

Lecture 18 & 19: MSS and Spectrally Thin Trees

Lecturer: Shayan Oveis Gharan

June 3rd and 8th

Disclaimer: *These notes have not been subjected to the usual scrutiny reserved for formal publications.*

In this lecture we use the extension of Marcus, Spielman and Srivastava result to prove an upper bound of $\text{polyloglog}(n)$ on the integrality gap of the Held-Karp relaxation for ATSP. The materials of this lecture is based on the work of Oveis Gharan and Anari [AO15]. The following is the main technical theorem.

Theorem 18.1. *Any k -edge connected graph has a $O(\text{polyloglog}(n)/k)$ -thin spanning tree.*

Throughout this lecture we assume that G is an (unweighted) k connected graph.

The following theorem that follows from an extension of MSS that we discussed in Lecture 15 is our main tool.

Theorem 18.2. *Given a graph $G = (V, E)$ and a set $F \subseteq E$ such that (V, F) is k -edge connected. If we assign a vector x_e to any edge $e \in F$ such that $\|x_e\|^2 \leq \epsilon$ and*

$$\sum_{e \in F} x_e x_e^\top \preceq I,$$

then there exists a tree $T \subseteq F$ such that

$$\left\| \sum_{e \in T} x_e x_e^\top \right\| \leq O(\epsilon + 1/k).$$

Let us start with some simple ideas. First, let $x_e \propto b_e$ normalized such that $\sum_e x_e x_e^\top \preceq I$, e.g., let $x_e = L_G^{\dagger/2} b_e$. Then, by the above theorem, if for every edge

$$\|x_e\|^2 = b_e^\top L_G^\dagger b_e = \text{Reff}(e)$$

is small, then G has a spanning tree T where

$$\left\| \sum_{e \in T} x_e x_e^\top \right\| = \left\| L_G^{\dagger/2} L_T L_G^{\dagger/2} \right\| \leq O(1/k + \max_e \text{Reff}(e)).$$

Such a tree is (spectrally) thin.

Unfortunately, the above idea does not work in general because many edges of G may have large effective resistance. In other words, k -connectivity does not imply small effective resistance. In fact, the converse of this is true as we saw in Problem 4 of Assignment 2.

The second idea is to utilize **Theorem 18.2**, and let F be the edges of small effective resistance. It follows that if F has several edges in every cut, then G has a spectrally thin tree. Such a claim is not totally out of the question. This is because the average effective resistance of the edges of G is $O(1/k)$.

$$\text{avg}_e \text{Reff}(e) = \frac{n-1}{|E|} \leq \frac{n-1}{nk/2} = O(1/k),$$

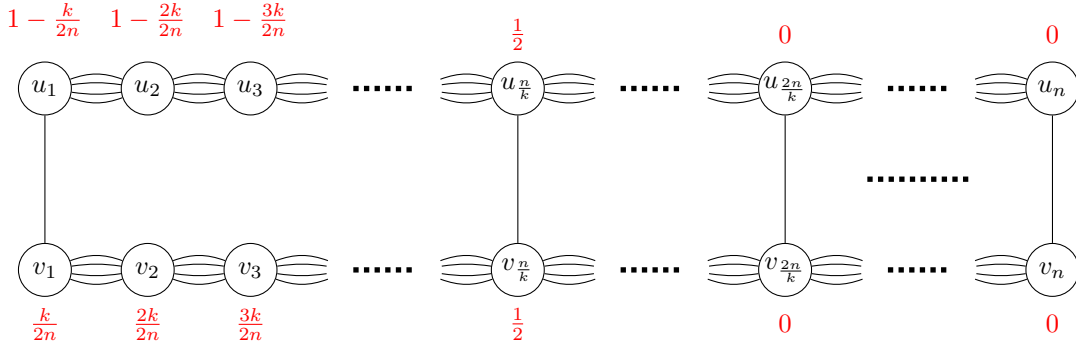


Figure 18.1: In this figure every vertical edge has effective resistance about $1 - k^2/n$ which is about 1 for small values of k . In other words, if we send one unit of electrical flow from u_1 to v_1 then at least $1 - k^2/n$ of the flow would go along the edge (u_1, v_1) , or in other words the energy of the flow is very close to 1. This is because $\text{Reff}(v_1, v_{n/k}) = \text{Reff}(u_1, u_{n/k}) \approx n/k^2$. Although there are k parallel paths from u_1 to v_1 they are very far from each other so $\text{Reff}(u_1, u_{n/k})$ is very large. So, the above graph has no spectrally thin tree

where the inequality uses that the degree of every vertex is at least k since G is k -connected. So, by choosing the right constants, we can make sure that F has at least 99% of the edges of G . Although this does not show that F has at least $\Omega(k)$ edges in every cut, it shows that F has an $\Omega(k)$ -connected subgraph with at most $n/10$ connected components. Using [Theorem 18.2](#) we can show that G has a linear size spectrally thin forest. Recall that we already proved this statement in Lecture 13 using the techniques of [\[BSS14\]](#).

Unfortunately, we can not expect F to have at several edges in *every* cut. Consider the graph [Figure 18.1](#). In this graph every vertical edge has effective resistance very close to 1. So, there is a cut with no edge of small effective resistance. We can make this example much worse by constructing a k -connected graph with vertex set $V = V_1 \cup V_2 \cup \dots \cup V_\ell$ for $\ell \approx n/\text{poly}(k)$ such that every edge between V_i 's have effective resistance very close to 1.

As we mentioned in Lecture 13, when there are no edges of small effective resistance in a cut (S, \bar{S}) of G then G has no spectrally thin tree. For example, in the graph of [Figure 18.1](#) say

$$T = \{(u_1, v_1), (u_1, u_2), \dots, (u_{n-1}, u_n), (v_1, v_2), \dots, (v_{n-1}, v_n)\}. \quad (18.1)$$

Although T is $1/k$ -(combinatorially) thin, it is only $1 - k/n$ -spectrally thin. Consider the vector x corresponding to red numbers written next to each vertex. It is easy to see that

$$\begin{aligned} x^\top L_T x &\approx 1 + O(k/n), \\ x^\top L_G x &\approx 1 + O(k^2/n). \end{aligned}$$

So, T is not spectrally thin.

In general, as alluded to in Lecture 13, the spectral thinness of any spanning tree is at least the maximum effective resistance of edges of T .

Lemma 18.3. *For any graph $G = (V, E)$, the spectral thinness of any spanning tree $T \subseteq E$ is at least $\max_{e \in T} \text{Reff}_{L_G}(e)$.*

Proof. Say the spectral thinness of T is α , i.e., $L_T \preceq \alpha \cdot L_G$. Obviously, by the downward closedness of spectral thinness, the spectral thinness of any subset of edges of T is at most α , so, for any edge $e \in T$,

$$b_e b_e^\top = L_e \preceq L_T \preceq \alpha \cdot L_G.$$

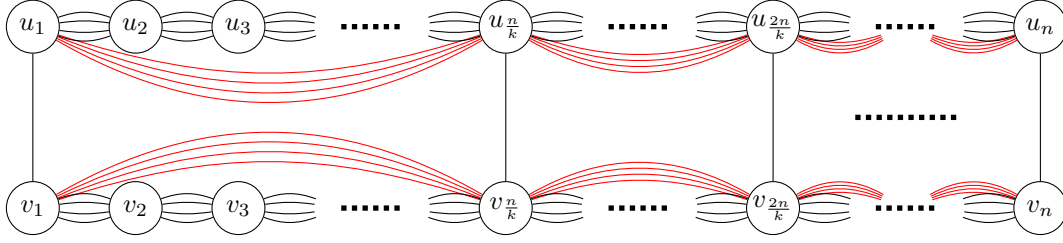


Figure 18.2: Here we added k shortcut edges between each consecutive pair of vertical edges of Figure 18.1. The size of every cut is at most twice of what it was before. But, the effective resistance of every black edge is $O(1/\sqrt{k})$. It is easy to see that in this graph the Hamiltonian path $v_n, v_{n-1}, \dots, v_1, u_1, u_2, \dots, u_n$ is an $O(1/\sqrt{n})$ spectrally thin tree.

But, the spectral thinness of an edge is indeed its effective resistance. More precisely, multiplying $L_G^{\dagger/2}$ on both sides of the above inequality we get

$$L_G^{\dagger/2} b_e b_e^\top L_G^{\dagger/2} = L_G^{\dagger/2} L_{\{e\}} L_G^{\dagger/2} \preceq \alpha \cdot L_G^{\dagger/2} L_G L_G^{\dagger/2} \preceq \alpha \cdot I.$$

Since the matrix in the LHS has rank one, its only eigenvalue is equal to its trace; therefore,

$$\text{Tr}(L_G^{\dagger/2} b_e b_e^\top L_G^{\dagger/2}) = b_e^\top L_G^\dagger b_e = \text{Reff}(e) \leq \alpha.$$

□

So, to prove Theorem 18.1 we have to get rid edges with large effective resistance. The basic idea is simple and similar to what we discussed in the last lecture. Observe that edges of G have small effective resistance if there are many short paths between the endpoints of each edge of G , in that case when we send on unit of electrical flow between the endpoints of an edge the electricity would distribute equally on the short paths so the energy would be much smaller than 1 (see Lectures 4 and 5 for the properties of effective resistance). On the other hand, k -connectivity is equivalent to having many edge disjoint paths between each pair of vertices of G . Unfortunately, these paths may be long, as in the example of Figure 18.1, so the effective resistances are not necessarily small. The basic idea is to use a semidefinite programming to shortcut the long paths in order to reduce the effective resistances while not changing the structure of cuts of G . This is analogous to the correction of the cycle graph that we talked about in Lecture 17.

