Sketching Meth

Yuqi Liu,

Intro, uses of RNL

Randomized Numerical Linear Algebra (RNLA

uses randomization to solve traditional numerical

hods for Randomized

, Leon Mikulinsky, Konstanti

LA

A) is a novel technique that

linear algebra problems. At

Figure 3: Least Squares for i.i.d. Gaussian $A^{1024x50}$

d Linear Algebra

tin Zörner

Applications

n b^{1024},

Figure 4: Low-rank approximation on Breast-MNIST Dataset

uses randomization to solve traditional numerical
its core, for a linear algebra problem featuring a
projecting ("sketching") **A** to **SA** using a sketching
an $(1\pm\varepsilon)\ell_p$ embedding with high probability, that is,

$$(1-\varepsilon)||\mathbf{x}||^2_p \leq ||\mathbf{Sx}||^2_p \leq (1+\varepsilon$$

After sketching, the classical deterministic algo
is run on **SA,** producing the desired result (ske
applications in ML, data science, signal processing

Sketching Method

Type	Sketching Algorithm
*Orthogonal [JL84]	Generate an entrywise i.i.d. stan normal matrix, run QR, and store
*Normal/ Uniform	Generate entrywise i.i.d. standar normal matrix / uniform over [-1

linear algebra problems. At
a matrix **A**, RNLA works by
g matrix **S** that functions as
, for all vectors **x**,

$\varepsilon)||\mathbf{x}||^2_p$

orithm to solve the problem
etch-and-solve). RNLA has
ng, statistics, and more.

ds

	Notes
ndard e Q.	Much slower than others due to QR.
rd 11	In high-dim, equivalent to

