

UNIVERSIDAD NACIONAL DE ASUNCIÓN
FACULTAD POLITÉCNICA

MAESTRÍA EN CIENCIAS DE LA COMPUTACIÓN



**UN MÉTODO PARA CONSTRUIR MATRICES DISPERSAS
QUE SON ESPECTRALMENTE PRÓXIMAS A MATRICES
SIMÉTRICAS CON ELEMENTOS NO DIAGONALES NO
NEGATIVOS**

SERGIO ORLANDO MERCADO BENÍTEZ

Trabajo de Maestría presentado en conformidad a los
requisitos para obtener el grado de Máster en Ciencias de la Computación

San Lorenzo - Paraguay

Noviembre - 2020

**UN MÉTODO PARA CONSTRUIR MATRICES DISPERSAS
QUE SON ESPECTRALMENTE PRÓXIMAS A MATRICES
SIMÉTRICAS CON ELEMENTOS NO DIAGONALES NO
NEGATIVOS**

Sergio Orlando Mercado Benítez

Tesis de Maestría aprobada el 27 de noviembre de 2020 por los siguientes miembros del Jurado de Defensa:

Prof. Dr. Pablo Romero (Universidad de la República, Uruguay)

Prof. Dr. Inocencio Ortíz (FP-UNA)

Prof. Dr. José Colbes (FP-UNA)

Prof. Dr. Marcos Villagra (FP-UNA), orientador

Prof. Dr. Horacio A. Legal Ayala

Coordinador Académico

Postgrado en Ciencias de la Computación

Facultad Politécnica

Universidad Nacional de Asunción

Prof. Dr. Marcos Villagra

Orientador

UNIVERSIDAD NACIONAL DE ASUNCIÓN
FACULTAD POLITÉCNICA

MAESTRÍA EN CIENCIAS DE LA COMPUTACIÓN



UN MÉTODO PARA CONSTRUIR MATRICES DISPERSAS QUE SON
ESPECTRALMENTE PRÓXIMAS A MATRICES SIMÉTRICAS CON
ELEMENTOS NO DIAGONALES NO NEGATIVOS

Sergio Orlando Mercado Benítez

San Lorenzo - Paraguay
Noviembre - 2020

UNIVERSIDAD NACIONAL DE ASUNCIÓN
FACULTAD POLITÉCNICA

MAESTRÍA EN CIENCIAS DE LA COMPUTACIÓN



**UN MÉTODO PARA CONSTRUIR MATRICES DISPERSAS QUE SON
ESPECTRALMENTE PRÓXIMAS A MATRICES SIMÉTRICAS CON
ELEMENTOS NO DIAGONALES NO NEGATIVOS**

SERGIO ORLANDO MERCADO BENÍTEZ

Orientador:

Prof. Dr. Marcos Daniel Villagra

Trabajo de Maestría presentado en conformidad a los
requisitos para obtener el grado de Máster en Ciencias de la Computación

San Lorenzo - Paraguay

Noviembre - 2020

*A mis padres
y mis hermanos.*

AGRADECIMIENTOS

Al profesor Marcos Villagra por la confianza puesta en mí, por aceptar ser mi orientador en esta tesis, y por todo el apoyo que me ha brindado.

A los profesores Santiago Gómez y Christian Schaerer, por haberme dado la oportunidad de trabajar con ellos y así dar mis primeros pasos en la actividad científica.

Al Consejo Nacional de Ciencia y Tecnología (Conacyt), que en el marco de sus diversos programas de incentivo a la actividad académica, ha brindado su apoyo económico a través de becas y proyectos a este programa de maestría.

Por último, a mis compañeros de la maestría, por el apoyo brindado y por los interesantes y fructíferos momentos de intercambio de ideas que han propiciado.

¡MUCHAS GRACIAS!

UN MÉTODO PARA CONSTRUIR MATRICES DISPERSAS QUE SON ESPECTRALMENTE PRÓXIMAS A MATRICES SIMÉTRICAS CON ELEMENTOS NO DIAGONALES NO NEGATIVOS

Autor: SERGIO ORLANDO MERCADO BENÍTEZ

Orientador: Prof. Dr. MARCOS DANIEL VILLAGRA

RESUMEN

Decimos que una matriz real M es ODN (del inglés *off-diagonal nonnegative*) si y solo si todos sus elementos fuera de la diagonal son mayores o iguales a cero. En este trabajo demostramos que para toda matriz simétrica y ODN M , existe una matriz dispersa \widehat{M} que es próxima en espectro a M . La idea clave que utilizamos para la obtención de dichas matrices dispersas, es la noción de dispersión espectral de grafos introducida por Spielman y Teng (2011).

Palabras Clave: Dispersión espectral, matrices simétricas, matrices no negativas, teoría espectral de grafos.

A METHOD TO BUILD SPARSE MATRICES SPECTRALLY CLOSE TO OFF-DIAGONAL NONNEGATIVE SYMMETRIC MATRICES

Author: SERGIO ORLANDO MERCADO BENÍTEZ

Advisor: Prof. Dr. MARCOS DANIEL VILLAGRA

ABSTRACT

We say that a square real matrix M is *off-diagonal nonnegative* if and only if all entries outside its diagonal are nonnegative real numbers. In this thesis we show that for any off-diagonal nonnegative symmetric matrix M , there exists \widehat{M} which is sparse and close in spectrum to M . The main approach of this work is to borrow some ideas from spectral sparsification technique which was introduced by Spielman and Teng (2011).

Keywords: spectral sparsification, symmetric matrices, nonnegative matrices, spectral graph theory.

ÍNDICE

Página

| | |
|---|-----------|
| 1. INTRODUCCIÓN | 1 |
| 1.1. JUSTIFICACIÓN Y MOTIVACIÓN | 2 |
| 1.2. OBJETIVOS | 3 |
| 1.2.1. Objetivo General | 3 |
| 1.2.2. Objetivos Específicos | 3 |
| 1.3. ORGANIZACIÓN DE ESTE LIBRO | 3 |
| 2. PRELIMINARES | 5 |
| 2.1. Álgebra Lineal | 5 |
| 2.1.1. Vectores y Subespacios Generados | 5 |
| 2.1.2. Matrices | 6 |
| 2.1.3. Autovalores y Autovectores | 7 |
| 2.1.4. Normas Vectoriales y Matriciales | 8 |
| 2.1.5. Formas Cuadráticas | 8 |
| 2.1.6. Teoría de la Perturbación | 9 |
| 2.2. Teoría de Grafos | 11 |
| 2.2.1. Matrices que Representan Grafos | 11 |
| 2.2.2. Construcción Alternativa de la Matriz Laplaciana | 12 |
| 2.2.3. Algunas Propiedades de la Matriz Laplaciana | 14 |
| 3. DISPERSIÓN ESPECTRAL DE GRAFOS | 17 |
| 3.1. Intuición de la Técnica de Dispersión Espectral de Grafos | 17 |
| 3.2. Algoritmos para la Construcción de Dispersores Espectrales | 18 |
| 3.2.1. Algoritmo BSS | 19 |
| 3.2.2. Dispersión Espectral vía Resistencia Efectiva | 21 |
| 3.2.3. BSS Probabilístico | 24 |
| 4. CONSTRUCCIÓN DE MATRICES DISPERSAS | 26 |
| 4.1. Resultado Principal | 27 |
| 4.2. Lemas Técnicos | 28 |
| 4.3. Prueba del Resultado Principal (Teorema 1.1) | 29 |
| 4.4. Aplicaciones | 31 |
| 4.4.1. PCA | 31 |
| 4.4.2. Optimización | 31 |
| 5. CONCLUSIONES Y TRABAJOS FUTUROS | 33 |
| BIBLIOGRAFÍA | 35 |

| | |
|---|-----------|
| 6. APÉNDICE | 38 |
| 6.1. Publicaciones y Difusión | 38 |
| 6.2. Conjetura | 38 |

LISTA DE FIGURAS

Página

- | | |
|--|----|
| 2.1. Grafo simple, no dirigido y ponderado de 5 vértices | 12 |
| 2.2. Grafo simple, no dirigido y ponderado de 3 vértices | 15 |

LISTA DE SÍMBOLOS

| | | | |
|------------------------|---|-------------------------|--|
| $x(i)$ | i -ésima coordenada del vector x | $\ x\ _2$ | norma vectorial definida como $\left(\sum_{i=1}^n x(i) ^2\right)^{1/2}$ |
| $X \subseteq Y$ | X es un subconjunto de Y | | |
| $X \subset Y$ | X es un subconjunto propio de Y | $\ x\ _\infty$ | norma vectorial definida como $\ x\ _\infty = \max_{i \leq n} x(i) $ |
| $ X $ | cardinalidad del conjunto X | | |
| $gen(X)$ | espacio generado por el conjunto de vectores $X = \{x_1, x_2, \dots, x_m\}$ | $(M)_{ij}$ | elemento ubicado en la i -ésima fila y j -ésima columna de la matriz M |
| \cup | unión | I | matriz identidad |
| \cap | intersección | $\mathbf{0}$ | matriz nula |
| $-$ | diferencia | M^k | matriz elevada a la potencia k |
| \emptyset | conjunto vacío | M^T | transpuesta de la matriz M |
| \mathbb{R} | conjunto de los números reales | Δ_M | elemento ubicado en la diagonal de M de mayor valor numérico |
| \mathbb{R}^+ | conjunto de los números reales positivos | δ_M | elemento ubicado en la diagonal de M de menor valor numérico |
| \mathbb{C} | conjunto de los números complejos | $diag(r_1, \dots, r_n)$ | matriz diagonal de $n \times n$ cuyo elemento ubicado en j -ésima fila y j -ésima columna es igual a r_j |
| $\min_{x \in S}$ | elemento mínimo del conjunto S | | |
| $\max_{x \in S}$ | elemento máximo del conjunto S | | |
| $\mathbf{0} \preceq M$ | M es positiva semidefinida | $\ M\ $ | norma matricial inducida por la norma vectorial $\ \cdot\ _2$ |
| $A \preceq B$ | $B - A$ es positiva semidefinida, donde A, B son matrices de $n \times n$ | $\mathcal{C}(M)$ | espacio columna de la matriz M |
| $\rho(M)$ | radio espectral de la matriz simétrica M | $\text{Tr}(M)$ | traza de la matriz M |
| \sin | función trigonométrica seno | $\mathcal{N}(M)$ | espacio nulo de la matriz M |
| \wedge | y (lógico) | $\lambda_{\min}(M)$ | autovalor de menor valor numérico de la matriz M |
| \vee | o (lógico) | $\lambda_{\max}(M)$ | autovalor de mayor valor numérico de la matriz M |
| \Rightarrow | implicancia | $G =$ | grafo simple, no dirigido y ponderado |
| \Leftrightarrow | si y solo si | (V, E, w) | |
| $ $ | tal que | $\{u, v\}$ | arista que conecta los vértices u y v |
| $\ x\ _1$ | norma vectorial definida como $\sum_{i \leq n} x(i) $ | | |

| | | | |
|-----------------------------|--|-----------------------------|--|
| $\overrightarrow{\{u, v\}}$ | arista orientada de u a v | M^+ | pseudoinversa de Moore-Penrose de la matriz M |
| $w_{u,v}$ | peso de la arista $\{u, v\}$ | | |
| $u \rightsquigarrow v$ | u y v son vértices adyacentes | $Id(\mathcal{V})$ | operador identidad del espacio vectorial \mathcal{V} |
| $d(u)$ | grado del vértice u | | |
| χ_u | vector característico del vértice u | v | vértice |
| $\chi_{u,v}$ | $\chi_u - \chi_v$ | e | arista |
| Ψ | vector de \mathbb{R}^n cuyas coordenadas son todas iguales a 1 | $\sum_{i \leq n} a_i$ | sumatoria |
| A_G | matriz de adyacencias del grafo G | $\mathcal{O}(f(n))$ | notación asintótica “o grande” |
| D_G | matriz de grados del grafo G | $\tilde{\mathcal{O}}(f(n))$ | $\mathcal{O}(f(n) \log(f(n)))$ |
| L_G | matriz laplaciana del grafo G | $\Omega(f(n))$ | notación asintótica “omega grande” |
| V | conjunto de vértices | $\mathbf{E}(X)$ | esperanza de la variable aleatoria X |
| E | conjunto de aristas | | |
| M^{-1} | inversa de la matriz M | $Var(X)$ | varianza de la variable aleatoria X |

CAPÍTULO 1

INTRODUCCIÓN

Es bien conocido que en muchos problemas matemáticos y de ciencias de la computación en las que se realizan cálculos con matrices, se pueden obtener ventajas cuando las matrices son dispersas (matrices con muchos elementos nulos). Estas ventajas consisten principalmente en una reducción en el tiempo de cómputo de los algoritmos que resuelven dichos problemas, así como también en el uso de recursos de almacenamiento. Por ejemplo, el algoritmo de Lanczos [1, 2] para el cálculo de autovalores y autovectores de matrices simétricas, tiene un tiempo de cómputo de $\mathcal{O}(\mu n^2)$, donde n es el orden y μ el promedio de elementos no nulos de las filas de la matriz, respectivamente. Por tanto, esta complejidad puede ir desde $\mathcal{O}(n^2)$ (cuando μ toma un valor muy pequeño), hasta $\mathcal{O}(n^3)$, en el peor caso. En esta tesis demostraremos que para ciertos tipos de matrices, el valor de μ puede hacerse pequeño.

El tipo de matrices que consideramos en este trabajo, tiene la característica de que sus elementos fuera de la diagonal son números no negativos, y nos referiremos a estas matrices como **matrices ODN** (*del inglés off-diagonal nonnegative*). En particular, centraremos nuestro estudio en las matrices que son **ODN y simétricas**.

Básicamente, el problema que afrontamos es el siguiente: Dada una matriz $M \in \mathbb{R}^{n \times n}$, simétrica y ODN, encontrar una matriz dispersa \widehat{M} que sea próxima en espectro a M . Con la palabra *espectro* nos referiremos al conjunto de autovalores y autovectores de la matriz.

La principal idea que empleamos para obtener tales matrices dispersas, es una técnica denominada *dispersión espectral de grafos*, que fue introducida por Spielman y Teng en [3]. Dicha técnica consiste básicamente en construir un grafo disperso \widehat{G} (grafo con pocas aristas), a partir de un grafo conexo, no dirigido y ponderado G , tal que las respectivas matrices laplacianas $L_{\widehat{G}}$ y L_G , sean próximas en espectro. Así, con esta técnica en mente, desde un punto de vista general podemos decir que el método que proponemos consiste en asociar una matriz simétrica ODN con la matriz de adyacencia y la matriz de grados de un grafo, y luego llevar a cabo el

proceso de dispersión espectral utilizando el algoritmo de Lee y Sun [4].

