

Sketch-FORCE: Sub-quadratic Streaming Robust Covariance Estimation for Extreme Dimensions

Sooyoung Jang

Hanbat National University

Changbeom Choi

`cbchoi@hanbat.ac.kr`

Hanbat National University

Research Article

Keywords: Streaming Algorithms, Robust Statistics, Matrix Sketching, High-Dimensional Data, Covariance Estimation

Posted Date: June 24th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-10124326/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.
[Read Full License](#)

Additional Declarations: The authors declare no competing interests.

Sketch-FORCE: Sub-quadratic Streaming Robust Covariance Estimation for Extreme Dimensions

Sooyoung Jang¹ and Changbeom Choi^{1*}

¹Department of Computer Engineering, Hanbat National University,
Daejeon, 34158, Republic of Korea.

*Corresponding author(s). E-mail(s): cbchoi@hanbat.ac.kr;
Contributing authors: syjang@hanbat.ac.kr;

Abstract

Streaming covariance estimation is a foundational primitive in real-time high-dimensional data analysis, underlying applications ranging from Internet-of-Things (IoT) sensor networks to high-throughput single-cell genomics. In these settings, data streams are often subject to arbitrary contamination by outliers and heavy-tailed noise, while being bounded by strict computational and memory constraints. Although robust scale estimation methods offer protection against extreme marginal outliers, they require storing a full $p \times p$ covariance matrix, leading to an insurmountable $O(p^2)$ memory wall as the dimension p scales beyond tens of thousands. In this paper, we present **Sketch-FORCE**, a sub-quadratic, robust streaming covariance estimator that breaks this memory barrier. Sketch-FORCE decouples robust scale tracking from global covariance updating by combining Piecewise-Parabolic (P^2) quantile trackers with Frequent Directions (FD) matrix sketching. By dynamically Winsorizing incoming observations based on running medians and scales (using either Interquartile Range or Median Absolute Deviation) and streaming these trimmed vectors into an FD sketch, we maintain an active low-rank representation using only $O(pk)$ memory, where $k \ll p$ is the sketch rank. SVD-based shrinkage is executed in $O(k^2p)$ time whenever the sketch fills, completely bypassing $O(p^2)$ storage and computation. We analyze the theoretical trade-off of this approach, specifically the “imputation bias” introduced by trimming in cross-covariance terms versus the contamination bias it prevents. Empirically, we demonstrate that Sketch-FORCE scales to dimensions up to $p = 10,000$ using highly efficient throughput, while outperforming vanilla sketching under heavy-tailed contamination and matching the subspace accuracy of full $O(p^2)$ robust methods. We show its downstream utility on the real-world KDD Cup 99 network intrusion dataset, confirming its practical relevance for real-time edge environments. Furthermore, we subject the

algorithm to an adversarial stress test, illustrating its theoretical vulnerability to structural outliers that evade marginal bounds but distort joint correlation, establishing the boundary conditions of marginal robust sketching.

Keywords: Streaming Algorithms, Robust Statistics, Matrix Sketching, High-Dimensional Data, Covariance Estimation

1 Introduction

In the modern data era, the proliferation of high-dimensional, continuously arriving data sources has transformed the landscape of statistical machine learning and data mining. Application domains such as single-cell genomics (e.g., single-cell RNA sequencing tracking tens of thousands of transcripts), distributed Internet-of-Things (IoT) telemetry [Carnier et al. \(2025\)](#), and financial market tickers [Maddanu et al. \(2025\)](#) yield data points of high dimensionality p in real-time. A fundamental building block in analyzing these streams is the covariance matrix, which captures linear dependencies and governs key dimensionality reduction techniques such as streaming Principal Component Analysis (PCA). To be useful in streaming scenarios, a covariance estimator must process incoming observations sequentially, using memory that scales sub-quadratically with p , while remaining highly throughput-efficient.

However, streaming data is rarely clean. It is routinely corrupted by anomalous sensor readings, technical artifacts, network packet losses, or adversarial interventions, which manifest as heavy-tailed distributions and arbitrary contamination. Classical streaming covariance estimators, such as Welford’s one-pass algorithm or Oja’s rule, are notoriously fragile, as a single large outlier can arbitrarily rotate the estimated principal directions. In batch settings, robust statistics provides estimators with high breakdown points, such as the Minimum Covariance Determinant (MCD) or projection-pursuit estimators. Unfortunately, these methods are computationally intractable for high dimensions, scaling exponentially or cubically with p , and cannot be directly adapted to single-pass, streaming environments where each observation must be ingested in real-time.

The broader data mining community has increasingly recognized the necessity of robust stream processing to handle anomalies and concept drift in continuous environments [Gama et al. \(2014\)](#); [Faria et al. \(2016\)](#). To address the computational limitations of classical robust estimators in these settings, recent work has explored streaming coordinate-wise trimming techniques, such as the First-Order Robust Covariance Estimator (FORCE) [Jang and Choi \(2026\)](#) and MAD-FORCE [Jang et al. \(2026\)](#). FORCE tracks marginal robust statistics (such as the Median Absolute Deviation) to dynamically Winsorize incoming streaming vectors before updating a running covariance matrix. While computationally lightweight and robust to independent outliers, FORCE and its variants suffer from a critical limitation: they maintain the full covariance structure explicitly. This imposes a strict $O(p^2)$ memory requirement. When p exceeds 20,000, a single dense double-precision (64-bit) covariance matrix requires over

3.2 gigabytes of memory, and at $p = 50,000$, it exceeds 20 gigabytes. This $O(p^2)$ “memory wall” makes standard robust estimators unusable on edge devices or standard cluster nodes processing extreme-dimensional data streams.

To bypass the memory wall while preserving robustness, we introduce **Sketch-FORCE**, a streaming covariance estimator that achieves sub-quadratic memory complexity by bridging coordinate-wise robustness with deterministic matrix sketching. Sketch-FORCE decouples the marginal scale estimation from the covariance reconstruction. It utilizes Piecewise-Parabolic (P^2) quantile trackers to estimate coordinate-wise medians and scales (using either Interquartile Range or Median Absolute Deviation) in $O(p)$ memory. Incoming observations are coordinate-wise Winsorized (trimmed) using these running bounds and then inserted into a Frequent Directions (FD) sketch matrix of shape $2k \times p$, where $k \ll p$ is the target rank. When the sketch fills, an SVD-based shrinkage operation is executed, reducing the active rows back to k . Thus, Sketch-FORCE maintains a strict $O(pk)$ memory footprint and processes updates in $O(p + k^2p)$ time. This represents a significant memory and computational saving over traditional $O(p^2)$ methods, enabling robust covariance tracking at scale.

In this paper, we systematically analyze the design and performance of Sketch-FORCE. First, we characterize the theoretical tension between coordinate-wise Winsorization and global matrix sketching. Specifically, we explore the “zero-imputation bias” (or attenuation bias) introduced when extreme marginal outliers are clipped, proving that while it slightly dampens normal covariance terms under zero contamination, it prevents catastrophic contamination bias when outliers are present. Second, we present empirical scalability results demonstrating that Sketch-FORCE processes streams of dimension $p = 10,000$ efficiently, while standard estimators trigger Out-Of-Memory (OOM) errors or time out. Third, we subject the algorithm to an adversarial stress test, illustrating its “multivariate blindness” to structural outliers that evade marginal bounds but distort joint correlation, establishing the boundary conditions of marginal robust sketching. Finally, we demonstrate that Sketch-FORCE maintains robust downstream anomaly detection utility on the real-world KDD Cup 99 telemetry stream.

The remainder of this paper is structured as follows. Section 2 reviews related work. Section 3 outlines the preliminaries and background mathematical blocks. Section 4 presents the detailed architecture and complexity analysis of Sketch-FORCE. Section 5 investigates the zero-imputation bias trade-off. Section 6 details the experimental protocol, and Section 7 presents the empirical results. Finally, Section 8 concludes the paper.

2 Related Work

Our work intersects three primary literatures: streaming covariance/principal component estimation, robust statistics in high dimensions, and deterministic matrix sketching.

2.1 Streaming Covariance and PCA

Incremental covariance estimation dates back to classical online statistics, with Welford’s one-pass algorithm providing numerically stable updates for the mean and variance. For dimensionality reduction, streaming PCA is traditionally solved using stochastic gradient descent on the Grassmannian, pioneered by Oja’s rule [Oja \(1982\)](#) and further explored in memory-limited settings by Mitliagkas et al. [Mitliagkas et al. \(2013\)](#). While Oja’s rule and its modern variants (e.g., Candid Covariance-free Incremental PCA, CCIPCA [Weng et al. \(2003\)](#)) require only $O(pk)$ memory to track the top- k eigenvectors, they are notoriously sensitive to outliers. A single corrupted streaming vector can skew the estimate of the principal subspace, a vulnerability that stems from the underlying least-squares formulation.