For example, if we add k shortcut edges between each consecutive pair of vertical edges as shown in Figure 18.2, then the effective resistance of every vertical edge will be $O(1/\sqrt{k})$. Call this new graph H . In addition, the size of every cut in H is at most twice of what it was in G . Now, the tree T of (18.1) that was not a spectrally thin tree with respect to G , is $O(1/\sqrt{k})$ spectrally thin with respect to H

$$L_T \preceq O(1/\sqrt{k}) L_H.$$

But, since the size of every cut of H is at most twice of G , L_T is also $O(1/\sqrt{k})$ (combinatorially) thin with respect to G .

Now, the main question that we should answer is how to choose these shortcut edges, and if we can always reduce the effective resistance of edges without violating the size of the cuts.

18.1 Bounded Degree Spanning Trees

In this section we use the above idea to give a simple proof that any k -edge-connected graph has a tree that is thin only with respect to the degree cuts.

Lemma 18.4. Any k -edge connected graph $G = (V, E)$ has a spanning tree $T \subseteq E$ such that for any $v \in V$,

$$d_T(v) \leq O(1/k)d_G(v),$$

where $d_T(v), d_G(v)$ are the degree of v in T and G respectively.

Let $T_1, \dots, T_{k/2}$ be $k/2$ edge disjoint spanning trees of G and let z be the average of these trees,

$$z = \frac{1}{k/2} \sum_{i=1}^{k/2} \mathbf{1}_{T_i}.$$

Note that z is in the spanning tree polytope by definition.

In two beautiful works Goemans [Goe06] and Lau-Singh [LS15] show that for any point z in the spanning tree polytope there is an *integral* spanning tree such that for any vertex v ,

$$d_T(v) \leq 1 + \left\lceil \sum_{e \sim v} z(e) \right\rceil.$$

Since the fractional degree of every vertex in z is $O(d_G(v)/k)$ this proves the above lemma. The seminal work of [LS15] exploit the iterated rounding method in a clever way. Here, we would like to give a *spectral* proof of the above lemma using the technique that we mentioned in the previous section.

The idea is to add edges to G and construct a new graph H where the effective resistance of every edge of G is $O(1/k)$ in H and then use Theorem 18.2. Since we are only interested in thinness with respect to the degree cuts it is enough that

$$|E_H(\{v\}, \overline{\{v\}})| \leq 2|E_G(\{v\}, \overline{\{v\}})|,$$

for every vertex v . This makes our life much very easy. Let $D = (V, E')$ be a complete graph on V where the weight of every edge is $\frac{k}{n-1}$ and let $H = G + D$, i.e., the edges of H is just a union of the edges of G, D .

First, since the (fractional) degree of every vertex is k in D , and by k -connectivity of G the degree of every vertex of G is at least k , the above equation holds obviously. Secondly, the effective resistance of each pair of vertices in D is $O(1/k)$. To see that first note that the effective resistance of each pair of vertices in a complete graph on n vertices is $O(1/n)$ because there are $n-1$ edge disjoint paths of length at most 2. Since D is $k/(n-1)$ fraction of a complete graph, the effective resistances are $O(1/k)$. But, by the monotonicity property of effective resistance, the effective resistance of each pair of vertices in $G + D$ can only be smaller, so for any edge $e \in E$,

$$\text{Reff}_H(e) \leq O(1/k).$$

Now, let $x_e = (L_G + L_D)^{\dagger/2} b_e$ for any edge $e \in E$. Then

$$\|x_e\|^2 = b_e (L_G + L_D)^{\dagger} b_e \leq O(1/k),$$

and

$$\sum_{e \in E} x_e x_e^{\top} = (L_G + L_D)^{\dagger/2} \sum_{e \in E} b_e b_e^{\top} (L_G + L_D)^{\dagger/2} = (L_G + L_D)^{\dagger/2} L_G (L_G + L_D)^{\dagger/2} \preceq I.$$

Therefore, by Theorem 18.2, there is a spanning tree $T \subseteq E$ such that

$$L_T \preceq O(1/k) L_H.$$

Since, the degree of every vertex in H is $k + d_G(v)/k$, the degree of every vertex of T is at most $O(1 + d_G(v)/k)$ as desired.

Note that although the tree T is $O(1/k)$ -(spectrally) with respect to H it may have almost all edges of some of the cuts of G . This is because there is no guarantee on the sizes of the cuts of H which are not singleton cuts.

18.2 The Main Idea

The main idea is to construct a graph D such that the value of every cut $D(S, \bar{S})$ is at most $G(S, \bar{S})$

$$L_D \preceq_C L_G$$

and such that in every cut of G there are at least k edges with small effective resistance with respect to D . Recall that similar to the previous lecture we write, $A \preceq_C B$ to denote

$$\mathbf{1}_S^T A \mathbf{1}_S \leq \mathbf{1}_S^T B \mathbf{1}_S, \forall S \subseteq V.$$

Then, we let $F \subseteq E$ to be the edges of small effective resistance and we use D to assign a vector $x_e = (L_D + L_G)^{\dagger/2} b_e$ to every edge of G . It is easy to see that these vectors have small norm and are sub-isotropic. So, by [Theorem 18.2](#) there is a spanning tree T that is spectrally thin with respect to $D + G$,

$$L_T \preceq O(1/k)L_G + L_D,$$

But since $L_D \preceq_C L_G$, any spectrally thin tree of $G + D$ is combinatorially thin with respect to G ,

$$L_T \preceq_C O(1/k)L_G.$$

We would like to think of D as a demand graph of a multicommodity flow with congestion 1 on the edges of G . This is how we bypass the spectral thinness barrier. The graph D symmetrizes the spectrum of G , i.e., reduces the effective resistance of edges while preserving its L_1 structure, i.e., the cut structure. Although G does not necessarily have a spectrally thin tree, $G + D$ does. This is analogous to using multicommodity flows to bypass the limitation of Cheeger's inequality in approximating the sparsest cut as we discussed in the last lecture.

The following is the main theorem of [\[AO15\]](#) that we would like to prove.

Theorem 18.5. *For any $k \gtrsim \log(n)$ -connected graph $G = (V, E)$, there is a PD matrix $0 \prec D \preceq_C L_G$ and a set $F \subseteq E$ such that (V, F) is $\Omega(k)$ -connected and for any edge $e \in F$,*

$$\text{Reff}_D(e) = b_e^T D^{-1} b_e \leq \tilde{O}(1/k).$$

Note that in the above theorem D is not a graph, it is rather a PD matrix. It turns out that for our purpose of constructing vectors x_e as the input of [Theorem 18.2](#) that is enough, we can simply let $x_e = (L_G + D)^{-1} b_e$. We will see later how generalizing to a matrix can help us proving the above theorem.

The lower bound $\Omega(\log(n))$ on k is because of a limitation of the techniques in [\[AO15\]](#). It is expected that the above theorem holds for any value of k not depending on n . Such a result would give improved bound on the integrality gap of the Held-Karp relaxation.

It is now evident that if the optimum value of the above theorem implies [Theorem 18.1](#) using [Theorem 18.2](#). So, from now on we will only talk about the above theorem.

The main question that we need to answer is how to construct D, F or to prove their existence. Building the ideas of ARV, we can write a SDP to minimize the effective resistance by finding a multicommodity with

congestion 1 on the edges of G . Let $\text{kmin}\{.\}$ be the k -th smallest number in a set.

$$\begin{aligned}
& \min \quad \gamma \\
& \text{subject to} \quad \text{kmin}\{e \in (S, \bar{S}) : b_e^\top L_D^\dagger b_e\} \leq \gamma, \quad \forall S \subset V, \\
& \quad L_D = \sum_{u,v} \sum_{P \in P_{u,v}} y_P L_{u,v} \\
& \quad \sum_{P: e \in P} y_P \leq 1 \quad \forall e, \\
& \quad y_P \geq 0 \quad \forall P.
\end{aligned} \tag{18.2}$$

Unfortunately, the above program is not convex. First, note that $b_e^\top L_D^\dagger b_e$ is a convex function of L_D as we saw in Problem 5 of Assignment 2. But the minimum (or the kmin) of a set of number is a concave function. So, let us relax this constraint by reducing the maximum effective resistance of edges.

$$\begin{aligned}
& \min \quad \gamma \\
& \text{subject to} \quad b_e^\top L_D^\dagger b_e \leq \gamma, \quad \forall e \in E, \\
& \quad L_D = \sum_{u,v} \sum_{P \in P_{u,v}} y_P L_{u,v} \\
& \quad \sum_{P: e \in P} y_P \leq 1 \quad \forall e, \\
& \quad y_P \geq 0 \quad \forall P.
\end{aligned}$$

Note that $b_e^\top L_D^\dagger b_e \leq \gamma$ if and only if

$$\gamma \cdot b_e^\top b_e = \gamma \cdot L_e \preceq L_D.$$

So, in a sense the above convex program is very similar to the dual of ARV. We want to route a flow in G such that every edge is spectrally thin with respect to the demand graph. In ARV we wanted to route a flow such that the complete graph K_n is spectrally thin with respect to the demand graph.

Since we are looking for existential results we can simplify the above SDP, instead of looking for a D that is routable in G we can simply impose a constraint $L_D \preceq_C L_G$. In addition, as alluded to above, D does not have to be a graph, it is sufficient that D is PD.

$$\begin{aligned}
& \min \quad \gamma \\
& \text{subject to} \quad \max_{e \in E} b_e^\top D^{-1} b_e \leq \gamma, \\
& \quad D \preceq_C L_G, \\
& \quad D \succ 0.
\end{aligned} \tag{18.3}$$

Unfortunately, the optimum value of the above SDP can be very close to 1 even if G is $\log(n)$ -connected. As we will see this is mainly because we relaxed the $\text{kmin}\{.\}$ constraint to a $\max\{.\}$. If G is the k -connected graph shown in [Figure 18.3](#), and $k \leq h = \log(n)$, for any PD matrix $D \preceq_C L_G$, there is an edge with effective resistance very close to 1.

