Sparse Recovery

Application of Count-Sketch

Hu Fu @SHUFE, Oct 14, 2023

The Sparse Recovery Problem

The Sparse Recovery Problem

- Recall: In a streaming algorithm setting, we are sometimes interested in having a sparse vector that approximates the frequency vector $\mathbf{x} \in \mathbb{R}^d$

The Sparse Recovery Problem

- Recall: In a streaming algorithm setting, we are sometimes interested in having a sparse vector that approximates the frequency vector $\mathbf{x} \in \mathbb{R}^d$
  - A vector is sparse if it has few non-zero entries

The Sparse Recovery Problem

- Recall: In a streaming algorithm setting, we are sometimes interested in having a sparse vector that approximates the frequency vector $\mathbf{x} \in \mathbb{R}^d$
  - A vector is sparse if it has few non-zero entries
  - We may measure the quality of approximation by ℓ_2 distance

The Sparse Recovery Problem

- Recall: In a streaming algorithm setting, we are sometimes interested in having a sparse vector that approximates the frequency vector $\mathbf{x} \in \mathbb{R}^d$
  - A vector is sparse if it has few non-zero entries
  - We may measure the quality of approximation by ℓ_2 distance
- So given $k \in \mathbb{N}$ and $\epsilon \in (0, \frac{1}{2})$, we are interested in finding $\mathbf{y} \in \mathbb{R}^d$, with

The Sparse Recovery Problem

- Recall: In a streaming algorithm setting, we are sometimes interested in having a sparse vector that approximates the frequency vector $\mathbf{x} \in \mathbb{R}^d$
  - A vector is sparse if it has few non-zero entries
  - We may measure the quality of approximation by ℓ_2 distance
- So given $k \in \mathbb{N}$ and $\epsilon \in (0, \frac{1}{2})$, we are interested in finding $\mathbf{y} \in \mathbb{R}^d$, with
  - $\|\mathbf{y}\|_0 \leq k$

The Sparse Recovery Problem

- Recall: In a streaming algorithm setting, we are sometimes interested in having a sparse vector that approximates the frequency vector $\mathbf{x} \in \mathbb{R}^d$
  - A vector is sparse if it has few non-zero entries
  - We may measure the quality of approximation by ℓ_2 distance
- So given $k \in \mathbb{N}$ and $\epsilon \in (0, \frac{1}{2})$, we are interested in finding $\mathbf{y} \in \mathbb{R}^d$, with
  - $\|\mathbf{y}\|_0 \leq k$
  - $\|\mathbf{y} - \mathbf{x}\|_2 \leq (1+\epsilon)E_2^k(\mathbf{x})$, where $E_2^k(\mathbf{x}) := \min_{\mathbf{z} \in \mathbb{R}^d, \|z\|_0 \leq k} \|\mathbf{z} - \mathbf{x}\|_2$

The Sparse Recovery Problem

- Recall: In a streaming algorithm setting, we are sometimes interested in having a sparse vector that approximates the frequency vector $\mathbf{x} \in \mathbb{R}^d$
  - A vector is sparse if it has few non-zero entries
  - We may measure the quality of approximation by ℓ_2 distance
- So given $k \in \mathbb{N}$ and $\epsilon \in (0, \frac{1}{2})$, we are interested in finding $\mathbf{y} \in \mathbb{R}^d$, with
  - $\|\mathbf{y}\|_0 \leq k$
  - $\|\mathbf{y} - \mathbf{x}\|_2 \leq (1+\epsilon)E_2^k(\mathbf{x})$, where $E_2^k(\mathbf{x}) := \min_{\mathbf{z} \in \mathbb{R}^d, \|z\|_0 \leq k} \|\mathbf{z} - \mathbf{x}\|_2$

Quantifying the error using $E_2^k(\mathbf{x})$ is necessary. It can be as large as comparable to $\|\mathbf{x}\|_2$

Sparse Recovery with Count-Sketch

Sparse Recovery with Count-Sketch

- Offline optimum solution: pick the k entries of $\mathbf{x}$ with the largest absolute values

Sparse Recovery with Count-Sketch

- Offline optimum solution: pick the k entries of $\mathbf{x}$ with the largest absolute values
- Recall Count-Sketch:

Sparse Recovery with Count-Sketch

- Offline optimum solution: pick the k entries of $\mathbf{x}$ with the largest absolute values
- Recall Count-Sketch:
  - Draw $\ell = O(\log d)$ hash functions $h_1, \cdots, h_\ell : [d] \to [w]$, independently from a pairwise independent hash family

Sparse Recovery with Count-Sketch

- Offline optimum solution: pick the k entries of $\mathbf{x}$ with the largest absolute values
- Recall Count-Sketch:
  - Draw $\ell = O(\log d)$ hash functions $h_1, \cdots, h_\ell : [d] \to [w]$, independently from a pairwise independent hash family
  - Draw ℓ hash functions $g_1, \cdots, g_\ell : [d] \to \{-1, +1\}$, independently from a pairwise independent hash family

Sparse Recovery with Count-Sketch

- Offline optimum solution: pick the k entries of $\mathbf{x}$ with the largest absolute values
- Recall Count-Sketch:
  - Draw $\ell = O(\log d)$ hash functions $h_1, \cdots, h_\ell : [d] \to [w]$, independently from a pairwise independent hash family
  - Draw ℓ hash functions $g_1, \cdots, g_\ell : [d] \to \{-1, +1\}$, independently from a pairwise independent hash family
  - At input i_t, increase counter $C_j[h_j(i_t)]$ by $g_j(i_t)\Delta_t$, for $j = 1, \ldots, \ell$

Sparse Recovery with Count-Sketch

- Offline optimum solution: pick the k entries of $\mathbf{x}$ with the largest absolute values
- Recall Count-Sketch:
  - Draw $\ell = O(\log d)$ hash functions $h_1, \cdots, h_\ell : [d] \to [w]$, independently from a pairwise independent hash family
  - Draw ℓ hash functions $g_1, \cdots, g_\ell : [d] \to \{-1, +1\}$, independently from a pairwise independent hash family
  - At input i_t, increase counter $C_j[h_j(i_t)]$ by $g_j(i_t)\Delta_t$, for $j = 1, \ldots, \ell$
  - Output: for coordinate i, report $\tilde{x}_i :=$ median $\{g_j(i)C_j[h_j(i)]\}$