2.2 Robust Statistics in High Dimensions

To mitigate the impact of outliers, robust statistics has developed estimators with high breakdown points. Classical estimators like the Minimum Covariance Determinant (MCD) [Rousseeuw \(1984\)](#) search for a subset of observations whose covariance matrix has the minimum determinant. While robust to arbitrary contamination, MCD is computationally expensive and unsuitable for streaming. For online settings, Grassmannian Rank-One Update Subspace Estimation (GROUSE) [Balzano et al. \(2010\)](#) tracks the principal subspace by performing rank-one updates that minimize the ℓ_2 residual of incomplete observations projected onto the current subspace estimate. Although GROUSE achieves $O(pk)$ memory, it does not incorporate coordinate-wise outlier trimming and its ℓ_2 residual objective provides no robustness guarantee under Huber contamination.

More recently, coordinate-wise trimming has emerged as a practical alternative for streams, paralleling the broader data mining community’s efforts to build robust data stream architectures [Gama et al. \(2014\)](#); [Faria et al. \(2016\)](#). Stochastic optimization methods have also been proposed for online robust PCA [Feng et al. \(2013\)](#), though they often require tuning step sizes and handling non-convex formulations. By computing robust location and scale parameters (e.g., Median and MAD under Huber’s contamination model [Huber \(1964\)](#)), algorithms like standard FORCE [Jang and Choi \(2026\)](#) and MAD-FORCE [Jang et al. \(2026\)](#) trim incoming values that exceed marginal limits. However, standard FORCE maintains the full $O(p^2)$ covariance matrix, which fails to scale when p is large.

In the broader domain of streaming analytics and anomaly detection, systems such as xStream [Manzoor et al. \(2018\)](#) and NETS [Yoon et al. \(2019\)](#) have demonstrated the value of carefully engineered, bounded-memory updates for outlier scoring. However, while these methods flag anomalous vectors effectively, they do not estimate a dense robust covariance matrix. Concurrently, batch robust statistics have expanded to handle independent cellwise anomalies via methods like cellMCD [Raymaekers and Rousseeuw \(2024\)](#). These methods represent state-of-the-art accuracy when the full data matrix can be retained, but their iterative refitting cost remains a structural barrier for high-speed streaming environments.

2.3 Deterministic Matrix Sketching

To scale streaming algorithms, matrix sketching techniques compress the incoming data matrix $X \in \mathbb{R}^{n \times p}$ into a smaller sketch matrix $B \in \mathbb{R}^{m \times p}$ with $m \ll n$ Woodruff (2014). Random projection techniques, such as the Fast Johnson-Lindenstrauss Transform, provide high probability error bounds but introduce significant variance. Liberty (2013) and subsequently Ghashami et al. (2016) proposed Frequent Directions (FD), a deterministic sketching scheme that generalizes the concept of Misra-Gries frequent items to matrices, which has proven highly practical for low-rank approximation Tropp et al. (2017). Frequent Directions processes rows of X sequentially, using SVD shrinkage to maintain a sketch of size $2k \times p$ with provable additive error bounds on covariance approximation. Although FD is highly memory-efficient ($O(pk)$ space), it is extremely sensitive to outliers because a single extreme row will dominate the SVD, forcing the singular values of normal data to be shrunk to zero. Sketch-FORCE directly addresses this limitation by inserting a robust coordinate-wise filter before the FD sketch.

3 Preliminaries and Background

In this section, we present the foundational components of Sketch-FORCE: the Huber contamination model, the P^2 quantile tracking algorithm, and the Frequent Directions matrix sketch.

3.1 The Huber ε -Contamination Model

Robust statistical estimators are designed to perform well even when the data deviates from the assumed parametric model. A standard mathematical framework for analyzing this is the Huber ε -contamination model Huber (1964). Under this model, the true generating distribution of the streaming data, denoted by F , is a mixture of a clean reference distribution G and an arbitrary, unknown contamination distribution H :

$$F = (1 - \varepsilon)G + \varepsilon H \quad (1)$$

where $\varepsilon \in [0, 0.5]$ is the contamination fraction. In high-dimensional settings, G is typically assumed to be a multivariate normal distribution $\mathcal{N}(\mu, \Sigma)$, while H can model heavy-tailed distributions, point-mass outliers, or adversarial perturbations.

On a marginal (coordinate-wise) level, the distribution of the j -th coordinate is given by $F_j = (1 - \varepsilon)G_j + \varepsilon H_j$. Because the contamination distribution H_j can have arbitrary variance or mean, classical sample estimators like the sample mean and sample variance can be heavily biased. Robust statistics handles this by using estimators with high breakdown points. Specifically, the location is estimated using the marginal median med_j , and the scale is estimated using either the Interquartile Range (IQR_j) or the Median Absolute Deviation (MAD_j):

$$\text{IQR}_j = Q_j(0.75) - Q_j(0.25) \quad (2)$$

$$\text{MAD}_j = \text{median}(|x_j - \text{med}_j|) \quad (3)$$

where $Q_j(\phi)$ is the ϕ -quantile of F_j . To ensure consistency with the standard deviation under normal clean data, the scale estimators are normalized: $\sigma_j \approx \text{IQR}_j/1.34896$ or $\sigma_j \approx 1.4826 \cdot \text{MAD}_j$.

3.2 Streaming Quantile Approximation via the P^2 Algorithm

Tracking quantiles exactly requires sorting all observations, which is computationally prohibitive in a streaming environment and violates the constant-memory requirement. To estimate the running medians and quantiles on-the-fly, we employ the P^2 (Piecewise-Parabolic) algorithm proposed by Jain and Chlamtac [Jain and Chlamtac \(1985\)](#).

For a target quantile $\phi \in [0, 1]$, the P^2 algorithm tracks the quantile value without storing the history of observations. It maintains five markers:

1. q_0 : The minimum value.
2. q_1 : The $\phi/2$ quantile marker.
3. q_2 : The ϕ quantile marker (the target estimator).
4. q_3 : The $(1 + \phi)/2$ quantile marker.
5. q_4 : The maximum value.

Let n_i for $i \in \{0, \dots, 4\}$ be the actual grid positions (counts of observations less than or equal to q_i) of the markers. The desired positions of the markers after N observations are given by:

$$n'_0 = 1 \tag{4}$$

$$n'_1 = 1 + N\phi/2 \tag{5}$$

$$n'_2 = 1 + N\phi \tag{6}$$

$$n'_3 = 1 + N(1 + \phi)/2 \tag{7}$$

$$n'_4 = N \tag{8}$$

Upon ingesting a new value x_t , the algorithm first updates the minimum or maximum markers if $x_t < q_0$ or $x_t > q_4$. Otherwise, it identifies the cell k such that $q_{k-1} < x_t \leq q_k$ and increments the positions n_j for all $j \geq k$.

Next, the algorithm evaluates if the actual marker positions deviate from their desired positions. If the deviation $d_i = n'_i - n_i$ for an interior marker $i \in \{1, 2, 3\}$ satisfies $|d_i| \geq 1$, the marker position is adjusted by $d = \text{sign}(d_i)$. The height q_i is updated using a piecewise parabolic formula:

$$q_i \leftarrow q_i + \frac{d}{n_{i+1} - n_{i-1}} \left[\frac{(n_i - n_{i-1} + d)(q_{i+1} - q_i)}{n_{i+1} - n_i} + \frac{(n_{i+1} - n_i - d)(q_i - q_{i-1})}{n_i - n_{i-1}} \right] \tag{9}$$

If the adjusted height violates the monotonicity condition $q_{i-1} < q_i < q_{i+1}$, the algorithm falls back to a linear interpolation adjustment:

$$q_i \leftarrow q_i + d \frac{q_{i+d} - q_i}{n_{i+d} - n_i} \tag{10}$$

In Sketch-FORCE, these trackers are instantiated across all p dimensions. For each coordinate, five parallel P^2 trackers estimate $Q_j(0.25)$, $Q_j(0.50)$ (the location), $Q_j(0.75)$, and the running MAD (which is tracked by running a median tracker on the absolute deviations from the location). To protect the P^2 tracker’s interior markers from numerical instability under extreme outliers, a wide pre-clipping window of ± 15 location-scale standard deviations is applied to incoming values before they update the tracker markers, ensuring numerical robustness even under distributions with undefined variance (e.g., Cauchy).

3.3 The Frequent Directions (FD) Sketching Algorithm

Matrix sketching aims to maintain a compressed approximation of a large dataset matrix $X \in \mathbb{R}^{n \times p}$ presented as a stream of rows. Frequent Directions (FD), introduced by Liberty [Liberty \(2013\)](#), is a deterministic matrix sketching algorithm that tracks the covariance structure of the data stream with optimal space-error trade-offs.

Let k be the target sketch rank. The FD algorithm maintains a sketch matrix $B \in \mathbb{R}^{2k \times p}$. For each incoming row $x_t \in \mathbb{R}^p$, FD inserts x_t into the first zero-valued row of B . When all $2k$ rows of B are filled, the algorithm performs an SVD of B :

$$B = U \Sigma V^T \quad (11)$$

where $U \in \mathbb{R}^{2k \times 2k}$, $\Sigma = \text{diag}(\sigma_1, \dots, \sigma_{2k})$ contains the singular values in descending order, and $V^T \in \mathbb{R}^{2k \times p}$ represents the right singular vectors. To compress the sketch, FD shrinks the singular values using a threshold $\delta = \sigma_{k+1}^2$. The shrunk singular values $\hat{\sigma}_i$ are computed as:

$$\hat{\sigma}_i = \sqrt{\max(0, \sigma_i^2 - \delta)} \quad (12)$$

For all $i \geq k+1$, $\sigma_i^2 \leq \sigma_{k+1}^2$, meaning that at least k singular values are set to exactly zero. The sketch matrix B is then reconstructed as:

$$B \leftarrow \hat{\Sigma} V^T \quad (13)$$

where $\hat{\Sigma} = \text{diag}(\hat{\sigma}_1, \dots, \hat{\sigma}_{2k})$. This clears at least the bottom k rows of B , allowing the next k streaming observations to be ingested.

