The University of Hong Kong

FYP18026 Final Year Project

Final Report

Diffusion Process with Mediator on Hypergraph

Author
Yuan Gao 3035234545

Supervisor
Dr. Hubert Chan

April 2019

Abstract

In this project, we study the spectrum of the hypergraph Laplacian and submodular transformation Laplacian. Due to the non-linearity of the hypergraph Laplacian and submodular transformation Laplacian, we resort to mathematical tools other than linear algebra to identify and characterize the spectrum of these operators. It was conjectured that the two operators possess non-trivial eigenvalues (in addition to the known second least eigenvalue)[1][2]. This project presents proofs for the conjectures over both operators.

We also consider the natural optimization problem arising from the study of the spectrum of the hypergraph Laplacian, in particular, the approximation of the maximum eigenvalue of the hypergraph Laplacian, whose existense is proved in this project.

Acknowledgement

I would like to extend my sincere gratitude to my supervisor Dr. Hubert Chan, who has always provided me unfailing supports throughout the project. I would also like to thank Mr. Zhibin Liang and Mr. Bintao Sun for their generous helps. Without them, this project could not have been possible.

Contents

1 Introduction

The spectral theory studies the spectral properties of graphs and their relation to the combinatorial properties of graph. Central to this rich theory, the Cheeger's inequality bounds the expansion or conductance of the graph by the second least eigenvalue of the Laplacian of the graph[3]. More specifically, for a graph $G = (V, E)$, its expansion is defined to be:

$$\phi_G := \min_{S \subset V} \frac{|\partial S|}{\min\{vol(S), vol(\bar{S})\}},$$

where $vol(S)$ is the sum of the degree of vertices in S and ∂S is the cut of S. The Cheeger's inequality[3] is as the following:

$$\frac{\lambda_2}{2} \leq \phi_G \leq \sqrt{2\lambda_2}$$

where λ_2 is the second least eigenvalue of the normalized Laplacian $\mathcal{L}_G$. This inequality can be utilized in the design of various algorithms.

In this project, we will study a mathematical object that generalizes graph, namely hypergraphs. An edge $e \in E$ can have more than two vertices in the hypergraph $H = (V, E)$. Hypergraph can be used to model various kinds of real-life relations and some of its combinatorial properties are both interesting in practice and in theory. Among these combinatorial properties of hypergraphs, expansions and max-cut are closely related to the spectral properties of hypergraph. Recently, the Cheeger's inequality was generalized to the case of hypergraph with the notion of hypergraph Laplacian[1][2]. This hypergraph Laplacian, denoted by L_ω, is defined to be an operator that determines a diffusion process with mediators over hypergraphs. In [1][2], it is shown that this Laplacian operator L_ω possesses a second least eigenvalue λ_2 that satisfy an inequality similar to that of Cheeger's:

$$\frac{\lambda_2}{2} \leq \phi_H \leq \sqrt{2\lambda_2},$$

where ϕ_H is the expansion of hypergraph, whose definition is similar to ϕ_G except that for any $S \subset V$, the cut of S consists of edges that intersects both

S and $V - S$. However, unlike graph Laplacians, which are symmetric linear operators, the hypergraph Laplacians are non-linear and non-smooth, which makes its spectrum hard to characterize. Hypergraph Laplacian is shown to have a trivial eigenvalue and a second least eigenvalue[1][2] and in this project, the conjecture that the hypergraph Laplacian possesses a maximum eigenvalue is proved. Furthermore, the maximum eigenvalue of the hypergraph Laplacian is related to the approximation of the hypergraph max cut, the computation of which is NP-hard. It is thus interesting to consider the approximation of the maximum eigenvalue and its corresponding eigenvectors. One of the goals of the project would therefore be studying possible approximation algorithms of the maximum eigenvalue of the hypergraph Laplacian.

A more generalized notion of Laplacian can be defined with respect to submodular transformations[4]. Hypergraph Laplacians and graph Laplacians can be viewed as special cases of the submodular transformation Laplacian by considering different submodular transformations. This more abstract construction gives rise to some novel Cheeger's inequalities under the same unifying framework. It is thus interesting to consider the spectral properties of this Laplacian. In [4], the generalized Laplacian is shown to possess two eigenpairs, similar to the case of hypergraph Laplacians. One possible research direction along this line would therefore be the study of other eigenpairs of the generalized Laplacians. The conjecture is that the generalized Laplacian also possesses a maximum eigenvalue. Hence a proof for this conjecture is studied in this project.

For the study of the spectral theory of the hypergraph Laplacian, generalized quadratic form plays a central role, both in theory and in practice. The generalized quadratic form for hypergraph is defined as the following:

$$Q(f) := \sum_{e \in E} \omega_e \{\beta_0^e \max_{s,i \in e}(f_s - f_i)^2 + \sum_{j \in e} \beta_j^e [(\max_{s \in e} f_s - f_j)^2 + (\min_{i \in e} f_i - f_j)^2]\},$$

where the parameters β and ω will be introduced in later sections. A basic task that has to be resolved before considering the application of the maximum eigenvalue of the hypergraph Laplacian is to design algorithms that solves optimization problems involving the generalized quadratic form. The challenging issue is the presence of the maximum and minimum operators in the equation. Generally speaking, one would consider semi-definite programming when the objective of the optimization problem is a quadratic form, however, even though in some cases semi-definite programming is indeed applicable to programs involving the generalized quadratic form (at the cost of the size of the program), it seems that there are no obvious ways to use semi-definite programming for the approximation of the maximum eigenvalue of the hypergraph Laplacian. In this project, we consider a simple transformation that models the approximation of the maximum eigenvalue of the hypergraph Laplacian as an optimization problem with polynomial constraints and objective (that is, a transformation that eliminates the maximum operators in the generalized quadratic form) and hence propose to solve the problem using the sum-of-squares method which will be reviewed in later sections.

The remainder of this final report is as follows. First, the report will give more explanations on the notion of hypergraph Laplacian, submodular transformation Laplacian and the related concepts. Then the methodology of manipulating these mathematical concepts will be described, with proofs of several conjectures about the spectrum of these operators explained. The report will proceed to discuss the related issues on the computational aspect of the hypergraph Laplacian, where some necessary preliminaries on optimization theories will be reviewed. It will close with a conclusion, with possible directions of further work emphasized.

2 Problem Definitions and Methodology

2.1 Hypergraph Laplacian

Hypergraph is the generalization of graphs where an edge can have more than two vertices. In this project we consider edge-weighted graphs $H = (V, E, \omega)$ where ω is a consistent weight function, that is,

$$\omega_v = \sum_{e \in E: v \in e} \omega_e, \forall v \in V.$$

We assume that all vertex weights are positive. Given $f \in \mathbb{R}^V$, $f_u, u \in V$ is the coordinate corresponding to $u \in V$. In addition, the space $\mathbb{R}^V$ is endowed with the inner product $\langle \cdot, \cdot \rangle_\omega$ defined as:

$$\langle f, f \rangle_\omega = \langle f, Wf \rangle,$$

where W is the diagonal matrix of the weights of the vertices. This inner product space is called the density space.

In spectral graph theory, the graph Laplacian can be interpreted as an operator that defines a diffusion process over the graph according to some rules ensuring that the process is consistent. Inspired by this, [1] [2] proposed to define a diffusion process over the hypergraph, then the Laplacian L_ω can be defined by:

$$-L_\omega f = \frac{df}{dt},$$

where f is the measure at each vertices of the hypergraph in the diffusion process, i.e. $-L_\omega$ is the derivative of the diffusion process with respect to time. We refer to [1] and [2] for the detailed definitions of the diffusion process and the proof of well-definedness. In this project, we deal with the diffusion process with mediator, which, in essence, is a process in which the measure will go through all vertices over the edge as it flow from the vertex with the highest measure to the ones with the lowest measure (this is a very rough and incomplete description, details can be found in [2]). To describe the participation of the intermediate vertices in the flow of measures, we have the following definitions. For each edge $e \in E$, let $[e] := e \cup \{0\}$, where 0 corresponds to the flow from the vertices with the highest measure to the vertices with lowest measure while the others correspond to the flow through each vertices of the edge. A set of constants β^e is defined over $[e]$ for each $e \in E$, where β^e_j controls the flow through $j \in [e]$. Naturally, $\beta^e_j \geq 0, \forall j \in [e]$ and $\sum_{j \in [e]} \beta^e_j = 1, \forall e \in E$.