Theorem 18.6. *The optimum of (18.3) is $\Omega(1)$ for the k -connected graph of [Figure 18.3](#) when $k \leq \log(n)$.*

We prove the above theorem by constructing a solution of large value for the dual of (18.3). We emphasize that the graph of [Figure 18.3](#) is also a bad example for using electrical flows to solve maximum flow problems as is done in the seminal work of Christiano, Kelner, Madry, Spielman and Teng [\[Chr+11\]](#). If we send $\log(n)$ units of electrical flow from vertex 0 to 2^h then most of the flow would go through the long edges connecting the two. So, the electrical flow has a congestion of $\log(n)$ where as the maximum flow would have a congestion of 1.

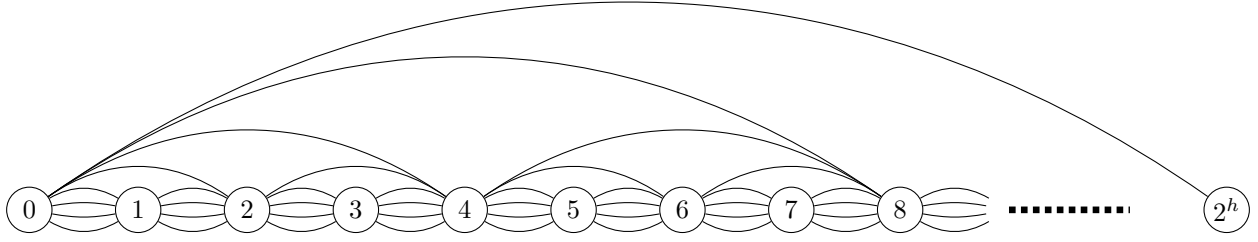


Figure 18.3: This graph has $2^h + 1$ vertices labeled with $\{0, 1, \dots, 2^h\}$. There are k parallel edges connecting each pair of consecutive vertices. In addition, for any $1 \leq i \leq h$ and any $0 \leq j < 2^{h-i}$ there is an edge $\{j \cdot 2^i, (j+1) \cdot 2^i\}$.

18.3 The Dual of SDP (18.3)

In this section first we write the dual of (18.3) and then we use it to prove Theorem 18.6.

First, we show SDP (18.3) satisfies Slater's condition, i.e., that (18.3) has a nonempty interior which implies that the duality gap is 0. It is easy to see that $D = \frac{1}{2}L_G + \frac{1}{3n^2}J$ is a positive definite matrix that satisfies all constraints strictly. In particular, since G is connected, for any set S , $\mathbf{1}_S^\top L_G \mathbf{1}_S \geq 1$, so

$$\frac{1}{3n^2} \mathbf{1}_S^\top J \mathbf{1}_S \leq \frac{1}{3} < \frac{1}{2} \mathbf{1}_S^\top L_G \mathbf{1}_S.$$

Therefore, $\mathbf{1}_S^\top D \mathbf{1}_S < \mathbf{1}_S^\top L_G \mathbf{1}_S$ for all S . Hence, Slater's condition is satisfied, and the strong duality is satisfied and the primal optimum (see Lecture 7 for more information).

For every $t \in T$ we associate a Lagrange multiplier w_e corresponding to the constraint $b_e^\top D^{-1} b_e \leq \gamma$, and for every set S we associate a nonnegative Lagrange multiplier y_S corresponding to the constraint $\mathbf{1}_S^\top D \mathbf{1}_S \leq \mathbf{1}_S^\top L_G \mathbf{1}_S$. The Lagrangian is defined as follows:

$$g(w, y) = \inf_{D \succ 0, \gamma} \gamma + \sum_{e \in E} w_e (b_e^\top D^{-1} b_e - \gamma) + \sum_{S \subset V} y_S (\mathbf{1}_S^\top D \mathbf{1}_S - \mathbf{1}_S^\top L_G \mathbf{1}_S)$$

Differentiating $g(w, y)$ w.r.t. γ we obtain that

$$\sum_{e \in E} w_e = 1. \tag{18.4}$$

Let

$$\begin{aligned} A &:= B^\top W B = \sum_e w_e b_e b_e^\top, \\ Z &:= \sum_{\emptyset \subset S \subset V} y_S \mathbf{1}_S \mathbf{1}_S^\top. \end{aligned}$$

where as usual B has the vectors b_e as its rows,

$$B = \begin{pmatrix} b_{e_1} \\ \vdots \\ b_{e_m} \end{pmatrix}$$

and W is the diagonal matrix of the edge weights.

Note that by definition A and Z are symmetric PSD matrices. So, the Lagrangian simplifies to

$$g(A, Z) = \inf_{D \succ 0} A \bullet D^{-1} + Z \bullet D - Z \bullet L_G,$$

subject to $\sum_e w_e = 1$. Now, we find the optimum D for fixed A, Z . First, we can assume that A and Z are nonsingular. This is without loss of generality by the continuity of $g(\cdot)$ and because the assumption $\sum_e w_e = 1$ can be satisfied by adding arbitrarily small perturbations. Differentiating with respect to D we obtain

$$D^{-1}AD^{-1} = Z.$$

Since, A, D are nonsingular there is a unique solution to the above equation,

$$D = Z^{-1/2}(Z^{1/2}AZ^{1/2})^{1/2}Z^{-1/2}$$

This is the solution which makes $A \bullet D^{-1} = Z \bullet D$. It is easy to see such a solution is an optimal solution [SLB74]. Using

$$D^{-1} = Z^{1/2}(Z^{1/2}AZ^{1/2})^{-1/2}Z^{1/2},$$

we have

$$\begin{aligned} A \bullet D^{-1} + Z \bullet D &= \text{Tr}(AZ^{1/2}(Z^{1/2}AZ^{1/2})^{-1/2}Z^{1/2}) + \text{Tr}(Z^{1/2}(Z^{1/2}AZ^{1/2})^{1/2}Z^{-1/2}) \\ &= 2 \text{Tr}((Z^{1/2}AZ^{1/2})^{1/2}). \end{aligned}$$

Therefore,

$$g(A, Z) = 2 \text{Tr}((Z^{1/2}AZ^{1/2})^{1/2}) - Z \bullet L_G$$

Let γ^* be the optimum value of (18.3). By the strong duality,

$$\gamma^* = \sup_{w, y \geq 0, \sum_e w_e = 1} g(A, Z) = \sup_{w, y \geq 0, \sum_e w_e = 1} 2 \text{Tr}((Z^{1/2}AZ^{1/2})^{1/2}) - Z \bullet L_G.$$

To make the notation simpler, from now on we assume $\sum_e w_e^2 = 1$ and we let W be the diagonal matrix of these new weights that are square root of what we had before. With this notation

$$Z^{1/2}AZ^{1/2} = Z^{1/2}B^\top W^2 BZ^{1/2}.$$

Recall that for any (nonsymmetric) matrix C , the nonzero eigenvalues of CC^\top are the same as the nonzero eigenvalues of $C^\top C$ (this follows from the SVD decomposition). So, the nonzero eigenvalues of $Z^{1/2}B^\top W W BZ^{1/2}$ are the same as the nonzero eigenvalues of $W^{1/2}BZB^\top W^{1/2}$. In addition, $Z \bullet L_G$ scales linearly with y while $\text{Tr}((WBZB^\top W)^{1/2})$ scales with \sqrt{y} . Optimizing the scaling of y we have,

$$\gamma^* = \sup_{w, y \geq 0, \sum_e w_e^2 = 1} 2 \text{Tr}((WBZB^\top W)^{1/2}) - Z \bullet L_G = \sup_{w, y \geq 0, \sum_e w_e^2 = 1} \frac{\text{Tr}((WBZB^\top W)^{1/2})^2}{Z \bullet L_G}.$$

Observe that although for a fixed W the quantity at the middle is convex, the RHS is no longer a convex function, because we are maximizing a quadratic form. But, we would like to work with the RHS quantity because it is scale free.

Cut Metrics. Write $Z = X^\top X$ where $X \in \mathbb{R}^{2^n \times n}$ and each row of X corresponds to a vector $y_S \mathbf{1}_S$ for a set $S \subseteq V$. Such a matrix X defines a weighted cut metric on the vertices of G . Let X_v be the v column of X . We have

$$\|X_u - X_v\|_1 = \sum_S y_S |\mathbf{1}_S(u) - \mathbf{1}_S(v)| = \mathbb{P}[u \text{ and } v \text{ are in different sides}].$$

It is easy to see that any weighted cut metric can be embedded into an unweighted cut metric simply by having multiple copies of each cut. For example, a cut metric of the form $\frac{3}{4}\mathbf{1}_{S_1} + \frac{1}{4}\mathbf{1}_{S_2}$ can be written as follows

$$\begin{pmatrix} \frac{3}{4}\mathbf{1}_{S_1} \\ \frac{1}{4}\mathbf{1}_{S_2} \end{pmatrix} = \frac{1}{4} \begin{pmatrix} \mathbf{1}_{S_1} \\ \mathbf{1}_{S_1} \\ \mathbf{1}_{S_1} \\ \mathbf{1}_{S_2} \end{pmatrix}.$$

Since Z is scale free, we can drop the scalar $1/4$. So, in general, we can assume $X \in \{0,1\}^{h \times n}$ for an h possibly larger than 2^n . To have a good intuition, the mapping X is embedding our graph into the vertices of a giant hypercube. With such a normalization we have the following useful identity

$$\|X_u - X_v\|_1 = \|X_u - X_v\|_2^2, \quad \forall u, v \in V. \quad (18.5)$$

We can write the objective as follows

$$\gamma^* = \sup_{w, \sum w^2 = 1, X \in \{0,1\}^{n \times h}} \frac{\text{Tr}((WBX^\top XB^\top W)^{1/2})^2}{\sum_{u \sim v} \|X_u - X_v\|^2}$$

where in the denominator we used that

$$Z \bullet L_G = \sum_{u \sim v} \|X_u - X_v\|^2,$$

as we shown in the last lecture.

