is nonempty, summing the geometric tail from its largest threshold, which is less than $64\rho H$, gives

$$\sum_{j:\, r_j < 64\rho H} r_j < 128\rho H.$$

The sum is zero if the range is empty. Thus this case contributes at least $-16\rho^2 H/L$ after averaging.

Combining the three cases yields

$$\begin{aligned}\mathbb{E}\left[1 - \frac{\mathcal{W}(q^+)}{\mathcal{W}(q)} \,\middle|\, \mathcal{F}\right] &\geq \frac{\rho\big(f(q) - f^\star\big)}{2^{18} L} - \frac{16\rho^2 H}{L} \\ &\geq \frac{f(q) - f^\star}{L}\left(\frac{\rho}{2^{18}} - 48\rho^2\right) \\ &\geq \frac{2^{-44}\big(f(q) - f^\star\big)}{L} \geq \frac{2^{-45}\epsilon}{L}.\end{aligned}$$

For the penultimate inequality, $\rho/2^{18} - 48\rho^2 = (128 - 48)2^{-50} = 80 \cdot 2^{-50} \geq 2^{-44}$. The last uses $f(q) - f^\star \geq \epsilon/2$. Multiplying by the fixed positive $\mathcal{W}(q)$ proves conditional contraction. Individual small-threshold updates may increase $\mathcal{W}$, but gains from larger thresholds outweigh their total expected cost. Thus contraction holds after averaging over thresholds and search outcomes. □

B.3 High-probability convergence

Theorem 5.3 (Search budget for high-probability convergence). *Under the conditions of Lemma B.1, let q_t be the iterate after t searches and updates from $q_0 = u = (1/n, \ldots, 1/n)$.*

Choose

$$T := \left\lceil \frac{2^{45}L}{\epsilon} \left[\log \frac{8}{\epsilon\delta} + \rho \log n \right] \right\rceil = \widetilde{O}(\epsilon^{-1} \log(1/\delta)), \qquad \rho := 2^{-25}. \tag{28}$$

Then

$$\mathbb{P}\big(f(q_T) - f^\star \geq \epsilon/2\big) \leq \delta/4. \tag{15}$$

Proof. We track the potential until the suboptimality gap falls below $\epsilon/2$. Let $\mathcal{F}_t$ be the history after t searches and their resulting updates, and define

$$a := \frac{2^{-45}\epsilon}{L} \in (0,1), \qquad Y_t := \mathbf{1}[f(q_t) - f^\star \geq \epsilon/2]\mathcal{W}(q_t).$$

When $f(q_t) - f^\star < \epsilon/2$, the indicator sets $Y_t = 0$. Loss monotonicity from Lemma B.1 keeps the gap below $\epsilon/2$, so $Y_{t+1} = Y_t = 0$. When $f(q_t) - f^\star \geq \epsilon/2$, Lemma 5.2 applies to the current history. Since $Y_{t+1} \leq \mathcal{W}(q_{t+1})$, it gives

$$\mathbb{E}[Y_{t+1} \mid \mathcal{F}_t] \leq (1-a)\mathcal{W}(q_t) = (1-a)Y_t.$$

The bound holds in both cases. Iterating with the initial bounds in Lemma B.1 gives

$$\mathbb{E}[Y_T] \leq (1-a)^T \mathbb{E}[Y_0] \leq (1+\epsilon)n^\rho e^{-aT}.$$

On the event $f(q_T) - f^\star \geq \epsilon/2$, the nonnegativity of $D(q_T)$ gives $Y_T = (f(q_T) - f^\star + \epsilon)e^{\rho D(q_T)} \geq 3\epsilon/2$. Markov's inequality therefore yields

$$\begin{aligned}\mathbb{P}(f(q_T) - f^\star \geq \epsilon/2) &\leq \frac{2(1+\epsilon)n^\rho}{3\epsilon} e^{-aT} \leq \frac{2n^\rho}{\epsilon} e^{-aT} \\ &\leq \frac{2n^\rho}{\epsilon} \exp\left(-\log \frac{8}{\epsilon\delta} - \rho \log n \right) = \frac{\delta}{4}.\end{aligned}$$

The second inequality uses $\epsilon \leq 1/8$. The third follows from Equation (28). Finally,

$$T = O\left(\frac{L}{\epsilon} \left[\log \frac{1}{\epsilon\delta} + \log n \right] \right) = \widetilde{O}(\epsilon^{-1} \log(1/\delta)),$$

where $\widetilde{O}$ hides logarithms in n and $1/\epsilon$ but keeps the logarithmic dependence on $1/\delta$ explicit. □

C Proofs for Section 6: the quantum implementation

The simplex point has full support, but storing all of its coordinates would be costly. We instead prepare it from the Frank–Wolfe update history. Independently mark each accepted update active with its Bernoulli probability, then choose the last active update. If none is active, choose a uniform atom. This interpretation gives the conditional probabilities for a binary-tree preparation.

We first derive the exact distribution, then implement its tree with a bounded circular buffer. We bound truncation and finite-precision errors before composing the resulting fixed circuits with estimation and search. The lookup–rotation–uncomputation pattern follows tree-based state preparation for Lasso (Chen & de Wolf, 2023, full version, Section 2.4). Storing subtree inactivity probabilities avoids materializing all n coordinate masses of the full-support point. Input queries follow Section 3. Working-memory conventions are specified in Appendix E.

Throughout this appendix, use

$$\nu = 2^{-20} \min\{1/n, \epsilon\}, \qquad \eta = \frac{1}{64n}. \tag{29}$$

The preparation tolerance ν serves two purposes. Its $1/n$ dependence keeps each atom's sampling probability within a constant factor of its ideal value (Equation (38)). Its ϵ dependence controls the violation and loss biases (Equation (36)). The coherent-test error η is chosen on the $1/n$ scale so that, whenever some atom has $v_a(q) \geq r$, flagged mass on atoms below $5r/8$ is at most one eighth of all flagged mass (Step 3 of Appendix C.5). These constants leave slack in the error bounds. A preparation is *input-query-free* if it calls neither O_X nor O_y. Working-memory reads, arithmetic, and rotations are charged separately. Purification labels remain throughout each coherent subroutine.

C.1 EXACT PREPARATION FROM THE UPDATE HISTORY

Proposition C.1 (Exact history preparation). *Let $\mathcal{H} = ((a_i, \gamma_i))_{i=1}^{U}$ be a fixed update sequence with $0 < \gamma_i < 1$. Starting from $q_0 = u$, define $q_i = (1-\gamma_i)q_{i-1} + \gamma_i e_{a_i}$. Its exact iterate is*

$$q_{\mathcal{H}} = c_U u + \sum_{i=1}^{U} w_i e_{a_i}, \qquad c_U = \prod_{i=1}^{U}(1-\gamma_i), \qquad w_i = \gamma_i \prod_{j=i+1}^{U}(1-\gamma_j). \tag{30}$$

An ideal input-query-free circuit prepares a purification of $q_{\mathcal{H}}$ using $O(1+\log(U+1))$ coherent tree lookup-and-rotation layers, together with a uniform-atom preparation.

Proof. **The distribution: the last active update.** Expanding the update recursion proves Equation (30). To interpret its coefficients, let B_i be independent Bernoulli variables with $\Pr(B_i = 1) = \gamma_i$, and call update i active when $B_i = 1$. If no update is active, output an independent uniform atom. Otherwise output a_J, where $J = \max\{i : B_i = 1\}$. Then

$$\Pr(\text{no active update}) = c_U, \qquad \Pr(J = i) = \gamma_i \prod_{j=i+1}^{U}(1-\gamma_j) = w_i.$$

Conditional probabilities on a logical tree in update order. Arrange all U update records as leaves of a balanced binary tree in update order, padding with empty leaves to a power of two. Thus every update in a left subtree is older than every update in the corresponding right subtree. For each node v, let I_v be the set of update indices stored in the subtree rooted at v. If v is internal, denote its left and right children by v_L and v_R, respectively. Their index sets partition I_v. Using the independent indicators B_i above, define the event

$$E_v := \{\text{there exists } i \in I_v \text{ with } B_i = 1\}.$$

Thus E_v means that the subtree rooted at v contains at least one active update. Let P_v be the probability that this subtree has no active update:

$$\begin{aligned} P_v &:= \Pr(E_v^c) = \prod_{i \in I_v}(1-\gamma_i), \\ P_v &= P_{v_L} P_{v_R} \qquad \text{for an internal node } v. \end{aligned} \tag{31}$$

In particular, $P_{v_L} = \Pr(E_{v_L}^c)$ and $P_{v_R} = \Pr(E_{v_R}^c)$ are the probabilities of no active update in the left and right subtrees. The product identity follows because the two child events depend on disjoint sets of independent indicators. If $I_v = \varnothing$, then E_v is impossible and $P_v = 1$.

On E_v, let $J_v = \max\{i \in I_v : B_i = 1\}$ be the last active update in this subtree. Choosing the right or left child means that J_v lies in I_{v_R} or I_{v_L}, respectively. Since right-subtree updates are newer, J_v lies on the right exactly when E_{v_R} occurs. It lies on the left exactly when $E_{v_L} \cap E_{v_R}^c$ occurs. Define the conditional right-branch probability p_v by

$$\begin{aligned} p_v &:= \Pr(J_v \in I_{v_R} \mid E_v) = \Pr(E_{v_R} \mid E_v) = \frac{1 - P_{v_R}}{1 - P_{v_L} P_{v_R}}, \\ 1 - p_v &= \Pr(J_v \in I_{v_L} \mid E_v) = \Pr(E_{v_L} \cap E_{v_R}^c \mid E_v) = \frac{P_{v_R}(1 - P_{v_L})}{1 - P_{v_L} P_{v_R}}. \end{aligned} \tag{32}$$

Both denominators are $\Pr(E_v) = 1 - P_v$, which is positive whenever $I_v \neq \varnothing$ because every $\gamma_i > 0$. If only one child contains records, choose that child with probability one. At an empty node, the circuit acts as the identity without evaluating a conditional probability.

The preparation operator. Let L hold the branch bit and retained path, and A the atom. Temporary work registers start at zero and are reset after each layer. We suppress them below. First rotate the branch bit with inactivity probability $c_U = P_{\text{root}}$, then prepare the uniform state $|\psi_u\rangle = n^{-1/2}\sum_a |a\rangle$ only on the tail branch:

$$\begin{aligned} |0\rangle_L |0\rangle_A &\longmapsto \big(\sqrt{c_U}\,|\text{tail}\rangle + \sqrt{1-c_U}\,|\text{root}\rangle\big)|0\rangle_A \\ &\longmapsto \sqrt{c_U}\,|\text{tail}\rangle|\psi_u\rangle + \sqrt{1-c_U}\,|\text{root}\rangle|0\rangle_A. \end{aligned}$$

The labels tail and root have different branch bits and an empty path.

On the root branch, descend the tree. A path prefix π selects $v(\pi)$. Append a fresh bit by

$$|\pi\rangle|0\rangle \longmapsto |\pi\rangle\big(\sqrt{1-p_{v(\pi)}}\,|0\rangle + \sqrt{p_{v(\pi)}}\,|1\rangle\big),$$

where 0 selects the left child and 1 the right child. At depth ℓ, the lookup-and-rotation operator is $O_{\mathcal{T},\ell}^{\dagger} C_\ell^{\dagger} \mathcal{R}_\ell C_\ell O_{\mathcal{T},\ell}$, applied from right to left: $O_{\mathcal{T},\ell}$ reads the node's child products and addresses, C_ℓ computes the rotation data, and $\mathcal{R}_\ell$ rotates the new path bit. The inverses clear the temporary work using the unchanged old prefix, retaining the new bit. Each layer acts coherently on all path prefixes and leaves the tail branch unchanged.

