Other Results

- **Theorem**: For $0 < p < 2$, can approximate $F_p(x)$ up to $1 \pm \varepsilon$ using **optimal** $O(\varepsilon^{-2} \log n)$ bits of space and $O(\log n)$ **update time**

- Valid only for $\varepsilon < 1/n^c$

- Improves on $O(\log^2 n \log\log n)$ update time of [KNPW '11]

- **CountSketch:** Given t and r, a streaming algorithm which can compute $\hat{x}[i]$ such that for $\alpha \leq 1$

- Obtained by derandomizing [Minton and Price '14]

- The algorithm uses $O(tr\log(n) + \log^2 n)$ bits of space

- $O(r \log n)$ update time

$$\Pr\left[\,|x[i]-\hat{x}[i]| > \alpha\frac{\|x\|_2}{\sqrt{t}}\right] \le 2\exp(-\alpha^2 r) + 1/\mathrm{poly}(n)$$

39

Other Results

- **Theorem**: For $0 < p < 2$, can approximate $F_p(x)$ up to $1 \pm \varepsilon$ using **optimal** $O(\varepsilon^{-2} \log n)$ bits of space and $O(\log n)$ **update time**
  - Valid only for $\varepsilon < 1/n^c$
  - Improves on $O(\log^2 n \log \log n)$ update time of [KNPW '11]
- **CountSketch:** Given t and r, a streaming algorithm which can compute $\hat{x}[i]$ such that for $\alpha \leq 1$

$$\Pr[\,|x[i] - \hat{x}[i]| > \alpha \frac{\|x\|_2}{\sqrt{t}}] \leq 2\exp(-\alpha^2 r) + 1/\text{poly}(n)$$

  - Obtained by derandomizing [Minton and Price '14]
  - The algorithm uses $O(tr\log(n) + \log^2 n)$ bits of space
    - $O(r \log n)$ update time

Other Applications of Sketching

- **Classic**
  - Ridge Regression [KW, AISTATS '20], [KW, ICML '22]
  - Dimensionality Reduction for Sum-of-Distances [FKW, ICML '21]
  - Reduced Rank Regression [KW, COLT '21]
  - Fast and Small Subspace Embeddings [CCKW, SODA '22]
  - PolySketchFormer : Linear Time Transformers obtained via sketching Polynomial Kernels [KMZ, ICML '24]
  - Lower Bounds for Adaptive Matrix Recovery [KW , NeurIPS '23]
  - Fast algorithms for Schatten-p Low Rank Approximation [KW, '24]