Differentially Private CountSketch

Improved utility analysis

Rasmus Pagh and Mikkel Thorup

NeurIPS 2022

CountSketch

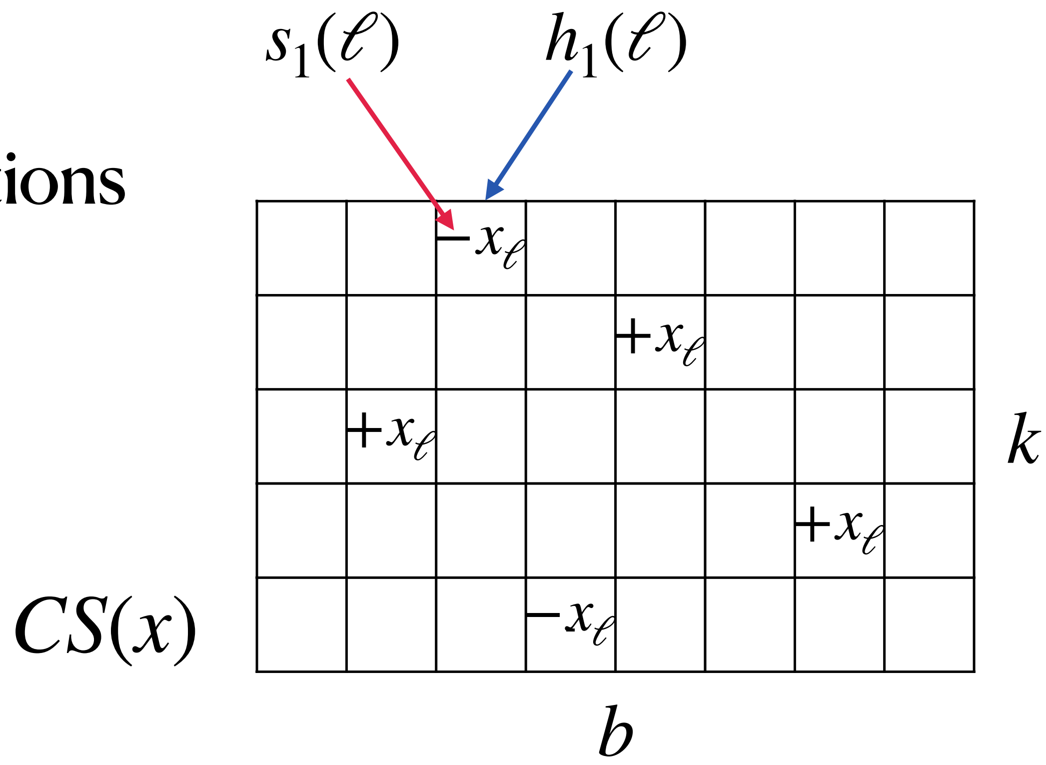
[Charikar, Chen, Farach-Colton 2002]

- Linear sketch, $CS : \mathbf{R}^d \to \mathbf{R}^{k \times b}$
- Defined using random hash functions