Sparse Recovery with Count-Sketch

- Offline optimum solution: pick the k entries of $\mathbf{x}$ with the largest absolute values
- Recall Count-Sketch:
  - Draw $\ell = O(\log d)$ hash functions $h_1, \cdots, h_\ell : [d] \to [w]$, independently from a pairwise independent hash family
  - Draw ℓ hash functions $g_1, \cdots, g_\ell : [d] \to \{-1, +1\}$, independently from a pairwise independent hash family
  - At input i_t, increase counter $C_j[h_j(i_t)]$ by $g_j(i_t)\Delta_t$, for $j = 1,\ldots,\ell$
  - Output: for coordinate i, report $\tilde{x}_i :=$ median $\{g_j(i)C_j[h_j(i)]\}$
- To solve sparse recovery, take $w = 3k/\epsilon^2$, take the k largest coordinates of $\tilde{\mathbf{x}}$

Sparse Recovery with Count-Sketch

- Offline optimum solution: pick the k entries of $\mathbf{x}$ with the largest absolute values
- Recall Count-Sketch:
  - Draw $\ell = O(\log d)$ hash functions $h_1, \cdots, h_\ell : [d] \to [w]$, independently from a pairwise independent hash family
  - Draw ℓ hash functions $g_1, \cdots, g_\ell : [d] \to \{-1, +1\}$, independently from a pairwise independent hash family
  - At input i_t, increase counter $C_j[h_j(i_t)]$ by $g_j(i_t)\Delta_t$, for $j = 1,\ldots,\ell$
  - Output: for coordinate i, report $\tilde{x}_i :=$ median $\{g_j(i)C_j[h_j(i)]\}$
- To solve sparse recovery, take $w = 3k/\epsilon^2$, take the k largest coordinates of $\tilde{\mathbf{x}}$

Note the dependence on k

Ideas of Proof

Ideas of Proof

- Main idea:

Ideas of Proof

- Main idea:
  - If we chose the k "correct" entries, since total error should be ϵE_2^k, the error allowed for each entry should be controlled to $\frac{\epsilon}{\sqrt{k}} E_2^k$

Ideas of Proof

- Main idea:
  - If we chose the k "correct" entries, since total error should be ϵE_2^k, the error allowed for each entry should be controlled to $\frac{\epsilon}{\sqrt{k}} E_2^k$
  - But the k entries we chose may differ from the "correct" ones. We should argue that, when all entries are estimated accurately enough, this doesn't introduce too much error.

Ideas of Proof

- Main idea:
  - If we chose the k "correct" entries, since total error should be ϵE_2^k, the error allowed for each entry should be controlled to $\frac{\epsilon}{\sqrt{k}} E_2^k$

Lemma. Count-Sketch with $w = 3k/\epsilon^2, \ell = O(\log n)$ guarantees $|x_j - \tilde{x}_j| \leq \frac{\epsilon}{\sqrt{k}} E_2^k$ for each j with high probability.

Ideas of Proof

- Main idea:
  - If we chose the k "correct" entries, since total error should be ϵE_2^k, the error allowed for each entry should be controlled to $\frac{\epsilon}{\sqrt{k}} E_2^k$
  - But the k entries we chose may differ from the "correct" ones. We should argue that, when all entries are estimated accurately enough, this doesn't introduce too much error.

Lemma. For $\mathbf{x}, \mathbf{y} \in \mathbb{R}^d$, if $\|\mathbf{x} - \mathbf{y}\|_\infty \leq \frac{\epsilon}{\sqrt{k}} E_2^k(\mathbf{x})$, let $\mathbf{z}$ be the k-sparse recovery of $\mathbf{y}$, then

$\|\mathbf{x} - \mathbf{z}\| \leq (1 + 5\epsilon) E_2^k(\mathbf{x})$

Proof of First Lemma

Lemma. Count-Sketch with $w = 3k/\epsilon^2, \ell = O(\log n)$ guarantees $|x_j - \tilde{x}_j| \le \frac{\epsilon}{\sqrt{k}} E_2^k$ for each j with high probability.

Proof of First Lemma

Lemma. Count-Sketch with $w = 3k/\epsilon^2, \ell = O(\log n)$ guarantees $|x_j - \tilde{x}_j| \leq \frac{\epsilon}{\sqrt{k}} E_2^k$ for each j with high probability.

Recall the analysis of Count-Sketch. For each $j \in [d]$, $i \in [\ell]$, the i-th estimate is $z_i := C_i[h_i(j)]g_i(j)$,

Proof of First Lemma

Lemma. Count-Sketch with $w = 3k/\epsilon^2, \ell = O(\log n)$ guarantees $|x_j - \tilde{x}_j| \leq \frac{\epsilon}{\sqrt{k}} E_2^k$ for each j with high probability.

Recall the analysis of Count-Sketch. For each $j \in [d], i \in [\ell]$, the i-th estimate is $z_i := C_i[h_i(j)]g_i(j)$, then $\mathbb{E}[z_i] = x_i$.

Proof of First Lemma

Lemma. Count-Sketch with $w = 3k/\epsilon^2$, $\ell = O(\log n)$ guarantees $|x_j - \tilde{x}_j| \le \frac{\epsilon}{\sqrt{k}} E_2^k$ for each j with high probability.

Recall the analysis of Count-Sketch. For each $j \in [d]$, $i \in [\ell]$, the i-th estimate is $z_i := C_i[h_i(j)]g_i(j)$, then $\mathbb{E}[z_i] = x_i$.

To apply Chebyshev's inequality, we bound the variance of z_i.

Proof of First Lemma

Lemma. Count-Sketch with $w = 3k/\epsilon^2, \ell = O(\log n)$ guarantees $|x_j - \tilde{x}_j| \leq \frac{\epsilon}{\sqrt{k}} E_2^k$ for each j with high probability.

Recall the analysis of Count-Sketch. For each $j \in [d]$, $i \in [\ell]$, the i-th estimate is $z_i := C_i[h_i(j)]g_i(j)$, then $\mathbb{E}[z_i] = x_i$.

To apply Chebyshev's inequality, we bound the variance of z_i.