Con el fin de medir las aproximaciones entre los espectros de las matrices M y \widehat{M} , utilizamos dos teoremas bien conocidos de la teoría de perturbación, que son: El Teorema de Perturbación de Weyl, para medir la proximidad de los respectivos autovalores, y el Teorema de Davis-Kahan, para la proximidad entre los autovectores.

Utilizando los símbolos Δ_M y δ_M para denotar al elemento de la diagonal de $M \in \mathbb{R}^{n \times n}$ de mayor y menor valor numérico, respectivamente; y $\rho(M)$ para denotar el radio espectral de M , nuestro resultado principal puede ser enunciado de la siguiente manera.

Teorema 1.1. Sea $M \in \mathbb{R}^{n \times n}$ una matriz simétrica y ODN con autovalores $\lambda_1 \geq \dots \geq \lambda_n$ y respectivos autovectores x_1, x_2, \dots, x_n . Definamos $\lambda_0 = -\infty$ y $\lambda_{n+1} = \infty$. Entonces, para todo $\epsilon > 0$ tal que $0 < \epsilon \leq 1/120$, existe una matriz simétrica y ODN $\widehat{M} \in \mathbb{R}^{n \times n}$ con $\mathcal{O}(\frac{n}{\epsilon^2})$ elementos distintos de cero, tal que para todo $i = 1, \dots, n$ se cumple que

$$|\lambda_i - \widehat{\lambda}_i| \leq \epsilon \sqrt{n} \rho(L_M) + \frac{\Delta_M - \delta_M}{2},$$

donde $\widehat{\lambda}_1 \geq \dots \geq \widehat{\lambda}_n$ son los autovalores de \widehat{M} . Además, si $\widehat{x}_1, \widehat{x}_2, \dots, \widehat{x}_n$ son los autovectores respectivos de $\widehat{\lambda}_1 \geq \dots \geq \widehat{\lambda}_n$ y θ_i es el ángulo (agudo) entre los autoespacios generados por autovectores x_i y \widehat{x}_i , entonces

$$\sin \theta_i \leq \frac{\epsilon \sqrt{n} \rho(L_M) + (\Delta_M - \delta_M)/2}{\min\{|\widehat{\lambda}_{i-1} - \lambda_i|, |\lambda_i - \widehat{\lambda}_{i+1}|\}}.$$

Zouzias en [5] desarrolló un algoritmo probabilístico que como entrada recibe una matriz simétrica A tal que $\|A\|_\infty \leq \sqrt{\theta} \|A\|$, y en un tiempo lineal, retorna una matriz \widehat{A} con $\mathcal{O}(\theta n \log n / \epsilon^2)$ elementos no nulos, tal que $\|A - \widehat{A}\| \leq \epsilon \|A\|$, donde $\epsilon \in (0, 1)$. Si bien el método de Zouzias funciona para un conjunto más general de matrices, se puede observar que presenta una desventaja comparado con nuestro método, cuando la matriz considerada es simétrica y ODN. Dicha desventaja consiste en que la cantidad de elementos no nulos de la matriz resultante, depende de los parámetros extras: $\log n$ y θ .

1.1. JUSTIFICACIÓN Y MOTIVACIÓN

El prominente paradigma para la construcción de grafos dispersos denominado **dispersión espectral de grafos**, ha dado lugar a numerosos trabajos de investigación en diversas áreas, tanto de las matemáticas como de ciencias de la computación, a lo largo de la última década. Tales trabajos se centraron más bien en el desarrollo de algoritmos más eficientes para construir grafos dispersos que preserven ciertas propiedades, y en la utilización de dichos grafos para resolución de problemas.

Luego de realizar una revisión del estado del arte de dicha técnica, hemos notado que también podría ser utilizada para la construcción de matrices dispersas que aproximen espectralmente

a ciertos tipos de matrices. Con base en el estudio bibliográfico realizado, hasta antes de este trabajo, no se ha desarrollado ningún método que utilice la técnica de dispersión espectral de grafos para construir matrices dispersas, que involucren matrices diferentes a la matriz laplaciana.

Proporcionar un método que construya matrices dispersas espectralmente próximas a matrices densas, adquiere relevancia en aquellos problemas en los que se consideran matrices de alto orden, y que a la vez solamente importa el espectro de la matriz. Así, en lugar de utilizar la matriz densa, podemos utilizar su *versión dispersada* de tal manera a realizar cálculos con menor costo computacional.

1.2. OBJETIVOS

1.2.1. Objetivo General

Proporcionar un método para construir matrices dispersas que sean espectralmente próximas a matrices que cumplan la condición de ser: real, simétrica y cuyos elementos fuera de la diagonal sean números no negativos.

1.2.2. Objetivos Específicos

- Utilizar la técnica de dispersión espectral de grafos de tal manera a construir matrices dispersas.
- Determinar la proximidad espectral entre la matriz original y la matriz dispersa generada, utilizando algunos teoremas de la teoría de la perturbación.
- Proveer una demostración constructiva que valide el método citado en el objetivo general.

1.3. ORGANIZACIÓN DE ESTE LIBRO

Este libro está organizado de la siguiente manera:

Capítulo 2

Se presentan las notaciones, los conceptos y los resultados del álgebra lineal y la teoría de grafos que utilizaremos en esta tesis.

Capítulo 3

Aquí se presenta la técnica de dispersión espectral de grafos y un desarrollo sobre los tres principales algoritmos empleados en la construcción de grafos dispersos. Estos algoritmos serán útiles para la comprensión del resultado principal de esta tesis.

Capítulo 4

En este capítulo se estudia el problema central de esta tesis. Contiene los lemas y la demostración del teorema 1.1 que resuelve el problema planteado en el objetivo general. Además se dan ejemplos de aplicaciones a problemas como PCA y Optimización.

Capítulo 5

En este capítulo se presenta la conclusión de este trabajo y algunos problemas abiertos relevantes.

CAPÍTULO 2

PRELIMINARES

En este capítulo introduciremos las notaciones, algunos conceptos y los resultados del álgebra lineal y la teoría de grafos, que utilizaremos en esta tesis. Las demostraciones serán omitidas pero junto a cada resultado, iremos indicando la referencia correspondiente.

2.1. Álgebra Lineal

Empezaremos con lo que respecta a las notaciones y resultados del álgebra lineal. En todo este trabajo, utilizaremos \mathbb{R} y \mathbb{C} para denotar al conjunto de los números reales y al conjunto de los números complejos, respectivamente.

2.1.1. Vectores y Subespacios Generados

Los vectores reales x de n componentes (o elementos del espacio vectorial \mathbb{R}^n), serán denotados por

$$x = \begin{pmatrix} x(1) \\ x(2) \\ \vdots \\ x(n) \end{pmatrix}.$$

Así, $x(i)$ denota al i -ésimo componente del vector x . La traspuesta de $x \in \mathbb{R}^n$ es

$$x^T = (x(1), x(2), \dots, x(n))^T.$$

Utilizaremos preferentemente las letras minúsculas a, b, x, y, z para denotar vectores.

Si $\{x_1, x_2, \dots, x_m\} \subset \mathbb{R}^n$ es un conjunto de vectores linealmente independiente, entonces al espacio generado por $\{x_1, x_2, \dots, x_m\}$ lo denotaremos por $gen(\{x_1, x_2, \dots, x_m\})$.

2.1.2. Matrices

Utilizaremos preferentemente las letras mayúsculas A, B, M y N para denotar a las matrices reales (o elementos del espacio vectorial $\mathbb{R}^{m \times n}$). Si $M \in \mathbb{R}^{n \times n}$, diremos que M es cuadrada de orden $n \times n$; o simplemente que es de orden n . Al elemento ubicado en la i -ésima fila y en la j -ésima columna de la matriz M , lo denotaremos por $(M)_{ij}$. Utilizaremos I para denotar la matriz identidad y $\mathbf{0}$ para la matriz nula. Además, $\text{diag}(\alpha_1, \alpha_2, \dots, \alpha_n)$ denotará una matriz diagonal de $\mathbb{R}^{n \times n}$ cuyo elemento ubicado en la j -ésima fila y j -ésima columna es igual a α_j .

Para una matriz $M \in \mathbb{R}^{m \times n}$, utilizaremos $\mathcal{C}(M) = \{Mx : x \in \mathbb{R}^n\}$ y $\mathcal{N}(M) = \{x \in \mathbb{R}^n : Mx = \mathbf{0}\}$ para denotar a su espacio columna y su espacio nulo, respectivamente. La traspuesta de una matriz M lo denotaremos por M^T . Si $M = M^T$, decimos que la matriz es simétrica. La traza de una matriz $M \in \mathbb{R}^{n \times n}$ se define como

$$\text{Tr}(M) = \sum_{i \leq n} (M)_{ii},$$

es decir, es la suma de todos los elementos de la diagonal de M .

En ocasiones estaremos interesados en la cantidad de columnas (vectores) linealmente independientes de una matriz, por ello introducimos la siguiente definición.

Definición 1. Sea $M \in \mathbb{R}^{m \times n}$. Al número de columnas linealmente independientes de M , lo denominaremos el **rango** de M .

Es bien conocido que una matriz $M \in \mathbb{R}^{n \times n}$ es invertible si y solo si, su rango es igual a n . A continuación presentamos un tipo especial de matrices de rango 1 y que utilizaremos bastante en este trabajo.

Definición 2. Sean x, z vectores de \mathbb{R}^n . El producto exterior entre x y z , es la matriz $xz^T \in \mathbb{R}^{n \times n}$, cuyos elementos son

$$(xz^T)_{ij} = x(i)z(j).$$

Proposición 1. Sean $A, B \in \mathbb{R}^{n \times n}$ matrices tales que a_1, a_2, \dots, a_n son las columnas de A y $b_1^T, b_2^T, \dots, b_n^T$ las filas de B . Entonces

$$AB = \sum_{i \leq n} a_i b_i^T.$$

Es decir, el producto AB puede ser expresado como suma de matrices de rango 1, en la que cada sumando es el producto exterior entre una columna de A por su respectiva fila de B ¹.

Proposición 2. Si $\{x_1, x_2, \dots, x_r\}$ es un conjunto ortonormal de vectores tal que

$$\text{gen}(\{x_1, x_2, \dots, x_r\}) = \mathcal{V} \subset \mathbb{R}^n,$$

¹Es decir, el producto exterior entre la i -ésima columna de A por la i -ésima fila de B .

entonces la matriz que proyecta de forma ortogonal sobre el espacio vectorial \mathcal{V} , es igual a

$$\sum_{i \leq r} x_i x_i^T.$$

2.1.3. Autovalores y Autovectores

Sea $M \in \mathbb{R}^{n \times n}$. El número $\lambda \in \mathbb{C}$ se denomina autovalor de M si existe un vector $x \in \mathbb{C}^n$ diferente del vector cero tal que

$$Mx = \lambda x.$$

En este caso, el vector $x \neq 0$, se denomina autovector de M asociado al autovalor λ . Es bien conocido que si $M \in \mathbb{R}^{n \times n}$ es simétrica, entonces sus autovalores son reales. En todo este material, los autovalores de las matrices simétricas serán ordenados de mayor a menor, es decir $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n$. Así, si $Mx_i = \lambda_i x_i$, diremos que λ_i es el i -ésimo autovalor y x_i es el i -ésimo autovector de M . En ocasiones, para indicar el autovalor de menor valor numérico de una matriz M , utilizaremos $\lambda_{\min}(M)$, y para el de mayor valor numérico, $\lambda_{\max}(M)$. Si \mathcal{E} es el espacio generado por los autovectores asociados al autovalor λ , diremos que \mathcal{E} es el *autoespacio asociado* a λ . Es bien conocido que para toda matriz $M \in \mathbb{R}^{n \times n}$, se cumple que

$$\text{Tr}(M) = \sum_{i \leq n} \lambda_i$$

Definición 3. Sea $M \in \mathbb{R}^{n \times n}$. El **radio espectral** de M es el número $\rho(A) = \max\{|\lambda| : \lambda \text{ es un autovalor de } A\}$.

Teorema 2.1 (Teorema Espectral). Sea $M \in \mathbb{R}^{n \times n}$ una matriz simétrica. Entonces, existen números reales $\lambda_1, \lambda_2, \dots, \lambda_n$ y vectores $x_1, x_2, \dots, x_n \in \mathbb{R}^n$, tales que para todo $i = 1, \dots, n$,

$$Mx_i = \lambda_i x_i,$$

y

$$M = \sum_{i \leq n} \lambda_i x_i x_i^T.$$

El siguiente teorema consiste en un resultado similar al teorema espectral, pero se aplica a matrices de $\mathbb{R}^{m \times n}$. Es conocido como *descomposición en valores singulares* o **SVD**, por sus siglas en inglés. Para más detalles de este teorema, ver [2].

Teorema 2.2 (SVD [2]). Sea $M \in \mathbb{R}^{m \times n}$ una matriz tal que $m \geq n$. Entonces existen matrices $U \in \mathbb{R}^{m \times m}$, $V \in \mathbb{R}^{n \times n}$ y $\Sigma = \text{diag}(\sigma_1, \dots, \sigma_n)$, tales que $M = U\Sigma V^T$, $UU^T = I$, $VV^T = I$ y $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_n \geq 0$. Las columnas u_1, u_2, \dots, u_n de V son denominados vectores singulares derechos de M , y los valores σ_i son los valores singulares de M . (Si $m < n$, la descomposición en valores singulares se realiza considerando la matriz M^T).

2.1.4. Normas Vectoriales y Matriciales

Sea x un vector de \mathbb{R}^n . Utilizaremos las siguientes normas vectoriales:

- $\|x\|_1 = \sum_{i \leq n} |x(i)|$,
- $\|x\|_2 = \left(\sum_{i \leq n} |x(i)|^2 \right)^{1/2}$,
- $\|x\|_\infty = \max_{i \leq n} |x(i)|$.

Estas normas también son conocidas como *norma 1*, *norma Euclídea* (o *norma-2*) y *norma infinita*, respectivamente.

Por otro lado, sea la matriz $A \in \mathbb{R}^{m \times n}$. Las normas matriciales que utilizaremos son:

- $\|A\|_1 = \max_{\|x\|_1=1} \|Ax\|_1 = \max_{j \leq n} \sum_{i \leq m} |(A)_{ij}|$,
- $\|A\|_2 = \max_{\|x\|_2=1} \|Ax\|_2 = \sqrt{\rho(A^T A)} = \sqrt{\rho(AA^T)}$,
- $\|A\|_\infty = \max_{\|x\|_\infty=1} \|Ax\|_\infty = \max_{i \leq m} \sum_{j \leq n} |(A)_{ij}|$.

De aquí en adelante, la norma matricial $\|\cdot\|_2$ será denotada simplemente por $\|\cdot\|$. Es bien conocido que si A es simétrica, entonces $\|A\|_1 = \|A\|_\infty$ y $\|A\| = \rho(A)$. Algunas relaciones entre estas normas matriciales, se establecen en el siguiente resultado.

