Matrix Sketching over Sliding Windows

Zhewei Wei[1], Xuancheng Liu[1], Feifei Li[2], Shuo Shang[1], Xiaoyong Du[1], Ji-Rong Wen[1]

[1] *School of Information, Renmin University of China*

[2] *School of Computing, The University of Utah*

Contact: zhewei@ruc.edu.cn

Motivation and Problem Statement

Matrix Sketching

- Modern data sets are modeled as large matrices, computing SVD is slow.
- Matrix sketching: approximate large matrix $A \in R^{n \times d}$ with B$\in R^{l \times d}$, $l \ll n$
- Row-update stream: each update receives a_i, a row of A.
- Covariance error: $||A^TA - B^TB|| \le \varepsilon ||A||_F^2$.
- Frequent Direction (FD) [Liberty2013]: B$\in R^{l \times d}$ s.t. $||A^TA - B^TB|| \le \frac{1}{l}||A||_F^2$.

Sliding Window Summaries:

- Model time-sensitive data.
- Sequence-based: past N items; Time-based: items in a past time period.

Matrix Sketching over Sliding Windows

- Maintain (approximately) $A_W^TA_W$ for time/sequence-based window W.
- Applications: sliding window PCA; analyzing text data for a past time period.

Lower bounds: Challenges and Assumptions

Theorem 4.1 An algorithm that returns A^TA for any sequence- based sliding window must use $\Omega(Nd)$ bits space.

- Unbounded stream solution: use $O(d^2)$ space to store A^TA. This solution does not work on sliding window.

Theorem 4.2 An algorithm that returns B_W such that

$$Pr[||A_W^TA_W - B_W^TB_W|| \le \frac{1}{8d}||A_W||_F^2] \ge \frac{1}{2}$$

for any sequence- based sliding window W must use $\Omega(Nd)$ bits space.

- Need to assume the ratio R between maximum squared norm and minimum squared norms is bounded.

Baseline: Sampling-based algorithm

Insight: Sample each row a_i with probability proportional to its squared norm $||a_i||^2$ and rescale with proper factors.

Sample with replacement (SWR)