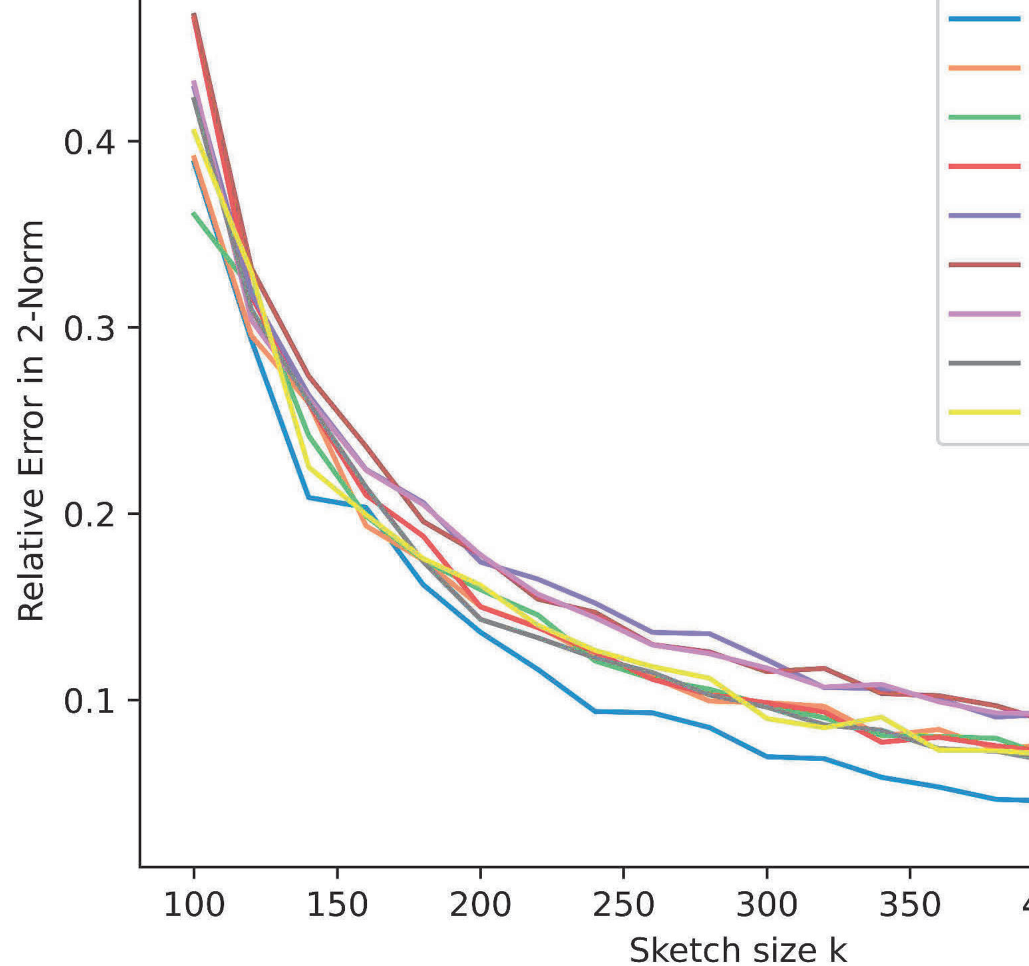
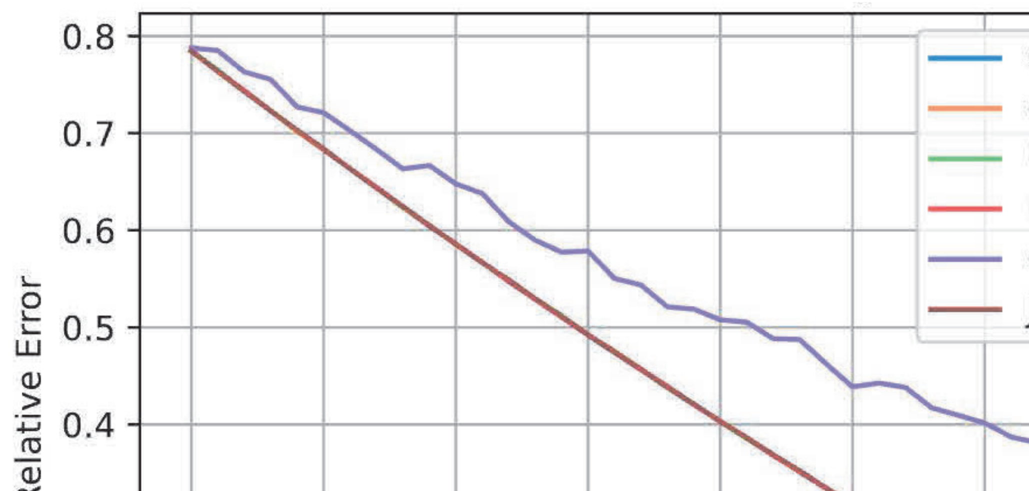
