CSCE 689: Special Topics in Modern Algorithms for Data Science

Lecture 15

Samson Zhou

Previously in the Streaming Model

- Reservoir sampling
- Heavy-hitters
  - Misra-Gries
  - CountMin
  - CountSketch
- Moment estimation
  - AMS algorithm

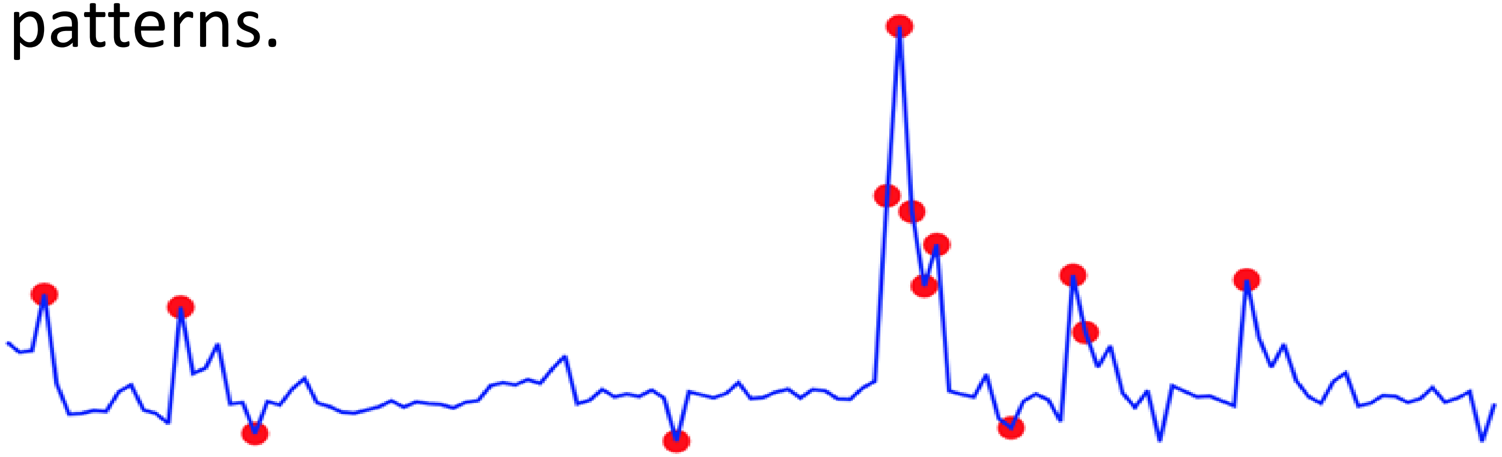
Sparse Recovery

- Suppose we have an insertion-deletion stream of length $m = \Theta(n)$ and at the end we are promised there are at most k nonzero coordinates
- Goal: Recover the k nonzero coordinates and their frequencies

Applications of Sparse Recovery

- Anomaly detection: Noiseless sparse recovery can be used to identify anomalies or outliers in streaming data

- By modeling normal behavior as a sparse signal, deviations from this model can be detected in real-time. This is valuable for cybersecurity, fraud detection, and monitoring network traffic for unusual patterns.