The Nuclear Norm. So, all we need to do is to relate $\text{Tr}((WBX^\top XB^\top W)^{1/2})$ to the properties of the mapping X . The main difficulty is to deal with the eigenvalues of the square root of $WBX^\top XB^\top W$.

First observe that the eigenvalues of $(WBX^\top XB^\top W)^{1/2}$ are the same as the singular values of the matrix $XB^\top W$ (see Problem 4 of Assignment 4 for the properties of the singular values of a matrix). Unlike symmetric matrices, $\text{Tr}(XB^\top W)$ is not equal to the sum of its singular values. Sum of the singular values of a matrix is also known as the *nuclear norm* of that matrix and it has many applications in sparse recovery and compressed sensing (see e.g., [RFP10]),

$$\text{Tr}((WBX^\top XB^\top W)^{1/2}) = \|XB^\top W\|_*.$$

As mentioned above $\|XB^\top W\|_* \neq \text{Tr}(XB^\top W)$. But under the best rotation of the matrix in the space, the trace will give us the sum of the singular values. See part (f) of Problem 4 of Assignment 4 for the proof. That is we have,

$$\|XB^\top W\|_* = \sup_{\text{unitary } U} U \bullet XB^\top W$$

Therefore,

$$\gamma^* = \sup_{\substack{w \geq 0, \sum_e w_e^2 = 1 \\ X \in \{0,1\}^{h \times n}, \text{unitary } U}} \frac{(U \bullet XB^\top W)^2}{\sum_{\{u,v\} \in E} \|X_u - X_v\|_2^2}$$

Cauchy-Schwarz. Finally, we can translate $U \bullet XB^\top W$ into the metric language. Let e_1, \dots, e_m be the edges of G and let $b_{e_i} = \mathbf{1}_{u_i} - \mathbf{1}_{v_i}$. The matrix XB^\top has E columns where the i -th column is $X_{u_i} - X_{v_i}$. The matrix $XB^\top W$ also has E columns where the i -th column is $w_{e_i}(X_{u_i} - X_{v_i})$.

$$XB^\top W = \left(w_{e_1}(X_{u_1} - X_{v_1}) \mid w_{e_2}(X_{u_2} - X_{v_2}) \mid \dots \mid w_{e_m}(X_{u_m} - X_{v_m}) \right)$$

Therefore,

$$\gamma^* = \sup_{\substack{w \geq 0, \sum_e w_e^2 = 1 \\ X \in \{0,1\}^{h \times n}, \text{unitary } U}} \frac{(\sum_{i=1}^m w_{e_i} (UX_{u_i} - UX_{v_i})_i)^2}{\sum_{i=1}^m \|X_{u_i} - X_{v_i}\|^2}, \quad (18.6)$$

where $(UX_{u_i} - UX_{v_i})_i$ denotes the i -th coordinate of the vectors $UX_{u_i} - UX_{v_i}$.

Finally, using the Cauchy-Schwarz inequality we can write

$$\gamma^* = \sup_{\substack{w \geq 0, \sum_e w_e^2 = 1 \\ X \in \{0,1\}^{h \times n}, \text{unitary } U}} \frac{(\sum_{e \in E} w_e^2) \cdot \sum_{i=1}^m (UX_{u_i} - UX_{v_i})_i^2}{\sum_{i=1}^m \|X_{u_i} - X_{v_i}\|^2}$$

The other side of the above equality follows by letting

$$w_{e_i} \propto (UX_{u_i} - UX_{v_i})_i.$$

such that $\sum_e w_e^2 = 1$. The following theorem follows.

Theorem 18.7. Let $E = \{e_1, \dots, e_m\}$ be the edges of G , where for each i , $e_i = \{u_i, v_i\}$. The dual of (18.3) is as follows

$$\sup_{\substack{X \in \{0,1\}^{h \times n}, \\ \text{unitary } U \in \mathbb{R}^{h \times h}}} \frac{\sum_{i=1}^m (UX_{u_i} - UX_{v_i})_i^2}{\sum_{i=1}^m \|X_{u_i} - X_{v_i}\|^2} = \sup_{\substack{X \in \{0,1\}^{h \times n}, \\ \text{orthonormal } z_{e_1}, \dots, z_{e_m}}} \frac{\sum_{i=1}^m \langle z_{e_i}, X_{u_i} - X_{v_i} \rangle^2}{\sum_{i=1}^m \|X_{u_i} - X_{v_i}\|^2}. \quad (18.7)$$

As a sanity check observe that the optimal value of the dual is always at most 1. This is because for any edge and any unit vector z_{e_i} by Cauchy-Schwarz,

$$\langle z_{e_i}, X_{u_i} - X_{v_i} \rangle^2 \leq \|X_{u_i} - X_{v_i}\|^2.$$

18.4 Intuitions on the Dual

In this section we give several intuitions on the properties of the dual and we will prove [Theorem 18.6](#).

Two Player Game. It is instructive to think of a two player game: Say the first player chooses a matrix D , or a demand graph of a multicommodity flow that is routable in G such that $\max_e \text{Reff}_D(e)$ is small; the second player wants to show that the maximum effective resistance is large w.r.t. D , so he chooses the potential vector corresponding to the largest effective resistance as his proof. The optimum strategy for the second player is to show that there are weights w_e for the edges and a set of vectors p_e such that the sum of the potential differences of the endpoints of all edges e scaled by w_e is large, i.e.,

$$\left(\sum_e w_e \langle p_e, b_e \rangle \right)^2$$

is large. The second player encodes the potential vectors in the cut metric X and uses orthonormal vectors z_e to denote the coordinates of the potential vector chosen for each edge e . Note that the weights w_e are not present because of an application of Cauchy-Schwarz inequality in (18.7).

Geometric View. Let us also understand the dual geometrically. Since X is a cut metric, it maps the vertices of G to a huge hypercube. Think of the z_1, \dots, z_m as the coordinate system of the space. The objective is to find a coordinate system of the space such that the sum of the square of the projections of the edges on the corresponding coordinates is as large as possible

$$\max_{\text{orthonormal } z_1, \dots, z_m} \sum_{i=1}^m \langle z_{e_i}, X_{u_i} - X_{v_i} \rangle^2,$$

while the sum of the squared length of the edges under X is equal to 1.

The reason that we have a cut metric in the dual is because of the constraint

$$D \preceq_C L_G$$

in the primal. If instead we had a multicommodity flow constraint, then analogous to ARV relaxation, in the dual we would have had a L_2^2 metric. Recall that $X : V \rightarrow \mathbb{R}^h$ is a L_2^2 metric if for each triple of vertices u_1, u_2, u_3 ,

$$\|X_{u_1} - X_{u_2}\|^2 \leq \|X_{u_1} - X_{u_3}\|^2 + \|X_{u_3} - X_{u_2}\|^2.$$

Note that a proof that only exploits the squared triangle inequality can assume that $z_{e_i} = \mathbf{1}_i$ for all i . This is because the L_2 distances are invariant under a unitary transformation (see Problem 4 of Assignment 4). Therefore,

$$\sup_{X \text{ is } L_2^2, U \text{ is unitary}} \frac{\sum_{i=1}^m (UX_{u_i} - UX_{v_i})_i^2}{\sum_{i=1}^m \|X_{u_i} - X_{v_i}\|^2} = \sup_{Y \text{ is } L_2^2} \frac{\sum_{i=1}^m (Y_{u_i} - Y_{v_i})_i^2}{\sum_{i=1}^m \|Y_{u_i} - Y_{v_i}\|^2},$$

where $Y = UX$. As we will see both of the positive statements [Lemma 18.8](#) and [Proposition 18.10](#) only assume that X is a L_2^2 metric.

In the rest of this section we prove two simple statements about the dual and then we prove [Theorem 18.6](#). Let us start with an easy example.

Lemma 18.8. *For any k -connected graph G and $1 \leq j \leq m$,*

$$\left\{ \begin{array}{l} \min \quad b_{e_j}^\top D^{-1} b_{e_j}, \\ D \preceq_C L_G, \\ D \succ 0 \end{array} \right\} = \sup_{X \in \{0,1\}^{h \times n}, \|z_{e_j}\|=1} \frac{\langle z_{e_j}, X_{u_j} - X_{v_j} \rangle^2}{\sum_{i=1}^m \|X_{u_i} - X_{v_i}\|^2} \leq O(1/k),$$

i.e., we can always reduce the effective resistance of any edge of G to $1/k$ without violating the size of the cuts.

Proof. The proof is relatively simply. First note that in the worst case the vector z_{e_j} is parallel to $X_{u_j} - X_{v_j}$. Therefore, the numerator is exactly $\|X_{u_j} - X_{v_j}\|^2$. The proof simply follows from the triangle inequality of the cut metrics.

Since G is k -connected there are k edge disjoint paths from u_j to v_j . For any such path P we have

$$\sum_{e_\ell \in P} \|X_{u_\ell} - X_{v_\ell}\|^2 \geq \|X_{u_j} - X_{v_j}\|^2.$$

□

The above claim can also be proved relatively easy in G . Since G is k -connected we can simply shortcut the k edge disjoint paths connecting the endpoints of e_j and have these k edges in D . Then $\text{Reff}_D(e_j) = 1/k$ as desired and $D \preceq_C L_G$ because it is routable in G .

In the next example, we show that the optimum of (18.7) is $\Omega(1)$ if G is a simple cycle.

Lemma 18.9. *The optimum of (18.7) is $\Omega(1)$ if G is a cycle of length n .*

Proof. All we need to do is to construct a dual solution of large value. Let v_1, \dots, v_n be the vertices of G and for each i , $e_i = v_i, v_{i+1}$. Now $X_{v_i} \in \{0, 1\}^n$ be the vector where the first $i - 1$ coordinates are 1 and the rest are 0. In other words, X is just an average of all threshold cuts of G .