A useful tool in the study of the spectral theory of hypergraphs is the Rayleigh quotient:

$$R(f) := \frac{\langle f, L_\omega f \rangle_\omega}{\langle f, f \rangle_\omega}, \text{for any } 0 \neq f \in \mathbb{R}^V.$$

If L_ω is a graph Laplacian operator which is a positive semi-definite linear operator, then by some easy arguments (for example, by using Lagrange multipliers), we have that the maximum and the minimum of the Rayleigh quotient of a graph Laplacian are eigenvalues. In any case, it can be shown that a graph Laplacian operator has $|V|$ real eigenvalues (possibly with multiplicities). It is clear that, for graph Laplacian, $\langle f, L_\omega f\rangle_\omega$ is a quadratic form. However, the situation for hypergraph Laplacians is different. Recall that the generalized quadratic form is:

$$Q(f) := \sum_{e\in E} \omega_e \{\beta_0^e \max_{s,i\in e}(f_s - f_i)^2 + \sum_{j\in e} \beta_j^e [(\max_{s\in e} f_s - f_j)^2 + (\min_{i\in e} f_i - f_j)^2]\}.$$

For $0 \neq f \in \mathbb{Q}^V$, the discrepancy ratio is:

$$D(x) := \frac{Q(f)}{\sum_{u\in V} \omega_u f_u^2}.$$

As a consequence of the definition of the diffusion process, it is shown that $\langle f, L_\omega f\rangle_\omega = Q(f)$ and hence $R(f) = D(f)$ [1] [2]. This enables us to characterize the eigenvalues of the hypergraph Laplacian via Rayleigh quotient.

2.2 Generalized Laplacian of Submodular Transformations

In this section we define a model that subsumes both hypergraphs and graphs. First, we define some basic terminologies. A map f from the power set of V, i.e. $\{0,1\}^V$, into $\mathbb{R}$, is called a submodular function if

$$f(S) + f(T) \geq f(S\cap T) + f(S\cup T), \forall S, T \subset V.$$

A map $F : \{0,1\}^V \to \mathbb{R}^E$ is said to be a submodular transformation if each of its component function $F_e, e \in E$ is submodular [5]. To see that this model generalizes hypergraphs, we can set each component function $F_e, e \in E$ to be the cut value of e as a subset of V [5].

[5] considers the submodular polyhedron $P(F)$ and the base polytope $B(F)$ of

F:

$$P(F) = \{x \in \mathbb{R}^V | \sum_{v \in S} x(v) \leq F(S), \forall S \in V\} \text{ and}$$
$$B(F) = \{x \in P(F) | \sum_{v \in V} x(v) = F(V)\}$$

The lovász extension (or convex closure) $f_e : \mathbb{R}^V \to \mathbb{R}$ of a submodular function F_e is a continuous extension of the submodular function:

$$f_e(x) = \max_{\omega \in B(F_e)} \langle \omega, x \rangle.$$

With this, we define

$$\partial f_e = \text{argmax}_{\omega \in B(F_e)} \langle \omega, x \rangle.$$

This definition is consistent with the usual notation that ∂f_e is the sub-gradient of f_e. The Lovász extension f of a submodular transformation is defined to be the product of Lovász extensions of each component functions of F.

With these definitions, we can proceed to define the Laplacian. In [4], the notion of Laplacian operator $L_F : \mathbb{R}^V \to 2^{\mathbb{R}^V}$ is defined for the submodular transformation F as the following:

$$L_F(x) := \{\sum_{e \in E} \langle \omega_e, x \rangle \omega_e | \omega_e \in \partial f_e(x)\}.$$

We note that this is a set-valued function and it requires a new definition of eigenpairs of the Laplacian, we defer this to later sections.

Similar to the case with hypergraph Laplacian, the Rayleigh quotient of the submodular transformation Laplacian is considered. At first sight, it is not very clear how to define the inner product of a vector x in $\mathbb{R}^V$ with an element (a set) in the image of the Laplacian. However, by noting that L_F in some sense projects x into the span of the sum of the sub-gradient $\partial f_e, e \in E$, [4] gives the following definition on the inner product between x and $L_F(x)$:

$$\langle x, L_F(x) \rangle = \langle x, \omega \rangle, \forall \omega \in L_F(x).$$

This is well-defined because of the following easy fact:

$$\langle x, \omega \rangle = \sum_{e \in E} f_e(x)^2, \forall \omega \in L_F(x).$$

Then naturally the Rayleigh quotient of submodular transformation Laplacian is:

$$R_F(x) := \frac{\langle x, L_F(x) \rangle}{\langle x, x \rangle} = \frac{\|f(x)\|_2^2}{\|x\|_2^2}, \forall x \neq \mathbf{0},$$

where $\|\cdot\|_2$ is the 2-norm over $\mathbb{R}^V$. Graph Laplacian and hypergraph Laplacian are special cases of this generalized Laplacian. All graphs and hypergraphs can be identified with submodular transformations whose Laplacian coincide with the graph and hypergraph Laplacian respectively in a natural way.

2.3 Least Non-trivial Eigenvalue of the Laplacian Operators

It is obvious that the hypergraph Laplacian obtains a trivial eigenpair. More specifically, $\mathbf{1} \in \mathbb{R}^V$ is an eigenvector of the Laplacian with 0 being its corresponding eigenvalue. It is not as trivial to find other eigenpairs of the Laplacian. In [1][2], it is shown that $\lambda_2 := \min_{\mathbf{1} \perp f \in \mathbb{R}^V} R(f)$ is an eigenvalue with any $\mathbf{1} \perp f$ attaining this value being an eigenvector. The proof of λ_2 being an eigenvalue is inspired by the idea that the diffusion process over hypergraph tends to mix the measures across the vertices and reduces the discrepancy ratio (precisely speaking, the diffusion process reduces the discrepancy ratio when the current state is not an eigenvector).

In more details, [1][2] first showed that the diffusion process is smooth with respect to the time t of the diffusion process (note that this does not imply the smoothness of the hypergraph Laplacian over the density space), which implies the existence of the right-hand derivative of the generalized quadratic form of the hypergraph. With careful computations, it is shown that

$$\frac{dR(f)}{dt} = -\frac{1}{\|f\|_\omega^4}(\|f\|_\omega^2 \|Lf\|_\omega^2 - \langle f, Lf \rangle_\omega) \leq 0$$

where the inequality follows from the Cauchy inequality. It follows that as the diffusion progresses for sufficiently small amount of time, the discrepancy ratio will reduce the discrepancy ratio if the initial state is not an eigenvector of the hypergraph Laplacian.

It should be pointed out that the submodular transformation Laplacian L_F is a set-valued operator. Therefore the definition of the spectrum of L_F should be stated as the following[4]:

$$(\lambda, x) \text{ is an eigenpair of the } L_F \text{ if and only if } \lambda x \in L_F(x).$$

It is immediate that $(0, \mathbf{1})$ is a trivial eigenpair of L_F. However, it is not entirely obvious how to generalize the concept of diffusion process over hypergraph to submodular transformations on which the proof of the existence of a non-trivial eigenpair of L_F should be based. Nevertheless, [4] defined a diffusion process similar to the case of hypergraph, by defining the time evolution of any vector $x \in \mathbb{R}^V$ as:

$$\frac{dx}{dt} \in -L_F(x),$$

whose well-definedness requires justifications. Essentially, [4] showed that for any initial state x, there exists some choices L_t such that $\frac{dx}{dt} = -L_t x$ is well-defined for any time $t \geq 0$. With this diffusion process defined, by generalizing the proof that hypergraph Laplacian possesses a non-trivial eigenpair, it can be shown that L_F possesses a non-trivial eigenpair as well.