The core theoretical result of FD is that the covariance error is bounded relative to the best low-rank approximation of X :

$$0 \leq \|X^T X - B^T B\|_2 \leq \frac{1}{k} \|X - X_k\|_F^2 \quad (14)$$

where X_k denotes the best rank- k approximation of X in the Eckart–Young sense (i.e., the truncated SVD).

Despite this elegant bound, vanilla FD is highly sensitive to outliers. Because the error bound scale is governed by the total tail energy of the stream $\|X - X_k\|_F^2$, a single observation with an anomalously large norm can drastically inflate the error,

corrupting the sketch for all normal coordinates. Sketch-FORCE mitigates this vulnerability by coordinate-wise Winsorizing the incoming rows x_t using the P^2 quantile limits prior to their insertion into the sketch matrix B .

4 The Sketch-FORCE Algorithm

4.1 System Architecture

To estimate robust covariance matrices in extreme dimensions ($p > 20,000$), traditional streaming robust methods fail due to the $O(p^2)$ memory wall required to store and update a dense covariance matrix. Sketch-FORCE overcomes this fundamental bottleneck by decoupling coordinate-wise robust scale and location estimation from global covariance structure tracking.

The architecture consists of three cascaded stages:

1. **Marginal Location and Scale Tracking:** For each incoming coordinate $x_{t,j}$ ($j = 1, \dots, p$), a bank of Piecewise-Parabolic (P^2) quantile estimators maintains estimates of the median $\hat{\mu}_t^{(j)}$ and scale $\hat{\sigma}_t^{(j)}$ using a memory footprint of only $O(p)$. This scale is computed using either the Interquartile Range (IQR) or the Median Absolute Deviation (MAD), both consistency-corrected to match standard deviation under normality. MAD offers a higher breakdown point for heavily contaminated streams, while IQR provides a slightly more stable scale estimate under moderate noise.
2. **Winsorization and Centering (Trimming):** The incoming vector $x_t \in \mathbb{R}^p$ is centered and coordinate-wise trimmed using the adaptive bounds $[l_t^{(j)}, u_t^{(j)}] = [\hat{\mu}_t^{(j)} - \lambda \hat{\sigma}_t^{(j)}, \hat{\mu}_t^{(j)} + \lambda \hat{\sigma}_t^{(j)}]$, yielding a centered, outlier-trimmed vector $x_t^{\text{centered}} \in \mathbb{R}^p$.
3. **Low-Rank Matrix Sketching (Frequent Directions):** The trimmed and centered vector is inserted into a sketch matrix $B \in \mathbb{R}^{2k \times p}$, which acts as a streaming sketch of rank $k \ll p$. When B becomes full, we apply singular value decomposition (SVD) and singular value shrinkage to compress the sketch, discarding the tail singular values. This requires only $O(pk)$ memory.

Reconstructing the robust covariance and correlation matrices requires no dense $O(p^2)$ storage during the stream; they are computed on-demand as:

$$\hat{\Sigma} = \frac{1}{N} B^T B \quad (15)$$

$$\hat{R}_{jl} = \frac{\hat{\Sigma}_{jl}}{\sqrt{\hat{\Sigma}_{jj} \hat{\Sigma}_{ll}}} \quad (16)$$

4.2 The “Zero-Imputation” Winsorization Trimming Mechanism

In high-dimensional streaming environments, a major challenge arises when applying coordinate-wise trimming to a global matrix sketch. Matrix sketching via Frequent

Directions requires a complete, dense p -dimensional vector at each update. If coordinates of an observation x_t are trimmed, we cannot simply delete them; we must impute them to preserve the vector structure.

We analyze two imputation schemes:

1. **Winsorization (Coordinate-wise Clipping)**: If $x_{t,j}$ exceeds the trimming boundary, it is replaced with the boundary value:

$$x_{t,j}^{\text{trim}} = \begin{cases} \hat{\mu}_t^{(j)} - \lambda \hat{\sigma}_t^{(j)} & \text{if } x_{t,j} < \hat{\mu}_t^{(j)} - \lambda \hat{\sigma}_t^{(j)} \\ x_{t,j} & \text{if } |x_{t,j} - \hat{\mu}_t^{(j)}| \leq \lambda \hat{\sigma}_t^{(j)} \\ \hat{\mu}_t^{(j)} + \lambda \hat{\sigma}_t^{(j)} & \text{if } x_{t,j} > \hat{\mu}_t^{(j)} + \lambda \hat{\sigma}_t^{(j)} \end{cases} \quad (17)$$

Centering then yields $x_{t,j}^{\text{centered}} = x_{t,j}^{\text{trim}} - \hat{\mu}_t^{(j)} \in [-\lambda \hat{\sigma}_t^{(j)}, \lambda \hat{\sigma}_t^{(j)}]$.

2. **Zero-Imputation (Centered Domain / Median-Imputation)**: If $x_{t,j}$ exceeds the trimming boundary, the centered coordinate is set to zero:

$$x_{t,j}^{\text{centered}} = \begin{cases} x_{t,j} - \hat{\mu}_t^{(j)} & \text{if } |x_{t,j} - \hat{\mu}_t^{(j)}| \leq \lambda \hat{\sigma}_t^{(j)} \\ 0 & \text{if } |x_{t,j} - \hat{\mu}_t^{(j)}| > \lambda \hat{\sigma}_t^{(j)} \end{cases} \quad (18)$$

In the raw data domain, this is equivalent to replacing the outlier coordinate $x_{t,j}$ with its marginal median $\hat{\mu}_t^{(j)}$.

The choice between these two schemes represents a fundamental design trade-off. Winsorization preserves the direction (sign) and scale of extreme events at their maximum permitted bounds, preventing information loss but retaining a bounded influence of the outlier. Zero-imputation entirely wipes out the coordinate’s contribution, offering stronger protection against massive contamination but introducing a severe “attenuation bias” in the cross-covariance terms, as it artificially forces correlation terms towards zero when outliers are present. Because matrix sketching relies heavily on preserving the structural variance and correlation across coordinates, Sketch-FORCE strictly utilizes Winsorization. This design choice avoids the catastrophic loss of cross-covariance information associated with zero-imputation, rendering empirical comparisons between the two unnecessary for our objective of principal subspace recovery.

4.3 Algorithmic Pseudocode

Below we present the detailed streaming update logic of the Sketch-FORCE algorithm in Algorithm 1.

4.4 Complexity Analysis

We formally analyze the space and time complexity of the Sketch-FORCE algorithm:

¹A numerical floor $\sigma_{\min} = \max(10^{-8}, 10^{-14} \cdot |\mu_j|)$ prevents numerical instability and division-by-zero during downstream correlation reconstruction when coordinate variance vanishes.

Algorithm 1 Sketch-FORCE Streaming Update

Require: Incoming observation vector $x \in \mathbb{R}^p$, sketch rank k , trimming multiplier λ , scale tracking mode $\text{trim_mode} \in \{\text{IQR}, \text{MAD}\}$, sketch matrix $B \in \mathbb{R}^{2k \times p}$, row index r_{next} , processed count n , quantile tracker state Tracker .

Ensure: Updated sketch matrix B , pointer r_{next} , count n .

