

## Lecture 8: Spectral Algorithms Wrap-up

*In which we talk about even more generalizations of Cheeger's inequalities, and we analyze the power method to find approximate eigenvectors, thus having a complete description of a polynomial-time approximation algorithm for sparsest cut*

### 1 Irregular Graphs

For simplicity, we proved our results on  $\lambda_2$  and  $\lambda_k$  for regular graphs. Those results extend, essentially with the same proofs, to the case of irregular undirected graphs. In an irregular graph  $G = (V, E)$ , the notion that generalizes edge expansion is called *conductance*. If  $d_v$  is the degree of vertex  $v$ , then the conductance of set  $S$  of vertices is

$$\phi(S) := \frac{E(S, V - S)}{\sum_{v \in S} d_v}$$

We will call the sum of the degrees of a set of vertices the *volume* of the set, and denote it  $\text{vol}(S) := \sum_{v \in S} d_v$ . The conductance of the graph  $G$  is

$$\phi(G) := \min_{S: \text{vol}(S) \leq \frac{1}{2} \text{vol}(V)} \phi(S)$$

Higher-order conductance is defined as higher-order expansion, but with conductance replacing expansion in the definition.

The Cheeger inequalities

$$\frac{\lambda_2}{2} \leq \phi(G) \leq \sqrt{2\lambda_2}$$

still hold, with the same proof. With some abuse of notation, we will call the following quantity the Rayleigh quotient of  $\mathbf{x}$

$$R_L(\mathbf{x}) := \frac{\sum_{\{u,v\} \in E} (x_u - x_v)^2}{\sum_{v \in V} d_v x_v^2}$$

even if, technically, it is the Rayleigh quotient of  $D^{1/2}\mathbf{x}$ , where  $D$  is the diagonal matrix of degrees.

We can also adapt the proof of the higher-order Cheeger inequality to show

$$\frac{\lambda_k}{2} \leq \phi_k(G) \leq O(k^{3.5}) \cdot \sqrt{\lambda_k}$$

## 2 More Cheeger-type Bounds

We proved that if  $(S_F, V - S_F)$  is the cut found by Fiedler's algorithm using the eigenvector of  $\lambda_2$ , then

$$\phi(S_F, V - S_F) \leq 2\sqrt{\phi(G)}$$

which is a good bound, although it usually underestimates the quality of the solutions found in practice. (There are graphs, however, for which the above inequality is tight within a constant factor.)

One case in which we can improve the analysis is when there are not too many eigenvalues close to  $\lambda_2$

**Theorem 1** *There is a constant  $c$  such that, if  $(S_F, V - S_F)$  is the cut obtained by Fiedler's algorithm using an eigenvector for  $\lambda_2$ , then, for every  $k \geq 2$ ,*

$$\phi(S_F, V - S_F) \leq c \cdot k \cdot \frac{\lambda_2}{\sqrt{\lambda_k}}$$

So we have

$$\phi(S_F, V - S_F) \leq 2c \cdot \phi(G) \cdot \min_{k \geq 2} \frac{k}{\sqrt{\lambda_k}}$$

which is a better bound for families of graphs in which, for some  $k$ ,  $\lambda_k \gg k^2\lambda_2$ .

We will not have time to prove Theorem 1, but we will state the two main pieces of its proof.

**Lemma 2** *Let  $\mathbf{x} \in \mathbb{R}_{\geq 0}^V$  be a non-negative vector. Then, for every  $k$ , there is a non-negative vector  $\mathbf{y} \in \mathbb{R}_{\geq 0}^V$  whose entries take at most  $2k$  distinct values and such that*

$$\|\mathbf{x} - \mathbf{y}\|^2 \leq \frac{R_L(\mathbf{x})}{\lambda_k} \|\mathbf{x}\|^2$$

That is, if  $R_L(\mathbf{x}) \gg \lambda_k$ , then there are  $2k$  values such that most entries of  $\mathbf{x}$  are close to one of those  $2k$  values.

