Quantum Algorithms for the Pathwise Lasso

João F. Doriguello, Debbie Lim, Chi Seng Pun, Patrick Rebentrost, Tushar Vaidya

June 1, 2024

Extended abstract We present a novel quantum high-dimensional linear regression algorithm with an ℓ_1-penalty based on the classical Least Angle Regression pathwise algorithm. More precisely, given $\mathbf{X} \in \mathbb{R}^{n\times d}$ and $\mathbf{y} \in \mathbb{R}^n$ with $d \gg n$, consider the Lasso regression

$$\hat{\boldsymbol{\beta}} \in \underset{\boldsymbol{\beta}\in\mathbb{R}^d}{\operatorname{argmin}} \frac{1}{2}\|\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\|_2^2 + \lambda\|\boldsymbol{\beta}\|_1 \quad \text{for all } \lambda > 0.$$

The ℓ_1-penalty term is $\|\boldsymbol{\beta}\|_1 = \sum_{i=1}^d |\beta_i|$, whereas the predictive loss is the standard Euclidean squared norm. In this work, we are interested in the Lasso solution $\hat{\boldsymbol{\beta}}(\lambda)$ as a function of the regularisation parameter $\lambda > 0$. For such we define the optimal regularisation path $\mathcal{P}$:

$$\mathcal{P} \triangleq \{\hat{\boldsymbol{\beta}}(\lambda) : \lambda > 0\}.$$

Lasso-type methods are used across disciplines [3, 2, 7, 9, 10, 11]. Efron *et al.* [6] showed that the optimal regularisation path $\mathcal{P}$ is *piecewise linear* and *continuous* with respect to λ. This means that there exist an $m \in \mathbb{N}$ and $\infty > \lambda_0 > \cdots > \lambda_{m-1} > \lambda_m = 0$ and $\boldsymbol{\theta}_0, \ldots, \boldsymbol{\theta}_m \in \mathbb{R}^d$ such that

$$\hat{\boldsymbol{\beta}}(\lambda) = \hat{\boldsymbol{\beta}}(\lambda_t) + (\lambda_t - \lambda)\boldsymbol{\theta}_t \quad \text{for } \lambda_{t+1} < \lambda \leqslant \lambda_t \quad (t \in \{0, \ldots, m-1\}).$$

The $m+1$ points $\lambda_0, \ldots, \lambda_m$ where $\partial\hat{\boldsymbol{\beta}}(\lambda)/\partial\lambda$ changes are called *kinks*, and the path $\{\hat{\boldsymbol{\beta}}(\lambda) : \lambda_{t+1} < \lambda \leqslant \lambda_t\}$ between two consecutive kinks λ_{t+1} and λ_t defines a linear segment. The fact that the regularisation path is piecewise linear and the locations of the kinks can be derived from the Karush-Kuhn-Tucker (KKT) optimality conditions.

Since the regularisation path is piecewise linear, one needs only to compute all kinks $(\lambda_t, \hat{\boldsymbol{\beta}}(\lambda_t))$ for $t \in \{0, \ldots, m-1\}$, from which the whole regularisation path follows by linear interpolation. That is exactly what the Least Angle Regression (LARS) algorithm proposed and named by Efron *et al.* [6] does. Starting at $\lambda = \infty$ where the Lasso solution is $\hat{\boldsymbol{\beta}}(\lambda) = \mathbf{0} \in \mathbb{R}^d$, the LARS algorithm decreases the regularisation parameter λ and computes the regularisation path by finding the kinks along the way with the aid of the KKT conditions. Each kink corresponds to an iteration of the algorithm. More specifically, at each iteration t, the LARS algorithm computes the next kink λ_{t+1} by finding the closest point (to the previous kink) where the KKT conditions break and thus need to be updated. This is done by performing a search over the *active set* $\mathcal{A} \triangleq \{i \in [d] : \hat{\beta}_i \neq 0\}$ and another search over the *inactive set* $\mathcal{I} \triangleq [d]\backslash\mathcal{A}$ (let $[d] \triangleq \{1, \ldots, d\}$). The search over $\mathcal{A}$ finds the next point $\lambda_{t+1}^{\text{cross}}$, called *crossing time*, where a variable i_{t+1}^{cross} must leave $\mathcal{A}$ and join $\mathcal{I}$. The search over $\mathcal{I}$ finds the next point $\lambda_{t+1}^{\text{join}}$, called *joining time*, where a variable i_{t+1}^{join} must leave $\mathcal{I}$ and join $\mathcal{A}$. The next kink is thus $\lambda_{t+1} = \max\{\lambda_{t+1}^{\text{cross}}, \lambda_{t+1}^{\text{join}}\}$. The overall complexity of the LARS algorithm per iteration is $O(nd + |\mathcal{A}|^2)$: the two searches over the active and inactive sets require $O(n|\mathcal{A}|)$ and $O(n|\mathcal{I}|)$ time, respectively, while computing the new direction $\partial\hat{\boldsymbol{\beta}}/\partial\lambda = \boldsymbol{\theta}_{t+1}$ of the regularisation path, which involves the computation of the pseudo-inverse of a submatrix of $\mathbf{X}$ specified by $\mathcal{A}$, requires $O(n|\mathcal{A}| + |\mathcal{A}|^2)$ time.

