code

Householder Sketch for Accurate and Accelerated Least-Mean-Squares Solvers

Jyotikrishna Dass and Rabi Mahapatra
Department of Computer Science and Engineering
Texas A&M University, College Station

TEXAS A&M UNIVERSITY

Introduction

Sketching – a technique to summarize data X into S to preserve or approximate covariance matrix, i.e. $S^T S = X^T X$

Least-Mean-Squares (LMS)

$$\min_{\mathbf{w}} f(\|\mathbf{X}\mathbf{w}-\mathbf{y}\|_2)+g(\mathbf{w}).$$

Linear Regression, $f(z)=z^2$, and $g(\mathbf{w})=0$.

$$(\mathbf{X}^T\mathbf{X})\mathbf{w}=\mathbf{X}^T\mathbf{y}$$

Ridge Regression, $f(z)=z^2$, and $g(\mathbf{w})=\lambda\|\mathbf{w}\|_2$, where, $\lambda>0$,

$$(\mathbf{X}^T\mathbf{X}+\lambda\mathbf{I})\mathbf{w}=\mathbf{X}^T\mathbf{y}$$

Focus

Create theoretically accurate summary of input data which could be directly plugged to accelerate common scikit-learn LMS solvers

Inspiration

LMS-BOOST [1] – Coreset-sketch fusion algorithm based on faster implementation of Caratheodory Theorem (1907)

- Summarizes input data X into S of size $O(d^2)$ x d
- Preserves input covariance, $S^T S = X^T X$
- Has computational time complexity, $O(nd^2 + \log n \times d^8)$
- **CLAIMS against QR decomposition**
  1. **Relatively time-consuming**
  2. **Unsuitable for exact factorization for streaming data**

Contributions

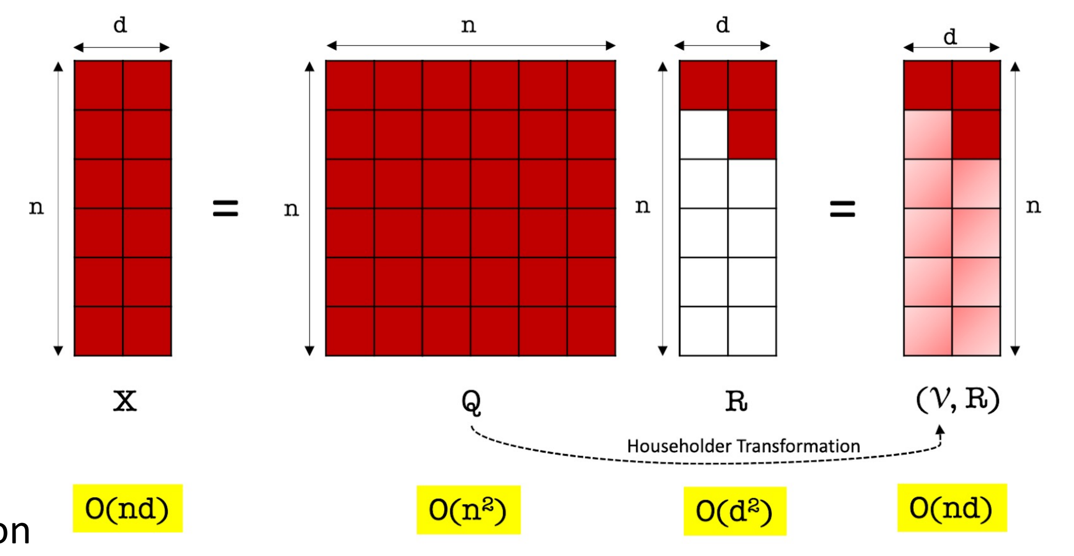
Test and check validity of the above claims against QR decomposition as a candidate for data summary via extensive theoretical and empirical analysis missing in literature

We pose **TWO** questions:

1. *Whether a classical and simple approach such as QR decomposition could (theoretically) accurately solve and accelerate common LMS solvers compared to the SOTA recursive and clustering-based fusion algorithm ?*
2. *Whether a numerically stable algorithm could generate accurate distributed sketches via exact factorization on streaming data ?*

Householder-QR [2]