**Lemma 3** *There is a constant  $c'$  such that, for every non-negative vectors  $\mathbf{x} \in \mathbb{R}_{\geq 0}^V$  and  $\mathbf{y} \in \mathbb{R}_{\geq 0}^V$ , if  $\mathbf{y}$  is such that its entries contain only  $k$  distinct values, then there is a threshold  $t > 0$  such that*

$$\phi(\{v : x_v \geq t\}) \leq c' \cdot k \cdot \left( R_L(\mathbf{x}) + \sqrt{R_L(\mathbf{x})} \cdot \frac{\|\mathbf{x} - \mathbf{y}\|}{\|\mathbf{x}\|} \right)$$

The above lemma should be compared to the fact, which was a major piece in the proof of Cheeger's inequality, that if  $\mathbf{x} \in \mathbb{R}_{\geq 0}^V$  is an arbitrary non-negative vector, then there is a threshold  $t > 0$  such that

$$\phi(\{v : x_v \geq t\}) \leq \sqrt{2R_L(\mathbf{x})}$$

One obtains Theorem 1 in the following. Start from an eigenvector  $\mathbf{x}$  of  $\lambda_2$  and, using the first step in the proof of Cheeger's inequality, obtain a vector  $\mathbf{x}' \in \mathbb{R}_{\geq 0}^V$  with non-negative entries such that  $R_L(\mathbf{x}') \leq R_L(\mathbf{x}) = \lambda_2$  and such that the support of  $\mathbf{x}'$  contains at most  $|V|/2$  vertices.

Use Lemma 2 to find a vector  $\mathbf{y}$  with non-negative entries and with at most  $2k$  distinct values among its entries such that  $\|\mathbf{x}' - \mathbf{y}\|^2 \leq \frac{\lambda_2}{\lambda_k} \|\mathbf{x}'\|^2$ . Then use Lemma 3 and the fact that  $\lambda_k \leq 2$  to conclude that there exists a  $t > 0$  such that

$$\phi(\{v : x'_v \geq t\}) \leq O(k) \cdot \frac{\lambda_2}{\sqrt{\lambda_k}}$$

The set  $\{v : x'_v \geq t\}$  contains at most  $|V|/2$  vertices, it is one of the cuts considered by Fiedler's algorithm on input  $\mathbf{x}$ .

Another property of graphs in which  $\lambda_k$  is large for small  $k$  is that they contain large expanders as induced subgraphs.

**Theorem 4** *There is a constant  $c$  such that, for every graph  $G$  and every  $k$ , there exists a partition of the vertices into  $\ell \leq k$  sets  $(S_1, \dots, S_\ell)$  such that, if we call  $G_i$  the subgraph induced by the vertex set  $S_i$ , we have*

$$\phi_{G_i} \geq c \frac{\lambda_k}{k^2}$$

**Theorem 5** *If  $\phi_{k+1} > (1+\epsilon)\phi_k$ , then there is a partition of the vertices into  $k$  subsets  $(S_1, \dots, S_k)$  such that*

$$\forall i \in \{1, \dots, k\} : \quad \phi_{G_i} \geq \Omega\left(\frac{\epsilon}{k}\right) \cdot \phi_{k+1}, \quad \phi(S_i) \leq k\phi_k$$

### 3 The Power Method

Earlier in this class, we showed that, if  $G = (V, E)$  is a  $d$ -regular graph, and  $L$  is its normalized Laplacian matrix with eigenvalues  $0 = \lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_n$ , given an eigenvector of  $\lambda_2$ , Fiedler's algorithm finds, in nearly-linear time  $O(|E| + |V| \log |V|)$ , a cut  $(S, V - S)$  such that  $\phi(S) \leq 2\sqrt{\phi(G)}$ .

More generally, if, instead of being given an eigenvector  $\mathbf{x}$  such that  $L\mathbf{x} = \lambda_2\mathbf{x}$ , we are given a vector  $\mathbf{x} \perp \mathbf{1}$  such that  $\mathbf{x}^T L\mathbf{x} \leq (\lambda_2 + \epsilon)\mathbf{x}^T \mathbf{x}$ , then the algorithm finds a cut such that  $\phi(S) \leq \sqrt{4\phi(G) + 2\epsilon}$ . We will now see how to compute such a vector using  $O((|V| + |E|) \cdot \frac{1}{\epsilon} \cdot \log \frac{|V|}{\epsilon})$  arithmetic operations.