$$h_1, \ldots, h_k : [d] \to [b]$$

$$s_1, \ldots, s_k : [d] \to \{-1, +1\}$$

> **This talk**: Assume hash functions are *fully* independent



CountSketch estimator

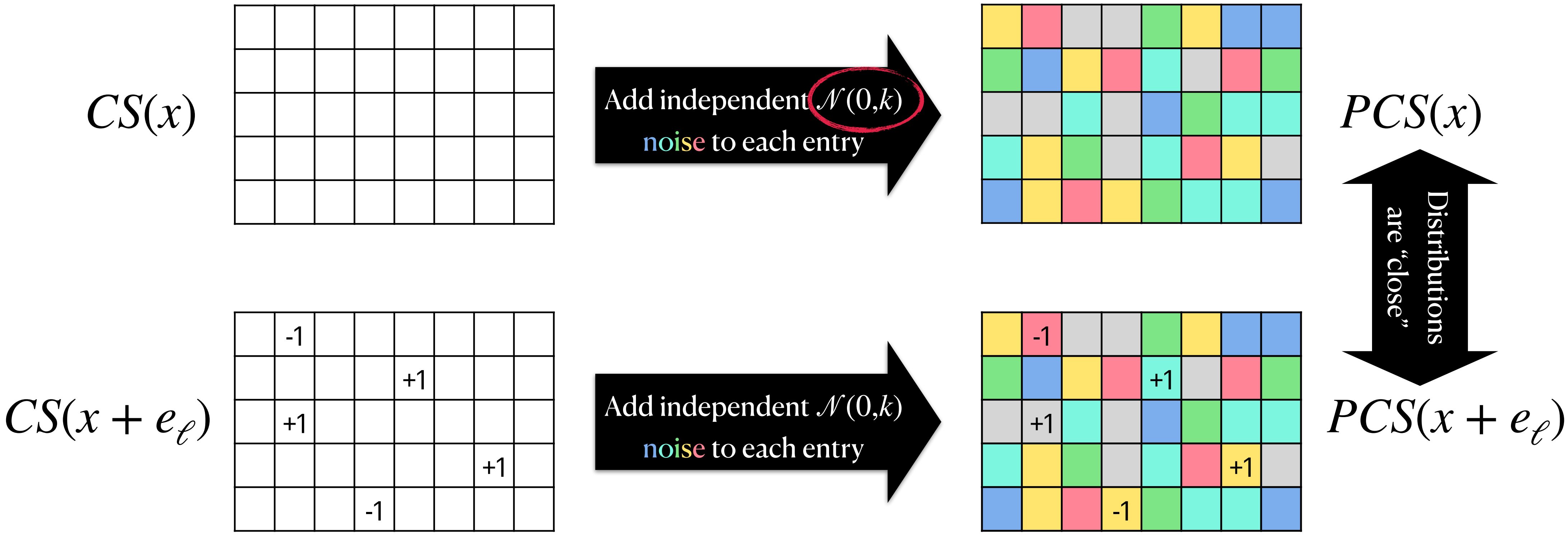
- Simple estimators: $s_1(\ell)CS(x)_{1,h_1(\ell)}, \ldots, s_k(\ell)CS(x)_{k,h_k(\ell)}$
- Median estimator: $\hat{x}_\ell = \text{median}(s_i(\ell)CS(x)_{i,h_i(\ell)} \mid i \in [k])$

Theorem (Minton & Price, 2014) For every $\alpha \in [0,1]$ and $\Delta = ||\text{tail}_b(x)||_2/\sqrt{b}$,

$$\Pr\left[|\hat{x}_\ell - x_\ell| > \alpha\,\Delta\right] < 2\exp\left(-\Omega\left(\alpha^2 k\right)\right) \quad ,$$

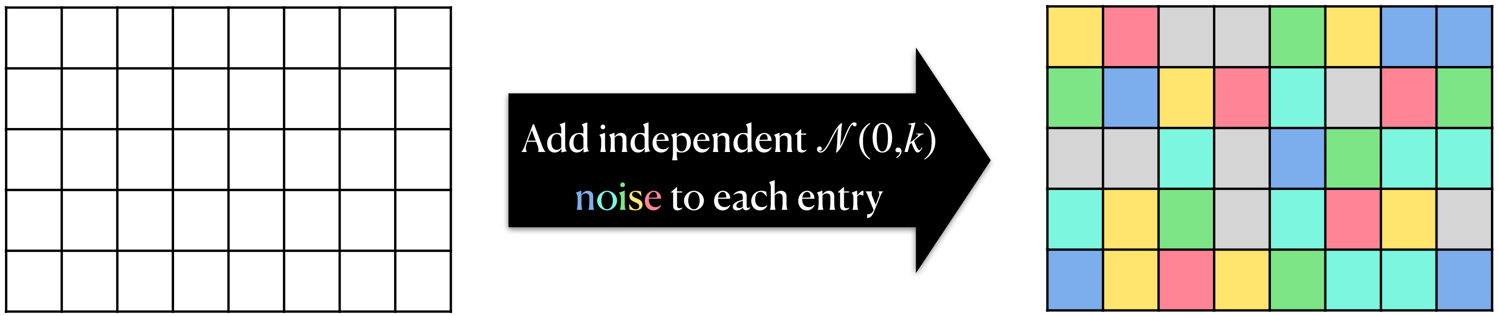
Δ is "maximum error of CountSketch"

Making CountSketch differentially private



Estimation from Private CountSketch

$CS(x)$



$PCS(x)$

$$\hat{x}_\ell = \text{median}(s_i(\ell)CS(x)_{i,h_i(\ell)} \mid i \in [k])$$

$$\bar{x}_\ell = \text{median}(s_i(\ell)PCS(x)_{i,h_i(\ell)} \mid i \in [k])$$

The question: How much worse is the private estimator $\bar{x}_\ell$ compared to $\hat{x}_l$?

