

Expanders and Eigenvalues

1 Overview of this lecture

We give a lower bound on the size of the second largest eigenvalue of an adjacency matrix. Then we relate the expansion properties of a d -regular graph to the size of the eigenvalues of its adjacency matrix. In the last section we consider how the amount of randomness used by a class of probabilistic algorithms can be reduced using a random walk on an expander graph.

2 Notation and Definitions

We denote by $G = (V, E)$ a d -regular graph on $n = |V|$ nodes with no multiple edges and no self-loops. We denote by A its adjacency matrix. Thus, A is a symmetric $n \times n$ -matrix in which each element either is zero or one and each row and column contains exactly d ones and $n - d$ zeros.

We denote by $d = \mu_1 \geq \dots \geq \mu_n$ the eigenvalues of A , and we denote by $d = \lambda_1 \geq \dots \geq \lambda_n$ the corresponding absolute values (possibly in different order).

Given subsets $X, Y \subset V$, we denote by $E(X, Y)$ the set of edges $(x, y) \in E$ such that $x \in X$ and $y \in Y$, and we write $e(X, Y) = |E(X, Y)|$.

2.1 The Trace of a Matrix

We write $\text{Tr} A$ for the trace of the matrix $A = (a_{ij})$, i.e. the sum $\sum_{i=1}^n a_{ii}$ of the elements along its diagonal. The trace obviously is linear, i.e. $\text{Tr}(A + B) = \text{Tr} A + \text{Tr} B$ and $\text{Tr}(kA) = k\text{Tr}(A)$ for matrices A and B and scalars k .

Fact 1. $\text{Tr}(AB) = \text{Tr}(BA)$ whenever the products make sense.

Proof. Let A be an $n \times m$ -matrix and let B be an $m \times n$ -matrix. Then we have $\text{Tr}(AB) = \sum_{i=1}^n \sum_{l=1}^m a_{il} b_{li} = \sum_{l=1}^m \sum_{i=1}^n b_{li} a_{il} = \text{Tr}(BA)$. $\square$

Recall that two $n \times n$ -matrices A and B are called similar if $A = PAP^{-1}$ for some invertible matrix P . The fact above implies that similar matrices have the equal trace.

Fact 2. Let A be an $n \times n$ matrix over $\mathbb{C}$. Then $\text{Tr} A = \sum_{i=1}^n \mu_i$, where $\mu_1, \dots, \mu_n$ are the eigenvalues of A .

Proof. Any square matrix A is similar to an upper triangular matrix. This is just a matter of performing gauss elimination, since every elementary row operation is invertible. The trace of an upper triangular matrix is the sum of the elements on its diagonal, and for an upper triangular matrix the diagonal elements are the eigenvalues. $\square$

3 Bounding λ_2 From Below

It is possible to derive the following bound.

Theorem 3. $\lambda_2 \geq 2\sqrt{d-1} - o_n(1)$.

We settle for a weaker bound that is easier to derive and then discuss informally how the stronger bound can be achieved.

Theorem 4. $\lambda_2 \geq \sqrt{d} - o_n(1)$.

Proof. Consider the square $S = A^2$ of A . The eigenvectors of S are obviously the same as those of A . This implies that the eigenvalues of S are given by λ_i^2 for $i = 1, \dots, n$.

Furthermore, since A is symmetric we have $s_{ii} = \sum_{l=1}^n a_{il}a_{li} = d$, i.e. all elements on the diagonal of S equal d . This implies that $\text{Tr}S = nd$. On the other hand we know from Fact 2 that $\text{Tr}S = \sum_{i=1}^n \lambda_i^2$. This implies that

$$\sum_{i=2}^n \lambda_i^2 = dn - d^2, \quad \text{and} \quad \lambda_2^2 \geq \frac{dn - d^2}{n-1}.$$

A simple calculation gives the claim. $\square$

The idea behind the proof of the sharper bound is to generalize squaring to taking any even power of A . For example, we can consider fourth powers $Q = A^4$ instead. The eigenvalues of Q are given by λ_i^4 , so $\text{Tr}Q = \sum_{i=1}^n \lambda_i^4$. We must bound this from below, i.e. we want to bound the sum $\sum_{i=1}^n q_{ii}$. Each element $q_{i,i}$ corresponds to the number of paths of length four from node i to itself. Thus, to bound the trace we can bound the number of such paths. To do this we consider two types of paths illustrated below.