Let $Y_{j,j'}$ be the indicator variable for the event $h_i(j) = h_i(j')$, then by pairwise independence of the hash family, $\mathbb{P}[Y_{j,j'}] = \frac{1}{w}$.

Proof of First Lemma

Lemma. Count-Sketch with $w = 3k/\epsilon^2, \ell = O(\log n)$ guarantees $|x_j - \tilde{x}_j| \leq \frac{\epsilon}{\sqrt{k}} E_2^k$ for each j with high probability.

Recall the analysis of Count-Sketch. For each $j \in [d]$, $i \in [\ell]$, the i-th estimate is $z_i := C_i[h_i(j)]g_i(j)$, then $\mathbb{E}[z_i] = x_i$.

To apply Chebyshev's inequality, we bound the variance of z_i.

Let $Y_{j,j'}$ be the indicator variable for the event $h_i(j) = h_i(j')$, then by pairwise independence of the hash family, $\mathbb{P}[Y_{j,j'}] = \frac{1}{w}$.

$$\text{Var}(z_i) = \mathbb{E}[(z_i - x_i)^2] = \mathbb{E}\left[\left(\sum_{j' \neq j} g_i(j)g_i(j')Y_{j,j'}x_{j'}\right)^2\right] = \sum_{j' \neq j} x_{j'}^2 \mathbb{E}[Y_{j,j'}^2] \leq \frac{\|x\|^2}{w}$$

Refining the Analysis

Refining the Analysis

- What is $E_2^k(\mathbf{x})$?

Refining the Analysis

- What is $E_2^k(\mathbf{x})$?
  - Let T be the set of k entries of $\mathbf{x}$ with the largest absolute values, then $E_2^k(\mathbf{x}) = \sqrt{\sum_{j' \notin T} x_{j'}^2}$

Refining the Analysis

- What is $E_2^k(\mathbf{x})$?
  - Let T be the set of k entries of $\mathbf{x}$ with the largest absolute values, then $E_2^k(\mathbf{x}) = \sqrt{\sum_{j' \notin T} x_{j'}^2}$
- The error introduced by collision with entries not in T is controllable by E_2^k

Refining the Analysis

- What is $E_2^k(\mathbf{x})$?
  - Let T be the set of k entries of $\mathbf{x}$ with the largest absolute values, then $E_2^k(\mathbf{x}) = \sqrt{\sum_{j' \notin T} x_{j'}^2}$
- The error introduced by collision with entries not in T is controllable by E_2^k
- What about collision with the entries in T?

Refining the Analysis

- What is $E_2^k(\mathbf{x})$?
  - Let T be the set of k entries of $\mathbf{x}$ with the largest absolute values, then $E_2^k(\mathbf{x}) = \sqrt{\sum_{j' \notin T} x_{j'}^2}$
- The error introduced by collision with entries not in T is controllable by E_2^k
- What about collision with the entries in T?
  - This is where we make use of $w = \Omega(k/\epsilon^2)$

Refining the Analysis

- What is $E_2^k(\mathbf{x})$?
  - Let T be the set of k entries of $\mathbf{x}$ with the largest absolute values, then $E_2^k(\mathbf{x}) = \sqrt{\sum_{j' \notin T} x_{j'}^2}$
- The error introduced by collision with entries not in T is controllable by E_2^k
- What about collision with the entries in T?
  - This is where we make use of $w = \Omega(k/\epsilon^2)$
  - With w growing linearly with k, this can be made to happen with small probability.

Proof of First Lemma

page 1

Lemma. Count-Sketch with $w = 3k/\epsilon^2, \ell = O(\log n)$ guarantees $|x_j - \tilde{x}_j| \leq \frac{\epsilon}{\sqrt{k}} E_2^k$ for each j with high probability.

Proof of First Lemma

page 1

Lemma. Count-Sketch with $w = 3k/\epsilon^2, \ell = O(\log n)$ guarantees $|x_j - \tilde{x}_j| \leq \frac{\epsilon}{\sqrt{k}} E_2^k$ for each j with high probability.

Recall the analysis of Count-Sketch. For each $j \in [d]$, $i \in [\ell]$, the i-th estimate is $z_i := C_i[h_i(j)]g_i(j)$, then $\mathbb{E}[z_i] = x_j$.

Proof of First Lemma

page 1

Lemma. Count-Sketch with $w = 3k/\epsilon^2, \ell = O(\log n)$ guarantees $|x_j - \tilde{x}_j| \leq \frac{\epsilon}{\sqrt{k}} E_2^k$ for each j with high probability.

Recall the analysis of Count-Sketch. For each $j \in [d]$, $i \in [\ell]$, the i-th estimate is $z_i := C_i[h_i(j)]g_i(j)$, then $\mathbb{E}[z_i] = x_j$.

Lemma. $\mathbb{P}[|z_i - x_j| \geq \frac{\epsilon}{\sqrt{k}} E_2^k(\mathbf{x})] \leq \frac{2}{5}$

Proof of First Lemma

page 1

Lemma. Count-Sketch with $w = 3k/\epsilon^2, \ell = O(\log n)$ guarantees $|x_j - \tilde{x}_j| \leq \frac{\epsilon}{\sqrt{k}} E_2^k$ for each j with high probability.

Recall the analysis of Count-Sketch. For each $j \in [d]$, $i \in [\ell]$, the i-th estimate is $z_i := C_i[h_i(j)]g_i(j)$, then $\mathbb{E}[z_i] = x_j$.

Lemma. $\mathbb{P}[\,|z_i - x_j| \geq \frac{\epsilon}{\sqrt{k}} E_2^k(\mathbf{x})] \leq \frac{2}{5}$

Proof of First Lemma

page 1

Lemma. Count-Sketch with $w = 3k/\epsilon^2, \ell = O(\log n)$ guarantees $|x_j - \tilde{x}_j| \leq \frac{\epsilon}{\sqrt{k}} E_2^k$ for each j with high probability.

Recall the analysis of Count-Sketch. For each $j \in [d]$, $i \in [\ell]$, the i-th estimate is $z_i := C_i[h_i(j)]g_i(j)$, then $\mathbb{E}[z_i] = x_j$.