It was conjectured that $\lambda^* = \max_{f \neq \mathbf{0} \in \mathbb{R}^V} R(f)$ is the maximum eigenvalue of the hypergraph Laplacian, with any $x \in \mathbb{R}^V$ attaining λ^* being an eigenvector. A similar conjecture can be made for the submodular transformation Laplacian. However, if one tries to generalize the proof strategy based on diffusion process, s/he will have to resolve the problem of reversibility of the diffusion process. The intuition is that, given f^* reaching the maximum for Rayleigh quotient, if we were to replicate the arguments used for the least non-trivial eigenvalue, we would be relying on the diffusion process to find vector giving higher value of the discrepancy ratio for the sake of contradiction. In the original argument we proved that if a state in the diffusion process is not an eigenvector of the Laplacian then after a sufficiently small amount of time the diffusion process will reach at a state obtaining a smaller discrepancy ratio, therefore, one would naturally expect for state that is not an eigenvector, by reversing the diffusion process for a sufficiently small amount of time, we will reach at a state giving larger discrepancy ratio. In other words, in the framework of the diffusion process, we would hope that, given some non-eigenvector states, we can find a state just before the given state in the diffusion process having larger discrepancy ratio. However, there are problems in this intuitive argument. We cannot show that any given state is reachable by the diffusion process in the sense that it is possible that starting with any state other than the given state, the diffusion process cannot reach this given state. Hence the diffusion process argument does not work for the maximum eigenvalue. It does not seem possible to bypass this problem in the framework of diffusion process. Therefore, in the next subsection, we will study a new proof strategy for both of the Laplacian operators.

2.4 Sketch Proof of the Existence of Maximum Eigenvalue

I will first illustrate the main frame of the proof using the hypergraph Laplacian and then move on to the more complicated case of submodular transformation Laplacian which requires some results from linear programming on $\epsilon-$perturbation [6]. This subsection will close with a brief comparison between the proposed proof and the proof based on diffusion process.

2.4.1 Maximum Eigenvalue of Hypergraph Laplacian

For the case of hypergraph, consider any $\mathbf{0} \neq f \in \mathbb{R}^V$ and any $\mathbf{0} \neq v \in \mathbb{R}^V$ which defines the trajectory: $\epsilon \mapsto f + \epsilon v$. The basic idea will be considering the change in the value of the Rayleigh quotient along this trajectory. To this end, for any functional T over $\mathbb{R}^V$, we define the left-hand derivative $\partial_v^- T(f)$ of the function $T(f + \epsilon v)$ with respect to ϵ at 0 as the following:

$$\partial_v^- T(f) = \lim_{\epsilon \to 0^-} \frac{T(f^* + \epsilon v) - T(f^*)}{\epsilon}$$

Of course, for some particular T, for example the Rayleigh quotient, the existence of the limit requires justification. Taking for granted the existence of the limit for the moment, let's consider the consequence of this definition. If for some $\mathbf{0} \neq v \in \mathbb{R}^V, \partial_v^- T(f) > 0$, then there exists $\epsilon < 0$ with $|\epsilon|$ sufficiently small such that $R(f + \epsilon v) > R(f)$, that is, we can find a state giving larger discrepancy ratio than $R(f)$ along the trajectory $\epsilon \mapsto \epsilon v$. Recall that diffusion process gives us a direction $\frac{df}{dt} = L_\omega f$ along which the discrepancy ratio decreases strictly for non-eigenvectors. Therefore, it makes sense to compute $\partial_v^- R(f)$ for $v = -L_\omega f$

Let $\lambda^* = \max_{0 \neq f \in \mathbb{R}^V} R(f)$ and f^* be any vector that attains the maximum Rayleigh quotient. Previous discussions show that to prove that (λ^*, f^*) is an eigenpair, we might proceed to compute $\partial_{L_\omega f}^- R(f)$ (and hence give a justification for the existence of this limit). To do this, we first consider $\partial_{-L_\omega}^- Q(f)$, the left-hand derivative of $Q(f^* + \epsilon v)$ with respect to ϵ at 0, where Q is the generalized quadratic form of hypergraph. We give the following definitions and identity to aid the computation.

For any $e \in E$,:

$$c_e^I := \omega_e[\beta_0^e(\max_{s \in e} f_s - \min_{i \in e} f_i) + \sum_{j \in e} \beta_j^e (f_j - \min_{i \in e} f_i)],$$

$$c_e^S := \omega_e[\beta_0^e(\max_{s \in e} f_s - \min_{i \in e} f_i) + \sum_{j \in e} \beta_j^e (\max_{s \in e} f_s - f_j^*)],$$

$$S_e := \text{argmax}_{s \in e} f_s,$$

$$I_e := \text{argmin}_{i \in e} f_i$$

and for each $j \in e$

$$c_j := \sum_{e \in E : j \in e} \beta_j^e \omega_e (\max_{s \in e} f_s + \min_{i \in e} f_i - 2 f_j),$$

and write $r := \frac{df}{dt}$. Certainly as one can see, these definitions are originated from the definition of the diffusion process (which will not be discussed in details here), however, they are presented here only for the sake of computation.

Then we give the following identity:

$$\sum_{e\in E} c_e^S \max_{s\in S_e} r_s - \sum_{e\in E} c_e^I \min_{i\in I_e} r_i - \sum_{j\in V} c_j r_j = -\|r\|_\omega^2.$$

This is identity is proved by considering the maximal densest subset recursively.

With these definitions and identity, we can proceed to show that for any $\mathbf{0} \neq v, \partial_v^- Q(f)$ exists. In particular, we have for $\epsilon < 0$:

$$\begin{aligned} Q(f+\epsilon v) = &\sum_{e\in E} \omega_e \{\beta_0^e [\max_{s\in e} f_s + \epsilon \min_{s'\in S_e} v_{s'} - (\min_{i\in e} f_i + \epsilon \max_{j'\in I_e} v_{i'})]^2 + \\ &\sum_{j\in e} \beta_j^e [(\max_{s\in e} f_s + \epsilon \min_{s'\in S_e} v_{s'} - f_j - \epsilon v_j)^2 + (f_j + \epsilon v_j - \min_{i\in e} f_i - \epsilon \max_{i'\in I_e} v_{i'})^2]\} \end{aligned}$$

therefore for $\epsilon < 0$ and $|\epsilon|$ sufficiently small,

$$\begin{aligned} Q(f+\epsilon v) - Q(f) = &\sum_{e\in E} \omega_e \{2\beta_0^e \epsilon \max_{s,i\in e}(f_s - f_i)(\min_{s'\in S_e} v_{s'} - \max_{i'\in I_e} v_{i'}) + \\ &\beta_0^e \epsilon^2 (\min_{s'\in S_e} v_{s'} - \max_{i'\in I_e} v_{i'})^2 + \sum_{j\in e} \beta_j^e [2\epsilon(\max_{s\in e} f_s - f_j)(\min_{s'\in S_e} v_{s'} - v_j) + \\ &\epsilon^2 (\min_{s'\in S_e} v_{s'} - v_j)^2 + 2\epsilon(f_j - \min_{i\in e} f_i)(v_j - \max_{i'\in I_e} v_{i'}) + \epsilon^2 (v_j - \max_{i'\in I_e} v_{i'})^2]\}. \end{aligned}$$

It follows that:

$\lim_{\epsilon\to 0^-} \frac{Q(f+\epsilon v)-Q(f)}{\epsilon}$

$= \lim_{\epsilon\to 0^-} \sum_{e\in E} \omega_e \{2\beta_0^e \max_{s,i\in e}(f_s - f_i)(\min_{s'\in S_e} v_{s'} - \max_{i'\in I_e} v_{i'}) + \beta_0^e \epsilon (\min_{s'\in S_e} v_{s'} - \max_{i'\in I_e} v_{i'})^2 + \sum_{j\in e} \beta_j^e [2(\max_{s\in e} f_s - f_j)(\min_{s'\in S_e} v_{s'} - v_j) + \epsilon(\min_{s'\in S_e} v_{s'} - v_j)^2 + 2(f_j - \min_{i\in e} f_i)(v_j - \max_{i'\in I_e} v_{i'}) + \epsilon(v_j - \max_{i'\in I_e} v_{i'})^2]\}$