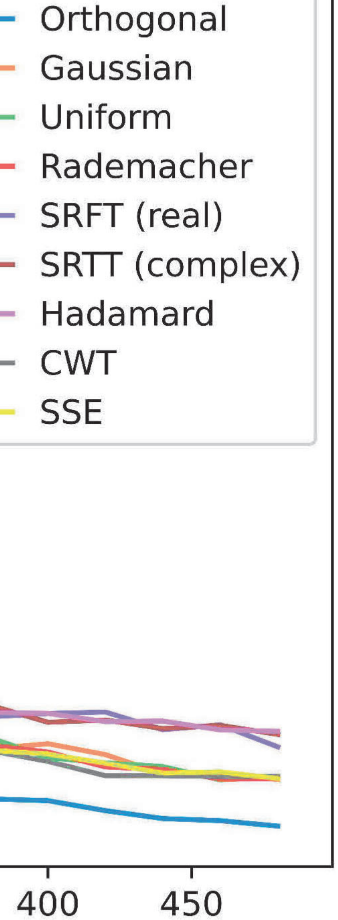
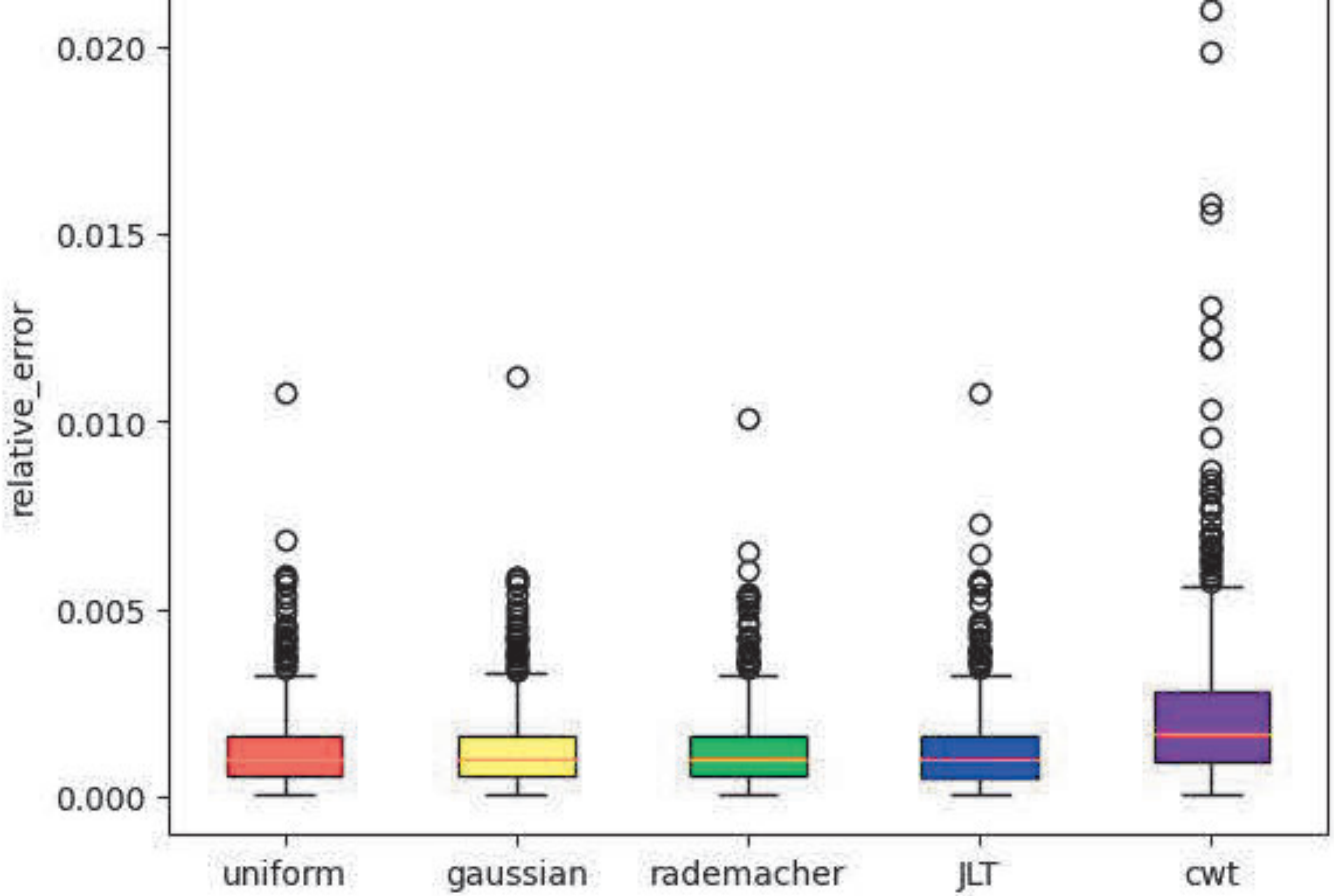