Lemma. $\mathbb{P}[\,|z_i - x_j| \geq \frac{\epsilon}{\sqrt{k}} E_2^k(\mathbf{x})] \leq \frac{2}{5}$

Chernoff bound

Proof of First Lemma

page 1

Lemma. Count-Sketch with $w = 3k/\epsilon^2, \ell = O(\log n)$ guarantees $|x_j - \tilde{x}_j| \leq \frac{\epsilon}{\sqrt{k}} E_2^k$ for each j with high probability.

Recall the analysis of Count-Sketch. For each $j \in [d]$, $i \in [\ell]$, the i-th estimate is $z_i := C_i[h_i(j)]g_i(j)$, then $\mathbb{E}[z_i] = x_j$.

Lemma. $\mathbb{P}[|z_i - x_j| \geq \frac{\epsilon}{\sqrt{k}} E_2^k(\mathbf{x})] \leq \frac{2}{5}$

Chernoff bound

Let $z_i' = \sum_{j' \notin T, j' \neq j} g_i(j) g_i(j') x_{j'}$, then $\text{Var}[z_i'] \leq \frac{(E_2^k(\mathbf{x}))^2}{w} = \frac{3k}{\epsilon^2}(E_2^k(\mathbf{x}))^2$.

Proof of First Lemma

page 1

Lemma. Count-Sketch with $w = 3k/\epsilon^2, \ell = O(\log n)$ guarantees $|x_j - \tilde{x}_j| \leq \frac{\epsilon}{\sqrt{k}} E_2^k$ for each j with high probability.

Recall the analysis of Count-Sketch. For each $j \in [d]$, $i \in [\ell]$, the i-th estimate is $z_i := C_i[h_i(j)]g_i(j)$, then $\mathbb{E}[z_i] = x_j$.

Lemma. $\mathbb{P}[|z_i - x_j| \geq \frac{\epsilon}{\sqrt{k}} E_2^k(\mathbf{x})] \leq \frac{2}{5}$

Chernoff bound

Let $z_i' = \sum_{j' \notin T, j' \neq j} g_i(j) g_i(j') x_{j'}$, then $\mathrm{Var}[z_i'] \leq \frac{(E_2^k(\mathbf{x}))^2}{w} = \frac{3k}{\epsilon^2}(E_2^k(\mathbf{x}))^2$.

By Chebyshev inequality, $\mathbb{P}[|z_i' - x_i| > \frac{\epsilon}{\sqrt{k}} E_2^k(\mathbf{x})] \leq \frac{1}{3}$.

Proof of First Lemma

page 1

Lemma. Count-Sketch with $w = 3k/\epsilon^2, \ell = O(\log n)$ guarantees $|x_j - \tilde{x}_j| \le \frac{\epsilon}{\sqrt{k}} E_2^k$ for each j with high probability.

Recall the analysis of Count-Sketch. For each $j \in [d]$, $i \in [\ell]$, the i-th estimate is $z_i := C_i[h_i(j)]g_i(j)$, then $\mathbb{E}[z_i] = x_j$.

Lemma. $\mathbb{P}[|z_i - x_j| \ge \frac{\epsilon}{\sqrt{k}} E_2^k(\mathbf{x})] \le \frac{2}{5}$

Chernoff bound

Let $z_i' = \sum_{j' \notin T, j' \neq j} g_i(j) g_i(j') x_{j'}$, then $\mathrm{Var}[z_i'] \le \frac{(E_2^k(\mathbf{x}))^2}{w} = \frac{3k}{\epsilon^2}(E_2^k(\mathbf{x}))^2$.

By Chebyshev inequality, $\mathbb{P}[|z_i' - x_i| > \frac{\epsilon}{\sqrt{k}} E_2^k(\mathbf{x})] \le \frac{1}{3}$.

Let A be the event that none of entries in T collide with j under h_i, then $\mathbb{P}[A] \ge 1 - \frac{\epsilon^2}{3}$

Proof of First Lemma

page 2

Lemma. $\mathbb{P}[\,|z_i - x_j| \geq \frac{\epsilon}{\sqrt{k}} E_2^k(\mathbf{x})] \leq \frac{2}{5}$

Let $z_i' = \sum_{j' \notin T, j' \neq j} g_i(j) g_i(j') x_{j'}$, then $\mathrm{Var}[z_i'] \leq \frac{(E_2^k(\mathbf{x}))^2}{w} = \frac{3k}{\epsilon^2}(E_2^k(\mathbf{x}))^2$.

By Chebyshev inequality, $\mathbb{P}[\,|z_i' - x_i| > \frac{\epsilon}{\sqrt{k}} E_2^k(\mathbf{x})] \leq \frac{1}{3}$.

Let A be the event that some entry in T collides with j under h_i, then $\mathbb{P}[A] \leq \frac{k}{w} = \frac{\epsilon^2}{3}$

$$\mathbb{P}[\,|z_i - x_j| \geq \frac{\epsilon}{\sqrt{k}} E_2^k(\mathbf{x})] \leq \mathbb{P}[A] + \mathbb{P}[\,|z_i' - x_i| > \frac{\epsilon}{\sqrt{k}} E_2^k(\mathbf{x}) \mid \bar{A}] \cdot \mathbb{P}[\bar{A}]$$

$$\leq \frac{\epsilon^2}{3} + \frac{1}{3} \leq \frac{2}{5}$$

Proof of First Lemma

page 2

Lemma. $\mathbb{P}[|z_i - x_j| \geq \frac{\epsilon}{\sqrt{k}} E_2^k(\mathbf{x})] \leq \frac{2}{5}$

Let $z_i' = \sum_{j' \notin T, j' \neq j} g_i(j) g_i(j') x_{j'}$, then $\mathrm{Var}[z_i'] \leq \frac{(E_2^k(\mathbf{x}))^2}{w} = \frac{3k}{\epsilon^2}(E_2^k(\mathbf{x}))^2$.

By Chebyshev inequality, $\mathbb{P}[|z_i' - x_i| > \frac{\epsilon}{\sqrt{k}} E_2^k(\mathbf{x})] \leq \frac{1}{3}$.