It follows that for any i ,

$$X_{v_{i+1}} - X_{v_i} = \mathbf{1}_i.$$

In addition,

$$\sum_{i=1}^n \|X_{v_{i+1}} - X_{v_i}\|^2 = 2n.$$

Finally, let $z_{\{i, i+1\}} = \mathbf{1}_i$ for all i . Then,

$$\langle z_{\{i, i+1\}}, X_{v_{i+1}} - X_{v_i} \rangle^2 = 1,$$

for all $1 \leq i < n$. Therefore, the optimum of (18.7) is at least $n/2n = 1/2$. \square

Although the statement of the above lemma is in a sense obvious the proof gives a simple rigorous argument of the fact.

Next, we prove **Theorem 18.6**.

Proof of Theorem 18.6. We define a mapping similar to the previous lemma. For vertex i , let X_i be the vector where the first i coordinates are 1 and the rest are 0. Then,

$$\sum_e \|X_{b_e}\|^2 = n \cdot k + n \cdot h$$

Next, we show that for a carefully chosen orthonormal vectors $\{z_e\}_{e \in E}$, $\sum_e \langle z_e, X_{b_e} \rangle^2 = \Theta(n \cdot h)$. Assuming $k \leq h$ this completes the proof.

Now, we need to construct orthonormal vectors one for each edge. We only assign vectors to the long edges. For each $1 \leq j \leq h$, we assign a vector to each edge $\{0, 2^j\}, \{2 \cdot 2^j, 3 \cdot 2^j\}, \{4 \cdot 2^j, 5 \cdot 2^j\}, \dots$ where for every the vector that we assign to $\{2i \cdot 2^j, (2i + 1) \cdot 2^j\}$ is defined as follows:

$$z_{\{2i \cdot 2^j, (2i+1) \cdot 2^j\}} = \begin{bmatrix} 1 & & 2i \cdot 2^j & & (2i+1) \cdot 2^j & & (2i+2) \cdot 2^j \\ 0 & \dots & \frac{+1}{\sqrt{2^j}} & \dots & \frac{-1}{\sqrt{2^j}} & \dots & 0 & \dots \end{bmatrix}$$

and we leave the rest of edges of layer j unassigned (or we may assign an arbitrary vectors orthogonal to all other vectors to them). The following matrix shows the vectors that we are using. Observe the vectors $z_{\{u, v\}}$ are orthonormal.

$$\begin{array}{l} z_{\{0, n\}} \\ z_{\{0, n/2\}} \\ z_{\{0, n/4\}} \\ z_{\{n/2, 3n/4\}} \end{array} \begin{bmatrix} 1 & & n/4 & & n/2 & & 3n/4 \\ \frac{+1}{\sqrt{n}} & \dots & \frac{+1}{\sqrt{n}} & \dots & \frac{+1}{\sqrt{n}} & \dots & \frac{+1}{\sqrt{n}} \\ \frac{+1}{\sqrt{n}} & \dots & \frac{+1}{\sqrt{n}} & \dots & \frac{-1}{\sqrt{n}} & \dots & \frac{-1}{\sqrt{n}} \\ \frac{+\sqrt{2}}{\sqrt{n}} & \dots & \frac{-\sqrt{2}}{\sqrt{n}} & \dots & 0 & \dots & 0 \\ 0 & \dots & 0 & \dots & \frac{+\sqrt{2}}{\sqrt{n}} & \dots & \frac{-\sqrt{2}}{\sqrt{n}} \end{bmatrix}$$

Now we are ready to bound the numerator.

$$\begin{aligned} \sum_{j=1}^h \sum_{e=\{2i \cdot 2^j, (2i+1) \cdot 2^j\}} \langle z_e, Xb_e \rangle^2 &= \sum_{j=1}^h \sum_{e=\{2i \cdot 2^j, (2i+1) \cdot 2^j\}} 2^j \\ &= \sum_{j=1}^h n/2 = n \cdot h/2. \end{aligned}$$

The lemma follows from the fact that the numerator is at least the above quantity. \square

18.5 Reducing the Average Effective Resistance

As we mentioned earlier, the main challenge in proving [Theorem 18.5](#) is that (18.2) is not a convex program. The closest we can get to the objective of this program is by minimizing average effective resistance over subsets of edges. In the following proposition we show that we can reduce the average effective resistance of any arbitrary subset of edges of G . But, unfortunately, this is not enough to bypass the $\log(n)$ barrier.

Proposition 18.10. *For k -connected graph $G = (V, E)$ and any set $F = \{e_1, e_2, \dots\} \subseteq E$ we have*

$$\left\{ \begin{array}{l} \min \mathbb{E}_{e \sim F} b_e D^{-1} b_e, \\ D = L_G + \sum_{u,v} \sum_{P \in P_{u,v}} y_P (L_{u,v} - L_P), \\ y_P \geq 0. \end{array} \right\} = \sup_{X \text{ is } L_2^2 \text{ metric}} \frac{\frac{1}{|F|} \left(\sum_{e_i \in F} \langle \mathbf{1}_i, Yb_{e_i} \rangle \right)^2}{\sum_{e \in E} \|Yb_{e_i}\|^2} \leq O\left(\frac{1}{\sqrt{k}}\right),$$

where we use $\mathbb{E}[\cdot]$ to denote the average under uniform distribution.

Note that we haven't shown the above supremum is the dual to the CP in the LHS but the proof immediately follows from the proof of [Theorem 18.7](#). We also would like to point out that with a more careful analysis one can show an upper bound of $\tilde{O}(1/k)$ on the value of the above convex program.

We emphasize that the proof of the above proposition does not exploit that X is a cut metric and it holds when X is just an L_2^2 metric. It is instructive to find a primal proof of the above statement. That is to construct a multicommodity flow whose demand graph reduces the average effective resistance of the edges of F .

Our main tool to upper bound the dual is by constructing disjoint balls. Suppose we have disjoint L_2^2 balls $B_1(X_{v_1}, r_1), \dots, B_\ell(X_{v_\ell}, r_\ell)$ each centered at one of the vertices. Since G is k connected there at least k edge disjoint paths from v_i to outside of the ball B_i . By the squared triangle inequality, the sum of the squared length of the edges of each of these paths is at least r_i . Since B_1, \dots, B_ℓ are disjoint this argument does not over count any of the edges of G . Therefore,

$$\sum_{i=1}^{\ell} r_i \cdot k \leq \sum_{e \in E} \|Xb_e\|^2.$$

See [Figure 18.4](#) for an example.

Fact 18.11. *For any $X : V \rightarrow \mathbb{R}^h$ that satisfies the squared triangle inequality and any set of disjoint balls B_1, \dots, B_ℓ centered at vertices of G such that the radius of B_i is r_i we have*

$$\sum_{i=1}^{\ell} r_i \cdot k \leq \sum_{e \in E} \|Xb_e\|^2.$$

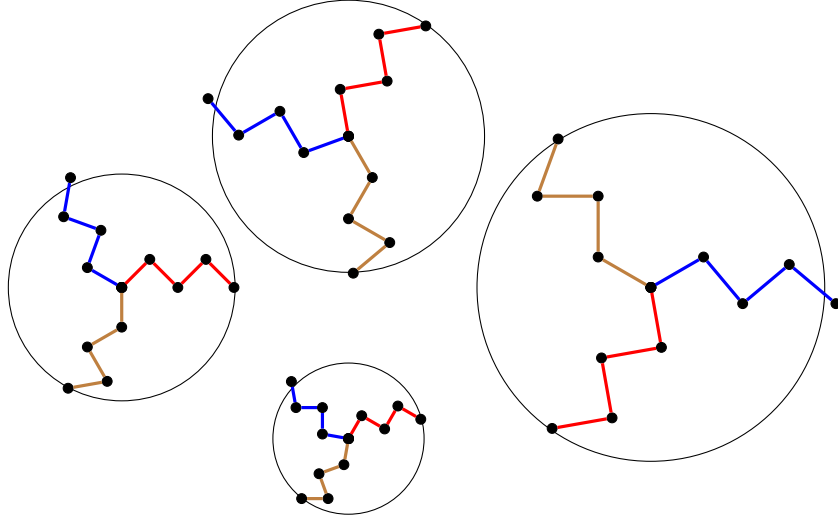


Figure 18.4: Since there are k edge disjoint paths connecting the center of each ball to the outside, by the triangle inequality, the sum of the squared length of the edges of the graph is at least k times the sum of the radii of the balls.

Therefore, to prove the above proposition it is enough to construct many disjoint balls such that the sum of their radii is at least $1/\sqrt{k}$ of $\frac{1}{|F|} \left(\sum_{e_i \in F} \langle \mathbf{1}_i, Yb_{e_i} \rangle \right)^2$. We use the following lemma to construct these balls.

Lemma 18.12. *For any $Y : V \rightarrow \mathbb{R}^h$ and $F = \{e_1, e_2, \dots\} \subseteq E$ such that*

$$(\mathbb{E}_{e_i \in F} \langle \mathbf{1}_i, Yb_{e_i} \rangle)^2 = \left(\frac{1}{|F|} \text{Tr}(YB_F^\top) \right)^2 \geq \alpha \cdot \frac{\|YB_F^\top\|_F^2}{|F|} = \alpha \cdot \mathbb{E}_{e_i \in F} \|Yb_{e_i}\|^2,$$

for some $0 < \alpha < 1$ where $\|\cdot\|_F$ is the Frobenius norm, there are b disjoint L_2^2 balls with (L_2^2) radius r such that the center of each ball is an endpoint of an edge of F and

$$b \cdot r \gtrsim \alpha \cdot |F| \cdot (\mathbb{E}_{e_i \in F} \langle \mathbf{1}_i, Yb_{e_i} \rangle)^2.$$

Proof. For a radius $r > 0$, run the following greedy algorithm to construct disjoint balls. Scan the endpoints of the edges in an arbitrary order; for each point Y_u , if the L_2^2 ball $B(Y_u, r)$ doesn't touch the balls that we have already construct, add $B(Y_u, r)$. Suppose we manage to select b balls. We say the algorithm succeeds if the lemma's conclusions is satisfied. In the rest of the proof we show that this algorithm always succeeds when $r \gtrsim (\mathbb{E}_{e_i \in F} \langle \mathbf{1}_i, Yb_{e_i} \rangle)^2$.

