Leveraging Well-Conditioned Bases: Streaming and Distributed Summaries in Minkowski p-Norms

Graham Cormode[1,2] Charlie Dickens[1,2], David P. Woodruff[3]

[1]University of Warwick, [2]The Alan Turing Institute, [3]Carnegie Mellon University

Motivation

There are many efficient randomized streaming and distributed algorithms for problems such as approximate matrix multiplication, low rank approximation, and regression.

We explore whether similar guarantees for these problems can be achieved using only deterministic algorithms.

Prior deterministic algorithms focus only on the 2-norm [1].

We give the first efficient deterministic distributed and streaming algorithms for ℓ_p-regression for any $p \geq 1$, including $p = \infty$.

We exploit the sketch-and-solve paradigm to prove approximation guarantees for problems such as regression.

Our experimental results validate the approach and show that it can be beneficial in practice.

Obtaining a Summary in a Data Stream

We use a *well-conditioned basis* [2] (wcb) which has properties similar to an orthonormal basis.

A matrix U is a wcb for the column space of A if $\|U\|_p \leq \alpha$ and for all z, $\|z\|_q \leq \beta \|Uz\|_q$ where q is the dual norm to p, and α and β are at most $\texttt{poly}(d)$ [2].

Let $U = AR$ for change of basis matrix R. Then the *full ℓ_p-leverage scores* are $w_i = \|(AR)_i\|_p^p$.

Our summary is obtained by reading blocks of the matrix, keeping rows of high leverage and pruning out rows of low leverage.

Problem: How do ℓ_p-leverage scores change when they are computed on a block of the input instead of the full input?

The hue of each row in Figures 1 and 2 indicates the leverage of that particular row (darker is higher leverage).

Figure 1: Global Leverage Scores

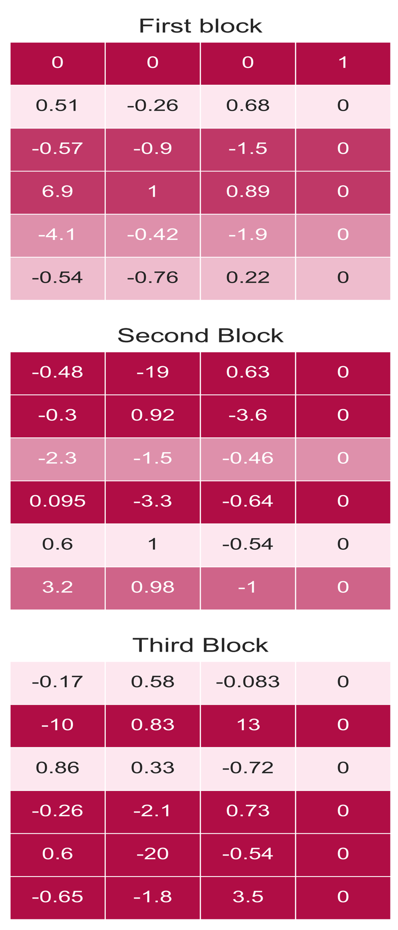

Figure 2: Local Leverage Scores

We prove that rows with high global leverage scores have high local leverage scores up to $\texttt{poly}(d)$ factors, $\hat{w}_i \geq w_i / \texttt{poly}(d)$.

The sum of the leverage scores is bounded by $\texttt{poly}(d)$ so there cannot exist too many rows of high leverage which ensures the summary does not grow too large.

ℓ_∞-Regression and Other Results

Problem: ℓ_∞-regression: Find $\varepsilon \|b\|_p$ additive error approximation to $\min_{x \in \mathbb{R}^d} \|Ax - b\|_\infty$.

Idea: Store all rows in A of high leverage and set all others to zero, call this A'. Keep b on these indices, otherwise set to zero, call this b'.

Solve the ℓ_∞-regression on A' and b'. This provides a good approximation (theoretically and empirically) as we can prove low leverage rows which are dropped do not contribute much to the ℓ_∞ cost.

High leverage rows are kept in their entirety so ℓ_∞ cost is same as in full problem.

Low leverage rows known not to contribute much to ℓ_∞ score so don't lose much by setting to zero.

Lower bound shows that relative error approximation cannot be obtained in sublinear space via reduction to $\texttt{Indexing}$ in communication complexity.