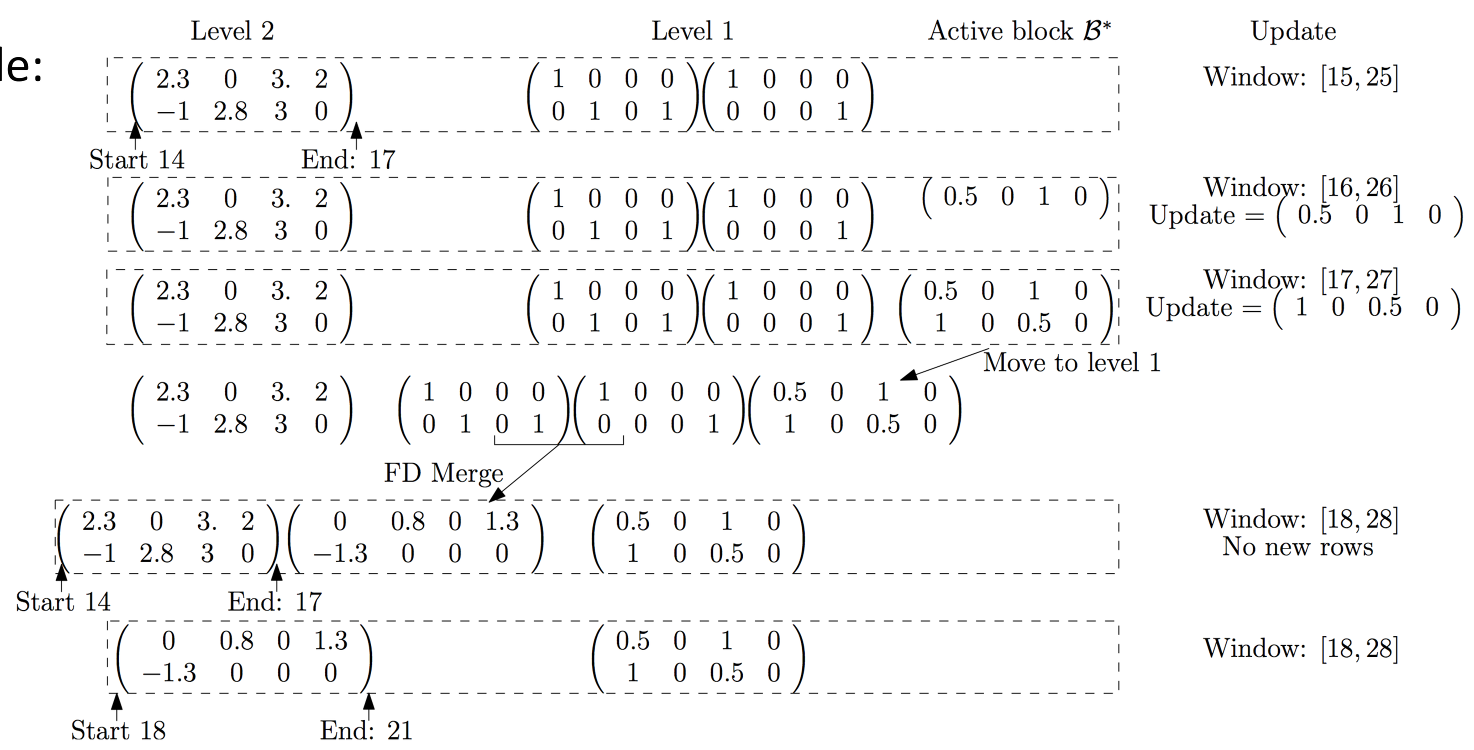
- "Magical" priority: $u^{1/||a_i||^2}$.
- Top-1 priority: sample proportional to $||a_i||^2$.
- Rescale sampled row a_i back by a factor of $\sqrt{l}||a_i||/||A||_F$.
- Run l independent samplers.

Algorithm 5.1 Update algorithm of SWR at time t

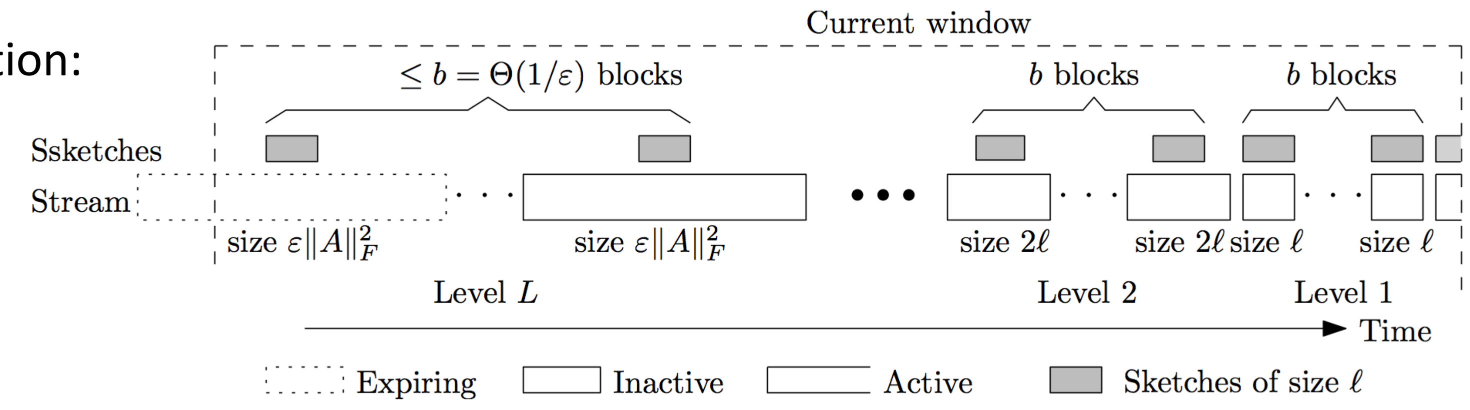
```
1: Remove all (a_j, t_j, ρ_j) in Q with t_j < t − Δ
2: if update = (a_t, t) then
3:    Choose u_t ∈ Unif(0,1) and set ρ_t ← u_t^{1/||a_t||^2}
4:    while ρ < ρ_t do
5:       (a_j, t_j, ρ_j) = Q.back
6:       if ρ < ρ_t then
7:          Remove (a_j, t_j, ρ_j)
8:    Append (a_t, t, ρ_t) to the end of Q
```