Let A be the event that some entry in T collides with j under h_i, then $\mathbb{P}[A] \leq \frac{k}{w} = \frac{\epsilon^2}{3}$

$$\mathbb{P}[|z_i - x_j| \geq \frac{\epsilon}{\sqrt{k}} E_2^k(\mathbf{x})] \leq \mathbb{P}[A] + \mathbb{P}[|z_i' - x_i| > \frac{\epsilon}{\sqrt{k}} E_2^k(\mathbf{x}) \mid \bar{A}] \cdot \mathbb{P}[\bar{A}]$$

$$\leq \frac{\epsilon^2}{3} + \frac{1}{3} \leq \frac{2}{5}$$

Recall the proof we gave for the performance of SkipList. We had a similar use of union bound.

Proof of Second Lemma

Lemma. If for $\mathbf{x}, \mathbf{y} \in \mathbb{R}^d$ we have $\|\mathbf{x} - \mathbf{y}\|_\infty \leq \frac{\epsilon}{\sqrt{k}} E_2^k(\mathbf{x})$, let $\mathbf{z}$ be the k-sparse recovery of $\mathbf{y}$, then

$$\|\mathbf{x} - \mathbf{z}\| \leq (1 + 5\epsilon) E_2^k(\mathbf{x})$$

Proof of Second Lemma

Lemma. If for $\mathbf{x}, \mathbf{y} \in \mathbb{R}^d$ we have $\|\mathbf{x} - \mathbf{y}\|_\infty \leq \frac{\epsilon}{\sqrt{k}} E_2^k(\mathbf{x})$, let $\mathbf{z}$ be the k-sparse recovery of $\mathbf{y}$, then

$$\|\mathbf{x} - \mathbf{z}\| \leq (1 + 5\epsilon) E_2^k(\mathbf{x})$$

Let $T \subseteq [d]$ be the set of k "big" entries of $\mathbf{x}$, and T' be that for $\mathbf{y}$, then $\|\mathbf{x} - \mathbf{z}\|_2^2$ has three parts:

Proof of Second Lemma

Lemma. If for $\mathbf{x}, \mathbf{y} \in \mathbb{R}^d$ we have $\|\mathbf{x} - \mathbf{y}\|_\infty \leq \frac{\epsilon}{\sqrt{k}} E_2^k(\mathbf{x})$, let $\mathbf{z}$ be the k-sparse recovery of $\mathbf{y}$, then

$$\|\mathbf{x} - \mathbf{z}\| \leq (1 + 5\epsilon) E_2^k(\mathbf{x})$$

Let $T \subseteq [d]$ be the set of k "big" entries of $\mathbf{x}$, and T' be that for $\mathbf{y}$, then $\|\mathbf{x} - \mathbf{z}\|_2^2$ has three parts:

$$\|\mathbf{x} - \mathbf{z}\|_2^2 = \sum_{j \in T \cap T'} |x_j - z_j|^2 + \sum_{j \notin T \cup T'} x_j^2 + \sum_{j \in T \setminus T'} x_j^2 + \sum_{j \in T' \setminus T} |x_j - z_j|^2$$

Proof of Second Lemma

Lemma. If for $\mathbf{x}, \mathbf{y} \in \mathbb{R}^d$ we have $\|\mathbf{x} - \mathbf{y}\|_\infty \leq \frac{\epsilon}{\sqrt{k}} E_2^k(\mathbf{x})$, let $\mathbf{z}$ be the k-sparse recovery of $\mathbf{y}$, then

$$\|\mathbf{x} - \mathbf{z}\| \leq (1 + 5\epsilon) E_2^k(\mathbf{x})$$

Let $T \subseteq [d]$ be the set of k "big" entries of $\mathbf{x}$, and T' be that for $\mathbf{y}$, then $\|\mathbf{x} - \mathbf{z}\|_2^2$ has three parts:

$$\|\mathbf{x} - \mathbf{z}\|_2^2 = \sum_{j \in T \cap T'} |x_j - z_j|^2 + \sum_{j \notin T \cup T'} x_j^2 + \sum_{j \in T \setminus T'} x_j^2 + \sum_{j \in T' \setminus T} |x_j - z_j|^2$$

- entries in $T \cap T'$ and $T' \setminus T$: by assumption, each entry contributes $\leq \frac{\epsilon^2}{k} (E_2^k(\mathbf{x}))^2$, and there are k of them

Proof of Second Lemma

Lemma. If for $\mathbf{x}, \mathbf{y} \in \mathbb{R}^d$ we have $\|\mathbf{x} - \mathbf{y}\|_\infty \leq \frac{\epsilon}{\sqrt{k}} E_2^k(\mathbf{x})$, let $\mathbf{z}$ be the k-sparse recovery of $\mathbf{y}$, then

$$\|\mathbf{x} - \mathbf{z}\| \leq (1 + 5\epsilon) E_2^k(\mathbf{x})$$

Let $T \subseteq [d]$ be the set of k "big" entries of $\mathbf{x}$, and T' be that for $\mathbf{y}$, then $\|\mathbf{x} - \mathbf{z}\|_2^2$ has three parts:

$$\|\mathbf{x} - \mathbf{z}\|_2^2 = \sum_{j \in T \cap T'} |x_j - z_j|^2 + \sum_{j \notin T \cup T'} x_j^2 + \sum_{j \in T \setminus T'} x_j^2 + \sum_{j \in T' \setminus T} |x_j - z_j|^2$$

- entries in $T \cap T'$ and $T' \setminus T$: by assumption, each entry contributes $\leq \frac{\epsilon^2}{k} (E_2^k(\mathbf{x}))^2$, and there are k of them
- entries in $\overline{T} \cap \overline{T}'$: by definition, these are original components of $(E_2^k(\mathbf{x}))^2$

Proof of Second Lemma

Lemma. If for $\mathbf{x}, \mathbf{y} \in \mathbb{R}^d$ we have $\|\mathbf{x}-\mathbf{y}\|_\infty \leq \frac{\epsilon}{\sqrt{k}} E_2^k(\mathbf{x})$, let $\mathbf{z}$ be the k-sparse recovery of $\mathbf{y}$, then

$$\|\mathbf{x}-\mathbf{z}\| \leq (1+5\epsilon) E_2^k(\mathbf{x})$$

Let $T \subseteq [d]$ be the set of k "big" entries of $\mathbf{x}$, and T' be that for $\mathbf{y}$, then $\|\mathbf{x}-\mathbf{z}\|_2^2$ has three parts:

$$\|\mathbf{x}-\mathbf{z}\|_2^2 = \sum_{j \in T \cap T'} |x_j - z_j|^2 + \sum_{j \notin T \cup T'} x_j^2 + \sum_{j \in T \setminus T'} x_j^2 + \sum_{j \in T' \setminus T} |x_j - z_j|^2$$

