Geometric Representations of Graphs

László Lovász
Institute of Mathematics
Eötvös Loránd University, Budapest
e-mail: lovasz@cs.elte.hu

December 11, 2009

Contents

Chapter 0

Introduction

Chapter 1

Planar graphs and polytopes

1.1 Planar graphs

A graph $G = (V, E)$ is *planar*, if it can be drawn in the plane so that its edges are Jordan curves and they intersect only at their endnodes[1]. A *plane map* is a planar graph with a fixed embedding. We also use this phrase to denote the image of this embedding, i.e., the subset of the plane which is the union of the set of points representing the nodes and the Jordan curves representing the edges.

The complement of a plane map decomposes into a finite number of arcwise connected pieces, which we call the *faces* (or *countries*) of the planar map. We usually denote the number of nodes, edges and faces of a planar graph by n, m and f.

Every planar map $G = (V, E)$ has a *dual map* $G^* = (V^*, E^*)$ (Figure 1.1) As an abstract graph, this can be defined as the graph whose nodes are the faces of G. If the two faces share k edges, then we connect them in G^* by k edges, so that each edge $e \in E$ will correspond to an edge e^* of G^*. So $|E^*| = |E|$.

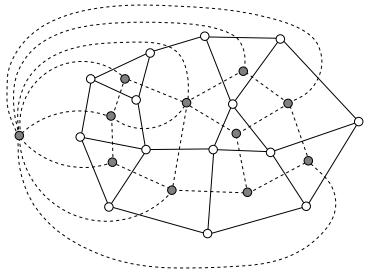

Figure 1.1: A planar map and its dual.

This dual has a natural drawing in the plane: in the interior of each face F of G we select

[1]We use the word *node* for the node of a graph, the word *vertex* for the vertex of a polytope, and the word *point* for points in the plane or in other spaces.

a point v_F (which can be called its *capital* if we use the country terminology), and on each edge $e \in E$ we select a point u_e (this will not be a node of G^*, just an auxiliary point). We connect v_F to the points u_e for each edge on the boundary of F by nonintersecting Jordan curves inside F. If the boundary of F goes through e twice (i.e., both sides of e belong to F), then we connect v_F to u_e by two curves, entering e from two sides. The two curves entering u_e form a single Jordan curve representing the edge e^*. It is not hard to see that each face of G^* will contain a unique node of G, and so $(G^*)^* = G$.

A planar map is called a *triangulation* if every face has 3 edges. Note that a triangulation may have parallel edges, but no two parallel edges can bound a face.

In every simple planar map we can introduce new edges to turn all faces into triangles while keeping the graph simple. In fact, let us draw new edges as long as we can preserving the planar embedding and creating no parallel edges. The resulting graph G has no cutpoints; for then it would have a face meeting two blocks of G, and two nodes of the face in different blocks could be connected by an edge inside this face. So G is 2-connected and, therefore, each face is a circuit. Suppose that C is the boundary of a face with at least 4 nodes. One of the two diagonals of C must be missing, since they all ought to run outside C and therefore, they would cross. This diagonal could be added inside C.

We often need the following basic fact about planar graphs, which we quote without proof (see Exercise 1.1.

Theorem 1.1.1 (Euler's Formula) *For every connected planar graph, $n-m+f = 2$ holds.* □

Some important consequences of Euler's Formula are the following.

Corollary 1.1.2 (a) *A simple planar graph with n nodes has at most $3n-6$ edges.*

(b) *A simple bipartite planar graph with n nodes has at most $2n-4$ edges.*

(c) *Every simple planar graph has a node with degree at most 5.*

(d) *Every simple bipartite planar graph has a node with degree at most 3.*

Corollary 1.1.3 *The graphs K_5 and $K_{3,3}$ (see Figure 1.2) are not planar.*

The following theorem gives a characterization of planar graphs.

Theorem 1.1.4 (Kuratowski's Theorem) *A graph G is embedable in the plane if and only if it does not contain a subgraph homeomorphic to the complete graph K_5 or the complete bipartite graph $K_{3,3}$.* □

Among planar graphs, 3-connected planar graphs are especially important. A cycle C in a graph G is called *non-separating*, if it has no chords, and the removal of its nodes does not disconnect the graph.

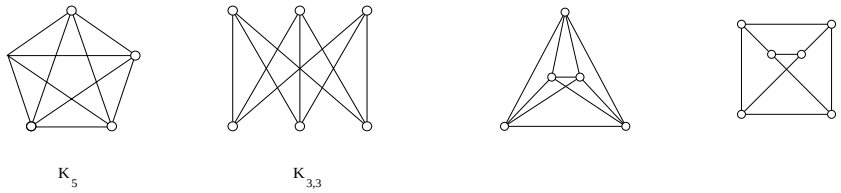

Figure 1.2: The two Kuratowski graphs. The two drawings on the right hand side show that both graphs can be drawn in the plane with a single crossing.

Proposition 1.1.5 *Let G be a 3-connected planar graph, and C a cycle in G. Then C is non-separating if and only if it bounds a face.* □

Corollary 1.1.6 *Every simple 3-connected planar graph has an essentially unique embedding in the plane in the sense that the set of cycles that bound faces is uniquely determined.* □

The following characterization of 3-connected planar graphs was proved by Tutte [197]. We will not use it and formulate it without proof.

Theorem 1.1.7 *Let G be a 3-connected graph. Then every edge of G is contained in at least two non-separating cycles. G is planar if and only if every edge is contained in exactly two non-separating cycles.* □

As a further application of Euler's Formula, we prove the following lemma, which will be useful in several arguments later on.

Lemma 1.1.8 *In a simple planar map whose edges are 2-colored with red and blue there is always a node where the red edges (and so also the blue edges) are consecutive.*

Proof. We may assume that the graph is connected (else, we can apply the lemma to each component). Define a *happy corner* as a pair of edges that are consecutive on the boundary of a face and one of them is red, the other blue. We are going to estimate the number N of happy corners in two different ways.

Suppose that the conclusion does not hold. Then for every node v, the color of the edge changes at least four times if we consider the edges incident with v in their cyclic order according to the embedding. (In particular, every node must have degree at least four.) So every node is incident with at least 4 happy corners, which implies that

$$N \geq 4n. \tag{1.1}$$

On the other hand, let f_r be the number of faces with r edges on their boundary. (To be precise: since we did not assume that the graph is 2-connected, it may happen that an

edge has the same face on both sides. In this case we count this edge twice.) The maximum number of happy corners a face with r edges can have is

$$\begin{cases} r-1, & \text{if } r \text{ is odd,} \\ r, & \text{if } r \text{ is even.} \end{cases}$$

Thus the number of happy corners is at most

$$N \leq 2f_3 + 4f_4 + 4f_5 + 6f_6 + 6f_7 + \dots \tag{1.2}$$

To relate the upper and lower bounds, we use the trivial identities

$$f = f_3 + f_4 + f_5 + \dots \qquad \text{and} \qquad 2m = 3f_3 + 4f_4 + 5f_5 + \dots$$

By (1.1) and Euler's Formula,

$$\begin{aligned} N \geq 4n = 4(m+2-f) &= 8 + (6f_3 + 8f_4 + 10f_5 + \dots) - (4f_3 + 4f_4 + 4f_5 + \dots) \\ &= 8 + (2f_3 + 4f_4 + 6f_5 + \dots) \end{aligned}$$

This clearly contradicts the upper bound (1.2) on N. □

Exercise 1.1 (a) Prove Euler's Formula. (b) Find a relationship between the numbers of nodes, edges and faces of a map in the projective plane and on the torus.

Exercise 1.2 Let G be a simple planar graph. Prove that you can add edges to G so that you make it 3-connected while keeping it simple and planar.

Exercise 1.3 Construct a 2-connected simple planar graph on n nodes which has an exponential (in n) number of different embeddings in the plane. (Recall: two embeddings are considered different if there is a cycle that is the boundary cycle of a face in one embedding but not in the other.)

Exercise 1.4 Show that the statement of Lemma 1.1.8 remains valid for maps on the projective plane, but not for maps on the torus.

Exercise 1.5 (a) Let L be a finite set of lines in the plane, not all going through the same point. Prove that there is a point where exactly two lines in L go through.

(b) Let S be a finite set of points in the plane, not all on a line. Prove that there is a line which goes through exactly two points in S.

Exercise 1.6 (a) Let L be a finite set of lines in the plane, not all going through the same point. Color these lines red and blue. Prove that there is a point where at least two lines in L intersect and all the lines through this point have the same color.

(b) Let S be a finite set of points in the plane, not all on a line. Color these points red and blue. Prove that there is a line which goes through at least two points in S and all whose points have the same color.

1.2 Planar separation

The following important theorem about planar graphs was proved by Lipton and Tarjan:

Theorem 1.2.1 *Every planar graph $G = (V, E)$ on n nodes contains a set $S \subseteq V$ such that $|S| \leq 4\sqrt{n}$, and every connected component of $G \setminus S$ has at most $2n/3$ nodes.*

We do not prove this theorem here; a proof (with a weaker bound of $3n/4$ on the sizes of the components) based on circle packing will be presented in Section 6.1.9.

Exercise 1.7 If G is a $k \times k$ grid, then every set S of nodes separating G into parts smaller that $(1 - c)k^2$ has at least $\sqrt{2c}k$ nodes.

1.3 Straight line representation and 3-polytopes

Theorem 1.3.1 (Fáry–Wagner Theorem) *Every simple planar map can be drawn in the plane with straight edges.*

Proof. We use induction on the number of nodes. If this is at most 3, the assertion is trivial. It suffices to prove the assertion for triangulations.

There is an edge xy which is contained in two triangles only. For let x be a node which is contained inside some triangle T (any point not on the outermost triangle has this property) and choose x, T so that the number of faces inside T is minimal. Let y be any neighbor of x. Now if xy belongs to three triangles (x, y, z_1), (x, y, z_2) and (x, y, z_3), then all these triangles would be properly contained in T and, say, (x, y, z_1) would contain z_3 inside it, contrary to the minimality of T.

Contract xy to a node p and remove one edge from both arising pairs of parallel edges. This way we get a new simple triangulation G_0 and by the induction hypothesis, there is an isomorphic triangulation G_0' with straight edges.

Now consider the edges pz_1 and pz_2 of G_0'. They split the angle around p into two angles; one of these contains the edges whose pre-images in G are adjacent to x, the other one those whose pre-images are adjacent to y. Therefore, we can "pull x and y apart" and get an a straight line drawing of G. □

Let P be a convex 3-polytope. (See Section 13.7 for basic facts about polytopes). The vertices and edges of P form a graph G_P, which we call the *skeleton* of P.