Repeat until reaching leaves, skipping the descent if $U = 0$. Multiplying the active-root probability $1 - c_U$ by the conditional probabilities along the path to record i cancels the intermediate factors $1 - P_v$. The remaining factors are γ_i at the leaf and the inactivity probabilities of all newer records, giving $\gamma_i \prod_{j=i+1}^{U}(1-\gamma_j) = w_i$. Thus, before reading any leaf atom, the state is

$$\sqrt{c_U}\,|\mathrm{tail}\rangle|\psi_u\rangle + \sum_{i=1}^{U} \sqrt{w_i}\,|i\rangle|0\rangle_A.$$

Finally, a leaf lookup maps $|i\rangle|0\rangle_A$ to $|i\rangle|a_i\rangle_A$ and leaves the tail branch unchanged, producing

$$|\psi_{\mathrm{full}}\rangle = \sqrt{c_U}\,|\mathrm{tail}\rangle|\psi_u\rangle + \sum_{i=1}^{U} \sqrt{w_i}\,|i\rangle|a_i\rangle.$$

The record labels retain their leaf paths and are mutually orthogonal and orthogonal to the tail label. Records sharing an atom therefore contribute probabilities that add, so measuring A gives $q_{\mathcal{H}}$. A balanced tree has depth $\lceil \log_2 \max\{1, U\} \rceil$. Completed paths are padded with identities. Each layer uses a constant number of reads of stored node data and reversible arithmetic and rotation operations, giving $O(1 + \log(U+1))$ tree layers and no input queries. □

C.2 A BOUNDED TREE FOR THE RECENT UPDATES

Proposition C.2 (Recent-history preparation and truncation error)**.** *Let $\mathcal{H}$ be as in Proposition C.1, with $\gamma_i = r_i/32$ and $r_i \in \{8, 4, \ldots, \epsilon/16\}$. With ν from Equation* (29)*, choose*

$$\gamma_{\min} = \epsilon/512, \qquad K_0 = \left\lceil \frac{512}{\epsilon} \log \frac{256}{\nu^2} \right\rceil, \qquad K = 2^{\lceil \log_2 K_0 \rceil}, \tag{33}$$
$$D_{\mathrm{tr}} = \log_2 K + \left\lceil \log_2(4 \log_2 K + 2) \right\rceil = O(\log K).$$

Store the most recent $m = \min\{U, K\}$ updates and prepare the law

$$q_{\mathcal{H}}^{(K)} = c_U^{(K)} u + \sum_{i=U-m+1}^{U} w_i e_{a_i}, \qquad c_U^{(K)} = \prod_{i=U-m+1}^{U} (1-\gamma_i). \tag{34}$$

Here $c_U^{(K)}$ is the inactivity probability for the retained updates. When $U \le K$, it equals c_U. The recent law has an ideal input-query-free preparation using $O(D_{\mathrm{tr}} + 1)$ coherent tree lookup-and-rotation layers, together with a uniform-atom preparation. The representation uses $O(K)$ classical words, costs $O(K)$ word operations to initialize, and costs $O(\log K)$ word operations per accepted update, including maintenance of the update-order cache. These guarantees hold for every accepted-update sequence, including erroneous acceptances. Moreover, $q_{\mathcal{H}}$ admits a purification $|\psi_{q_{\mathcal{H}}}\rangle$ on the same record-label and atom registers as the ideal recent-history state $|\psi_{\mathrm{rec}}\rangle$, with

$$\big\| |\psi_{q_{\mathcal{H}}}\rangle - |\psi_{\mathrm{rec}}\rangle \big\| \le \nu/8.$$

The two states can be chosen identical when $U \le K$.

Proof. **The recent-history state.** Apply Proposition C.1 to the last m updates, starting from u. Each retained record has coefficient w_i because all later updates are retained. The ideal preparation gives

$$|\psi_{\mathrm{rec}}\rangle = \sqrt{c_U^{(K)}}\,|\mathrm{tail}\rangle|\psi_u\rangle + \sum_{i=U-m+1}^{U} \sqrt{w_i}\,|i\rangle|a_i\rangle.$$

The record labels use the paths and physical slots described below. The index i identifies the record, without requiring a stored full-history index.

The data structure: physical slots and an update-order cache. A complete physical tree has K leaves. Record i occupies slot $(i-1) \bmod K$. Unused slots have factor one and a fixed valid

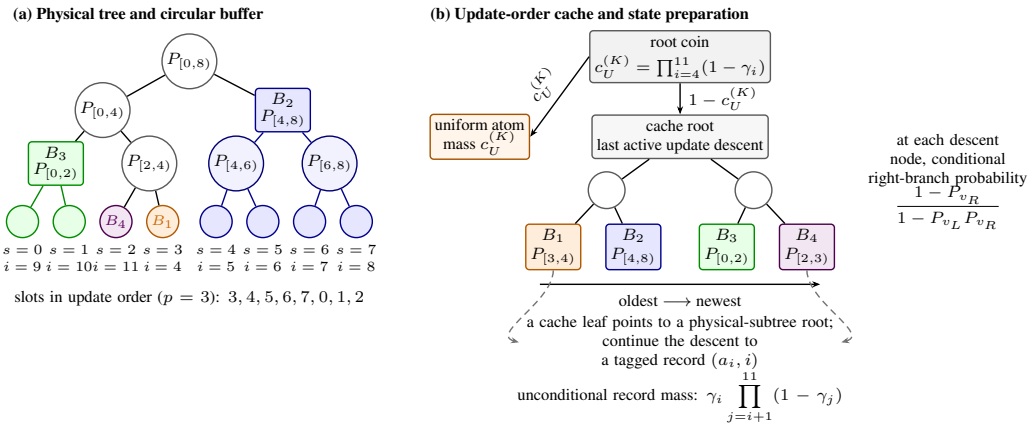

Figure 1: Recent-update preparation after wraparound ($K = 8$, $U = 11$, and $p = 3$). Physical slots $3, \ldots, 7, 0, 1, 2$ contain records $4, \ldots, 11$ in update order. The update-order cache links the canonical blocks $[3, 4)$, $[4, 8)$, $[0, 2)$, and $[2, 3)$ from oldest to newest. The root branches assign mass $c_U^{(K)}$ to a uniform atom and mass $1 - c_U^{(K)}$ to the traversal that selects the last active update. A cache leaf points to a physical-subtree root. A terminal record i has unconditional mass $\gamma_i \prod_{j=i+1}^{U}(1 - \gamma_j)$. All working data are frozen during preparation. Finite circuits use Proposition C.3.

atom label. We keep occupancy information exact. Each internal node stores the product in Equation (31) in the ideal description. Proposition C.3 specifies the finite stored values. Overwriting one slot changes only one leaf-to-root path. This makes product maintenance inexpensive. After wraparound, physical slot order differs from update order, whereas selecting the last active update requires the latter. The cache below maintains that order for traversal.

Before the buffer fills, the retained slots are $[0, U)$ in update order. Afterward, with $p = U \bmod K$, the retained slots appear in update order as $[p, K)$ followed by $[0, p)$. Decompose these intervals into canonical physical subtrees and list their roots from oldest to newest. There are at most $4 \log_2 K + 2$ such roots. A balanced cache stores their addresses and products, with empty positions having product one. Rebuilding this cache reads and combines $O(\log K)$ roots without rebuilding the physical tree. Consequently initialization costs $O(K)$ and an insertion or overwrite, including the cache, costs $O(\log K)$ word operations.

A cache leaf points to a physical subtree. Descending first through the cache and then through that subtree gives a logical tree whose leaves follow update order and whose depth is at most D_{tr}. Figure 1 shows a wraparound example. At each depth, a constant number of coherent word reads, arithmetic operations, rotations, and inverse reads implement the layer above. Completed paths are padded with identities. All stored data are frozen during a quantum subroutine. The root bit and the traversal therefore use $O(D_{\mathrm{tr}} + 1)$ lookup-and-rotation layers and no input queries.

Truncation error. Let $z = q_{U-m}$ and $|\psi_z\rangle = \sum_a \sqrt{z_a}|a\rangle$. Expanding the last m updates gives $q_{\mathcal{H}} = c_U^{(K)} z + \sum_{i=U-m+1}^{U} w_i e_{a_i}$, so we may take

$$|\psi_{q_{\mathcal{H}}}\rangle = \sqrt{c_U^{(K)}}\,|\mathrm{tail}\rangle|\psi_z\rangle + \sum_{i=U-m+1}^{U} \sqrt{w_i}\,|i\rangle|a_i\rangle.$$

If $U \le K$, then $z = u$ and the states coincide. Otherwise $c_U^{(K)} \le e^{-\gamma_{\min} K} \le \nu^2/256$ by Equation (33). Only the normalized tail states differ, so

$$\begin{aligned}\big\||\psi_{q_{\mathcal{H}}}\rangle - |\psi_{\mathrm{rec}}\rangle\big\| &= \sqrt{c_U^{(K)}}\,\big\||\psi_z\rangle - |\psi_u\rangle\big\| \\ &\le 2\sqrt{c_U^{(K)}} \le \nu/8.\end{aligned}$$

□

C.3 FINITE IMPLEMENTATION OF UPDATE-HISTORY STATE PREPARATION

We use the finite-circuit convention of Appendix A.2. Rounding the stored products changes the conditional branch probabilities and hence the prepared state. The next proposition bounds this error uniformly over all update histories, so the guarantee remains valid after each overwrite.

Proposition C.3 (Finite update-history state preparation)**.** *For each history $\mathcal{H}$ in Proposition C.2, with exact iterate q, a finite circuit implements a unitary $\widehat{U}_q$, and some purification $|\psi_q\rangle$ of q satisfies*

$$\big\|\widehat{U}_q|0\rangle - |\psi_q\rangle\big\| \le \nu.$$

It uses no input queries, $O(D_{\mathrm{tr}}+1)$ coherent tree layers, and $b = O(\log n + \log(1/\epsilon))$ fractional bits, with overhead polynomial in Appendix E's word length. The uniform-atom branch supports only valid indices. For FINDATOM*'s tagged equal mixture, the implemented atom law satisfies*

$$|\widehat{\sigma}_a - \sigma_a| \le 2\nu, \qquad \sigma_a = (q_a + 1/n)/2.$$

Proof. **Stored products.** Set $h = 2^{-b}$ and take $b \ge k+9$, so all leaf factors $1-\gamma_i$ are exact. Store empty products as one and round each other product downward after multiplying its current children. Since the cache combines disjoint physical subtrees, induction gives $0 \le P_v - \widehat{P}_v \le Kh$ at every current node. More precisely, a subtree with m occupied leaves has error at most $(m-1)_+ h$.

Overwriting a leaf recomputes the affected ancestors and cache. Unchanged nodes still represent their current leaves. Hence the error depends on K, not on the total history length U.

Branch probabilities. At a node with two occupied children, compute

$$\widehat{d} = 1 - \widehat{P}_{v_L}\widehat{P}_{v_R}, \qquad \widehat{p}_v = \left\lfloor \frac{1-\widehat{P}_{v_R}}{\widehat{d}} \right\rfloor_b,$$

using $2b$ fractional scratch bits for the product and rounding the quotient downward to b fractional bits. Downward rounding gives $\widehat{d} \ge 1 - P_{v_L}P_{v_R} \ge \gamma_{\min}$ and $0 \le 1 - \widehat{P}_{v_R} \le \widehat{d}$. The numerator and denominator errors are at most Kh, so $|\widehat{p}_v - p_v| \le 3Kh/\gamma_{\min}$, including division error.

Exact occupancy selects empty and single-child cases without division. Probabilities zero and one use identity or bit flip.

Error per tree layer. The rotation angles satisfy, for $x, y \in [0,1]$,

$$|\arcsin\sqrt{x} - \arcsin\sqrt{y}| \le 2\sqrt{|x-y|}. \tag{35}$$

First account for rounding the branch probabilities. Compare each complete lookup, rotation, and uncomputation layer on the subspace of valid paths with zero temporary registers. The induced operation, including the initial tail rotation, differs from its ideal counterpart by at most $4\sqrt{Kh/\gamma_{\min}}$ in operator norm on this subspace.

The rotations preserve the old prefix and control data, so each implementation exactly uncomputes its own temporary values. Each subsequent layer again starts with zero temporary registers. Retained path labels make these logical operations block diagonal, so the same error bound holds on a superposition of valid paths.

Precision and state error. Let τ_{unif} bound uniform-state preparation error and τ_{syn} bound the total angle-evaluation and gate-synthesis error. Adding these allowances to the layer errors gives

$$\big\|\widehat{U}_q|0\rangle - |\psi_{\mathrm{rec}}\rangle\big\| \le 4(D_{\mathrm{tr}}+1)\sqrt{\frac{K2^{-b}}{\gamma_{\min}}} + \tau_{\mathrm{unif}} + \tau_{\mathrm{syn}},$$

Choose both allowances as $\nu/8$ and

$$b \ge \max\left\{k+9, \left\lceil \log_2 \frac{256K(D_{\mathrm{tr}}+1)^2}{\gamma_{\min}\nu^2} \right\rceil\right\}.$$

The first term in the state-error bound is then at most $\nu/4$, and $b = O(\log n + \log(1/\epsilon))$. Adding the uniform-state, synthesis, and truncation allowances of $\nu/8$ each gives error at most $5\nu/8$ relative to the exact-point purification $|\psi_q\rangle$.

