CNMAC 2019
XXXIX Congresso Nacional de Matemática Aplicada e Computacional
16 a 20 de Setembro de 2019 - Universidade Federal de Uberlândia - Uberlândia - MG

# A Study of the Optimality of PCA under Spectral Sparsification

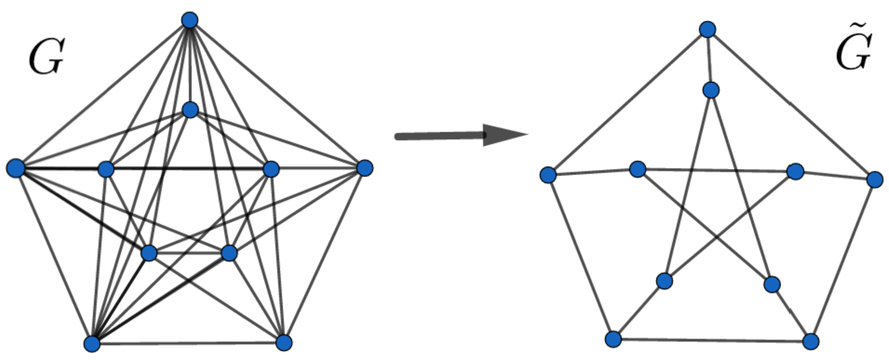
Sergio Mercado, Marcos Villagra

Núcleo de Investigación y Desarrollo Tecnológico, Facultad Politécnica, UNA - Paraguay.

## Introduction

Principal component analisys (PCA) is a data analysis technique for mapping points in $\mathbb{R}^n$ to a two or three dimensional space. This dimensionality reduction preserves the natural grouping of points and information of data. This is done optimally by an ortogonal projection of the points in $\mathbb{R}^n$ over the subspace generated by eigenvectors associated to the two or three greatests eigenvalues of the covariance matrix. It is well known that computing eigenvalues in general is computationally expensive, and therefore, several authors use techniques of numerical approximation. Furthermore, computations are more efficient whenever the matrices are sparse and memory costs can be reduced.
It can be proved that adding zeros in a symmetric matrix $M$ is equivalent to delete edges of a graph that represent $M$. This way, we can study this problem using graph theory.

## Spectral Sparsification




Let $G = (V, E, w)$ be an undirected weighted graph. We want to approximate $G$ by a sparse subgraph $\widetilde{G} = (V, \widetilde{E}, \widetilde{w})$, such that, given an $\epsilon \in (0, 1)$, and for all $x \in \mathbb{R}^n$

$$(1 - \epsilon)x^T L_G x \leq x^T L_{\widetilde{G}} x \leq (1 + \epsilon)x^T L_G x,$$

where $L_G$ and $L_H$ are the laplacian matrix of $G$ and $\widetilde{G}$, respectively.

**Theorem (Zouzias)** Suppose $a < \epsilon < 1$ and $A = \sum_{i=1}^{m} v_i v_i^T$ are given, with column vectors $v_i \in \mathbb{R}^n$. Then there are non-negative real weights $\{s_i\}$ at most $\lceil n/\epsilon^2 \rceil$ of which are non-zero, such taht

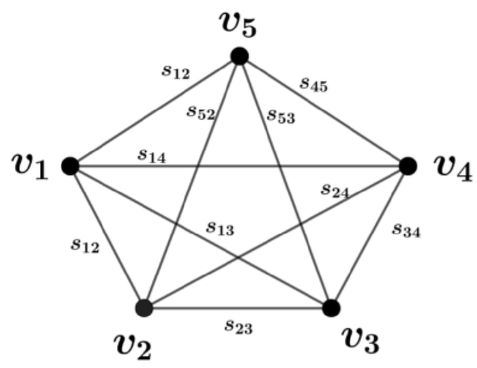
$$(1 - \epsilon)^3 A \preceq \widetilde{A} \preceq (1 + \epsilon)^3 A,$$

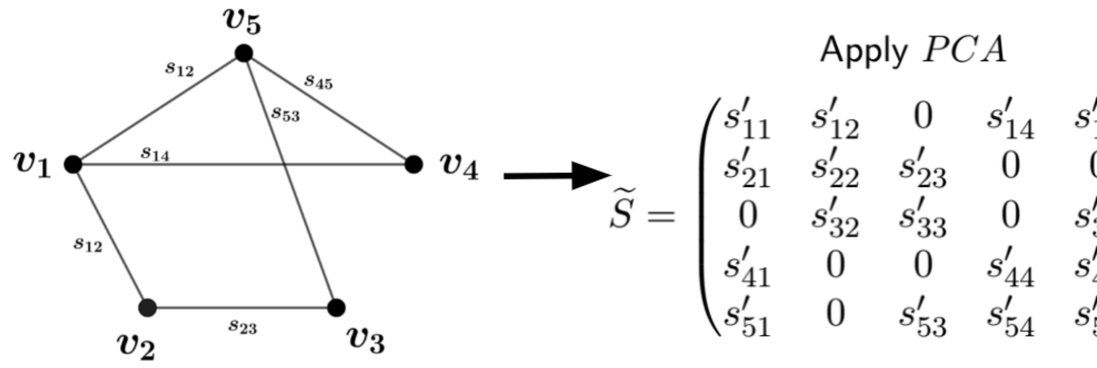
where, $\widetilde{A} = \sum_{i=1}^{m} s_i v_i v_i^T$.