Proposition 1.3.2 *The skeleton of every 3-polytope is a 3-connected planar graph.*

We describe the simple proof, because this is our first example of how a geometric representation can be used to derive a purely graph-theoretic property, namely 3-connectivity.

Proof. Let F be any facet of P, and let x be a point that is outside P but very close to F; more precisely, assume that the plane Σ of F separates x from P, but for every other facet F', x is on the same side of the plane of F' as P. Let us project the skeleton of P from x to the plane Σ. Then we get an embedding of G_P in the plane.

To see that G_P is 3-connected, it suffices to show that for any four nodes a, b, c, d there is a path from a to b which avoids c and d.

If a, b, c, d are not coplanar, then let Π be a plane that separates $\{a, b\}$ from $\{c, d\}$; then we can connect a and b by a polygon consisting of edges of P that stays on the same side of Π as a and b, and so avoids c and d (see Appendix 13.7, Corollary 13.7.6).

If a, b, c, d are coplanar, let Π be a plane that contains them. One of the open halfspaces bounded by Π contains at least one vertex of P. We can then connect a and b by a polygon consisting of edges of P that stays on this side of Π (except for its endpoints a and b), and so avoids c and d. □

The embedding of G_P in the plane constructed above is called the *Schlegel diagram* of the polytope (with respect to the facet F).

The converse of this last proposition is an important and much more difficult theorem, proved by Steinitz [188]:

Theorem 1.3.3 (Steinitz's Theorem) *A simple graph is isomorphic to the skeleton of a 3-polytope if and only if it is 3-connected and planar.*

A bijection between the nodes of a simple graph G and the vertices of a convex polytope P in $\mathbb{R}^3$ that gives a bijection between the edges of G and the edges of P is called a *Steinitz representation* of the graph G. We don't prove Steinitz's Theorem here, but later constructions of representations by polytopes, and in fact with special properties, will follow from the material in chapters 4, 6.1 and 8.

There may be many representations of a graph by a 3-polytope; in 3-space the variety of these representations has nice properties, but in higher dimension it can be very complicated. We refer to [167].

The construction of the planar embedding of G_P in the proof of Proposition 1.3.2 gives an embedding with straight edges. Therefore Steinitz's Theorem also proves the Fáry–Wagner theorem, at least for 3-connected graphs. It is easy to see that the general case can be reduced to this by adding new edges so as to make the graph 3-connected (see exercise 1.2).

Finally, we note that the Steinitz representation is also related to planar duality.

Proposition 1.3.4 *Let P be a convex polytope with the origin in its interior, and let P^* be its polar. Then the skeletons G_P are G_{P^*} are dual planar graphs.*

Proof.

□

1.4 Crossing number

If G is a non-planar graph, we may want to draw it so as to minimize the number of intersections of the curves representing the edges. To be more precise, we assume that the drawing satisfies the following conditions: none of the edges is allowed to pass through a node, no three edges pass through the same point, any two edges intersect in a finite number of points, and every common point of two edges other than a common endpoint is an intersection point (so edges cannot touch). The minimum number of intersection points in such a drawing is called the *crossing number* of the graph, and denoted by $\mathrm{cr}(G)$.

The question whether $\mathrm{cr}(G) = 0$, i.e., whether G is planar, is decidable in polynomial time; but to compute $\mathrm{cr}(G)$ in general is NP-hard.

It is ot even known what the crossing numbers of complete are complete bipartite graphs are. Guy proved (by constructing an appropriate drawing of K_n in the plane) that

$$\mathrm{cr}(K_n) \le \frac{1}{4}\lfloor\frac{n}{2}\rfloor \cdot \lfloor\frac{n-1}{2}\rfloor \cdot \lfloor\frac{n-2}{2}\rfloor \cdot \lfloor\frac{n-3}{2}\rfloor < \frac{3}{8}\binom{n}{4}. \tag{1.3}$$

Zarankiewicz proved that

$$\mathrm{cr}(K_{n,m}) \le \lfloor\frac{n}{2}\rfloor \cdot \lfloor\frac{n-1}{2}\rfloor \cdot \lfloor\frac{m}{2}\rfloor \cdot \lfloor\frac{m-1}{2}\rfloor < \frac{1}{4}\binom{n}{2}\binom{m}{2}. \tag{1.4}$$

In both cases, it is conjectured that equality holds in the first inequality.

We will need lower bounds on the crossing number. We start with an easy observation.

Lemma 1.4.1 *Let G be a simple graph with n nodes and m edges, then*

$$\mathrm{cr}(G) \ge m - 3n + 6.$$

Proof. Consider a drawing of G with a minimum number of crossings. If $m \le 3n - 6$, then the assertion is trivial. If $m > 3n - 6$, then by Corollary 1.1.2(a), there is at least one crossing. Deleting one of the edges participating in a crossing, the number of crossings drops by at least one, and the lemma follows by induction. □

If m is substantially larger than n, then this bound becomes very week. The following theorem of Ajtai, Chvátal, Newborn, and Szemerédi [3] and Leighton [123] gives a much better bound when m is large, which is best possible except for the constant.

Theorem 1.4.2 *Let G be a simple graph with n nodes and m edges, where $m \ge 4n$. Then*

$$\mathrm{cr}(G) \ge \frac{m^3}{64n^2}.$$

Proof. Consider a drawing of G with a minimum number of crossings. Let $p = (4n)/m$, and delete each node of G independently with probability $1 - p$. The remaining graph H satisfies

$$\mathrm{cr}(H) \geq |E(H)| - 3|V(H)| + 6 > |E(H)| - 3|V(H)|.$$

Take the expectation of both sides. Then

$$\mathsf{E}(|V(H)|) = pn, \qquad \mathsf{E}(|E(H)) = p^2 m, \qquad \mathsf{E}(\mathrm{cr}(H)) \leq p^4 \mathrm{cr}(G).$$

(For the last inequality: a crossing of G survives if and only if all four endpoints of the two participating edges survive, so the drawing of H we get has, in expectation, $p^4\mathrm{cr}(G)$ crossings. However, this may not be an optimal drawing of H, hence we only get an inequality.) Thus

$$\mathrm{cr}(G) \geq \frac{m}{p^2} - \frac{3n}{p^3} = \frac{m^3}{64n^2}.$$

□

Chapter 2

Graphs from point sets

2.1 Unit distance graphs

Given a set S of points in $\mathbb{R}^d$, we construct the *unit distance graph* on S by connecting two points by an edge if and only if their distance is 1. Many basic properties of these graphs have been studied, most of which turn out quite difficult.

2.1.1 The number of edges

Let S be a finite set of n points in the plane. A natural problem is the following [61]: how many times can the same distance (say, distance 1) occur between the points of a set S of size n. In this first paper Erdős gave the following upper bound.

Theorem 2.1.1 *The unit distance graph of n points in the plane has at most $n^{3/2}$ edges.*

Proof. The graph of unit distances contains no $K_{2,3}$. □

It was not easy to improve these simple bounds. The only substantial progress was made by Spencer, Szemerédi and Trotter [185], who proved the following bound, which is still the best known.

Theorem 2.1.2 *The unit distance graph of n points in the plane has at most $n^{4/3}$ edges.*

Proof. We describe a recent elegant proof of this bound by Székely [194]. □

The proof just given also illuminates that this bound depends not only on graph theoretic properties of $G(S)$ but, rather, on properties of its actual embedding in the plane. The difficulty of improving the bound $n^{4/3}$ is in part explained by a construction of Brass [31], which shows that in some Banach norm, n points in the plane *can* determine as many as $\Theta(n^{3/2})$ unit distances.

The situation in $\mathbb{R}^3$ is probably even less understood, but in dimensions 4 and higher, an easy construction of Lenz shows that every finite complete k-partite graph can be realized in $\mathbb{R}^{2k}$ as the graph of unit distances: place the points on two orthogonal planes, at distance $1/\sqrt{2}$ from the origin. Hence the maximum number of unit distances between n points in $\mathbb{R}^d$ ($d \geq 4$) is $\Theta(n^2)$.

Another intriguing unsolved question is the case when the points in S form the vertices of a convex n-gon. Erdős and Moser conjectured the upper bound to be $O(n)$. The best construction [58] gives $2n - 7$ and the best upper bound is $O(n \log n)$ [78].

We get different, and often more accessible questions if we look at the largest distances among n points. In 1937 Hopf and Pannwitz [103] proved the following.

Theorem 2.1.3 *The largest distance among n points in the plane occurs at most n times.*

Proof. The proof of this is left to the reader as Exercise 2.1. □

Exercise 2.1 Let S be a set of n points in the plane.

(a) Prove that any two longest segments connecting points in S have a point in common.

(b) Prove that the largest distance between points in S occurs at most n times.

For 3-space, Heppes and Révész [93] proved that the largest distance occurs at most $2n - 2$ times. In higher dimensions the number of maximal distances can be quadratic: Lenz's construction can be carried out so that all the $n^2/4$ unit distances are maximal. The exact maximum is known in dimension 4 [30].

2.1.2 Chromatic number and independence number

The problem of determining the maximum chromatic number of a unit distance graph in the plane was raised by Hadwiger [86]. He proved the following bounds; we don't know more about this even today.

Theorem 2.1.4 *Every planar unit distance graph has chromatic number at most* 7. *There is a planar unit distance graph with chromatic number* 4.

Proof. Figure 2.1(a) shows a coloring of all points of the plane with 7 colors such that no two points at distance 1 have the same color. Figure 2.1(b) shows a unit distance graph with chromatic number 4: Suppose it can be colored with three colors, and let the top point v have color red (say). The other two points in the left hand side triangle containing v must be green and blue, so the left bottom point must be red again. Similarly, the right bottom point must be red, so we get two adjacent red points.

□

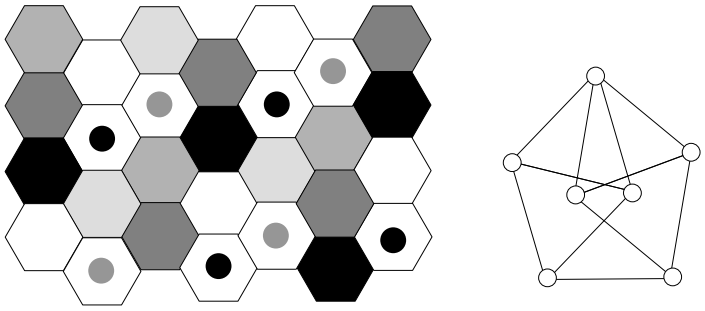

Figure 2.1: (a) The plane can be colored with 7 colors. (b) Three colors are not enough.