$$i \xrightarrow{d} j \xrightarrow{d-1} k \xrightarrow{1} j \xrightarrow{1} i$$

$$i \xrightarrow{d} j \xrightarrow{1} i \xrightarrow{d} j \xrightarrow{1} i$$

We assume that i, j, k are pairwise distinct nodes. Starting from i we have d choices, since G is d -regular. From j we have $d-1$ choices, since we are not allowed to go back to i . Then from k , we may always take the same way back to i . Thus, there are at least $d^2 - d$ distinct paths of the first type. A

similar analysis for the second type of path give that we have at least d^2 such paths. Note that there may or may not be more paths of each respective type. The lower bound on q_{ii} implies that $\text{Tr}Q \geq (2d^2 - d)n$. Using a similar argument as above we get $\lambda_2^4 \geq \frac{(2d^2 - d)n - d^4}{n-1}$. Thus, the bound we get using the fourth power of A is somewhat better than the bound in Theorem 4.

To generalize the argument we must only come up with a way to bound the size of the diagonal elements of A^{2t} for arbitrary t .

4 Edge Expansion and the Second Eigenvalue

Recall the definition of edge expansion.

Definition 5 (Edge Expansion). *Given a graph G , the edge expansion $\alpha(G)$ is defined as the largest α such that $e(S, S^c) \geq \alpha|S|$ for all $S \subset V$ such that $|S| \leq n/2$.*

It turns out that the expansion of a graph is related to the gap between the largest and second largest eigenvalue of its adjacency matrix. More precisely the following theorem holds.

Theorem 6. $\frac{d-\mu_2}{2} \leq \alpha(G) \leq (2d(d-\mu_2))^{1/2}$.

Note that the relation is not terribly tight, but it does show that the expansion $\alpha(G)$ goes to zero when the second eigenvalue μ_2 goes to d . We divide the proof of the theorem into two lemmas.

Lemma 7. $\frac{d-\mu_2}{2} \leq \alpha(G)$.

Proof. We are looking for a lower bound on μ_2 , in terms of $\alpha(G)$. Let S be any subset of V of size at most $n/2$. Recall (from Jakob's lecture notes on Raleigh-quotas) that $e_1 = (1, 1, \dots, 1)$ is the eigenvector corresponding to the largest eigenvalue $\mu_1 = d$, and that

$$\mu_2 = \max_{x \perp e_1} \frac{x^\top A x}{\|x\|^2}.$$

Thus, to get a lower bound on μ_2 we can plug any vector $x \perp e_1$ into the expression on the right. We consider the fixed vector $x = (x_i)_{i=1}^n$ defined by

$$x_i = \begin{cases} |S^c| & \text{if } i \in S \\ -|S| & \text{if } i \in S^c \end{cases}.$$

We clearly have $(e_1, x) = \sum_{i=1}^n x_i = |S| \cdot |S^c| - |S^c| \cdot |S| = 0$, so $x \perp e_1$ as required. Furthermore,

$$\|x\|^2 = \sum_{i=1}^n x_i^2 = |S| \cdot |S^c|^2 + |S^c| \cdot |S|^2 = n|S| \cdot |S^c|.$$

Unfortunately, we cannot compute $x^\top Ax$ directly, so we bound it from below. By definition we have

$$x^\top Ax = 2 \sum_{(i,j) \in E} x_i x_j . \quad (1)$$

We expect that there are edges between S and S^c , but it is a good starting point to assume that there are no such edges, i.e., we assume that $e(S, S^c) = 0$. Then each node in S has exactly d neighbors all of which are in S . Thus, there are $d|S|/2$ edges between nodes in S . Similarly there are $d|S^c|/2$ edges between nodes in S^c . Combined with the special structure of the vector x it is not hard to see that

$$x^\top Ax = 2 \left(\frac{d|S|}{2} |S^c|^2 + \frac{d|S^c|}{2} |S|^2 \right) = dn|S| \cdot |S^c| .$$

Let us now return to reality. We are given some S such that $e(S, S^c) > 0$.