Proposición 3. *Sea $M \in \mathbb{R}^{m \times n}$, entonces*

1. $\|A\|^2 \leq \|A\|_1 \|A\|_\infty$,
2. $\frac{1}{\sqrt{n}} \|A\|_\infty \leq \|A\| \leq \sqrt{m} \|A\|_\infty$.

Estas propiedades pueden ser consultadas en [2, 6].

2.1.5. Formas Cuadráticas

Sea $M \in \mathbb{R}^{n \times n}$ una matriz simétrica y $x \in \mathbb{R}^n$. La forma cuadrática de M en x está dada por la expresión

$$x^T M x.$$

Decimos que una matriz simétrica M es *positiva semidefinida* (psd) si para todo $x \in \mathbb{R}^n$, se cumple

$$x^T M x \geq 0.$$

Para indicar que M es positiva semidefinida, escribiremos $\mathbf{0} \preceq M$, o $M \succeq \mathbf{0}$. De forma similar, para dos matrices simétricas M, N escribiremos

$$M \preceq N \text{ si y solo si, } \mathbf{0} \preceq M - N.$$

De la definición se deduce inmediatamente que si $c \geq 0$, entonces

$$M \preceq N \text{ si y solo si, } cM \preceq cN.$$

Además

$$M \preceq N \text{ y } N \preceq P, \text{ implica } M \preceq P.$$

A continuación damos algunas caracterizaciones de las matrices psd.

Teorema 2.3. [7] Una matriz simétrica $M \in \mathbb{R}^{n \times n}$ es positiva semidefinida si y solo si:

1. todos los autovalores de M son mayores o iguales a cero, o
2. existe una matriz B tal que $M = B^T B$.

Sea $M \in \mathbb{R}^{n \times n}$ una matriz simétrica y $x \in \mathbb{R}^n$. El **cociente de Rayleigh** de M en x , es definido como

$$\frac{x^T M x}{x^T x}.$$

Así, si λ es el autovalor de M asociado al autovector x , entonces

$$\frac{x^T M x}{x^T x} = \frac{x^T \lambda x}{x^T x} = \frac{\lambda \|x\|^2}{\|x\|^2} = \lambda.$$

El siguiente teorema utiliza el cociente de Rayleigh para dar una caracterización de los autovalores de una matriz simétrica.

Teorema 2.4 (Courant-Fischer [8, 9]). Sea $M \in \mathbb{R}^{n \times n}$ una matriz simétrica con autovalores $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n$. Entonces

$$\lambda_k = \max_{\substack{S \subset \mathbb{R}^n \\ \dim(S)=k}} \min_{x \in S, x \neq 0} \frac{x^T M x}{x^T x} = \min_{\substack{T \subset \mathbb{R}^n \\ \dim(T)=n-k+1}} \max_{x \in T, x \neq 0} \frac{x^T M x}{x^T x},$$

donde la maximización y la minimización se realizan sobre los subespacios S y T de \mathbb{R}^n .

2.1.6. Teoría de la Perturbación

El problema central de la teoría de la perturbación es determinar el cambio producido en los autoespacios de una matriz simétrica A , luego de aplicarle una perturbación definida como $\hat{A} = A + B$, donde B se considera “pequeña”. Para medir este *cambio* o *distancia* entre los respectivos autoespacios, se utiliza la noción de *ángulos principales*. En efecto, sean V, \hat{V} matrices de $\mathbb{R}^{n \times r}$, con columnas ortonormales. Entonces, los ángulos $\cos^{-1} \sigma_1, \dots, \cos^{-1} \sigma_r$ son los r ángulos principales entre los espacios columnas de V y \hat{V} , donde los valores $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_r$ son los valores singulares de la matriz $\hat{V}^T V$.

Nótese que esta definición es una generalización del método usual para hallar el ángulo agudo entre dos vectores.

Así, la distancia entre los espacios $\mathcal{C}(V)$ y $\mathcal{C}(\widehat{V})$, se define como

$$\|\sin \Theta(V, \widehat{V})\|,$$

donde $\sin \Theta(V, \widehat{V})$ es una matriz diagonal definida como

$$\sin \Theta(V, \widehat{V}) = \begin{pmatrix} \sin(\cos^{-1} \sigma_1) & & \\ & \ddots & \\ & & \sin(\cos^{-1} \sigma_r) \end{pmatrix}$$

Uno de los resultados más importantes de la Teoría de la Perturbación, es el *teorema del sin θ* de Davis-Kahan [10]. Dicho teorema proporciona una cota superior para la distancia entre los subespacios generados por los correspondientes conjuntos de autovectores de las matrices V y \widehat{V} . Una versión más simple del mismo fue dada por Yu, Wang y Samworth en [11] y es la que utilizaremos aquí, debido a que se ajusta mejor a los propósitos de este trabajo.

Teorema 2.5 (Davis-Kahan, 1970 [11]). Sean $A, \widehat{A} \in \mathbb{R}^{n \times n}$ matrices simétricas con autovalores $\lambda_1 \geq \dots \geq \lambda_n$ y $\widehat{\lambda}_1 \geq \dots \geq \widehat{\lambda}_n$, respectivamente. Sean r, s enteros tales que $1 \leq r \leq s \leq n$ y $d = s - r + 1$. Sean $A_0 = (a_r, a_{r+1}, \dots, a_s)$, $\widehat{A}_0 = (\widehat{a}_r, \widehat{a}_{r+1}, \dots, \widehat{a}_s) \in \mathbb{R}^{n \times d}$ matrices tales que para todo $i = r, \dots, s$, las columnas a_i y \widehat{a}_i , son los autovectores normalizados asociados a los autovalores λ_i y $\widehat{\lambda}_i$ de las matrices A y \widehat{A} , respectivamente. Definimos $\delta = \inf\{|\lambda - \widehat{\lambda}| : \lambda \in [\lambda_s, \lambda_r], \widehat{\lambda} \in (-\infty, \widehat{\lambda}_{s+1}] \cup [\widehat{\lambda}_{r-1}, \infty)\}$, donde $\widehat{\lambda}_0 = -\infty$ y $\widehat{\lambda}_{n+1} = \infty$. Asumimos que $\delta > 0$. Entonces se cumple que

$$\|\sin \Theta(A_0, \widehat{A}_0)\| \leq \frac{\|A - \widehat{A}\|}{\delta}, \quad (2.1)$$

Si $s = r = j$, entonces la desigualdad (2.1) se reduce a una cota superior para el ángulo entre los subespacios generados por los autovectores a_j y \widehat{a}_j . Es decir

$$\sin \Theta(a_j, \widehat{a}_j) \leq \frac{\|A - \widehat{A}\|}{\min\{|\widehat{\lambda}_{j-1} - \lambda_j|, |\widehat{\lambda}_{j+1} - \lambda_j|\}}. \quad (2.2)$$

Si θ es el ángulo entre estos últimos subespacios, entonces $0 \leq \theta \leq \pi/2$, pues siempre podemos elegir dos vectores de los mencionados subespacios, de tal manera que el $a_j^T \widehat{a}_j \geq 0$.

El siguiente teorema proporciona una cota para estimar la proximidad entre los respectivos autovalores de dos matrices simétricas A y B , a partir de la norma $\|A - B\|$.

Teorema 2.6 (Teorema de Perturbación de Weyl [8]). Sean $A, B \in \mathbb{R}^{n \times n}$ matrices simétricas tal que $\alpha_1, \alpha_2, \dots, \alpha_n$, y $\beta_1, \beta_2, \dots, \beta_n$ sean los autovalores de A y B , respectivamente. Entonces, para todo $i = 1, \dots, n$,

$$|\alpha_i - \beta_i| \leq \|A - B\|.$$

Para más detalles sobre estos los teoremas 2.5 y 2.6, ver los libros de Bhatia [8] y Stewart-Sun [9].

2.2. Teoría de Grafos

En esta sección introduciremos las notaciones y los conceptos básicos de la teoría de grafos que necesitaremos para este trabajo. Para una consulta detallada de los mismos, se recomienda el libro de Diestel [12].

Definición 2.1. Un **grafo simple no dirigido** G , es un par ordenado (V, E) donde V es un conjunto distinto de vacío y E un subconjunto de $V \times V$ que satisface $E \subseteq \{\{u, v\} \mid u, v \in V \wedge u \neq v\}$. Los elementos de V son denominados vértices, y los elementos $\{u, v\}$ de E se denominan aristas.

Si $\{u, v\}$ es una arista del grafo, decimos que los vértices u y v son *adyacentes*, o que $\{u, v\}$ es la arista que *conecta* los vértices u y v .

En todo este trabajo, asumiremos que los conjuntos de vértices siempre son conjuntos finitos y que sus elementos son etiquetados con los números del 1 al $|V|$.²

Definición 2.2. Un grafo **ponderado** es una terna ordenada $G = (V, E, w)$, en la que (V, E) es un grafo simple y no dirigido, y w una función de costo definida como

$$w : V \times V \mapsto \mathbb{R}^+ \cup \{0\},$$
$$\{u, v\} \mapsto w_{u,v},$$

tal que $w_{u,v} > 0$ si $\{u, v\} \in E$ y $w_{u,v} = 0$ si $\{u, v\} \notin E$. El número $w_{u,v}$ es el *peso* de la arista $\{u, v\}$.

Así, en un grafo ponderado, dos vértices u y v son adyacentes si y solo si $w_{u,v} > 0$. En ocasiones utilizaremos la notación $u \rightsquigarrow v$ para indicar que dichos vértices son adyacentes.

En lo que resta de este trabajo, consideraremos solamente grafos ponderados, y en algunas ocasiones nos referiremos a ellos simplemente con la palabra **grafo**.

Definición 2.3. Sean $G = (V, E, w)$ un grafo. El grado del vértice $i \in V$ denotado por $d(i)$, es

$$d(i) = \sum_{j \rightsquigarrow i} w_{i,j}.$$

2.2.1. Matrices que Representan Grafos

A continuación introducimos las matrices que normalmente son utilizadas para la representación de un grafo.

Sea $G = (V, E, w)$ un grafo ponderado. La matriz de adyacencia de G denotada por A_G , se define como

$$(A_G)_{ij} = \begin{cases} w_{i,j} & \text{si } i \rightsquigarrow j \\ 0 & \text{en otro caso.} \end{cases}$$

² $|V|$ denota el cardinal del conjunto V .

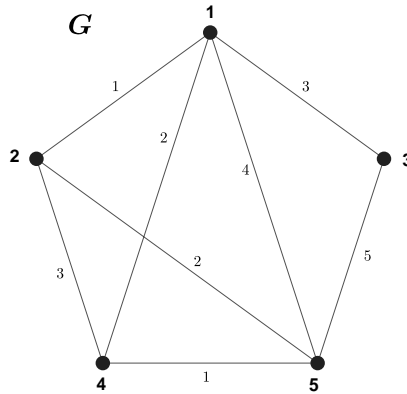


Figura 2.1: Grafo simple, no dirigido y ponderado de 5 vértices

Por otro lado, la matriz de grados de G denotada por D_G , se define como

$$(D_G)_{ij} = \begin{cases} d(i) & \text{si } i = j \\ 0 & \text{si } i \neq j. \end{cases}$$

Por último, la matriz laplaciana de G se define como $L_G = D_G - A_G$.

A partir de las definiciones anteriores, tenemos que A_G y L_G son simétricas y D_G es diagonal (también simétrica). La matriz laplaciana tal vez sea la matriz más importante de representación de un grafo, pues muchas propiedades de los grafos, pueden ser explicadas a través del espectro de esta matriz. Cabe mencionar que esta conexión entre grafos y matrices, es el objeto de estudio de la denominada *Teoría Espectral de Grafos*. Para más detalles sobre esta área, se puede consultar [13].

A continuación ilustramos un ejemplo de los conceptos recién mencionados. Consideremos el grafo de la figura 2.1.

Entonces, tenemos que

$$A_G = \begin{pmatrix} 0 & 1 & 3 & 2 & 4 \\ 1 & 0 & 0 & 3 & 2 \\ 3 & 0 & 0 & 0 & 5 \\ 2 & 3 & 0 & 0 & 1 \\ 4 & 2 & 5 & 1 & 0 \end{pmatrix}, D_G = \begin{pmatrix} 10 & 0 & 0 & 0 & 0 \\ 0 & 6 & 0 & 0 & 0 \\ 0 & 0 & 8 & 0 & 0 \\ 0 & 0 & 0 & 6 & 0 \\ 0 & 0 & 0 & 0 & 12 \end{pmatrix} \text{ y } L_G = \begin{pmatrix} 10 & -1 & -3 & -2 & -4 \\ -1 & 6 & 0 & -3 & -2 \\ -3 & 0 & 8 & 0 & -5 \\ -2 & -3 & 0 & 6 & -1 \\ -4 & -2 & -5 & -1 & 12 \end{pmatrix}.$$

2.2.2. Construcción Alternativa de la Matriz Laplaciana

Hemos visto que la matriz laplaciana se define como $L_G = D_G - A_G$, sin embargo, daremos una manera alternativa para construirla. La importancia de esta construcción radica en el hecho de que cumple un papel fundamental en los algoritmos de dispersión espectral que veremos en

el siguiente capítulo. Antes de realizar la nueva construcción, necesitamos introducir algunas definiciones.

Definición 2.4. Sea $G = (V, E, w)$ un grafo de n vértices y $u \in V$. Definimos el vector $\chi_u \in \mathbb{R}^n$ como $\chi_u(t) = 1$ si $t = u$, y $\chi_u(t) = 0$ en otro caso. El vector χ_u es el **vector característico del vértice u** .

Definición 2.5. Sea G un grafo de n vértices. Orientamos todas las aristas de G de forma arbitraria y obtenemos las aristas orientadas $\overrightarrow{\{u, v\}}$, de manera que u sea el vértice inicial y v el vértice final de dicha arista. Entonces el vector $\chi_{u,v} \in \mathbb{R}^n$ definido por $\chi_{u,v} = \chi_u - \chi_v$, es el **vector característico de la arista $\overrightarrow{\{u, v\}}$** . Por otra parte, la matriz $\chi_{u,v}\chi_{u,v}^T$ es la **matriz característica de la arista $\{u, v\}$** .

La matriz característica de una arista está bien definida, pues independientemente de la orientación que se le asigne, tendremos que

$$\chi_{u,v}\chi_{u,v}^T = \chi_{v,u}\chi_{v,u}^T.$$

Fijemos una arista $\{u, v\}$. Entonces por construcción, los elementos de $\chi_{u,v}\chi_{u,v}^T$ satisfacen lo siguiente:

- $(\chi_{u,v}\chi_{u,v}^T)_{mn} = (\chi_{u,v}\chi_{u,v}^T)_{nm} = -1$, si $m = u$ y $n = v$ o si $m = v$ y $n = u$,
- $(\chi_{u,v}\chi_{u,v}^T)_{mm} = (\chi_{u,v}\chi_{u,v}^T)_{nn} = 1$, si $m = n = u$ o $m = n = v$, y
- $(\chi_{u,v}\chi_{u,v}^T)_{mn} = 0$ en otro caso.