Logarithmic Method: LM-FD algorithm

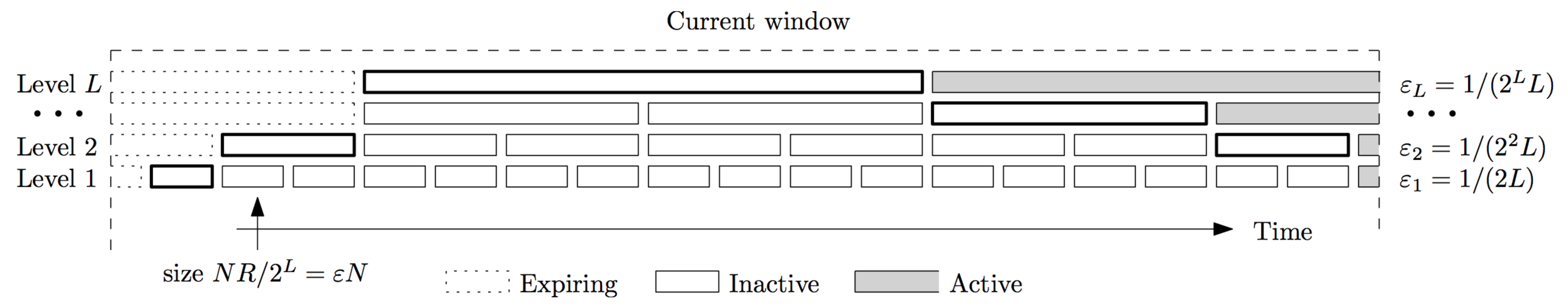
- Work for time/sequence-based window.
- Mergeablility: $B_1 = FD(A_1, \varepsilon), B_2 = FD(A_2, \varepsilon)$, we have $B = FD([B_1, B_2], \varepsilon)$ is a FD sketch for A.
- Combine FD with Exponential Histogram [Datar2002].
- Logarithmic number of levels, each contains $1/\varepsilon$ sketches.
- Merge all blocks to form B.
- Example:



- Illustration:



Dyadic Interval: DI-FD algorithm



- Work for sequence-based window.
- Window of size N can be decomposed into $\log N$ dyadic intervals.
- Maintain a sketch for each interval.
- Sketches at different levels have different error parameters.
- Combine $\log N$ sketches to form B.