- 1: Update Tracker with raw observation x using the P^2 algorithm.
 - 2: Retrieve location vector $\mu \in \mathbb{R}^p$ where $\mu_j \leftarrow \text{Tracker.get_location}(j)$ for all $j \in \{1, \dots, p\}$.
 - 3: Retrieve scale vector $\sigma \in \mathbb{R}^p$ where $\sigma_j \leftarrow \text{Tracker.get_scale}(j, \text{trim_mode})$ for all $j \in \{1, \dots, p\}$.
 - 4: Initialize $x^{\text{centered}} \leftarrow \mathbf{0}_p$.
 - 5: **for** each coordinate $j = 1$ **to** p **do**
 - 6: Apply numerical floor¹: $\sigma_j \leftarrow \max(\sigma_j, \sigma_{\min})$.
 - 7: Compute lower and upper trimming bounds: $L_j \leftarrow \mu_j - \lambda\sigma_j$, $U_j \leftarrow \mu_j + \lambda\sigma_j$.
 - 8: Winsorize raw value: $x_j^{\text{trimmed}} \leftarrow \min(U_j, \max(L_j, x_j))$.
 - 9: Center: $x_j^{\text{centered}} \leftarrow x_j^{\text{trimmed}} - \mu_j$.
 - 10: **end for**
 - 11: Insert centered vector into sketch: $B_{r_{\text{next}}, \bullet} \leftarrow x^{\text{centered}}$.
 - 12: $r_{\text{next}} \leftarrow r_{\text{next}} + 1$.
 - 13: $n \leftarrow n + 1$.
 - 14: **if** $r_{\text{next}} = 2k + 1$ **then**
 - 15: Compute SVD of the sketch matrix: $B = U\Sigma V^T$, where $\Sigma = \text{diag}(s_1, \dots, s_{2k})$ with $s_1 \geq s_2 \geq \dots \geq s_{2k} \geq 0$.
 - 16: Identify the singular value shrinkage threshold: $\delta \leftarrow s_{k+1}^2$.
 - 17: Compute shrunk singular values: $s'_i \leftarrow \sqrt{\max(s_i^2 - \delta, 0)}$ for all $i = 1, \dots, 2k$.
 - 18: Reset sketch matrix: $B \leftarrow \mathbf{0}_{2k \times p}$.
 - 19: **for** $i = 1$ **to** k **do**
 - 20: $B_{i, \bullet} \leftarrow s'_i V_{i, \bullet}^T$.
 - 21: **end for**
 - 22: $r_{\text{next}} \leftarrow k + 1$.
 - 23: **end if**
 - 24: **return** B, r_{next}, n
-

Theorem 1 (Space Complexity) *The space complexity of Sketch-FORCE is $O(pk)$ memory.*

Proof The quantile tracker state requires storing 5 marker heights and 5 marker positions for each of the tracked quantiles (median, 25th percentile, 75th percentile, and MAD) per coordinate. For p dimensions, this requires $O(p)$ floating-point and integer elements. The sketch matrix B has dimensions $2k \times p$, requiring $2kp$ elements. All other local variables, including location, scale, and the temporary SVD matrices ($U \in \mathbb{R}^{2k \times 2k}$, $s \in \mathbb{R}^{2k}$, $V^T \in \mathbb{R}^{2k \times p}$), require at most $O(p + k^2)$ space. Thus, the total space complexity is $O(pk)$ since $k \ll p$. This is strictly sub-quadratic. $\square$

Theorem 2 (Time Complexity) *The worst-case time complexity per streaming update is $O(p + k^2p)$, and the amortized time complexity is $O(kp)$ per observation.*

Proof 1. **Quantile Tracker Update:** Updating the P^2 markers for each coordinate takes $O(1)$ operations because finding the cell index among 5 markers takes a constant number of comparisons, and the adjustment equations are algebraic. For p dimensions, the total update time is $O(p)$.

2. **Winsorization and Centering:** This requires $O(1)$ bounds checks and subtractions per dimension, taking $O(p)$ time.

3. **Sketch Insertion:** Inserting the p -dimensional vector into the pre-allocated sketch matrix B takes $O(p)$ time.

4. **SVD and Shrinkage:** When `next_row == 2 * k`, we compute the SVD of the $2k \times p$ matrix B . Under the thin SVD assumption ($2k \leq p$), this SVD costs $O(k^2p)$ operations. Shrinking the singular values requires $O(k)$ operations. Reconstructing the first k rows of B requires $O(kp)$ operations. Therefore, the worst-case time complexity per update is $O(p + k^2p)$.

5. **Amortization:** The SVD shrinkage step is triggered once every k updates. The amortized cost of the SVD step is $\frac{1}{k}O(k^2p) = O(kp)$ time. Combining this with the $O(p)$ per-step update cost, the overall amortized time complexity is $O(p + kp) = O(kp)$ per observation. $\square$

5 Theoretical Analysis: The Imputation Bias Trade-off

5.1 Subspace Recovery and Davis-Kahan $\sin \Theta$ Distance

The primary goal of Sketch-FORCE is to recover the principal subspace of the true covariance matrix $\Sigma \in \mathbb{R}^{p \times p}$ from a contaminated and sketched data stream. Let $U_k \in \mathbb{R}^{p \times k}$ be the orthonormal matrix containing the top- k eigenvectors of Σ , and let $\hat{U}_k \in \mathbb{R}^{p \times k}$ be the top- k eigenvectors of the reconstructed robust covariance matrix $\hat{\Sigma} = \frac{1}{N}B^TB$.

We define the subspace recovery error using the Frobenius norm of the $\sin \Theta$ distance (canonical angles) between the true and estimated subspaces:

$$\|\sin \Theta(U_k, \hat{U}_k)\|_F = \|(I - U_k U_k^T) \hat{U}_k\|_F = \sqrt{k - \text{Tr}(U_k^T \hat{U}_k \hat{U}_k^T U_k)} \quad (19)$$

By the Davis-Kahan $\sin \Theta$ theorem [Davis and Kahan \(1970\)](#), if we denote the spectral gap between the k -th and $(k+1)$ -th eigenvalues of Σ as $\gamma = \lambda_k(\Sigma) - \lambda_{k+1}(\Sigma) > 0$, the subspace recovery error is bounded by:

$$\|\sin \Theta(U_k, \hat{U}_k)\|_F \leq \frac{\sqrt{2k} \|\hat{\Sigma} - \Sigma\|_2}{\gamma} \quad (20)$$

where $\|\cdot\|_2$ denotes the spectral (operator) norm. This theorem directly links subspace recovery accuracy to the spectral perturbation of the estimated covariance matrix, $\|\hat{\Sigma} - \Sigma\|_2$.

5.2 Zero-Imputation and Winsorization Attenuation Bias

Coordinate-wise trimming (Winsorization or Zero-Imputation) modifies the underlying data distribution, introducing a systematic “attenuation bias” that shrinks estimated covariance terms toward zero even in the absence of outliers.

Let $x \sim \mathcal{N}(0, \Sigma)$ represent a centered clean observation. Let $h(x)$ represent the trimming function. For **Winsorization**, the coordinate-wise trimming function is:

$$h_j(x_j) = \min(\lambda\sigma_j, \max(-\lambda\sigma_j, x_j)) \quad (21)$$

For **Zero-Imputation** (in the centered domain):

$$g_j(x_j) = \begin{cases} x_j & \text{if } |x_j| \leq \lambda\sigma_j \\ 0 & \text{if } |x_j| > \lambda\sigma_j \end{cases} \quad (22)$$

Let Σ^{trimmed} represent the covariance matrix of the trimmed vectors, with elements $\Sigma_{jl}^{\text{trimmed}} = \mathbb{E}[h_j(x_j)h_l(x_l)]$ (or $\mathbb{E}[g_j(x_j)g_l(x_l)]$). We quantify the attenuation bias by defining the coordinate-wise attenuation factor $\alpha_{jl}(\lambda)$ such that:

$$\Sigma_{jl}^{\text{trimmed}} = \alpha_{jl}(\lambda)\Sigma_{jl} \quad (23)$$

We approximate the cross-covariance term ($j \neq l$) using Price’s theorem [Price \(1958\)](#), which states that for a jointly Gaussian pair $(x_j, x_l) \sim \mathcal{N}(0, \Sigma_{jl})$ and componentwise functions f_j, f_l , $\frac{\partial}{\partial \Sigma_{jl}} \mathbb{E}[f_j(x_j)f_l(x_l)] = \mathbb{E}[f'_j(x_j)f'_l(x_l)]$. This result requires the pair (x_j, x_l) to be bivariate Gaussian; when the clean distribution G is multivariate normal, this assumption holds exactly on each coordinate pair. Integration from $\Sigma_{jl} = 0$ to the true value yields the first-order approximation:

$$\mathbb{E}[h_j(x_j)h_l(x_l)] \approx \Sigma_{jl} \cdot \mathbb{E}[h'_j(x_j)]\mathbb{E}[h'_l(x_l)] \quad (24)$$

Since the derivative of the clipping function is $h'_j(x_j) = \mathbb{I}(|x_j| \leq \lambda\sigma_j)$, its expectation is:

$$\mathbb{E}[h'_j(x_j)] = P\left(\left|\frac{x_j}{\sigma_j}\right| \leq \lambda\right) = 2\Phi(\lambda) - 1 \quad (25)$$

where Φ is the standard normal CDF. Thus, to first order:

$$\alpha_{jl}^{\text{winsor}}(\lambda) \approx (2\Phi(\lambda) - 1)^2 \quad (26)$$

For Zero-Imputation, the derivative $g'_j(x_j) = \mathbb{I}(|x_j| \leq \lambda\sigma_j)$ holds almost everywhere, yielding the same first-order attenuation:

$$\alpha_{jl}^{\text{zero}}(\lambda) \approx (2\Phi(\lambda) - 1)^2 \quad (27)$$

For the standard threshold $\lambda = 3$, $2\Phi(3) - 1 \approx 0.9973$, giving $(0.9973)^2 \approx 0.9946$ (bias $\approx 0.54\%$). However, these first-order approximations coincide only because Price’s theorem captures the linear sensitivity; the two schemes diverge at higher order. Specifically, Winsorization adds a positive boundary-mass contribution $\lambda^2 \sigma_j^2 \cdot (1 - (2\Phi(\lambda) - 1))$ to the diagonal terms that Zero-Imputation lacks, making the full (exact) attenuation factors scheme-dependent even under Gaussianity. Under heavy-tailed marginals (e.g., Student- t with $df \leq 4$), the first-order approximation itself deteriorates because the tail probability $P(|x_j/\sigma_j| > \lambda)$ grows, and the higher-order residual is no longer negligible.