For higher dimensions, we know in a sense more. Klee noted that $\chi(G(\mathbb{R}^d))$ is finite for every d, and Larman and Rogers [121] proved the bound $\chi(G(\mathbb{R}^d)) < (3 + o(1))^d$. Erdős conjectured and Frankl and Wilson [76] proved that the growth of $\chi(\mathbb{R}^d)$ is indeed exponential in d: $\chi(\mathbb{R}^d) > (2 - o(1))^d$. The exact base of this exponential function is not known.

One gets interesting problems by looking at graph formed by largest distances. The chromatic number of the graph of the largest distances is equivalent to the well known Borsuk problem: how many sets with smaller diameter can cover a convex body in $\mathbb{R}^d$? Borsuk conjectured that the answer is $d+1$, which is not difficult to prove this if the body is smooth. The bound is also proved for $d \leq 3$. (Interestingly the question in small dimensions was settled by simply using upper bound on the number of edges in the graph of the largest distances, discussed above). But the general case was settled in the negative by Kahn and Kalai [109], who proved that the minimum number of covering sets (equivalently, the chromatic number of the graph of largest distances) can be as large as $2^{\Omega(\sqrt{d})}$.

2.1.3 Unit distance representation

We can turn the above question around and ask: Given a graph G, how can we represent it by unit distances? To be precise, a *unit distance representation* of a graph $G = (V, E)$ is a mapping $u : V \to \mathbb{R}^d$ for some $d \geq 1$ such that $|u_i - u_j| = 1$ for every $ij \in E$ (we allow that $|u_i - u_j| = 1$ for some nonadjacent nodes i, j).

Every finite graph has a unit distance representation in a sufficiently high dimension. (For example, we can map its nodes onto the vertices of a regular simplex with edges of length 1.) The first problem that comes to mind, first raised by Erdős, Harary and Tutte [66], is to find the minimum dimension $\dim(G)$ in which a graph has a unit distance representation. Figure 2.2 shows a 2-dimensional unit distance representation of the Petersen graph [66]. It

turns out that this minimum dimension is linked to the chromatic number of the graph:

$$\lg \chi(G) \leq d(G) \leq 2\chi(G).$$

The lower bound follows from the results of Larman and Rogers [121] mentioned above; the upper bound follows from a modification of Lenz's construction.

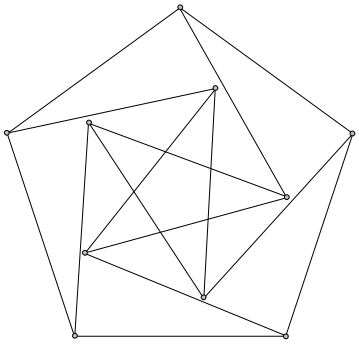

Figure 2.2: A unit distance representation of the Petersen graph.

There are many other senses in which we may want to find the most economical unit distance representation. We only discuss one: what is the smallest radius of a ball containing a unit distance representation of G (in any dimension)? This question can be answered using semidefinite programming (see Appendix 13.16).

Considering the Gram matrix $A = (u_i^{\mathsf{T}} u_j)$, it is easy to obtain the following reduction to semidefinite programming:

Proposition 2.1.5 *A graph G has a unit distance representation in a ball of radius R (in some appropriately high dimension) if and only if there exists a positive semidefinite matrix A such that*

$$A_{ii} \leq R^2 \quad (i \in V)$$
$$A_{ii} - 2Aij + A_{jj} = 1 \quad (ij \in E).$$

In other words, the smallest radius R is the square root of the optimum value of the semidefinite program

$$\begin{aligned} &\text{minimize } w \\ &\text{subject to } A \succeq 0 \\ &\qquad A_{ii} \leq w \quad (i \in V) \\ &\qquad A_{ii} - 2Aij + A_{jj} = 1 \quad (ij \in E). \end{aligned}$$

The unit distance embedding of the Petersen graph in Figure 2.2 is not an optimal solution of this problem. Let us illustrate how semidefinite optimization can find the optimal embedding by determining this for the Petersen graph. In the formulation above, we have to find a 10×10 positive semidefinite matrix A satisfying the given linear constraints. For a given w, the set of feasible solutions is convex, and it is invariant under the automorphisms of the Petersen graph. Hence there is an optimum solution which is invariant under these automorphisms (in the sense that if we permute the rows and columns by the same automorphism of the Petersen graph, we get back the same matrix).

Now we know that the Petersen graph has a very rich automorphism group: not only can we transform every node into every other node, but also every edge into every other edge, and every nonadjacent pair of nodes into every other non-adjacent pair of nodes. A matrix invariant under these automorphisms has only 3 different entries: one number in the diagonal, another number in positions corresponding to edges, and a third number in positions corresponding to nonadjacent pairs of nodes. This means that this optimal matrix A can be written as

$$A = xP + yJ + zI,$$

where P is the adjacency matrix of the Petersen graph, J is the all-1 matrix, and I is the identity matrix. So we only have these 3 unknowns x, y and z to determine.

The linear conditions above are now easily translated into the variables x, y, z. But what to do with the condition that A is positive semidefinite? Luckily, the eigenvalues of A can also be expressed in terms of x, y, z. The eigenvalues of P are well known (and easy to compute): they are 3, 1 (5 times) and -2 (4 times). Here 3 is the degree, and it corresponds to the eigenvector $\mathbf{1} = (1, \ldots, 1)$. This is also an eigenvector of J (with eigenvalue 10), and so are the other eigenvectors of P, since they are orthogonal to $\mathbf{1}$, and so are in the nullspace of J. Thus the eigenvalues of $xP + yJ$ are $3x + 10y$, x, and $-2x$. Adding zI just shifts the spectrum by z, so the eigenvalues of A are $3x+10y+z$, $x+z$, and $-2x+z$. Thus the positive semidefiniteness of A, together with the linear constraints above, gives the following linear program for x, y, z, w:

$$\begin{aligned} & \text{minimize } w \\ \text{subject to } & 3x + 10y + z \geq 0, \\ & x + z \geq 0, \\ & -2x + z \geq 0, \\ & y + z \leq w, \\ & 2z - 2x = 1. \end{aligned}$$

It is easy to solve this: $x = -1/4$, $y = 1/20$, $z = 1/4$, and $w = 3/10$. Thus the smallest radius of a ball in which the Petersen graph has a unit distance representation is $\sqrt{3/10}$. The corresponding matrix A has rank 4, so this representation is in 4 dimension.

It would be difficult to draw a picture of this representation, but we can offer the following nice matrix, whose columns will realize this representation (the center of the smallest ball containing it is not in the origin!):

$$\frac{1}{2}\begin{pmatrix} 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 1 & 1 \end{pmatrix} \tag{2.1}$$

(This matrix reflects the fact that the Petersen graph is the complement of the line-graph of K_5.)

2.2 Bisector graphs

2.2.1 The number of edges

For a set S of $n = 2m$ points in $\mathbb{R}^2$, we construct the *bisector graph* G_S on S by connecting two points by an edge if and only if the line through them has $m - 2$ points on each side.

Lemma 2.2.1 *Let $v \in S$, let e, f be two edges of G_S incident with v that are consecutive in the cyclic order of such edges, let W be the wedge between e and f, and let W' be the negative of W. Then there is exactly one edge of G_S incident with v in the wedge W'.*

Proof. Draw a line ℓ, with orientation, through v, and rotate it counterclockwise by π. Consider the number of points of S on its left hand side. This number changes from less than $m - 2$ to larger than $m - 2$ whenever the negative half of e passes through an edge of G_S incident with v, and it changes from larger than $m - 2$ to less than $m - 2$ whenever the positive half of e passes through an edge of G_S incident with v. □

Corollary 2.2.2 *Let $v \in S$, and let ℓ be a line through v that does not pass through any other point in S. Let H be the halfplane bounded by e that contains at most $m - 1$ points. Then for some $d \geq 0$, d edges incident with v are contained in H and $d + 1$ of them are on the opposite side of e.*

Corollary 2.2.3 *Every node in G_S has odd degree.*

Corollary 2.2.4 *Let ℓ be a line that does not pass through any point in S. Let k and $n - k$ of the points be contained on the two sides of ℓ, where $k \leq m$. Then ℓ intersects exactly k edges of G_S.*

From this corollary it is easy to deduce:

Theorem 2.2.5 *The number of edges of G_S is at most $2n^{3/2}$.*

Proof. Consider a line that is not parallel to any segment connecting two points of S, and call its direction "vertical". Let $k = \lceil \sqrt{n} \rceil$, and let us divide S by $k-1$ vertical lines $\ell_1, \ldots, \ell_{k-1}$ into k parts, each containing at most k points.

By Corollary 2.2.4, each line ℓ_i intersects at most m edges of G_S, so the number of those edges which intersect any of the lines ℓ_i is at most $(k-1)m \leq (k-1)k^2/2$. On the other hand, the number of those edges of G_S that do not intersect any ℓ_i is obviously at most $k\binom{k}{2} = (k-1)k^2/2$. Hence the number of edges of G_S is at most $(k-1)k^2 \leq 2n^{3/2}$. □

While the proof above seems to have a lot of room for improvement, the easy arguments only lead to an improvement in the constant. But using the "crossing number method", a substantially better bound can be derived [51]:

Theorem 2.2.6 *The number of edges of G_S is at most $4n^{4/3}$.*

Proof. We prove that

$$\mathrm{cr}(G_S) \leq \binom{n}{2}. \tag{2.2}$$

Theorem 1.4.2 implies then that

$$|E(G_S)| \leq (64n^2 \mathrm{cr}(G_S))^{1/3} < 4n^{4/3}.$$

To prove (2.2), note that we already have a drawing of G_S in the plane, and it suffices to estimate the number of crossings between edges in this drawing. Fix a coordinate system so that no segment connecting two points of S is vertical. Let $P = v_0 v_1 \ldots v_k$ be a path in G_S. We call P a *convex chain*, if

(i) the x-coordinates of the v_i are increasing,

(ii) the slopes of the edges $v_i v_{i+1}$ are increasing,

(iii) for $1 \leq i \leq k-1$, there is no edge incident with v_i whose slope is between the slopes of $v_{i-1}v_i$ and $v_i v_{i+1}$,

(iv) P is maximal with respect to this property.

Note that we could relax (iii) by excluding such an edge going to the right from v_i; indeed, if an edge $v_i u$ has slope between the slopes of $v_{i-1}v_i$ and $v_i v_{i+1}$, and the x-coordinate of u is larger than that of v_i, then there is also an edge wv_i that has slope between the slopes of $v_{i-1}v_i$ and $v_i v_{i+1}$, and the x-coordinate of w is smaller than that of v_i. This follows from Lemma 2.2.1.

Claim 2.2.7 *Every edge of G_S belongs to exactly one convex chain.*

Indeed, the edge as a path of length 1 satisfies (i)-(iii), and so it will be contained in a maximal such path. On the other hand, if two convex chains share and edge, then they must "branch" somewhere, which is impossible as they both satisfy (iii).