Experiments and Conclusion

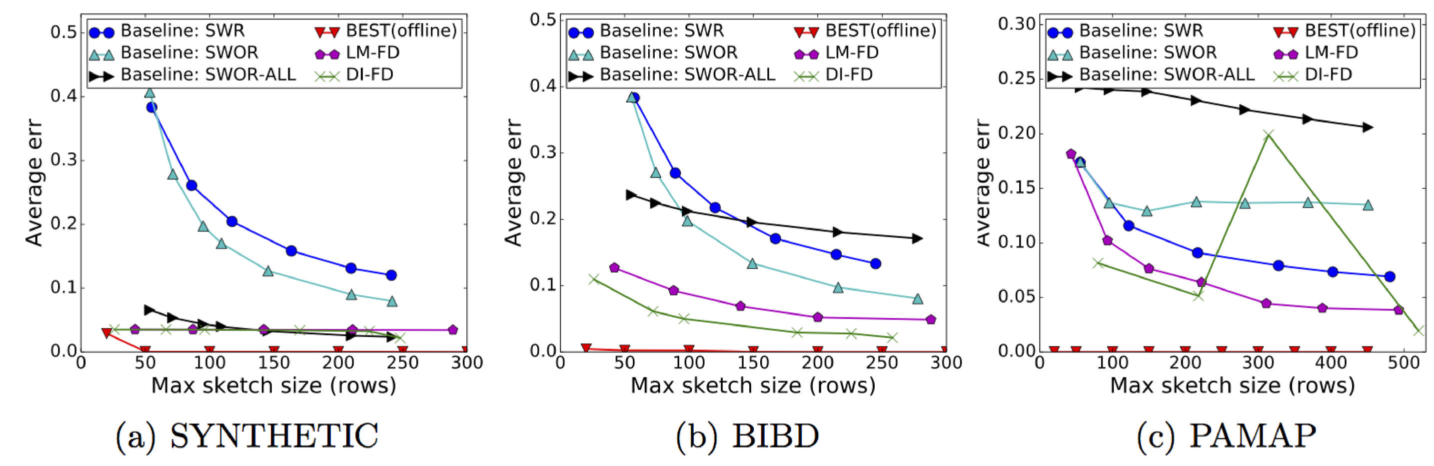
Datasets:

Data Set	total rows n	d	N	ratio R
SYNTHETIC	10^6	300	10,000	8.35
BIBD	319,770	231	10,000	1
PAMAP	198,000	35	10,000	90089

Table 2: Data Sets for sequence-based window.

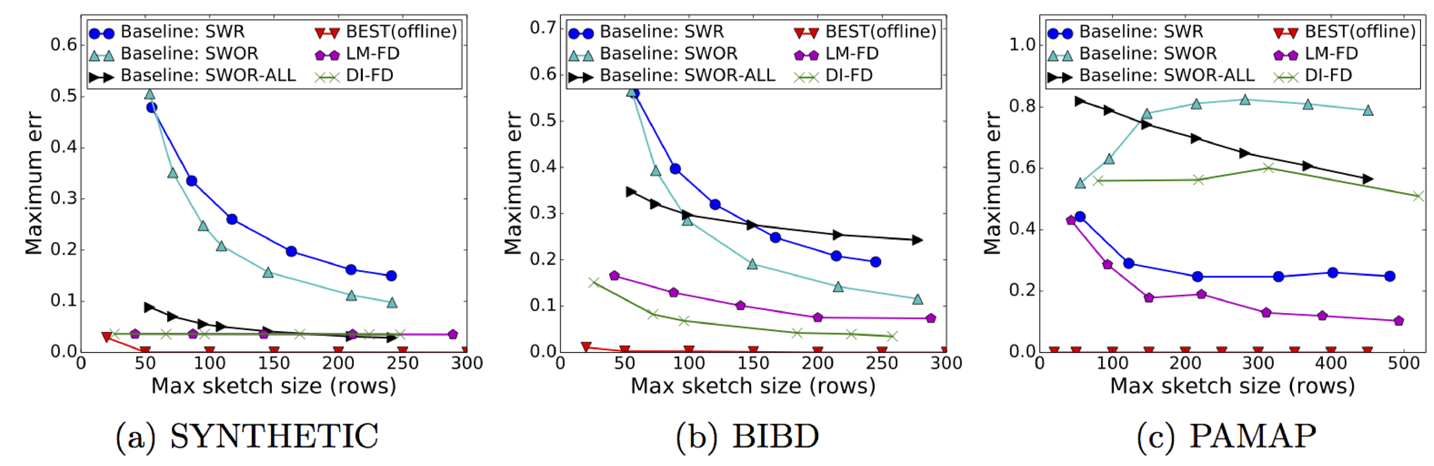
Observations

- FD vs. Sampling: DI-FD and LM-FD provide better space-error tradeoffs.
- DI-FD vs. LM-FD: depends on the ratio R between maximum squared norm and minimum squared norms in the data set.
- SWOR vs. SWR: depends on data set.