- entries in $T \cap T'$ and $T' \setminus T$: by assumption, each entry contributes $\leq \frac{\epsilon^2}{k} (E_2^k(\mathbf{x}))^2$, and there are k of them
- entries in $\overline{T} \cap \overline{T}'$: by definition, these are original components of $(E_2^k(\mathbf{x}))^2$
- entries in $T - T'$ and $T' - T$:

Proof of Second Lemma

Lemma. If for $\mathbf{x}, \mathbf{y} \in \mathbb{R}^d$ we have $\|\mathbf{x} - \mathbf{y}\|_\infty \leq \frac{\epsilon}{\sqrt{k}} E_2^k(\mathbf{x})$, let $\mathbf{z}$ be the k-sparse recovery of $\mathbf{y}$, then

$$\|\mathbf{x} - \mathbf{z}\| \leq (1 + 5\epsilon) E_2^k(\mathbf{x})$$

Let $T \subseteq [d]$ be the set of k "big" entries of $\mathbf{x}$, and T' be that for $\mathbf{y}$, then $\|\mathbf{x} - \mathbf{z}\|_2^2$ has three parts:

$$\|\mathbf{x} - \mathbf{z}\|_2^2 = \sum_{j \in T \cap T'} |x_j - z_j|^2 + \sum_{j \notin T \cup T'} x_j^2 + \sum_{j \in T \setminus T'} x_j^2 + \sum_{j \in T' \setminus T} |x_j - z_j|^2$$

- entries in $T \cap T'$ and $T' \backslash T$: by assumption, each entry contributes $\leq \frac{\epsilon^2}{k} (E_2^k(\mathbf{x}))^2$, and there are k of them
- entries in $\overline{T} \cap \overline{T}'$: by definition, these are original components of $(E_2^k(\mathbf{x}))^2$
- entries in $T - T'$ and $T' - T$:
  - in $(E_2^k(\mathbf{x}))^2$ we should have $\sum_{j \in T' \setminus T} x_j^2$

Proof of Second Lemma

Lemma. If for $\mathbf{x}, \mathbf{y} \in \mathbb{R}^d$ we have $\|\mathbf{x}-\mathbf{y}\|_\infty \le \frac{\epsilon}{\sqrt{k}} E_2^k(\mathbf{x})$, let $\mathbf{z}$ be the k-sparse recovery of $\mathbf{y}$, then

$$\|\mathbf{x}-\mathbf{z}\| \le (1+5\epsilon)E_2^k(\mathbf{x})$$

Let $T \subseteq [d]$ be the set of k "big" entries of $\mathbf{x}$, and T' be that for $\mathbf{y}$, then $\|\mathbf{x}-\mathbf{z}\|_2^2$ has three parts:

$$\|\mathbf{x}-\mathbf{z}\|_2^2 = \sum_{j \in T \cap T'} |x_j - z_j|^2 + \sum_{j \notin T \cup T'} x_j^2 + \sum_{j \in T \setminus T'} x_j^2 + \sum_{j \in T' \setminus T} |x_j - z_j|^2$$

- entries in $T \cap T'$ and $T' \backslash T$: by assumption, each entry contributes $\le \frac{\epsilon^2}{k}(E_2^k(\mathbf{x}))^2$, and there are k of them
- entries in $\overline{T} \cap \overline{T}'$: by definition, these are original components of $(E_2^k(\mathbf{x}))^2$
- entries in $T - T'$ and $T' - T$:
  - in $(E_2^k(\mathbf{x}))^2$ we should have $\sum_{j \in T' \backslash T} x_j^2$
  - note that $|T - T'| = |T' - T|$ since $|T| = |T'| = k$.

Proof of Second Lemma

Lemma. If for $\mathbf{x}, \mathbf{y} \in \mathbb{R}^d$ we have $\|\mathbf{x} - \mathbf{y}\|_\infty \leq \frac{\epsilon}{\sqrt{k}} E_2^k(\mathbf{x})$, let $\mathbf{z}$ be the k-sparse recovery of $\mathbf{y}$, then

$$\|\mathbf{x} - \mathbf{z}\| \leq (1 + 5\epsilon) E_2^k(\mathbf{x})$$

Let $T \subseteq [d]$ be the set of k "big" entries of $\mathbf{x}$, and T' be that for $\mathbf{y}$, then $\|\mathbf{x} - \mathbf{z}\|_2^2$ has three parts:

$$\|\mathbf{x} - \mathbf{z}\|_2^2 = \sum_{j \in T \cap T'} |x_j - z_j|^2 + \sum_{j \notin T \cup T'} x_j^2 + \sum_{j \in T \setminus T'} x_j^2 + \sum_{j \in T' \setminus T} |x_j - z_j|^2$$

- entries in $T \cap T'$ and $T' \setminus T$: by assumption, each entry contributes $\leq \frac{\epsilon^2}{k} (E_2^k(\mathbf{x}))^2$, and there are k of them
- entries in $\overline{T} \cap \overline{T}'$: by definition, these are original components of $(E_2^k(\mathbf{x}))^2$
- entries in $T - T'$ and $T' - T$:
  - in $(E_2^k(\mathbf{x}))^2$ we should have $\sum_{j \in T' \setminus T} x_j^2$
  - note that $|T - T'| = |T' - T|$ since $|T| = |T'| = k$.