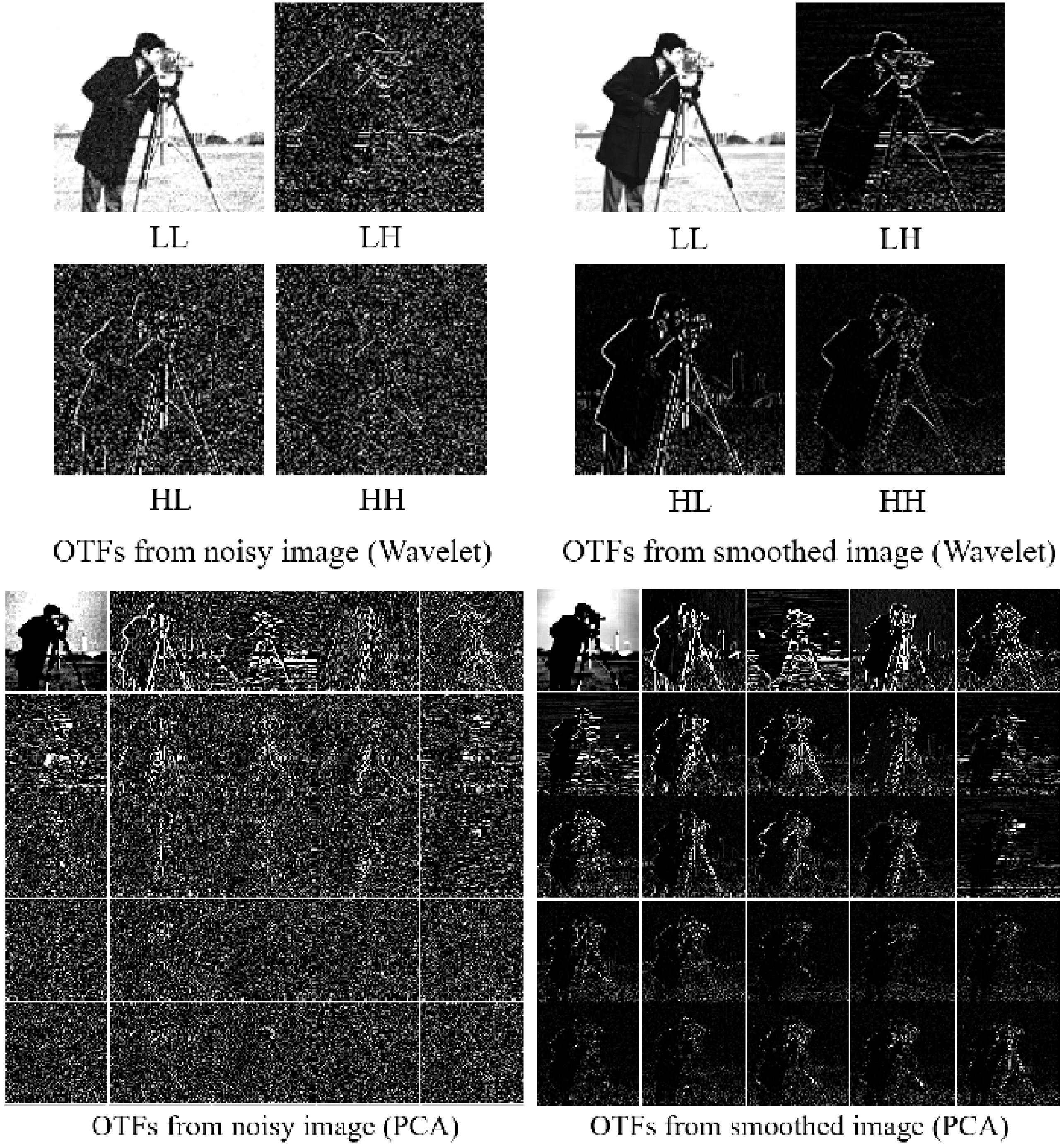


Applications of Sparse Recovery

- Network traffic analysis: Noiseless sparse recovery can be applied to analyze network traffic in real-time, identifying patterns and trends, and helping in network management, intrusion detection, and quality of service (QoS) optimization

Applications of Sparse Recovery

- Real-time compressive imaging: Compressive imaging techniques can be applied to streaming video or image data. By capturing and processing fewer measurements, noiseless sparse recovery can provide real-time reconstruction of high-resolution images or videos.



"Deep Orthogonal Transform Feature for Image Denoising", Shin, et. al. [2020]

Applications of Sparse Recovery

- Online natural language processing (NLP): In real-time natural language processing tasks, noiseless sparse recovery can assist in extracting relevant features or patterns from streaming text data, making it useful for sentiment analysis, topic modeling, and summarization

Sparse Recovery

- Suppose we have an insertion-deletion stream of length $m = \Theta(n)$
- Suppose at the end we are promised there are at most k nonzero coordinates
- How do we recover the vector?