A symmetric matrix is *positive semi-definite* (abbreviated PSD) if all its eigenvalues are nonnegative. We begin by describing an algorithm that approximates the largest eigenvalue of a given symmetric PSD matrix. This might not seem to help very much because we want to compute the second smallest, not the largest, eigenvalue. We will see, however, that the algorithm is easily modified to accomplish what we want.

#### 3.1 The Power Method to Approximate the Largest Eigenvalue

The algorithm works as follows

Algorithm *Power*

Input: PSD matrix  $M$ , parameter  $k$

- Pick uniformly at random  $\mathbf{x}_0 \sim \{-1, 1\}^n$
- for  $i := 1$  to  $k$   
     $\mathbf{x}_i := M \cdot \mathbf{x}_{i-1}$
- return  $\mathbf{x}_k$

That is, the algorithm simply picks uniformly at random a vector  $\mathbf{x}$  with  $\pm 1$  coordinates, and outputs  $M^k \mathbf{x}$ .

Note that the algorithm performs  $O(k \cdot (n + m))$  arithmetic operations, where  $m$  is the number of non-zero entries of the matrix  $M$ .

**Theorem 6** *For every PSD matrix  $M$ , positive integer  $k$  and parameter  $\epsilon > 0$ , with probability  $\geq 3/16$  over the choice of  $\mathbf{x}_0$ , the algorithm Power outputs a vector  $\mathbf{x}_k$  such that*

$$\frac{\mathbf{x}_k^T M \mathbf{x}_k}{\mathbf{x}_k^T \mathbf{x}_k} \geq \lambda_1 \cdot (1 - \epsilon) \cdot \frac{1}{1 + 4n(1 - \epsilon)^{2k}}$$

where  $\lambda_1$  is the largest eigenvalue of  $M$ .

Note that, in particular, we can have  $k = O(\log n/\epsilon)$  and  $\frac{\mathbf{x}_k^T M \mathbf{x}_k}{\mathbf{x}_k^T \mathbf{x}_k} \geq (1 - O(\epsilon)) \cdot \lambda_1$ .

Let  $\lambda_1 \geq \dots \geq \lambda_n$  be the eigenvalues of  $M$ , with multiplicities, and  $\mathbf{v}_1, \dots, \mathbf{v}_n$  be a system of orthonormal eigenvectors such that  $M\mathbf{v}_i = \lambda_i \mathbf{v}_i$ . Theorem 6 is implied by the following two lemmas

**Lemma 7** *Let  $\mathbf{v} \in \mathbb{R}^n$  be a vector such that  $\|\mathbf{v}\| = 1$ . Sample uniformly  $\mathbf{x} \sim \{-1, 1\}^n$ . Then*

$$\mathbb{P} \left[ |\langle \mathbf{x}, \mathbf{v} \rangle| \geq \frac{1}{2} \right] \geq \frac{3}{16}$$

**Lemma 8** *For every  $\mathbf{x} \in \mathbb{R}^n$ , for every positive integer  $k$  and positive  $\epsilon > 0$ , if we define  $\mathbf{y} := M^k \mathbf{x}$ , we have*

$$\frac{\mathbf{y}^T M \mathbf{y}}{\mathbf{y}^T \mathbf{y}} \geq \lambda_1 \cdot (1 - \epsilon) \cdot \left( 1 + \frac{\|\mathbf{x}\|^2}{\langle \mathbf{x}, \mathbf{v}_1 \rangle^2} (1 - \epsilon)^{2k} \right)^{-1}$$

It remains to prove the two lemmas.