Proof of Second Lemma

Lemma. If for $\mathbf{x}, \mathbf{y} \in \mathbb{R}^d$ we have $\|\mathbf{x} - \mathbf{y}\|_\infty \leq \frac{\epsilon}{\sqrt{k}} E_2^k(\mathbf{x})$, let $\mathbf{z}$ be the k-sparse recovery of $\mathbf{y}$, then

$$\|\mathbf{x} - \mathbf{z}\| \leq (1 + 5\epsilon) E_2^k(\mathbf{x})$$

Let $T \subseteq [d]$ be the set of k "big" entries of $\mathbf{x}$, and T' be that for $\mathbf{y}$, then $\|\mathbf{x} - \mathbf{z}\|_2^2$ has three parts:

$$\|\mathbf{x} - \mathbf{z}\|_2^2 = \sum_{j \in T \cap T'} |x_j - z_j|^2 + \sum_{j \notin T \cup T'} x_j^2 + \sum_{j \in T \setminus T'} x_j^2 + \sum_{j \in T' \setminus T} |x_j - z_j|^2$$

- entries in $T \cap T'$ and $T' \setminus T$: by assumption, each entry contributes $\leq \frac{\epsilon^2}{k}(E_2^k(\mathbf{x}))^2$, and there are k of them
- entries in $\overline{T} \cap \overline{T}'$: by definition, these are original components of $(E_2^k(\mathbf{x}))^2$
- entries in $T - T'$ and $T' - T$:
  - in $(E_2^k(\mathbf{x}))^2$ we should have $\sum_{j \in T' \setminus T} x_j^2$
  - note that $|T - T'| = |T' - T|$ since $|T| = |T'| = k$.
  - Key observation: entries in $T - T'$ and $T' - T$ must all be close (in absolute value)

Lemma. If for $\mathbf{x}, \mathbf{y} \in \mathbb{R}^d$ we have $\|\mathbf{x}-\mathbf{y}\|_\infty \le \frac{\epsilon}{\sqrt{k}} E_2^k(\mathbf{x})$, let $\mathbf{z}$ be the k-sparse recovery of $\mathbf{y}$, then

$$\|\mathbf{x}-\mathbf{z}\| \le (1+5\epsilon) E_2^k(\mathbf{x})$$

Let $T \subseteq [d]$ be the set of k "big" entries of $\mathbf{x}$, and T' be that for $\mathbf{y}$, then $\|\mathbf{x}-\mathbf{z}\|_2^2$ has three parts:

$$\|\mathbf{x}-\mathbf{z}\|_2^2 = \sum_{j \in T \cap T'} |x_j - z_j|^2 + \sum_{j \notin T \cup T'} x_j^2 + \sum_{j \in T \setminus T'} x_j^2 + \sum_{j \in T' \setminus T} |x_j - z_j|^2$$

- entries in $T - T'$ and $T' - T$:
  - in $(E_2^k(\mathbf{x}))^2$ we should have $\sum_{j \in T' \setminus T} x_j^2$
  - note that $|T - T'| = |T' - T|$ since $|T| = |T'| = k$.
  - Key observation: entries in $T - T'$ and $T' - T$ must all be close (in absolute value)

Claim. If $j \in T \setminus T'$ and $j' \in T' \setminus T$, then $x_j \le x_{j'} + \frac{2\epsilon}{\sqrt{k}} E_2^k(\mathbf{x})$

Lemma. If for $\mathbf{x}, \mathbf{y} \in \mathbb{R}^d$ we have $\|\mathbf{x}-\mathbf{y}\|_\infty \leq \frac{\epsilon}{\sqrt{k}} E_2^k(\mathbf{x})$, let $\mathbf{z}$ be the k-sparse recovery of $\mathbf{y}$, then

$$\|\mathbf{x}-\mathbf{z}\| \leq (1+5\epsilon) E_2^k(\mathbf{x})$$

Let $T \subseteq [d]$ be the set of k "big" entries of $\mathbf{x}$, and T' be that for $\mathbf{y}$, then $\|\mathbf{x}-\mathbf{z}\|_2^2$ has three parts:

$$\|\mathbf{x}-\mathbf{z}\|_2^2 = \sum_{j \in T \cap T'} |x_j - z_j|^2 + \sum_{j \notin T \cup T'} x_j^2 + \sum_{j \in T \setminus T'} x_j^2 + \sum_{j \in T' \setminus T} |x_j - z_j|^2$$

- entries in $T - T'$ and $T' - T$:
  - in $(E_2^k(\mathbf{x}))^2$ we should have $\sum_{j \in T' \setminus T} x_j^2$
  - note that $|T - T'| = |T' - T|$ since $|T| = |T'| = k$.
  - Key observation: entries in $T - T'$ and $T' - T$ must all be close (in absolute value)

Claim. If $j \in T \setminus T'$ and $j' \in T' \setminus T$, then $x_j \leq x_{j'} + \frac{2\epsilon}{\sqrt{k}} E_2^k(\mathbf{x})$

$$\sum_{j \in T \setminus T'} x_j^2 \leq \sum_{j \in T' \setminus T} \left(|x_j| + \frac{2\epsilon}{\sqrt{k}} E_2^k(\mathbf{x})\right)^2 \leq \sum_{j \in T' \setminus T} x_j^2 + \frac{4\epsilon^2}{k} (E_2^k(\mathbf{x}))^2 + \frac{4\epsilon |x_j|}{\sqrt{k}} E_2^k(\mathbf{x}) \leq \sum_{j \in T' \setminus T} x_j^2 + 8\epsilon (E_2^k(\mathbf{x}))^2$$

Lemma. If for $\mathbf{x}, \mathbf{y} \in \mathbb{R}^d$ we have $\|\mathbf{x}-\mathbf{y}\|_\infty \le \frac{\epsilon}{\sqrt{k}} E_2^k(\mathbf{x})$, let $\mathbf{z}$ be the k-sparse recovery of $\mathbf{y}$, then

$$\|\mathbf{x}-\mathbf{z}\| \le (1+5\epsilon)E_2^k(\mathbf{x})$$

Let $T \subseteq [d]$ be the set of k "big" entries of $\mathbf{x}$, and T' be that for $\mathbf{y}$, then $\|\mathbf{x}-\mathbf{z}\|_2^2$ has three parts:

$$\|\mathbf{x}-\mathbf{z}\|_2^2 = \sum_{j \in T \cap T'} |x_j - z_j|^2 + \sum_{j \notin T \cup T'} x_j^2 + \sum_{j \in T \setminus T'} x_j^2 + \sum_{j \in T' \setminus T} |x_j - z_j|^2$$

- entries in $T - T'$ and $T' - T$:
  - in $(E_2^k(\mathbf{x}))^2$ we should have $\sum_{j \in T' \setminus T} x_j^2$
  - note that $|T - T'| = |T' - T|$ since $|T| = |T'| = k$.
  - Key observation: entries in $T - T'$ and $T' - T$ must all be close (in absolute value)