$= 2(\sum_{e\in E} \omega_e \{\beta_0^e \max_{s,i\in e}(f_s - f_i)(\min_{s'\in S_e} v_{s'} - \max_{i'\in I_e} v_{i'}) + \sum_{j\in e} \beta_j^e [(\max_{s\in e} f_s - f_j)(\min_{s'\in S_e} v_{s'} - v_j) + (f_j - \min_{i\in e} f_i)(v_j - \max_{i'\in I_e} v_{i'})]\})$

$= 2(\sum_{e\in E} \{\omega_e [\beta_0^e \max_{s,i\in e}(f_s - f_i) + \sum_{j\in e} \beta_j^e (\max_{s\in e} f_s - f_j)] \min_{s'\in S_e} v_{s'} - \omega_e [\beta_0^e \max_{s,i\in e}(f_s - f_i) + \sum_{j\in e} \beta_j^e (f_j - \min_{i\in e} f_i)] \max_{i'\in I_e} v_{i'} - \sum_{j\in e} \omega_e \beta_j^e (\max_{s\in e} f_s + \min_{i\in e} f_i - 2f_j)\})$

$= 2(\sum_{e\in E}\{\omega_e[\beta_0^e \max_{s,i\in e}(f_s - f_i) + \sum_{j\in e}\beta_j^e(\max_{s\in e} f_s - f_j)] \min_{s'\in S_e} v_{s'} - \omega_e[\beta_0^e \max_{s,i\in e}(f_s - f_i) + \sum_{j\in e}\beta_j^e(f_j - \min_{i\in e} f_i)] \max_{i'\in I_e} v_{i'}\} - \sum_{j\in V}\sum_{e\in E: j\in e}\beta_j^e\omega_e(\max_{s\in e} f_s + \min_{i\in e} f_i - 2f_j)v_j)$

$= 2(\sum_{e\in E} c_e^S \min_{s'\in S_e} v_{s'} - \sum_{e\in E} c_e^I \max_{i'\in I_e} v_{i'} - \sum_{j\in V} c_j v_j)$

Hence the existence of $\partial_v^- Q(f)$ for any $v \in \mathbb{R}^V$. Since both c_e^S and c_e^I are greater or equal to zero, we have that:

$$\partial_v^- Q(f) \leq 2(\sum_{e\in E} c_e^S \max_{s\in S_e} v_s - \sum_{e\in E} c_e^I \min_{i\in I_e} v_i - \sum_{j\in V} c_j v_j)$$

Now we proceed to compute $\partial_{-L_\omega}^- R(f)$. We first note a standard result:

$$\partial_v^- \langle f, f\rangle_\omega = 2\langle f, v\rangle, \forall f, v \in \mathbb{R}^V$$

On the other hand, we have:

$$\partial_{-L_\omega f}^- Q(f) \leq 2(\sum_{e\in E} c_e^S r_S(e) - \sum_{e\in E} c_e^I r_I(e) - \sum_{j\in V} c_j r_j),$$

where $r := -L_\omega f$. It follows that

$$\partial_{-L_\omega f}^- Q(f) \leq -2\|r\|_\omega^2.$$

Therefore,

$$\partial_{-L_\omega f}^- R(f) = \frac{1}{\|f\|_\omega^4}(\|f\|_\omega^2 \partial_{-L_\omega f}^- Q(f) - 2\langle f, L_\omega f\rangle_\omega \langle f, -L_\omega f\rangle_\omega)$$

$$\leq \frac{1}{\|f\|_\omega^4}(\|f\|_\omega^2(-2\|L_\omega f\|_\omega^2) - 2\langle f, L_\omega f\rangle_\omega \langle f, -L_\omega f\rangle_\omega)$$

$$= -\frac{2}{\|f\|_\omega^4}(\|f\|_\omega^2 \|L_\omega f\|_\omega^2 - \langle f, L_\omega f\rangle_\omega^2) \leq 0$$

where the last inequality follows from the Cauchy Inequality, with the equality holds if and only if $L_\omega f \in span(f)$. Therefore, if $L_\omega \notin span(f)$, then $\partial^-_{-L_\omega f}(f) < 0$.

What remains is to formalize the arguments provided at the start of this section. To complete the argument, assume that $L_\omega \notin span(f^*)$, then for the sake of contradiction, it suffices to show that there exists $\epsilon < 0$ such that $R(f + \epsilon(-L_\omega f)) > R(f)$. Assume that for all $\epsilon < 0, R(f + \epsilon(-L_\omega f))) \leq R(f)$, then $\partial^-_{-L_\omega f} R(f) = \lim_{\epsilon \to 0^-} \frac{R(f+\epsilon(-L_\omega f))-R(f)}{\epsilon} \geq 0$, a contradiction. Hence we have that (λ^*, f^*) is an eigenpair of L_ω.

It should be pointed out that in this section $\partial^-_v R(f)$ is computed explicitly by the definition of left-hand side derivative, that is, an explicit formula of $Q(f + \epsilon v)$ for $\epsilon < 0$ is found. In the following section, we will proceed to study the maximum eigenvalue of submodular transformation Laplacian. However, computing the derivatives for Rayleigh quotient of submodular transformation Laplacian seems to be infeasible. This problem is solved with the ϵ-perturbation method in linear programming [6] (to eliminate confusion with notations, please note that in the following subsection for submodular transformation, ϵ and f will not carry the same meaning as in the discussion for hypergraph).

2.4.2 Maximum Eigenvalue of Submodular Transformation Laplacian

Recall the definition of f as the Lovász extension of a submodular transformation F and L_F as the Laplacian of the submodular transformation F. Also, recall that for any $x \in \mathbb{R}^V, R_F(x) = \frac{\|f(x)\|_2^2}{\|x\|_2^2}$. We note briefly in the preceding subsection that the main difficulty for the discussion of submodular transformation is that $\|f(x)\|_2^2$ is difficult to handle. Precisely speaking, it is not very clear how to compute the following expression explicitly for $\delta \in \mathbb{R}$, with $|\delta|$ sufficiently small and any $v \in \mathbb{R}^V$:

$$\|f(x + \delta v)\|_2^2 - \|f(x)\|_2^2 = \sum_{e \in E} f_e(x + \delta v)^2 - f_e(x)^2.$$

This subsection will therefore focus on $f_e(x+\delta v)^2 - f_e(x)^2$ for some $e \in E$. A natural guess is that there exists some operators L_x such that for any $\delta \in \mathbb{R}$, with $|\delta|$ sufficiently small and any $v \in \mathbb{R}^V$, $f_e(x)^2 = \langle x, L_x x\rangle$ and $f_e(x+\delta v)^2 = \langle x+\delta v, L_x(x+\delta v)\rangle$. To this end, I rewrite the definition of $f_e(x)$ as a linear program such that $f_e(x)$ equals to:

$$\begin{array}{ll} \text{Maximize} & \omega^T x \\ \text{s.t.} & \mathbf{1_S}^T \omega \leq F_e(S), \forall S \subset V \\ & \mathbf{1_V}^T \omega = F_e(V) \end{array}$$

Suppose we order all elements except for $\emptyset$ and V in the power set of V and mark them as $S_1, S_2, \cdots, S_{2^{|V|}-2}$, then equivalently, $f_e(x)$ can be defined by the dual problem of the above linear program as $P_e(x)$:

$$\begin{array}{ll} \text{Minimize} & F_e^T y \\ \text{s.t.} & Ay = x \\ & y \geq 0 \end{array}$$

where

$$A = \begin{pmatrix} \mathbf{1_{S_1}} & \cdots & \mathbf{1_{S_{2^{|V|}-2}}} & \mathbf{1_V} & -\mathbf{1_V} \end{pmatrix} \in \mathbb{Z}^{|V| \times 2^{|V|}}$$

is of full row rank and

$$F_e = \begin{pmatrix} F_e(S_1) & \cdots & F_e(S_{2^{|V|}-2}) & F_e(V) & -F_e(V) \end{pmatrix}^T \in \mathbb{R}^{2^{|V|}}.$$

Then naturally, we have transformed the discussion on $f_e(x)$ and $f_e(x+\delta v)$ into the discussion on the sensitivity analysis of the linear program $(P_e(x))$.