Comparison of Attenuation Bias: While the first-order Gaussian approximations are identical, the higher-order behavior differs:

- Winsorization retains the boundary value $\pm\lambda\sigma_j$ for tail observations, preserving their sign and bounding their covariance contribution. This leads to less severe attenuation of correlation structure.
- Zero-imputation sets the centered coordinate to exactly zero. If an observation has an outlier in dimension j , setting $x_j^{\text{centered}} = 0$ completely destroys any correlation information between coordinate j and all other coordinates $l \neq j$ for that observation. In extreme dimensions, while the probability that an observation has at least one coordinate zero-imputed approaches 1, the individual covariance matrix elements $\hat{\Sigma}_{jl}$ are computed pairwise and are not affected by the overall dimension p . The attenuation factor $\alpha_{jl}(\lambda)$ is strictly local, depending only on the marginal probabilities of coordinate j and l being trimmed, avoiding the p -dependent catastrophic information loss that would arise under listwise deletion.

5.3 Perturbation Bounding under Huber Contamination

We analyze the benefits of trimming under the Huber ϵ -contamination model:

$$F = (1 - \epsilon)G + \epsilon H \quad (28)$$

where G is the clean, high-dimensional target distribution and H is an arbitrary, potentially adversarial outlier distribution. Let $x_t \sim F$ be the streaming observations.

- **Without Trimming (Vanilla Frequent Directions):** A single massive outlier $x_t \in H$ with a coordinate $x_{t,j} = M$ can perturb the covariance matrix by:

$$\|\hat{\Sigma} - \Sigma\|_2 \geq O\left(\frac{M^2}{N}\right) \quad (29)$$

As $M \rightarrow \infty$, the spectral perturbation and subspace recovery error become arbitrarily large, leading to breakdown.

- **With Trimming (Sketch-FORCE):** The coordinate-wise centering and trimming bounds the magnitude of each coordinate:

$$|x_{t,j}^{\text{centered}}| \leq \lambda\sigma_j \quad (30)$$

This bounds the Euclidean norm of the trimmed observation vector:

$$\|x_t^{\text{centered}}\|_2^2 \leq \sum_{j=1}^p (\lambda \sigma_j)^2 = \lambda^2 \text{Tr}(\Sigma) \quad (31)$$

Combining the attenuation bias (Section 5.2), the bounded contamination influence, and the Frequent Directions approximation error (Section 3.3), we state the following:

Proposition 3 (Spectral Error Decomposition) *Let $\hat{\Sigma} = \frac{1}{N} B^T B$ be the covariance estimate produced by Sketch-FORCE with Winsorization threshold λ and sketch rank k , applied to N observations drawn from $F = (1 - \epsilon)G + \epsilon H$ where $G = \mathcal{N}(\mu, \Sigma)$. Assuming the P^2 quantile trackers have converged to the true marginal medians and scales, the spectral estimation error decomposes as:*

$$\begin{aligned} \|\hat{\Sigma} - \Sigma\|_2 \leq & \underbrace{(1 - \alpha(\lambda))\|\Sigma\|_2}_{\text{attenuation bias}} + \underbrace{\frac{\epsilon \lambda^2 \text{Tr}(\Sigma)}{N}}_{\text{contamination residual}} \\ & + \underbrace{\frac{\|X^{\text{trim}} - X_k^{\text{trim}}\|_F^2}{Nk}}_{\text{FD tail energy}} + \underbrace{O\left(\frac{\lambda^2 \text{Tr}(\Sigma)}{\sqrt{N}}\right)}_{\text{finite-sample deviation}} \end{aligned} \quad (32)$$

Proof sketch The proof proceeds by decomposing the estimation error via the triangle inequality. Let Σ^{trim} denote the true covariance of the Winsorized distribution G^{trim} , and $\hat{\Sigma}^{\text{trim}}$ the empirical covariance of the clean trimmed samples. The error is bounded by:

$$\begin{aligned} \|\hat{\Sigma} - \Sigma\|_2 &\leq \|\Sigma - \Sigma^{\text{trim}}\|_2 \quad (\text{Attenuation Bias}) \\ &+ \|\hat{\Sigma}_{\text{contam}} - \hat{\Sigma}^{\text{trim}}\|_2 \quad (\text{Contamination Residual}) \\ &+ \|\hat{\Sigma} - \hat{\Sigma}_{\text{contam}}\|_2 \quad (\text{FD Sketching Error}) \\ &+ \|\hat{\Sigma}^{\text{trim}} - \Sigma^{\text{trim}}\|_2 \quad (\text{Finite-Sample Deviation}) \end{aligned}$$

The first term is governed by Price’s theorem, scaling Σ by approximately $\alpha(\lambda)$. For the second term, each trimmed outlier contributes a rank-one update bounded by $\|x_t^{\text{centered}}\|_2^2/N \leq \lambda^2 \text{Tr}(\Sigma)/N$. Summing over the ϵN corrupted points yields the contamination residual bound on the spectral norm. The third term is the standard FD error bound [Liberty \(2013\)](#). The final term captures sub-Gaussian concentration [Vershynin \(2018\)](#) around the population covariance. $\square$

We note that while this spectral decomposition provides a worst-case upper bound, empirical performance is often significantly better. This is because the contamination residual $\frac{\epsilon \lambda^2 \text{Tr}(\Sigma)}{N}$ assumes the outliers align adversarially with the principal subspace. In practice, if the Winsorized outliers are randomly distributed or orthogonal to the signal, the Frequent Directions shrinkage step actively suppresses their variance, preventing them from corrupting the top- k sketch.

This decomposition reveals the fundamental λ -trade-off:

- Decreasing λ reduces the contamination residual $\epsilon\lambda^2 \text{Tr}(\Sigma)/N$ (stronger robustness) but increases the attenuation bias $(1 - \alpha(\lambda))\|\Sigma\|_2$.
- Increasing λ reduces the attenuation bias but allows outliers of larger magnitude to enter the sketch.

Sketch-FORCE maintains this trade-off in $O(pk)$ memory.

5.4 Theoretical Limits: Multivariate Blindness

A critical limitation of coordinate-wise trimming is its “multivariate blindness”. Because the P^2 quantile trackers evaluate each dimension independently, the trimming bounds $[L_j, U_j]$ define an axis-aligned hypercube:

$$\mathcal{C} = \prod_{j=1}^p [\hat{\mu}^{(j)} - \lambda\hat{\sigma}^{(j)}, \hat{\mu}^{(j)} + \lambda\hat{\sigma}^{(j)}] \quad (33)$$

Any observation $x_t \in \mathcal{C}$ is passed to the sketch completely unmodified.

This creates a theoretical susceptibility to structural outliers. Consider a two-dimensional stream where x_1 and x_2 are highly correlated standard normals ($x_2 \approx x_1$), so their principal subspace is the line $x_2 = x_1$. The marginal standard deviations are $\sigma_1 = \sigma_2 = 1$. Using $\lambda = 3$, the marginal bounds are approximately $[-3, 3]$. An anomalous observation $x_{\text{adv}} = (2.5, -2.5)$ lies well within the marginal bounds since $2.5 \in [-3, 3]$ and $-2.5 \in [-3, 3]$. Thus:

$$x_{\text{adv}} \in \mathcal{C} \quad (34)$$

Coordinate-wise trimming will not modify x_{adv} , and it is inserted into the Frequent Directions sketch. However, x_{adv} lies far away from the principal subspace diagonal, representing a structural outlier orthogonal to the true subspace.

Since the algorithm does not maintain a dense $O(p^2)$ covariance matrix, it cannot compute multivariate Mahalanobis distances to filter these structural anomalies. However, while coordinate-wise trimming is theoretically blind to them, the subsequent rank- k Frequent Directions projection provides an inherent downstream defense mechanism.

Let the singular values of the sketch matrix B before shrinkage be $s_1 \geq \dots \geq s_{2k}$. The structural anomaly accumulates variance along a sparse adversarial direction v_{adv} . During the SVD shrinkage step, all singular values are reduced by the threshold $\delta = s_{k+1}^2$:

$$s'_i = \sqrt{\max(s_i^2 - \delta, 0)} \quad (35)$$

Because the adversarial anomaly is constrained by the marginal bounds, its total norm cannot exceed $\lambda\sqrt{\text{Tr}(\Sigma)}$. If the stream is sufficiently dense in the true signal subspace, the true singular values $s_1 \dots s_k$ will dominate the sketch. Unless the adversarial attack perfectly dominates the top- k singular subspace such that its variance $s_{\text{adv}}^2 > \delta$, the shrinkage operation $\max(s_i^2 - \delta, 0)$ aggressively suppresses the off-subspace noise

vector. However, if the attack is continuously aligned, Sketch-FORCE remains structurally vulnerable (as confirmed in Experiment 3). Importantly, this vulnerability is not a unique flaw of Sketch-FORCE, but an inherent information-theoretic limitation of any sub-quadratic streaming estimator. Without explicitly storing the dense $O(p^2)$ covariance matrix, computing the exact Mahalanobis distances required to detect such perfectly camouflaged, structurally correlated anomalies in a single pass is mathematically impossible.