Other Results: We also give deterministic small-space streaming and distributed algorithms for *ℓ_p-subspace embeddings, ℓ_p-regression, ℓ_p-low-rank approximation, and approximate matrix product.*

Experimental Evaluations

Methods: We tested the ℓ_∞-regression algorithm on US Census and YearsPredictionsMSD datasets by varying the space bound.

Compared ℓ_2 wcb (Orth), and ℓ_1 wcb (SPC3) [3] to baseline methods of no conditioning (Identity), and uniform sampling (Uniform).

Performance is similar in both datasets and the results for the US census data are presented. The census data has 5 million rows and 11 features.

Time Complexity: Figure 3 shows that there is only a small absolute time difference between finding the two well-conditioned bases.

Figure 4 shows that this difference is almost negligible over the entire stream.

The time to stream the data, generate a summary, and solve the approximate reduced problem is less than the time taken by the brute force solver.

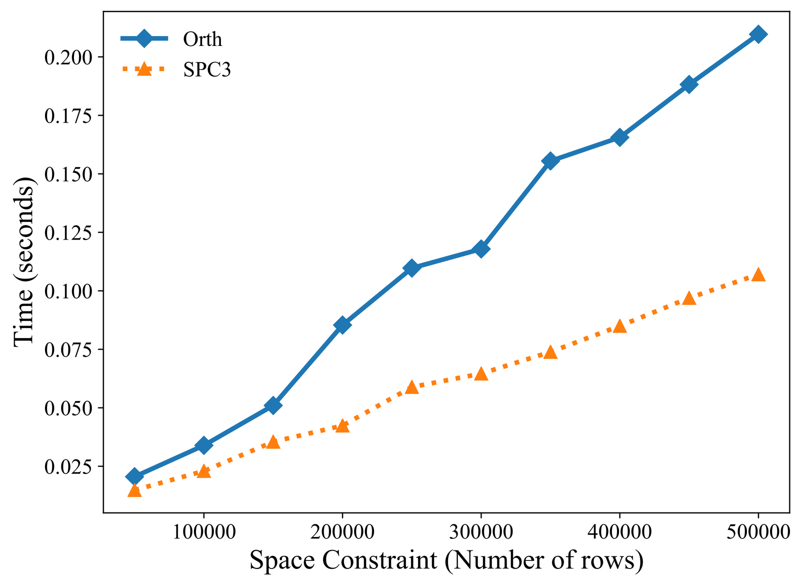

Figure 3: Time to compute WCB

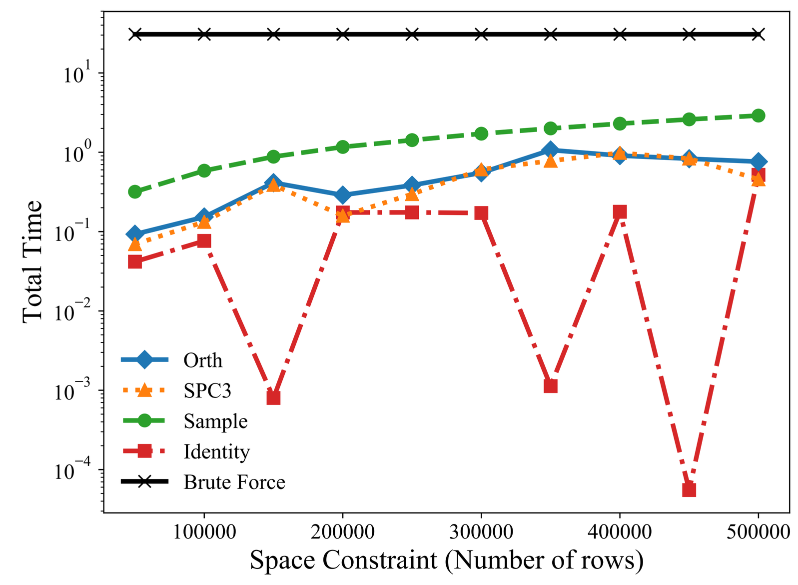

Figure 4: Total Time for WCB + Regression

Space Complexity: Figure 5 shows that both conditioning methods admit summaries which are significantly sublinear in the input size.

The Identity method always outputs a small summary and Sample always generates a summary of size equal to the space constraint, however, both of these can be suboptimal.

Figure 4 shows that the Sample summary is larger so takes longer to query, and Figure 6 shows Identity has no consistent error behaviour.