We can use convex chains to charge every intersection of two edges of G_S to a pair of nodes of G_S as follows. Suppose that the edges ab and cd intersect at a point p, where the notation is such that a is to the left from b, c is to the left of d, and the slope of ab is larger than the slope of cd. Let $P = v_0v_1 \dots ab \dots v_k$ and $Q = u_0u_1 \dots cd \dots u_l$ be the convex chains containing ab and cd, respectively. Let v_i be the first point on P for which the semiline v_iv_{i+1} intersects Q, and let u_j be the last point on Q for which the semiline $u_{i-1}u_i$ intersects P. We charge the intersection of ab and cd to the pair (v_i, u_j).

Claim 2.2.8 *Both P and Q are contained in the same side of the line v_iu_j.*

Assume (for convenience) that v_iu_j lies on the x-axis. We may also assume that P goes at least as deep below the x-axis as Q.

If the slope of $v_{i-1}v_i$ is positive, then the semiline $v_{i-1}v_i$ intersects Q, contrary to the definition of i. Similarly, the slope of u_ju_{j+1} (if $j < l$) is positive.

If the slope of v_iv_{i+1} is negative, then the semiline v_iv_{i+1} stays below the x-axis, and P is above this semiline. So if it intersects Q, then Q must reach farther below the x-axis than P, a contradiction.

Thus we know that the slope of $v_{i-1}v_i$ is negative but the slope of v_iv_{i+1} is positive, which implies by convexity that P stays above the x-axis. By our assumption, this implies that Q also stays above the x-axis, which proves the claim.

Claim 2.2.9 *At most one convex chain starts at each node.*

Assume that $P = v_0v_1 \dots ab \dots v_k$ and $Q = u_0u_1 \dots cd \dots u_l$ are two convex chains with $u_0 = v_0$. The edges v_0v_1 and u_0u_1 are different, so we may assume that the slope of u_0u_1 is larger. By Lemma 2.2.1, there is at least one edge tu_0 entering u_0 whose slope is between the slopes of v_0v_1 and u_0u_1. We may assume that tu_0 has the largest slope among all such edges. But then appending the edge tu_0 to Q conditions (i)-(iii) are preserved, which contradicts (iv).

To complete the proof of (2.2), it suffices to show:

Claim 2.2.10 *At most one intersection point is charged to every pair (v, u) of points of S.*

Suppose not. Then there are two pairs of convex chains P, Q and P', Q' such that p and P' pass through v, Q and Q' pass through u, some edge of P to the right of v intersects an edge of Q to the left of P, and similarly for P' and Q'. We may assume that $P \neq P'$. Let

vw and vw' be the edges of P and P' going from v to the right. Claim 2.2.7 implies that these edges are different, and so we may assume that the slope of uw is larger than the slope of uw'; by Claim 2.2.8, these slopes are positive. As in the proof of Claim 2.2.9, we see that Q does not start at v, and in fact it continues to the left by an edge with positive slope. But this contradicts Claim 2.2.8. □

2.2.2 k-sets and j-edges

A k-set of S is a subset $T \subseteq S$ of cardinality k such that T can be separated from its complement $T \setminus S$ by a line. An i-set with $1 \leq i \leq k$ is called an $(\leq k)$-*set*.

A j-*edge* of S is an ordered pair uv, with $u, v \in S$ and $u \neq v$, such that there are exactly j points of S on the right hand side of the line uv. Let $e_j = e_j(S)$ denote the number of j-edges of S; it is well known and not hard to see that for $1 \leq k \leq n-1$, the number of k-sets is e_{k-1}. An i-edge with $i \leq j$ will be called a $(\leq j)$-*edge*; we denote the number of $(\leq j)$-edges by $E_j = e_0 + \ldots + e_j$.

It is not hard to derive a sharp lower bound for each individual e_j. We will use the following theorem from [60], which can be proved extending the proof of Corollary 2.2.4.

Theorem 2.2.11 *Let S be a set of n points in the plane in general position, and let T be a k-set of S. Then for every $0 \leq j \leq (n-2)/2$, the number of j-edges uv with $u \in T$ and $v \in S \setminus T$ is exactly* $\min(j+1, k, n-k)$.

As an application, we prove the following bounds on e_j (the upper bound is due to Dey [51], the lower bound, as we shall see, is an easy consequence of Theorem 2.2.11).

Theorem 2.2.12 *For every set of n points in the plane in general position and for every $j < \frac{n-2}{2}$,*

$$2j + 3 \leq e_j \leq 8nj^{1/3}.$$

For every $j \geq 0$ and $n \geq 2j+3$, this bound is attained.

Proof. The proof of the upper bound is a rather straightforward extension of the proof of Theorem 2.2.6.

To prove the lower bound, consider any j-edge uv, and let e be the line obtained by shifting the line uv by a small distance so that u and v are also on the right hand side, and let T be the set of points of S on the smaller side of e. This will be the side containing u and v unless $n = 2j+3$; hence $|T| \geq j+1$. By Theorem 2.2.11, the number of j-edges xy with $x \in T$, $y \in S \setminus T$ is exactly $j+1$. Similarly, the number of j-edges xy with $y \in T$, $x \in S \setminus T$ is exactly $j+1$, and these are distinct from the others since $n \geq 2j+3$. Together with uv, this gives a total of $2j+3$ such pairs. □

The following construction shows that the lower bound in the previous theorem is sharp. Despite significant progress in recent years, the problem of determining the (order of magnitude of) the best upper bound remains open. The current best construction, due to Tóth [196], gives a set of points with $e_j \approx ne^{\Theta(\sqrt{\log k})}$.

Example 2.2.13 Let S_0 be a regular $(2j+3)$-gon, and let S_1 be any set of $n-2j-3$ points in general position very near the center of S_0.

Every line through any point in S_1 has at least $j+1$ points of S_0 on both sides, so the j-edges are the longest diagonals of S_0, which shows that their number is $2j+3$.

We now turn to estimating the number E_j of $\leq j$-edges. Theorem 2.2.12 immediately implies the following lower bound on the number of $(\leq j)$-edges:

$$\sum_{i=0}^{j} e_i \geq (j+2)^2 - 1. \tag{2.3}$$

This is, however, not tight, since the constructions showing the tightness of the lower bound $e_j \geq 2j+3$ are diffierent for different values of j. The following bounds can be proved by more complicated arguments.

Theorem 2.2.14 *Let S be a set of n points in the plane in general position, and $j < (n-2)/2$. Then*

$$3\binom{j+2}{2} \leq E_j \leq n(j+1).$$

.

The upper bound, due to Alon and Győri [10], is tight, as shown by a convex polygon; the lower bound [1, 142] is also tight for $k \leq n/3$, as shown by the following example:

Example 2.2.15 Let r_1, r_2 and r_3 be three rays emanating from the origin with an angle of 120° between each pair. Suppose for simplicity that n is divisible by 3 and let S_i be a set of $n/3$ points in general position, all very close to r_i but at distance at least 1 from each other and from the origin. Let $S = S_1 \cup S_2 \cup S_3$.

Then for $1 \leq k \leq n/3$, every k-set of S contains the i points farthest from 0 in one S_a, for some $1 \leq i \leq k$, and the $(k-i)$ points farthest from 0 in another S_b. Hence the number of k-sets is $3k$ and the number of $(\leq k)$-sets equals $3\binom{k+1}{2}$.

2.3 Rectilinear crossing number

The *rectilinear crossing number* of a graph G is the minimum number of crossings in a drawing of G in the plane with straight edges and nodes in general position. The Fáry–Wagner

Theorem 1.3.1 says that if the crossing number of a graph is 0, then so is its rectilinear crossing number. Is it true more generally that these two crossing numbers are equal? The answer is negative: the rectilinear crossing number may be strictly larger than the crossing number [183]. As a general example, the crossing number of the complete graph K_n is less than $\frac{3}{8}\binom{n}{4} = 0.375\binom{n}{4}$ (see (1.3)), while its rectilinear crossing number is at least $0.37501\binom{n}{4}$ if n is large enough [142].

The rectilinear crossing number of the complete n-graph K_n can also be defined as follows. Let S be a set of n points in general position in the plane, i.e., no three points are collinear. Four points in S may or may not form the vertices of a convex quadrilateral; if they do, we call this subset of 4 elements *convex*. We are interested in the number $c_S(S)$ of convex 4-element subsets. This can of course be as large as $\binom{n}{4}$, if S is in convex position, but what is its minimum?

The graph K_5 is not planar, and hence every 5-element set has at least one convex 4-element subset, from which it follows by straightforward averaging that at least $1/5$ of the 4-element subsets are convex for every $n \geq 5$.

The best upper bound on the minimum number of convex quadrilaterals is $0.3807\binom{n}{4}$, obtained using computer search by Aichholzer, Aurenhammer and Krasser [2]. As an application of the results on k-sets in Section 2.2.2, we prove the following lower bound [1, 142]:

Theorem 2.3.1 *Let S be a set of n points in the plane in general position. Then the number $c_S(S)$ of convex quadrilaterals determined by S is at least $0.375\binom{n}{4}$.*

The constant in the bound can be improved to 0.37501 [142]. This tiny improvement is significant, as we have seen, but we cannot give the proof here.

The first ingredient of our proof is an expression of the number of convex quadrilaterals in terms of j-edges of S.

Let c_S denote the number of 4-tuples of points in S that are in convex position, and let b_S denote the number of those in concave (i.e., not in convex) position. We are now ready to state the crucial lemma [205, 142], which expresses c_S as a positive linear combination of the numbers e_j (one might say, as the second moment of the distribution of j-edges).

