Online PCA with Spectral Bounds

Zohar Karnin and Edo Liberty, July 4th 2015

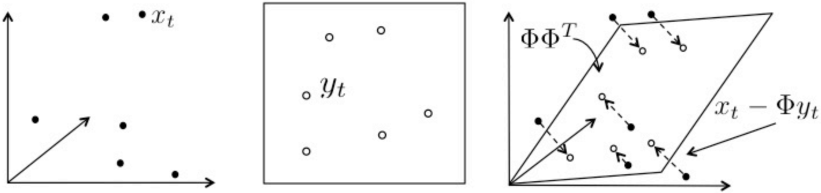
Principal Component Analysis

Problem Definition

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

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

We think of $X = [x_1, x_2, \ldots]$ and $Y = [y_1, y_2, \ldots]$ as collections of column vectors.



Optimal Offline Solution

Let U_k span the top k left singular vectors of X

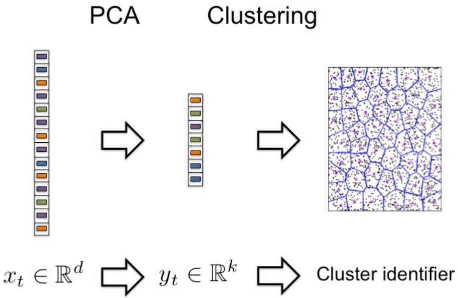
Setting $Y = U_k^T X$ and $\Phi = U_k$ minimizes $\min_\Phi \|X - \Phi Y\|_F^2$

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 Φ turns out to be an isometry.

Online PCA

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



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]

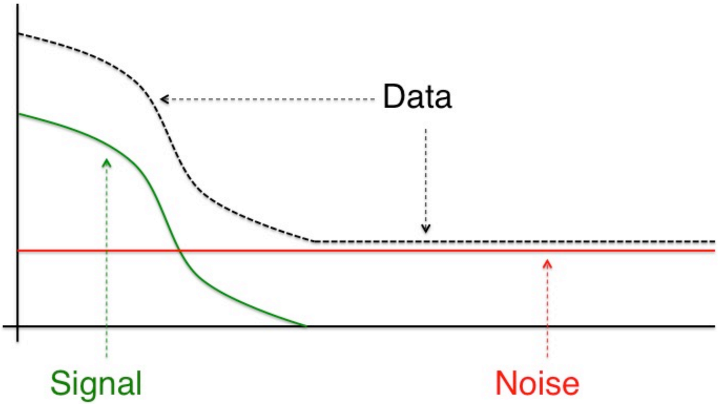
Random projection: can guarantee online that $\|(X - (XY^{+})Y\|_F^2$ is small.
[Sar06, CW09]

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

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

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

Noisy Data Spectra



Online PCA Problem Definitions

Main Contribution [KL15]

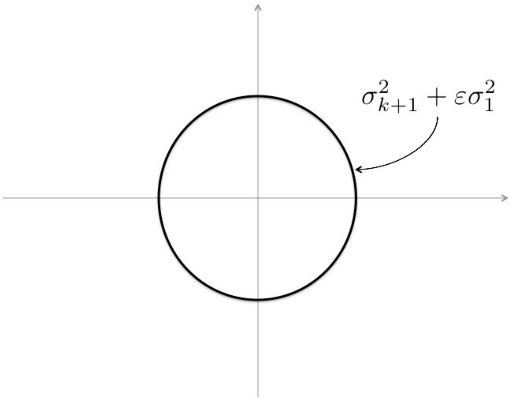
There exists an algorithm that receives $x_t \in \mathbb{R}^d$ and $k < d$ and

- 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 Φ.

$\Delta = \sigma_{k+1}^2 + \varepsilon \sigma_1^2$
$U \leftarrow$ all zeros matrix
for $x_t \in X$ **do**
    if $\|(I - UU^T)X_{1:t}\|^2 \geq \Delta$
        Add the top left singular vector of $(I - UU^T)X_{1:t}$ to U
    yield $y_t = U^T x_t$

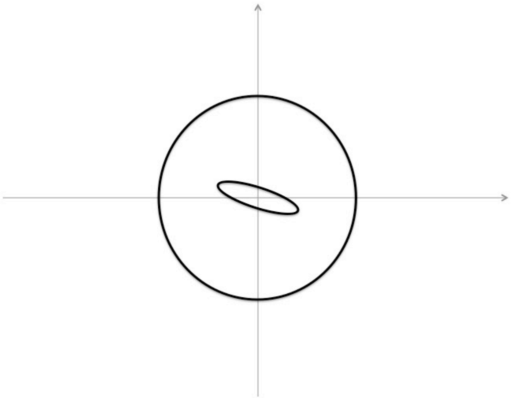
There are obvious problems with this algorithm. We will be fixed those later...