Suppose that B is an optimal basis of $(P_e(x))$, then the goal is to justify that B is also an optimal basis of $(P_e(x+\delta v))$ for $|\delta|$ sufficiently small. At first sight, this is rather trivial due to the principle of sensitivity analysis. However, one will run into problems because of the possible degeneracy of $(P_e(x))$ and will notice that B might not be optimal for $(P_e(x+\delta v))$, no matter how small $|\delta|$ is. Then the main difficulty here is to avoid degeneracy in the discussion.

Here we state a result on degeneracy avoidance in linear programming using $\epsilon - perturbation$ method by [6]. However, the theorem we state here is slightly more general. The difference is that in [6] the studied linear program is of integer coefficients. The reason for such restriction is that [6] attempts to show that degeneracy can be avoided in polynomial time for any linear program with integer coefficients. However, since the discussion here is only theoretical, we can relax this restriction. The proofs for the following results in our settings are almost identical to those in [6] and only some simple alternations to their arguments are needed, therefore we omit the proofs here.

Consider the linear programming problem (P):

$$\begin{aligned} \text{Maximize} \quad & c^T x \\ \text{s.t.} \quad & Ax = b \\ & x \geq 0 \end{aligned}$$

where $A \in \mathbb{Z}^{m \times n}, b \in \mathbb{R}^m$ and $c \in \mathbb{R}^n$. Assume that A is of full row rank. Let M be the maximum absolute value of any entry of A.

Consider the linear programming problem $(P(\epsilon))$, for some $0 \leq \epsilon \in \mathbb{R}$:

$$\begin{aligned} \text{Maximize} \quad & c^T x \\ \text{s.t.} \quad & Ax = b + \boldsymbol{\epsilon} \\ & x \geq -\epsilon K \mathbf{1} \end{aligned}$$

or equivalently:

$$\begin{aligned} \text{Maximize} \quad & c^T x \\ \text{s.t.} \quad & Ax = b + \epsilon K \mathbf{1} + \boldsymbol{\epsilon} \\ & x \geq 0 \end{aligned}$$

where $\boldsymbol{\epsilon} = (\epsilon, \epsilon^2, \cdots, \epsilon^m)^T$ and $K = (m!)^2 M^{2m-1}$.

With these, we can state the result from [6]. We note that [6] requires that $A \in \mathbb{Z}^{m \times n}, b \in \mathbb{Z}^m$ and $c \in \mathbb{Z}^n$, while we only need $A \in \mathbb{Z}^{m \times n}$. The consequence is that [6] can give an explicit and polynomial-computable bound on ϵ in the following theorem while we can only have it as “sufficiently small”.

Theorem. *For any sufficiently small $\epsilon > 0$, $(P(\epsilon))$ is non-degenerate and* *have the same status in terms of feasibility and boundedness with (P). Every optimal basis of $(P(\epsilon))$ is an optimal basis of (P).*

Because of this theorem, we have $(P_e(x,\epsilon))$ as the following:

$$\begin{array}{ll} \text{Minimize} & F_e^T y \\ \text{s.t.} & Ay = x + \epsilon K \mathbf{1_V} + \boldsymbol{\epsilon} \\ & y \geq 0 \end{array}$$

where $K = (2^{|V|}!)^2$. $(P_e(x + \delta v, \epsilon))$ is:

$$\begin{array}{ll} \text{Minimize} & F_e^T y \\ \text{s.t.} & Ay = x + \epsilon K \mathbf{1_V} + \boldsymbol{\epsilon} + \delta v \\ & y \geq 0 \end{array}$$

By the preceding theorem, there exists sufficiently small $\epsilon > 0$ such that the linear programs $(P_e(x,\epsilon))$ and $(P_e(x+\delta v,\epsilon))$ defined above, are non-degenerate and have the same status in terms of feasibility and boundedness with $(P_e(x))$ and $(P_e(x+\delta v))$ respectively and every optimal basis of $(P_e(x,\epsilon))$ is an optimal basis of $(P_e(x))$ and every optimal basis of $(P_e(x+\delta v,\epsilon))$ is an optimal basis of $(P_e(x+\delta v))$. Suppose $B_{e,x}$ is an optimal basis for $(P_e(x,\epsilon))$ and hence an optimal basis for $(P_e(x))$, we have that $f_e(x) = \langle F_e, B_{e,x}^{-1}(x)\rangle$.

Now, consider a situation of sensitivity analysis in which the optimal basis $B_{e,x}$ is obtained for the linear program $(P_e(x,\epsilon))$. Then we change the RHS of the constraints of the linear program $(P_e(x,\epsilon))$ from $(x + \epsilon K\mathbf{1_V} + \boldsymbol{\epsilon})$ into $(x+\epsilon K\mathbf{1_V}+\boldsymbol{\epsilon}+\delta v)$. In a simplex-type algorithm, such a change does not affect the optimality of the basis $B_{e,x}$, hence the only concern is the feasibility of the basis $B_{e,x}$. According to the preceding theorem, the linear program $(P_e(x,\epsilon))$ is non-degenerate, that is, $B^{-1}(x + \epsilon K\mathbf{1_V} + \boldsymbol{\epsilon}) > 0$. Then clearly for δ with $|\delta|$ sufficiently small, we have that $B^{-1}(x + \epsilon K\mathbf{1_V} + \boldsymbol{\epsilon} + \delta v) > 0$. Therefore $B_{x,\epsilon}$ is a feasible basis for the linear program $(P_e(x+\delta v,\epsilon))$. Hence $B_{x,\epsilon}$ is an optimal basis for the linear program $(P_e(x+\delta v,\epsilon))$. By the preceding theorem, $B_{e,x}$ is therefore an optimal basis for the linear program $(P_e(x+\delta v))$. Then $f_e(x) = \langle F_e, B_{e,x}^{-1}x\rangle$ and $f_e(x+\delta v) = \langle F_e, B_{e,x}^{-1}(x+\delta v)\rangle$. Let $\omega_{e,x} = B_{e,x}^{-T}F_e$, we have that $f_e(x) = \langle \omega_{e,x}, x\rangle$ and $f_e(x+\delta v) = \langle \omega_{e,x}, x+\delta v\rangle$.

Let $W_x \in \mathbb{R}^{V \times E}$ be the matrix whose columns are $\omega_{e,x}$ as defined above. Let $L_x = W_x W_x^T \in \mathbb{R}^{V \times V}$. Then for each $x \in \mathbb{R}^V, R_F(x) = \frac{\sum_{e \in E} f_e(x)^2}{\|x\|^2} = \frac{\langle x, L_x x \rangle}{\|x\|^2}$ and for δ with $|\delta|$ sufficiently small and any $v \in \mathbb{R}^V, R_F(x + \delta v) = \frac{\sum_{e \in E} f_e(x+\delta v)^2}{\|x+\delta v\|^2} = \frac{\langle x+\delta v, L_x(x+\delta v) \rangle}{\|x+\delta v\|^2}$. Therefore, if we are only interested in the value of R_F, we can just consider $L_x(x)$ instead of $L_F(x)$. Note that L_x is symmetric.

Consider $L_F^* : \mathbb{R}^V \to \mathbb{R}^V$, defined by $L_F^*(x) = L_x(x)$ and $Q_F^*(x) = \langle x, L_F^*(x) \rangle$. The following facts are easy to prove:

Fact 1. *For any* $x, v \in \mathbb{R}^V, \lim_{\delta \to 0} \frac{Q_F^*(x+\delta v) - Q_F^*(x)}{\delta} = 2\|L_x x\|^2$.

Fact 2. *For any* $x \in \mathbb{R}^V$, $\frac{d(Q_F^*(x+\delta L_F^*(x)))}{d\delta}|_{\delta=0} = 2\|L_F^* x\|^2$. *If we let* $R_F^*(x) = \frac{Q_F^*(x)}{\|x\|^2}$, *then* $\frac{dR_F^*(x+\delta L_F^* x)}{d\delta}|_{\delta=0} \geq 0$, *with the equality holds if and only if* $L_F^*(x) = \lambda x$, *for some* $\lambda \in \mathbb{R}$.