Mixture probabilities. For the tagged equal mixture, preparing the additional uniform branch and mixture-bit rotation each to error $\nu/8$ keeps the state error at most $7\nu/8 < \nu$. Thus each atom probability changes by at most 2ν. All preparations preserve valid atom indices. □

C.4 Preparing the point and calibrating the tests

Proposition 6.1 (State preparation)**.** *Let q be the exact iterate after any accepted-update sequence with step sizes $r_i/32$, where $r_i \in \{8, 4, \dots, \epsilon/16\}$. Use ν, η from Equation* (29) *and K from Equation* (33)*. The recent-update representation prepares, without input queries, a state $\widehat{U}_q|0\rangle$ within norm ν of some purification $|\psi_q\rangle = \sum_a \sqrt{q_a}|a\rangle|\xi_a\rangle$, with unit labels ξ_a. It uses $O(K) = \widetilde{O}(\epsilon^{-1})$ classical words, $O(K)$ initialization word operations, and $O(\log K)$ word operations per accepted update including cache maintenance. A preparation uses $O(\log K)$ tree-read and controlled-rotation layers with addresses in superposition, together with a uniform-atom preparation.*

Proof. The construction follows from Propositions C.2 and C.3. Uniform index preparations follow Lemma A.3. The tree-layer count concerns working-memory reads. The additional $O(\log n)$ uniform-index arithmetic and rotation layers are included in Appendix E's word overhead. Both truncation and numerical bounds hold for every history. □

Proposition 6.2 (Violation estimation)**.** *Let $z \in \Delta_n$, allowing zero coordinates, and suppose a fixed finite preparation uses no input queries and produces a state within norm ν of a purification $|\psi_z\rangle$ of z. Use ν, η from Equation* (29) *and $\epsilon/16 \le r \le 8$, with $0 < \beta < 1/2$. In the sample circuits of Lemma A.3, prepare two independent copies on separate registers. Give the uniform-row preparation and sample flag rotation error allowances ν each, with the rotation bound uniform over its controlling data. Preparations use only valid indices, retain their labels, and preserve a controlling atom. All circuits and stored data remain fixed during each subroutine. The means encoded by these sample circuits satisfy*

$$|\bar{v}_a - v_a(z)| \le 128\nu \le r/32, \qquad |\bar{f} - f(z)| \le 64\nu \le \epsilon/32. \tag{36}$$

(i) Coherent test. *Estimating $\bar{v}_a$ to tolerance $r/16$ gives total error below $r/8$ relative to $v_a(z)$ with failure probability at most $\eta = 1/(64n)$. The test V_r sets $F = 1$ when the estimate is at least $3r/4$. Writing $\mu_a = \Pr(F = 1 \mid a)$, we have $\mu_a \le \eta$ for $v_a(z) < 5r/8$ and $\mu_a \ge 1 - \eta$ for $v_a(z) \ge r$. Each use or inverse costs $O(r^{-1} \log n)$ input queries.*

(ii) Independent post-check. *An independent estimate of $\bar{v}_a$ with tolerance $r/32$ has total error at most $r/16$ relative to $v_a(z)$ with conditional failure probability at most β. The post-check accepts estimates at least $9r/16$ and costs $O(r^{-1} \log(2/\beta))$ input queries. Accurate checks accept all atoms with $v_a(z) \ge 5r/8$ and reject those with $v_a(z) < r/2$. Every atom accepted by an accurate check also satisfies*

$$f(z) - f\big((1 - r/32)z + (r/32)e_a\big) \ge 3r^2/256. \tag{37}$$

For the optimization loop, take $z = q$ with the preparation of Proposition 6.1.

Proof. **Finite-circuit bias.** For the sample unitaries $\widehat{U}_{v,a}, \widehat{U}_f$ implemented in Lemma A.3, the actual targets are

$$\bar{v}_a = 16\|\Pi_F \widehat{U}_{v,a}|0\rangle\|^2 - 8, \qquad \bar{f} = 8\|\Pi_F \widehat{U}_f|0\rangle\|^2 - 4.$$

State error. To bound their bias, first consider the three independent sample preparations. Write $|u_N\rangle = N^{-1/2} \sum_{i=1}^N |i\rangle$ for the ideal row state and $|\psi_z\rangle$ for the assumed purification of z, including its retained labels. Their implemented states $|\widehat{u}_N\rangle$ and $|\widehat{\psi}_z\rangle$ each have error at most ν. Replacing the three normalized tensor factors one at a time gives

$$|\Psi\rangle = |u_N\rangle \otimes |\psi_z\rangle^{\otimes 2}, \qquad |\widehat{\Psi}\rangle = |\widehat{u}_N\rangle \otimes |\widehat{\psi}_z\rangle^{\otimes 2},$$
$$\big\||\widehat{\Psi}\rangle - |\Psi\rangle\big\| \le \big\||\widehat{u}_N\rangle - |u_N\rangle\big\| + 2\big\||\widehat{\psi}_z\rangle - |\psi_z\rangle\big\| \le 3\nu.$$

All errors internal to a point preparation are included in its ν allowance. Reversible evaluation and uncomputation preserve these distances. The flag rotation, including its numerical evaluation, contributes at most one further ν. Its error bound holds for every controlling data value. The data

labels are retained and orthogonal, so the bound also holds on their superposition. Thus the ideal and implemented final sample states satisfy

$$\big\||\widehat{\Phi}\rangle - |\Phi\rangle\big\| \le 3\nu + \nu = 4\nu.$$

Flag probabilities. For either sample circuit, let $p = \|\Pi_F|\Phi\rangle\|^2$ and $\widehat{p} = \|\Pi_F|\widehat{\Phi}\rangle\|^2$ be its ideal and actual flag probabilities. Since both states are normalized and Π_F is a projector,

$$\begin{aligned}|\widehat{p} - p| &\le \big(\|\Pi_F|\widehat{\Phi}\rangle\| + \|\Pi_F|\Phi\rangle\|\big)\big\|\Pi_F(|\widehat{\Phi}\rangle - |\Phi\rangle)\big\| \\ &\le 2\big\||\widehat{\Phi}\rangle - |\Phi\rangle\big\| \le 8\nu.\end{aligned}$$

Mean bias. The affine factors 16 and 8 therefore give $|\bar{v}_a - v_a(z)| \le 128\nu$ and $|\bar{f} - f(z)| \le 64\nu$. By Equation (29) and $r \ge \epsilon/16$,

$$128\nu \le \epsilon/8192 \le r/32, \qquad 64\nu \le \epsilon/16384 \le \epsilon/32.$$

Thus the finite-circuit biases fit within the allowances in Equation (36), even at the smallest threshold. Amplitude estimation targets these actual circuit means. Adding the deterministic bias above then gives an error bound for the true violation or loss.

Proof of (i). On an accurate coherent estimate the true error is at most $r/16 + r/32 = 3r/32 < r/8$. Comparison with $3r/4$ gives both flag implications. Theorem A.2 gives the query cost with $\eta = 1/(64n)$.

Proof of (ii). On an accurate post-check the true error is at most $r/32 + r/32 = r/16$. Thus acceptance at $9r/16$ implies $v_a(z) \ge r/2$. An accurate post-check accepts whenever $v_a(z) \ge 5r/8$.

After the candidate is measured, the post-check uses fresh preparation and estimator registers. Its error bound holds conditional on the full preceding transcript, including the selected candidate. Theorem A.2 gives $O(r^{-1}\log(2/\beta))$ input queries.

With $\gamma = r/32$, Equation (5) gives

$$f(z) - f((1-\gamma)z + \gamma e_a) \ge \gamma(r/2) - 4\gamma^2 = 3r^2/256.$$

□

C.5 Implementing FindAtom and proving its guarantees

Proposition 6.3 (Guarantees of the FindAtom implementation)**.** *Fix the transcript before a call of Algorithm 2, with exact point q, $\epsilon = 2^{-k} \le 1/8$, and $r = 8 \cdot 2^{-j} \in [\epsilon/16, 8]$. Set $c = \min\{r, 1/n\}$ and $\sigma_a = (q_a + 1/n)/2$. Use preparation tolerance $\nu = 2^{-20}\min\{1/n, \epsilon\}$, ensuring $2\nu \le 1/(4n) \le \sigma_a/2$ for every atom. Use the finite preparations and tests above, the independent post-check with $0 < \beta < 1/2$, and*

$$R = R_\epsilon(r) = 2^{\left\lceil \log_2 \max\left\{1, \min\left\{\frac{r}{\sqrt{\epsilon}c}, \sqrt{n}\right\}\right\}\right\rceil}.$$

In the accurate analysis process, each returned atom satisfies $v_a(q) \ge r/2$. Let $\pi_a := \Pr(A_{\text{out}} = a \mid \mathcal{F}, r)$. For every $H \ge \epsilon$ and every atom with $v_a(q) \ge r$,

$$r^2 \Pr(A_{\text{out}} \neq \perp \mid \mathcal{F}, r) + \frac{Hc}{\sigma_a}\pi_a \ge \frac{1}{16}\min\{r^2, Hnc\}. \tag{10}$$

The algorithm and its output distribution do not depend on H. On every branch, the call uses at most $2R-1$ applications of its fixed W_r or inverse and at most one post-check, costing $O((\sqrt{n/\epsilon} + \epsilon^{-1})\log n + \epsilon^{-1}\log(2/\beta))$ input queries.

Proof. We bound two events: the search returns any atom, and it returns a particular atom whose violation reaches the threshold. The proof uses the accurate analysis process, in which measured post-checks are accurate and quantum transitions use the actual finite circuits. All probabilities below are conditional on the fixed transcript and threshold r.

Step 1: the sampling probability of each atom. Let $\widehat{\sigma}$ be the atom law of the implemented equal mixture of q and the uniform distribution, with all preparation labels retained. Proposition C.3 bounds its error by 2ν. Since $q_a \ge 0$ and $\nu = 2^{-20}\min\{1/n, \epsilon\}$,

$$\sigma_a = \frac{q_a + 1/n}{2} \ge \frac{1}{2n}, \qquad 2\nu \le \frac{2^{-19}}{n} \le \frac{1}{4n} \le \frac{\sigma_a}{2}.$$

Subtracting the error allowance from σ_a therefore gives

$$\begin{aligned} |\widehat{\sigma}_a - \sigma_a| &\le 2\nu, \\ \widehat{\sigma}_a &\ge \sigma_a - 2\nu \ge \sigma_a/2 \ge 1/(4n). \end{aligned} \tag{38}$$

In particular, every atom has positive probability under the finite preparation.

Step 2: the fixed search circuit and its budget. Apply the coherent test in fresh work registers, controlled by the atom and acting trivially on the retained preparation labels. Let μ_a be the probability that its flag is one conditional on atom a. Before amplification, the joint and total flag probabilities are

$$\Pr(a, F = 1) = \widehat{\sigma}_a \mu_a, \qquad \bar{\mu} = \Pr(F = 1) = \sum_a \widehat{\sigma}_a \mu_a.$$

Let W_r be this fixed preparation-and-test unitary. Its Grover operator is

$$G = W_r(2|0\rangle\langle 0| - I)W_r^\dagger(I - 2\Pi_F). \tag{39}$$

The zero-state reflection includes all registers initialized by W_r and excludes frozen working memory. The operator $W_r^\dagger$ is the exact inverse of the implemented circuit.

For the budget, write $t = \max\{1, \min\{r/\sqrt{\epsilon c}, \sqrt{n}\}\}$. The definition of R rounds t up to a power of two, so $t \le R < 2t$. Since $t \le \sqrt{n}$, we obtain

$$\begin{aligned} R^2 &\ge \min\{r^2/(\epsilon c), n\}, \qquad R < 2\sqrt{n}, \\ HcR^2 &\ge \min\{(H/\epsilon)r^2, Hnc\} \ge \omega_H := \min\{r^2, Hnc\}, \qquad H \ge \epsilon. \end{aligned} \tag{40}$$

The last inequality uses $H/\epsilon \ge 1$. We will use it when combining the output probabilities. The parameter H is only used for analysis.

Step 3: the flagged mass that passes the post-check. Proposition 6.2 gives total post-check error at most $r/16$. Thus acceptance at $9r/16$ implies $v_a(q) \ge 9r/16 - r/16 = r/2$. Conversely, if $v_a(q) \ge 5r/8$, every accurate post-check accepts, since $5r/8 - r/16 = 9r/16$.

These implications hold for every record in the support of K_h from Equation (19). The argument below uses no acceptance probability for atoms in $[r/2, 5r/8)$. Changing their decisions preserves the probability lower bounds, provided the two implications above are maintained. Every such accepted atom still obeys Lemma 5.1. In Lemma 5.2, its negative potential term is absorbed for $r \ge 64\rho H$ and covered by the geometric-tail bound for smaller r. Thus the convergence proof does not require a particular decision rule within this band.