Algorithm Intuition



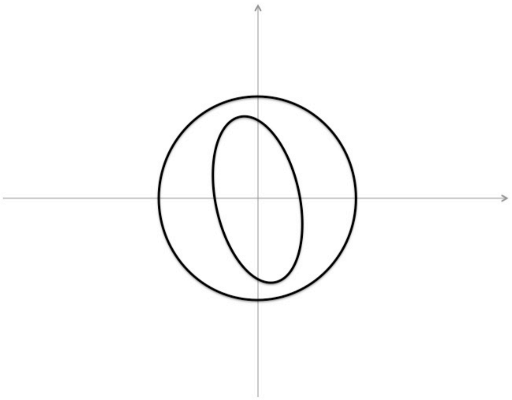
Assume we know $\Delta = \sigma_{k+1}^2 + \varepsilon\sigma_1^2$.

Algorithm Intuition



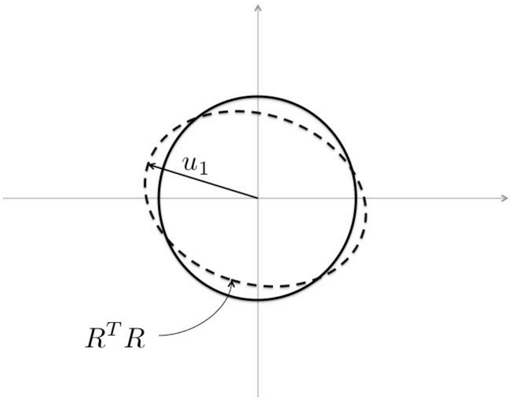
We start with mapping $x_t \mapsto 0$ and $R_{[1:t]} = X_{[1:t]}$

Algorithm Intuition



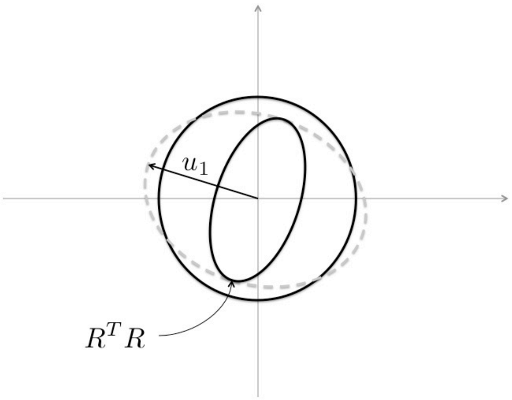
This is continued as long as $||R^T R|| \leq \Delta$

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!

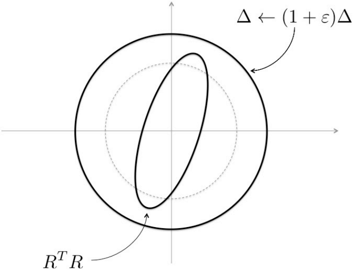
Theorem 1 in [KL15]

Number of direction added by the algorithm is at most k/ε.

(This is actually not very hard to show)

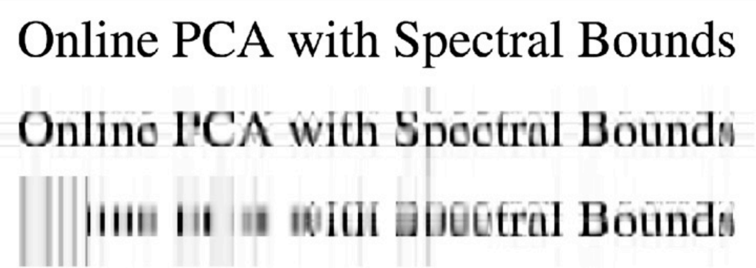
Fixing the Algorithm

- Exponentially search for the right Δ.
  If we added more than k/ε 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''.