Ejemplo 2.1 Sea G un grafo de 5 vértices y consideremos la matriz característica de la arista $\{2, 5\}$. Entonces tenemos que

$$\chi_{2,5}\chi_{2,5}^T = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & -1 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 & 1 \end{pmatrix}$$

Como se puede observar, los elementos 2-5 y 5-2 son iguales a -1, los ubicados en las posiciones 2-2 y 5-5 son iguales a 1, y cero en el resto.

Lo interesante de las matrices características es que a través de ellas podemos construir la matriz laplaciana de un grafo.

En efecto, sea $G = (V, E, w)$ un grafo de n vértices y m aristas. Si asignamos una orientación arbitraria a cada arista de G , entonces la matriz laplaciana de G puede ser expresada como

$$L_G = \sum_{\{u,v\} \in E} w_{u,v}\chi_{u,v}\chi_{u,v}^T. \quad (2.3)$$

Nótese que los elementos de la matriz laplaciana cumplen lo siguiente: Si $u \neq v$,

$$(L_G)_{uv} = (L_G)_{vu} = -w_{u,v} \cdot (\chi_{u,v} \chi_{u,v}^T)_{uv}, \quad (2.4)$$

y si $u = v$,

$$(L_G)_{uu} = \sum_{u \leq v} (\chi_{u,v} \chi_{u,v}^T)_{uu} = d(v_u).$$

Otra manera de ver esta construcción es la siguiente: Consideremos nuevamente un grafo $G = (V, E, w)$ de n vértices y m aristas y asignemos una orientación arbitraria a todas las aristas de G . Luego construimos la matriz $B \in \mathbb{R}^{m \times n}$, tal que sus filas estén indexadas por las aristas de G , sus columnas indexadas por los vértices de G , y sus elementos definidos como

$$(B)_{ev} = \begin{cases} 1 & \text{si } v \text{ es el vértice final de } e \\ -1 & \text{si } v \text{ es el vértice inicial de } e \\ 0 & \text{en otro caso.} \end{cases}$$

La matriz B recibe el nombre de **matriz de incidencia**. Si b_e es la columna de B^T asociada a la arista $e = \{u, v\}$, entonces dependiendo de la orientación de e , se cumple que

$$b_e = \chi_{u,v} \quad \text{o} \quad b_e = \chi_{v,u}.$$

Entonces,

$$b_e b_e^T = \chi_{u,v} \chi_{u,v}^T.$$

Por tanto, si $W \in \mathbb{R}^{m \times m}$ es la matriz diagonal tal que sus elementos son $(W)_{e,e} = w_e$, tenemos

$$L_G = \sum_{e \in E} w_e b_e b_e^T = B^T W B. \quad (2.5)$$

De esta manera, a partir de la última expresión, podemos demostrar que la matriz laplaciana es psd, pues basta observar para todo $x \in \mathbb{R}^n$, se cumple que

$$x^T L x = x^T B^T W B x = \|W^{1/2} B x\|^2 \geq 0,$$

donde $W^{1/2}$ es la matriz diagonal cuyos elementos son $(W^{1/2})_{uu} = (W)_{uu}^{1/2}$.

2.2.3. Algunas Propiedades de la Matriz Laplaciana

Primeramente ilustraremos una notable propiedad de la forma cuadrática de la matriz laplaciana de un grafo. Consideremos el grafo T de la Figura 2.2.

Nótese que para cada $x \in \mathbb{R}^n$, su i -ésima coordenada $x(i)$, puede ser expresada como

$$x(i) = x^T \chi_i,$$

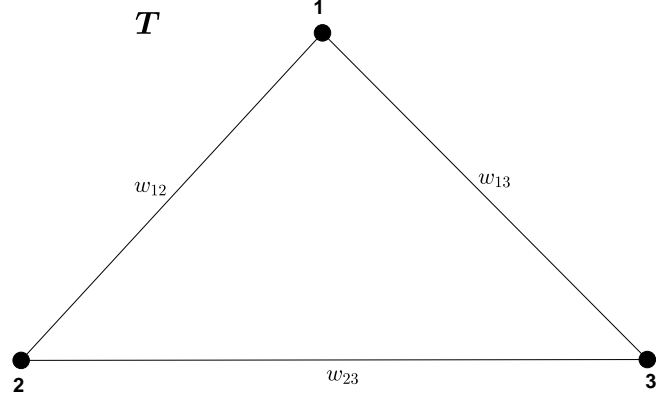


Figura 2.2: Grafo simple, no dirigido y ponderado de 3 vértices

entonces se sigue que

$$\begin{aligned}
& w_{1,2}(x(1) - x(2))^2 + w_{1,3}(x(1) - x(3))^2 + w_{2,3}(x(2) - x(3))^2 \\
&= w_{1,2}(x^T \chi_1 - x^T \chi_2)^2 + w_{1,3}(x^T \chi_1 - x^T \chi_3)^2 + w_{2,3}(x^T \chi_2 - x^T \chi_3)^2 \\
&= w_{1,2}(x^T(\chi_1 - \chi_2))^2 + w_{1,3}(x^T(\chi_1 - \chi_3))^2 + w_{2,3}(x^T(\chi_2 - \chi_3))^2 \\
&= w_{1,2}(x^T \chi_{1,2})^2 + w_{1,3}(x^T \chi_{1,3})^2 + w_{2,3}(x^T \chi_{2,3})^2 \\
&= w_{1,2}x^T \chi_{1,2} \chi_{1,2}^T x + w_{1,3}x^T \chi_{1,3} \chi_{1,3}^T x + w_{2,3}x^T \chi_{2,3} \chi_{2,3}^T x \\
&= w_{1,2}x^T \begin{bmatrix} \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix} \begin{pmatrix} 1 & -1 & 0 \end{pmatrix}^T \end{bmatrix} x + w_{1,3}x^T \begin{bmatrix} \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix} \begin{pmatrix} 1 & 0 & -1 \end{pmatrix}^T \end{bmatrix} x + \\
&\quad w_{2,3}x^T \begin{bmatrix} \begin{pmatrix} 0 \\ 1 \\ -1 \end{pmatrix} \begin{pmatrix} 0 & 1 & -1 \end{pmatrix}^T \end{bmatrix} x \\
&= w_{1,2}x^T \begin{pmatrix} 1 & -1 & 0 \\ -1 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix} x + w_{1,3}x^T \begin{pmatrix} 1 & 0 & -1 \\ 0 & 0 & 0 \\ -1 & 0 & 1 \end{pmatrix} x + w_{2,3}x^T \begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & -1 \\ 0 & -1 & 1 \end{pmatrix} x \\
&= x^T \begin{pmatrix} w_{1,2} & -w_{1,2} & 0 \\ -w_{1,2} & w_{1,2} & 0 \\ 0 & 0 & 0 \end{pmatrix} x + x^T \begin{pmatrix} w_{1,3} & 0 & -w_{1,3} \\ 0 & 0 & 0 \\ -w_{1,3} & 0 & w_{1,3} \end{pmatrix} x + x^T \begin{pmatrix} 0 & 0 & 0 \\ 0 & w_{2,3} & -w_{2,3} \\ 0 & -w_{2,3} & w_{2,3} \end{pmatrix} x \\
&= x^T \begin{pmatrix} d(v_1) & -w_{1,2} & -w_{1,3} \\ -w_{1,2} & d(v_2) & -w_{2,3} \\ -w_{1,3} & -w_{2,3} & d(v_3) \end{pmatrix} x \\
&= x^T L_G x.
\end{aligned}$$

De manera general, si $G = (V, E, w)$ es un grafo de n vértices y m aristas, entonces para todo $x \in \mathbb{R}^n$, se cumple que

$$x^T L_G x = \sum_{\{i,j\} \in E} w_{i,j} (x(i) - x(j))^2. \quad (2.6)$$

Entre otras propiedades importantes de la matriz laplaciana se destaca el hecho de poseer siempre un autovalor igual a cero, por tanto no es invertible. Además, la multiplicidad algebraica del autovalor cero de L_G es igual a la cantidad de componentes conexos de G . En particular, si G es conexo y posee n vértices, entonces el autovalor cero de L_G tiene multiplicidad 1 y $\mathcal{N}(L_G) = \text{gen}(\{\Psi\})$, donde Ψ es el vector de \mathbb{R}^n tal que $\Psi(i) = 1$, para todo $i = 1, \dots, n$ [14].

La Pseudoinversa de la Matriz Laplaciana

Consideremos un grafo conexo G de n vértices. Como L_G es simétrica, entonces por el Teorema Espectral puede ser expresada como

$$L_G = \sum_{i \leq n-1} \lambda_i u_i u_i^T,$$

donde los u_i son los autovectores normalizados asociados a los autovalores $\lambda_i \neq 0$ de la matriz L_G . La *pseudoinversa* de L_G se define como

$$L_G^+ = \sum_{i \leq n-1} \frac{1}{\lambda_i} u_i u_i^T.$$

Entonces, es inmediato que $\mathcal{C}(L_G) = \mathcal{C}(L_G^+) = \text{gen}(\{u_1, u_2, \dots, u_{n-1}\})$, y que $\mathcal{N}(L_G) = \mathcal{N}(L_G^+)$. Además,

$$L_G L_G^+ = L_G^+ L_G = \sum_{i \leq n-1} u_i u_i^T. \quad (2.7)$$

Nótese que la matriz de la Ecuación (2.7) es la que proyecta de manera ortogonal sobre el espacio $\mathcal{C}(L_G)$. Por tanto, si $x \in \mathcal{C}(L_G)$ entonces

$$\left(\sum_{i \leq n-1} u_i u_i^T \right) x = x.$$

Es decir,

$$\sum_{i \leq n-1} u_i u_i^T = \text{Id}(\mathcal{C}(L_G)). \quad (2.8)$$

La pseudoinversa de la matriz laplaciana cumple un papel fundamental en los algoritmos de dispersión espectral que veremos en el siguiente capítulo. También más adelante consideraremos las siguientes matrices definidas como

$$L_G^{1/2} = \sum_{i \leq n-1} \sqrt{\lambda_i} u_i u_i^T, \text{ y } (L_G^+)^{1/2} = \sum_{i \leq n-1} \frac{1}{\sqrt{\lambda_i}} u_i u_i^T.$$

CAPÍTULO 3

DISPERSIÓN ESPECTRAL DE GRAFOS

La dispersión espectral de grafos es una técnica introducida por D. Spielman y S. H. Teng en [3] y consiste en construir un grafo disperso \widehat{G} a partir de un grafo denso G de tal manera que los autovalores de las respectivas matrices laplacianas L_G y $L_{\widehat{G}}$ sean próximas. Más formalmente, sea $G = (V, E, w)$ un grafo simple, ponderado y no dirigido de n vértices. Decimos que un grafo $\widehat{G} = (V, \widehat{E}, \widehat{w})$ es un *dispersor espectral de G* si para todo $x \in \mathbb{R}^n$, se cumple que

$$a \cdot L_G \preceq L_{\widehat{G}} \preceq b \cdot L_G, \quad (3.1)$$

y $0 < a \leq b$. Así, si $\lambda_1, \lambda_2, \dots, \lambda_n$ y $\widehat{\lambda}_1, \widehat{\lambda}_2, \dots, \widehat{\lambda}_n$ son los autovalores de L_G y $L_{\widehat{G}}$, respectivamente, entonces de la Ecuación (3.1) y del teorema de Courant-Fischer (2.4) tenemos que

$$a\lambda_i \leq \widehat{\lambda}_i \leq b\lambda_i.$$

Si los escalares a y b dependen de un valor ϵ , decimos que \widehat{G} es un ϵ -*dispersor espectral de G* . Por lo general el valor de $\epsilon \in (0, 1)$.

Esta técnica ha sido utilizada en problemas tales como la resolución de sistemas lineales de ecuaciones en la que la matriz de coeficientes del sistema es una matriz laplaciana [15], y en modelos de comunicación [16, 17, 18], por mencionar unos pocos.

3.1. Intuición de la Técnica de Dispersión Espectral de Grafos

Ilustraremos el método que aplica la técnica de dispersión espectral, de tal manera a obtener ceros en la matriz laplaciana. Este método aprovecha las propiedades estructurales que tiene

dicha matriz.

Consideremos nuevamente el grafo T de la Figura 2.2. Según la Ecuación (2.3), la matriz laplaciana de T puede ser expresada como

$$\begin{aligned} L_T &= w_{1,2} \cdot \begin{pmatrix} 1 & -1 & 0 \\ -1 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix} + w_{1,3} \cdot \begin{pmatrix} 1 & 0 & -1 \\ 0 & 0 & 0 \\ -1 & 0 & 1 \end{pmatrix} + w_{2,3} \cdot \begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & -1 \\ 0 & -1 & 1 \end{pmatrix} \\ &= \begin{pmatrix} w_{1,2} + w_{1,3} & -w_{1,2} & -w_{1,3} \\ -w_{1,2} & w_{1,2} + w_{2,3} & -w_{2,3} \\ -w_{1,3} & -w_{2,3} & w_{1,3} + w_{2,3} \end{pmatrix}. \end{aligned} \quad (3.2)$$

Consideremos una arista de T , por ejemplo $\{1, 3\}$. En la ecuación (3.2), se puede ver claramente que los elementos $(L_T)_{13}$ y $(L_T)_{31}$, dependen exclusivamente del peso de la arista $\{1, 3\}$, es decir, de $w_{1,3}$ (de hecho, esto es lo que expresa la Ecuación (2.4)). Por tanto, al multiplicar $w_{1,3}$ por el valor cero, tenemos

$$\begin{aligned} L_{\hat{T}} &= w_{1,2} \cdot \begin{pmatrix} 1 & -1 & 0 \\ -1 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix} + 0 \cdot w_{1,3} \cdot \begin{pmatrix} 1 & 0 & -1 \\ 0 & 0 & 0 \\ -1 & 0 & 1 \end{pmatrix} + w_{2,3} \cdot \begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & -1 \\ 0 & -1 & 1 \end{pmatrix} \\ &= \begin{pmatrix} w_{1,2} & -w_{1,2} & 0 \\ -w_{1,2} & w_{1,2} + w_{2,3} & -w_{2,3} \\ 0 & -w_{2,3} & w_{2,3} \end{pmatrix}. \end{aligned}$$

Es decir, los elementos $(L_{\hat{T}})_{13}$ y $(L_{\hat{T}})_{31}$ de la nueva matriz son iguales a cero. Esto es equivalente a eliminar la arista $\{1, 3\}$ de T y obtener un grafo \hat{T} .