Let $\sigma_1, \dots, \sigma_{|F|}$ be the singular values of YB_F^\top . In the next claim, we show that if the above algorithm finds b balls for a value of r , that implies an upper bound on the singular values of YB_F^\top .

Claim 18.13. *For any $r > 0$, if the above greedy algorithm finds b disjoint balls of radius r , then*

$$r \geq \frac{1}{16|F|} \sum_{i=b+1}^{|F|} \sigma_i^2.$$

Proof. First, we construct a low-rank matrix $C \in \mathbb{R}^{h \times |F|}$ assuming b is small. Then, we use the following theorem to prove the claim.

Theorem 18.14 (Hoffman-Wielandt Inequality). *For any $A, B \in \mathbb{R}^{n \times n}$ with singular values $\sigma_1 \leq \sigma_2 \leq \dots, \sigma_n$ and $\sigma'_1 \leq \sigma'_2 \leq \dots \leq \sigma'_n$ we have,*

$$\sum_{i=1}^n (\sigma_i - \sigma'_i)^2 \leq \|A - B\|_F^2.$$

Let Y_{w_1}, \dots, Y_{w_b} be the centers of the chosen balls. For any endpoint v of an edge in F , let $c(v)$ be the closest center to Y_v , i.e.,

$$c(v) := \operatorname{argmin}_{w_i} \|Y_{w_i} - Y_v\|_2^2$$

We construct a matrix $C \in \mathbb{R}^{b \times |F|}$; say $b_{e_i} = \mathbf{1}_{u_i} - \mathbf{1}_{v_i}$, we let $Y_{c(u_i)} - Y_{c(v_i)}$ be i -th column of C . By definition, $\operatorname{rank}(C) \leq b$.

First, notice that

$$\begin{aligned} \|YB_F^\top - C\|_F^2 &= \sum_{\{u_i, v_i\} \in F} \|(Y_{u_i} - Y_{v_i}) - (Y_{c(u_i)} - Y_{c(v_i)})\|_2^2 \\ &\leq \sum_{\{u_i, v_i\} \in F} (\|Y_{u_i} - Y_{c(u_i)}\|_2 + \|Y_{v_i} - Y_{c(v_i)}\|_2)^2 \\ &\leq \sum_{\{u_i, v_i\} \in F} 2\|Y_{u_i} - Y_{c(u_i)}\|_2^2 + 2\|Y_{v_i} - Y_{c(v_i)}\|_2^2 \leq 16r \cdot |F|, \end{aligned}$$

where the first inequality follows by the triangle inequality and the last inequality follows by definition of greedy algorithm, i.e., $\|Y_v - Y_{c(v)}\|_2^2 \leq 4r$ for all endpoints of edges of F . Therefore, by [Theorem 18.14](#),

$$16r \cdot |F| \geq \|YB_F^\top - C\|_F^2 \geq \sum_{i=b+1}^{|F|} \sigma_i^2.$$

where the second inequality uses the fact that $\operatorname{rank}(C) \leq b$. □

First, by the lemma's assumption,

$$\frac{1}{|F|} \sum_{i=1}^{|F|} \sigma_i^2 = \frac{1}{|F|} \|YB_F^\top\|_F^2 \leq \frac{1}{\alpha} \left(\frac{\operatorname{Tr}(YB_F^\top)}{|F|} \right)^2 \leq \frac{1}{\alpha} \left(\frac{\sum_{i=1}^{|F|} \sigma_i}{|F|} \right)^2. \quad (18.8)$$

Note that the last follows by the characterization of the nuclear norm (see Problem 4 of Assignment 4). The above inequality is the inverse of the Cauchy-Schwarz inequality on the singular vectors of YB_F^\top . In particular, for $\alpha = 1$, the RHS is always less than or equal to the LHS and the equality occurs when the singular values of YB_F^\top are equal. Therefore, for values of α bounded away from zero the singular values of YB_F^\top are “almost” equal, so, using the above lemma we can argue that the number of balls b is large which completes the proof.

Let $C > 4$ be a constant and suppose we have found b balls for

$$\begin{aligned} r &= \frac{1}{C} (\mathbb{E}_{e_i \in F} \langle \mathbf{1}_i, Yb_{e_i} \rangle)^2 \\ b &\leq \frac{\alpha}{C} \cdot |F|. \end{aligned}$$

We will reach a contradiction with the Lemma's assumption. Note that if $b \geq \alpha|F|/C$, then, $r \cdot b \gtrsim \alpha \cdot |F| \cdot (\mathbb{E}_{e_i \in F} \langle \mathbf{1}_i, Yb_{e_i} \rangle)^2$ and we are done. By the above claim,

$$\frac{1}{|F|} \sum_{i=b+1}^{|F|} \sigma_i^2 \leq r = \frac{1}{C} (\mathbb{E}_{e_i \in F} \langle \mathbf{1}_i, Yb_{e_i} \rangle)^2 \leq \frac{1}{C} \left(\frac{1}{|F|} \sum_{i=1}^{|F|} \sigma_i \right)^2 \quad (18.9)$$

where the second inequality follows by the lemma's assumption. Therefore,

$$\begin{aligned} \left(\frac{1}{|F|} \sum_{i=1}^{|F|} \sigma_i \right)^2 &\leq 2 \left(\frac{1}{|F|} \sum_{i=1}^b \sigma_i \right)^2 + 2 \left(\frac{1}{|F|} \sum_{i=b+1}^{|F|} \sigma_i \right)^2 \\ &\leq \frac{2b}{|F|^2} \sum_{i=1}^b \sigma_i^2 + \frac{2}{|F|} \sum_{i=b+1}^{|F|} \sigma_i^2 \\ &\leq \frac{2\alpha}{C \cdot |F|} \sum_{i=1}^b \sigma_i^2 + \frac{2}{C} \left(\frac{1}{|F|} \sum_{i=1}^{|F|} \sigma_i \right)^2 \end{aligned}$$

where the second inequality follows by Cauchy-Schwarz, the third inequality uses (18.9) and the definition of b .

But for $C > 4$, we have

$$\frac{1}{2} \left(\frac{1}{|F|} \sum_{i=1}^{|F|} \sigma_i \right)^2 < \frac{2\alpha}{C \cdot |F|} \sum_{i=1}^{|F|} \sigma_i^2.$$

This is contradiction with (18.8) (and the lemma's assumption). This completes the proof of Lemma 18.12. \square

We remark that the dependency on α in the conclusion of the above lemma can be improved to α^ϵ at the cost of $1/\epsilon^2$ loss. We refer interested readers to [AO15, Lemma 5.2].

Now, we are ready to prove Proposition 18.10.

Proof of Proposition 18.10. Fix an L_2^2 metric X . If

$$\frac{\frac{1}{|F|} \left(\sum_{e_i \in F} \langle \mathbf{1}_i, Y b_{e_i} \rangle \right)^2}{\sum_{e \in E} \|Y b_e\|^2} \leq O(1/\sqrt{k}),$$

then we are done. So, assume the LHS is at least α for some $\alpha = O(1/\sqrt{k})$. By Lemma 18.12, we can construct b L_2^2 balls of radius r such that

$$b \cdot r \gtrsim \alpha \cdot |F| \left(\mathbb{E}_{e_i \in F} \langle \mathbf{1}_i, Y b_{e_i} \rangle \right)^2$$

But, by Fact 18.11 we have

$$b \cdot r \cdot k \leq \sum_{e \in E} \|Y b_e\|^2.$$

The proposition follows from the above two inequality. \square

If we repeatedly apply the above proposition we can reduce the maximum effective resistance of the edges of G to $\log(n)/\sqrt{k}$, and if we use the optimized statement we can reduce it to $\tilde{O}(\log(n)/k)$. Such a result is the best possible for the graph G illustrated in Figure 18.3 as we proved in Theorem 18.6. However, it is not possible to use the statement of the above proposition to prove Theorem 18.5.

Perhaps, the next idea that come to mind is to minimize the average effective resistance over all cuts of G , i.e., to show that the optimum of the following convex program is small.

$$\begin{aligned} \min \quad & \gamma, \\ & \mathbb{E}_{e \in E(S, \bar{S})} \text{Reff}_D(e) \leq \gamma, \quad \forall S \subset V, \\ & D \preceq_C L_G, \\ & D \succ 0. \end{aligned} \tag{18.10}$$

Unfortunately, the same example of [Figure 18.3](#) shows that the optimum of the above program can be $\Omega(1)$ even if G is $\log(n)$ -connected.

Nonetheless, in the next section, we show that we can reduce the average effective resistance of all the degree cuts of G . Unlike the proof of this section, we will take advantage of the L_1 structure of X .

We also note that using the ideas of the above theorem we can prove the following statement.

Theorem 18.15. *For any k -connected graph G , we have*

$$\left\{ \begin{array}{l} \min_{u,v} \max \text{Reff}_D(u,v), \\ D = L_G + \sum_{u,v} \sum_{P \in P_{u,v}} y_P (L_{u,v} - L_P), \\ y_P \geq 0. \end{array} \right\} \leq \frac{\text{polylog}(n)}{k},$$

i.e., we can reduce the maximum effective resistance of all pairs of vertices of G to $O(\text{polylog}(n)/k)$.

As an interesting application of the above theorem, one can route a multicommodity flow in a cycle (of length n) to reduce the effective resistance of all pairs of vertices of G to $O(\log^2 n)$.

18.6 Reducing the Average Effective Resistance of Degree Cuts

In the following theorem we show that we can reduce the *average* effective resistance of all of the degree cuts (a.k.a., singleton cuts) of G . This is the first step towards proving [Theorem 18.5](#). The proof of the following theorem has many of the ideas of the main proof. The proof of [Theorem 18.5](#) is tedious and beyond the scope of this course. We just point out that the high-level idea is to decompose the given k -connected graph G into expanders and use repeated application (of a stronger variant) of the following theorem to reduce the effective resistance of all of the degree cuts of the expanders. This is sufficient to prove [Theorem 18.5](#) because in an expander there are at least k edges of small effective resistance in every cut if the average effective resistance of all the degree cuts is small.