Figure 5: Low-rank approximation for i.i.d Gau







aussian $A^{1024x512}$

<u>Problem types:</u>

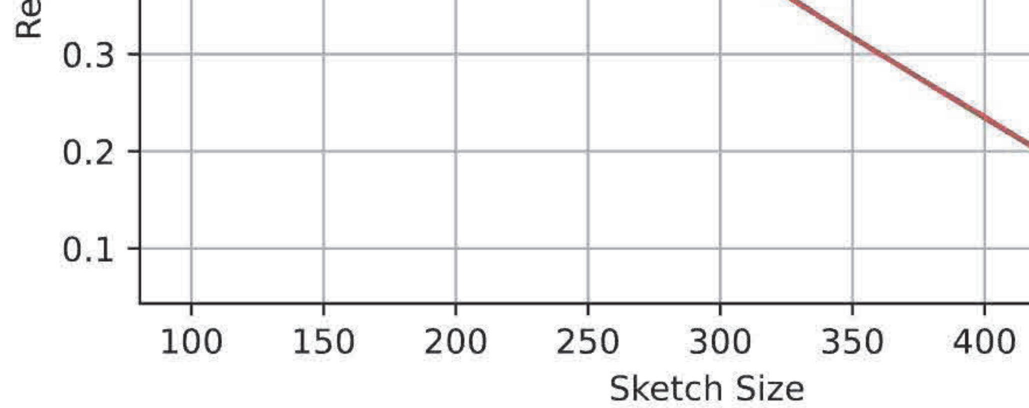
Overdetermined least squares:

$$\min_{\mathbf{x}} ||\mathbf{Ax} - \mathbf{b}||_2 \text{ for } \mathbf{A}^{mxn}, m > n$$

Low rank approximation:

Uniform [DG03]	normal matrix / uniform over [-1,
*Rademacher [Ach03]	Generate a matrix with entries of each ½ probability.
*Subsampled Random Trigonometric Transforms (SRTT) [Mur+23]	S = RTD, D diagonal with entries uniformly distributed on unit circl ±1 in real case), T trigonometric transform (Hadamard, discrete Fourier, discrete sine…), R samp random subset of k columns fron
Clarkson-Woodruff Transform (CWT) [CW12]	Generate a matrix that has one nonzero element in each column is ±1, each with equal probability
Sparse Sign Embedding (SSE)	Given a sparsity parameter s, generate a matrix that has s non elements in each column that are

, 1].	equivalent to rademacher [OT17].
of ± 1,	No multiplication in the matmul.
s le (or ples a m m.	Operates column-wise on A, so for $\mathbf{A}^{m \times n}$, $S^{k \times m}$, sketching is at most $O(km \log n)$, not requiring matmul.
n that y.	S is sparse, so sketching takes $O(nnz(A))$.
nzero re	S is sparse, so sketching takes $O(nnz(A))$.



Least Squares:

The sketching methods car (compare Figure 3):

1. Orthogonal sketching r
2. Sparse embeddings (S uniform, Rademacher are preferable due to t
3. Trigonometric Transfo but they preserve norn

$$\operatorname{argmin}_{\operatorname{rank}(\mathbf{A}') \leq k} \|\mathbf{A} - \mathbf{A}'\|_2$$
$$\text{for } k \ll \min(m, n)$$

Note: In Figure 5, all lines but CWT overlap.

Conclusion

n be divided into three groups by their accuracy

minimizes error & distortion but is expensive.

(SSE & CWT) & entrywise i.i.d. Matrices (normal, r) have similar accuracy, but sparse embeddings their lower computational cost.

orms have the lowest accuracy for least squares, ms better than any other method (Figure 1)

(SSE) [Hu+21]	elements in each column that are $\pm 1/\sqrt{s}$, each with equal probabilit

* these methods may require rescaling the sketc embedding property

Properties of sketching n

In practice, guaranteeing that the embedding always necessary, instead it might be sufficient picked) to ensure that relative norms are preserve

$$||\mathbf{Sx}||_2||\mathbf{u}||_2 \approx ||\mathbf{Su}||_2\ ||\mathbf{x}$$

We showcase this property in Figure 1. Stro distortion ε, which is the smallest number so that

$$(1-\varepsilon)||\mathbf{x}|| \leq ||\mathbf{Sx}|| \leq (1+\varepsilon)$$

holds for all $\mathbf{x} \in \mathrm{col}(\mathbf{A})$. In Figure 2 we depict the above sketching methods

re ty.	O(nnz(A)).

ching matrix to preserve the

matrices

g property holds true is not (as in the applications we ed, i.e.