Lo importante aquí es notar que modificando el peso de una arista $\{u, v\}$, también se modifican los elementos (y solamente estos elementos) asociados a esta arista en la matriz laplaciana. En particular, si el peso de la arista es multiplicado por cero, entonces equivale a eliminar dicha arista del grafo.

Es claro que las modificaciones de los pesos de las aristas, se debe realizar de forma “inteligente”, de tal manera que las propiedades espectrales de la matriz laplaciana sea preservada. Para ello, se debe cuidar que para todo vector x , la forma cuadrática $x^T L_G x$ no sufra cambios considerables. En la dispersión espectral, estos cambios están acotados según los factores de aproximación que se indica en la Ecuación (3.1).

3.2. Algoritmos para la Construcción de Dispersores Espectrales

En esta sección describiremos los algoritmos de dispersión espectral de grafos que utilizaremos para obtener nuestro resultado principal. Primeramente describiremos el algoritmo

BSS [19], luego el algoritmo de dispersión espectral vía *resistencia efectiva* [20] y finalmente el algoritmo de Lee y Sun [4]. Este último combina algunas ideas de los dos métodos anteriores.

3.2.1. Algoritmo BSS

Este algoritmo fue introducido por Batson, Spielman y Srivastava en [19], y lo denominaremos BSS por practicidad. Recibe como entrada un grafo $G = (V, E, w)$ de n vértices y m aristas, junto con un escalar $\epsilon \in (0, 1)$ y retorna un subgrafo $\widehat{G} = (V, \widehat{E}, \widehat{w})$ de $\mathcal{O}(n/\epsilon^2)$ aristas tal que

$$(1 - \epsilon)^2 L_G \preceq L_{\widehat{G}} \preceq (1 + \epsilon)^2 L_G.$$

La idea central de este algoritmo se basa en el siguiente teorema.

Teorema 3.1. Sean $d > 1$ y v_1, v_2, \dots, v_m vectores de \mathbb{R}^n tal que

$$\sum_{i \leq m} v_i v_i^T = I.$$

Entonces, existe un conjunto de escalares $s_i \geq 0$, con $|\{s_i : s_i \neq 0\}| = \mathcal{O}(dn)$, tal que

$$\left(1 - \frac{1}{\sqrt{d}}\right)^2 I \preceq \sum_{i \leq m} s_i v_i v_i^T \preceq \left(1 + \frac{1}{\sqrt{d}}\right)^2 I. \quad (3.3)$$

El tiempo de ejecución de este algoritmo es $\mathcal{O}(mn^3)$.

La demostración de este teorema es constructiva y provee un método iterativo para hallar los escalares s_i . Desde un punto de vista general, dicho método procede como sigue:

Se empieza definiendo una matriz $A = \mathbf{0}$, junto con unos escalares l_0 y u_0 denominados barreras, tal que $l_0 < 0 < u_0$. En la i -ésima iteración, se elige un apropiado vector v junto con un escalar $t > 0$ y se construye la matriz $A_i = A_{i-1} + tvv^T$ de tal manera que el invariante

$$l_i I \preceq A_i \preceq u_i I,$$

siempre se cumpla. En cada iteración, la elección del vector v depende de las funciones $\Phi^u = \text{Tr}(uI - A)^{-1}$ y $\Phi_l = \text{Tr}(A - lI)^{-1}$, que son denominadas *upper potentials* y *lower potentials*, respectivamente. Las funciones Φ_l y Φ^u junto con las barreras, controlan las propiedades espectrales de A . Luego de actualizar la matriz A , se actualizan las barreras a $u_{i+1} = u_i + \delta_u$ y $l_{i+1} = l_i + \delta_l$, donde δ_u y δ_l son unas constantes fijadas al inicio del algoritmo. Así, después de $\Theta(n/\epsilon^2)$ iteraciones, se llega la garantía indicada en la Ecuación (3.3). Si al final del proceso un vector v_i es elegido k veces, entonces hacemos

$$t_1 v_i + t_2 v_i + \dots + t_k v_i = \left(\sum_{h \leq k} t_h \right) v_i = s_i v_i.$$

La obtención de las s_i constituye el mayor costo computacional del algoritmo.

Ahora, con esta idea en mente, podemos enunciar el siguiente teorema.

Teorema 3.2 (Dispersión Espectral de Grafos). Sean $G = (V, E, w)$ un grafo ponderado de n vértices, m aristas y ϵ un número real tal que $0 < \epsilon < 1$. Entonces, existe un grafo \widehat{G} con $\mathcal{O}(n/\epsilon^2)$ aristas tal que

$$(1 - \epsilon)^2 L_G \preceq L_{\widehat{G}} \preceq (1 + \epsilon)^2 L_G.$$

Demostración. Sin pérdida de generalidad supongamos que G es conexo, entonces $\mathcal{N}(L_G)$ es igual al espacio generado por el vector $\Psi \in \mathbb{R}^n$, y el autovalor cero de L_G tiene multiplicidad algebraica igual a 1.

Por otro lado, hemos visto en la Ecuación (2.5) que la matriz laplaciana de G puede ser expresada como

$$L_G = B^T W B = \sum_{e \in E} w_e b_e b_e^T.$$

Entonces si para toda arista $e \in E$, definimos $z_e = \sqrt{w_e} b_e$ y $v_e = (L_G^+)^{1/2} z_e$, tenemos que

$$L_G = \sum_{e \in E} z_e z_e^T,$$

y

$$\begin{aligned} \sum_{e \in E} v_e v_e^T &= \sum_{e \in E} (L_G^+)^{1/2} z_e z_e^T (L_G^+)^{1/2} \\ &= (L_G^+)^{1/2} \left(\sum_{e \in E} z_e z_e^T \right) (L_G^+)^{1/2} \\ &= (L_G^+)^{1/2} L_G (L_G^+)^{1/2} \\ &= \sum_{i \leq n-1} u_i u_i^T, \end{aligned}$$

donde los u_i son los autovectores ortonormales asociados a los autovalores $\lambda_i \neq 0$ de la matriz L_G . Así, de las ecuaciones (2.7) y (2.8), tenemos que

$$\sum_{e \in E} v_e v_e^T = Id(\mathcal{C}(L_G)).$$

Entonces por el Teorema 3.1, eligiendo $\epsilon = 1/\sqrt{d}$, existe un conjunto de escalares $s_i \geq 0$, con $|\{s_i : s_i \neq 0\}| = \mathcal{O}(dn)$ tal que

$$(1 - \epsilon)^2 y^T y \leq y^T \left(\sum_{e \in E} s_e v_e v_e^T \right) y \leq (1 + \epsilon)^2 y^T y, \quad \forall y \in \mathcal{C}(L_G). \quad (3.4)$$

Dado que $\mathcal{C}(L_G) = \mathcal{C}(L_G^{1/2})$, entonces cuantificar sobre todos los vectores $y \in \mathcal{C}(L_G)$ es igual a cuantificar sobre todos los vectores $L_G^{1/2} x$, donde $x \in \mathbb{R}^n$. Así, al sustituir y por $L_G^{1/2} x$ en la

Ecuación (3.4) tenemos

$$(1 - \epsilon)^2 x^T L_G x \leq x^T L_G^{1/2} \left(\sum_{e \in E} s_e v_e v_e^T \right) L_G^{1/2} x \leq (1 + \epsilon)^2 x^T L_G x \quad \forall x \in \mathbb{R}^n.$$

Atendiendo que $v_e = (L_G^+)^{1/2} z_e$, la última expresión es igual a

$$(1 - \epsilon)^2 x^T L_G x \leq x^T L_G^{1/2} \left(\sum_{e \in E} s_e (L_G^+)^{1/2} z_e z_e^T (L_G^+)^{1/2} x \right) L_G^{1/2} x \leq (1 + \epsilon)^2 x^T L_G x \quad \forall x \in \mathbb{R}^n,$$

que al desarrollar nos queda

$$(1 - \epsilon)^2 x^T L_G x \leq x^T L_G^{1/2} (L_G^+)^{1/2} \left(\sum_{e \in E} s_e z_e z_e^T \right) (L_G^+)^{1/2} L_G^{1/2} x \leq (1 + \epsilon)^2 x^T L_G x \quad \forall x \in \mathbb{R}^n.$$

Definiendo $L_{\hat{G}} = \sum_e s_e z_e z_e^T$ y sabiendo que

$$L_G^{1/2} (L_G^+)^{1/2} = (L_G^+)^{1/2} L_G^{1/2} = \sum_{i \leq n-1} u_i u_i^T,$$

tenemos

$$(1 - \epsilon)^2 x^T L_G x \leq x^T L_{\hat{G}} x \leq (1 + \epsilon)^2 x^T L_G x \quad \forall x \in \mathbb{R}^n,$$

lo cual es equivalente al resultado deseado. □

Por tanto, si definimos la matriz digonal $S \in \mathbb{R}^{m \times m}$ tal que $(S)_{ee} = s_e$, entonces en términos de operaciones con matrices,

$$L_{\hat{G}} = \sum_{e \in E} s_e w_e b_e b_e^T = B^T S W B.$$

Es decir, el dispersor espectral se obtiene computando $B^T S W B$.

3.2.2. Dispersión Espectral vía Resistencia Efectiva

Este algoritmo es probabilístico y fue desarrollado por Spielman y Srivastava en [20]. Recibe como entrada un grafo $G = (V, E, w)$ de n vértices y m aristas, y un valor $\epsilon \in (1/\sqrt{n}, 1)$. En un tiempo de $\tilde{O}(m)$, retorna un subgrafo $\hat{G} = (V, \hat{E}, \hat{w})$ con $\mathcal{O}(n \log n / \epsilon^2)$ aristas, tal que con alta probabilidad, se cumple que

$$(1 - \epsilon) L_G \preceq L_{\hat{G}} \preceq (1 + \epsilon) L_G.$$

El concepto principal utilizado para este algoritmo es el de la *resistencia efectiva de una*

arista, que matemáticamente se define como

$$R_{u,v} = (\chi_u - \chi_v)^T L_G^+(\chi_u - \chi_v).$$

$R_{u,v}$ denota la resistencia efectiva de la arista $\{u, v\}$.

El algoritmo en cuestión se denomina **Sparsify** y funciona de la siguiente manera:

$$\widehat{G} = \mathbf{Sparsify}(G, q)$$

- 1: Elegir de forma aleatoria una arista e de G con probabilidad $p_e = w_e R_e / \sum_{e \in E} w_e R_e$ y añadir a \widehat{G} . Tomar q muestras independientes con reemplazos. En el caso en que una arista sea elegida más de una vez, sumar los pesos de la arista.

La distribución de probabilidad utilizada implica que la esperanza $\mathbf{E}(\widehat{w}_e) = w_e$, y $\mathbf{E}(L_{\widehat{G}}) = L_G$. Spielman y Srivastava [20] demostraron que si el tamaño de la muestra es $\mathcal{O}(n \log n / \epsilon^2)$, entonces con probabilidad al menos $1/2$, se cumple

$$(1 - \epsilon)L_G \preceq L_{\widehat{G}} \preceq (1 + \epsilon)L_G.$$

Para llevar a cabo este proceso en términos de operaciones con matrices, debemos notar que la distribución de probabilidad empleada para la selección de las aristas, implica que el peso de la arista e en el grafo \widehat{G} es $w_e = \frac{k_e w_e}{p_e q}$, donde k_e es la cantidad de veces que la arista e fue seleccionada. Entonces definiendo la matriz diagonal $S \in \mathbb{R}^{m \times m}$, tal que su elemento $(S)_{ee} = \frac{k_e w_e}{p_e q}$, tenemos que la matriz laplaciana del dispersor espectral, se puede calcular simplemente haciendo

$$L_{\widehat{G}} = B^T S W B.$$

El principal desafío de este algoritmo es el cómputo de las resistencias efectivas, pues involucra el cómputo de la matriz L_G^+ , que tomaría un tiempo de $\mathcal{O}(n^3)$. Para superar este problema, Spielman y Srivastava propusieron un método que permita aproximar las resistencias efectivas en un tiempo $\widetilde{\mathcal{O}}(m/\epsilon^2)$. Para entender este método, primeramente debemos observar que la resistencia efectiva de la arista $\{u, v\}$ puede ser expresada como

$$\begin{aligned} R_{u,v} &= (\chi_u - \chi_v)^T L_G^+(\chi_u - \chi_v) \\ &= (\chi_u - \chi_v)^T L_G^+ L_G L_G^+(\chi_u - \chi_v) \\ &= ((\chi_u - \chi_v)^T L_G^+ B^T W^{1/2}) (W^{1/2} B L_G^+(\chi_u - \chi_v)) \\ &= \|W^{1/2} B L^+(\chi_u - \chi_v)\|^2. \end{aligned}$$

De esta manera, para hallar $R_{u,v}$, simplemente tenemos que calcular la distancia entre los vectores $W^{1/2} B L^+ \chi_u$ y $W^{1/2} B L^+ \chi_v$, que es equivalente a hallar la distancia entre los vectores columnas u y v de la matriz $W^{1/2} B L^+ \chi_u$. Para computar esta distancia de forma eficiente, se combina un método de reducción de dimensión, junto con un algoritmo de tiempo casi lineal

de resolución de sistemas de ecuaciones lineales en la que la matriz del sistema es una matriz laplaciana.

El método de reducción de dimensión mencionado fue proporcionado por Achlioptas en [21]. Se trata de una versión alternativa del conocido Lema de Johnson–Lindenstrauss, el cual básicamente establece que dado un conjunto de n puntos de dimensión d , siempre podemos encontrar un espacio de dimensión $\mathcal{O}(\log n/\epsilon^2)$ sobre el cual se pueda representar los puntos, de tal manera que las distancias entre cada par de puntos en el espacio original, sea próxima en un factor que depende de ϵ , a la distancia entre los respectivos pares de puntos sobre el nuevo espacio.

Teorema 3.3 (Achlioptas [21]). Sean v_1, v_2, \dots, v_n vectores de \mathbb{R}^d y $\epsilon > 0$. Sea Q una matriz aleatoria de orden $k \times d$, donde $k \geq 24 \log n/\epsilon^2$, tal que sus elementos $(Q)_{ij}$ cumplen la siguiente ley de distribución de probabilidad

$$(Q)_{ij} = \begin{cases} 1/\sqrt{k} & \text{con probabilidad } 1/2 \\ -1/\sqrt{k} & \text{con probabilidad } 1/2. \end{cases}$$

Entonces, con probabilidad al menos $1 - 1/n$ se cumple

$$(1 - \epsilon)\|v_i - v_j\|^2 \leq \|Qv_i - Qv_j\|^2 \leq (1 + \epsilon)\|v_i - v_j\|^2$$

para todos los pares $i, j \leq n$.