Theorem 18.16. *For any k -regular k -connected graph G ,*

$$\left\{ \begin{array}{l} \min \quad \gamma, \\ \mathbb{E}_{e \sim v} b_e D^{-1} b_e \leq \gamma \quad \forall v, \\ D \preceq_C L_G, \\ D \succ 0. \end{array} \right\} = \sup_{\substack{X \in \{0,1\}^{h \times n} \\ \{z_e\}_{e \in E} \text{ are orthonormal}}} \frac{\sum_{v \in V} (\mathbb{E}_{e \sim v} \langle z_e, X b_e \rangle)^2}{\sum_{v \in V} \mathbb{E}_{e \sim v} \|X b_e\|^2} \leq \tilde{O}(1/\sqrt{k}).$$

One of the interesting consequences of the above theorem is the existence of thin edge covers.

Theorem 18.17. *For any k -connected graph G there is a set $T \subseteq E$ such that each vertex v is incident to at least one edge of T and that T is $\tilde{O}(1/\sqrt{k})$ -thin w.r.t. G .*

Note that unlike [Theorem 18.5](#) the conclusion of the above theorem holds for any arbitrary value of k that does not depend on n .

Proof. Let γ^*, D^* be the optimum of primal of [Theorem 18.16](#). Let

$$F = \{e : b_e^\top D^{-1} b_e \leq 2\gamma^*\}.$$

By the constraints of the SDP, F has at least $k/2$ edges incident to every vertex of G . For every edge $e \in F$, let $x_e = (L_G + D)^{-1/2} b_e$. Observe that for any edge $e \in F$, $\|x_e\|^2 \leq 2\gamma^*$.

Let $V = \{v_1, \dots, v_n\}$. Let μ be a strong Rayleigh distribution on multisets $\{E_1, E_2, \dots, E_n\}$ where E_i is a uniformly random edge of F incident to v_i . It follows that the marginal probability of each edge is at most $2/k$ because every vertex has at least $k/2$ edges in F . Therefore, by the extension of [MSS13] that we discussed in Lecture 16, there is a sample of μ such that

$$\sum_{i=1}^n b_{E_i}^\top b_{E_i} \preceq O(\gamma^*)D + L_G \preceq_C O(\gamma^*)L_G$$

The conclusion of the theorem follows by that $\gamma^* = O(1/\sqrt{k})$ as shown in [Theorem 18.16](#). \square

In the rest of this section we prove [Theorem 18.16](#). Fix X and orthonormal vectors $\{z_e\}_{e \in E}$. Define

$$U = \left\{ v : \frac{(\mathbb{E}_{e \sim v} \langle z_e, X b_e \rangle)^2}{\mathbb{E}_{e \sim v} \|X b_e\|^2} \geq \alpha \right\},$$

be the set of “bad” vertices for some $\alpha = \frac{C}{\sqrt{k}}$. If $U = \emptyset$, then there is nothing to prove and we are done. So, we just need to take care of the bad vertices.

We show that

$$\sum_{v \in U} (\mathbb{E}_{e \sim v} \langle z_e, X b_e \rangle)^2 \leq \alpha \cdot \sum_{v \in V} \mathbb{E}_{e \sim v} \|X b_e\|^2. \quad (18.11)$$

Note that in the RHS we have the sum of the squared length of all edges of G .

Let $s = \max_{v \in U} \mathbb{E}_{e \sim v} \langle z_e, X b_e \rangle$. We partition the vertices of U into $\log(n)$ groups, where for each i ,

$$U_i = \{v \in U : \mathbb{E}_{e \sim v} \langle z_e, X b_e \rangle \approx s/10^i\}.$$

If there is any vertex where $\mathbb{E}_{e \sim v} \langle z_e, X b_e \rangle \ll s/n^2$ we simply remove v from U as it has no contribution in the LHS of (18.11).

Let $F_i = \cup_{v \in U_i} E(\{v\}, \overline{\{v\}})$ be the union of edges incident to vertices of U_i . By [Lemma 18.12](#), for each i , there is a family \mathcal{B}_i of disjoint L_2^2 balls of radius r_i all centered at endpoints of edges of F_i such that

$$\begin{aligned} |\mathcal{B}_i| \cdot r_i &\gtrsim \alpha \cdot \frac{1}{|F_i|} \left(\sum_{e \in F_i} \langle z_e, X b_e \rangle \right)^2 \\ &\asymp \frac{\alpha \cdot |U_i|}{|F_i|} \cdot \sum_{v \in U_i} \left(\sum_{e \sim v} \langle z_e, X b_e \rangle \right)^2, \\ &= k \cdot \alpha \sum_{v \in U_i} (\mathbb{E}_{e \sim v} \langle z_e, X b_e \rangle)^2. \end{aligned} \quad (18.12)$$

where the second equation follows by the definition of U_i and the third equation follows by the k -regularity of G . Summing up the above over all i ,

$$\sum_i |\mathcal{B}_i| \cdot r_i \gtrsim k \cdot \alpha \cdot \sum_{v \in U} (\mathbb{E}_{e \sim v} \langle z_e, X b_e \rangle)^2.$$

So, to prove (18.11) all we need to show is that

$$\sum_i |\mathcal{B}_i| \cdot r_i \lesssim \sum_{v \in V} \mathbb{E}_{e \sim v} \|X b_e\|^2. \quad (18.13)$$

L_1 translation. Next, we translate the above goal into an L_1 statement. Since X is a (unweighted) cut metric, for any edge e , $\|Xb_e\|^2 = \|Xb_e\|_1$. In addition, if we replace each (L_2^2) ball in \mathcal{B}_i with an L_1 ball of the same radius r_i and the same center, the balls of \mathcal{B}_i remain disjoint. To see that note that for any two balls $B_1(X_{v_1}, r_i), B_2(X_{v_2}, r_i) \in \mathcal{B}_i$, since B_1, B_2 are disjoint,

$$\|X_{v_1} - X_{v_2}\| \geq 2\sqrt{r_i},$$

so

$$\|X_{v_1} - X_{v_2}\|_1 = \|X_{v_1} - X_{v_2}\|^2 \geq 4r_i.$$

So, all we need to show is for families of disjoint L_1 balls $\mathcal{B}_1, \mathcal{B}_2, \dots$ defined as above

$$\sum_i k \cdot |\mathcal{B}_i| \cdot r_i \lesssim \sum_{v \in V} \|Xb_e\|^2. \quad (18.14)$$

Another Bad Example. First observe that if there is only one family of balls in this sequence, i.e., $\mathcal{B}_i = \emptyset$ for $i \geq 2$, the above inequality follows immediately by [Fact 18.11](#). The main challenge is that balls in different \mathcal{B}_i 's may intersect.

For a vertex $v \in U_i$ let $\mathcal{B}_{i,v}$ be the balls of \mathcal{B}_i that are centered at neighbors of v . We assume that for each $v \in U_i$,

$$|\mathcal{B}_{i,v}| \cdot r_i \gtrsim k \cdot \alpha (\mathbb{E}_{e \sim v} \langle z_e, Xb_e \rangle)^2 \quad (18.15)$$

and we remove v from $U - i$ and the balls of $\mathcal{B}_{i,v}$ from \mathcal{B}_i if v does not satisfy this constraint. By [\(18.12\)](#), the LHS of [\(18.14\)](#) is (almost) invariant under this removal. For the simplicity of the argument we assume that for each $v \in U$, and all edges $f \sim v$,

$$\|Xb_f\|^2 \approx \mathbb{E}_{e \sim v} \|Xb_e\|^2. \quad (18.16)$$

The above assumption can be justified with a simple bucketing argument that incurs a $\log(k)$ loss, we omit the details.

Before proving [\(18.14\)](#) we want to emphasize the importance of the above construction of $\mathcal{B}_{i,v}$'s. In general, if we do not assume a specific structure on the balls of \mathcal{B}_i other than they are disjoint [\(18.14\)](#) is not necessarily true, see [Figure 18.5](#) for a bad example.

Compact Bags of Balls. We say a family of disjoint L_1 balls \mathcal{B} is *compact*, if the sum of the radii of the balls in \mathcal{B} is at least 10 times the L_1 diameter of the centers of balls of \mathcal{B} . In the following claim we show that [\(18.14\)](#) holds if all of the sets $\mathcal{B}_{i,v}$ are compact.

First, let us prove that these sets are compact. Fix $v \in U_i$; since the balls of \mathcal{B}_i are centered at neighbors of v , by the assumption [\(18.16\)](#), the L_1 diameter of $\mathcal{B}_{i,v}$ is at most

$$2\mathbb{E}_{e \sim v} \|Xb_e\|^2 \leq \frac{2}{\alpha} (\mathbb{E}_{e \sim v} \langle z_e, Xb_e \rangle)^2.$$

using the definition of U_i . But, by the pruning assumption [\(18.15\)](#), the sum of the radii of balls of $\mathcal{B}_{i,v}$ is at least

$$|\mathcal{B}_{i,v}| \cdot r_i \gtrsim k \cdot \alpha (\mathbb{E}_{e \sim v} \langle z_e, Xb_e \rangle)^2.$$

So, for a large enough C , $\mathcal{B}_{i,v}$ is compact.

The following standalone claim holds for any sequence of family of compact bags balls.