$X=QR$, where, $Q^TQ=QQ^T=I$



Householder-Sketch

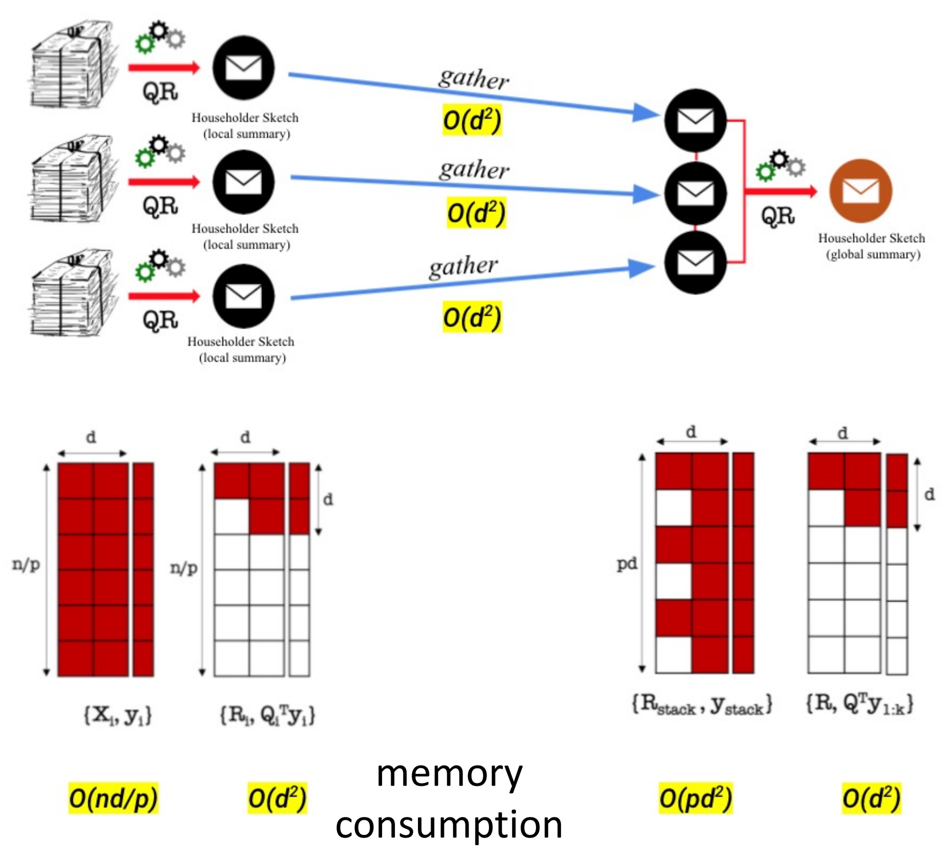
$$\underbrace{\|\mathbf{X}\mathbf{w}-\mathbf{y}\|_2}_{\text{(LMS)}}=\|\mathbf{QR}\mathbf{w}-\mathbf{y}\|_2=\|\mathbf{QR}\mathbf{w}-\mathbf{Q}\mathbf{Q}^T\mathbf{y}\|_2=\|\mathbf{Q}\|_2\,\|\mathbf{R}\mathbf{w}-\mathbf{Q}^T\mathbf{y}\|_2=\underbrace{\|\mathbf{R}\mathbf{w}-\mathbf{Q}^T\mathbf{y}\|_2}_{\text{(LMS-QR)}}$$

Accurate Sketch
$R^T R = X^T X$

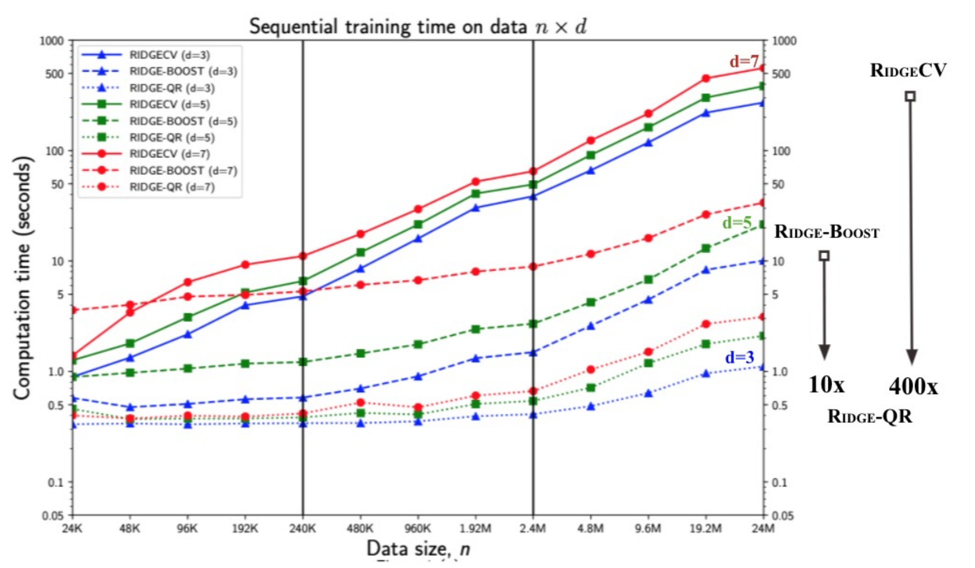
Distributed Householder-Sketch

$$\mathbf{X}=\begin{pmatrix}\mathbf{X}_1\\ \mathbf{X}_2\\ \vdots\\ \mathbf{X}_p\end{pmatrix}=\begin{pmatrix}\mathbf{Q}_1\mathbf{R}_1\\ \mathbf{Q}_2\mathbf{R}_2\\ \vdots\\ \mathbf{Q}_p\mathbf{R}_p\end{pmatrix}=\mathrm{diag}(\mathbf{Q}_1,\ldots,\mathbf{Q}_p)\begin{pmatrix}\mathbf{R}_1\\ \mathbf{R}_2\\ \vdots\\ \mathbf{R}_p\end{pmatrix}\qquad \mathbf{R}_{stack}=\begin{pmatrix}\mathbf{R}_1\\ \mathbf{R}_2\\ \vdots\\ \mathbf{R}_p\end{pmatrix}=\mathbf{Q}_M\mathbf{R}_M$$