PROOF: (Of Lemma 7) Let  $\mathbf{v} = (v_1, \dots, v_n)$ . The inner product  $\langle \mathbf{x}, \mathbf{v} \rangle$  is the random variable

$$S := \sum_i x_i v_i$$

Let us compute the first, second, and fourth moment of  $S$ .

$$\mathbb{E} S = 0$$

$$\mathbb{E} S^2 = \sum_i v_i^2 = 1$$

$$\mathbb{E} S^4 = 3 \left( \sum_i v_i^2 \right) - 2 \sum_i v_i^4 \leq 3$$

Recall that the Paley-Zygmund inequality states that if  $Z$  is a non-negative random variable with finite variance, then, for every  $0 \leq \delta \leq 1$ , we have

$$\mathbb{P}[Z \geq \delta \mathbb{E} Z] \geq (1 - \delta)^2 \cdot \frac{(\mathbb{E} Z)^2}{\mathbb{E} Z^2} \quad (1)$$

which follows by noting that

$$\mathbb{E} Z = \mathbb{E}[Z \cdot 1_{Z < \delta \mathbb{E} Z}] + \mathbb{E}[Z \cdot 1_{Z \geq \delta \mathbb{E} Z}] ,$$

that

$$\mathbb{E}[Z \cdot 1_{Z < \delta \mathbb{E} Z}] \leq \delta \mathbb{E} Z ,$$

and that

$$\begin{aligned} \mathbb{E}[Z \cdot 1_{Z \geq \delta \mathbb{E} Z}] &\leq \sqrt{\mathbb{E} Z^2} \cdot \sqrt{\mathbb{E} 1_{Z \geq \delta \mathbb{E} Z}} \\ &= \sqrt{\mathbb{E} Z^2} \sqrt{\mathbb{P}[Z \geq \delta \mathbb{E} Z]} \end{aligned}$$

We apply the Paley-Zygmund inequality to the case  $Z = S^2$  and  $\delta = 1/4$ , and we derive

$$\mathbb{P} \left[ S^2 \geq \frac{1}{4} \right] \geq \left( \frac{3}{4} \right)^2 \cdot \frac{1}{3} = \frac{3}{16}$$

□

**Remark 9** *The proof of Lemma 7 works even if  $\mathbf{x} \sim \{-1, 1\}^n$  is selected according to a 4-wise independent distribution. This means that the algorithm can be derandomized in polynomial time.*

PROOF: (Of Lemma 8) Let us write  $\mathbf{x}$  as a linear combination of the eigenvectors

$$\mathbf{x} = a_1 \mathbf{v}_1 + \cdots + a_n \mathbf{v}_n$$

where the coefficients can be computed as  $a_i = \langle \mathbf{x}, \mathbf{v}_i \rangle$ . We have

$$\mathbf{y} = a_1 \lambda_1^k \mathbf{v}_1 + \cdots + a_n \lambda_n^k \mathbf{v}_n$$

and so

$$\mathbf{y}^T M \mathbf{y} = \sum_i a_i^2 \lambda_i^{2k+1}$$

and

$$\mathbf{y}^T \mathbf{y} = \sum_i a_i^2 \lambda_i^{2k}$$

We need to prove a lower bound to the ratio of the above two quantities. We will compute a lower bound to the numerator and an upper bound to the denominator in terms of the same parameter.

Let  $\ell$  be the number of eigenvalues larger than  $\lambda_1 \cdot (1 - \epsilon)$ . Then, recalling that the eigenvalues are sorted in non-increasing order, we have

$$\mathbf{y}^T M \mathbf{y} \geq \sum_{i=1}^{\ell} a_i^2 \lambda_i^{2k+1} \geq \lambda_1 (1 - \epsilon) \sum_{i=1}^{\ell} a_i^2 \lambda_i^{2k}$$

We also see that

$$\begin{aligned} & \sum_{i=\ell+1}^n a_i^2 \lambda_i^{2k} \\ & \leq \lambda_1^{2k} \cdot (1 - \epsilon)^{2k} \sum_{i=\ell+1}^n a_i^2 \\ & \leq \lambda_1^{2k} \cdot (1 - \epsilon)^{2k} \cdot \|\mathbf{x}\|^2 \\ & \leq a_1^2 \lambda_1^{2k} (1 - \epsilon)^{2k} \cdot \frac{\|\mathbf{x}\|^2}{a_1^2} \\ & \leq \frac{\|\mathbf{x}\|^2}{a_1^2} (1 - \epsilon)^{2k} \sum_{i=1}^{\ell} a_i^2 \lambda_i^{2k} \end{aligned}$$