Let $V = \{a : v_a(q) \ge r\}$. If V is empty, the atom-wise joint inequality has no instances. The acceptance guarantee above and the cost bound below still apply. For the probability argument, assume $V \ne \varnothing$. The coherent-test guarantees are

$$\mu_a \ge 1 - \eta \quad (a \in V), \qquad \mu_a \le \eta \quad \text{if } v_a(q) < 5r/8, \qquad \eta = 1/(64n).$$

Choose $a_0 \in V$. Since $1 - \eta \ge 1/2$, Equation (38) implies

$$\bar{\mu} \ge \widehat{\sigma}_{a_0}\mu_{a_0} \ge \frac{1}{4n} \cdot \frac{1}{2} = \frac{1}{8n}.$$

The total flagged mass on atoms below $5r/8$ is at most

$$\sum_{a : v_a(q) < 5r/8} \widehat{\sigma}_a \mu_a \le \eta \sum_{a : v_a(q) < 5r/8} \widehat{\sigma}_a \le \eta = \frac{1}{64n} \le \frac{\bar{\mu}}{8}.$$

Consequently, conditional on the flag being one, at least $7/8$ of the mass lies on atoms that pass every accurate post-check. This bound uses the actual flag distribution of the fixed circuit, including coherent-test errors. It does not condition the coherent test on being accurate.

Step 4: output probabilities after randomized amplification. Lemma A.4(i) preserves the normalized flagged state. For any count with nonzero flag probability, the conditional atom law is

$$\Pr(a \mid F = 1) = \frac{\widehat{\sigma}_a \mu_a}{\bar{\mu}}.$$

Averaging over the uniform iteration count $i \in \{0, \dots, R-1\}$ preserves the same conditional distribution. By Lemma A.4(ii), the resulting total flag probability g satisfies

$$g = s_R(\bar{\mu}) \ge \frac{1}{4} \min\{1, R^2\bar{\mu}\}.$$

At least $7/8$ of this flagged mass passes the post-check by Step 3. Moreover, each $a \in V$ always passes, and $\widehat{\sigma}_a \mu_a \geq (\sigma_a/2)(1/2) = \sigma_a/4$. Writing $x = R^2\bar{\mu} > 0$, we therefore obtain

$$\begin{aligned} \Pr(A_{\text{out}} \neq \perp) &\geq \frac{7}{8} g \geq \frac{7}{32} \min\{1, x\} \geq \frac{1}{8} \min\{1, x\}, \\ \pi_a = g \frac{\widehat{\sigma}_a \mu_a}{\bar{\mu}} &\geq \frac{\sigma_a}{16\bar{\mu}} \min\{1, x\} = \frac{R^2 \sigma_a}{16} \frac{\min\{1, x\}}{x}, \qquad a \in V. \end{aligned} \tag{41}$$

Thus the coherent-test error is already accounted for in the flagged mass and its conditional distribution throughout amplification.

Step 5: the joint guarantee. Fix any $H \geq \epsilon$ and $a \in V$. Recall $\omega_H = \min\{r^2, Hnc\}$ from Equation (40). If $x \leq 1$, the bound on the individual output probability gives

$$\frac{Hc}{\sigma_a} \pi_a \geq \frac{HcR^2}{16} \geq \frac{\omega_H}{16}.$$

If $x \geq 1$, the overall return probability gives

$$r^2 \Pr(A_{\text{out}} \neq \perp) \geq \frac{r^2}{8} \geq \frac{\omega_H}{8} \geq \frac{\omega_H}{16}.$$

In the first case, the probability of returning the individual atom gives Equation (10). In the second, the probability of returning any atom gives it. Thus one of the two nonnegative terms suffices in either case. The preparation, test, budget, and post-check do not use H, so the same output distribution satisfies the inequality simultaneously for every $H \geq \epsilon$.

Step 6: cost on every measurement branch. The initial preparation uses one call to W_r, and each application of G uses one call each to W_r and $W_r^\dagger$. Hence $i < R$ iterations use $1 + 2i \leq 2R - 1$ such calls. After measurement, the algorithm performs at most one post-check. These counts apply to every measurement branch, including branches with inaccurate post-checks.

Each use of W_r or its inverse costs $O(r^{-1} \log n)$ input queries. Its mixture preparation uses none. The measured post-check costs $O(r^{-1} \log(2/\beta))$ queries. To bound their total, use $r \geq \epsilon/16$ and $R < 2t \leq 2 + 2r/\sqrt{\epsilon c}$:

$$\frac{R}{r} < \frac{2}{r} + \frac{2}{\sqrt{\epsilon c}} \leq \frac{40}{\epsilon} + 2\sqrt{\frac{n}{\epsilon}}, \qquad \frac{1}{r} \leq \frac{16}{\epsilon}.$$

Indeed, $2/r \leq 32/\epsilon$. The other term is at most $8/\epsilon$ if $c = r$ and equals $2\sqrt{n/\epsilon}$ if $c = 1/n$. Charging the coherent calls and the single post-check separately gives

$$Q_{\text{call}}(r) = O\left(\frac{(2R-1)\log n + \log(2/\beta)}{r}\right) = O\Big((\sqrt{n/\epsilon} + \epsilon^{-1}) \log n + \epsilon^{-1} \log(2/\beta)\Big).$$

The post-check confidence factor therefore does not multiply the Grover iteration count. □

D VERIFIED SPARSE OUTPUT AND PROOF OF THE MAIN THEOREM

D.1 SPARSIFICATION BY SAMPLING AND VERIFIED OUTPUT

The loop represents the full-support iterate by its update history. For an explicit sparse vector, we sample an empirical distribution. Sampling controls the expected loss increase. Fresh candidate loss estimates are compared with one stored reference to give a high-probability guarantee.

For each sparse candidate $\widehat{q}$, store its exact integer counts in a sparse binary sum tree over the n atoms. Child-count/parent-count ratios give its sampling law. Empty subtrees are excluded exactly, and the tree has depth at most $\lceil \log_2 n \rceil$. Use the finite rotations and per-layer error allocation in the proof of Lemma A.3, with total error budget ν, and keep the counts and preparation circuit fixed throughout each loss-estimation subroutine. This gives an input-query-free preparation within norm ν of a purification of $\widehat{q}$, including when some coordinates vanish. It therefore satisfies the hypotheses of Proposition 6.2.

Proposition D.1 (Empirical sparsification and verified sparse output)**.** *Let $\epsilon = 2^{-k}$, $k \geq 3$, and $0 < \delta < 1/3$. Fix a final optimization transcript whose exact point q satisfies $f(q) - f^\star < \epsilon/2$.*

Algorithm 3: $\mathrm{SparseOutput}(q, \epsilon, \delta)$: sparsification and loss verification

Input: finite preparation of the final iterate q, oracle access to X, y, accuracy $\epsilon = 2^{-k}$, failure parameter δ

Output: either $(\perp, \text{failure})$ or $(\widehat{\theta}, \text{success})$

$M_{\text{out}} \leftarrow \lceil \log(8/\delta)/\log(5/4) \rceil$;
$\beta_{\text{ref}} \leftarrow \delta/16$; $\beta_{\text{out}} \leftarrow \delta/(16M_{\text{out}})$;
$\widetilde{f}_{\text{ref}} \leftarrow$ estimate $f(q)$ once to total error $\epsilon/16$ with failure probability at most β_{ref};
for $b = 1, \ldots, M_{\text{out}}$ **do**
  draw $16/\epsilon$ new independent atoms from the finite preparation of q, forming $\widehat{q}$; {Sparse candidate}
  estimate $f(\widehat{q})$ to total error $\epsilon/16$ with conditional failure probability at most β_{out};
  if $\widetilde{f}(\widehat{q}) - \widetilde{f}_{\text{ref}} \le 3\epsilon/8$ **then return** $(B\widehat{q}, \text{success})$; {Loss verification}
return $(\perp, \text{failure})$;

Preparation. *Use its finite preparation with ν from Equation* (29) *and actual sampling law $\widetilde{q}$. By Proposition C.3 and Equation* (20) *and the choice of ν,*

$$\|\widetilde{q} - q\|_1 \le 2\nu, \qquad f(\widetilde{q}) \le f(q) + 8\nu \le f(q) + \epsilon/32. \tag{42}$$

Sampling. *Use $s = 16/\epsilon$ fresh independent samples from $\widetilde{q}$ per candidate, with $\widehat{q}_a$ the fraction equal to atom a. Generate at most*

$$M_{\text{out}} := \left\lceil \frac{\log(8/\delta)}{\log(5/4)} \right\rceil \tag{43}$$

independent candidates.

Verification. *Use the sample circuit of Equation* (23) *with tolerance $\epsilon/32$ and the finite bias bound of Equation* (36) *at q and $\widehat{q}$. Each loss estimate has total error at most $\epsilon/16$ with the failure probability specified below.*

Before the trials, estimate $f(q)$ once with failure probability at most $\beta_{\text{ref}} = \delta/16$ and store $\widetilde{f}_{\text{ref}}$. In each trial, estimate only $f(\widehat{q})$ with failure probability at most $\beta_{\text{out}} = \delta/(16M_{\text{out}})$, conditional on the complete preceding transcript, including the stored reference and current candidate. Accept the first candidate satisfying

$$\widetilde{f}(\widehat{q}) - \widetilde{f}_{\text{ref}} \le 3\epsilon/8. \tag{44}$$

If none passes, return failure.

Output guarantees. *Algorithm 3 satisfies both guarantees below.*

(i) Conditional on the fixed optimization transcript, it returns an explicit feasible vector $\widehat{\theta} = B\widehat{q}$ with loss at most $f^\star + \epsilon$ with probability at least $1 - \delta/4$.

(ii) On every vector-output branch, $\|\widehat{\theta}\|_1 \le 1$, its support size is at most $\min\{d, 16/\epsilon\}$, and its coefficients are multiples of $\epsilon/16$.

Proof. **Proof of (i). Expected loss.** For $z \in \Delta_n$ and independent $A_1, \ldots, A_s \sim z$, $\widehat{q} = s^{-1} \sum_{i=1}^{s} e_{A_i}$ is feasible and s-sparse, and

$$\mathbb{E}[f(\widehat{q})] - f(z) = \frac{1}{s} \left(\sum_a z_a \|e_a\|_Z^2 - \|z\|_Z^2 \right) \le 1/s. \tag{45}$$

The linear term in the quadratic expansion at z has zero expectation. Independence cancels the off-diagonal quadratic terms, and $\|e_a\|_Z^2 \le 1$ bounds the variance.

Taking $z = \widetilde{q}$ and $s = 16/\epsilon$ in Equation (45), then applying Equation (42), gives

$$\begin{aligned} \mathbb{E}[f(\widehat{q}) - f^\star] &\le f(\widetilde{q}) - f^\star + \epsilon/16 \\ &\le f(q) - f^\star + \epsilon/32 + \epsilon/16 = f(q) - f^\star + 3\epsilon/32. \end{aligned}$$

Candidate quality. Markov's inequality applies to this nonnegative gap:

$$\mathbb{P}(f(\widehat{q}) - f(q) > \epsilon/4) \le \frac{f(q) - f^\star + 3\epsilon/32}{f(q) - f^\star + \epsilon/4} \le \frac{19}{24} < \frac{4}{5}. \tag{46}$$

The last inequality uses $f(q) - f^\star < \epsilon/2$ and the fact that the ratio increases with this gap.

For the analysis, generate all independent candidates in advance. The probability that none has true loss increase at most $\epsilon/4$ is then bounded by $(4/5)^{M_{\mathrm{out}}} \le \delta/8$.

Verification. The verification errors contribute a separate failure-probability budget. There is one reference estimate and at most M_{out} candidate estimates. Their conditional error guarantees and a union bound give

$$\mathbb{P}(\text{any verification estimate is inaccurate}) \le \beta_{\mathrm{ref}} + M_{\mathrm{out}}\beta_{\mathrm{out}} = \delta/16 + \delta/16 = \delta/8.$$

Reusing the reference correlates the verification decisions, but the argument requires neither independent verification errors nor independent decisions. Fresh candidate draws are independent of the reference and of earlier draws.

On the event that every estimate is accurate, a good candidate passes Equation (44), and every passing candidate satisfies $f(\widehat{q}) - f(q) \le 3\epsilon/8 + 2(\epsilon/16) = \epsilon/2$. Its excess is therefore below ϵ.

Proof of (ii). The candidate count tree has at most $\min\{n, s\}$ occupied leaves. Each insertion uses $\widetilde{O}(1)$ word tasks and no input queries. Decode by subtracting opposite-sign counts and dividing by the power of two s. This proves feasibility, sparsity, and the coefficient grid on every vector-output branch, irrespective of estimate accuracy.