Our result

Theorem For every $\alpha \in [0,1]$ and $\Delta = ||\text{tail}_b(x)||_2/\sqrt{b}$,

$$\Pr\left[|\bar{x}_\ell - x_\ell| > \alpha \max\{\Delta, \sigma\}\right] < 2\exp\left(-\Omega\left(\alpha^2 k\right)\right)$$

Low noise ($\sigma \leq \Delta$):
Same tail bound as CountSketch

High noise ($\sigma > \Delta$), $k = \sigma^2$:
Tail like $\mathcal{N}(0,1)$ noise + $\exp(-\Omega(k))$

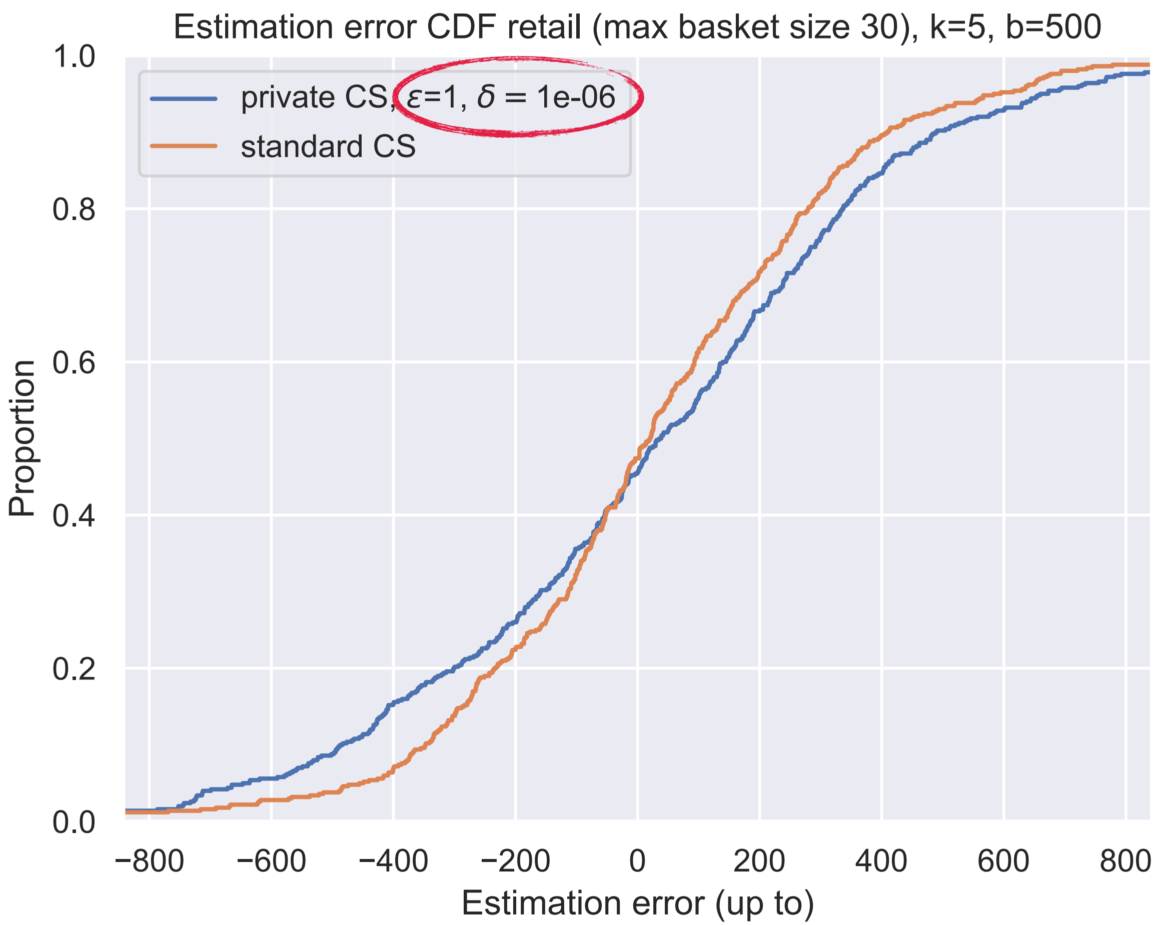
Message of our work: Estimation error of Private CountSketch is either the CountSketch error or the error needed for DP, whichever is larger

Proof ingredients

(about 1 page)

- Two cases:
  - Adding noise with $\sigma \leq \Delta$ maintains the probability of a good simple estimator up to a constant factor
  - Adding noise with $\sigma > \Delta$, the probability of a good simple estimator can be bounded up to a constant factor in terms of σ
- Lemma from Minton & Price, using symmetry of estimators, finishes the argument

Experiments — market basket data



Related work in NeurIPS 2022

Differentially Private Linear Sketches: Efficient Implementations and Applications

Fuheng Zhao[*,†]
fuheng_zhao@ucsb.edu

Dan Qiao[*,†]
danqiao@ucsb.edu

Rachel Redberg[*]
rredberg@ucsb.edu

Divyakant Agrawal[*]
agrawal@cs.ucsb.edu

Amr El Abbadi[*]
amr@cs.ucsb.edu

Yu-Xiang Wang[*]
yuxiangw@ucsb.edu