So we have

$$\mathbf{y}^T \mathbf{y} \leq \left( 1 + \frac{\|\mathbf{x}\|^2}{a_1^2} (1 - \epsilon)^{2k} \right) \cdot \sum_{i=1}^{\ell} a_i^2$$

giving

$$\frac{\mathbf{y}^T M \mathbf{y}}{\mathbf{y}^T \mathbf{y}} \geq \lambda_1 \cdot (1 - \epsilon) \cdot \left( 1 + \frac{\|\mathbf{x}\|^2}{a_1^2} (1 - \epsilon)^{2k} \right)^{-1}$$

□

**Remark 10** *Where did we use the assumption that  $M$  is positive semidefinite? What happens if we apply this algorithm to the adjacency matrix of a bipartite graph?*

## 3.2 Approximating the Second Largest Eigenvalue

Suppose now that we are interested in finding the second largest eigenvalue of a given PSD matrix  $M$ . If  $M$  has eigenvalues  $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n$ , and we know the eigenvector  $\mathbf{v}_1$  of  $\lambda_1$ , then  $M$  is a PSD linear map from the orthogonal space to  $\mathbf{v}_1$  to itself, and  $\lambda_2$  is the largest eigenvalue of this linear map. We can then run the previous algorithm on this linear map.

Algorithm *Power2*  
 Input: PSD matrix  $M$ , vector  $\mathbf{v}_1$  parameter  $k$

- Pick uniformly at random  $\mathbf{x} \sim \{-1, 1\}^n$
- $\mathbf{x}_0 := \mathbf{x} - \mathbf{v}_1 \cdot \langle \mathbf{x}, \mathbf{v}_1 \rangle$
- for  $i := 1$  to  $k$ 
  - $\mathbf{x}_i := M \cdot \mathbf{x}_{i-1}$
- return  $\mathbf{x}_k$

If  $\mathbf{v}_1, \dots, \mathbf{v}_n$  is an orthonormal basis of eigenvectors for the eigenvalues  $\lambda_1 \geq \dots \geq \lambda_n$  of  $M$ , then, at the beginning, we pick a random vector

$$\mathbf{x} = a_1 \mathbf{v}_1 + a_2 \mathbf{v}_2 + \dots + a_n \mathbf{v}_n$$

that, with probability at least  $3/16$ , satisfies  $|a_2| \geq 1/2$ . (Cf. Lemma 7.) Then we compute  $\mathbf{x}_0$ , which is the projection of  $\mathbf{x}$  on the subspace orthogonal to  $\mathbf{v}_1$ , that is

$$\mathbf{x}_0 = a_2 \mathbf{v}_2 + \dots + a_n \mathbf{v}_n$$

Note that  $\|\mathbf{x}\|^2 = n$  and that  $\|\mathbf{x}_0\|^2 \leq n$ .

The output is the vector  $\mathbf{x}_k$

$$\mathbf{x}_k = a_2 \lambda_2^k \mathbf{v}_2 + \dots + a_n \lambda_n^k \mathbf{v}_n$$

If we apply Lemma 8 to subspace orthogonal to  $\mathbf{v}_1$ , we see that when  $|a_2| \geq 1/2$  we have that, for every  $0 < \epsilon < 1$ ,

$$\frac{\mathbf{x}_k^T M \mathbf{x}_k}{\mathbf{x}_k^T \mathbf{x}_k} \geq \lambda_2 \cdot (1 - \epsilon) \cdot \frac{1}{4n(1 - \epsilon)^{2k}}$$

We have thus established the following analysis.