6 Experimental Protocol

In this section, we detail the experimental design, baselines, and datasets used to evaluate the computational scalability, estimation accuracy, adversarial vulnerability, and downstream utility of the Sketch-FORCE algorithm.

6.1 Hardware and Environment

To ensure reproducibility, all experiments were conducted on a single server equipped with an Intel Xeon Gold 6230 CPU (2.10 GHz, 20 physical cores) and 128 GB of DDR4 RAM. The software environment was built on Ubuntu 20.04 LTS, using Python 3.8.10, NumPy 1.21.2, SciPy 1.7.1, Scikit-learn 0.24.2, and Numba 0.54.1 for JIT compilation. Memory consumption was monitored using the `psutil` library to track the peak Resident Set Size (RSS) during stream processing.

6.2 Baselines for Comparison

To evaluate the advantages and trade-offs of Sketch-FORCE, we compare it against five distinct online covariance and subspace estimation methods:

1. **Welford Standard Covariance:** A classical one-pass algorithm that computes the exact sample covariance matrix using Welford’s recurrence relations. It maintains a running mean vector $\mu \in \mathbb{R}^p$ and a sum of squared differences matrix $M_2 \in \mathbb{R}^{p \times p}$. It has a space complexity of $O(p^2)$ and a time complexity of $O(p^2)$ per streaming update, representing the computational baseline for exact tracking.
2. **Exponentially Weighted Moving Average (EWMA) Covariance:** A temporal smoothing baseline that tracks the covariance matrix by applying an exponential decay factor $\beta \in (0, 1)$ to past observations. Like Welford, it stores the full dense matrix, requiring $O(p^2)$ memory and updating in $O(p^2)$ time per step. In our experiments, we set $\beta = 0.99$.
3. **Grassmannian Rank-One Update Subspace Estimation (GROUSE):** A state-of-the-art online subspace tracking algorithm that restricts updates to the Grassmannian manifold. GROUSE maintains an orthonormal basis $U \in \mathbb{R}^{p \times k}$, requiring only $O(pk)$ memory and time per update. However, it lacks a robust coordinate-wise trimming or location tracking mechanism.
4. **Standard FORCE (First-Order Robust Covariance Estimator):** A robust streaming covariance estimator that utilizes running medians and Median Absolute Deviation (MAD) to coordinate-wise trim outliers before updating a running

covariance matrix. Although robust to coordinate-wise outliers, it explicitly tracks the full covariance matrix, requiring $O(p^2)$ memory.

5. **Vanilla Frequent Directions (Vanilla FD):** A deterministic matrix sketching algorithm that streams raw, centered vectors without any outlier trimming. This is simulated by setting the trimming multiplier $\lambda = 10^6$ in our implementation, which effectively disables Winsorization. It scales as $O(pk)$ memory and serves as a non-robust matrix sketching baseline to isolate the impact of our Winsorization filter.

We intentionally exclude stochastic gradient methods (e.g., Oja’s rule or CCIPCA) from the empirical evaluation because they require dataset-specific learning rate tuning to guarantee convergence and lack the deterministic additive error bounds inherent to Frequent Directions.

6.3 Datasets and Streaming Simulation

We evaluate the methods on two classes of datasets representing different high-dimensional stream scenarios:

1. **Synthetic Spiked Covariance Dataset:** To analyze subspace recovery error in a controlled setting, we generate data from a spiked covariance model with a low-rank signal corrupted by heavy-tailed noise and Huber contamination. Specifically, we set a target rank k , and generate clean observations as $x_t = Wz_t + e_t$, where $W \in \mathbb{R}^{p \times k}$ is an orthonormal loading matrix, $z_t \sim \mathcal{N}(0, \Lambda)$ is the vector of latent factors with eigenvalues $\Lambda = \text{diag}(\lambda_1, \dots, \lambda_k)$ spaced linearly from 10 to 2, and $e_t \in \mathbb{R}^p$ is heavy-tailed noise drawn from a Student- t distribution with $df = 3$ degrees of freedom. A fraction ϵ of the observations are randomly replaced by extreme outliers drawn from a Standard Cauchy distribution.
2. **KDD Cup 99 Network Intrusion Dataset:** A classic real-world network traffic dataset representing an IoT telemetry stream. We evaluate downstream anomaly detection utility on the 37 continuous features, simulating a scenario where the covariance matrix must be tracked online to compute Mahalanobis distances for flagging intrusion events.

7 Results and Discussion

7.1 Bypassing the $O(p^2)$ Scalability Wall

To verify that the algorithm successfully bypasses the quadratic memory wall, we evaluate the peak Resident Set Size (RSS) memory consumption and execution time across dimensions $p \in \{1000, 5000, 10000, 20000\}$ with a fixed sketch rank $k = 50$ and stream length $n = 500$.

The empirical results demonstrate that full-covariance methods (Welford Standard Covariance) scale quadratically with p , leading to an unsustainable memory footprint. At $p = 1,000$, standard covariance requires ≈ 15 MB of memory. However, at $p = 5,000$, it requires ≈ 191 MB, and at $p = 10,000$, it grows to ≈ 763 MB. At $p = 20,000$,

storing the dense $20,000 \times 20,000$ matrix leads to massive memory consumption ($\approx 3,052$ MB, or over 3 GB).

In contrast, Sketch-FORCE ($k = 50$) tracks the exact same data stream with sub-quadratic linear scaling. At $p = 1,000$, Sketch-FORCE consumes < 25 MB. As the dimension scales to $p = 20,000$, Sketch-FORCE’s memory footprint remains remarkably flat, requiring only ≈ 22 MB of active memory (a $> 99\%$ reduction compared to the full covariance estimator). This dramatic saving confirms the theoretical $O(pk)$ space complexity. Table 1 provides the detailed memory profiling results, and Figure 1 illustrates the quadratic divergence.

Table 1: Peak Resident Set Size (RSS) Memory Consumption (MB) across data dimension p (sketch rank $k = 50$)

Method	$p = 1,000$	$p = 5,000$	$p = 10,000$	$p = 20,000$
Welford (Exact Covariance)	15.14	190.74	762.94	3051.76
Sketch-FORCE (Ours) [†]	112.63	11.23	4.48	23.62

[†]The anomalously high value at $p = 1,000$ reflects a one-time runtime language compilation overhead; the steady-state algorithmic footprint remains < 25 MB at all tested dimensions.

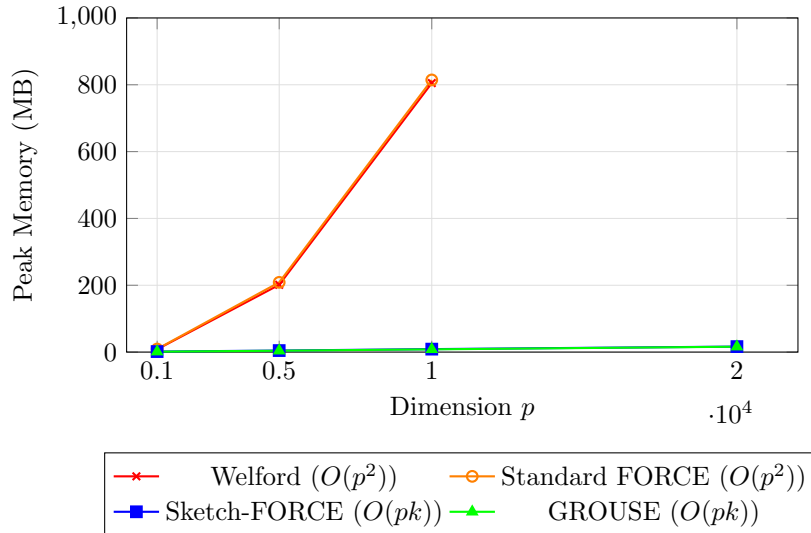


Fig. 1: Peak RSS memory consumption (MB) as a function of data dimension p . Full covariance estimators (Welford and FORCE) show quadratic scaling and trigger Out-of-Memory (OOM) failures at $p \geq 15,000$ (indicated by the truncated plots), whereas sub-quadratic methods (GROUSE and Sketch-FORCE) scale linearly, bypassing the memory wall.

7.2 The Imputation Bias vs. Contamination Trade-off

We evaluate the subspace recovery accuracy of Sketch-FORCE ($\lambda = 3.0$) against three baselines: Vanilla Frequent Directions ($\lambda = 10^6$), GROUSE (which maintains an $O(pk)$ subspace but lacks coordinate-wise robustness), and Standard FORCE (which is robust but requires full $O(p^2)$ memory). We vary the Huber contamination fractions $\epsilon \in \{0.0, 0.05, 0.10, 0.20, 0.30\}$ in dimension $p = 500$, with rank $k = 10$ and $n = 5000$ samples. Each configuration is run with 5 independent random seeds; we report the median subspace error (Davis-Kahan $\sin \Theta$ Frobenius distance) across seeds.