$\mathbf{x}||_2$

ongly related to that is the

$)||\mathbf{x}||$

ne distortion for some of the

but they preserve norm

The same structuring can b distortion for rating sketching

Low rank approximation:

The non-sparse sketching m 1), with the gaussian sketc Gaussian dataset and the sketching method does signi

[Ach03] D. Achlioptas. “Databa Lindenstrauss with binary coin 66.4 (2003), pp. 671–687.

[CW12] Kenneth L. Clarkson a Regression in Input Sparsity T 1207.6365. url: http://arxiv.org/

ms better than any other method (Figure 1)

be inferred from Figure 2, showing the power of g methods.

methods are all roughly equivalent (Figure 4, Table ch matrix doing slightly better. For both the i.i.d. (poorly-conditioned) MNIST dataset, the sparse nificantly worse than all others.

References

base-friendly random projections: Johnson-ns". In: Journal of Computer and System Sciences

and David P. Woodruff. "Low Rank Approximation and Time". In: CoRR abs/1207.6365 (2012). arXiv: g/abs/1207.6365.

above sketching methods.

Figure 1: Embedding Property

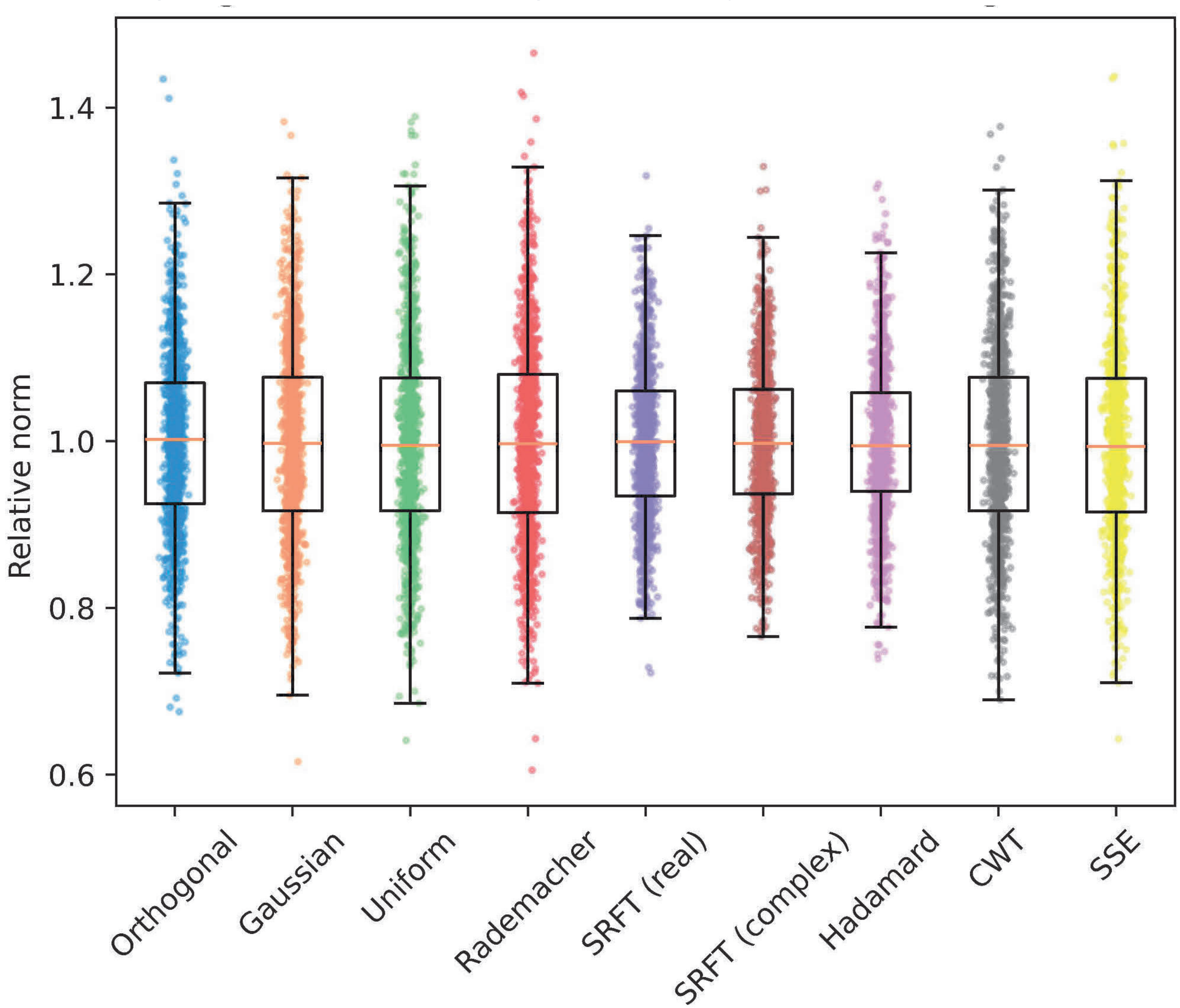
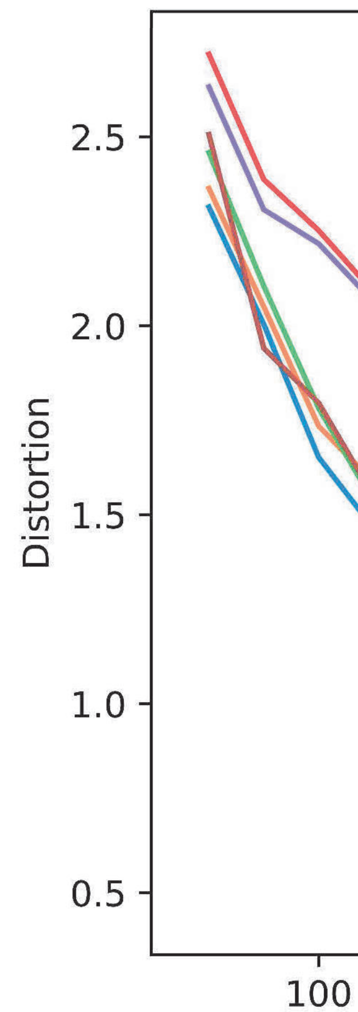
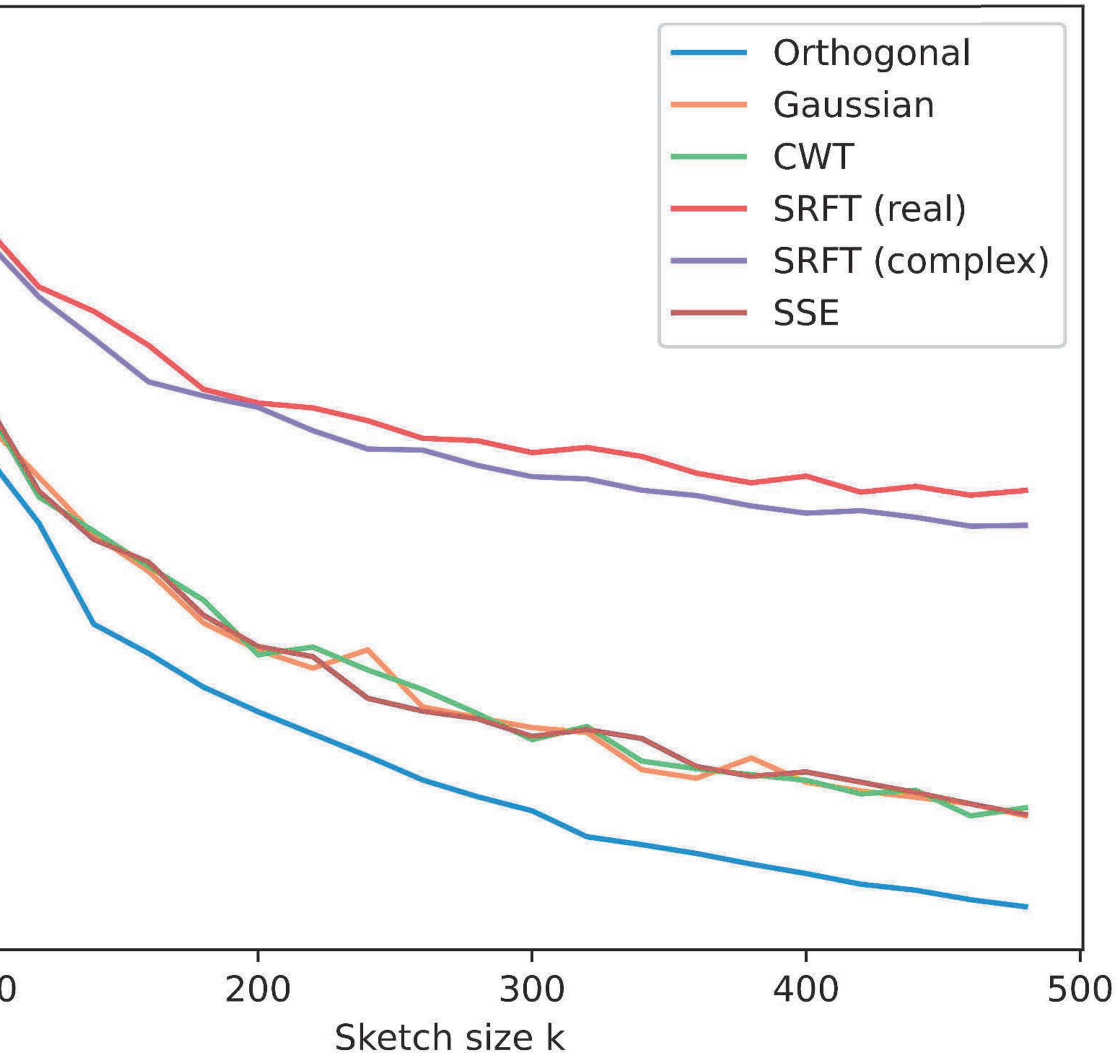


Figure 2:



2: Distortion



[DG03] S. Dasgupta and A. Gu
and Lindenstrauss". In: Rando

[Hu+21] Dong Hu et al. "Spars
Algebra". In: ICASSP 2021 - 2
Speech and Signal Processing
10.1109/ICASSP39728.2021.9

[JL84] W. Johnson and J. Lind
Hilbert space". In: Contempora

[Mur+23] Riley Murray et al. R
on the Field With an Eye to Sc
https://arxiv.org/abs/2302.1147

[OT17] S. Oymak and J. A. Tro
reduction, with applications". In
IMA 7.3 (2017), pp. 337–446.

Gupta. “An elementary proof of a theorem of Johnson
om Structures and Algorithms 22.1 (2003), pp. 60–65.

se Graph Based Sketching for Fast Numerical Linear
2021 IEEE International Conference on Acoustics,
ng (ICASSP). 2021, pp. 3255–3259. doi:
9414030.

denstrauss. “Extensions of Lipshitz mapping into
rary Mathematics 26 (1984), pp. 189–206.

Randomized Numerical Linear Algebra : A Perspective
oftware. 2023. arXiv: 2302.11474 [math.NA]. url:
74.

ropp. “Universality laws for randomized dimension
In: Information and Inference: A Journal of the