Online PCA with Spectral Bounds
Zohar Karnin and Edo Liberty, July 4th 2015
Principal Component Analysis

Problem Definition

Given $X \in \mathbb{R}^{d \times n}$ and $k < d$ minimize over $Y \in \mathbb{R}^{k \times n}$

$$\min_{\Phi} \|X - \Phi Y\|^2_F \quad \text{or} \quad \min_{\Phi} \|X - \Phi Y\|^2_2$$

We think of $X = [x_1, x_2, \ldots]$ and $Y = [y_1, y_2, \ldots]$ as collections of column vectors.
Let $U_k$ span the top $k$ left singular vectors of $X$.

Setting $Y = U_k^T X$ and $\Phi = U_k$ minimizes $\min_{\Phi} \| X - \Phi Y \|_F^2$. 

Setting $Y = U_k^T X$ and $\Phi = U_k$ minimizes $\min_{\Phi} \| X - \Phi Y \|_2^2$.

- Computing $U_k$ is possible offline using the Singular Value Decomposition.
- The same solution achieves the optimal value for both objectives.
- The optimal reconstruction $\Phi$ turns out to be an isometry.
Online PCA

Consider clustering the reduced dimensional vectors online (e.g. [Mey01, LSS14])

\[ x_t \in \mathbb{R}^d \quad \rightarrow \quad y_t \in \mathbb{R}^k \quad \rightarrow \quad \text{Cluster identifier} \]

The PCA algorithm must output \( y_t \) before receiving \( x_{t+1} \).
### Online PCA, Possible Problem Definitions

**Regret minimization:** Minimizes $\sum_t \| x_t - P_{t-1} x_t \|^2$. Commits to $P_{t-1}$ before observing $x_t$.  
[WK06, NKW13]

**Random projection:** can guarantee online that $\| (X - (XY^+) Y) \|^2_F$ is small.  
[Sar06, CW09]

**Stochastic model:** Bounds $\| X - \Phi Y \|^2_F$, assumes $x_t$ are i.i.d. from an unknown distribution.  
[OK85, ACS13, MCJ13, BDF13]

**Adversarial model:** Bounds $\| X - \Phi Y \|^2_F$ in the adversarial setting.  
[BGKL15]

Regardless of problem definition, all previous work focused on the Frobenius loss.
Noisy Data Spectra

- **Signal**
- **Data**
- **Noise**
Online PCA Problem Definitions

Main Contribution [KL15]

There exists an algorithm that receives $x_t \in \mathbb{R}^d$ and $k < d$ and

- yields $y_t \in \tilde{O}(k/\varepsilon^2)$ before observing $x_{t+1}$.
- guarantees that $\|X - \Phi Y\|_2^2 \leq \sigma_k^2 + \varepsilon \sigma_1^2$ for some isometry $\Phi$.

\[
\Delta = \sigma_{k+1}^2 + \varepsilon \sigma_1^2 \\
U \leftarrow \text{all zeros matrix}
\]

\begin{align*}
\text{for } x_t \in X & \text{ do} \\
& \text{if } \|(I - UU^T)X_{1:t}\|_2^2 \geq \Delta \\
& \quad \text{Add the top left singular vector of } (I - UU^T)X_{1:t} \text{ to } U \\
& \text{yield } y_t = U^T x_t
\end{align*}

There are obvious problems with this algorithm. We will be fixed those later...
Algorithm Intuition

Assume we know $\Delta = \sigma^2_{k+1} + \varepsilon \sigma^2_1$. 
Algorithm Intuition

We start with mapping $x_t \mapsto 0$ and $R_{[1:t]} = X_{[1:t]}$
Algorithm Intuition

This is continued as long as $\|R^T R\| \leq \Delta$
Algorithm Intuition

When $\|R^T R\| > \Delta$ we commit to a new online PCA direction $u_i$. 
Algorithm Intuition

This prevents $R^T R$ from growing more in the direction $u_i$. 
Algorithm Properties

Theorems 2,5 and 6 in [KL15]

\[ \|X - UY\|^2_2 \leq \|R\|^2_2 \leq \sigma_k^2 + \varepsilon \sigma_1^2 + o(\sigma_1^2) . \]

``Proof by drawing'' above is deceptively simple. This is the main difficulty!

Theorem 1 in [KL15]

Number of direction added by the algorithm is at most \( k/\varepsilon \).

(This is actually not very hard to show)
Fixing the Algorithm

- Exponentially search for the right $\Delta$.
  If we added more than $k/\varepsilon$ direction to $U$ we can conclude that $\Delta < \sigma_{k+1}^2 + \varepsilon\sigma_1^2$.

- Instead of keeping $X_{1:t}$ use covariance sketching [Lib13].
  This keeps $B$ such that $XX^T \sim BB^T$ and $B$ required $o(d^2)$ to store.

- Only compute the top singular value of $(I - UU^T)X_{1:t}$ "once in a while".
Can we reduce the target dimension while keeping the approximation guarantee?

Would allowing *scaled* isometric registration help reduce the target dimension?

Can we avoid the exponential search for \( \Delta \)?

Is there a simple way to update \( U \) that is more accurate than only adding columns?

Can we reduce the running time of online PCA? Currently the bottleneck is covariance sketching.
Thank you
Raman Arora, Andy Cotter, and Nati Srebro.
Stochastic optimization of pca with capped msg.

Akshay Balsubramani, Sanjoy Dasgupta, and Yoav Freund.
The fast convergence of incremental pca.

Christos Boutsidis, Dan Garber, Zohar Shay Karnin, and Edo Liberty.
Online principal components analysis.

Kenneth L. Clarkson and David P. Woodruff.
Numerical linear algebra in the streaming model.

Kenneth L. Clarkson and David P. Woodruff.
Low rank approximation and regression in input sparsity time.

Alan Frieze, Ravi Kannan, and Santosh Vempala.
Fast monte-carlo algorithms for finding low-rank approximations.

Mina Ghashami, Edo Liberty, Jeff M. Phillips, and David P. Woodruff.
Frequent directions : Simple and deterministic matrix sketching.

Mina Ghashami and Jeff M. Phillips.
Relative errors for deterministic low-rank matrix approximations.
Zohar Shay Karnin and Edo Liberty.
Online PCA with spectral bounds.
COLT, 2015.

Edo Liberty, Ram Sriharsha, and Maxim Sviridenko.
An algorithm for online k-means clustering.

Ioannis Mitliagkas, Constantine Caramanis, and Prateek Jain.
Memory limited, streaming pca.

Adam Meyerson.
Online facility location.

Jiazhong Nie, Wojciech Kotlowski, and Manfred K. Warmuth.
Online PCA with optimal regrets.

Erkki Oja and Juha Karhunen.
On stochastic approximation of the eigenvectors and eigenvalues of the expectation of a random matrix.

Roberto Imbuzeiro Oliveira.
Sums of random hermitian matrices and an inequality by rudelson.
Mark Rudelson and Roman Vershynin.
Sampling from large matrices: An approach through geometric functional analysis.

Tamas Sarlos.
Improved approximation algorithms for large matrices via random projections.

Roman Vershynin.
A note on sums of independent random matrices after ahlswede-winter.
*Lecture Notes*.

Manfred K. Warmuth and Dima Kuzmin.
Randomized PCA algorithms with regret bounds that are logarithmic in the dimension.

Franco Woolfe, Edo Liberty, Vladimir Rokhlin, and Mark Tygert.
A fast randomized algorithm for the approximation of matrices.

David P. Woodruff.
Low rank approximation lower bounds in row-update streams.

Edo Liberty.
Simple and deterministic matrix sketching.