Lemma 2.3.2 *For every set of n points in the plane in general position,*

$$c_S = \sum_{j<\frac{n-2}{2}} e_j \left(\frac{n-2}{2} - j\right)^2 - \frac{3}{4}\binom{n}{3}.$$

Proof. Clearly we have

$$c_S + b_S = \binom{n}{4}. \tag{2.4}$$

To get another equation between these quantities, let us count, in two different ways, ordered 4-tuples (u, v, w, z) such that w is on the right of the line uv and z is on the left of this line. First, if $\{u, v, w, z\}$ is in convex position, then we can order it in 4 ways to get such an ordered quadruple; if $\{u, v, w, z\}$ is in concave position, then it has 6 such orderings. Hence the number of such ordered quadruples is $4c_S + 6b_S$. On the other hand, any j-edge uv it can be completed to such a quadruple in $j(n - j - 2)$ ways. So we have

$$4c_S + 6b_S = \sum_{j=0}^{n-2} e_j j(n - j - 2). \tag{2.5}$$

From (2.4) and (2.5) we obtain

$$c_S = \frac{1}{2}\left(6\binom{n}{4} - \sum_{j=0}^{n-2} e_j(n - j - 2)j\right).$$

Using that

$$\sum_{j=0}^{n-2} e_j = n(n-1), \tag{2.6}$$

we can write

$$6\binom{n}{4} = \sum_{j=0}^{n-2} e_j \frac{(n-2)(n-3)}{4}$$

to get

$$c_S = \frac{1}{2}\sum_{j=0}^{n-2} e_j \left(\frac{(n-2)(n-3)}{4} - j(n-j-2)\right),$$

from which the lemma follows by simple computation, using (2.6) and that $e_j = e_{n-2-j}$. □

Having expressed c_S (up to some error term) as a positive linear combination of the e_j's, we can substitute any lower estimates for the numbers e_j to obtain a lower bound for c_S. Using (2.3), we get

$$c_S \geq \frac{1}{4}\binom{n}{4} + O(n^3).$$

This lower bound for c_S is quite weak. To obtain the stronger lower bound stated in Theorem 2.3.1, we do "integration by parts", i.e., we pass from j-facets to $(\leq j)$-facets. We substitute $e_j = E_j - E_{j-1}$ in Lemma 2.3.2 and rearrange to get the following:

Lemma 2.3.3 *For every set of n points in the plane in general position,*

$$c_S = \sum_{j<\frac{n-2}{2}} E_j(n - 2j - 3) - \frac{3}{4}\binom{n}{3} + \varepsilon_n,$$

where

$$\varepsilon_n = \begin{cases} \frac{1}{4}E_{\frac{n-3}{2}}, & \text{if } n \text{ is odd,} \\ 0, & \text{if } n \text{ is even.} \end{cases}$$

Note that the last two terms in the formula of this lemma are $O(n^3)$. Using these bounds, Theorem 2.3.1 follows.

2.4 Orthogonality graphs

For a given d, let G_d denote the graph whose nodes are all unit vectors in $\mathbb{R}^d$ (i.e., the points of the unit sphere S^{d-1}), and we connect two nodes if and only if they are orthogonal.

Example 2.4.1 The graph G_2 consists of disjoint 4-cycles.

Theorem 2.4.2 (Erdős–De Bruijn Theorem) *Let G be an infinite graph and $k \geq 1$, and integer. Then G is k-colorable if and only if every finite subgraph of G is k-colorable.*

Proof. See [129], Exercise 9.14. □

Theorem 2.4.3 *There exists a constant $c > 1$ such that*

$$c^d \leq \chi(G_d) \leq 4^d.$$

Chapter 3

Harmonic functions on graphs

In this section we consider general graphs, but some of the most important applications will concern planar graphs.

The notion of a harmonic function is basic in analysis, usually defined as smooth functions $f : \mathbb{R}^d \to \mathbb{R}$ (perhaps defined only in some domain) satisfying the differential equation $\Delta f = 0$, where $\Delta = \sum_{i=1}^{d} \partial^2/\partial x_i^2$ is the Laplace operator. It is a basic fact that such functions can be characterized by the "mean value property" that their value at any point equals to their average on a ball around this point.

Taking this second characterization as our starting point, we can define an analogous notion of harmonic functions defined on the nodes of a graph.

Harmonic functions play an important role in the study of random walks (after all, the averaging in the definition can be interpreted as expectation after one move). They also come up in the theory of electrical networks, and also in statics. This provides a connection between these fields, which can be exploited. In particular, various methods and results from the theory of electricity and statics, often motivated by physics, can be applied to provide results about random walks. From our point of view, their main applications will be rubber band representations and discrete analytic functions. In this chapter we develop this simple but useful theory.

3.1 Definition and uniqueness

Let $G = (V, E)$ be a connected simple graph and $S \subseteq V$. A function $f : V \to \mathbb{R}$ is called a *harmonic* at a node $v \in V$ if

$$\frac{1}{d_v} \sum_{u \in N(v)} f(u) = f(v) \qquad \forall v \in V \setminus S. \tag{3.1}$$

This equation can also be written as

$$\sum_{u \in N(v)} (f(u) - f(v)) = 0 \tag{3.2}$$

A node where a function is not harmonic is called a *pole* of the function.

We can extend this notion to multigraphs by replacing (3.1) by

$$\frac{1}{d_v} \sum_{\substack{e \in E \\ V(e) = \{u,v\} \ni v}} f(u) = f(v) \qquad \forall v \in V \setminus S. \tag{3.3}$$

Another way of writing this is

$$\sum_{u \in V} a_{uv}(f(u) - f(v)) = 0 \qquad \forall v \in V \setminus S, \tag{3.4}$$

where a_{uv} is the multiplicity of the edge uv. We could further generalize this by assigning arbitrary nonnegative weights β_{uv} to the edges of G ($uv \in E(G)$), and require the condition

$$\sum_{u \in V} \beta_{uv}(f(u) - f(v)) = 0 \qquad \forall v \in V \setminus S, \tag{3.5}$$

Every constant function is harmonic at each node. On the other hand,

Proposition 3.1.1 *Every non-constant harmonic function has at least two poles.*

Proof. Let S be the set where the function assumes its maximum, and let S' be the set of those nodes in S that are connected to any node outside S. Then every node in S' must be a pole, since in (3.1), every value $f(u)$ on the left had side is at most $f(v)$, and at least one is less, so the average is less than $f(v)$. Since the function is nonconstant, S is a nonempty proper subset of V, and since the graph is connected, S' is nonempty. So there is a pole where the function attains its maximum. Similarly, there is another pole where it attains its minimum. □

For any two nodes there will be a nonconstant harmonic function that is harmonic everywhere else. More generally, we have the following theorem.

Theorem 3.1.2 *For every nonempty set $S \subseteq V$ and every function $f_0 : S \to \mathbb{R}$ there is a unique function $f : V \to \mathbb{R}$ extending f_0 that is harmonic at each node of $V \setminus S$.*

We call this function f the *harmonic extension* of f_0.

The uniqueness of the harmonic extension is easy by the argument in the proof of Proposition 3.1.1. Suppose that f and f' are two harmonic extensions of f_0. Then $g = f - f'$ is harmonic on $V \setminus S$, and satisfies $g(v) = 0$ at each $v \in S$. If g is the identically 0 function,

then $f = f'$ as claimed. Else, either is minimum or its maximum is different from 0. But we have seen that both these sets contain at least one pole, which is a contradiction.

Before proving the existence of harmonic extensions in the next section, we make a few remarks.

If $|S| = 1$, then this unique extension must be the constant function. Thus Theorem 3.1.2 implies Proposition 3.1.1 (but it was instructive to prove that separately).

If $S = \{a, b\}$, then a function that is harmonic outside S is uniquely determined by two of its values, $f(a)$ and $f(b)$. Scaling the function by a real number and translating by a constant preserves harmonicity at each node, so if we know the function f with (say) $f(a) = 0$ and $f(b) = 1$, harmonic at $v \neq a, b$, then $g(v) = A+(B-A)f(v)$ describes the harmonic extension with $g(a) = A$, $g(b) = B$.

Note that equation (3.2) is equivalent to saying that $F_{ij} = f(j) - f(i)$ defines a flow from a to b.

3.2 Constructing harmonic functions

There are several ways to construct a harmonic function with given poles, and we describe four of them. This abundance of proof makes sense, because each of these constructions illustrates an application area of harmonic functions.

3.2.1 Linear algebra

First we construct a harmonic function with two given poles a and b. Let $\chi_a : V \to \mathbb{R}$ denote the function which is 1 on a and 0 everywhere else. Consider the equation

$$Lf = \chi_b - \chi_a,$$

(where L is the Laplacian of the graph). Every function satisfying this is harmonic outside $\{a, b\}$.

The matrix L is not quite invertible, but it has a one-dimensional nullspace spanned by the vector $\mathbf{1} = (1, \ldots, 1)^\mathsf{T}$, and so it determines f up to adding the same scalar to every entry. So we may assume that $\mathbf{1}^\mathsf{T} f = 0$. If $J \in \mathbb{R}^{V \times V}$ denotes the all-1 matrix, then $L + J$ is invertible, and

$$(L + J)f = Lf = \chi_b - \chi_a,$$

and so we can express f as

$$f = (L + J)^{-1}(\chi_b - \chi_a). \tag{3.6}$$

All other functions harmonic outside $\{a, b\}$ can be obtained from f by multiplying by a constant and adding another constant.

Second, suppose that $|S| \geq 3$, let $a, b \in S$, and let $g_{a,b}$ denote the function with is harmonic outside a and b, and satisfies $g_{a,b}(a) = 0$, $g_{a,b}(b) = 1$. Note that every linear combination of the functions $g_{a,b}$ is harmonic outside S.

Let $g'_{a,b}$ denote the restriction of $g_{a,b}$ to S. Fix a. We claim that the functions $g'_{a,b}$, $b \in S \setminus \{a\}$, are linearly independent. Indeed, a relation $\sum_b \alpha_{a,b} g'_{a,b} = 0$ implies that $\sum_b \alpha_{a,b} g_{a,b} = 0$ (by the uniqueness of the harmonic extension), which in turn implies $\sum_b \alpha_{a,b} L g_{a,b} = 0$. But the value of this last function at b is $\alpha_{a,b} L g_{a,b}(b)$, so every $\alpha_{a,b} = 0$.

So the functions $g'_{a,b}$ generate the $(|S| - 1)$-dimensional space of all functions $f \in \mathbb{R}^S$ with $f(a) = 0$. If we throw in the constant functions, they generate the whole space $\mathbb{R}^V$. This means that f_0 is a linear combination of functions $g'_{a,b}$ and a constant function. The corresponding linear combination of the $g_{a,b}$ is the harmonic extension of f.

3.2.2 Random walks

Let $G = (V, E)$ be a connected graph and $a, b \in V$. Let $f(v)$ denote the probability that a random walk starting at node v hits b before it hits a. Then f is harmonic at all nodes except a and b. Clearly we have $f(a) = 0$ and $f(b) = 1$.

More generally, given a set $S \subseteq V$ and a function $f_0 : S \to \mathbb{R}$, we define $f(v)$ for $v \in V \setminus S$ as the expectation of $f_0(s)$, where s is the (random) node where a random walk starting at v first hits S. (One needs some facts from the theory of random walks, for example, that a random walk hits the set S with probability 1.)

Then $f(v)$ is harmonic at $V \setminus S$. Indeed, let $v^0 = v, v^1, \ldots$ be the random walk. If $v \notin S$, then, conditioning on the first step,

$$\mathsf{E}(f(s)) = \sum_{j \in N(v)} \mathsf{P}(v^1 = j) \mathsf{E}(f(s) \mid v^1 = j) = \sum_{j \in N(v)} \frac{1}{d_v} f(j).$$

Moreover, it is clear that $f(v) = f_0(v)$ for all $v \in S$.