De esta manera, el problema se reduce a computar las distancias entre las proyecciones $\{QW^{1/2}BL^+\chi_v\}$. Para ello, utilizaremos el algoritmo **STsolve** de tiempo casi lineal que fue desarrollado por Spielman y Teng en [15].

Teorema 3.4 (STsolve). Existe un algoritmo $x = \mathbf{STsolve}(L, y, \delta)$ el cual toma una matriz laplaciana L , un vector columna y , y un parámetro de error $\delta > 0$, y retorna un vector columna x tal que

$$\|x - L^+y\|_L \leq \epsilon\|L^+y\|_L,$$

donde $\|y\|_L = \sqrt{y^T L y}$. La esperanza del tiempo de ejecución es $\tilde{O}(m \log(1/\delta))$, donde m es el número de elementos distintos de cero de L .

Con este algoritmo se construye una matriz \hat{Z} de tal manera que cada una de sus filas \hat{z}_i , sea próxima a la respectiva fila z_i de la matriz $Z = QW^{1/2}BL^+$. El proceso es el siguiente:

1. Utilizar el método de Achlioptas para construir una matriz $Q \in \mathbb{R}^{k \times m}$, donde $k = 24 \log n/\epsilon^2$.
2. Computar $Y = QW^{1/2}B$, el cual se realiza en tiempo $\tilde{O}(m/\epsilon^2)$.
3. Computar $\hat{z}_i = \mathbf{STsolve}(L, y_i, \delta)$, por cada fila y_i de Y . Tomar $\delta = \frac{\epsilon}{3} \left(\frac{2(1 - \epsilon)w_{\min}}{n^3(1 + \epsilon)w_{\max}} \right)^{1/2}$.

Según el Lema 9 de [20] esto genera una matriz \widehat{Z} tal que para toda arista $\{u, v\}$,

$$(1 - \epsilon)^2 R_{uv} \leq \|\widehat{Z}(\chi_u - \chi_v)\|^2 \leq (1 + \epsilon)^2 R_{uv}.$$

3.2.3. BSS Probabilístico

Este algoritmo fue introducido por Lee y Sun en [4]. Como entrada recibe un grafo $G = (V, E, w)$ de n vértices y m aristas, y un parámetro $\epsilon \in (0, 1/120]$, de tal manera que en un tiempo de $\widehat{O}(m/\epsilon^2)$, retorna un subgrafo $\widehat{G} = (V, \widehat{E}, \widehat{w})$ de $\mathcal{O}(n/\epsilon^2)$ aristas tal que con alta probabilidad,

$$(1 - \epsilon)L_G \preceq L_{\widehat{G}} \preceq (1 + \epsilon)L_G.$$

Si comparamos el algoritmo de Lee y Sun con el algoritmo de resistencia efectiva, podemos ver que el primero genera un dispersor espectral con menor cantidad aristas, aunque en un tiempo levemente mayor. Por otro lado, comparado con el BSS, la cantidad de aristas generadas son del mismo orden, pero el tiempo de ejecución del algoritmo de Lee y Sun es considerablemente menor. Lo novedoso de este algoritmo es la combinación de los dos algoritmos mencionados previamente, además de una idea similar sobre la *función potencial* introducida en [22].

Desde un punto de vista general, el algoritmo de Lee y Sun es similar al BSS en el sentido que también consiste en la construcción de manera iterativa de una matriz A por medio de suma de matrices de rango 1. Sin embargo, se diferencia del BSS por el hecho de que en cada iteración, computa las denominadas *resistencias efectivas relativas* definidas como

$$R_i(A_j, u_j, l_j) = v_i^T (u_j I - A_j)^{-1} v_i + v_i^T (A_j - l_j I)^{-1} v_i,$$

para todo v_i . Además, en cada iteración, ya no se selecciona simplemente un vector v para actualizar la matriz A , sino una cantidad N_j de vectores de forma aleatoria, donde la probabilidad de que cada vector sea seleccionado, depende de su resistencia efectiva relativa. El escalar N_j que representa la cantidad de vectores seleccionados en la fase j , se define como

$$N_j = \frac{1}{n^{2/q}} \left(\sum_{i \leq m} R_i(A_j, u_j, l_j) \right) \cdot \min\{\lambda_{\min}(u_j I - A_j), \lambda_{\min}(A_j - l_j I)\}.$$

Con el fin de controlar las propiedades espectrales de A , se utiliza la siguiente función

$$\Phi_{u,l}(A) = \text{Tr}(uI - A)^{-q} + \text{Tr}(A - lI)^{-q},$$

donde $q \geq 10$. Finalmente, para proceder a la siguiente fase, las barreras inferior y superior son actualizadas por medio de los factores $\Delta_{l,j}$ y $\Delta_{u,j}$, respectivamente.

A continuación damos una descripción del algoritmo de Lee y Sun.

Input: $\epsilon \leq 1/120$, $q \geq 10$.

- 1: $j = 0$;
 - 2: $l_0 = -(2n)^{1/q}$, $u_0 = (2n)^{1/q}$, $A_0 = 0$;
 - 3: **while** $u_j - l_j < 4(2n)^{1/q}$, **do**
 - 4: $W_j = 0$;
 - 5: Computar $R_i(A_j, u_j, l_j)$ para todos los vectores v_i ;
 - 6: Tomar una muestra de N_j vectores de forma independiente con reemplazamiento, tal que cada v_i sea elegido con probabilidad proporcional a $R_i(A_j, u_j, l_j)$. Por cada vector v seleccionado, sumar $(\epsilon/q)(R_i(A_j, u_j, l_j))^{-1} \cdot vv^T$ a W_j ;
 - 7: $A_{j+1} = A_j + W_j$;
 - 8: $u_{j+1} = u_j + \Delta_{u,j}$, $l_{j+1} = l_j + \Delta_{l,j}$;
 - 9: $j = j + 1$;
 - 10: **Return** A_j ;
-

Como se puede observar, el cómputo en cada fase de las resistencias efectivas relativas de todos los vectores, constituye la parte con mayor complejidad computacional del algoritmo. Para superar este inconveniente, Lee y Sun proceden de manera muy similar a Spielman y Srivastava, pues también utilizan la combinación del método de reducción de Achlioptas y el algoritmo **STsolve**. Omitimos esta parte por ser muy similar al proceso de Spielman y Srivastava. Para más detalles ir a la sección 4 de [4].

Formalmente, el resultado de Lee y Sun se resume en el siguiente teorema:

Teorema 3.5 (Lee-Sun). Sea $G = (V, E, w)$ un grafo simple, no dirigido y ponderado de n vértices y m aristas. Sean $0 < \epsilon \leq 1/120$ y $q \geq 10$ dos escalares. Entonces, existe un algoritmo de tiempo $\tilde{O}\left(\frac{qmn^{5/q}}{\epsilon^{4+4/q}}\right)$, que retorna un grafo $\widehat{G} = (V, \widehat{E}, w)$ de $\mathcal{O}\left(\frac{qn}{\epsilon^2}\right)$ aristas, tal que

$$(1 - \epsilon)L_G \preceq L_{\widehat{G}} \preceq (1 + \epsilon)L_G.$$

CAPÍTULO 4

CONSTRUCCIÓN DE MATRICES DISPERSAS

En este capítulo estudiaremos el problema central de esta tesis. Para ello, recordemos primeramente que una matriz $M \in \mathbb{R}^{n \times n}$ es ODN si para todo $i \neq j$, $(M)_{ij} \geq 0$. Así, con esta definición en mente, nuestro problema central puede ser enunciado de la siguiente manera: *Dada una matriz $M \in \mathbb{R}^{n \times n}$ simétrica y ODN, encontrar una matriz dispersa $\widehat{M} \in \mathbb{R}^{n \times n}$ que sea próxima en espectro a M .* Una matriz $M \in \mathbb{R}^{n \times n}$ es considerada *dispersa* si tiene $\mathcal{O}(n)$ elementos no nulos. Por otro lado, decimos que M es *densa* si su cantidad de elementos no nulos es $\Omega(n^2)$.

Para resolver este problema central, debemos responder las siguientes preguntas fundamentales:

1. ¿cómo dispersar¹ la matriz de tal manera que las propiedades espectrales sean preservadas?, y
2. ¿cómo acotar o controlar el cambio que el proceso de dispersión genera en el espectro de la matriz?

La respuesta a la primera pregunta es casi inmediata a partir de la dispersión espectral de grafos, pues, hemos visto que dado un grafo G , siempre es posible construir un grafo disperso \widehat{G} , tal que las respectivas matrices laplacianas L_G y $L_{\widehat{G}}$, sean próximas en espectro. Dado que $L_G = D_G - A_G$, entonces al dispersar L_G , también estamos dispersando la matriz de adyacencias A_G , obteniendo $A_{\widehat{G}}$; por tanto, es natural esperar que las matrices de adyacencias *hereden* una cierta proximidad espectral. De ser así, es decir, si existiera una relación espectral que las matrices A_G y $A_{\widehat{G}}$ heredan luego del proceso de dispersión, entonces ya estaríamos muy cerca

¹Dispersar una matriz M significa que se está construyendo una matriz dispersa a partir de M

de responder la primera pregunta, pues las matrices A_G y $A_{\widehat{G}}$ son *ODN*, y sería suficiente aplicar el siguiente proceso.

- Tomar una matriz simétrica *ODN* tal que sus elementos de la diagonal sean todos iguales a cero y considerarlo como la matriz de adyacencias de un grafo.
- Construir su matriz laplaciana correspondiente.
- Llevar a cabo el proceso de dispersión espectral de grafos.
- Extraemos las matrices de adyacencias A_G y $A_{\widehat{G}}$, las cuales son próximas en espectro.

El problema para llevar a cabo esta idea es que según nuestro entendimiento actual, no existe ninguna relación espectral entre las matrices A_G y $A_{\widehat{G}}$. Además, proceder de esta manera, solo daría una solución al caso en el que la matriz simétrica *ODN* considerada, tenga diagonal nula (que sus elementos de la diagonal sean todos iguales a cero), lo cual es muy restrictivo. Por ello, también estudiamos la relación espectral entre las matrices de adyacencias A_G y $A_{\widehat{G}}$, obteniendo una cota relativamente buena. Luego, utilizamos este resultado para responder nuestro segundo problema principal. Además, en nuestro Teorema principal (1.1) daremos una respuesta a un caso más general de este problema; es decir, cuando la matriz simétrica *ODN* tiene elementos no nulos en la diagonal.

La segunda pregunta consiste básicamente en cómo cambian los autovectores de la matriz simétrica considerada, en comparación con sus respectivos autovectores de la matriz dispersada. Para responder esta pregunta, nos apoyaremos en el teorema de Davis-Kahan el cual hemos visto en el capítulo 2.

Trabajar con matrices dispersas supone algunas ventajas pues, requieren menos costo de almacenamiento y los cálculos que dependen de la cantidad de elementos no nulos de la matriz, se pueden realizar de manera más eficiente. Por ejemplo, para el cálculo de autovalores de matrices simétricas, lo más usual es utilizar el algoritmo de Lanczos, el cual tiene un tiempo de cálculo de $\mathcal{O}(n^3)$. Entonces, dispersando la matriz, podemos reducir considerablemente esta complejidad. Demostraremos de hecho que puede ser reducida a $\mathcal{O}(n^2/\epsilon^2)$, donde $\epsilon \in (0, 1/120]$.

4.1. Resultado Principal

Como hemos mencionado en el capítulo 1, el resultado principal de esta tesis se enuncia en el siguiente teorema.

Teorema (Copia del Teorema 1.1). Sea $M \in \mathbb{R}^{n \times n}$ una matriz simétrica y *ODN* con autovalores $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n$ y respectivos autovectores x_1, x_2, \dots, x_n . Definamos $\lambda_0 = -\infty$ y $\lambda_{n+1} = \infty$. Entonces, para todo $\epsilon > 0$ tal que $0 < \epsilon \leq 1/120$, existe una matriz simétrica y *ODN* $\widehat{M} \in \mathbb{R}^{n \times n}$ con $\mathcal{O}(\frac{n}{\epsilon^2})$ elementos distintos de cero, tal que para todo $i = 1, \dots, n$ se cumple que

$$|\lambda_i - \widehat{\lambda}_i| \leq \epsilon \sqrt{n} \rho(L_M) + \frac{\Delta_M - \delta_M}{2},$$

donde $\widehat{\lambda}_1 \geq \widehat{\lambda}_2 \geq \dots \geq \widehat{\lambda}_n$ son los autovalores de \widehat{M} . Además, si $\widehat{x}_1, \widehat{x}_2, \dots, \widehat{x}_n$ son autovectores respectivos de $\widehat{\lambda}_1 \geq \widehat{\lambda}_2 \geq \dots \geq \widehat{\lambda}_n$ y θ_i es el ángulo (agudo) entre los autoespacios generados por autovectores x_i y \widehat{x}_i , entonces

$$\sin \theta_i \leq \frac{\epsilon \sqrt{n} \rho(L_M) + (\Delta_M - \delta_M)/2}{\min\{|\widehat{\lambda}_{i-1} - \lambda_i|, |\lambda_i - \widehat{\lambda}_{i+1}|\}}.$$

La idea clave para la demostración del Teorema 1.1 consiste en asociar la matriz M simétrica y ODN dada, con la matriz de adyacencia y la matriz de grados de un grafo. Luego, construir la matriz laplaciana $L_M = D_M - A_M$ y llevar a cabo la técnica de dispersión espectral por medio del algoritmo de Lee y Sun. La razón por la que utilizamos el algoritmo de Lee y Sun, se sustenta en que es el algoritmo que produce un dispersor espectral con menos aristas, lo cual implica la obtención de una matriz de adyacencia más dispersa. La importancia de lo último, yace en que la matriz \widehat{M} buscada, la obtendremos a partir de la matriz de adyacencias que resulta del proceso de dispersión espectral. Para comparar los espectros entre M y \widehat{M} , utilizaremos dos teoremas provenientes de la Teoría de la Perturbación que son: el Teorema de Perturbación de Weyl para comparar los autovalores, y el Teorema de Davis-Kahan para la comparación entre los autovectores.