Sparse Recovery

- Suppose at the end we are promised there are at most k nonzero coordinates
- Suppose $k = 1$ and we are promised the coordinate has frequency 1

Sparse Recovery

- Suppose at the end we are promised there are at most k nonzero coordinates
- Suppose $k = 1$ and we are promised the coordinate has frequency 1

u_1: "Increase f_6"

Sparse Recovery

- Suppose at the end we are promised there are at most k nonzero coordinates
- Suppose $k = 1$ and we are promised the coordinate has frequency 1

u_2: "Increase f_5"

Sparse Recovery

- Suppose at the end we are promised there are at most k nonzero coordinates
- Suppose $k = 1$ and we are promised the coordinate has frequency 1

u_3: "Increase f_2"

Sparse Recovery

- Suppose at the end we are promised there are at most k nonzero coordinates
- Suppose $k = 1$ and we are promised the coordinate has frequency 1

u_4: "Increase f_7"

Sparse Recovery

- Suppose at the end we are promised there are at most k nonzero coordinates
- Suppose $k = 1$ and we are promised the coordinate has frequency 1

u_5: "Increase f_3"

Sparse Recovery

- Suppose at the end we are promised there are at most k nonzero coordinates
- Suppose $k = 1$ and we are promised the coordinate has frequency 1

u_6: "Increase f_3"

Sparse Recovery

- Suppose at the end we are promised there are at most k nonzero coordinates
- Suppose $k = 1$ and we are promised the coordinate has frequency 1

u_7: "Increase f_2"

Sparse Recovery

- Suppose at the end we are promised there are at most k nonzero coordinates
- Suppose $k = 1$ and we are promised the coordinate has frequency 1

u_8: "Increase f_8"

Sparse Recovery

- Suppose at the end we are promised there are at most k nonzero coordinates
- Suppose $k = 1$ and we are promised the coordinate has frequency 1

u_9: "Decrease f_3"

Sparse Recovery

- Suppose at the end we are promised there are at most k nonzero coordinates
- Suppose $k = 1$ and we are promised the coordinate has frequency 1

u_{10}: "Decrease f_5"

Sparse Recovery

- Suppose at the end we are promised there are at most k nonzero coordinates
- Suppose $k = 1$ and we are promised the coordinate has frequency 1

u_{11}: "Increase f_1"

Sparse Recovery

- Suppose at the end we are promised there are at most k nonzero coordinates
- Suppose $k = 1$ and we are promised the coordinate has frequency 1

u_{12}: "Increase f_7"

Sparse Recovery

- Suppose at the end we are promised there are at most k nonzero coordinates
- Suppose $k = 1$ and we are promised the coordinate has frequency 1

u_{13}: "Decrease f_6"

Sparse Recovery

- Suppose at the end we are promised there are at most k nonzero coordinates
- Suppose $k = 1$ and we are promised the coordinate has frequency 1

u_{14}: “Decrease f_8”

Sparse Recovery

- Suppose at the end we are promised there are at most k nonzero coordinates
- Suppose $k = 1$ and we are promised the coordinate has frequency 1

u_{15}: "Decrease f_1"

Sparse Recovery

- Suppose at the end we are promised there are at most k nonzero coordinates
- Suppose $k = 1$ and we are promised the coordinate has frequency 1

u_{16}: "Decrease f_7"

Sparse Recovery

- Suppose at the end we are promised there are at most k nonzero coordinates
- Suppose $k = 1$ and we are promised the coordinate has frequency 1

u_{17}: "Decrease f_3"

Sparse Recovery

- Suppose at the end we are promised there are at most k nonzero coordinates
- Suppose $k = 1$ and we are promised the coordinate has frequency 1

u_{18}: "Decrease f_2"

Sparse Recovery

- Suppose at the end we are promised there are at most k nonzero coordinates
- Suppose $k = 1$ and we are promised the coordinate has frequency 1

u_{19}: "Decrease f_7"

Sparse Recovery

- Suppose at the end we are promised there are at most k nonzero coordinates
- Suppose $k = 1$ and we are promised the coordinate has frequency 1
- What is left?