3.2.3 Electrical networks

Consider the graph G as an electrical network, where each edge represents a unit resistance. Assume that an electric current is flowing through G, entering at a and leaving at b. Let $f(v)$ be the potential of node v. Then f is a harmonic except at a and b.

More generally, given a set $S \subseteq V$ and a function $f_0 : S \to \mathbb{R}$, let a current flow through the network while keeping each $v \in S$ at potential $f_0(v)$. Then the potential $f(v)$ defined for all nodes v is an extension of f_0. Furthermore, the current through an edge uv is $f(u) - f(v)$ by Ohm's Law, and hence Kirchhoff's Current Law, (3.2) holds for every $v \in V \setminus S$.

3.2.4 Rubber bands

Consider the edges of the graph G as ideal rubber bands (or springs) with unit Hooke constant (i.e., it takes h units of force to stretch them to length h). Grab the nodes a and b and stretch the graph so that their distance is 1. Then the distance f(v) of node v from a is a function that is harmonic for $v \neq a, b$, and satisfies $f(a) = 0$ and $f(v) = 1$.

More generally, if we have a set $S \subseteq V$ and we fix the positions of each node $v \in S$ at a given point $f_0(v)$ of the real line, and let the remaining nodes find their equilibrium, then the position $f(v)$ of a node will be a function that extends f_0 (trivially), and that is harmonic at each $v \notin S$. To see the last fact, note that the edge uv pulls v with force $f(u) - f(v)$ (the force is positive if it pulls in the positive direction), and so (3.2) is just the condition of the equilibrium.

We'll come back to this construction repeatedly, extending it to higher dimensions. We'll prove that the equilibrium exists and is unique without any reference to physics.

3.2.5 Connections

A consequence of the uniqueness property in Theorem 3.1.2 is that the harmonic functions constructed in sections 3.2.1, 3.2.2, 3.2.3 and 3.2.4 are the same. As an application of this idea, we show the following interesting identities (see Nash-Williams [156], Chandra at al. [38]).

Considering the graph G as an electrical network, let R_{st} denote the effective resistance between nodes s and t. Considering the graph G as a spring structure in equilibrium, with two nodes s and t nailed down at 1 and 0, let F_{ab} denote the force pulling the nails. Doing a random walk on the graph, let $\kappa(a, b)$ denote the commute time between nodes a and b (i.e., the expected time it takes to start at a, walk until you first hit b, and then walk until you first hit a again).

Theorem 3.2.1

$$R_{ab} = \frac{1}{F_{ab}} = \frac{\kappa_{ab}}{2m}.$$

Proof. Consider the function f satisfying $Lf(a) = 1$, $\sum_v f(v) = 0$ and harmonic for $v \neq a, b$. By the discussion in section 3.2.3, $f(v)$ is equal to the potential of v if we push a unit current through G from a to b. So the effective resistance is $R_{ab} = f(b) - f(a)$.

The second equality is similarly easily derived from Hooke's Law.

Finally, we know that for an appropriate A and B, $Af(u) + B$ is the probability that a random walk starting at u visits a before b. Checking this for $u = s$ and $u = t$, we get that $A = 1/(f(b) - f(a))$ and $B = -f(u)/(f(b) - f(a))$. Hence $p_0 = \frac{1}{d_a} \sum_{u \in N(a)} (A(f(u) - f(a))$ is the probability that a random walk starting at a hits b before returning to t.

Let T be the first time when a random walk starting at a returns to a and S, the first time when it returns to a after visiting b. We know from the theory of random walks that $E(T) = 2m/d_a$ and by definition, $E(S) = \kappa(a, b)$. Clearly $T \leq S$ and the probability of $T = S$ is exactly p_0. This implies that $E(S - T) = (1 - p_0)E(S)$, since if $T < S$, then after the first T steps, we have to walk from a until we reach b and then return to b. Hence

$$p_0 = \frac{E(T)}{E(S)} = \frac{2m}{d_a \kappa(a, d)}.$$

□

For the rubber band model, imagine that we slowly stretch the graph until nodes a and b will be at distance 1. When they are at distance t, the force pulling our hands is tF_{ab}, and hence the energy we have to spend is

$$\int_0^1 tF_{ab}\, dt = \frac{1}{2}F_{ab}.$$

By conservation of energy, we get the identity

$$\sum_{ij \in E} (f(i) - f(j))^2 = F_{ab}. \tag{3.7}$$

Using the "topological formulas" from the theory of electrical networks for the resistance, we get a further characterization of these quantities:

Corollary 3.2.2 *Let G' denote the graph obtained from G by identifying a and b, and let $\mathcal{T}(G)$ denote the number of spanning trees of G. Then*

$$R_{ab} = \frac{\mathcal{T}(G)}{\mathcal{T}(G')}.$$

Another application of this equivalence between our three models can be used to prove the following.

Theorem 3.2.3 *Let $G = (V, E)$ be a graph and $a, b \in V$. Let G' be the graph obtained from G by connecting two nodes by a new edge.*

(a) *The commute time between a and b does not increase.*

(b) **(Raleigh's Theorem)** *The effective resistance between nodes a and b does not increase.*

(c) *If nodes a and b nailed down at 0 and 1, the force pulling the nails in the equilibrium does not decrease.*

Proof. The three statements are equivalent by Theorem 3.2.1; we prove (c). By (3.7), it suffices to prove that the equilibrium energy does not decrease. Consider the equilibrium

position of G', and delete the new edge. The contribution of this edge to the energy was nonnegative, so the total energy decreases. The current position of the nodes may not be in equilibrium; but the equilibrium position minimizes the energy, so when they move to the equilibrium of G', the energy further decreases. □

Exercise 3.1 Let a, b be two nodes of a connected graph G. We define the *hitting time* $H(a, b)$ as the expected number of steps before a random walk, started at a, will reach b.

Consider the graph as a rubber band structure as above, but also attach a weight of $d(v)$ to each node v. Hold the graph at b and let it find its equilibrium. Prove that a will be at distance $H(a, b)$ below b.

Chapter 4

Rubber bands

4.1 Rubber band representation

Let $G = (V, E)$ be a connected graph and $\emptyset \neq S \subseteq V$. Fix an integer $d \geq 1$ and a map $x^0 : S \to \mathbb{R}^d$. We extend this to a map $x : V \to \mathbb{R}^d$ (a geometric representation of G) as follows.

First, let's give an informal description. Replace the edges by ideal rubber bands (satisfying Hooke's Law). Think of the nodes in S as nailed to their given position (node $i \in S$ to $x_i^0 \in \mathbb{R}^d$), but let the other nodes settle in equilibrium. We'll see that this equilibrium position is uniquely determined. We call it the *rubber band representation* of G in $\mathbb{R}^d$ extending x^0.

To be precise, let $x_i \in \mathbb{R}^d$ be the position of node $i \in V$. By definition, $x_i = x_i^0$ for $i \in S$. The *energy* of this representation is defined as

$$\mathcal{E}(x) = \sum_{ij \in E} |x_i - x_j|^2$$

We want to find the representation with minimum energy, subject to the boundary conditions:

$$\text{minimize } \mathcal{E}(x) \tag{4.1}$$

$$\text{subsect to } x_i = x_i^0 \text{ for all } i \in S. \tag{4.2}$$

Lemma 4.1.1 *If $S \neq \emptyset$, then the function $\mathcal{E}(x)$ is strictly convex.*

Proof. We have

$$\mathcal{E}(x) = \sum_{ij \in E} \sum_{k=1}^{d} (x_{ik} - x_{jk})^2.$$

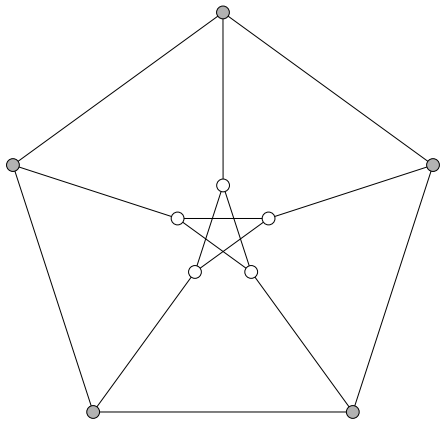

Figure 4.1: Rubber band representation of the Petersen graph, with one pentagon nailed (dark nodes).

Every function $(x_{ik} - x_{jk})^2$ is convex, so $\mathcal{E}$ is convex. Suppose that $\mathcal{E}(\frac{x+y}{2}) = \frac{1}{2}(\mathcal{E}(x) + \mathcal{E}(y))$ for some $x, y : V \to \mathbb{R}^d$. Then for every edge ij and every $1 \leq k \leq d$ we have

$$\left(\frac{x_{ik} + y_{ik}}{2} - \frac{x_{jk} + y_{jk}}{2}\right)^2 = \frac{(x_{ik} - x_{jk})^2 + (x_{ik} - x_{jk})^2}{2},$$

which implies that $x_{ik} - y_{ik} = x_{jk} - y_{jk}$. Since $x_{ik} = y_{ik}$ for every $i \in S$, it follows that $x_{ik} = y_{ik}$ for every $i \in V$. So $x = y$, which means that $\mathcal{E}$ is strictly convex. □

It is trivial that if any of the x_i tends to infinity, then $\mathcal{E}(x)$ tends to infinity (still assuming the boundary conditions 4.2 hold, where S is nonempty). With Lemma 4.1.1 this implies that the representation with minimum energy is uniquely determined. If $i \in V \setminus S$, then at the minimum point the partial derivative of $\mathcal{E}(x)$ with respect to any coordinate of x must be 0. This means that for every $i \in V \setminus S$,

$$\sum_{j \in N(i)} (x_i - x_j) = 0. \tag{4.3}$$

This we can rewrite as

$$x_i = \frac{1}{d_i} \sum_{j \in N(i)} x_j. \tag{4.4}$$

This equation means that *every free node is in the center of gravity of its neighbors.* Equation (4.3) also has a nice physical meaning: the rubber band connecting i and j pulls i with force $x_j - x_i$, so (4.3) states that the forces acting on i sum to 0 (as they should at the equilibrium).