4.2. Lemas Técnicos

Lema 4.1. Sean G un grafo simple, ponderado y no dirigido de n vértices, y \widehat{G} un ϵ -dispersor espectral de G . Sean $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n$ los autovalores de L_G con respectivos autovectores x_1, x_2, \dots, x_n , y sean $\widehat{\lambda}_1 \geq \widehat{\lambda}_2 \geq \dots \geq \widehat{\lambda}_n$ los autovalores de $L_{\widehat{G}}$ con respectivos autovectores $\widehat{x}_1, \widehat{x}_2, \dots, \widehat{x}_n$. Entonces:

1. $\|L_G - L_{\widehat{G}}\| \leq \epsilon \rho(L_G)$, y
2. si θ_i es el ángulo agudo entre x_i y \widehat{x}_i , entonces

$$\sin \theta_i \leq \frac{\epsilon \rho(L_G)}{\min\{\widehat{\lambda}_{i-1} - \lambda_i, \widehat{\lambda}_{i+1} - \lambda_i\}}.$$

Demostración. Como \widehat{G} es un ϵ -dispersor espectral de G , tenemos que

$$(1 - \epsilon)L_G \preceq L_{\widehat{G}} \preceq (1 + \epsilon)L_G,$$

lo cual implica que

$$L_G - L_{\widehat{G}} \preceq \epsilon L_G. \tag{4.1}$$

Dado que δ es un autovalor de $L_G - L_{\widehat{G}}$ si y solo si, $-\delta$ es un autovalor de $L_{\widehat{G}} - L_G$, sin pérdida de generalidad supongamos que $\rho(L_G - L_{\widehat{G}})$ coincide con el autovalor más grande de $L_G - L_{\widehat{G}}$

y sea z el autovector normalizado asociado a $\rho(L_G - L_{\widehat{G}})$. Entonces

$$\begin{aligned}
\|L_G - L_{\widehat{G}}\| &= \rho(L_G - L_{\widehat{G}}) && (L_G - L_{\widehat{G}} \text{ es simétrica}) \\
&= z^T(L_G - L_{\widehat{G}})z \\
&\leq z^T(\epsilon L_G)z && (\text{de la Ec. (4.1)}) \\
&\leq \rho(\epsilon L_G) \\
&= \epsilon \cdot \|L_G\| = \epsilon \cdot \rho(L_G).
\end{aligned}$$

De esta manera la primera parte queda demostrada. La segunda parte se obtiene directamente del teorema de Davis-Kahan. \square

Lema 4.2. Sean $L_G = D_G - A_G$ y $L_H = D_H - A_H$ las matrices laplacianas de los grafos G y H , respectivamente. Entonces

$$\|A_G - A_H\| \leq \sqrt{n}\|L_G - L_H\|.$$

Demostración. La matriz $A_G - A_H$ es simétrica, entonces $\|A_G - A_H\|_\infty = \|A_G - A_H\|_1$. Así, usando la desigualdad (1) de la Proposición 3, tenemos que $\|A_G - A_H\| \leq \|A_G - A_H\|_\infty$. Por otro lado, $\|L_G - L_H\|_\infty = \|A_G - A_H\|_\infty + \max_{i \leq n} |(D_G)_{ii} - (D_H)_{ii}|$, entonces

$$\begin{aligned}
\|A_G - A_H\|_\infty &\leq \|L_G - L_H\|_\infty \\
&\leq \sqrt{n}\|L_G - L_H\|,
\end{aligned}$$

donde la última desigualdad se obtiene a partir de la desigualdad (2) de la Proposición 3. De esta manera, el Lema queda demostrado. \square

4.3. Prueba del Resultado Principal (Teorema 1.1)

Sea M una matriz simétrica ODN y sea \overline{M} otra matriz definida como

$$(\overline{M})_{ij} = \begin{cases} (M)_{ij} & \text{si } i \neq j \\ d & \text{si } i = j, \end{cases}$$

donde $d = (\Delta_M + \delta_M)/2$. Entonces tenemos que

$$\|M - \overline{M}\| = \max\{\Delta_M - d, \delta_M - d\} = (\Delta_M - \delta_M)/2, \quad (4.2)$$

Nótese que el valor de d seleccionado es el que minimiza la norma $\|M - \overline{M}\|$.

Por otro lado, definamos las matrices A_M y D_M como

$$(A_M)_{ij} = \begin{cases} (M)_{ij} & \text{si } i \neq j \\ 0 & \text{si } i = j, \end{cases}$$

y

$$(D_M)_{ij} = \begin{cases} 0 & \text{si } i \neq j \\ \sum_{\substack{j \leq n \\ j \neq i}} (M)_{ij} & \text{si } i = j, \end{cases}$$

respectivamente. De esta manera tenemos que $A_M = \overline{M} - dI$. Nótese que las matrices A_M y D_M pueden ser vistas como la matriz de adyacencia y la matriz de grados de un grafo G_M . Consecuentemente, $L_M = D_M - A_M$ es la matriz laplaciana de G_M .

Del teorema de Lee y Sun sabemos que dado $0 < \epsilon \leq 1/120$, existe una matriz \widehat{L}_M con $\mathcal{O}(n/\epsilon^2)$ elementos distintos de cero, tal que \widehat{L}_M es un ϵ -dispersor espectral de L_M . Entonces por el Lema 4.1 se cumple que

$$\|L_M - \widehat{L}_M\| \leq \epsilon \rho(L_M),$$

y por el Lema 4.2

$$\|A_M - \widehat{A}_M\| \leq \epsilon \sqrt{n} \rho(L_M).$$

Definiendo la matriz $\widehat{M} = \widehat{A}_M + dI$ y utilizándola en la última en desigualdad tenemos

$$\|\overline{M} - \widehat{M}\| \leq \epsilon \sqrt{n} \rho(L_M). \quad (4.3)$$

Luego, de (4.2) y (4.3), se sigue que

$$\begin{aligned} \|M - \widehat{M}\| &= \|M - \widehat{M} + \overline{M} - \overline{M}\| \\ &\leq \|M - \overline{M}\| + \|\overline{M} - \widehat{M}\| \\ &\leq (\Delta_M - \delta_M)/2 + \epsilon \sqrt{n} \rho(L_M). \end{aligned}$$

Por tanto, si $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n$ y $\widehat{\lambda}_1 \geq \widehat{\lambda}_2 \geq \dots \geq \widehat{\lambda}_n$ son los autovalores de las matrices M y \widehat{M} , respectivamente, entonces por el Teorema de Perturbación de Weyl, es inmediato que

$$|\lambda_i - \widehat{\lambda}_i| \leq (\Delta_M - \delta_M)/2 + \epsilon \sqrt{n} \rho(L_M).$$

El hecho de que \widehat{M} sea ODN y tenga $\mathcal{O}(\frac{n}{\epsilon^2})$ elementos no nulos, se deduce directamente del teorema de Lee y Sun. La última parte del teorema se obtiene inmediatamente a partir del teorema de Davis-Kahan, esto es

$$\sin \theta_i \leq \frac{\|M - \widehat{M}\|}{\min\{|\widehat{\lambda}_{i-1} - \lambda_i|, |\lambda_i - \widehat{\lambda}_{i+1}|\}} \leq \frac{\epsilon \sqrt{n} \rho(L_M) + (\Delta_M - \delta_M)/2}{\min\{|\widehat{\lambda}_{i-1} - \lambda_i|, |\lambda_i - \widehat{\lambda}_{i+1}|\}}.$$

4.4. Aplicaciones

En esta sección presentamos dos ejemplos de aplicaciones del Teorema 1.1. Primeramente a una técnica de minería de datos denominada *PCA* y luego a un problema de optimización.

4.4.1. PCA

El Análisis de Componentes Principales o *PCA* por sus siglas en inglés, es una técnica de reducción de dimensión de datos. Dado un conjunto de datos con variables correlacionadas, PCA reduce el número de variables, mientras preserva la varianza de los datos lo máximo posible. Dicha reducción se realiza construyendo un nuevo conjunto de variables no correlacionadas, conocidas como *componentes principales* [23].

Sea x un vector de n variables aleatorias. El primer componente principal es definido como $z_1 = v_1^T x$, donde $v_1 \in \mathbb{R}^n$ es un vector que maximiza la varianza de z_1 , la cual es denotada por $Var(z_1)$. De manera similar, el i -ésimo componente principal, es la variable definida como $z_i = v_i^T x$, con $v_i \in \mathbb{R}^n$, de tal manera que no esté correlacionada con los componentes z_1, \dots, z_{i-1} , y la varianza $Var(z_i)$ sea máxima. Es bien conocido que si S es la matriz de covarianzas de x , con autovalores $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n$, entonces v_i es el autovector de S , correspondiente a λ_i ; además $Var(z_i) = \lambda_i$.

Una manera alternativa para aplicar PCA es utilizando la matriz de correlaciones en lugar de la matriz de covarianzas. Sea X es una matriz de datos de $m \times n$. Sin pérdida de generalidad, supongamos que la respectiva matriz de correlaciones sea $M = (1/n)X^T X$. Supongamos además que M es ODN. Por el Teorema 1.1, podemos construir una matriz dispersa \widehat{M} que es próxima en espectro a M . Sean z_i y \widehat{z}_i el i -ésimo componente principal de M y \widehat{M} , respectivamente. Entonces, para el primer componente principal tenemos

$$Var(z_1) - \epsilon\sqrt{n}\rho(L_M) \leq Var(\widehat{z}_1) \leq Var(z_1) + \epsilon\sqrt{n}\rho(L_M).$$

En general, para los primeros p componentes principales, tenemos

$$\sum_{i=1}^p Var(z_i) - i\epsilon\sqrt{n}\rho(L_M) \leq \sum_{i=1}^p Var(\widehat{z}_i) \leq \sum_{i=1}^p Var(z_i) + i\epsilon\sqrt{n}\rho(L_M).$$

La pérdida de precisión es compensada con una reducción en el tiempo de cómputo de autovalores, el cual es un paso clave en PCA. Por ejemplo, si usamos el algoritmo de Lanczos [1], el cómputo de autovalores y autovectores, puede ser realizado en $\mathcal{O}(n^2/\epsilon^2)$ operaciones aritméticas. Esto si asumimos que el promedio de elementos no nulos de la matriz es $\mathcal{O}(1/\epsilon^2)$.

4.4.2. Optimización

Decimos que una forma cuadrática $Q(x) = x^T A x$ es ODN si la matriz A es ODN. Cada forma cuadrática tiene una forma diagonal $Q(x) = \lambda_1 x_1^2 + \dots + \lambda_n x_n^2$, donde cada λ_i es un

autovalor de A . Es más, la Ley de Inercia de Sylvester nos dice que el número de coeficientes positivos y negativos en una forma diagonal de Q es un invariante de Q .

Sea $\widehat{Q}(x) = x^T \widehat{A}x$ donde \widehat{A} es una matriz obtenida de A por medio del Teorema 1.1. Si $\widehat{Q}(x) = \widehat{\lambda}_1 x_1^2 + \cdots + \widehat{\lambda}_n x_n^2$ donde $\widehat{\lambda}_i$ es un autovalor de \widehat{A} , tenemos que $|Q(x) - \widehat{Q}(x)| \leq \rho(A) + (\Delta_A - \delta_A)/2$, para un ϵ suficientemente pequeño. Así, si estamos interesados en la optimización de $Q(x)$, podemos usar \widehat{A} en lugar de A y obtener una solución con la garantía mencionada más arriba. Según el estado del arte de los algoritmos de optimización de formas cuadráticas, se puede sacar provecho de la *dispersidad* de la matriz considerada, lo cual es especialmente importante en los problemas no convexos.

Los problemas de optimización de formas cuadráticas son NP-difícil, incluso con variables binarias. Es más, un solo autovalor negativo es suficiente para que el problema se vuelva NP-difícil [24]. También está estrechamente relacionado con la optimización del modelo de Ising en mecánica estadística. No sabemos si la optimización de formas cuadráticas ODN es NP-difícil o no, por lo que actualmente es un problema abierto.

CAPÍTULO 5

CONCLUSIONES Y TRABAJOS FUTUROS

Hemos proporcionado un método para construir matrices dispersas que aproximen espectralmente a matrices que son simétricas y ODN. La idea clave que utilizamos para llevar a cabo el trabajo, es la técnica de *dispersión espectral de grafos* introducida por Spielman y Teng en [3]. Específicamente, el algoritmo de dispersión espectral que hemos utilizado es el de Lee y Sun [4], por ser de tiempo casi lineal y porque la cantidad de elementos no nulos que se obtiene es de orden $\mathcal{O}(n/\epsilon^2)$, donde $\epsilon \in (0, 1/120]$, lo cual fue adecuado para nuestros propósitos.

Básicamente el método que hemos proporcionado consiste en lo siguiente:

- Considerar una matriz simétrica y ODN M y construir las matrices A_M, D_M tal como se realizó en la prueba del Teorema 1.1.
- Construir la matriz $L_M = D_M - A_M$, la cual puede ser considerada como la matriz laplaciana de un grafo.
- Utilizar el algoritmo de Lee y Sun para obtener $\widehat{L}_M = \widehat{D}_M - \widehat{A}_M$, que es la matriz laplaciana de un grafo disperso con $\mathcal{O}(n/\epsilon^2)$ aristas.
- A partir de la matriz \widehat{A}_M , obtener la matriz \widehat{M} (tal como está indicado en la demostración del Teorema 1.1)

Para medir la proximidad espectral entre las matrices M y \widehat{M} , hemos utilizado dos teoremas conocidos de la teoría de la perturbación. Por un lado, el Teorema de Davis-Kahan para medir la proximidad de autovectores, y por el otro el Teorema de Perturbación de Weyl, para medir la proximidad entre los autovalores. Además hemos dado una demostración constructiva de nuestro método.

Nuestro método presenta una ventaja con respecto al método desarrollado por Zouzias en [5], en los casos en que la matriz considerada es simétrica y ODN, pues la cantidad de elementos no

nulos obtenidos con el de Zouzias es de $\mathcal{O}(\theta n \log n / \epsilon^2)$, donde θ es un factor extra que depende de la *norma-2* de la matriz considerada.

Si bien las cotas dadas en el Teorema 1.1 son ajustadas, creemos que podrían ser más ajustadas, eliminando el factor \sqrt{n} , pues, según lo observado en los resultados experimentales, si tenemos dos matrices laplacianas $L_1 = D_1 - A_1$ y $L_2 = D_2 - A_2$, entonces la desigualdad $\|A_1 - A_2\| \leq \|L_1 - L_2\|$ siempre es válida. De ser así, el factor \sqrt{n} sería automáticamente eliminado. En el Apéndice 6.2, damos algunos gráficos que ilustran la desigualdad conjeturada.

La importancia de obtener una cota que no dependa del factor \sqrt{n} yace en que eso nos daría mayor libertad de elegir un ϵ más grande, con lo cual obtendríamos una matriz más dispersa.