Our work

In the high-dimensional setting where $d \gg n$, the search over the inactive set $\mathcal{I}$, and thus the computation of the joining time, is by far the most costly step per iteration. In this work, we propose quantum algorithms for the pathwise Lasso regression problem based on the Least Angle Regression (LARS) algorithm. To the best of our knowledge, this work is the first to present a quantum version of the LARS algorithm. We propose mainly two quantum algorithms based on the LARS algorithm to speedup the computation of the joining time. Our first quantum algorithm, called simple quantum LARS algorithm, is a straightforward improvement that utilizes the well-known quantum minimum-finding subroutine from Dürr and Høyer [5] to perform the search over $\mathcal{I}$. We assume that the design matrix $\mathbf{X}$ is stored in a quantum-readable read-only memory (QROM) and can be accessed in time $O(\operatorname{poly}\log(nd))$. We also assume that data can be written into quantum-readable classical-writable classical memories (QRAM) of size $O(n)$, which can be later queried in time $O(\operatorname{poly}\log n)$. The complexity of computing the joining time is now $O(n\sqrt{|\mathcal{I}|})$, where $O(n)$ is the time required to compute, via classical circuits, the joining times to be maximised over $\mathcal{I}$. The final runtime of our simple quantum algorithm is $\widetilde{O}(n\sqrt{|\mathcal{I}|}+n|\mathcal{A}|+|\mathcal{A}|^2)$ per iteration, where we omit polylog factors in n and d.

We then improve upon our simple quantum algorithm by approximately computing the joining times to be maximised over $\mathcal{I}$, thus reducing the factor $O(n)$ to $O(\sqrt{n})$. This is done by constructing a quantum unitary based on quantum amplitude estimation [1] that approximately computes the joining times of all the variables in the inactive set $\mathcal{I}$. We then employ the recently developed approximate quantum minimum-finding subroutine from Chen and de Wolf [4] to find an estimate ϵ-close to the true maximum joining time. The result is our approximate quantum LARS algorithm that achieves a quadratic speedup in both the number of features d and the number of observations n. We also propose a sampling-based classical LARS algorithm. Our improved quantum algorithm (and the sampling-based classical algorithm), however, introduces an error in computing the joining time $\lambda_{t+1}^{\text{join}}$. As one of our main contributions, we prove that the LARS algorithm is robust to errors. More specifically, we show that, by slightly adapting the LARS algorithm, it returns (and so our approximate quantum algorithm) an *approximate regularisation path* with error proportional to the error from computing $\lambda_{t+1}^{\text{join}}$. We first define an approximate regularisation path.

Definition 1. *Given $\epsilon : \mathbb{R}_{>0} \to \mathbb{R}_{\geqslant 0}$, $\widetilde{\boldsymbol{\beta}} \in \mathbb{R}^d$ is an approximate Lasso solution with error $\epsilon(\lambda)$ if*

$$\frac{1}{2}\|\mathbf{y}-\mathbf{X}\widetilde{\boldsymbol{\beta}}\|_2^2+\lambda\|\widetilde{\boldsymbol{\beta}}\|_1-\left(\min_{\boldsymbol{\beta}\in\mathbb{R}^d}\frac{1}{2}\|\mathbf{y}-\mathbf{X}\boldsymbol{\beta}\|_2^2+\lambda\|\boldsymbol{\beta}\|_1\right)\leqslant\epsilon(\lambda)\quad \text{where } \lambda>0.$$

A set $\widetilde{\mathcal{P}} \triangleq \{\widetilde{\boldsymbol{\beta}}(\lambda) \in \mathbb{R}^d : \lambda > 0\}$ is an approximate regularisation path with error $\epsilon(\lambda)$ if, for all $\lambda > 0$, the point $\widetilde{\boldsymbol{\beta}}(\lambda)$ of $\widetilde{\mathcal{P}}$ is an approximate Lasso solution with error $\epsilon(\lambda)$.