We saw in Chapter 3 that every 1-dimensional rubber band representation gives rise to a harmonic function. Equation (4.4) extends this to higher dimensional rubber band representations:

Proposition 4.1.2 *In a rubber band representation, each coordinate function is harmonic at every free node.*

It will be useful to extend the rubber band construction to the case when the edges of G have arbitrary positive weights. Let w_{ij} denote the weight of the edge ij. We then define the energy function of a representation $i \mapsto x_i$ by

$$\mathcal{E}_w(x) = \sum_{ij \in E} w_{ij}|x_i - x_j|^2.$$

The simple arguments above remain valid: $\mathcal{E}_w$ is strictly convex if at least one node is nailed, there is a unique optimum, and for the optimal representation every $i \in V \setminus S$ satisfies

$$\sum_{j \in N(i)} w_{ij}(x_i - x_j) = 0. \tag{4.5}$$

This we can rewrite as

$$x_i = \frac{1}{\sum_{j \in N(i)} w_{ij}} \sum_{j \in N(i)} w_{ij} x_j. \tag{4.6}$$

Thus x_i is no longer in the center of gravity of its neighbors, but it is still a convex combination of them with positive coefficients. In other words, it is in the relative interior of the convex hull of its neighbors.

Exercise 4.1 Prove that $\mathcal{E}_{\min}(w) := \min_x \mathcal{E}_w(x)$ (where the minimum is taken over all representations x with some nodes nailed) is a concave function of w.

Exercise 4.2 Let $G = (V, E)$ be a connected graph, $\emptyset \neq S \subseteq V$, and $x^0 : S \to \mathbb{R}^d$. Extend x^0 to $x : V \setminus S \to \mathbb{R}^d$ as follows: starting a random walk at j, let i be the (random) node where S is first hit, and let x_j denote the expectation of the vector x^0_i. Prove that x is the same as the rubber band extension of x^0.

4.2 Rubber bands, planarity and polytopes

4.2.1 How to draw a graph?

The rubber band method was first analyzed by Tutte [197]. In this classical paper he describes how to use "rubber bands" to draw a 3-connected planar graph with straight edges and convex faces.

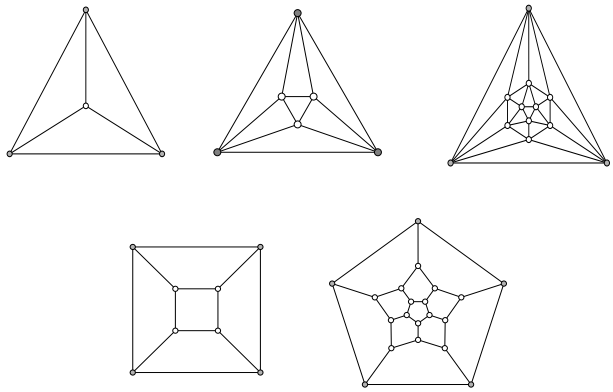

Figure 4.2: Rubber band representations of the skeletons of platonic bodies

Let $G = (V, E)$ be a 3-connected planar graph, and let F_0 be any face of it. Let C_0 be the cycle bounding F_0. Let us map the nodes of C_0 on the vertices of a convex polygon P_0 in the plane, in the same cyclic order. Let $i \mapsto v_i$ be the rubber band representation of G in the plane with extending this map. We also draw the edges of G as straight line segments connecting the appropriate endpoints. We call this mapping the *rubber band representation of G with C_0 nailed*. Figure 4.2 shows the rubber band representation of the skeletons of the five platonic bodies.

By the above, we know that each node not on C_0 is positioned at the center of gravity of its neighbors. Tutte's main result about this embedding is the following:

Theorem 4.2.1 *If G is a simple 3-connected planar graph, then its rubber band representation gives an embedding of G in the plane.*

Proof. The key to the proof is the following claim.

Claim 1. *Let ℓ be a line intersecting the polygon P_0, and let U be the set of nodes of G that are mapped onto a given (open) side of ℓ. Then U induces a connected subgraph of G.*

Clearly the nodes of C_0 in U form a path P. We may assume that ℓ is not parallel to any edge. Let $a \in S \setminus V(C_0)$, then v_a is in the center of gravity of its neighbors, and so either it is mapped onto the same point of the plane as all its neighbors (a degenerate case), or it has a neighbor a_1 such that v_{a_1} is on the same side of ℓ as a, but farther from ℓ. In the second case, we find a neighbor a_2 of a_1 such that v_{a_2} is on the same side of ℓ as v_{a_1}, but farther from ℓ, etc. This way we get a path Q in G that connects a to P, such that $V(Q) \subseteq S$.

In the degenerate case, consider all nodes mapped onto v_a, and a connected component H of the subgraph of G induced by them. If H contains a nailed node then it contains a path from a to P, all in U. Else, there must be an edge connecting a node a_1 in H to a node

a_2 with $v_{a_2} \neq v_a$, and then, since the system is in equilibrium, this node can be chosen so that v_{a_2} is at least as far from ℓ than v_a (here we use that no edge is parallel to ℓ). From here the proof goes just as in the nondegenerate case. (See Figure 4.3).

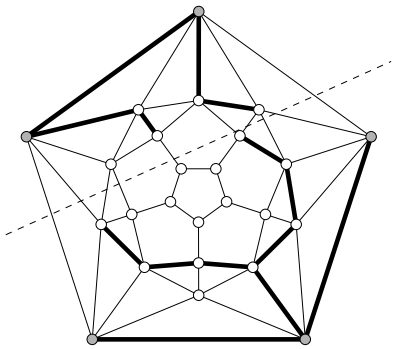

Figure 4.3: Every line cuts a rubber band embedding into connected parts.

Next, we exclude a possible degeneracy. Call a node *degenerate* if there is a line such that the node and all its neighbors get positioned on this line.

Claim 2. *No degenerate nodes exist.*

Suppose that there are degenerate nodes; then there is a line ℓ which contains a node and all its neighbors. Fix this line, and consider the subgraph induced by all degenerate nodes on ℓ with all neighbors also on ℓ, and let H be a connected component of this subgraph (H may be a single node). Let S be the set of neighbors of H (outside H). Then $|S| \geq 3$ by 3-connectivity.

Let U_1 and U_2 be the sets of nodes of G on the two sides of ℓ. We claim that each node a in S is connected to both U_1 and U_2. By the definition of S, a is positioned on ℓ, but it has a neighbor that is not on ℓ, and so it has a neighbor in $U_1 \cup U_2$. If $a \notin V(C_0)$, then a is the center of gravity of its neighbors, and so it cannot happen that it has a neighbor on one side of ℓ but not the other. If $a \in V(C_0)$, then its two neighbors along C_0 are on different sides of ℓ.

Now $V(H)$ induces a connected graph by definition, and U_1 and U_2 induce connected subgraphs by Claim 1. So we can contract these sets to single nodes. These three nodes will be adjacent to all nodes in S. So G can be contracted to $K_{3,3}$, which is a contradiction since it is planar. This proves Claim 2.

Claim 3. *Let ab be an edge that is not an edge of C_0, and let F_1 and F_2 be the two faces incident with ab. Then all other nodes of F_1 are mapped on one side of the line ℓ through v_a and v_b, and all other nodes of F_2 are mapped on the other side.*

Suppose not, then F_1 has a node c and F_2 has a node d such that v_c and v_d are both on (say) the positive side of ℓ, or on ℓ itself. In the latter case, they have a neighbor on the positive side of ℓ, by Claim 2. So by Claim 1, there is a path P connecting c and d whose internal nodes are positioned on the positive side of ℓ. Similarly, there is a path P' connecting a and b whose internal nodes are positioned on the negative side of ℓ. Thus P and P' are disjoint. But look at the planar embedding: the edge ab, together with P', forms a Jordan curve that separates b and d, so P cannot exist.

Claim 4. *The boundary of every face F is mapped onto a convex polygon P_F.*

This is immediate from Claim 3, since the line of an edge never intersects the interior of the face.

Claim 5. *The interiors of the polygons P_F (where F is a bounded face) are disjoint.*

Let x be a point inside P_0, we want to show that it is covered by one P_F only. Clearly we may assume that x is not on the image of any edge. Draw a line through x that does not go through the image any node, and see how many times its points are covered by interiors of such polygons. As we enter P_0, this number is clearly 1. Claim 2 says that as the line crosses an edge, this number does not change. So x is covered exactly once.

Now the proof is essentially finished. Suppose that the images of two edges have a common point. Then two of the faces incident with them would have a common interior point, which is a contradiction except if these faces are the same, and the two edges are consecutive edges of this face. □

Before going on, let's analyze this proof a little. The key step, namely Claim 1, is very similar to a basic fact concerning convex polytopes, namely Corollary 13.7.7. Let us call a geometric representation of a graph *section-connected*, if for every open halfspace, the subgraph induced by those nodes that are mapped into this halfspace is connected (or empty). The skeleton of a polytope, as a representation of itself, is section-connected; and so is the rubber-band representation of a planar graph. Note that the proof of Claim 1 did not make use of the planarity of G; in fact, the same proof gives:

Lemma 4.2.2 *Let G be a connected graph, and let w be a geometric representation of an induced subgraph H of G (in any dimension). If w is section-connected, then its rubber-band extension to G is also section-connected.*

4.2.2 How to lift a graph?

An old construction of Cremona and Maxwell can be used to "lift" Tutte's rubber band representation to a Steinitz representation.

Theorem 4.2.3 *Let $G = (V, E)$ be a 3-connected planar graph, and let T be a triangular face of G. Let*

$$i \mapsto v_i = \begin{pmatrix} v_{i1} \\ v_{i2} \end{pmatrix} \in \mathbb{R}^2$$

be a rubber band representation of G obtained by nailing T to any triangle in the plane. Then we can assign a number $\eta_i \in \mathbb{R}$ to each $i \in V$ such that $\eta_i = 0$ for $i \in V(T)$, $\eta_i > 0$ for $i \in V \setminus V(T)$, and the mapping

$$i \mapsto u_i = \begin{pmatrix} v_{i1} \\ v_{i2} \\ \eta_i \end{pmatrix} \in \mathbb{R}^3$$

is a Steinitz representation of G.

Before starting with the proof, we need a little preparation to deal with edges on the boundary triangle. Recall that we can think of $F_{ij} = v_i - v_j$ as the force with which the edge ij pulls its endpoint j. Equilibrium means that for every internal node j,

$$\sum_{i \in N(j)} F_{ij} = 0. \tag{4.7}$$

This does not hold for the nailed nodes, but we can modify the definition of F_{ij} along the three boundary edges so that (4.7) will hold for all nodes (this is the only point where we use that the outer face is a triangle). This is natural by a physical argument: let us replace the outer edges by rigid bars, and remove the nails. The whole structure will remain in equilibrium, so appropriate forces must act in the edges ab, bc and ac to keep balance.

To translate this to mathematics, one has to work a little; this is leaft to the reader as an exercise.