Fact 2 is a consequence of fact 1. These facts enables us to prove the existence of the maximum eigenvalue of the submodular transformation Laplacian. Consider $\lambda^* = \max_{0 \neq x \in \mathbb{R}^V} R_F^*(x)$. Note that $\forall a \in \mathbb{R}, L_F^*(ax) = aL_F^*(x)$. Then λ^* is well defined, i.e. there exists $0 \neq x^* \in \mathbb{R}^V$ that attains λ^*. Just as in the case of hypergraph Laplacian, fact 2 implies that x^* is an eigenvector of L_F^*, with λ^* being the eigenvalue. Therefore, (λ^*, x^*) is an eigenpair of L_F.

In conclusion, submodular transformation Laplacian does possess a maximum eigenpair.

2.4.3 Comparison with the Diffusion Process Argument

In this subsection, I will focus on the case of hypergraph Laplacian and adopt the notations for it. As mentioned in the above subsections, the main difficulty of applying the diffusion process argument to the maximum eigenvalues of the hypergraph Laplacian is the reversibility of the diffusion process. Intuitively speaking, if $\mathbf{0} \neq f \in \mathbb{R}^V$ is not an eigenvector of the hypergraph Laplacian, then reversing the diffusion process will give a state that has larger discrepancy ratio (value of the Rayleigh quotient). Such argument could lead to a proof that (λ^*, f^*) is an eigenpair of hypergraph Laplacian, if the diffusion process is indeed (locally) reversible. However, it is possible that f^* can only be an initial state of the diffusion process, which disputes the reversibility of the diffusion.

In comparison, the argument presented in this project bypasses the diffusion process and study the hypergraph Laplacian solely as an operator over $\mathbb{R}^V$. However, one might notice that the study of $\partial^-_{-L_\omega f} R(f)$ is essentially "simulating" the reverse of diffusion process. Such an argument gives more freedom in the manipulation of L than the diffusion process argument since we are not confined to the time evolution of the diffusion process, instead we can now look at the "evolution" in any direction.

2.5 Complexity Overhead

It has already been established that the Laplacians do possess a maximum eigenvalue. However, to really make these results useful, it is necessary to deal with the computational issue arising around the spectrum of the Laplacians. In this section, two major computational issues about the spectral properties of the Laplacians will be discussed. The first is on the computation of the generalized quadratic form, which is center to all the computational problems on the spectral properties. The second is about the approximation of the maximum eigenvalue of the the Laplacian. For simplicity, this section will only consider hypergraph Laplacians.

2.5.1 Spectral Sparsifier of Hypergraph

The size of the edge set of hypergraphs can be of order 2^n. Therefore, in the worst-case analysis, the evaluation of the hypergraph Laplacian and Rayleigh quotient (or the generalized quadratic form) is computationally intractable. It is therefore important to consider the approximation of the generalized quadratic form of a hypergraph with the generalized quadratic form of another hypergraph with a polynomial-sized edge set. Formally, a subgraph H' of the hypergraph H is an ϵ-spectral sparsifier of H if:

$$(1-\epsilon)Q_{H'}(f) \leq Q_H(f) \leq (1+\epsilon)Q_{H'}(f)$$

for every $f \in \mathbb{R}^V$. To make the sparsifier useful, it is important that the size of the edge set of H' is small. A randomized algorithm is proposed in [5] that can generate a sparsifier with $O(n^3logn/\epsilon^2)$ edges with high probability for any hypergraph, which makes the computation of various problems that involves the quadratic form of hypergraphs feasible. This algorithm is a simple (the analysis of the algorithm is certainly not simple) sampling algorithm that samples each edge of H with a probability. This sparsifier is of order $O(n^3logn/\epsilon^2)$ [5]. However, a hypergraph with edges that contains no more than 3 vertices has no more than $O(n^3)$ edges, which means that the proposed sparsifier will not sparsify this hypergraph at all. Hence it would be interesting to consider sparsifiers that might give better theoretical guarantees. In any case, it was established that the Rayleigh quotient of the hypergraph Laplacian is at least computable in polynomial time.

2.5.2 Approximation of the Maximum Eigenvalue

The eigenpair (λ^*, f^*) is closely related to the max-cut of hypergraphs, the computation of which is NP-hard. Hence knowing the maximum eigenvalue and its corresponding eigenvector might be useful for the approximation of the max cut, which requires an approximation algorithm for the maximum of the Rayleigh quotient of hypergraphs. Therefore, one of the objectives of this project is on the approximation of this eigenpair. For simplicity, in this section, we only consider the case where the diffusion process has no mediators, i.e.

$$\beta_0^e = 1 \text{ and } \beta_j^e = 0, \forall j \in e, \forall e \in E.$$

In [1], a semi-definite programming (SDP) relaxation and a rounding algorithm is proposed for the approximation of the procedural minimizers of the discrepancy ratio of hypergraphs, where the procedural minimizers are defined recursively as the following:

Given $\{f_i\}_{i\in[k-1]},$ for some $k \geq 2,$ a set of orthonormal vectors, define:
$\gamma_k = \min\{D(f) : 0 \neq f \perp_\omega \{f_i : i \in [k-1]\}\},$ and
f_k to be the vector orthogonal to $\{f_i\}_{[k-1]}$ that attains γ_k.

Given orthogonal vectors $f_1, \cdots, f_{k-1}$, the semi-definite program proposed in [1] is:

$$\begin{aligned} &\text{Minimize} && \textstyle\sum_{e \in E} \omega_e \max_{u,v \in e} \|g_u - g_v\|^2 \\ &\text{s.t.} && \textstyle\sum_{v \in V} \omega_v |g_v\|^2 = 1 \\ &&& \textstyle\sum_{v \in V} \omega_v f_i(v) g_v = 0, \qquad \forall i \in [k-1]. \end{aligned}$$

To obtain the f_k using vectors $g_v, v \in V$, it suffices to run a Gaussian rounding algorithm. From a purely theoretical perspective, if we are working with graphs, instead of hypergraphs, then this procedure can be applied to calculate all the eigenpairs of the graph Laplacian, which obviously includes the maximum eigenpair. However, the situation for hypergraph Laplacian is tricky, perhaps due to the fact that hypergraph Laplacian is only piece-wise linear. The first two eigenpairs of the hypergraph Laplacian are in fact the first two procedural minimizers of the discrepancy ratio. However, [1] gives a hypergraph whose procedural minimizer does not produce a unique γ_3. Therefore, we cannot use the procedural minimizers to compute the maximum eigenpair of the hypergraph Laplacian.

One possible idea for the approximation of the maximum eigenvalue of the hypergraph Laplacian might be formulating a similar SDP relaxation and rounding algorithm. However, even though semi-definite program can resolve maximum operators in minimization problems, maximum operators in maximization problem will cause problem. We do not know if there is any simple solution for this issue within the framework of semi-definite programming. For this reason, we have to consider some other optimization framework. In particular, in this project, we consider the Sum-of-Squares method. First, we will review some of the very basic definitions and properties in Sum-of-Squares method. For a more complete introduction to Sum-of-Squares method, please refer to [7]. The definitions and theorems regarding the Sum-of-Squares method, without further explanations, are all cited from [7].

Theorem. *Let $P_1, \cdots, P_m \in \mathbb{R}[x]$, then the system of polynomial equations $\mathcal{E} = \{P_1 = 0, \cdots, P_m = 0\}$ has no solution over $\mathbb{R}^n$ if and only if, $\exists Q_1, \cdots, Q_m \in$*

$\mathbb{R}[x]$ *and* $S \in \mathbb{R}[x]$ *a sum of squares of polynomial s.t.*

$$-1 = S + \sum Q_i \cdot P_i$$

This theorem a is corollary of the Positivstellensatz and $S, Q_1, \cdots, Q_m$ is called a Sum-of-Squares refutation of $\mathcal{E}$. It is a degree-ℓ refutation if $\max \deg Q_i P_i \leq \ell$.

We also have a dual object of the Sum-of-Square refutation.