Define $E_0 = E$. We define E_j from E_{j-1} as follows. We pick two edges $(u, v), (u', v') \in S \times S^c$ and define

$$E_j = (E_{j-1} - \{(u, v), (u', v')\}) \cup \{(u, v'), (u', v)\} .$$

If we write $E_j(S, S^c) = \{(u, v) \in E_j \mid u \in S, v \in S^c\}$ and $e_j(S, S^c) = |E_j(S, S^c)|$, we clearly have $e_j(S, S^c) = e_{j-1}(S, S^c) - 2$. This process must end in a finite number of steps k and give a graph (V, E_k) . If $e(S, S^c)$ is even there are no edges in E_k between S and S^c . If $e(S, S^c)$ is odd, then there may be a single edge in E_k between S and S^c . It is not hard to take care of also this edge, but for simplicity we assume that $e(S, S^c)$ is even.

We can obviously reverse the above process and go from (V, E_k) back to (V, E) in $k = e(S, S^c)/2$ steps. We know the value of $x^\top A_k x$, where A_k is the adjacency matrix of (V, E_k) . Thus, it suffices to consider how $x^\top A_{j-1} x$ is related to $x^\top A_j x$ to bound $x^\top Ax$.

In each step the first edge $(u, v) \in S \times S$ contributes $2|S|^2$ and the second edge $(u', v') \in S^c \times S^c$ contributes $2|S^c|^2$ to $x^\top A_j x$. Thus, when we remove these edges the value of $x^\top A_j x$ decreases by $2(|S|^2 + |S^c|^2)$. When the edges (u, v') and (u', v) are added, the value is decreased by an additional $4|S| \cdot |S^c|$. Thus, we have

$$x^\top A_{j-1} x = x^\top A_j x - 2(|S|^2 + |S^c|^2 + 2|S| \cdot |S^c|) = x^\top A_j x - 2n^2 .$$

This implies that

$$\begin{aligned} x^\top Ax &\geq dn|S| \cdot |S^c| - n^2 e(S, S^c) \\ &\geq dn|S| \cdot |S^c| - n^2 |S| \alpha(G) \quad (\text{use expansion } e(S, S^c) \geq \alpha(G)|S|) \\ &\geq n|S| \cdot |S^c| (d - 2\alpha(G)) . \quad (\text{use } 2|S^c| \geq n) \end{aligned}$$

Finally, we have

$$\mu_2 \geq \frac{n|S| \cdot |S^c|(d - 2\alpha(G))}{n|S| \cdot |S^c|} = d - 2\alpha(G)$$

and the first inequality follows. $\square$

Lemma 8. $\alpha(G) \leq (2d(d - \mu_2))^{1/2}$.

Proof. Denote by e_2 the eigenvector corresponding to the second largest eigenvalue μ_2 . We know that $\sum_{i=1}^n e_{2,i} = (e_2, e_1) = 0$, as the eigenvectors are orthogonal.

Intuitively, we would like to do the “reverse” of the previous argument, but this is not possible since e_2 may not have any particular nice structure.

We define a certain “truncated” vector $f = (f_i)_{i=1}^n$ by

$$f_i = \begin{cases} e_{2,i} & \text{if } e_{2,i} > 0 \\ 0 & \text{otherwise} \end{cases}.$$

Without loss we may assume that

$$|\{1 \leq i \leq n \mid e_{2,i} > 0\}| \leq n/2, \quad (2)$$

since if it is not the case we could use $-e_2$ as the eigenvector instead.

In some way, possibly through use of dark magic, we know that it is a good idea to investigate the expression $\sum_{(i,j) \in E} |f_i^2 - f_j^2|$. Recall that the Cauchy-Schwarz inequality states that $|(x, y)| \leq \|x\| \cdot \|y\|$ for $x, y \in \mathbb{R}^n$. Using this inequality we have

$$\begin{aligned} \sum_{(i,j) \in E} |f_i^2 - f_j^2| &= \sum_{(i,j) \in E} |(f_i - f_j)(f_i + f_j)| \\ &\leq \left(\sum_{(i,j) \in E} (f_i - f_j)^2 \sum_{(i,j) \in E} (f_i + f_j)^2 \right)^{1/2}. \end{aligned} \quad (3)$$

We now bound each of the sums in the last expression. We have

$$\sum_{(i,j) \in E} (f_i + f_j)^2 \leq \sum_{(i,j) \in E} 2(f_i^2 + f_j^2) = 2d \sum_{i=1}^n f_i^2,$$

where the inequality follows from the fact that $(x+y)^2 \leq (x+y)^2 + (x-y)^2 = 2(x^2 + y^2)$ for all $x, y \in \mathbb{R}$, and the equality from d -regularity of G . We are more careful when we bound the second sum. This is bounded by

$$\sum_{(i,j) \in E} (f_i - f_j)^2 = \sum_{(i,j) \in E} (f_i^2 + f_j^2 - 2f_i f_j) = d \sum_{i=1}^n f_i^2 - f^\top A f.$$

Thus, it suffices to bound $f^\top Af$ from above.