## Proposal

Apply Zouzias's theorem to a covariance matrix $S$ according to the following scheme.

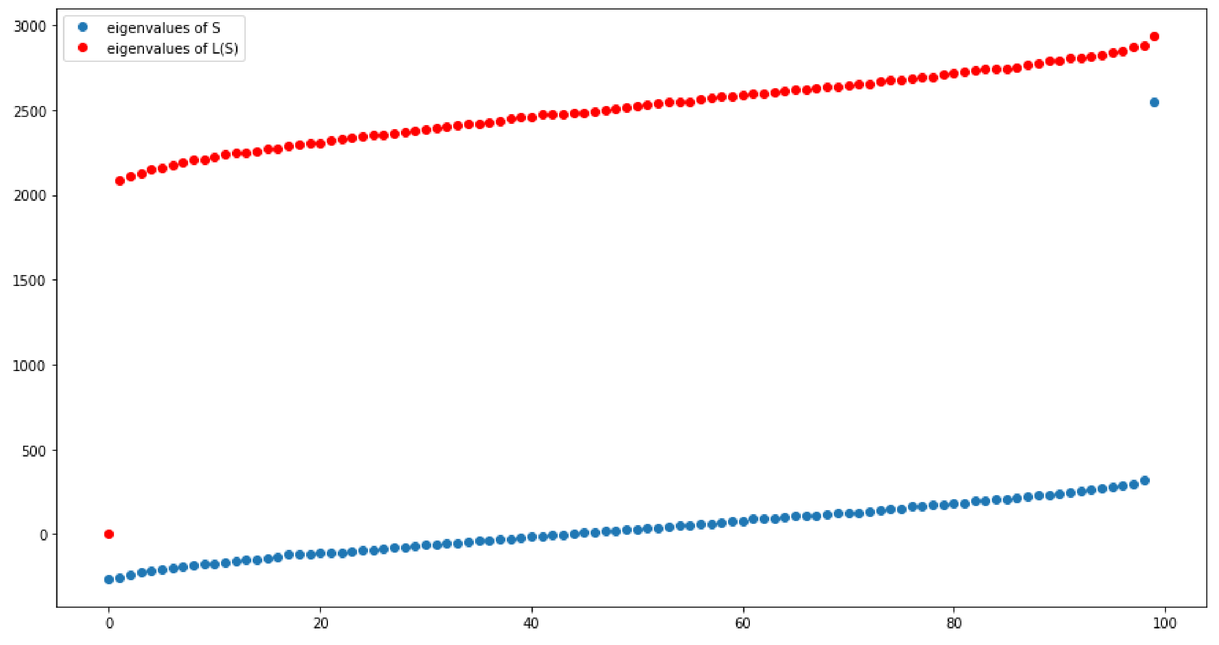
$$S = \begin{pmatrix} s_{11} & s_{12} & s_{13} & s_{14} & s_{15} \\ s_{21} & s_{22} & s_{23} & s_{24} & s_{25} \\ s_{31} & s_{32} & s_{33} & s_{34} & s_{35} \\ s_{41} & s_{42} & s_{43} & s_{44} & s_{45} \\ s_{51} & s_{52} & s_{53} & s_{54} & s_{55} \end{pmatrix}$$




## Experiments

We give an example of the behavior of the eigenvalues of a symetric matrix $S$ and its associated laplacian matrix $L(S)$, built according to the previous scheme.




## Conclusions

We observed that the eigenvalues greater of $S$ and $L(S)$ tend to have the same values. This suggests us that by this scheme, we could use the $L(S)$ matrix, to apply PCA using a principal component.

## Acknowledgment

S.M. is supported by FEEI- CONACyT-PROCIENCIA research grant PINV15-706 "COMIDENCO" and research grant POSG17-62. M.V. is supported by Conacyt research grant PINV15-208.

## References

[1] Batson, J., Spielman, D. A., & Srivastava, N. (2012). Twice-ramanujan sparsifiers. SIAM Journal on Computing, 41(6), 1704-1721.

[2] Jolliffe, I. (2011). Principal component analysis (pp. 1094-1096). Springer Berlin Heidelberg.

[3] Spielman, D. A., & Teng, S. H. (2011). Spectral sparsification of graphs. SIAM Journal on Computing, 40(4), 981-1025.

[4] Zouzias, A. (2012, July). A matrix hyperbolic cosine algorithm and applications. In International Colloquium on Automata, Languages, and Programming (pp. 846-858). Springer, Berlin, Heidelberg.