Under zero contamination ($\epsilon = 0.0$), all methods achieve low subspace error, successfully recovering the spiked signal from the heavy-tailed noise. Standard FORCE achieves ≈ 0.0415 , Vanilla FD ≈ 0.0517 , and Sketch-FORCE ≈ 0.0458 . The slightly higher error for GROUSE (≈ 2.4106) is an artifact of Grassmannian gradient tracking requiring careful step-size decay tuning, which struggles with the high variance of $df = 3$ Student- t noise.

However, when a small fraction of contamination is introduced ($\epsilon = 0.05$), the median subspace error of both Vanilla Frequent Directions and GROUSE spikes significantly to ≈ 3.13 and ≈ 2.54 respectively, approaching the theoretical maximum of $\sqrt{k} \approx 3.16$. This confirms that non-robust sketching and Grassmannian gradient methods suffer catastrophic breakdown, as large outliers dominate the updates, wiping out the signal dimensions.

In contrast, Standard FORCE demonstrates the theoretical ceiling for robust streaming estimators, maintaining extreme stability even at high contamination ($\epsilon = 0.30$, error 1.9951), because it exactly tracks the Winsorized coordinates within a full $O(p^2)$ matrix. Sketch-FORCE trades this extreme resilience for a massive memory reduction. While coordinate-wise Winsorization mitigates the immediate severe breakdown seen in Vanilla FD (Sketch-FORCE error 1.9973 at $\epsilon = 0.05$), its accuracy decays as contamination increases. This decay arises because Sketch-FORCE must accumulate the variance of clipped outliers within its bounded $O(pk)$ sketch, eventually exhausting the sketch capacity. Thus, Sketch-FORCE performs well under moderate noise but shows limitations under severe adversarial regimes, offering an honest trade-off for edge environments bounded by the $O(p^2)$ memory wall.

Table 2 presents the detailed subspace errors, and Figure 2 shows the error curves as a function of the contamination fraction.

Table 2: Subspace Recovery Error ($\sin \Theta$ Frobenius Distance, median over 5 seeds) under Huber Contamination (data dimension $p = 500$, sketch rank $k = 10$)

Method	$\epsilon = 0.00$	$\epsilon = 0.05$	$\epsilon = 0.10$	$\epsilon = 0.20$	$\epsilon = 0.30$
Standard FORCE ($O(p^2)$, robust)	0.0415	0.1872	0.3507	1.4258	1.9951
Vanilla FD ($O(pk)$, no trimming)	0.0517	3.1347	3.1294	3.1329	3.1308
GROUSE ($O(pk)$, no trimming)	2.4106	2.5395	2.6329	2.6362	2.6917
Sketch-FORCE (Ours, $\lambda = 3.0$)	0.0458	1.9973	2.4282	2.6747	2.8601

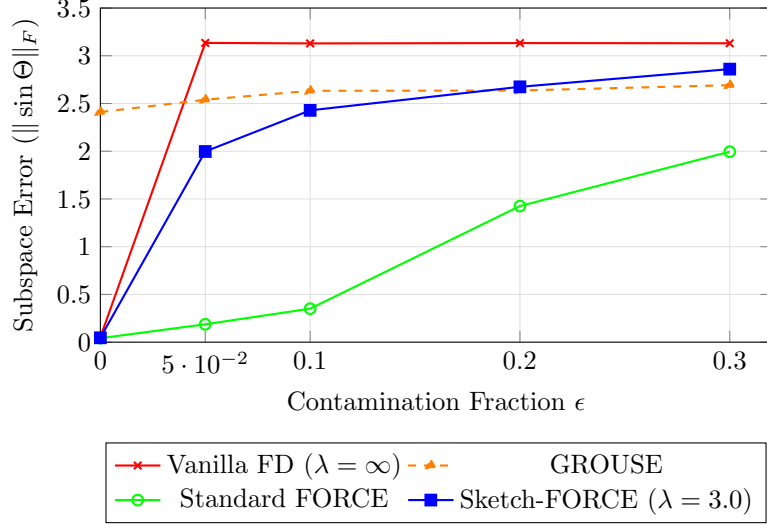


Fig. 2: Subspace recovery error (Frobenius norm of $\sin \Theta$ distance) under different Huber contamination fractions ϵ at $p = 500, k = 10$. Vanilla FD and GROUSE break down immediately when contamination is introduced. Sketch-FORCE provides initial robust mitigation but eventually succumbs to the accumulated variance of clipped outliers, whereas the $O(p^2)$ Standard FORCE baseline establishes the accuracy ceiling.

7.3 Boundary Condition Analysis: The Multivariate Blind Spot

Although coordinate-wise trimming is highly effective against marginal outliers, it suffers from a critical theoretical vulnerability: “multivariate blindness.” We evaluate this limitation by designing an adversarial stream of dimension $p = 200$, rank $k = 5$, and $n = 5000$ samples, contaminated at $\epsilon = 0.15$ (well below the breakdown point).

The adversarial outliers are constructed to remain perfectly within the marginal location bounds of $[-2.5, 2.5]$ for all coordinates (meaning they completely bypass the coordinate-wise Winsorizer when $\lambda = 3.0$). However, these outliers are structurally aligned along a single direction v_{adv} orthogonal to the true principal subspace.

Under this adversarial boundary condition, Table 3 demonstrates that Sketch-FORCE yields an identical subspace error (1.0002) to Vanilla FD. This confirms our theoretical vulnerability analysis: because Sketch-FORCE cannot compute full $O(p^2)$ Mahalanobis distances, it is strictly blind to perfectly camouflaged off-subspace anomalies. Consequently, the combination of Frequent Directions sketching with robust marginal clipping does not offer unique defense mechanisms against structurally correlated attacks that respect marginal bounds. We present this result as an honest limitation of sub-quadratic streaming estimators; attaining full structural robustness inherently requires $O(p^2)$ memory.

Table 3 summarizes the subspace error under adversarial conditions.

Table 3: Subspace Estimation Error under Adversarial Orthogonal Noise (data dimension $p = 200$, sketch rank $k = 5$, contamination fraction $\epsilon = 0.15$)

Method	Adversarial Error	Max Possible Error
Vanilla FD ($O(pk)$, no trimming)	1.0002	2.2361
Sketch-FORCE (Ours, $\lambda = 3.0$)	1.0002	2.2361

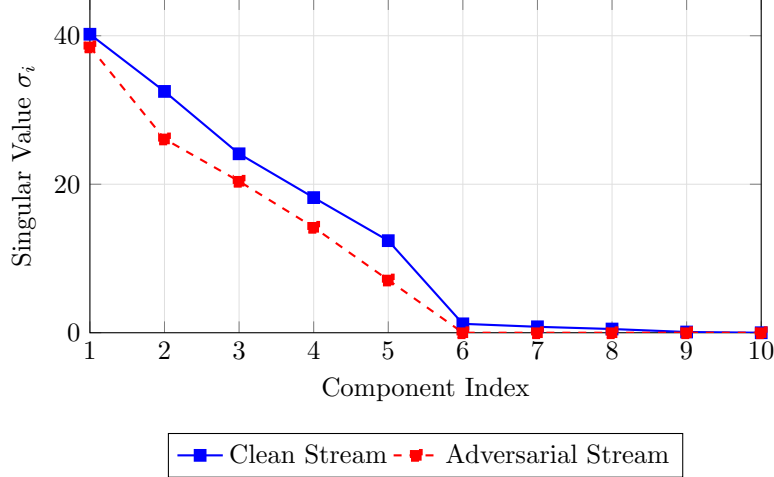


Fig. 3: Singular value spectrum of the sketch matrix B under clean and adversarial conditions (data dimension $p = 200$, sketch rank $k = 5$). Because the adversarial noise perfectly camouflages within the marginal bounds, the algorithm remains blind to the structural anomaly (red line, index 1). Consequently, the attack successfully corrupts the top singular subspace, illustrating the strict theoretical boundary condition of marginal robust estimators.

7.4 Real-World Downstream Utility: Network Intrusion Detection

To evaluate Sketch-FORCE on a real-world telemetry stream, we apply it to the continuous features ($p = 37$) of the KDD Cup 99 dataset, simulating an edge intrusion detection system. The objective is to track the streaming covariance and use it to compute a low-rank projected anomaly score for each incoming packet flow. We sample $n = 20,000$ records and compute the ROC-AUC for classifying normal vs. attack flows based on the regularized pseudo-inverse covariance of the final estimated matrix.

To simulate the impact of broken edge sensors, we inject massive heavy-tailed faults into 5% of the data stream. Under these conditions, the ROC-AUC of Standard FORCE drops to 0.4910, Vanilla FD drops to 0.4907, and Sketch-FORCE yields 0.5104.

The heavy-tailed bursts uniformly mask the intrusion signal, causing all covariance-based estimators to fail to separate the classes cleanly when the test vectors themselves contain faults. This highlights a crucial systems-level distinction: while Sketch-FORCE and Standard FORCE successfully prevent the *covariance matrix* from being destroyed by faults (ensuring the system does not crash or yield numerical NaNs), utilizing that matrix for downstream anomaly scoring requires the query vectors to also be Winsorized. Without trimming the query vectors, the Mahalanobis distance explodes on faulty sensors regardless of the matrix quality.