Since e_2 is the eigenvector corresponding to μ_2 we have $Ae_2 = \mu_2 e_2$. Denote by $(Af)_i$ the i th element of Af , i.e., $(Af)_i = \sum_{(i,j) \in E} f_j$. If we consider a fixed index i such that $e_{2,i} > 0$ in isolation we have

$$\mu_2 f_i = \mu_2 e_{2,i} = \sum_{j=1}^n a_{i,j} e_{2,j} = \sum_{(i,j) \in E} e_{2,j} \leq \sum_{(i,j) \in E} f_j = (Af)_i ,$$

since when we replace $e_{2,i}$ by f_i all negative $e_{2,i}$ become zero. We use this inequality to analyze $f^\top Af$ as follows

$$f^\top Af = \sum_{i=1}^n f_i (Af)_i = \sum_{e_{2,i} > 0} f_i (Af)_i \geq \sum_{e_{2,i} > 0} f_i \mu_2 f_i = \mu_2 \sum_{i=1}^n f_i^2 ,$$

This implies that the first sum of the expression in (3) is bounded by

$$\sum_{(i,j) \in E} (f_i - f_j)^2 \leq (d - \mu_2) \sum_{i=1}^n f_i^2 ,$$

and we conclude that

$$\sum_{(i,j) \in E} |f_i^2 - f_j^2| \leq (2d(d - \mu_2))^{1/2} \sum_{i=1}^n f_i^2 .$$

To finish the proof we must argue that

$$\sum_{(i,j) \in E} |f_i^2 - f_j^2| \geq \alpha(G) \sum_{i=1}^n f_i^2 .$$

Denote by $\beta_k > \beta_{k-1} > \dots > \beta_0 = 0$ the distinct values assumed by f_i^2 for $i = 1, \dots, n$. Define sets

$$S_l = \{i \mid f_i^2 \geq \beta_l\} ,$$

and set $S_{k+1} = \emptyset$. The following equations hold

$$\begin{aligned} \sum_{(i,j) \in E} |f_i^2 - f_j^2| &= \sum_{l=1}^k (\beta_l - \beta_{l-1}) e(S_l, S_l^c) \\ &\geq \sum_{l=1}^k (\beta_l - \beta_{l-1}) \alpha(G) |S_l| \\ &= \alpha(G) \sum_{l=0}^k \beta_l (|S_l| - |S_{l+1}|) = \alpha \sum_{l=0}^k f_i^2 . \end{aligned} \tag{4}$$

The last equality is just a matter of rearranging the sum. The inequality follows from the expansion property of G and the fact that $|S_l| \leq n/2$, which in turn follows from (2). The first equality requires some explanation.

It is not hard to see that something like the equality must hold. What may not be obvious are the weights $e(S_l, S_l^c)$. The idea is to think of the values β_l on a line with an edge between two values β_{l_i} and β_{l_j} if there is an edge $(i, j) \in E$, as is illustrated in Figure 1.

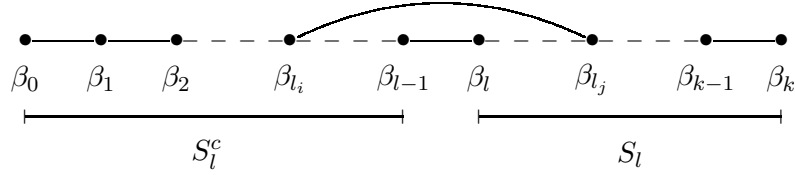


Figure 1: Illustration of the Equality (4).