**Theorem 11** For every PSD matrix  $M$ , positive integer  $k$  and parameter  $\epsilon > 0$ , if  $\mathbf{v}_1$  is a length-1 eigenvector of the largest eigenvalue of  $M$ , then with probability  $\geq 3/16$  over the choice of  $\mathbf{x}_0$ , the algorithm Power2 outputs a vector  $\mathbf{x}_k \perp \mathbf{v}_1$  such that

$$\frac{\mathbf{x}_k^T M \mathbf{x}_k}{\mathbf{x}_k^T \mathbf{x}_k} \geq \lambda_2 \cdot (1 - \epsilon) \cdot \frac{1}{1 + 4n(1 - \epsilon)^{2k}}$$

where  $\lambda_2$  is the second largest eigenvalue of  $M$ , counting multiplicities.

### 3.3 The Second Smallest Eigenvalue of the Laplacian

Finally, we come to the case in which we want to compute the second smallest eigenvalue of the normalized Laplacian matrix  $L = I - \frac{1}{d}A$  of a  $d$ -regular graph  $G = (V, E)$ , where  $A$  is the adjacency matrix of  $G$ .

Consider the matrix  $M := 2I - L = I + \frac{1}{d}A$ . Then if  $0 = \lambda_1 \leq \dots \leq \lambda_n \leq 2$  are the eigenvalues of  $L$ , we have that

$$2 = 2 - \lambda_1 \geq 2 - \lambda_2 \geq \dots \geq 2 - \lambda_n \geq 0$$

are the eigenvalues of  $M$ , and that  $M$  is PSD.  $M$  and  $L$  have the same eigenvectors, and so  $\mathbf{v}_1 = \frac{1}{\sqrt{n}}(1, \dots, 1)$  is a length-1 eigenvector of the largest eigenvalue of  $M$ .

By running algorithm Power2, we can find a vector  $\mathbf{x}$  such that

$$\mathbf{x}^T M \mathbf{x}^T \geq (1 - \epsilon) \cdot (2 - \lambda_2) \cdot \mathbf{x}^T \mathbf{x}$$

and

$$\mathbf{x}^T M \mathbf{x}^T = 2\mathbf{x}^T \mathbf{x} - \mathbf{x}^T L \mathbf{x}$$

so, rearranging, we have

$$\frac{\mathbf{x}^T L \mathbf{x}}{\mathbf{x}^T \mathbf{x}} \leq \lambda_2 + 2\epsilon$$

If we want to compute a vector whose Rayleigh quotient is, say, at most  $2\lambda_2$ , then the running time will be  $\tilde{O}((|V| + |E|)/\lambda_2)$ , because we need to set  $\epsilon = \lambda_2/2$ , which is not nearly linear in the size of the graph if  $\lambda_2$  is, say  $O(1/|V|)$ .

For a running time that is nearly linear in  $n$  for all values of  $\lambda_2$ , one can, instead, apply the power method to the *pseudoinverse*  $L^+$  of  $L$ . (Assuming that the graph is connected,  $L^+\mathbf{x}$  is the unique vector  $\mathbf{y}$  such that  $L\mathbf{y} = \mathbf{x}$ , if  $\mathbf{x} \perp (1, \dots, 1)$ , and  $L^+\mathbf{x} = 0$  if  $\mathbf{x}$  is parallel to  $(1, \dots, 1)$ .) This is because  $L^+$  has eigenvalues  $0, 1/\lambda_2, \dots, 1/\lambda_n$ , and so  $L^+$  is PSD and  $1/\lambda_2$  is its largest eigenvalue.

Although computing  $L^+$  is not known to be doable in nearly linear time, there are nearly linear time algorithms that, given  $\mathbf{x}$ , solve in  $\mathbf{y}$  the linear system  $L\mathbf{y} = \mathbf{x}$ ,

and this is the same as computing the product  $L^+\mathbf{x}$ , which is enough to implement algorithm *Power* applied to  $L^+$ .

(Such algorithms will be discussed in the third part of the course. The algorithms will find an approximate solution  $\mathbf{y}$  to the linear system  $L\mathbf{y} = \mathbf{x}$ , but this will be sufficient. In the following, we proceed as if the solution was exact.)