Keep one candidate and one measured-estimator scratch space at a time, together with one word for the stored reference estimate. □

D.2 Proof of Theorem 4.1

Theorem 4.1 (Main theorem). *Let $\epsilon = 2^{-k}$ for an integer $k \ge 3$, and let $0 < \delta < 1/3$. In the coordinate-query model of Section 3, there is a quantum algorithm that returns either $(\perp, \mathrm{failure})$ or $(\widehat{\theta}, \mathrm{success})$, where every returned $\widehat{\theta}$ is an explicitly stored vector with $\|\widehat{\theta}\|_1 \le 1$ and at most $\min\{d, 16/\epsilon\}$ nonzero entries, and*

$$\mathbb{P}\big[\mathrm{success}\ \textit{and}\ L_{X,y}(\widehat{\theta}) \le \min_{\|\theta\|_1 \le 1} L_{X,y}(\theta) + \epsilon\big] \ge 1 - \delta.$$

On every measurement branch the algorithm terminates and uses at most

$$Q_{X,y} = \widetilde{O}\Big(\sqrt{d}\,\epsilon^{-3/2}\log(1/\delta) + \epsilon^{-2}\log^2(1/\delta)\Big) \qquad \textit{input queries.} \tag{9}$$

With the additional classically writable, coherently readable working-memory assumption of Appendix E, and suppressing confidence factors and polynomial word-length overheads, the algorithm uses $\widetilde{O}(\sqrt{d}\,\epsilon^{-3/2} + \epsilon^{-2})$ classical work (including the cost of generating quantum circuit instructions), $\widetilde{O}(\epsilon^{-1})$ classical memory words, and $\widetilde{O}(1)$ quantum workspace words. Memory excludes input storage. Quantum workspace excludes input-oracle and QRAM hardware. Detailed resource bounds are in Appendix E.

Proof. **Step 1: success of the implemented optimization loop.** Start at $q_0 = u$ and use the horizon T from Equation (28). By Proposition 6.3, the accurate analysis process satisfies the FindAtom interface. With unlimited uniform threshold-index sampling, Theorem 5.3 therefore gives $\mathbb{P}(f(q_T) - f^\star \ge \epsilon/2) \le \delta/4$ for that process.

There are at most T measured post-checks, each with conditional failure probability at most $\beta = \delta/(4T)$. By Lemma 6.4, replacing the accurate analysis process by the implemented checks adds at most $T\beta = \delta/4$ to this failure probability.

This comparison is valid on every history. Even an inaccurate acceptance uses $0 < \gamma = r/32 \le 1/4$, so the exact iterate remains feasible and positive. The finite-preparation guarantees therefore continue to apply.

Threshold sampling. To bound the number of random bits on every execution path, implement the uniform threshold draw in line 4 of Algorithm 1 by bounded rejection sampling. Set $b_{\mathrm{idx}} = \lceil \log_2 L \rceil$ and $m = \lceil \log_2(4T/\delta) \rceil$.

For each threshold draw, generate b_{idx} independent fair bits and interpret them as $z \in \{0, \dots, 2^{b_{\mathrm{idx}}} - 1\}$. If $z < L$, accept $j = z$ and return r_j from Equation (7); otherwise, repeat with fresh bits. If all m trials are rejected, the algorithm returns $(\perp, \mathrm{failure})$.

Each valid index has the same probability on every trial, so conditional on acceptance the index is uniform on $\{0, \dots, L-1\}$. Each trial accepts with probability at least $1/2$, so the probability of any sampling abort over the T calls is at most $T2^{-m} \le \delta/4$.

Couple the bounded and unlimited samplers until the first such abort. This adds the abort allowance without conditioning the uniform threshold law used above. The Grover count is sampled exactly with $\log_2 R$ fair bits because R is a power of two.

Combining the three allowances gives

$$\mathbb{P}(\text{optimization aborts or finishes with } f(q_T) - f^\star \ge \epsilon/2) \le \frac{\delta}{4} + T\beta + T2^{-m} \le \frac{3\delta}{4}. \tag{47}$$

Step 2: verified output and the total failure probability. Fix any final optimization transcript on which no abort occurred and $f(q_T) - f^\star < \epsilon/2$. Proposition D.1 applies to the actual finite sampling law at this point. Conditional on this transcript, output fails to return an ϵ-optimal vector with probability at most $\delta/4$. Averaging over all such transcripts and using Equation (47) gives

$$\mathbb{P}(\text{failure or a returned vector with excess loss} > \epsilon) \le \frac{3\delta}{4} + \frac{\delta}{4} = \delta.$$

No independence between optimization and output failures is needed.

By Lemma 3.1, the lifted loss and optimum are exactly those of the original Lasso problem. By Proposition D.1, every returned vector is explicitly stored, has ℓ_1 norm at most one, and has at most $\min\{d, 16/\epsilon\}$ nonzero entries, including branches with inaccurate estimates.

Step 3: resources and termination. Proposition 6.3 bounds each of the at most T search calls by

$$O\Big((\sqrt{n/\epsilon} + \epsilon^{-1}) \log n + \epsilon^{-1} \log(2/\beta)\Big)$$

input queries. Here $T = \widetilde{O}(\epsilon^{-1} \log(1/\delta))$ and $\log(2/\beta) = \widetilde{O}(\log(1/\delta))$. Threshold-index sampling and candidate sampling use no input queries.

There are at most $M_{\text{out}} + 1 = O(\log(1/\delta))$ output loss estimates. The reference costs $O(\epsilon^{-1} \log(2/\beta_{\text{ref}}))$ input queries. Each candidate estimate costs $O(\epsilon^{-1} \log(2/\beta_{\text{out}}))$.

Multiplying these deterministic budgets and using $n = 2d$ gives

$$Q_{X,y} = \widetilde{O}\Big(\sqrt{d}\, \epsilon^{-3/2} \log(1/\delta) + \epsilon^{-2} \log^2(1/\delta)\Big).$$

Proposition E.1 and Corollary E.2 justify this query accounting without native QRAM. Under the additional working-memory assumption, Proposition E.1 also gives the stated classical work, classical memory, and quantum workspace bounds.

Termination. Finally, T bounds the optimization calls, m bounds each threshold sampler, and M_{out} bounds the output trials. Every quantum subroutine also has a fixed finite budget. Thus the algorithm terminates on every measurement branch. □

E Resource accounting

E.1 Queries, gates, classical work, and storage

Finite words. The finite-circuit convention in Appendix A.2 and the sampler precision of Proposition C.3 allow all values, addresses, and counters to fit in words of length

$$w = b_{\text{in}} + O(\log(Nn/\epsilon) + \log(2 + \log(1/\delta))). \tag{48}$$

The row and atom preparations use $O(\log(Nn/\epsilon))$ bits, and amplitude estimation increases confidence by repeating trials at fixed precision. Binary confidence exponents and horizon and repetition counters require only the additional $O(\log(2 + \log(1/\delta)))$ bits.

Proposition E.1 (Resources on every measurement branch)**.** *Let $\epsilon = 2^{-k}$ with $k \ge 3$ and $0 < \delta < 1/3$. The input has N examples, d features, and b_{in} bits per entry. Set $n = 2d$.*

Preparation. *Use T from Theorem 5.3, the recent-update preparation of Propositions C.2 and C.3, and ν, K from Equations* (29) *and* (33)*. The prepared law is Equation* (34)*, while the exact iterate is Equation* (30)*.*

Access and costs. *Use the coordinate-query model of Section 3. For auxiliary circuit bounds, assume efficient writable QRAM providing classically writable, coherently readable working memory. Its contents change only between circuits and stay fixed during each circuit. Classical word writes and coherent reads of this memory have unit cost. Input-oracle calls are charged separately. Word arithmetic and finite-gate synthesis cost* $\mathrm{poly}(w)$*, with* w *from Equation* (48)*.*

The algorithm terminates on every measurement branch and satisfies the following deterministic bounds.

(i) ***Input queries.*** *The number of input-oracle calls or inverses is*

$$Q_{X,y} = \widetilde{O}\Big(\sqrt{d}\,\epsilon^{-3/2}\log(1/\delta) + \epsilon^{-2}\log^2(1/\delta)\Big)\,.$$

(ii) ***Auxiliary quantum work.*** *Nonquery quantum gates and coherent working-memory reads total* $G_{\mathrm{aux}} = \widetilde{O}(Q_{X,y})\,\mathrm{poly}(w)$*. This count excludes gates inside the input oracles.*

(iii) ***Classical work.*** *Data-structure maintenance, random controls, and processing measured estimates and candidates cost* $\widetilde{O}(\epsilon^{-1}\log^2(1/\delta))\,\mathrm{poly}(w)$*. Including generation and issuance of quantum circuit instructions, total classical work is* $\widetilde{O}(Q_{X,y})\,\mathrm{poly}(w)$*.*

(iv) ***Storage.*** *Classical working memory uses* $\widetilde{O}(\epsilon^{-1} + \log(1/\delta))$ *words, and quantum workspace uses* $\widetilde{O}(1)$ *words. Their total is* $S_{\mathrm{aux}} = \widetilde{O}(\epsilon^{-1} + \log(1/\delta))$ *words. These bounds exclude input storage and omit polynomial word-length overheads. Quantum workspace also excludes input-oracle and QRAM hardware.*

Proof. **Common budgets.** The horizon and post-check budget give

$$T = \widetilde{O}(\epsilon^{-1}\log(1/\delta)), \qquad \beta = \frac{\delta}{4T}, \qquad \log(2/\beta) = \widetilde{O}(\log(1/\delta)). \tag{49}$$

The suppressed factors are size logarithms. The confidence dependence is displayed explicitly.

Each FindAtom call causes at most one update, including on inaccurate-estimate branches. The recent-update history has K records and a physical tree with K leaves, so

$$U \le T, \qquad K = \widetilde{O}(\epsilon^{-1}), \qquad D_{\mathrm{tr}} = O(\log K), \qquad \text{update-path work} = O(T\log K). \tag{50}$$

An overwrite refreshes one physical root-to-leaf path and an update-order cache of $O(\log K)$ nodes.

Propositions C.2 and C.3 prove the preparation guarantee on every update sequence. Discarded history has weight at most $e^{-\epsilon K/512} \le \nu^2/256$. When both purifications share the same tail label, the truncation state error is at most $\nu/8$. Including numerical tree error keeps total error below ν.

Proof of (i). Input queries. Optimization. The coherent test's error allowance $\eta = 1/(64n)$ is independent of δ, whereas the single measured post-check uses β. Multiplying their separately charged costs by T gives

$$\begin{aligned} Q_{\mathrm{opt}} &= O\Big(T\Big[(\sqrt{n/\epsilon} + \epsilon^{-1})\log n + \epsilon^{-1}\log(2/\beta)\Big]\Big) \\ &= \widetilde{O}\Big(\sqrt{n}\,\epsilon^{-3/2}\log(1/\delta) + \epsilon^{-2}\log^2(1/\delta)\Big)\,. \end{aligned} \tag{51}$$

Output. The output step makes one reference estimate and at most M_{out} candidate loss estimates. Its input-query cost is

$$O\big(\epsilon^{-1}\left[\log(2/\beta_{\mathrm{ref}}) + M_{\mathrm{out}}\log(2/\beta_{\mathrm{out}})\right]\big) = \widetilde{O}(\epsilon^{-1}\log^2(1/\delta)).$$

Its $sM_{\mathrm{out}} = \widetilde{O}(\epsilon^{-1}\log(1/\delta))$ state-preparation calls for atom sampling use no input queries.

Total. Thus a deterministic total budget, with $n = 2d$, is

$$Q_{X,y} = \widetilde{O}\Big(\sqrt{d}\,\epsilon^{-3/2}\log(1/\delta) + \epsilon^{-2}\log^2(1/\delta)\Big)\,. \tag{52}$$

At constant confidence, ϵ^{-2} is absorbed by $\sqrt{n}\,\epsilon^{-3/2}$ when $n\epsilon \ge 1$. All N are allowed.

Proof of (ii). Auxiliary quantum work. In the assumed memory model, each use of a sample circuit makes at most eight calls to the input oracles O_X, O_y or their inverses to read entries of X, y and uncompute these reads (Lemma A.3).

The auxiliary operations fall into two groups.

- **Preparation and arithmetic.** These operations include tree and uniform-index preparation, coherent working-memory reads, and polynomial-w reversible arithmetic. Both the recent-update and candidate count trees have logarithmic traversal depth.
- **Estimation and reflections.** Amplitude-estimation controls, finite auxiliary gates, reversible medians, and zero reflections have polynomial cost in their register lengths. The reflection acts on workspace, not on all QRAM cells.