Visual Illustration and Open Problem



- 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 Δ?
- 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.
In C.J.C. Burges, L. Bottou, M. Welling, Z. Ghahramani, and K.Q. Weinberger, editors, *Advances in Neural Information Processing Systems 26*, pages 1815--1823. 2013.

Akshay Balsubramani, Sanjoy Dasgupta, and Yoav Freund.
The fast convergence of incremental pca.
In C.J.C. Burges, L. Bottou, M. Welling, Z. Ghahramani, and K.Q. Weinberger, editors, *Advances in Neural Information Processing Systems 26*, pages 3174--3182. 2013.

Christos Boutsidis, Dan Garber, Zohar Shay Karnin, and Edo Liberty.
Online principal components analysis.
In *Proceedings of the Twenty-Sixth Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2015, San Diego, CA, USA, January 4-6, 2015*, pages 887--901, 2015.

Kenneth L Clarkson and David P Woodruff.
Numerical linear algebra in the streaming model.
In *Proceedings of the 41st annual ACM symposium on Theory of computing*, pages 205--214. ACM, 2009.

Kenneth L. Clarkson and David P. Woodruff.
Low rank approximation and regression in input sparsity time.
CoRR, abs/1207.6365, 2012.

Alan Frieze, Ravi Kannan, and Santosh Vempala.
Fast monte-carlo algorithms for finding low-rank approximations.
J. ACM, 51(6):1025--1041, November 2004.

Mina Ghashami, Edo Liberty, Jeff M. Phillips, and David P. Woodruff.
Frequent directions : Simple and deterministic matrix sketching.
CoRR, abs/1501.01711, 2015.

Mina Ghashami and Jeff M. Phillips.
Relative errors for deterministic low-rank matrix approximations.
In *Proceedings of the Twenty-Fifth Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2014, Portland, Oregon, USA, January 5-7, 2014*, pages 707--717, 2014.

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.
CoRR, abs/1412.5721, 2014.

Ioannis Mitliagkas, Constantine Caramanis, and Prateek Jain.
Memory limited, streaming pca.
In C.J.C. Burges, L. Bottou, M. Welling, Z. Ghahramani, and K.Q. Weinberger, editors, *Advances in Neural Information Processing Systems 26*, pages 2886--2894. 2013.

Adam Meyerson.
Online facility location.
In *42nd Annual Symposium on Foundations of Computer Science, FOCS 2001, 14-17 October 2001, Las Vegas, Nevada, USA*, pages 426--431, 2001.

Jiazhong Nie, Wojciech Kotlowski, and Manfred K. Warmuth.
Online PCA with optimal regrets.
In *ALT*, pages 98--112, 2013.

Erkki Oja and Juha Karhunen.
On stochastic approximation of the eigenvectors and eigenvalues of the expectation of a random matrix.
Journal of Mathematical Analysis and Applications, 106(1):69 -- 84, 1985.

Roberto Imbuzeiro Oliveira.
Sums of random hermitian matrices and an inequality by rudelson.
arXiv:1004.3821v1, April 2010.

Mark Rudelson and Roman Vershynin.
Sampling from large matrices: An approach through geometric functional analysis.
J. ACM, 54(4), July 2007.

Tamas Sarlos.
Improved approximation algorithms for large matrices via random projections.
In *FOCS '06: Proceedings of the 47th Annual IEEE Symposium on Foundations of Computer Science*, pages 143--152, Washington, DC, USA, 2006.

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.
In *Advances in Neural Information Processing Systems 19, Proceedings of the Twentieth Annual Conference on Neural Information Processing Systems, Vancouver, British Columbia, Canada, December 4-7, 2006*, pages 1481--1488, 2006.

Franco Woolfe, Edo Liberty, Vladimir Rokhlin, and Mark Tygert.
A fast randomized algorithm for the approximation of matrices.
Applied and Computational Harmonic Analysis, 25(3):335 -- 366, 2008.

David P. Woodruff.
Low rank approximation lower bounds in row-update streams.
In *Advances in Neural Information Processing Systems 27: Annual Conference on Neural Information Processing Systems 2014, December 8-13 2014, Montreal, Quebec, Canada*, pages 1781--1789, 2014.

Edo Liberty.
Simple and deterministic matrix sketching.
In *Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '13, pages 581--588, New York, NY, USA, 2013. ACM.