Result 2. *Let $\epsilon \in [0, 2)$. Consider an approximate* LARS *algorithm that returns $\widetilde{\mathcal{P}} = \{\widetilde{\boldsymbol{\beta}}(\lambda) \in \mathbb{R}^d : \lambda > 0\}$ and wherein, at each iteration t, the joining time $\lambda_{t+1}^{\text{join}}$ is approximated by $\widetilde{\lambda}_{t+1}^{\text{join}}$ such that $\lambda_{t+1}^{\text{join}} \leqslant \widetilde{\lambda}_{t+1}^{\text{join}} \leqslant (1-\epsilon/2)^{-1}\lambda_{t+1}^{\text{join}}$. Then $\widetilde{\mathcal{P}}$ is an approximate regularisation path with error $\lambda\epsilon\|\widetilde{\boldsymbol{\beta}}(\lambda)\|_1$.*

To prove the above result, we consider an approximate version of the KKT conditions and use a duality gap for the Lasso regression. As far as we are aware, this is the first direct result on the robustness of the LARS algorithm. Compared to the work of Mairal and Yu [8], who also introduced an approximate LARS algorithm and employed similar techniques to analyse its correctness, the computation of the joining time in their case is exact, and errors only arise by utilising a first-order optimisation method to find an approximate solution when kinks happen to be too close.

Result 2 guarantees the correctness of all our algorithms. Their time complexities, on the other hand, are given by the theorem below and summarised in Table 1. In the following, let $\mathbf{X}^+$ be the Moore–Penrose inverse of $\mathbf{X}$. Also, given some matrix $\mathbf{A}$, let $\|\mathbf{A}\|_2$ be its the spectral norm, $\|\mathbf{A}\|_1 = \max_j \sum_i |A_{ij}|$, and $\|\mathbf{A}\|_{\max} = \max_{i,j} |A_{ij}|$.

Result 3. *Let $\mathbf{X} \in \mathbb{R}^{n\times d}$ and $\mathbf{y} \in \mathbb{R}^n$. Assume that $\mathbf{X}$ is stored in a* QROM *and we have access to* QRAMs *and classical-samplable structures of memory size $O(n)$. Let $\delta, \epsilon \in (0,1)$, and $T \in \mathbb{N}$. Let $\alpha, \gamma \in (0,1]$ such that*

$$\max_{\mathcal{A}\subseteq[d]:|\mathcal{A}|\leqslant T} \|\mathbf{X}_{\mathcal{A}}^+ \mathbf{X}_{\mathcal{A}^c}\|_1 \leqslant 1-\alpha \quad \text{and} \quad \min_{\mathcal{A}\subseteq[d]:|\mathcal{A}|\leqslant T} \frac{\|\mathbf{X}^\top(\mathbf{I}-\mathbf{X}_{\mathcal{A}}\mathbf{X}_{\mathcal{A}}^+)\mathbf{y}\|_\infty}{\|\mathbf{X}\|_{\max}\|(\mathbf{I}-\mathbf{X}_{\mathcal{A}}\mathbf{X}_{\mathcal{A}}^+)\mathbf{y}\|_1} \geqslant \gamma,$$

where $\mathcal{A}^c = [d]\backslash\mathcal{A}$ and $\mathbf{X}_{\mathcal{A}} \in \mathbb{R}^{n\times|\mathcal{A}|}$ is the matrix formed by the columns of $\mathbf{X}$ in $\mathcal{A}$.

- *There is a quantum* LARS *algorithm that returns an optimal regularisation path with T kinks with probability at least $1-\delta$ and in time per iteration*

$$\widetilde{O}\big(n\sqrt{|\mathcal{I}|} + n|\mathcal{A}| + |\mathcal{A}|^2\big).$$

- *There is a quantum* LARS *algorithm that returns an approximate regularisation path with additive error $\lambda\epsilon\|\widetilde{\boldsymbol{\beta}}\|_1$ and T kinks with probability at least $1-\delta$ and in time per iteration*

$$\widetilde{O}\left(\frac{\gamma^{-1}+\sqrt{n}\|\mathbf{X}\|_{\max}\|\mathbf{X}^+\|_2}{\alpha\epsilon}\sqrt{|\mathcal{I}|} + n|\mathcal{A}| + |\mathcal{A}|^2\right).$$

- *There is a classical (dequantized)* LARS *algorithm that returns an approximate regularisation path with additive error $\lambda\epsilon\|\widetilde{\boldsymbol{\beta}}\|_1$ and T kinks with probability at least $1-\delta$ and in time per iteration*

$$\widetilde{O}\left(\frac{\gamma^{-2}+n\|\mathbf{X}\|_{\max}^2\|\mathbf{X}^+\|_2^2}{\alpha^2\epsilon^2}|\mathcal{I}| + n|\mathcal{A}| + |\mathcal{A}|^2\right).$$