We define the index $l(i, j)$ to be such that $\beta_{l(i, j)}$ equals the larger of f_i^2 and f_j^2 and we define $s(i, j)$ to be such that $\beta_{s(i, j)}$ equals the smaller of f_i^2 and f_j^2 . Using these indices we can rewrite a term of the sum $\sum_{(i, j) \in E} |f_i^2 - f_j^2|$ on the form

$$\begin{aligned} |f_i^2 - f_j^2| &= \beta_{l(i, j)} - \beta_{s(i, j)} = \sum_{t=s(i, j)+1}^{l(i, j)} (\beta_t - \beta_{t-1}) \\ &= \sum_{t=1}^k \psi_t(i, j) (\beta_t - \beta_{t-1}) , \end{aligned}$$

where ψ_t denotes the characteristic function of the set of edges $E(S_t, S_t^c)$. We substitute the expression for each term back into the original sum and get

$$\begin{aligned} \sum_{(i, j) \in E} |f_i^2 - f_j^2| &= \sum_{(i, j) \in E} \sum_{t=1}^k \psi_{S_t}(i, j) (\beta_t - \beta_{t-1}) \\ &= \sum_{t=1}^k (\beta_t - \beta_{t-1}) \sum_{(i, j) \in E} \psi_{S_t}(i, j) \\ &= \sum_{l=1}^k (\beta_l - \beta_{l-1}) e(S_l, S_l^c) . \end{aligned}$$

□

5 Probabilistic Algorithms and Expanders

In this section we describe a nice application of expander graphs. Let L be some language in NP and assume that M is a probabilistic polynomial time algorithm using R bits of randomness such that

$$\begin{aligned} x \in L &\Rightarrow \Pr[M(x) = 1] = 1 \\ x \notin L &\Rightarrow \Pr[M(x) = 1] \leq \beta . \end{aligned}$$

In other words, if an element x belongs to the language L , the algorithm always accepts, and if x does not belong to L it accepts with probability at most β . The probability is taken over the internal randomness of M . An example of such an algorithm is the primality test of Miller-Rabin. To make the dependence on internal randomness $r \in \{0, 1\}^R$ explicit we write $M(x, r)$ instead of $M(x)$.

In both theory and practice it is useful to decrease the probability of a false positive answer, and a simple way to do this is by repetition. Let M' be the machine that given input x runs $b_i = M(x, r_i)$ for $i = 1, \dots, k$ and outputs 1 if $b_i = 1$ for all i and 0 otherwise. The internal randomness of M' is $r = (r_1, \dots, r_k)$. Then from independence we have

$$\begin{aligned} x \in L &\Rightarrow \Pr[M'(x) = 1] = 1 \\ x \notin L &\Rightarrow \Pr[M'(x) = 1] = \beta^k . \end{aligned}$$

In other words we can decrease the probability of a false positive exponentially, β^k , at a linear cost, kR , in randomness.

We claim that using a random walk on an expander graph the probability of a false positive can be decreased to $O(\beta^k)$ using only $O(R + k)$ random bits.

5.1 The Construction

Let G be a d -regular graph on 2^R nodes. We identify the nodes with the set of strings $\{0, 1\}^R$. Consider now the following algorithm M'' .

M'' chooses $r_1 \in \{0, 1\}^R$ randomly and computes $b_1 = M(x, r_1)$ as before. However, instead of choosing a completely new random string r_2 for the next execution of M it chooses a random integer $i_2 \in \{1, \dots, d\}$ and defines r_2 as the i_2 th neighbor of r_1 in G . In the third execution it chooses a random integer $i_3 \in \{1, \dots, d\}$ and defines r_3 to be the i_3 th neighbor of r_2 , and so on until the last execution of M .

The construction clearly requires $R + k \lceil \log d \rceil = O(R + k)$ random bits. It remains to argue that the decrease in the probability for a false positive decreases exponentially in k .

Note that for the construction to make sense we cannot use an expander graph without structure, e.g., by choosing a random d -regular graph. The

problem is that it must be possible to compute the i th neighbor of any node in G quickly. Thus, the expander graph G must be “very explicit”, in the sense that there exists an algorithm that given the label r of a node and an integer $i \in \{1, \dots, d\}$ outputs the label r' of the i th neighbor of r in G in *polynomial time* in R .

5.2 The Analysis of the Construction

Theorem 9. *Let A be the adjacency matrix of G and let $\gamma = \lambda_2(A)/d$. Then for all $x \notin L$*

$$\Pr[M''(x) = 1] \leq (\beta + \gamma)^k .$$

Before we prove the theorem we consider its consequences. Given any β and $\gamma < 1$, we can run M' defined above and choose a constant k' such that $\beta^{k'} < 1 - \gamma$. Then we apply the construction in the previous section to M' instead of M and conclude that the probability of a false positive is bounded by $(\beta^{k'} + \gamma)^k = O(\beta^k)$.

