Exponentially Improved Dimensionality Reduction for ℓ_1 :

Subspace Embeddings and Independence Testing

Yi Li[1], David P. Woodruff[2], Taisuke Yasuda[2]

Nanyang Technological University[1], Carnegie Mellon University[2]

Overview

A celebrated result of Johnson and Lindenstrauss [JL84] gives highly efficient dimension reduction for the ℓ_2 norm. That is, a $r \times n$ Gaussian matrix S for $r = \Theta(\varepsilon^{-2}\log\delta^{-1})$ satisfies $\|Sx\|_2 = (1 \pm \varepsilon)\|x\|_2$ with probability at least $1-\delta$. However, the ℓ_1 norm is often more appropriate than the ℓ_2:

- For data mining applications, the ℓ_1 norm is more robust to outliers than the ℓ_2 norm
- The ℓ_1 norm is twice the total variance distance for distributions

We thus seek analogous results to the Johnson-Lindenstrauss lemma for the ℓ_1 norm.

Subspace Embeddings

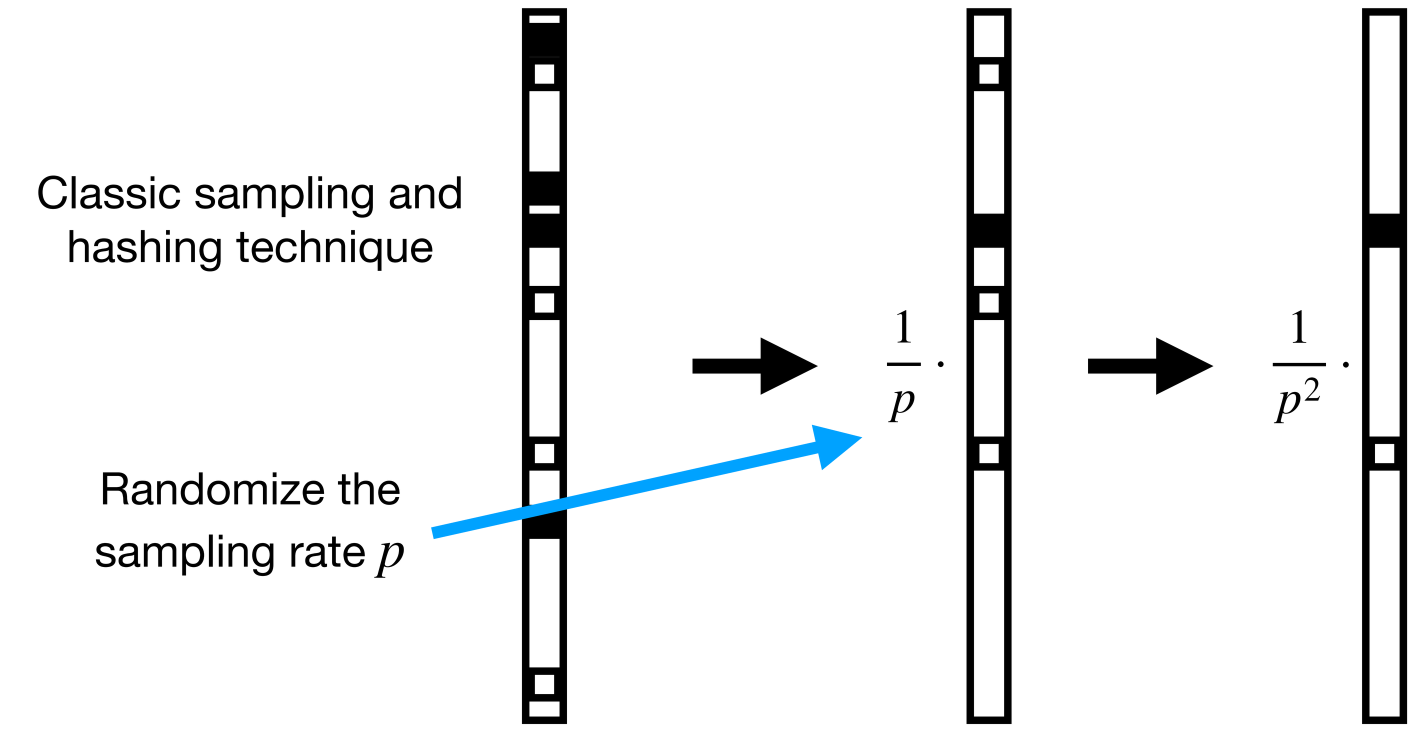
The known ℓ_1 analogue for the Johnson—Lindenstrauss lemma, i.e. linear oblivious dimension reduction maps S with $\|Sx\|_1 = (1 \pm \varepsilon)\|x\|_1$, has a doubly exponential dependency on ε^{-1} [WW19]. For subspace embeddings, i.e. the above guarantee for every vector in a d-dimensional subspace, [WW19] also achieves a doubly exponential dependence on d, ε^{-1}, and also shows a singly exponential lower bound in d. Can we close this gap? **We answer this question in the affirmative.**