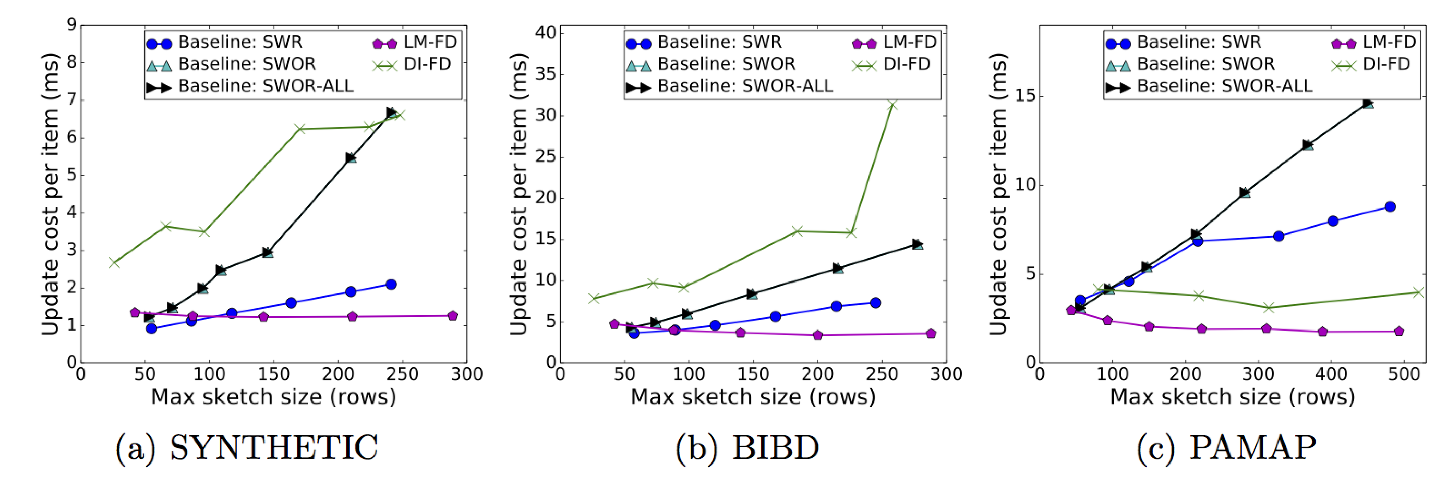
(a) SYNTHETIC (b) BIBD (c) PAMAP

Figure 3: average err vs. maximum sketch size.



(a) SYNTHETIC (b) BIBD (c) PAMAP

Figure 4: maximum err vs. maximum sketch size.



(a) SYNTHETIC (b) BIBD (c) PAMAP

Conclusions

- Sampling: interpretable, bad space usage. Slow update.
- DI-FD: best space usage for normalized, sequence-based windows. Slow update.
- LM-FD: space-efficient, work for time/sequence-based windows, insensitive to R. Fast update.

sketch κ	update time	sketch size	cova-err	ℓ	window	$B \subset A$?	Need R?
Sampling (SWR)	$(d/\varepsilon^2)\log\log NR$	$(d/\varepsilon^2)\log NR$ (Expected)	$\le \varepsilon$ w.h.p.	d/ε^2	sequence & time	✓	No
LM-FD	$d\log\varepsilon NR$	$(1/\varepsilon^2)\log\varepsilon NR$	$\le \varepsilon$	$1/\varepsilon$	sequence & time	×	Yes
DI-FD	$(d/\varepsilon)\log R/\varepsilon$	$(R/\varepsilon)\log R/\varepsilon$	$\le \varepsilon$	$1/\varepsilon$	sequence	×	Yes