Size logarithms are suppressed in $\widetilde{O}$, and the remaining word cost is $\mathrm{poly}(w)$. Hence the sample-circuit work over the whole run is at most $\widetilde{O}(Q_{X,y})\,\mathrm{poly}(w)$.

Extra Grover preparation and reflection work outside these samples is bounded by

$$\widetilde{O}(\sqrt{n}\,T)\,\mathrm{poly}(w) = \widetilde{O}(\sqrt{n}\,\epsilon^{-1}\log(1/\delta))\,\mathrm{poly}(w), \tag{53}$$

using $R < 2\sqrt{n}$.

The input-query-free output draws add $\widetilde{O}(\epsilon^{-1}\log(1/\delta))\,\mathrm{poly}(w)$. Both are dominated by the deterministic budget in Equation (52) times $\mathrm{poly}(w)$. This proves the G_{aux} bound.

These auxiliary bounds use the assumed input-oracle and efficient coherent working-memory primitives. They exclude the cost of implementing those primitives. The circuit-model lookup cost is treated in Corollary E.2.

Proof of (iii). Classical work. Maintenance and processing have four contributions.

- **Tree maintenance.** Preallocating the recent-update tree costs $O(K)$ word tasks. Its accepted updates and cache refreshes use $O(T\log K) = \widetilde{O}(\epsilon^{-1}\log(1/\delta))$ classical word tasks.
- **Random controls.** The bounded threshold sampler uses at most
  $$T\lceil\log_2(4T/\delta)\rceil\lceil\log_2 L\rceil = \widetilde{O}(\epsilon^{-1}\log^2(1/\delta))$$
  random-bit tasks. Exact Grover-count sampling uses $O(T\log n)$ more.
- **Measured estimates.** Each measured post-check processes $O(\log(2/\beta)) = \widetilde{O}(\log(1/\delta))$ scalar estimates, and the output step similarly processes at most $M_{\mathrm{out}}+1$ confidence-amplified loss estimates. Their classical summaries and linear-time medians fit $\widetilde{O}(\epsilon^{-1}\log^2(1/\delta))$ word tasks.
- **Candidate processing.** The $sM_{\mathrm{out}} = \widetilde{O}(\epsilon^{-1}\log(1/\delta))$ candidate samples and their count-tree updates fit the same bound.

Including arithmetic, classical maintenance and processing therefore cost

$$\widetilde{O}(\epsilon^{-1}\log^2(1/\delta))\,\mathrm{poly}(w).$$

At constant confidence, this is $\widetilde{O}(\epsilon^{-1})\,\mathrm{poly}(w)$.

Generating and issuing the finite gate programs in (ii) has the same polynomial word overhead. Including these instructions gives total classical work $\widetilde{O}(Q_{X,y})\,\mathrm{poly}(w)$.

Proof of (iv). Storage. Classical memory. The recent history, physical tree, and cache use $\widetilde{O}(K) = \widetilde{O}(\epsilon^{-1})$ words, and one empirical candidate uses $\widetilde{O}(\min\{n,\epsilon^{-1}\})$ words. Scalar confidence trials are measured sequentially, with $\widetilde{O}(\log(1/\delta))$ classical scratch for medians. The history, trees, caches, and candidate arrays are classical data in writable QRAM.

Quantum workspace. Preparation and estimation use address, path, finite-word arithmetic, and estimator registers. Only the coherent test's $O(\log n)$ trials keep simultaneous estimator workspaces. Measured trials reuse their quantum registers sequentially. This gives $\widetilde{O}(1)$ quantum workspace words, omitting polynomial word-length overheads and excluding input-oracle and QRAM hardware.

Word length and total storage. Addresses, precision, and counters fit the word length $w = b_{\mathrm{in}} + O(\log(Nn/\epsilon) + \log(2+\log(1/\delta)))$ of Equation (48). Quantum gates are generated and issued sequentially using counters and arithmetic scratch. Apart from polynomial word-length overheads, classical working memory and total auxiliary storage are each bounded by

$$S_{\mathrm{aux}} = \widetilde{O}(\epsilon^{-1} + \log(1/\delta)) \qquad \text{words outside input memory.}$$

Termination. All counts include inaccurate-estimate and abort branches. The horizon T, bounded samplers, and output-trial limit ensure termination on every branch. □

Corollary E.2 (Working-memory lookups in the circuit model)**.** *Under the hypotheses of Proposition E.1, keep the input oracles as black boxes but implement coherent reads of working memory by reversible multiplexers over the current classical tables, without native QRAM. The input-query bound is unchanged. A conservative nonquery gate bound is*

$$\widetilde{O}(KQ_{X,y})\,\mathrm{poly}(w) = \widetilde{O}\Big(\sqrt{d}\,\epsilon^{-5/2}\log(1/\delta) + \epsilon^{-3}\log^2(1/\delta)\Big)\,\mathrm{poly}(w).$$

This is an upper bound for this implementation, not a gate lower bound.

Proof. Each table accessed coherently has $O(K)$ classical words and is frozen during its quantum circuit. To read it coherently, compare the address with each table index and conditionally XOR that word into the output register. Then uncompute the comparison. This uses $O(K)$ gates up to polynomial word and address overheads. It also implements the inverse lookup exactly. Substitute this circuit for every working-memory read counted in Proposition E.1. The other auxiliary gates and the input-query-free output samples fit the same enlarged budget, so the total is $\widetilde{O}(KQ_{X,y})\,\mathrm{poly}(w)$. Classical updates change the compiled table constants between coherent circuits. The input-oracle calls are unaffected, and substituting $K = \widetilde{O}(\epsilon^{-1})$ gives the displayed gate bound. □

F Encoding and population transfer

F.1 Input encoding

The main theorem applies to the encoded values returned by the oracles. The following lemma controls the additional error when transferring its guarantee to the loss on unencoded data.

Lemma F.1 (Stability under input encoding)**.** *Let $N, d \geq 1$, and let (X, y) and $(\widetilde{X}, \widetilde{y})$ be $N \times d$ and N-entry arrays with entries in $[-1, 1]$. Write $L_{X,y}(\theta) = N^{-1}\sum_{i=1}^{N}(\langle X_{i,\cdot}, \theta\rangle - y_i)^2$, and define the encoded loss $L_{\widetilde{X},\widetilde{y}}$ analogously. Suppose their entrywise encoding error is bounded by $\nu_{\mathrm{enc}} \geq 0$:*

$$\max_{i,j}|X_{ij} - \widetilde{X}_{ij}| \leq \nu_{\mathrm{enc}}, \qquad \max_{i}|y_i - \widetilde{y}_i| \leq \nu_{\mathrm{enc}}.$$

Then the following statements hold.

(i) Uniformly over $\|\theta\|_1 \leq 1$,

$$|L_{X,y}(\theta) - L_{\widetilde{X},\widetilde{y}}(\theta)| \leq 8\nu_{\mathrm{enc}}.$$

(ii) Every feasible $\widehat{\theta}$ satisfies

$$\begin{aligned} L_{X,y}(\widehat{\theta}) - \min_{\|\theta\|_1\leq 1} L_{X,y}(\theta) \\ \leq L_{\widetilde{X},\widetilde{y}}(\widehat{\theta}) - \min_{\|\theta\|_1\leq 1} L_{\widetilde{X},\widetilde{y}}(\theta) + 16\nu_{\mathrm{enc}}. \end{aligned} \tag{54}$$

Proof. **Proof of (i).** For any feasible θ, the entrywise error bounds give, for each row i,

$$\big|(\langle X_{i,\cdot}, \theta\rangle - y_i) - (\langle \widetilde{X}_{i,\cdot}, \theta\rangle - \widetilde{y}_i)\big| \leq \nu_{\mathrm{enc}}\|\theta\|_1 + \nu_{\mathrm{enc}} \leq 2\nu_{\mathrm{enc}}.$$

Since all entries lie in $[-1, 1]$, both residuals have magnitude at most $\|\theta\|_1 + 1 \leq 2$. The identity $|a^2 - b^2| = |a-b|\,|a+b|$ therefore bounds each squared-loss difference by $(2\nu_{\mathrm{enc}})(2+2) = 8\nu_{\mathrm{enc}}$. Averaging over the rows proves (i).

Proof of (ii). Part (i) bounds the loss at $\widehat{\theta}$ from above. It also gives $L_{X,y}(\theta) \geq L_{\widetilde{X},\widetilde{y}}(\theta) - 8\nu_{\mathrm{enc}}$ for every feasible θ, so we may take minima over the same feasible set. Together these yield

$$\begin{aligned} L_{X,y}(\widehat{\theta}) &\leq L_{\widetilde{X},\widetilde{y}}(\widehat{\theta}) + 8\nu_{\mathrm{enc}}, \\ \min_{\|\theta\|_1\leq 1} L_{X,y}(\theta) &\geq \min_{\|\theta\|_1\leq 1}\big(L_{\widetilde{X},\widetilde{y}}(\theta) - 8\nu_{\mathrm{enc}}\big) \\ &= \min_{\|\theta\|_1\leq 1} L_{\widetilde{X},\widetilde{y}}(\theta) - 8\nu_{\mathrm{enc}}. \end{aligned}$$

The last equality holds because ν_{enc} is independent of θ. Subtracting this lower bound from the upper bound at $\widehat{\theta}$ gives Equation (54), with total error $8\nu_{\mathrm{enc}} + 8\nu_{\mathrm{enc}} = 16\nu_{\mathrm{enc}}$. □

F.2 Uniform deviation and population risk

The next lemma transfers empirical to population accuracy using a loss bound uniform over all feasible vectors. It therefore covers outputs depending on the sample and quantum measurements.

Lemma F.2 (Uniform transfer on the ℓ_1 ball)**.** *Let $N, d \geq 1$, let $\mathcal{D}$ be a distribution on $[-1,1]^d \times [-1,1]$ and let $\mathcal{S} = ((x_i, y_i))_{i=1}^N \sim \mathcal{D}^N$. Write*

$$L_{\mathcal{D}}(\theta) = \mathbb{E}_{(x,y)\sim\mathcal{D}}(\langle x, \theta\rangle - y)^2, \qquad L_{\mathcal{S}}(\theta) = \frac{1}{N}\sum_{i=1}^{N}(\langle x_i, \theta\rangle - y_i)^2.$$

(i) For every $0 < \delta_{\mathrm{stat}} < 1$, with probability at least $1 - \delta_{\mathrm{stat}}$,

$$\sup_{\|\theta\|_1 \leq 1} |L_{\mathcal{D}}(\theta) - L_{\mathcal{S}}(\theta)| \leq \gamma_N := 16\sqrt{\frac{\log(2d) + \log(2/\delta_{\mathrm{stat}})}{N}}. \tag{55}$$

(ii) For any $\epsilon_{\mathrm{alg}} \geq 0$, a possibly randomized vector $\widehat{\theta}$ is called an empirical ϵ_{alg}-minimizer when $\|\widehat{\theta}\|_1 \leq 1$ and $L_{\mathcal{S}}(\widehat{\theta}) \leq \min_{\|\theta\|_1\leq 1} L_{\mathcal{S}}(\theta) + \epsilon_{\mathrm{alg}}$. On the event in (i), every such vector obeys

$$L_{\mathcal{D}}(\widehat{\theta}) - \min_{\|\theta\|_1\leq 1} L_{\mathcal{D}}(\theta) \leq \epsilon_{\mathrm{alg}} + 2\gamma_N. \tag{56}$$

Proof. **Proof of (i).** For a loss $\ell \in [0, M]$ that is μ-Lipschitz in its first argument, Mohri et al. (2018, Theorem 11.3) gives, with probability at least $1 - \delta_{\mathrm{stat}}$, uniformly over $h \in \mathcal{H}$,

$$\mathbb{E}_{(x,y)\sim\mathcal{D}}\ell(h(x), y) \leq \frac{1}{N}\sum_{i=1}^{N}\ell(h(x_i), y_i) + 2\mu\mathfrak{R}_N(\mathcal{H}) + M\sqrt{\frac{\log(1/\delta_{\mathrm{stat}})}{2N}}.$$

Here $\mathfrak{R}_N(\mathcal{H})$ is the expected Rademacher complexity. For $\mathcal{H} = \{x \mapsto \langle x, \theta\rangle : \|\theta\|_1 \leq \Lambda_1\}$ and $\|x_i\|_\infty \leq r_\infty$, Mohri et al. (2018, Theorem 11.15) bounds its empirical counterpart by

$$\widehat{\mathfrak{R}}_{\mathcal{S}}(\mathcal{H}) \leq r_\infty\Lambda_1\sqrt{\frac{2\log(2d)}{N}}.$$