	ℓ_2 [JL84]	ℓ_1 LB [WW19]	ℓ_1 UB [WW19]	ℓ_1 UB [LWY21]
1 vector	ε^{-2}		$2^{2^{\varepsilon^{-2}}}$	$2^{\varepsilon^{-1}}$
m vectors	$\varepsilon^{-2}\log m$	$2^{\sqrt{m}}$	$2^{2^{\varepsilon^{-2}\log m}}$	$2^{\varepsilon^{-1}m}$
d-dim subspace	$\varepsilon^{-2}d$	$2^{\sqrt{d}}$	$2^{2^{\varepsilon^{-2}d}}$	$2^{\varepsilon^{-1}d}$

*Suppresses big Oh and log factors

Idea for Singly Exponential Dependence in ε^{-1}

To achieve our singly exponential dependence on ε^{-1}, we start with the M-sketch construction of [CW14]. M-sketch is based on sampling and hashing the coordinates of x, and achieves a distortion of $O(1)$. We modify this construction by **randomizing the sampling rates** themselves to achieve our result.



Idea for Singly Exponential Dependence in d

To achieve our singly exponential dependence on d, we cannot afford to union bound over a net, as done by [JL84]. Instead, we apply our earlier result on the ℓ_1 **leverage score vector** with distortion $(1 + \varepsilon/d)$.

Clarkson—Drineas—Magdon-Ismail—Mahoney—Meng—Woodruff (2013)

Let $A \in \mathbb{R}^{n\times d}$. Then, there exists a vector $\lambda = \lambda(A) \in \mathbb{R}^n$ such that $\|\lambda\|_1 = 1$ and for all $x \in \operatorname{span}(A)$,

$$\frac{|x_i|}{\|x\|_1} \le d \cdot \lambda_i$$

for every $i \in [n]$.

Independence Testing

Consider a distribution given by a data stream:

- Each stream element is $(i_1, \ldots, i_q)$ for $i_j \in [d]$
- **Empirical joint distribution P:**

$$p(i_1, \ldots, i_q) = \frac{\text{number of occurrences of } (i_1, \ldots, i_q)}{\text{length of stream}}$$

- **Empirical product distribution $Q = Q_1 \times \ldots \times Q_q$:**

$$q_j(i) = \frac{\text{number of occurrences of } (*, \ldots, *, i, *, \ldots, *)}{\text{length of stream}}$$

Our task is to estimate $\|P - Q\|_1$.

The previous known algorithm for this problem has a doubly exponential dependence on q:

Braverman—Ostrovsky (2010)

$\|P - Q\|_1$ can be estimated in

$$(\varepsilon^{-1}\log d)^{q^{O(q)}}$$

space.

We improve this to a singly exponential bound:

Li—Woodruff—Yasuda (2021)

$\|P - Q\|_1$ can be estimated in

$$2^{O(q^2)}\left(q\varepsilon^{-1}\log d\right)^{O(q)}$$

space.

Our Construction

We design a sketching matrix S which takes the tensor product structure $S = (S^1) \otimes \ldots (S^q)$, where each S^j sketches each mode of the tensor. This allows us to maintain $S^1Q_1, \ldots, S^qQ_q$ in the stream and compute $SQ = (S^1Q_1) \otimes \ldots \otimes (S^qQ_q)$. We also maintain SP, then estimate $\|P - Q\|_1$ based on $SP - SQ = S(P - Q)$.

As with the subspace embedding, our sketch construction for a single mode starts with sampling and hashing techniques. This is used recursively to handle all q modes.

References

- [BO10] Braverman, Vladimir, and Rafail Ostrovsky. "Measuring independence of datasets." *Proceedings of the forty-second ACM symposium on Theory of computing*. 2010.
- [CDMMMW13] Clarkson, Kenneth L., et al. "The fast cauchy transform and faster robust linear regression." *SIAM Journal on Computing* 45.3 (2016): 763-810.
- [CW14] Clarkson, Kenneth L., and David P. Woodruff. "Sketching for M-estimators: A unified approach to robust regression." *Proceedings of the twenty-sixth annual ACM-SIAM symposium on Discrete algorithms*. Society for Industrial and Applied Mathematics, 2014.
- [JL84] Johnson, William B., and Joram Lindenstrauss. "Extensions of Lipschitz mappings into a Hilbert space 26." *Contemporary mathematics* 26 (1984).
- [WW19] Wang, Ruosong, and David P. Woodruff. "Tight bounds for ℓp oblivious subspace embeddings." *Proceedings of the Thirtieth Annual ACM-SIAM Symposium on Discrete Algorithms*. Society for Industrial and Applied Mathematics, 2019.