In time  $O((V + |E|) \cdot (\log |V|/\epsilon)^{O(1)})$ , we can find a vector  $\mathbf{y}$  such that  $\mathbf{y} = (L^+)^k \mathbf{x}$ , where  $\mathbf{x}$  is a random vector in  $\{-1, 1\}^n$ , shifted to be orthogonal to  $(1, \dots, 1)$  and  $k = O(\log |V|/\epsilon)$ . What is the Rayleigh quotient of such a vector with respect to  $L$ ?

Let  $\mathbf{v}_1, \dots, \mathbf{v}_n$  be a basis of orthonormal eigenvectors for  $L$  and  $L^+$ . If  $0 = \lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_n$  are the eigenvalues of  $L$ , then we have

$$L\mathbf{v}_1 = L^+\mathbf{v}_1 = \mathbf{0}$$

and, for  $i = 1, \dots, n$ , we have

$$L\mathbf{v}_i = \lambda_i \mathbf{v}_i \quad L^+\mathbf{v}_i = \frac{1}{\lambda_i} \mathbf{v}_i$$

Write  $\mathbf{x} = a_2\mathbf{v}_2 + \dots + a_n\mathbf{v}_n$ , where  $\sum_i a_i^2 \leq n$ , and assume that, as happens with probability at least  $3/16$ , we have  $a_2^2 \geq \frac{1}{4}$ . Then

$$\mathbf{y} = \sum_{i=2}^n a_i \frac{1}{\lambda_i^k} \mathbf{v}_i$$

and the Rayleigh quotient of  $\mathbf{y}$  with respect to  $L$  is

$$\frac{\mathbf{y}^T L \mathbf{y}}{\mathbf{y}^T \mathbf{y}} = \frac{\sum_i a_i^2 \frac{1}{\lambda_i^{2k-1}}}{\sum_i a_i^2 \frac{1}{\lambda_i^{2k}}}$$

and the analysis proceeds similarly to the analysis of the previous section. If we let  $\ell$  be the index such that  $\lambda_\ell \leq (1+\epsilon) \cdot \lambda_2 \leq \lambda_{\ell+1}$  then we can upper bound the numerator as

$$\begin{aligned} \sum_i a_i^2 \frac{1}{\lambda_i^{2k-1}} &\leq \sum_{i \leq \ell} a_i^2 \frac{1}{\lambda_i^{2k-1}} + \frac{1}{(1+\epsilon)^{2k-1} \lambda_2^{2k-1}} \sum_{i > \ell} a_i^2 \\ &\leq \sum_{i \leq \ell} a_i^2 \frac{1}{\lambda_i^{2k-1}} + \frac{1}{(1+\epsilon)^{2k-1} \lambda_2^{2k-1}} \cdot n \\ &\leq \sum_{i \leq \ell} a_i^2 \frac{1}{\lambda_i^{2k-1}} + \frac{1}{(1+\epsilon)^{2k-1} \lambda_2^{2k-1}} \cdot 4na_2^2 \end{aligned}$$

$$\leq \left(1 + \frac{4n}{(1 + \epsilon)^{2k-1}}\right) \cdot \sum_{i \leq \ell} a_i^2 \frac{1}{\lambda_i^{2k-1}}$$

and we can lower bound the denominator as

$$\begin{aligned} \sum_i a_i^2 \frac{1}{\lambda_i^{2k}} &\geq \sum_{i \leq \ell} a_i^2 \frac{1}{\lambda_i^{2k}} \\ &\geq \frac{1}{(1 + \epsilon)\lambda_2} \cdot \sum_{i \leq \ell} a_i^2 \frac{1}{\lambda_i^{2k-1}} \end{aligned}$$

and the Rayleigh quotient is

$$\frac{\mathbf{y}^T L \mathbf{y}}{\mathbf{y}^T \mathbf{y}} \leq \lambda_2 \cdot (1 + \epsilon) \cdot \left(1 + \frac{4n}{(1 + \epsilon)^{2k-1}}\right) \leq (1 + 2\epsilon) \cdot \lambda_2$$

when  $k = O\left(\frac{1}{\epsilon} \log \frac{n}{\epsilon}\right)$ .