Sparse Recovery

- Suppose at the end we are promised there are at most k nonzero coordinates
- Suppose $k = 1$ and we are promised the coordinate has frequency 1
- What is left?

$$f_2 = 1$$

Sparse Recovery

- Suppose at the end we are promised there are at most k nonzero coordinates
- Suppose $k = 1$ and we are promised the coordinate has frequency 1

- Algorithm: Keep running sum of all the coordinates

Sparse Recovery

- Suppose at the end we are promised there are at most k nonzero coordinates
- Suppose $k = 1$ and we are promised the coordinate has frequency 1

- Algorithm: Keep running sum of all the coordinates

- Write each insertion to coordinate $c_i \in [n]$ as $u_i \leftarrow (s_i = 1, c_i)$
- Write each deletion to coordinate $c_i \in [n]$ as $u_i \leftarrow (s_i = -1, c_i)$

Sparse Recovery

- Suppose $k = 1$ and we are promised the coordinate j has frequency 1
- Algorithm: Keep running sum of all the coordinates

- Write each insertion to coordinate $c_i \in [n]$ as $u_i \leftarrow (s_i = 1, c_i)$
- Write each deletion to coordinate $c_i \in [n]$ as $u_i \leftarrow (s_i = -1, c_i)$

- Running sum of coordinates $\sum_{i \in [m]} s_i c_i = j$

Sparse Recovery

- Suppose at the end we are promised there are at most k nonzero coordinates
- Suppose $k = 1$ ~~and we are promised the coordinate j has frequency 1~~

- Algorithm: Keep running sum of all the coordinates?

Sparse Recovery

- Suppose at the end we are promised there are at most k nonzero coordinates
- Suppose $k = 1$ ~~and we are promised the coordinate j has frequency 1~~

- Algorithm: Keep running sum of all the coordinates AND a different linear combination of all the coordinates

Sparse Recovery

- Suppose at the end we are promised there are at most k nonzero coordinates
- Suppose $k = 1$ ~~and we are promised the coordinate j has frequency 1~~

- Algorithm: Keep running sum of all the coordinates AND a different linear combination of all the coordinates

- Keep $\sum_{i \in [m]} s_i c_i$ and $\sum_{i \in [m]} s_i c_i^2$

Sparse Recovery

- Suppose at the end we are promised there are at most k nonzero coordinates
- Suppose $k = 1$ ~~and we are promised the coordinate j has frequency 1~~

u_1: "Increase f_6"

Sparse Recovery

- Suppose at the end we are promised there are at most k nonzero coordinates
- Suppose $k = 1$ ~~and we are promised the coordinate j has frequency 1~~

u_2: "Increase f_5"

Sparse Recovery

- Suppose at the end we are promised there are at most k nonzero coordinates
- Suppose $k = 1$ ~~and we are promised the coordinate j has frequency 1~~

u_3: “Increase f_2”

Sparse Recovery

- Suppose at the end we are promised there are at most k nonzero coordinates
- Suppose $k = 1$ ~~and we are promised the coordinate j has frequency 1~~

u_4: "Increase f_7"

Sparse Recovery

- Suppose at the end we are promised there are at most k nonzero coordinates
- Suppose $k = 1$ ~~and we are promised the coordinate j has frequency 1~~

u_5: "Increase f_3"

Sparse Recovery

- Suppose at the end we are promised there are at most k nonzero coordinates
- Suppose $k = 1$ ~~and we are promised the coordinate j has frequency 1~~

u_6: "Increase f_3"

Sparse Recovery

- Suppose at the end we are promised there are at most k nonzero coordinates
- Suppose $k = 1$ ~~and we are promised the coordinate j has frequency 1~~

u_7: "Increase f_2"

Sparse Recovery

- Suppose at the end we are promised there are at most k nonzero coordinates
- Suppose $k = 1$ ~~and we are promised the coordinate j has frequency 1~~

u_8: “Increase f_8”

Sparse Recovery

- Suppose at the end we are promised there are at most k nonzero coordinates
- Suppose $k = 1$ ~~and we are promised the coordinate j has frequency 1~~

u_9: "Decrease f_3"

Sparse Recovery

- Suppose at the end we are promised there are at most k nonzero coordinates
- Suppose $k = 1$ ~~and we are promised the coordinate j has frequency 1~~

u_{10}: "Decrease f_5"

Sparse Recovery

- Suppose at the end we are promised there are at most k nonzero coordinates
- Suppose $k = 1$ ~~and we are promised the coordinate j has frequency 1~~

u_{11}: "Increase f_1"

Sparse Recovery

- Suppose at the end we are promised there are at most k nonzero coordinates
- Suppose $k = 1$ ~~and we are promised the coordinate j has frequency 1~~

u_{12}: "Increase f_7"

Sparse Recovery

- Suppose at the end we are promised there are at most k nonzero coordinates
- Suppose $k = 1$ ~~and we are promised the coordinate j has frequency 1~~

u_{13}: "Decrease f_6"