Claim. If $j \in T \setminus T'$ and $j' \in T' \setminus T$, then $x_j \le x_{j'} + \frac{2\epsilon}{\sqrt{k}} E_2^k(\mathbf{x})$

$$\sum_{j \in T \setminus T'} x_j^2 \le \sum_{j \in T' \setminus T} \left(|x_j| + \frac{2\epsilon}{\sqrt{k}} E_2^k(\mathbf{x})\right)^2 \le \sum_{j \in T' \setminus T} x_j^2 + \frac{4\epsilon^2}{k}(E_2^k(\mathbf{x}))^2 + \frac{4\epsilon |x_j|}{\sqrt{k}} E_2^k(\mathbf{x}) \le \sum_{j \in T' \setminus T} x_j^2 + 8\epsilon (E_2^k(\mathbf{x}))^2$$

By Cauchy-Schwartz, $\sum_{j \in T' \setminus T} |x_j| \le \sum_{j \notin T} |x_j| \le \sqrt{k \sum_{j \notin T} x_j^2} = \sqrt{k} E_2^k(\mathbf{x})$

Proof of Second Lemma

Lemma. If for $\mathbf{x}, \mathbf{y} \in \mathbb{R}^d$ we have $\|\mathbf{x} - \mathbf{y}\|_\infty \leq \frac{\epsilon}{\sqrt{k}} E_2^k(\mathbf{x})$, let $\mathbf{z}$ be the k-sparse recovery of $\mathbf{y}$, then

$$\|\mathbf{x} - \mathbf{z}\| \leq (1 + 5\epsilon) E_2^k(\mathbf{x})$$

Let $T \subseteq [d]$ be the set of k "big" entries of $\mathbf{x}$, and T' be that for $\mathbf{y}$, then $\|\mathbf{x} - \mathbf{z}\|_2^2$ has three parts:

$$\|\mathbf{x} - \mathbf{z}\|_2^2 = \sum_{j \in T \cap T'} |x_j - z_j|^2 + \sum_{j \notin T \cup T'} x_j^2 + \sum_{j \in T \setminus T'} x_j^2 + \sum_{j \in T' \setminus T} |x_j - z_j|^2$$

- entries in $T \cap T'$ and $T' \setminus T$: by assumption, each entry contributes $\leq \frac{\epsilon^2}{k} (E_2^k(\mathbf{x}))^2$, and there are k of them
- entries in $\overline{T} \cap \overline{T}'$: by definition, these are original components of $(E_2^k(\mathbf{x}))^2$
- entries in $T - T'$ and $T' - T$: $\sum_{j \in T \setminus T'} x_j^2 \leq \sum_{j \in T' \setminus T} x_j^2 + 8\epsilon (E_2^k(\mathbf{x}))^2$

Putting everything together, $\|\mathbf{x} - \mathbf{z}\|_2^2 \leq (1 + 9\epsilon)(E_2^k(\mathbf{x}))^2$, hence $\|\mathbf{x} - \mathbf{z}\| \leq \sqrt{1 + 9\epsilon} E_2^k(\mathbf{x}) \leq (1 + 5\epsilon) E_2^k(\mathbf{x})$.

Putting Things Together..

Lemma. Count-Sketch with $w = 3k/\epsilon^2, \ell = O(\log n)$ guarantees $|x_j - \tilde{x}_j| \leq \frac{\epsilon}{\sqrt{k}} E_2^k$ for each j w.h.p.

Lemma. If for $\mathbf{x}, \mathbf{y} \in \mathbb{R}^d$ we have $\|\mathbf{x} - \mathbf{y}\|_\infty \leq \frac{\epsilon}{\sqrt{k}} E_2^k(\mathbf{x})$, let $\mathbf{z}$ be the k-sparse recovery of $\mathbf{y}$, then

$$\|\mathbf{x} - \mathbf{z}\| \leq (1 + 5\epsilon) E_2^k(\mathbf{x})$$

Putting the two Lemmas together, we have that with high probability, the sparse recovery yielded by Count-Sketch has error $\leq (1 + 5\epsilon) E_2^k(\mathbf{x})$

Putting Things Together..

Lemma. Count-Sketch with $w = 3k/\epsilon^2, \ell = O(\log n)$ guarantees $|x_j - \tilde{x}_j| \leq \frac{\epsilon}{\sqrt{k}} E_2^k$ for each j w.h.p.

Lemma. If for $\mathbf{x}, \mathbf{y} \in \mathbb{R}^d$ we have $\|\mathbf{x} - \mathbf{y}\|_\infty \leq \frac{\epsilon}{\sqrt{k}} E_2^k(\mathbf{x})$, let $\mathbf{z}$ be the k-sparse recovery of $\mathbf{y}$, then

$$\|\mathbf{x} - \mathbf{z}\| \leq (1 + 5\epsilon) E_2^k(\mathbf{x})$$

Putting the two Lemmas together, we have that with high probability, the sparse recovery yielded by Count-Sketch has error $\leq (1 + 5\epsilon) E_2^k(\mathbf{x})$

One last thing: to give the sketch from $\tilde{\mathbf{x}}$, naïvely we need to go through all the coordinates, which takes time $O(d)$.

Putting Things Together..

Lemma. Count-Sketch with $w = 3k/\epsilon^2, \ell = O(\log n)$ guarantees $|x_j - \tilde{x}_j| \leq \frac{\epsilon}{\sqrt{k}} E_2^k$ for each j w.h.p.

Lemma. If for $\mathbf{x}, \mathbf{y} \in \mathbb{R}^d$ we have $\|\mathbf{x} - \mathbf{y}\|_\infty \leq \frac{\epsilon}{\sqrt{k}} E_2^k(\mathbf{x})$, let $\mathbf{z}$ be the k-sparse recovery of $\mathbf{y}$, then

$$\|\mathbf{x} - \mathbf{z}\| \leq (1 + 5\epsilon) E_2^k(\mathbf{x})$$

Putting the two Lemmas together, we have that with high probability, the sparse recovery yielded by Count-Sketch has error $\leq (1 + 5\epsilon) E_2^k(\mathbf{x})$

One last thing: to give the sketch from $\tilde{\mathbf{x}}$, naïvely we need to go through all the coordinates, which takes time $O(d)$.

We can do faster by maintaining a record as the input comes!