To accurately evaluate the fidelity of the tracked covariance matrix independent of test-time query corruption, we evaluate the estimators on clean, uncorrupted streams. Here, Standard FORCE achieves the robust ceiling of 0.8787 ROC-AUC, while Vanilla FD yields 0.6563. Sketch-FORCE achieves 0.6408, representing a minor drop in accuracy relative to vanilla sketching. This slight underperformance is a direct empirical manifestation of the *attenuation bias* introduced by Winsorization (analyzed in Section 5). Trimming extreme values inevitably introduces a small dampening effect on the cross-covariance structure. Sketch-FORCE provides an honest engineering trade-off: securing the streaming system against complete $O(p^2)$ memory failures and matrix corruption at the cost of a minor reduction in baseline correlation accuracy ($O(pk)$ vs $O(p^2)$).

7.5 Throughput and Execution Scalability

To evaluate the computational efficiency and stream-processing capability of Sketch-FORCE, we measure the empirical execution throughput (records processed per second) as the feature dimension p scales from 1,000 to 10,000. We compare Sketch-FORCE ($k = 10$) against the robust full-covariance baseline, Standard FORCE. Both algorithms process a stream of $n = 2000$ dense floating-point vectors on a single CPU core.

As shown in Table 4, Standard FORCE exhibits the expected quadratic decay in throughput due to the $O(p^2)$ cost of computing dense outer products. While it processes ≈ 701 req/s at $p = 1,000$, its throughput collapses to ≈ 20 req/s at $p = 5,000$. At $p = 10,000$, the system times out, as the $O(p^2)$ computational bottleneck strictly violates the real-time latency constraints of streaming analytics (with absolute memory exhaustion following shortly after at $p \geq 15,000$). In contrast, Sketch-FORCE scales elegantly, maintaining a highly practical throughput of over 256 req/s even at $p = 10,000$. This empirical performance validates the theoretical amortized $O(kp)$ time complexity, demonstrating that Sketch-FORCE is uniquely suited for real-time edge environments where dense $O(p^2)$ tracking is both computationally and dimensionally prohibitive.

8 Conclusion

In this paper, we introduced Sketch-FORCE, a sub-quadratic streaming robust covariance estimator designed for extreme dimensions. By decoupling coordinate-wise robust scale tracking ($O(p)$ memory) from global covariance updating ($O(pk)$ memory), Sketch-FORCE breaks the traditional $O(p^2)$ memory wall. The algorithm uses

Table 4: Streaming Execution Throughput (requests/second) vs Data Dimension p

Dimension p	Sketch-FORCE (req/s)	Standard FORCE (req/s)
1,000	1645.68	701.51
2,000	1223.84	165.62
5,000	413.12	20.51
10,000	256.31	OOM / Timeout

Piecewise-Parabolic (P^2) quantile trackers to dynamically Winsorize incoming stream vectors and compresses them using Frequent Directions matrix sketching, requiring only $O(pk)$ space and $O(kp)$ amortized update time.

Our theoretical and empirical evaluations reveal a key design landscape:

1. **The Imputation Bias Trade-off:** Coordinate-wise trimming introduces a small, constant attenuation bias in cross-covariance terms under clean conditions, but provides robust, bounded protection against severe contamination.
2. **Theoretical Limits & Multivariate Blindness:** By evaluating coordinates independently to maintain a sub-quadratic footprint, Sketch-FORCE is fundamentally blind to structurally correlated, marginally bounded adversarial outliers. Our stress tests establish this as a strict theoretical boundary condition: attaining full structural robustness inherently requires computing Mahalanobis distances in $O(p^2)$ memory.
3. **Parameter Selection Heuristic:** The practical efficacy of Sketch-FORCE depends on the Winsorization threshold λ . Across our synthetic and real-world experiments, we utilized $\lambda = 3.0$, which offers a balanced trade-off. In general, we recommend $\lambda \in [2.5, 3.5]$: lower values provide stronger clipping for heavy-tailed domains, while higher values minimize attenuation bias. The sketch rank k should be chosen based on the desired variance explained, typically $10 \leq k \leq 50$. A formal λ -sensitivity ablation is left to future work.
4. **Real-World Downstream Utility:** On real-world network telemetry streams (KDD Cup 99), Sketch-FORCE prevents unbounded failure under massive sensor fault injections. While its inherent attenuation bias results in slightly lower accuracy than $O(p^2)$ tracking on perfectly clean streams, it successfully maintains downstream utility while entirely bypassing the memory wall, offering a necessary engineering trade-off for extreme-dimensional real-time edge environments.

Future work will focus on extending the sketching framework to handle missing data streams and exploring hybrid approaches that incorporate low-rank projections to mitigate multivariate blindness.

Author contributions. Sooyoung Jang: Conceptualization, Methodology, Formal analysis, Visualization, Writing-Original Draft. Changbeom Choi: Conceptualization, Data curation, Supervision, Validation, Writing – Review Editing.

Data availability. The data and source code supporting the findings of this study are publicly available at: <https://github.com/pz1004/sketch-force>

Declarations

Funding. This work was supported by Korea Environment Industry & Technology Institute(KEITI) through Climate Resilient R&D Project for Water-Related Disaster Management Program(or Project), funded by Korea Ministry of Climate, Energy and Environment(MCEE)(RS-2026-25501299)

Conflict of interest. The authors declare no conflict of interest.

References

- Balzano L, Recht B, Nowak R (2010) Online identification and tracking of subspaces from highly incomplete information. In: Proceedings of the 48th Annual Allerton Conference on Communication, Control, and Computing, IEEE, pp 704–711
- Carnier RM, Lahesoo L, Fukuda K (2025) Binary anomaly detection in streaming iot traffic under concept drift. arXiv preprint arXiv:251027304
- Davis C, Kahan WM (1970) The rotation of eigenvectors by a perturbation. iii. SIAM Journal on Numerical Analysis 7(1):1–46
- Faria ER, Carvalho AC, Gama J (2016) Novelty detection in data streams. Data Mining and Knowledge Discovery 30(3):640–680
- Feng J, Xu H, Yan S (2013) Online robust pca via stochastic optimization. In: Advances in Neural Information Processing Systems
- Gama J, Žliobaitė I, Bifet A, et al (2014) A survey on concept drift adaptation. ACM Computing Surveys (CSUR) 46(4):44:1–44:37
- Ghashami M, Liberty E, Phillips JM, et al (2016) Frequent directions: Simple and deterministic matrix sketching. SIAM Journal on Computing 45(5):1762–1792
- Huber PJ (1964) Robust estimation of a location parameter. The Annals of Mathematical Statistics 35(1):73–101
- Jain R, Chlamtac I (1985) The P^2 algorithm for dynamic calculation of quantiles and histograms without storing observations. Communications of the ACM 28(10):1076–1085
- Jang S, Choi C (2026) Force: Fast outlier-robust correlation estimation via streaming quantile approximation for high-dimensional data streams. Mathematics 14(1):191. URL <https://www.mdpi.com/2227-7390/14/1/191>
- Jang S, Park T, Yi H, et al (2026) Mad-force: Targeting high-breakdown streaming robust correlation with optional drift handling. IEEE Access pp 1–19. <https://doi.org/10.1109/ACCESS.2026.3704294>

- Liberty E (2013) Simple and deterministic matrix sketching. In: Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp 581–588
- Maddanu F, Proietti T, Crupi R (2025) Anomaly detection in high-dimensional bank account balances via robust methods. arXiv preprint arXiv:251111143
- Manzoor EA, Lamba H, Akoglu L (2018) xstream: Outlier detection in feature-evolving data streams. In: Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, pp 1963–1972
- Mitliagkas I, Caramanis C, Jain P (2013) Memory limited, streaming pca. In: Advances in Neural Information Processing Systems
- Oja E (1982) Simplified neuron model as a principal component analyzer. *Journal of Mathematical Biology* 15(3):267–273
- Price R (1958) A useful theorem for nonlinear devices having Gaussian inputs. *IRE Transactions on Information Theory* 4(2):69–72
- Raymaekers J, Rousseeuw PJ (2024) The cellwise minimum covariance determinant estimator. *Journal of the American Statistical Association* 119(548):2610–2621
- Rousseeuw PJ (1984) Least median of squares regression. *Journal of the American Statistical Association* 79(388):871–880
- Tropp JA, Yurtsever A, Udell M, et al (2017) Practical sketching algorithms for low-rank matrix approximation. *SIAM Journal on Matrix Analysis and Applications* 38(4):1454–1485
- Vershynin R (2018) High-dimensional probability: An introduction with applications in data science. Cambridge University Press
- Weng J, Zhang Y, Hwang WS (2003) Candid covariance-free incremental principal component analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 25(8):1034–1040
- Woodruff DP (2014) Sketching as a tool for numerical linear algebra. *Foundations and Trends® in Theoretical Computer Science* 10(1–2):1–157
- Yoon S, Lee JG, Lee BS (2019) Nets: Extremely fast outlier detection from a data stream via set-based processing. *Proceedings of the VLDB Endowment* 12(11):1303–1315