Sparse Recovery

- Suppose at the end we are promised there are at most k nonzero coordinates
- Suppose $k = 1$ ~~and we are promised the coordinate j has frequency 1~~

$$u_{14}\text{: "Decrease } f_8\text{"}$$

Sparse Recovery

- Suppose at the end we are promised there are at most k nonzero coordinates
- Suppose $k = 1$ ~~and we are promised the coordinate j has frequency 1~~

$$u_{15}\text{: “Decrease } f_1\text{”}$$

Sparse Recovery

- Suppose at the end we are promised there are at most k nonzero coordinates
- Suppose $k = 1$ ~~and we are promised the coordinate j has frequency 1~~

u_{16}: "Decrease f_7"

Sparse Recovery

- Suppose at the end we are promised there are at most k nonzero coordinates
- Suppose $k = 1$ ~~and we are promised the coordinate j has frequency 1~~

u_{17}: "Decrease f_3"

Sparse Recovery

- Suppose at the end we are promised there are at most k nonzero coordinates
- Suppose $k = 1$ ~~and we are promised the coordinate j has frequency 1~~

u_{18}: "Decrease f_7"

Sparse Recovery

- Suppose at the end we are promised there are at most k nonzero coordinates
- Suppose $k = 1$ ~~and we are promised the coordinate j has frequency 1~~
- What is the state of our algorithm?

Sparse Recovery

- Suppose at the end we are promised there are at most k nonzero coordinates
- Suppose $k = 1$ ~~and we are promised the coordinate j has frequency 1~~
- What is the state of our algorithm?

$$\sum_{i \in [m]} s_i c_i = 4 \text{ and } \sum_{i \in [m]} s_i c_i^2 = 8$$

Sparse Recovery

- Suppose at the end we are promised there are at most k nonzero coordinates
- Suppose $k = 1$ ~~and we are promised the coordinate j has frequency 1~~
- What is the state of our algorithm?

$$\sum_{i \in [m]} s_i c_i = 4 \text{ and } \sum_{i \in [m]} s_i c_i^2 = 8$$

- We know $\sum_{i \in [m]} s_i c_i = j \cdot f_j$ and $\sum_{i \in [m]} s_i c_i^2 = j^2 \cdot f_j$

Sparse Recovery

- Suppose at the end we are promised there are at most k nonzero coordinates
- Suppose $k = 1$ ~~and we are promised the coordinate j has frequency 1~~
- What is the state of our algorithm?

$$\sum_{i \in [m]} s_i c_i = 4 \text{ and } \sum_{i \in [m]} s_i c_i^2 = 8$$

- We know $\sum_{i \in [m]} s_i c_i = j \cdot f_j$ and $\sum_{i \in [m]} s_i c_i^2 = j \cdot f_j^2$
- So $f_j = 2$ and $j = 2$

Sparse Recovery

- Suppose at the end we are promised there are at most k nonzero coordinates
- Suppose $k = 1$ ~~and we are promised the coordinate j has frequency 1~~
- What is the state of our algorithm?

$$\sum_{i \in [m]} s_i c_i = 4 \text{ and } \sum_{i \in [m]} s_i c_i^2 = 8$$

- We know $\sum_{i \in [m]} s_i c_i = j \cdot f_j$ and $\sum_{i \in [m]} s_i c_i^2 = j \cdot f_j^2$
- So $f_j = 2$ and $j = 2$

f_1	f_2	f_3	f_4	f_5	f_6	f_7
0	2	0	0	0	0	0

Sparse Recovery

- Suppose at the end we are promised there are at most k nonzero coordinates
- Algorithm for $k = 1$: Keep running sum of all the coordinates AND a different linear combination of all the coordinates

Sparse Recovery

- Suppose at the end we are promised there are at most k nonzero coordinates

- Algorithm: Keep $2k$ running sum of different linear combinations of all the coordinates

- We have $2k$ equations and $2k$ unknown variables

- Correctness can be shown (not quite linear algebra)

Sparse Recovery

- Suppose at the end we are promised there are at most k nonzero coordinates

- Algorithm: Keep $2k$ running sum of different linear combinations of all the coordinates

- Space: $O(k)$ words of space