Taking expectation over $\mathcal{S}$ gives the same bound for $\mathfrak{R}_N(\mathcal{H})$. Here $r_\infty = \Lambda_1 = 1$, and both $(t-y)^2$ and $4 - (t-y)^2$ are bounded by 4 and 4-Lipschitz on $t, y \in [-1, 1]$. Applying Theorem 11.3 to each loss with $M = \mu = 4$ and failure probability $\delta_{\mathrm{stat}}/2$, a union bound gives

$$\sup_{\|\theta\|_1\leq 1} |L_{\mathcal{D}}(\theta) - L_{\mathcal{S}}(\theta)| \leq 8\sqrt{\frac{2\log(2d)}{N}} + 4\sqrt{\frac{\log(2/\delta_{\mathrm{stat}})}{2N}} \leq \gamma_N.$$

Proof of (ii). On the uniform event in (i), for a population minimizer $\theta^\star_{\mathcal{D}}$,

$$L_{\mathcal{D}}(\widehat{\theta}) \leq L_{\mathcal{S}}(\widehat{\theta}) + \gamma_N \leq L_{\mathcal{S}}(\theta^\star_{\mathcal{D}}) + \epsilon_{\mathrm{alg}} + \gamma_N \leq L_{\mathcal{D}}(\theta^\star_{\mathcal{D}}) + \epsilon_{\mathrm{alg}} + 2\gamma_N.$$

□

Theorem F.3 (Population excess risk)**.** *Let $\mathcal{D}$ and $\mathcal{S} = ((x_i, y_i))_{i=1}^N$ be as in Lemma F.2, with ideal sample arrays $X_{ij} = (x_i)_j$ and $y = (y_i)_{i=1}^N$. Fix $0 < \epsilon < 1/2$ and $0 < \delta < 1/3$. Suppose the input oracles return arrays $(\widetilde{X}, \widetilde{y})$ with entries in $[-1, 1]$ and entrywise encoding error at most $\nu_{\mathrm{enc}} \leq \epsilon/64$. Choose an accuracy $\epsilon_{\mathrm{alg}} = 2^{-k_{\mathrm{alg}}}$ with integer k_{alg} satisfying $\epsilon/8 < \epsilon_{\mathrm{alg}} \leq \epsilon/4$, and run Theorem 4.1 with failure probability $\delta/2$ and this accuracy. If*

$$N \geq 4096\,\frac{\log(2d) + \log(4/\delta)}{\epsilon^2} \tag{57}$$

then, jointly over the sample and the algorithm's measurements,

$$\mathbb{P}\left[\mathrm{tag} = \mathrm{success}\ \textit{and}\ L_{\mathcal{D}}(\widehat{\theta}) \leq \min_{\|\theta\|_1\leq 1} L_{\mathcal{D}}(\theta) + \epsilon\right] \geq 1 - \delta.$$

The worst-case number of queries to the encoded sample oracles is

$$\widetilde{O}\Big(\sqrt{d}\,\epsilon_{\mathrm{alg}}^{-3/2}\log(2/\delta) + \epsilon_{\mathrm{alg}}^{-2}\log^2(2/\delta)\Big) = \widetilde{O}\Big(\sqrt{d}\,\epsilon^{-3/2}\log(1/\delta) + \epsilon^{-2}\log^2(1/\delta)\Big). \tag{58}$$

The confidence dependence matches Theorem 4.1.

Proof. **Accuracy and success probability.** Take $\epsilon_{\mathrm{alg}} = 2^{-\lceil \log_2(4/\epsilon) \rceil}$, which lies in $(\epsilon/8, \epsilon/4]$ and is admissible for Theorem 4.1. For every realized sample its encoded empirical excess is at most ϵ_{alg} with failure at most $\delta/2$. Lemma F.1 bounds its ideal empirical excess by $\epsilon_{\mathrm{alg}} + 16\nu_{\mathrm{enc}} \le \epsilon/2$.

Intersect with Equation (55) at $\delta_{\mathrm{stat}} = \delta/2$. The sample-size hypothesis gives $2\gamma_N \le \epsilon/2$. Thus Equation (56) and a union bound prove the risk claim.

Query cost. The accuracy changes by at most a factor eight, and $\log(2/\delta) = \log(1/\delta) + \log 2 = O(\log(1/\delta))$. Substitution in Equation (9) proves Equation (58).

Input encoding. Rounding entries toward zero on a dyadic grid of spacing at most $\epsilon/64$ realizes the encoding tolerance in $O(\log(1/\epsilon))$ bits while keeping the entries in $[-1, 1]$. □

F.3 Lower-bound comparison

Proposition F.4 (Population lower bound in the coordinate-query model)**.** *There are universal constants $c_1, c_2 > 0$ such that the following properties hold with $c_3 := 2048$ for $c_1/d < \epsilon < c_2$ and $0 < \delta < 1/20$. On the hard family of distributions on $\{-1, 1\}^d \times \{1\}$ of Chen & de Wolf (2023, full version, Section 4), every quantum algorithm that returns a population ϵ-minimizer with probability at least $1 - \delta$ on every input makes $\Omega(\sqrt{d}\,\epsilon^{-3/2})$ coordinate queries returning an entire encoded entry. For every $N \ge c_3\epsilon^{-2}\log(2d/\delta)$, the same lower bound holds for any quantum algorithm that solves the empirical problem of Theorem 4.1 on every $N \times d$ input with probability at least $1 - \delta$. At constant confidence, Theorem F.3 matches the population lower bound up to logarithmic and constant factors.*

Proof. Chen & de Wolf (2023, Section 4.3) establish the population Lasso lower bound $\Omega(\sqrt{d}\,\epsilon^{-3/2})$ at success probability $9/10$. After constant rescaling of accuracy and the endpoints c_1, c_2, this gives the population claim. Their hard inputs have entries in $\{-1, 1\}$ and labels 1, so a query returning an entire encoded entry, or its inverse, is simulated with $O(1)$ queries in their model. Their sampling reduction simulates each sample-entry query with $O(1)$ input queries for any sample size N (Chen & de Wolf, 2023, Theorem 22 and the following oracle simulation). Thus the query count gains no factor N.

For empirical optimization, apply Lemma F.2 with $\delta_{\mathrm{stat}} = \delta$. Then $2\gamma_N \le \epsilon$ whenever

$$N \ge 1024\,\frac{\log(2d) + \log(2/\delta)}{\epsilon^2},$$

This follows from $N \ge c_3\epsilon^{-2}\log(2d/\delta)$ with $c_3 = 2048$, since $d \ge 1$ and $\delta < 1/20$ imply

$$\log(2d) + \log(2/\delta) = \log(2d/\delta) + \log 2 \le 2\log(2d/\delta).$$

This numerical choice is sufficient and is not optimized. On that uniform event, an empirical ϵ-minimizer is a population 2ϵ-minimizer. Empirical success at least $1 - \delta$ for every realized sample and a uniform event of probability at least $1 - \delta$ give population success at least $1 - 2\delta > 9/10$ by a union bound. Applying the population lower bound at accuracy 2ϵ gives the same order of queries, with the constant factor absorbed into c_1, c_2.

The population upper bound follows from Theorem F.3 at constant confidence, since $d\epsilon > c_1$ absorbs its ϵ^{-2} term. □

Corollary 4.2 (High-dimensional optimality). *Fix a constant $0 < \delta < 1/20$, let $\epsilon = 2^{-k}$ for an integer $k \ge 3$, and set $c_3 := 2048$. There are universal constants $c_1, c_2 > 0$ such that, for $c_1/d < \epsilon < c_2$ and $N \ge c_3\epsilon^{-2}\log(2d/\delta)$, every quantum algorithm that solves the empirical problem of Theorem 4.1 with probability at least $1-\delta$ on every input makes $\Omega(\sqrt{d}\,\epsilon^{-3/2})$ coordinate queries, and Theorem 4.1 with failure-probability budget δ makes $\widetilde{O}(\sqrt{d}\,\epsilon^{-3/2})$ coordinate queries. In this regime the quantum coordinate query complexity of the Lasso is therefore $\widetilde{\Theta}(\sqrt{d}\,\epsilon^{-3/2})$.*

Proof. The empirical lower bound is Proposition F.4. For the upper bound, δ is fixed, so the confidence factors in Equation (9) are constants, and $d\epsilon > c_1$ absorbs its ϵ^{-2} term. □

For any target accuracy $0 < \epsilon \leq 1/8$, choose a power of two $\epsilon' \in (\epsilon/2, \epsilon]$ and run Theorem 4.1 at ϵ'. The query bound changes by only a constant factor, and the output has fewer than $32/\epsilon$ nonzero entries, still $O(1/\epsilon)$. The sharper $16/\epsilon$ support bound above uses the stated power-of-two accuracy.

Empirical lower-bound qualifications. The classical population bound of Chen & de Wolf (2023, full version, Appendix B, Corollary B.7) is $\widetilde{\Omega}(d\epsilon^{-2})$ for $1/\sqrt{d} < \epsilon < 1/200$. The uniform-risk argument in Lemma F.2, as used in Proposition F.4, transfers this bound to empirical optimization at constant confidence when $N \geq C\epsilon^{-2}\log(2d)$, after constant rescaling of accuracy. This sample size suffices to transfer the hard family through uniform risk control. It does not assert a necessary sample-size condition for every empirical lower bound.

G Comparison with prior work

Table 2 supports the Lasso bounds in Table 1. Table 3 separates other regression objectives and outputs. Table 4 records broader optimization results under their native assumptions.

G.1 Lasso bounds and oracle costs

Confidence and loss scale. For the expected bounds of Hazan & Koren (2012, Theorem 3.4) and Du et al. (2022, Theorem 3), Markov's inequality and a union bound convert expected excess loss $\epsilon/6$ and expected cost Q into ϵ accuracy with probability at least $2/3$ under a $6Q$ query cap. Return a feasible fallback on truncation. Any separate source failure event must also be budgeted, for example by using $\epsilon/12$ and $12Q$ to reserve another $1/6$. Since $L_{X/s,y/s} = L_{X,y}/s^2$, data scaling requires target error ϵ/s^2. Du et al.'s objective is $L_{X,y}/2$, so its target accuracy is $\epsilon/2$.

Oracle-cost conversion. Chen & de Wolf (2023, Theorems 10–11 and Corollary 12) give coherent estimates of loss values and gradient coordinates to accuracy τ using $\widetilde{O}(\tau^{-1})$ entry queries at constant confidence, given access to a Kerenidis–Prakash (KP) tree. This binary tree stores scaled coordinates and partial sums of their magnitudes for state preparation (Chen & de Wolf, 2023, Section 2.4). For explicit finite-precision points, indexed KP-tree lookups can be computed reversibly from the point registers without input queries. Their unit-ball restriction extends to $\|z\|_1 \leq 3/2$ by $L_{X,y}(z) = \frac{9}{4}L_{X,2y/3}(2z/3)$, with a constant fraction of the error budget reserved for rounding.

For the adaptation of Jordan (2005), write $L = L_{X,y}$. For feasible θ and $h \in [-1/(2d), 1/(2d)]^d$, we have $\|\theta \pm h\|_1 \leq 3/2$ and

$$F_\theta(h) := \frac{L(\theta + h) - L(\theta - h)}{2} = \langle \nabla L(\theta), h \rangle. \tag{59}$$

Two sufficiently accurate, amplified loss estimates implement a unitary τ-approximator of F_θ with retained work registers. Its value register is within τ of $F_\theta(h)$ with probability at least $2/3$ for every basis input h (van Apeldoorn et al., 2020, Definitions 2 and 28). van Apeldoorn et al. (2020, Corollary 29), applied with $r = 1/d$ and $B = 4$, gives gradient error $O(\tau/r)$ in the infinity norm using $O(\log(d/\rho))$ calls to this unitary and its inverse. Its hypotheses hold since F_θ is linear and $\|\nabla L(\theta)\|_\infty \leq 4$. Choosing a sufficiently small $\tau = \Theta(\alpha/d)$ gives gradient error α at cost $\widetilde{O}((d/\alpha)\log(d/\rho))$. For $0 < \rho \leq 1/3$, split the failure budget between the routine and finite gate synthesis.