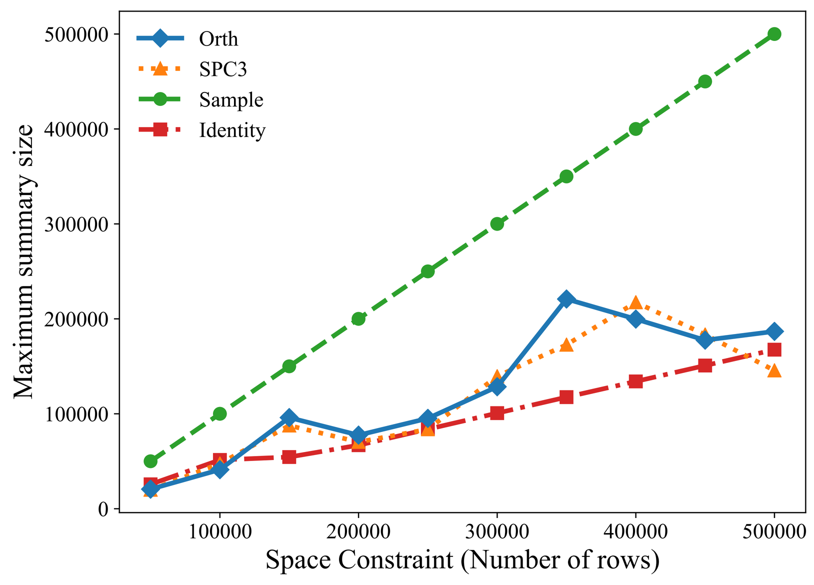

Figure 5: Space Complexity

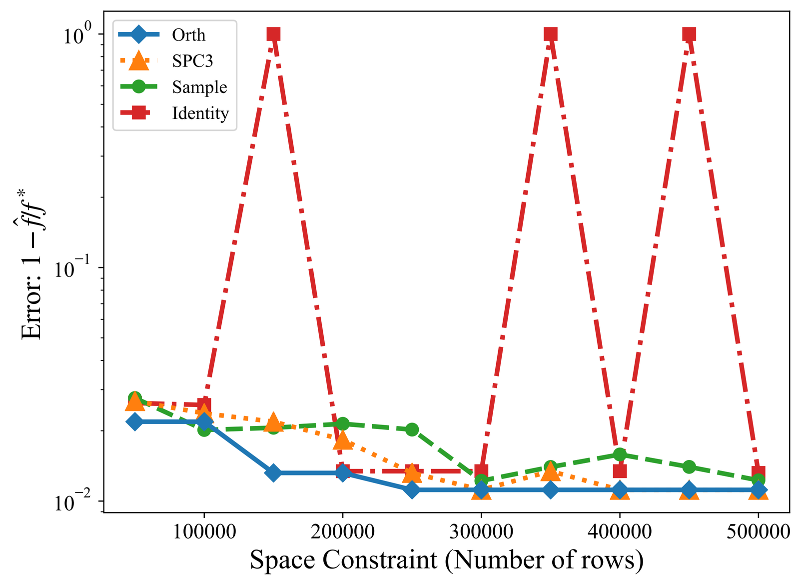

Figure 6: Error as Space is increased

Accuracy: The ℓ_∞-regression can be modelled as a linear program. The sketched LP has solution $\hat{f}$ and the true LP has solution f^*: we measure error as $1 - \hat{f}/f^*$.

Figure 6 shows that accurate summaries of the data which are a fraction of the total input size can be found by using the conditioning methods.

Summary: Conditioning methods perform well across each of space, time, and accuracy showing that they are a robust method for generating accurate and fast summaries!

References

[1]. Edo Liberty. Simple and deterministic matrix sketching. In *The 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD 2013, Chicago, IL, USA, August 11-14, 2013*, pages 581–588, 2013.

[2]. A. Dasgupta, P. Drineas, B. Harb, R. Kumar, and M. W. Mahoney. Sampling algorithms and coresets for ℓ_p-regression. In *Proc. of the 19th Annual SODA, pages 932–941, 2008.*

[3] Xiangrui Meng and Michael W. Mahoney. Low-distortion subspace embeddings in input-sparsity time and applications to robust linear regression. In *Symposium on Theory of Computing Conference, STOC'13, Palo Alto, CA, USA, June 1-4, 2013, pages 91–100, 2013.*

Acknowledgments

GC and CD acknowledge support from the European Research Council grant ERC-2014-CoG 647557 & The Alan Turing Institute under EPSRC grant EP/N510129/1. DW acknowledges the support from the National Science Foundation under Grant No. CCF-1815840.

LaTeX TikZposter