Bibliografía

- [1] Cornelius Lanczos. «An iteration method for the solution of the eigenvalue problem of linear differential and integral operators». En: *Journal of Research of the National Bureau of Standards* 45.4 (1950), págs. 255-282.
- [2] James W. Demmel. *Applied numerical linear algebra*. Society for Industrial and Applied Mathematics, 1997. ISBN: 0-89871-389-7.
- [3] Daniel A Spielman y Shang-Hua Teng. «Spectral sparsification of graphs». En: *SIAM Journal on Computing* 40.4 (2011), págs. 981-1025. DOI: 10.1137/08074489X.
- [4] Yin Tat Lee y He Sun. «Constructing linear-sized spectral sparsification in almost-linear time». En: *SIAM Journal on Computing* 47.6 (2018), págs. 2315-2336. DOI: 10.1137/16M1061850.
- [5] Anastasios Zouzias. «A matrix hyperbolic cosine algorithm and applications». En: *International Colloquium on Automata, Languages, and Programming*. Springer, 2012, págs. 846-858. DOI: 10.1007/978-3-642-31594-7_71.
- [6] H Golub Gene y F Charles. *Matrix computations. Third Edition*. 1996. ISBN: 0-8018-5413-X.
- [7] G. Strang. *Linear Algebra and Its Applications*. Thomson, Brooks/Cole, 2006. ISBN: 9780030105678. URL: <https://books.google.com.py/books?id=8QVdcRJyL2oC>.
- [8] Rejendra Bhatia. *Matrix analysis*. Springer Science & Business Media, 2013. ISBN: 0-387-94846-5.
- [9] G. Stewart y J. G. Sun. *Matrix perturbation theory*. Academic Press. San Diego, 1990.
- [10] Chandler Davis y W.M. Kahan. «The Rotation of Eigenvectors by a Perturbation. III». En: *SIAM Journal on Numerical Analysis* 7.1 (1970), págs. 1-46. DOI: 10.1137/0707001.
- [11] Yi Yu, Tengyao Wang y Richard J. Samworth. «A useful variant of the Davis-Kahan theorem for statisticians». En: *Biometrika* 102.2 (2015), págs. 315-323. DOI: 10.1093/biomet/asv008.
- [12] Reinhard Diestel. *Graph Theory. Fifth Edition*. Springer-Verlag Berlin Heidelberg, 2017. ISBN: 978-3-662-53621-6.
- [13] Andries E Brouwer y Willem H Haemers. *Spectra of graphs*. Springer Science & Business Media, 2011.

- [14] Ulrike Von Luxburg. «A tutorial on spectral clustering». En: *Statistics and computing* 17.4 (2007), págs. 395-416. DOI: 10.1007/s11222-007-9033-z.
- [15] Daniel A Spielman y Shang-Hua Teng. «Nearly-linear time algorithms for graph partitioning, graph sparsification, and solving linear systems». En: *Proceedings of the thirty-sixth annual ACM symposium on Theory of computing*. 2004, págs. 81-90. DOI: 10.1145/1007352.1007372.
- [16] Fabricio Mendoza-Granada y Marcos Villagra. «A Distributed Algorithm for Spectral Sparsification of Graphs with Applications to Data Clustering». En: *Proceedings of the 18th Cologne-Twente Workshop on Graphs and Combinatorial Optimization (CTW)*. Vol. 5. AIRO Springer Series, 2020. URL: <https://arxiv.org/abs/2003.10612>.
- [17] Fabricio A Mendoza Granada y Marcos Villagra. «Distributed spectral clustering on the coordinator model». En: *Proceeding Series of the Brazilian Society of Computational and Applied Mathematics* 6.2 (2018). URL: <https://proceedings.sbmec.emnuvens.com.br/sbmec/article/view/2307>.
- [18] Fabricio Mendoza-Granada y Marcos Villagra. «Number-On-Forehead Communication Complexity of Data Clustering with Sunflowers». En: *Anais do IV Encontro de Teoria da Computação*. SBC. 2019. URL: <https://doi.org/10.5753/etc.2019.6394>.
- [19] Joshua Batson, Daniel A Spielman y Nikhil Srivastava. «Twice-ramanujan sparsifiers». En: *SIAM Journal on Computing* 41.6 (2012), págs. 1704-1721. DOI: 10.1137/090772873.
- [20] Daniel A Spielman y Nikhil Srivastava. «Graph sparsification by effective resistances». En: *SIAM Journal on Computing* 40.6 (2011), págs. 1913-1926.
- [21] Dimitris Achlioptas. «Database-friendly random projections». En: *Proceedings of the twentieth ACM SIGMOD-SIGACT-SIGART symposium on Principles of database systems*. 2001, págs. 274-281. DOI: 10.1145/375551.375608.
- [22] Zeyuan Allen-Zhu, Zhenyu Liao y Lorenzo Orecchia. «Spectral sparsification and regret minimization beyond matrix multiplicative updates». En: *Proceedings of the forty-seventh annual ACM symposium on Theory of computing*. 2015, págs. 237-245. DOI: 10.1145/2746539.2746610.
- [23] Ian T. Jolliffe. *Principal Components Analysis. Second Edition*. Springer-Verlag New York, NY., 2002. ISBN: 0-387-95442-2.
- [24] P.M. Pardalos y S.A. Vavasis. «Quadratic programming with one negative eigenvalue is NP-hard». En: *Journal of Global optimization* 1.1 (1991), págs. 15-22. DOI: 10.1007/BF00120662.
- [25] Fabricio A Mendoza Granada, Sergio Mercado y Marcos Villagra. «Deterministic Graph Spectral Sparsification». En: *Proceeding Series of the Brazilian Society of Computational and Applied Mathematics* 6.2 (2018). URL: <https://proceedings.sbmec.org.br/sbmec/article/viewFile/2311/2327>.

- [26] Sergio Mercado y Marcos Villagra. «A Study of the Optimality of PCA under Spectral Sparsification». En: *Proceeding Series of the Brazilian Society of Computational and Applied Mathematics* 7.1 (2020). URL: <https://proceedings.sbmac.emnuvens.com.br/sbmac/article/view/3017>.
- [27] Sergio Mercado y Marcos Villagra. «Bounds on the Spectral Sparsification of Symmetric and Off-Diagonal Nonnegative Real Matrices». En: *arXiv:2009.11133* (2020). URL: <https://arxiv.org/abs/2009.11133>.

CAPÍTULO 6

APÉNDICE

6.1. Publicaciones y Difusión

Es oportuno mencionar que en el marco de esta tesis, se ha llevado a cabo otros trabajos preliminares, los cuales son citados en [25, 26]. Dichos trabajos fueron presentados en las conferencias CNMAC 2018 y CNMAC2019, respectivamente. Además, el artículo [27] que contiene el resultado principal de esta tesis, ha sido remitido para una revista.

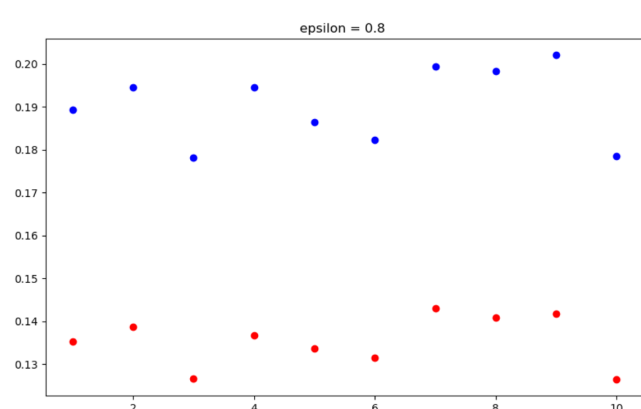
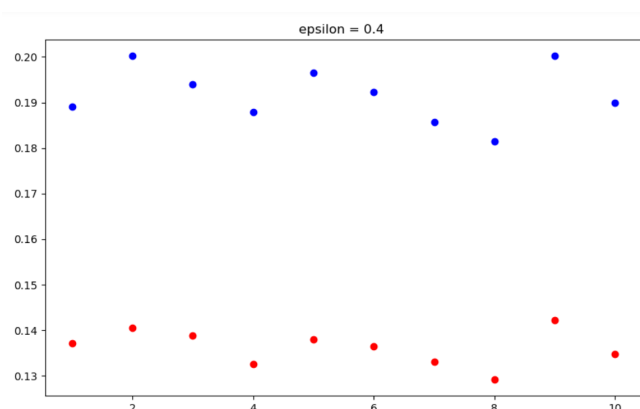
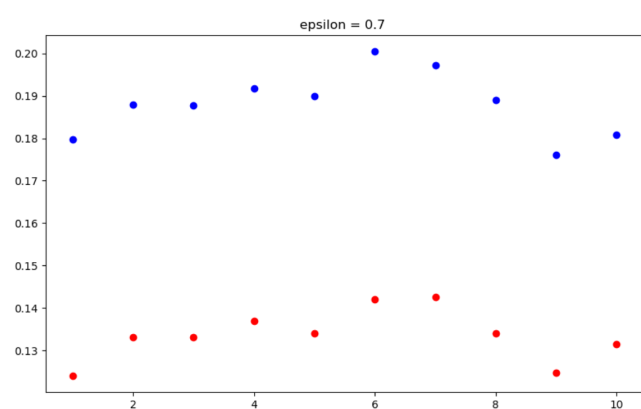
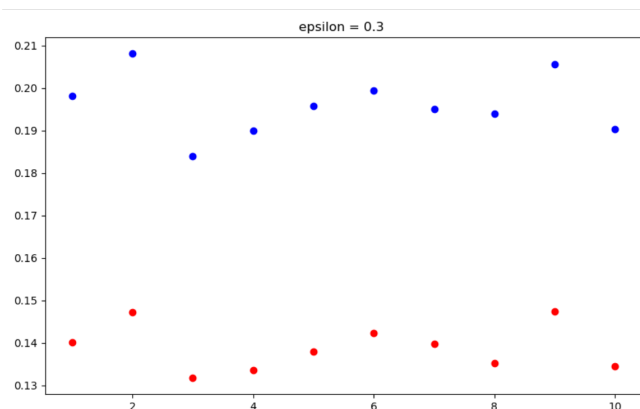
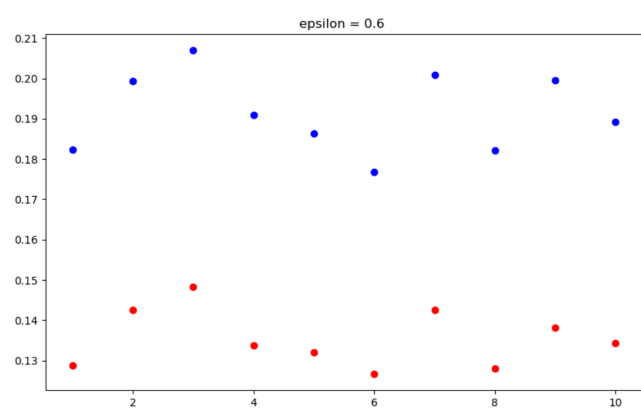
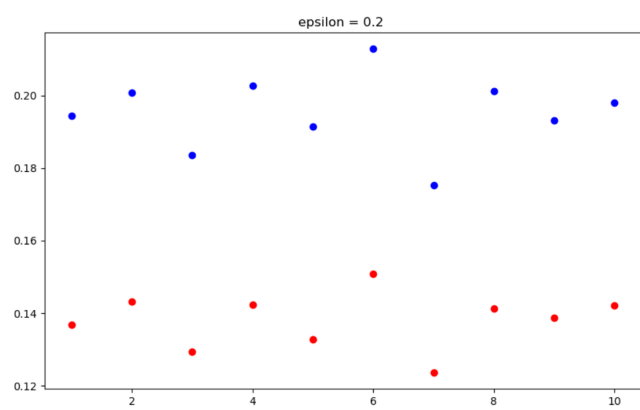
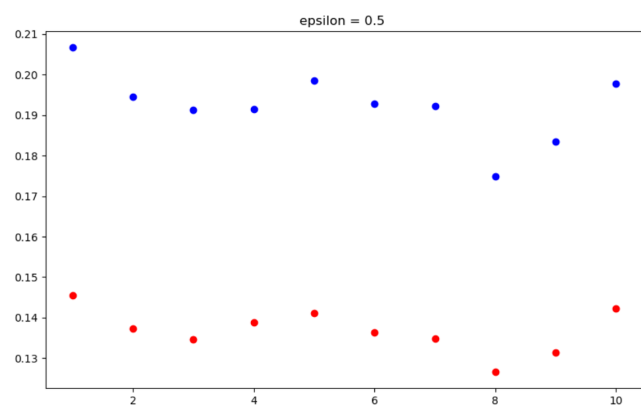
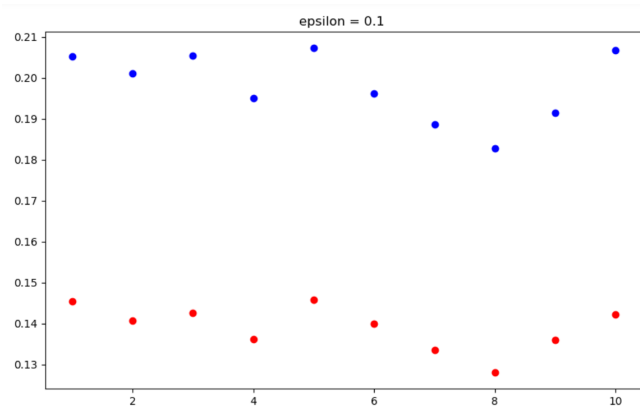
6.2. Conjetura

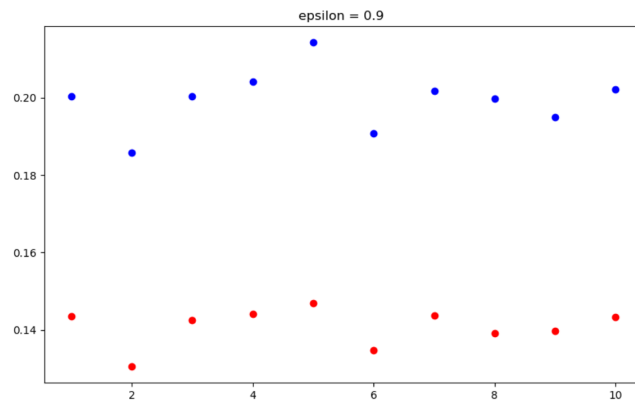
Sea $\widehat{L} = \widehat{D} - \widehat{A}$ un ϵ -dispersador espectral de $L = D - A$, entonces

$$\|A - \widehat{A}\| \leq \|L - \widehat{L}\|.$$

A continuación ilustramos el comportamiento de las normas mencionadas. Construimos un grafo completo G cuyas aristas toman valores $w_{ij} \in (0, 10)$ de manera uniforme. Cada gráfico se compone de la siguiente manera: En el eje horizontal se coloca la cantidad de veces que se aplicó dispersión espectral con el valor del épsilon indicado, y en el eje vertical, los valores de las normas $\|A - \widehat{A}\|$ y $\|L - \widehat{L}\|$, donde en cada caso, los puntos rojos corresponden a $\|A - \widehat{A}\|$ y los azules a $\|L - \widehat{L}\|$.

En cada caso, los puntos rojos corresponden a $\|A - \widehat{A}\|$ y los azules a $\|L - \widehat{L}\|$.





En todos los casos se observa que $\|A - \hat{A}\| \leq \|L - \hat{L}\|$, y la diferencia entre estos valores es muy pequeña.