Choose a sufficiently small $\alpha = \Theta(\epsilon)$ and conditional failure probability at most $1/(3T)$ per update. The approximate Frank–Wolfe bound in Proposition 3.2 then gives accuracy ϵ after $T = O(\epsilon^{-1})$ signed-coordinate updates from zero, with constant success probability and at most T nonzeros. This yields $\widetilde{O}(d\epsilon^{-2})$ input queries. For minimum finding, $L(\theta + e_j/2) - L(\theta - e_j/2) = \partial_j L(\theta)$ gives coordinate accuracy $O(\epsilon)$ at cost $\widetilde{O}(\epsilon^{-1})$. The bounded-error search of Chen & de Wolf (2023, Theorem 1) therefore costs $\widetilde{O}(\sqrt{d}/\epsilon)$ per update and $\widetilde{O}(\sqrt{d}\,\epsilon^{-2})$ in total, matching their Theorem 13. These are entry-query adaptations of the approaches in He & Lui (2025). Classical processing and circuit generation are outside these query bounds.

Table 2: Native costs and Lasso entry-query implementations at constant confidence, up to logarithmic factors.

Paper	Claimed complexity	Lasso implementation	Entry-query bound
Hazan & Koren (2012), Thm. 3.4	$\widetilde{O}(d\epsilon^{-2})$ observed attributes; expected excess loss.	Uniform empirical rows and the confidence conversion above.	$\widetilde{O}(d\epsilon^{-2})$ (classical).
Du et al. (2022), Thm. 3 (non-private)	$\widetilde{O}(T^3\sqrt{Nd}/\varsigma)$ time; error $O(C_f/T+\varsigma)$ for $L/2$.	Normalization above with $C_f = O(1)$, $T = \Theta(\epsilon^{-1})$, $\varsigma = \Theta(\epsilon)$.	$\widetilde{O}(\sqrt{Nd}\,\epsilon^{-4})$.
Chen & de Wolf (2023)	$\widetilde{O}(\sqrt{d}\,\epsilon^{-2})$ time for an explicit Lasso minimizer.	Same entry model and constant-confidence convention.	$\widetilde{O}(\sqrt{d}\,\epsilon^{-2})$.
He & Lui (2025), Thms. 1, 5	Reported value calls: $\widetilde{O}(\sqrt{d}/\epsilon)$ (minimum finding) or $O(1/\epsilon)$ (Jordan gradient).	Signed Frank–Wolfe directions and the value-oracle implementation above.	$\widetilde{O}(\sqrt{d}\,\epsilon^{-2})$ or $\widetilde{O}(d\epsilon^{-2})$, respectively.
van Apeldoorn et al. (2020), Cor. 29	$O(\log(d/\rho))$ value calls for gradient error $O(\tau/r)$, failure ρ.	Use $r = 1/d$, $\tau = \Theta(\epsilon/d)$, then Frank–Wolfe.	$\widetilde{O}(d\epsilon^{-2})$ for this optimization adaptation.
This work	Theorem 4.1: explicit feasible ϵ-optimal, $O(\epsilon^{-1})$-sparse output.	Direct entry access; no condition-number or incoherence promise.	$\widetilde{O}(\sqrt{d}\,\epsilon^{-3/2}+\epsilon^{-2})$.

Source qualification. Our adaptation of He & Lui (2025) uses the gradient sign omitted from their Algorithms 2 and 5. These entry-query bounds use the conversion above, including the precision-dependent cost of each value call. The table fixes confidence. Before the suppressed size logarithms, the first term of AQFW's bound dominates the second when $d\epsilon \gtrsim \log^2(1/\delta)$. These asymptotic query comparisons exclude input loading and do not establish a practical running-time crossover.

G.2 Other regression objectives and outputs

Regression data have size $N \times d$, while the input of Li et al. (2019) has size $m \times k$. All bounds below have constant success probability. Here, ξ is an accuracy parameter and r is maximum row sparsity. For least-angle regression (LARS) in Doriguello et al. (2025), A and I are active and inactive sets, T bounds the number of kinks, and X_A^+ is the Moore–Penrose inverse. Write $\mathcal{L}_\lambda(w) = \frac{1}{2}\|y - Xw\|_2^2 + \lambda\|w\|_1$. The approximate LARS row assumes $\alpha_A, \gamma_A \in (0,1)$ and, for every relevant A with $|A| \le T$,

$$\|X_A^+ X_{A^c}\|_1 \le \alpha_A, \qquad \frac{\|X^\top z_A\|_\infty}{\|X\|_{\max}\|z_A\|_1} \ge \gamma_A, \qquad z_A = (I_N - X_A X_A^+)y.$$

Here the matrix 1-norm is the maximum absolute column sum and $\|X_A^+\|_2$ is the spectral norm. For Lim et al. (2024), T counts online rounds. For quantum orthogonal matching pursuit (QOMP) in Bellante et al. (2025), m is dictionary size, K bounds exact sparsity, μ is mutual incoherence, $\gamma \le \sigma_{\min}(D_{\Lambda_{\mathrm{opt}}})$, $\eta \in (0,1)$ is incoherence slack, and $\varepsilon_Q \in (0,1)$ is residual tolerance. QOMP assumes a unit-norm target and unit-norm dictionary columns. For Liu & Ji (2025), η denotes the printed multiplicative-error parameter.

Table 3: Native complexities for other regression objectives or outputs.

Paper	Native complexity and scope
Li et al. (2019), Cor. 4.3	$\widetilde{O}(\sqrt{m}\,\xi^{-4} + \min\{k\xi^{-1}, \sqrt{k}\,\xi^{-7}\})$ gates for $\mathrm{OPT} = \max_{w\in\mathbb{R}^k} \min_{i\in[m]}(2X_i^\top w - \|w\|_2^2)$, with $\max_i \|X_i\|_2 \le 1$ and entry-oracle access. Returns a succinct classical representation of $\bar{w}$ attaining $\mathrm{OPT} - \xi$, for $0 < \xi < \mathrm{OPT}$.

Continued on next page

Table 3 (continued)

Paper	Native complexity and scope
Lim et al. (2024), Thm. 5.3	$\widetilde{O}(\sqrt{d}\,T^{5/2})$ example-entry queries for average regularized regret $O(T^{-1/2})$ in online least squares, with truncated ℓ_1 penalties. This fixes bounds on row norms, prediction errors, the shrinkage parameter, and comparator norms. Weights are available through coherent entry access.
Doriguello et al. (2025), Thm. 22	$\widetilde{O}(N\sqrt{\lvert I\rvert} + N\lvert A\rvert + \lvert A\rvert^2)$ time per kink for at most T kinks of the exact Lasso path. Assumes X in quantum read-only memory (QROM) and writable QRAMs of size $O(N)$.
Doriguello et al. (2025), Thm. 27	$\widetilde{O}\left(\frac{\gamma_A^{-1}+\sqrt{N}\lVert X\rVert_{\max}\lVert X_A^+\rVert_2}{(1-\alpha_A)\xi}\sqrt{\lvert I\rvert} + N\lvert A\rvert + \lvert A\rvert^2\right)$ time per kink under the conditions above and the same QROM/QRAM access. Returns an approximate path with at most T kinks satisfying $\mathcal{L}_\lambda(\widetilde{w}) - \min_w \mathcal{L}_\lambda(w) \le \lambda\xi\lVert\widetilde{w}\rVert_1$.
Bellante et al. (2025), Thm. 40	Reported costs are $\widetilde{O}(K^{3/2}\sqrt{m}/(\gamma\eta\varepsilon_Q))$ target-state calls and $\widetilde{O}(K^2\sqrt{m}/(\gamma\eta\varepsilon_Q))$ dictionary calls, including inverse and controlled state-preparation oracles. If $K < \frac{1-\eta}{2-\eta}(1+\mu^{-1})$, the output is a subset of the exact support whose span approximates the target within ε_Q.
Liu & Ji (2025), Cor. 26	Reports $\widetilde{O}(r\sqrt{Nd}/\eta + d^3/\eta^2)$ time and a classical coefficient vector for penalized least squares, with the guarantee qualified below. Assumes entry and row-nonzero-location oracles for X, and entry access to y.

Source qualifications. For the LARS bounds, we use Doriguello et al. (2025, Theorems 22 and 27) (arXiv version 3), fixing confidence and suppressing logarithmic dependence on T. Bellante et al. (2025, Theorem 40) (arXiv version 1) report γ^{-1} for dictionary calls. Substituting its inner-product tolerance into Theorem 33, Eq. (33), gives γ^{-2} instead. Liu & Ji (2025, Corollary 26) (arXiv version 1) state

$$\|Xw - y\|_2^2 + \lambda\|w\|_1 \le (1+\eta)\min_v\big(\|Xv - y\|_2^2 + \|v\|_1\big).$$

The right-hand objective omits λ, so the printed statement is not a relative-error guarantee for the same objective at general λ.

G.3 Broader optimization under native assumptions

Symbols in this comparison are local to each source. Here, N is the number of components, d the number of variables, and ϵ the excess-objective tolerance. The parameter R bounds distance to an optimizer or saddle point. For the finite-sum bound of Zhang et al. (2024, Corollary 2), ℓ is the component-smoothness constant and we abbreviate

$$C_{\mathrm{fs}} := R\sqrt{\ell/\epsilon}(N^{1/3}d^{1/3} + N^{-2/3}d^{5/6}).$$

The constants G^2 and σ_V^2 bound gradient second moments and gradient variance in Sidford & Zhang (2023, Theorem 5) and Marsden et al. (2025, Theorem 7), respectively. Matrix games use an $m \times k$ matrix A and saddle-gap tolerance η (Bouland et al., 2023; Gao et al., 2023; Karmarkar et al., 2026). For Liu et al. (2025, Theorem 4.5), k is the joint dimension and L_2 is the Hessian-Lipschitz constant. For Kim et al. (2025, Theorem C.3), L is the smoothness constant, μ is the mirror map's strong-convexity constant, and R^2 bounds Bregman distance.

Table 4: Broader optimization bounds under the original oracle models.

Work	Complexity and access	Guarantee and assumptions
Convex minimization		
Zhang et al. (2024) Cor. 2 Finite-sum optimization	$\widetilde{O}(N + \sqrt{d} + C_{\mathrm{fs}})$ Expected full component-gradient queries.	Expected excess objective ϵ. Convex ℓ-smooth components; known convex proximal term; optimizer within radius R.
Sidford & Zhang (2023) Thm. 5 Stochastic optimization	$\widetilde{O}(d^{5/8}(GR/\epsilon)^{3/2})$ Expected stochastic-gradient queries.	Expected excess loss ϵ. Unconstrained convex objective; global gradient second-moment bound G^2.

Continued on next page

Table 4 (continued)

Work	Complexity and access	Guarantee and assumptions
Marsden et al. (2025) Thm. 7 Cutting planes	$\widetilde{O}(dR\sigma_V/\epsilon)$ Stochastic-gradient queries.	Excess loss ϵ with probability $\geq 2/3$. Unconstrained, globally Lipschitz convex objective; global gradient variance bound $\sigma_V^2 > 0$.
Chakrabarti et al. (2020) Thm. 1.1 Convex-body optimization	$\widetilde{O}(d)$ queries of each type Value and membership access; oracle precision depends on geometry.	Approximate convex optimization. Convex body with known inner and outer radii and objective bounds.
Kim et al. (2025) Thm. C.3 Mirror prox	$\widetilde{O}(LR^2/(\mu\epsilon))$ Binary value queries at precision θ.	Excess loss ϵ with probability $\geq 2/3$. Closed bounded convex domain; Lipschitz, L-smooth objective; μ-strongly convex mirror map.
Games and saddle-point problems		
Bouland et al. (2023) Thm. 2.3 Gibbs sampling	$\widetilde{O}(\sqrt{m+k}\,\eta^{-5/2}+\eta^{-3})$ Time; quantum entry access.	Classical simplex strategies. Saddle gap η. $\|A\|_{\max} \leq 1$.
Gao et al. (2023) Cor. 1.2 Multiplicative weights	$\widetilde{O}(\sqrt{m+k}\,\eta^{-5/2})$ Time; quantum entry access.	Classical simplex strategies. Saddle gap η. $\|A\|_{\mathrm{op}} \leq 1$; source's classical fallback applies at small η.
Karmarkar et al. (2026) Thm. 1.2 (arXiv v3) Deterministic optimization	$\widetilde{O}(\eta^{-2/3})$ Full matrix–vector products $(A^\top y, Ax)$; classical algorithm.	Ball–simplex strategies; saddle gap η. Each row of A has Euclidean norm at most one.
Liu et al. (2025) Thm. 4.5 Minimax optimization	$\widetilde{O}(k+k^{2/3}L_2^{2/3}R^2\eta^{-2/3})$ Value queries.	Restricted saddle gap η. Unconstrained convex–concave objective; f, ∇f, and $\nabla^2 f$ globally Lipschitz.

All methods are quantum except the deterministic method of Karmarkar et al. (2026). Randomized guarantees have constant success probability unless stated in expectation. Bounds use each source's cost units; the table lists selected assumptions. Full oracle specifications and guarantees are in the cited papers.