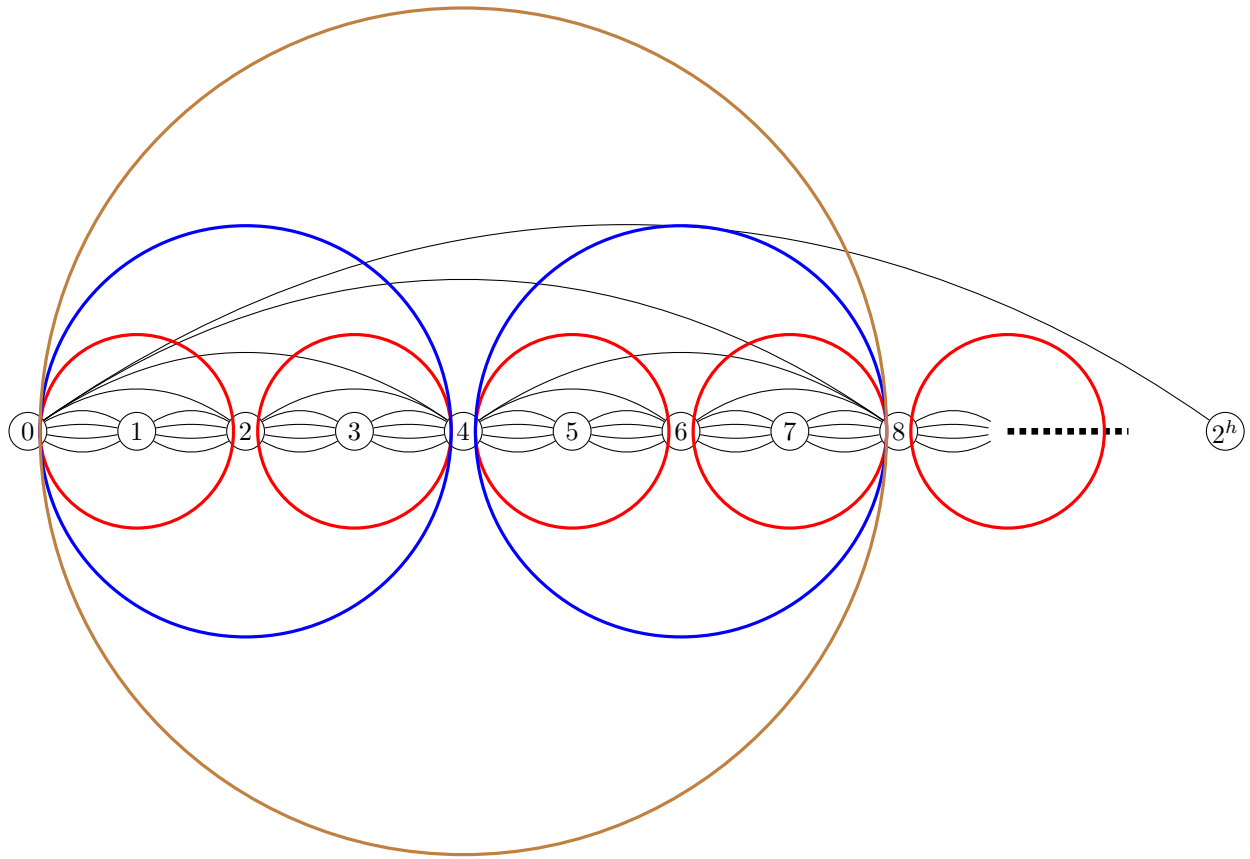


Figure 18.5: Consider the natural L_1 mapping of this graph where vertex i is mapped to the number i . Consider h layers of L_1 balls as shown where the radii of all balls in layer i is 2^i and they are disjoint. Although the sum of the radii of all balls in this family is $n \cdot h$, the sum of the L_1 length of the edges of G is $n \cdot (h + k)$.

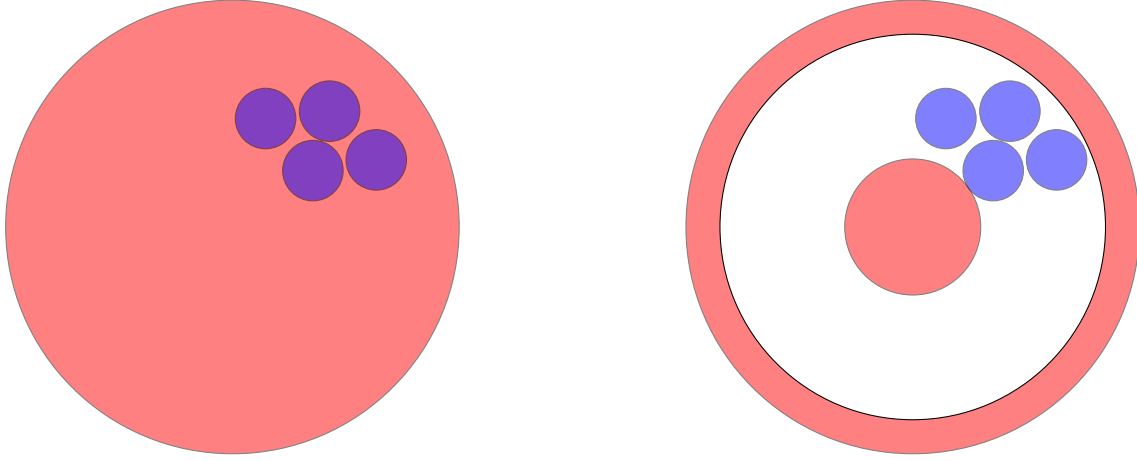


Figure 18.6: Construction of Hollowed ball. When a compact bag of balls lie inside an inserted ball B , i.e., the big red ball in the left, we decompose B into two hollowed balls that do not intersect any of the balls in the give compact set as shown in the right.

Claim 18.18. *Let G be a k -connected graph. Let $\mathcal{B}_1, \mathcal{B}_2, \dots$ be a sequence of families of L_1 balls such that balls in \mathcal{B}_i are disjoint, they have radius r_i and they are all centered at the vertices of G such that for each $i \geq 2$, $r_i \leq r_{i-1}/10$. If for each i , we can decompose \mathcal{B}_i into bags $\mathcal{B}_{i,v}$ where each bag $\mathcal{B}_{i,v}$ is compact, then,*

$$\sum_i k \cdot |\mathcal{B}_i| \cdot r_i \leq \sum_e \|Xb_e\|_1. \quad (18.17)$$

Proof. As usual, if there is only family \mathcal{B}_1 we are done by [Fact 18.11](#). The idea is to decompose the space into *hollowed* balls. An L_1 hollowed ball $B(X_v, r \| r')$ is the set of points at distance r to r' of v ,

$$B(X_v, r \| r') = \{y : r \leq \|X_v - y\|_1 \leq r'\}.$$

We say the *width* of the above hollowed ball is $r' - r$. From now on we also use the term hollowed ball to refer to an ordinary L_1 ball. Observe that the statement of [Fact 18.11](#) naturally extends to any family of *disjoint* hollowed balls.

So, all we need to do is to construct a disjoint family of hollowed balls such that the sum their widths is at least a constant factor of the LHS of (18.17). Here is the idea. We inductively insert the balls in the family for $i = 1 \rightarrow \infty$. Say we already have an L_1 ball $B \in \mathcal{B}_i$. By definition any other ball of \mathcal{B}_i does not intersect B . Now, suppose a bag of balls $\mathcal{B}_{i+1,v}$ lies completely inside B (see the left of [Figure 18.6](#)). Then, we decompose B into two hollowed balls that do not intersect with balls of $\mathcal{B}_{i+1,v}$ and we insert all balls of $\mathcal{B}_{i+1,v}$ (see the right of [Figure 18.6](#)).

This operation decreases the sum of the widths of the (hollowed) balls by at most twice the diameter of the centers of balls of \mathcal{B}_{i+1} and increases it by at least $|\mathcal{B}_{i+1,v}| \cdot r_i$. Since $\mathcal{B}_{i+1,v}$ is compact, in total, the sum of the widths of all of the (hollowed) balls increases by at least $|\mathcal{B}_{i+1,v}| \cdot r_i/2$. Therefore, by the end of this construction we get a family of disjoint (hollowed) balls where the sum of their widths is within a constant factor of the LHS of (18.17) as desired. \square

References

- [AO15] N. Anari and S. Oveis Gharan. “Effective-Resistance-Reducing Flows, Spectrally Thin Trees, and Asymmetric TSP”. to appear in FOCS. 2015. URL: <http://arxiv.org/abs/1411.4613> (cit. on pp. 18-1, 18-5, 18-16).
- [BSS14] J. D. Batson, D. A. Spielman, and N. Srivastava. “Twice-Ramanujan Sparsifiers”. In: *SIAM Review* 56.2 (2014), pp. 315–334 (cit. on p. 18-2).
- [Chr+11] P. Christiano, J. A. Kelner, A. Madry, D. A. Spielman, and S. Teng. “Electrical flows, laplacian systems, and faster approximation of maximum flow in undirected graphs”. In: *STOC*. 2011, pp. 273–282 (cit. on p. 18-6).
- [Goe06] M. X. Goemans. “Minimum Bounded Degree Spanning Trees”. In: *FOCS*. 2006, pp. 273–282 (cit. on p. 18-4).
- [LS15] L. C. Lau and M. Singh. “Approximating Minimum Bounded Degree Spanning Trees to within One of Optimal”. In: *J. ACM* 62.1 (2015), pp. 1–19 (cit. on p. 18-4).
- [MSS13] A. Marcus, D. A. Spielman, and N. Srivastava. “Interlacing Families II: Mixed Characteristic Polynomials and the Kadison-Singer Problem”. 2013. URL: <http://arxiv.org/abs/1306.3969> (cit. on p. 18-18).
- [RFP10] B. Recht, M. Fazel, and P. A. Parrilo. “Guaranteed Minimum-Rank Solutions of Linear Matrix Equations via Nuclear Norm Minimization”. In: *SIAM REVIEW* 52.3 (2010), pp. 471–501 (cit. on p. 18-9).
- [SLB74] G. L. Shurbet, T. O. Lewis, and T. L. Boullion. “Quadratic Matrix Equations”. In: *Ohio Journal of Science* 74.5 (1974), pp. 273–277 (cit. on p. 18-8).