Define. *A degree-d pseudo-distribution* μ *is a finitely supported function over* $\mathbb{R}^n$ *s.t.* $\tilde{\mathbb{E}}_\mu 1 = 1$ *and* $\tilde{\mathbb{E}}_\mu f^2 \geq 0, \forall f \in \mathbb{R}[x]_{\leq d/2}$*, where*

$$\tilde{\mathbb{E}}_\mu := \sum_{x \in \ supp(\mu)} f(x)\mu(x)$$

Given $\mathcal{E}$ *as defined in the preceding theorem, we say* μ *satisfies* $\mathcal{E}$*, denoted by* $\mu \models \mathcal{E}$*, if* $\forall P_i, \tilde{\mathbb{E}}_\mu Q P_i = 0, \forall Q \in \mathbb{R}[x]_{\leq (d - \deg P_i)}$.

The following theorem shows us why a constant degree pseudo-distribution that satisfies some system of polynomial equations is polynomial-time computable.

Theorem. *For a finitely supported functional* μ *with* $\tilde{\mathbb{E}}_\mu 1 = 1$*, the following two statements are equivalent:*

1) μ *is a degree-d pseudo-distribution.*

2) $\tilde{\mathbb{E}}_\mu ((1, x)^{\otimes d/2})((1, x)^{\otimes d/2})^T$ *is positive semi-definite,*

where $\tilde{\mathbb{E}}_\mu ((1, x)^{\otimes d/2})((1, x)^{\otimes d/2})^T$ is called the formal degree-d moment matrix of μ. Also, note that in the condition that a constant degree pseudo-distribution satisfies a system of polynomial equations, we are safe to only consider $Q \in \mathbb{R}[x]$ that are monomials, which implies that this condition imposes a polynomial-size set of linear constraints in terms of the formal degree-d moment matrix of μ. Together with the above theorem, it is easy to see that the condition that a finitely supported functional is a constant degree pseudo-distribution that satisfies a set of polynomial equations is equivalent to the condition that the formal degree-d moment matrix of μ lies in the intersection of some linear subspace and the positive semi-definite convex cone in $\mathbb{R}^{d\times d}$. Therefore, we can use semi-definite programming to find such a constant degree pseudo-distribution.

With these definition, it is very easy to state the degree-ℓ Sum-of-Squares algorithm. Suppose that we are given $P_0, \cdots, P_m$ and the objective is to minimize (or maximize) $P_0(x)$ under the constraints that x satisfies the polynomial equations $P_1 = \cdots = P_m = 0$. We simply compute the smallest (or largest, respectively) $\phi^{(\ell)}$ s.t that there exists degree-ℓ pseudo-distribution $\mu \models \{P_0 = \phi^{(\ell)}, P_1 = \cdots = P_m = 0\}$.

The final theorem gives the duality between Sum-of-Squares refutations and pseudo-distributions. Due to the corollary of Positivstellensatz, the following theorem implies that by considering pseudo-distributions of some degree (possibly exponential in the size of the problem or worse), we can accurately extremize a polynomial over polynomial constraints that are explicitly bounded (that is, there exists a linear combination of the polynomial constraints that equals $\|x\| \leq M$ for some $M > 0$).

Theorem. *Suppose that $\mathcal{E}$ is explicitly bounded. Then one and only one of the followings holds:*

1) there exists a degree-ℓ Sum-of-Squares proof refuting $\mathcal{E}$.

2) there exists a degree-ℓ pseudo-distribution $\mu \models \mathcal{E}$.

The above discussion implies that Sum-of-Squares method, at least in theory, can always verify whether a system of polynomial equations has solution or not. Although since we want the Sum-of-Square algorithm to run in polynomial time and hence have to require the degree of the Sum-of-Square algorithm to be a small constant (perhaps 4 or 6), this method at least points a direction for us along which an approximation algorithm with provable guarantees (although very difficult) for our optimization problems might be found. To this end, we will try to formulate the approximation of the maximum eigenpair of the hypergraph as a polynomial optimization problem first. Similar to the semi-definite programming formulation of the procedural minimizers, the approximation of the maximum eigenpair can be formulated as the following:

$$\begin{aligned} &\text{Minimize} \quad 2 - \textstyle\sum_{e \in E} \omega_e \max_{u,v \in e} (f_u - f_v)^2 \\ &\text{s.t.} \qquad\qquad \textstyle\sum_{v \in V} \omega_v f_v^2 = 1. \end{aligned}$$

As remarked at the beginning of this subsection, semi-definite programming can not solve this problem because we are minimizing an objective containing negative maximization operators. So now the main issue is to eliminate the maximum operator in the objective. This can be resolved as the following:

$$\begin{aligned} &\text{Minimize} \quad 2 - \textstyle\sum_{e \in E} \omega_e \sum_{u,v \in e} (\eta^e_{u,v})^2 (f_u - f_v)^2 \\ &\text{s.t.} \qquad\qquad \textstyle\sum_{v \in V} \omega_v f_v^2 = 1. \\ &\qquad\qquad\qquad \textstyle\sum_{u,v \in e} (\eta^e_{u,v})^2 = 1, \qquad\qquad \forall e \in E. \end{aligned}$$

Clearly the above optimization problem consists of only polynomial constraints and polynomial objective (which is of degree 4). To see that the two optimization problems are equivalent, one can note that for any $e \in E$ and $\eta^e_{u,v}$ that satisfies the constraint $\sum_{u,v \in e} (\eta^e_{u,v})^2 = 1$, we have that $\omega_e \sum_{u,v \in e} (\eta^e_{u,v})^2 (f_u - f_v)^2 \leq \omega_e \max_{u,v \in e} (f_u - f_v)^2$. In addition, we are also safe to put the constraints $\eta^e_{u,v} \geq 0, \forall e \in E, u, v \in e$ in, and use $\eta^e_{u,v}$ directly in other constraints and the objective. There exists a slightly more general version of the Sum-of-Squares method that deals with inequality polynomial constraints. However, even though doing so would reduce the degree of the objective to 3, we choose not to do so since this will introduce too many addition constraints.

Now let $P = 2 - \sum_{e \in E} \omega_e \sum_{u,v \in e} (\eta^e_{u,v})^2 (f_u - f_v)^2$ and $P_1 = \sum_{v \in V} \omega_v f_v^2 - 1$ and $P_e = \sum_{u,v \in e} (\eta^e_{u,v})^2 - 1$ for any $e \in E$. We define

$$\mathcal{A} := \{P - \gamma = 0, P_1 = 0, P_e = 0, \forall e \in E\}.$$

Then for any γ, we can in polynomial time determine that whether or not there exists a degree-4 pseudo-distribution $\mu \models \mathcal{A}$.

Recall the Sum-of-Square algorithm, then one of the issues would be how to find the smallest γ such that there exists a degree-4 pseudo-distribution that satisfies it (we will call this smallest value $\gamma^{(4)}$). This might not be a trivial problem for an arbitrary polynomial optimization problem. If we simply try to put γ (as a variable) into the semi-definite program that solves the existence of the pseudo-distribution, then we might notice that the constraints that requires the pseudo-distribution to satisfy the polynomial system might not be linear in the variables (formal moment matrix of pseudo-distribution and γ) anymore, meaning that this is no longer a semi-definite program. However, finding the smallest γ is relatively easy for our problem. Note that $\mathbb{S}^{n-1}$ the unit sphere in $\mathbb{R}^n$ is compact. Since the Rayleigh quotient of the hypergraph Laplacian R is continuous, then the image of the unit sphere under the Rayleigh quotient $R(\mathbb{S}^{n-1})$ is compact as well, in particular it is bounded, where its minimum is 0 and the maximum is less than or equal to 2. In addition, since $\mathbb{S}^{n-1}$ is connected, its image $R(\mathbb{S}^{n-1})$ is connected as well, that means R attains all the values between the minimum and the maximum over $\mathbb{S}^{n-1}$. On the other hand, we note that $R(\mathbb{S}^{n-1})$ is contained in the image of $\{P_1 = 0, P_e = 0, \forall e \in E\}$ under $(2 - P_0)$. Also, since $0 \leq P_0(x, \eta) \leq R(x), \forall x, \eta$, we see that the image of $\{P_1 = 0, P_e = 0, \forall e \in E\}$ under $2 - P_0$ is the same as $R(\mathbb{S}^{n-1})$. As a result, over $\{P_1 = 0, P_e = 0, \forall e \in E\}, 2 - P_0$ attains all the values between the minimum 0 and maximum (less than or equal to 2). This fact suggests that we can consider using a binary search between 0 and 2 to search for the smallest value of γ, i.e. $\gamma^{(4)}$.