Proof. For any fixed $x \notin L$ we define S , the set of “bad” random strings by

$$S = \{r \mid M(x, r) = 1\} .$$

The assumption implies that we for all x have $|S| \leq 2^R \beta$. We define the normalized adjacency matrix $B = \frac{1}{d}A$ and note that $\lambda_2(B) = \gamma$.

Consider the following probability

$$p_r^i = \Pr[r_1, \dots, r_i \in S \wedge r_i = r] ,$$

i.e., the probability that we after i steps have never been outside S and are standing in the node r . It follows that $\Pr[M''(x) = 1] = \sum_{s \in \{0,1\}^R} p_r^k$, since the events considered are mutually exclusive.

We define $N = 2^R$ and let P be the projection onto the set S . Thus, if $v = (v_1, \dots, v_N)$ and $w = (w_1, \dots, w_N) = Pv$, we have

$$w_i = \begin{cases} v_i & \text{if } i \in S \\ 0 & \text{otherwise} \end{cases} .$$

If we set $p^i = (p_r^i)_{r \in \{0,1\}^R}$ we have by definition

$$p^0 = (1/N, \dots, 1/N) , \quad \text{and} \quad p^{i+1} = (PBP)p^i . \quad (5)$$

Here we use the fact that $Pp^i = p^i$. This follows since all components p_r^i of p^i with $r \notin S$ are zero. The left equality above implies that $\|p^0\| = 1/\sqrt{N}$. If we could establish that

$$\|PBPv\| \leq (\beta + \gamma)\|v\| , \quad (6)$$

for all v then we could combine this with (5) and conclude that

$$\|p^i\| \leq (\beta + \gamma)^i \|p^0\| = (\beta + \gamma)^i / \sqrt{N} ,$$

from which it, using Cauchy-Schwarz, follows that

$$\sum_{r \in \{0,1\}^R} p_r^i \leq \left(\sum_{r \in \{0,1\}^R} 1 \right)^{1/2} \left(\sum_{r \in \{0,1\}^R} (p_r^i)^2 \right)^{1/2} = \sqrt{N} \|p^i\| \leq (\beta + \gamma)^i .$$

Thus, all that remains is to prove (6). Define $u = (1/\sqrt{N}, \dots, 1/\sqrt{N})$. Then we can write

$$Pv = v^{\parallel} + v^{\perp}$$

where $v^{\parallel}$ is parallel to u and $v^{\perp}$ is orthogonal to u . From linearity follows that

$$PBPv = PBv^{\parallel} + PBv^{\perp} .$$

We apply the triangle inequality to the above and get

$$\|PBPv\| \leq \|PBv^{\parallel}\| + \|PBv^{\perp}\| . \quad (7)$$

Then we bound the two terms on the right. To start with we have

$$\|PBv^{\perp}\| \leq \|Bv^{\perp}\| \leq \gamma \|v^{\perp}\| \leq \gamma \|v\| .$$

The first and last follows from orthogonality and the fact that $\|Pw\| \leq \|w\|$ holds for all projections P and $w \in \mathbb{R}^{2^R}$. The middle inequality follows, since γ is the second largest eigenvalue of B and $v^{\perp}$ is orthogonal to the eigenvector u corresponding to the largest eigenvalue 1 of B .

We now turn to the second term on the right in (7). We have $PBv^{\parallel} = Pv^{\parallel}$, since $v^{\parallel}$ is parallel to u and the eigenvalue of u is 1. Combined with the definition of P we have, using Cauchy-Schwarz, that

$$\|PBv^{\parallel}\| = \|Pv^{\parallel}\| \leq \sqrt{\beta} \|v^{\parallel}\| .$$

To finish the proof we show that $\|v^{\parallel}\| \leq \sqrt{\beta} \|v\|$. We have

$$\begin{aligned} \|v^{\parallel}\| &= (v^{\parallel}, u) = (Pv, u) = \sum_{i \in S} \frac{1}{\sqrt{N}} v_i \\ &\leq \left(\sum_{i \in S} v_i^2 \right)^{1/2} \left(\sum_{i \in S} \frac{1}{N} \right)^{1/2} \leq \|v\| \sqrt{\beta} . \end{aligned}$$

□