The notation $\widetilde{O}(\cdot)$ omits poly log *terms in n, d, T, and δ.*

Algorithm	**Error**	**Time complexity per iteration**
Classical LARS	None	$O(n\|\mathcal{I}\| + n\|\mathcal{A}\| + \|\mathcal{A}\|^2)$
Simple quantum LARS	None	$\widetilde{O}(n\sqrt{\|\mathcal{I}\|} + n\|\mathcal{A}\| + \|\mathcal{A}\|^2)$
Approximate classical LARS (dequantized)	$\lambda\epsilon\|\widetilde{\boldsymbol{\beta}}\|_1$	$\widetilde{O}\left(\frac{\gamma^{-2}+n\|\mathbf{X}\|_{\max}^2\|\mathbf{X}^+\|_2^2}{\alpha^2\epsilon^2}\|\mathcal{I}\| + n\|\mathcal{A}\| + \|\mathcal{A}\|^2\right)$
Approximate quantum LARS	$\lambda\epsilon\|\widetilde{\boldsymbol{\beta}}\|_1$	$\widetilde{O}\left(\frac{\gamma^{-1}+\sqrt{n}\|\mathbf{X}\|_{\max}\|\mathbf{X}^+\|_2}{\alpha\epsilon}\sqrt{\|\mathcal{I}\|} + n\|\mathcal{A}\| + \|\mathcal{A}\|^2\right)$

Table 1: Summary of results. Here n is the number of observations, d is the number of features, $|\mathcal{A}|$ is the size of the active set, and $|\mathcal{I}|$ is the size of the inactive set. In addition, $\epsilon \in (0,1)$, $\mathbf{X}^+$ is the Moore–Penrose inverse of $\mathbf{X}$, and the parameters $\alpha, \gamma \in (0,1]$ are defined in Result 3.

Compared to the previous work of Chen and de Wolf [4] on quantum algorithms for Lasso, our work is different in that we study the pathwise Lasso regression problem as the tuning parameter λ varies, and not focus on a single point λ. Both works are thus incomparable. The output solution from the quantum algorithm of [4] can actually be used to initialise our algorithms instead of the point $(\lambda = \infty, \hat{\boldsymbol{\beta}}(\infty) = \mathbf{0})$.

References

[1] Gilles Brassard, Peter Høyer, Michele Mosca, and Alain Tapp. Quantum amplitude amplification and estimation. *Contemporary Mathematics*, 305:53–74, 2002. 2

[2] Emmanuel J. Candès, Justin K. Romberg, and Terence Tao. Stable signal recovery from incomplete and inaccurate measurements. *Communications on Pure and Applied Mathematics*, 59(8):1207–1223, 2006. 1

[3] Scott Shaobing Chen, David L. Donoho, and Michael A. Saunders. Atomic decomposition by basis pursuit. *SIAM Review*, 43(1):129–159, 2001. 1

[4] 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, editors, *50th International Colloquium on Automata, Languages, and Programming (ICALP 2023)*, volume 261 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 38:1–38:21, Dagstuhl, Germany, 2023. Schloss Dagstuhl – Leibniz-Zentrum für Informatik. 2, 3

[5] Christoph Dürr and Peter Høyer. A quantum algorithm for finding the minimum. *arXiv preprint quant-ph/9607014*, 1996. 2

[6] Bradley Efron, Trevor Hastie, Iain Johnstone, and Robert Tibshirani. Least angle regression. *The Annals of Statistics*, 32(2):407 – 499, 2004. 1

[7] Mário A. T. Figueiredo, Robert D. Nowak, and Stephen J. Wright. Gradient projection for sparse reconstruction: Application to compressed sensing and other inverse problems. *IEEE Journal of Selected Topics in Signal Processing*, 1(4):586–597, 2007. 1

[8] Julien Mairal and Bin Yu. Complexity analysis of the lasso regularization path. In *Proceedings of the 29th International Coference on International Conference on Machine Learning*, ICML'12, page 1835–1842, Madison, WI, USA, 2012. Omnipress. 2

[9] Yiming Peng and Vadim Linetsky. Portfolio selection: A statistical learning approach. In *Proceedings of the Third ACM International Conference on AI in Finance*, ICAIF '22, page 257–263, New York, NY, USA, 2022. Association for Computing Machinery. 1

[10] M. Graziano Usai, Mike E. Goddard, and Ben J. Hayes. LASSO with cross-validation for genomic selection. *Genetics Research*, 91(6):427–436, 2009. 1

[11] Tong Tong Wu, Yi Fang Chen, Trevor Hastie, Eric Sobel, and Kenneth Lange. Genome-wide association analysis by lasso penalized logistic regression. *Bioinformatics*, 25(6):714–721, 01 2009. 1