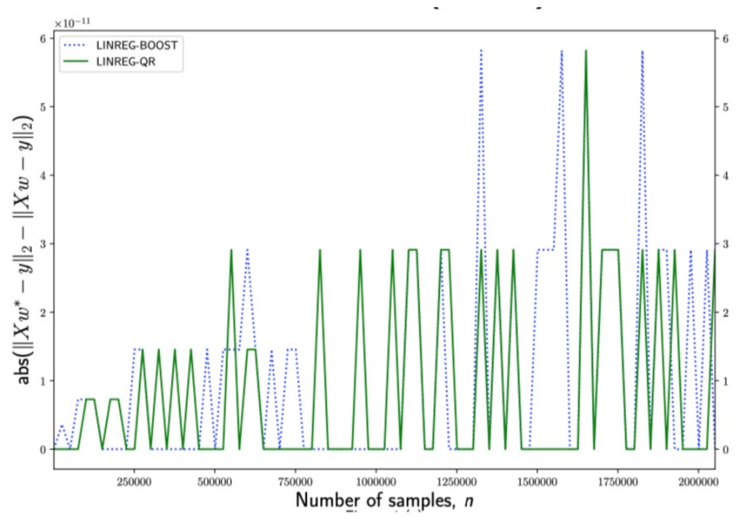
$$\mathbf{X}=\mathrm{diag}(\mathbf{Q}_1,\ldots,\mathbf{Q}_p)\mathbf{R}_{stack}=\underbrace{\mathrm{diag}(\mathbf{Q}_1,\ldots,\mathbf{Q}_p)\mathbf{Q}_M}_{\mathbf{Q}}\underbrace{\mathbf{R}_M}_{\mathbf{R}}$$



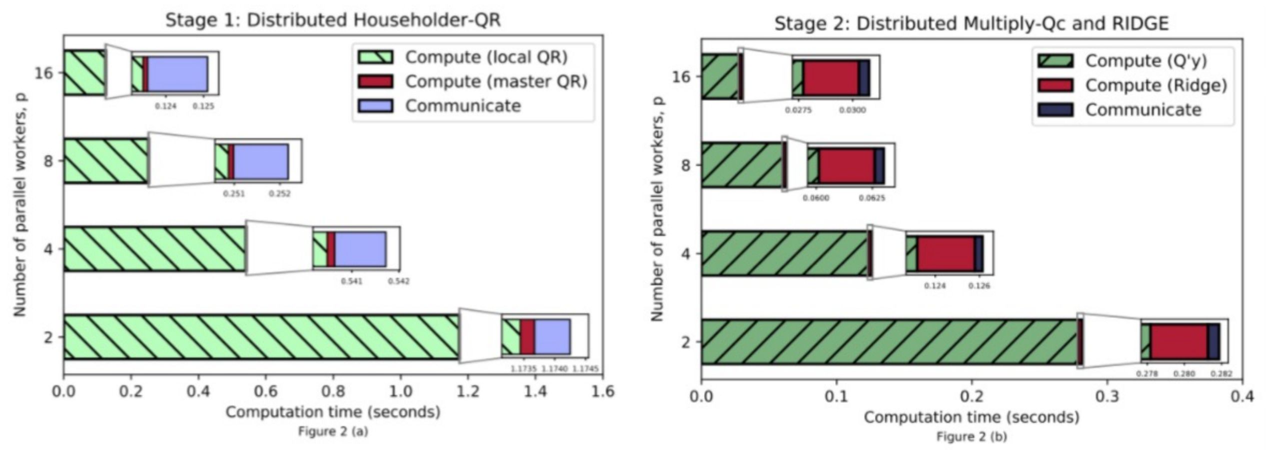
Results – proving Claims to be False



Sequential Training Time (Ridge-QR vs others)



Accuracy ($\times 10^{-11}$) comparison of LinReg-QR and LinReg-Boost on Household Power Consumption dataset (~ 2M × 8), $w*$ is solution from scikit-learn *LinearRegression*



Linear scalability with negligible communication overhead

Execution Time breakdown of Distributed Ridge-QR (on 10M x 10) with zoomed insets depicting communication time

References

1. Maalouf, A., Jubran, I., and Feldman, D. "*Fast and accurate least-mean-squares solvers*". in Advances in Neural Information Processing Systems, pp. 8305–8316, 2019
2. Golub, G. H. and Van Loan, C. F. "*Matrix computations*", volume 3. JHU press, 2012.