Now, what about the approximation quality? Recall that having a degree-4 pseudo-distribution μ such that $\mu \models \mathcal{A}$ is equivalent to that there does not exist a degree-4 Sum-of-Squares refutation for $\mathcal{A}$. If $\lambda^* = \max_{0 \neq x \in \mathbb{R}^V}$, then clearly the polynomial system $\{P_0 = 2 - \lambda^*, P_1 = 0, P_e = 0, \forall e \in E\}$ can be satisfied and does not have Sum-of-Squares refutation with any degree, and in particular it does not have a degree-4 Sum-of-Squares refutation, implying that there must exists a degree-4 pseudo-distribution that satisfies it. Therefore, $\gamma^{(4)}$ should be less than $2 - \lambda^*$. However, the real question is, with the Sum-of-Square method certifying $\gamma^{(4)}$, how can one get a vector $x \in \mathbb{R}^V$ such that $2 - R(x)$ gives a good approximation of $2 - \lambda^*$, or, how can one get a vector $x \in \mathbb{R}^V$ such that $2 - R(x)$ is close to $\gamma^{(4)}$ which is a value that degree-4 pseudo-distribution deems reachable? The common framework is to make use of the pseudo-distribution that is computed using the semi-definite program as the following [7]. Notice that $\Sigma = \tilde{\mathbb{E}}_\mu (x - v)(x - v)^T$, where $v = \tilde{\mathbb{E}}_\mu x$, the formal covariance of the pseudo-distribution is also a positive semi-definite matrix. Then the Gaussian distribution computed as $(\tilde{\mathbb{E}}_\mu x + \Sigma^{1/2}\mathcal{N})$, where $\mathcal{N}$ is a normal Gaussian distribution, has it that its first two moments matches that of the pseudo-distribution μ's (this is the reason why in this section we study the Sum-of-Square method that uses pseudo-distribution instead of Sum-of-Square refutation which can be computed using semi-definite program as well). Then we might be able to sample from this Gaussian distribution and hope that the sampled vector can give a close approximation of $2 - \lambda^*$, or perhaps $\gamma^{(4)}$. Since this Gaussian distribution's first two moments are the same as the pseudo-distribution, proving an approximation guarantee might amount to proving that there exists some function $f : \mathbb{R} \to \mathbb{R}$ such that

$$\tilde{\mathbb{E}}_{\mu(x)}(2 - R(x)) \leq f(\tilde{\mathbb{E}}_{\mu(x,\eta)} P_0(x, \eta)).$$

This function, in some ideal cases, could be the map $\alpha \mapsto c \cdot \alpha$, for some $c \geq 1$. However, this is a difficult task and so far we are still not able to prove it. This could therefore be our future research direction.

Finally, as is mentioned in the early sections, approximating the maximum eigenvalue of the hypergraph Laplacian is important because it can be related to the hypergraph max-cut which is NP-hard to compute. However, since, in theory, the Sum-of-Square method can be applied to any system of polynomial equations, we might also consider approximating max-cut directly using the Sum-of-Square method. In particular, we might consider rewriting the hypergraph max-cut as a polynomial optimization problem. This is achieved by noticing that

$$\sum_{e \in E} \omega_e \max_{u,v \in e} (f_u - f_v)^2 = \partial S, \text{ for } f = \mathbf{1}_S, S \subset V,$$

where $\mathbf{1}_S$ is the characteristic vector of S. Therefore, the hypergraph max-cut can be transformed into the maximization of the generalized quadratic form over the hypercube $\{0,1\}^V$. In particular, we have the following optimization problem:

$$\begin{aligned} \text{Maximize} \quad & \textstyle\sum_{e\in E} \omega_e \max_{u,v\in e}(f_u - f_v)^2 \\ \text{s.t.} \quad & f_v^2 - f_v = 0, \qquad \forall v \in V, \end{aligned}$$

where the constraints requires that either $f_v = 0$ or $f_v = 1$. Similar to the approximation of the maximum eigenpair of the hypergraph Laplacian, we can transform the above optimization problem to:

$$\begin{aligned} \text{Maximize} \quad & \textstyle\sum_{e\in E} \omega_e \sum_{u,v\in e}(\eta^e_{u,v})^2(f_u - f_v)^2 \\ \text{s.t.} \quad & f_v^2 - f_v = 0, \qquad \forall v \in V, \\ & \textstyle\sum_{u,v\in e}(\eta^e_{u,v})^2 = 1, \qquad \forall e \in E. \end{aligned}$$

Note that for this optimization problem, we cannot use the binary search strategy mentioned before since the hypercube is discreet. However, note that there are no more than n possible values for the hypergraph max-cut, we might just check all of them. Again we mention that it is very hard to prove any approximation guarantee for the Sum-of-Square algorithm with this optimization problem. However, this algorithm might only be interesting for hypergraphs with edges that are of small sizes. Goemans-Williamson Algorithm [8] is a 0.868 approximation algorithm and when phrased as a Sum-of-Squares algorithm similar to the above, this is still the best guarantee we have. Therefore we might not expect that the above optimization problem can do much better. However, for an r-uniform hypergraph (a hypergraph whose edges are of the same size r), a trivial random bipartition of V can already achieve a $(1 - \frac{1}{2^{r-1}})$ approximation of the max-cut in expectation, hence the above optimization problem might only be interesting for 3-uniform hypergraphs, perhaps. In any case, proving some approximation guarantees for the above optimization problem could also be a direction for future work.

3 Conclusion

This project studies both the theoretical and computational aspects of the spectral theory of hypergraphs and submodular transformations. The study of these theories can connect to various areas of interests, including approximation schemes and optimization theory.

This report described proofs for the conjectures on the spectrum of the Laplacians defined in the previous sections. The proof touches upon various interesting mathematical theories, including some theoretical results on linear programming. We also considered some computational issues on the spectral theories, namely the spectral sparsifier of hypergraph and the approximation of the maximum eigenvalue of the hypergraph Laplacian. We identified that it is difficult to reduce the complexity overhead induced by these computational problems. In particular, a possible useful tool, Sum-of-Square method is studied and we proposed a degree-4 Sum-of-Squares algorithm for the approximation of the maximum eigenvalue of the hypergraph Laplacian. However, for the moment, we are not able to give our algorithm provable approximation guarantees.

As is clear from the previous section, there are a plenty of room for future works. The approximation of the maximum eigenvalue is still an open problem. Future work might be on proving approximation guarantees for our Sum-of-Squares algorithms, in particular, on finding $c \geq 1$ such that $\tilde{\mathbb{E}}_{\mu(x)}(2 - R(x)) \leq c \cdot \tilde{\mathbb{E}}_{\mu(x,\eta)} P_0(x, \eta)$.

References

[1] T.-H. Hubert Chan, Anand Louis, Zhihao Gavin Tang, and Chenzi Zhang. Spectral properties of hypergraph laplacian and approximation algorithms. *CoRR*, abs/1605.01483, 2016.

[2] T.-H. Hubert Chan and Zhibin Liang. Generalizing the hypergraph laplacian via a diffusion process with mediators. *CoRR*, abs/1804.11128, 2018.

[3] F.R.K. Chung. *Spectral Graph Theory*. Number no. 92 in CBMS Regional Conference Series. Conference Board of the Mathematical Sciences, 1992.

[4] Yuichi Yoshida. Cheeger inequalities for submodular transformations. *CoRR*, abs/1708.08781, 2017.

[5] T. Soma and Y. Yoshida. Spectral Sparsification of Hypergraphs. *ArXiv e-prints*, July 2018.

[6] Nimrod Megiddo and R. Chandrasekaran. On the ϵ-perturbation method for avoiding degeneracy. *Operations Research Letters*, 8(6):305 – 308, 1989.

[7] Boaz Barak and David Steurer. Sum-of-squares proofs and the quest toward optimal algorithms. *CoRR*, abs/1404.5236, 2014.

[8] Michel X. Goemans and David P. Williamson. Improved approximation algorithms for maximum cut and satisfiability problems using semidefinite programming. *J. ACM*, 42(6):1115–1145, November 1995.