Exercise 4.3 Let $G = (V, E)$ be a simple 3-connected planar graph with a triangular face T. Let v be a rubber band representation of G in the plane with $\{a, b, c, \}$ nailed. Define $F_{ij} = v_i - v_j$ for all edges in $E \setminus E(T)$. Then we can define F_{ij} for $i, j \in V(T)$, $i \neq j$ so that

(a) $F_{ij} = -F_{ji}$ for all $ij \in E$,

(b) (4.7) holds for all nodes $i \in V$,

(c) F_{ij} is parallel to $v_j - v_i$.

Now we are ready to prove theorem 4.2.3.

Proof. Imagine that we have the proper lifting. Let's call the third coordinate direction "vertical". For each face F, let g_F be a normal vector. Since no face is parallel to a vertical line, we can normalize g_F so that its third coordinate is 1. Clearly for each face F, g_F will be an outer normal, except for $F = T$, when g_F is an inner normal.

Write $g_F = \binom{h_F}{1}$. Let ij be any edge of G, and let F_1 and F_2 be the two faces incident with ij. Then both g_{F_1} and g_{F_2} are orthogonal to the edge u_iu_j of the polytope, and therefore so is their difference. Since

$$(g_{F_1} - g_{F_2})^{\mathsf{T}}(u_i - u_j) = \left(\begin{pmatrix} h_{F_1} - h_{F_2} \\ 0 \end{pmatrix}\right)^{\mathsf{T}} \left(\begin{pmatrix} v_i - v_j \\ \eta_i - \eta_j \end{pmatrix}\right) = (h_{F_1} - h_{F_2})^{\mathsf{T}}(v_i - v_j),$$

we get

$$(h_{F_1} - h_{F_2})^{\mathsf{T}}(v_i - v_j) = 0. \tag{4.8}$$

We also have

$$h_T = 0,$$

since the facet T is not lifted.

Using that not only $g_{F_1} - g_{F_2}$, but also g_{F_1} is orthogonal to the edge v_iv_j, we get from

$$g_{F_1}^{\mathsf{T}}(u_i - u_j) = h_{F_1}^{\mathsf{T}}(v_i - v_j) + (\eta_i - \eta_j)$$

that

$$\eta_i - \eta_j = -h_{F_1}^{\mathsf{T}}(v_i - v_j). \tag{4.9}$$

This discussion allows us to explain the plan of the proof: given the Tutte representation, we first reconstruct the vectors h_i so that all equations (4.8) are satisfied, then using these, we reconstruct the numbers η_i so that equations (4.9) are satisfied. It will not be hard to verify then that we get a Steinitz representation.

Let R denote the counterclockwise rotation in the plane by 90°. We claim that we can replace (4.8) by the stronger equation

$$h_{F_1} - h_{F_2} = RF_{ij} \tag{4.10}$$

and still have a solution. Starting with $h_T = 0$, and moving from face to adjacent face, this equation will determine the value of h_F for every face. What we have to show is that we don't run into contradiction, i.e., if we get to the same face F in two different ways, then we get the same vector h_F. This is equivalent to saying that if we walk around a closed cycle of faces, then the total change in the vector h_F is zero. We can think of this closed cycle as a Jordan curve in the plane, that does not go through any nodes, and crosses every edge at most once. We want to show that the sum of RF_{ij} over all edges that it crosses is zero (where the order of the endpoints is determined so that the Jordan curve crosses the edge from right to left).

From (4.7) we have that

$$\sum_{i \in N(j)} RF_{ij} = 0.$$

Summing this over all nodes j for which x_j lies in the interior of the Jordan curve, the terms corresponding to edges with both endpoints inside cancel (since $F_{ij} + F_{ji} = 0$), and we get that the sum is 0 for the edges crossing the Jordan curve. This proves that we can define the vectors h_F.

Second, we construct numbers η_i satisfying (4.9) by a similar argument. We set $\eta_i = 0$ if i is an external node. Equation (4.9) tells us what the value at one endpoint of an edge must be, if we have it for the other endpoint.

Again, the main step is to prove that we don't get a contradiction when coming back to a value we already defined.

The first concern is that (4.9) gives two conditions for each, depending on which face incident with it we choose. But if F_1 and F_2 are two faces incident with the edge ij, then

$$h_{F_1}^{\mathsf{T}}(v_i - v_j) - h_{F_2}^{\mathsf{T}}(v_i - v_j) = (h_{F_1} - h_{F_2})^{\mathsf{T}}(v_i - v_j) = (RF_{ij})^{\mathsf{T}}(v_i - v_j) = 0,$$

since F_{ij} is parallel to $v_i - v_j$ and so RF_{ij} is orthogonal to it. Thus the two conditions on the difference $\eta_i - \eta_j$ are the same; in other words,

$$\chi_{ij} = -h_F^{\mathsf{T}}(v_i - v_j)$$

depends only on the edge ij.

Now consider a cycle C in G, and (for reference) orient it counterclockwise. We want to show that the total change of η_i we prescribed along this cycle is 0. For every face F with boundary cycle ∂F, we have

$$\sum_{ij \in E(\partial F)} \chi_{ij} = \sum_{ij \in E(\partial F)} h_F^{\mathsf{T}}(v_i - v_j) = 0.$$

Summing this over all faces inside C the contribution of every edge cancels except for the edges on C. This proves that

$$\sum_{ij \in E(C)} \chi_{ij} = 0$$

as claimed.

Now define $u_i = \binom{v_i}{\eta_i}$ for every node i and $g_F = \binom{h_F}{1}$ for every face F. It remains to prove that $i \mapsto u_i$ maps the nodes of G onto the vertices of a convex polytope, so that edges go to edges and faces go to faces. We start with observing that if F is a face and ij is an edge of F, then

$$g_F^{\mathsf{T}} u_i - g_F^{\mathsf{T}} u_j = h_F^{\mathsf{T}}(v_i - v_j) + (\eta_i - \eta_j) = 0,$$

and hence there is a scalar α_F so that all nodes of F are mapped onto the hyperplane $g_F^{\mathsf{T}} x = \alpha_F$. We know that the image of F under $i \mapsto v_i$ is a convex polygon, and so the same follows for the map $i \mapsto u_i$.

To conclude, it suffices to prove that if ij is any edge, then the two convex polygons obtained as images of faces incident with ij "bend" in the right way; more exactly, let F_1 and F_2 be the two faces incident with ij, and let Q_{F_1} and Q_{F_2} be two corresponding convex polygons. We claim that Q_{F_2} lies on the same side of the plane $g_{F_1}^{\mathsf{T}} x = \alpha_{F_1}$ as the bottom face. Let x be any point of the polygon Q_{F_2} not on the edge $u_i u_j$. We want to show that $g_{F_1}^{\mathsf{T}} x < \alpha_{F_1}$. Indeed,

$$g_{F_1}^{\mathsf{T}} x - \alpha_{F_1} = g_{F_1}^{\mathsf{T}} x - g_{F_1}^{\mathsf{T}} u_i = g_{F_1}^{\mathsf{T}}(x - u_i) = (g_{F_1} - g_{F_2})^{\mathsf{T}}(x - u_i)$$

(since both x and u_i lie on the plane $g_{F_2}^{\mathsf{T}} x = \alpha_{F_2}$),

$$= \begin{pmatrix} h_{F_1} - h_{F_2} \\ 0 \end{pmatrix}^{\mathsf{T}} (x - u_i) = (h_{F_1} - h_{F_2})^{\mathsf{T}}(x' - v_i)$$

(where x' is the projection of x onto the first two coordinates)

$$= (RF_{ij})^{\mathsf{T}}(x' - v_i) < 0$$

(since x' lies on the right hand side of the edge $v_i v_j$). This completes the proof. □

Theorem 4.2.3 proves Steinitz's theorem in the case when the graph has a triangular face. We are also home if the dual graph has a triangular face; then we can represent the dual graph as the skeleton of a 3-polytope, choose the origin in the interior of this polytope, and consider its polar; this will represent the original graph.

So the proof of Steinitz's theorem is complete, if we prove the following simple fact:

Lemma 4.2.4 *Let G be a 3-connected simple planar graph. Then either G or its dual has a triangular face.*

Proof. If G^* has no triangular face, then every node in G has degree at least 4, and so

$$|E(G)| \geq 2|V(G)|.$$

If G has no triangular face, then similarly

$$|E(G^*)| \geq 2|V(G^*)|.$$

Adding up these two inequalities and using that $|E(G)| = |E(G^*)|$ and $|V(G)| + |V(G^*)| = |E(G)| + 2$ by Euler's theorem, we get

$$2|E(G)| \geq 2|V(G)| + 2|V(G^*)| = 2|E(G)| + 4,$$

a contradiction. □

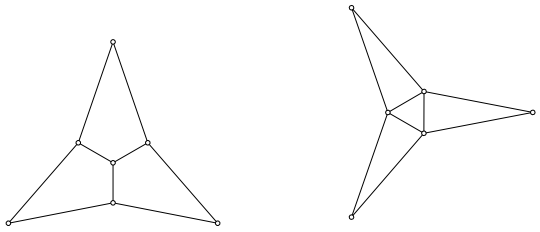

Figure 4.4: Rubber band representation of a cube with one node deleted, and of an octahedron with the edges of a triangle deleted. Corresponding edges are parallel and have the same length.

Exercise 4.4 Prove that every Schlegel diagram with respect to a face F can be obtained as a rubber band representation of the skeleton with the nodes of the face nailed (the strengths of the rubber bands must be chosen appropriately).

Exercise 4.5 Let G be a 3-connected planar graph with a triangular face abc. Let p denote the infinite face, and let q, r, s be the faces neighboring p. Let G^* be the dual graph. Consider a rubber band representation $x: V(G) \to \mathbb{R}^2$ of G with a, b, c nailed down, and also a rubber band representation $y: V(G^*) \setminus \{p\} \to \mathbb{R}^2$ of $G \setminus p$ with q, r, s nailed down (both with unit rubber band strengths). Prove that the positions of q, r, s can be chosen so that for every edge $ij \in E(G) \setminus \{ab, bc, ca\}$, and the corresponding edge $uv \in E(G^*)$ (where the labeling is chosen so that moving from i to j, the node u is on the left hand side), we have $x_i - x_j = y_u - y_v$ (Figure 4.4).

4.2.3 How to find the starting face?

Can we use Tutte's method to test a 3-connected graph for planarity? The main difficulty is that it assumes that we already know a face. So it may be a good method for designing actual drawings of these graphs, it seems to assume that we already know the combinatorial structure of the embedding.

But we only need to know *one* face to be able to apply Tutte's method, and it turns out that to find one face is easier than to find the whole embedding.

Recall that a *depth-first search spanning tree* in a connected graph G with root $r \in V$ is a spanning tree T with the property that for every edge ij of G, both endnodes of G lie on a path starting at r. It is very easy to find a depth-first search spanning tree in a graph by the depth-first search algorithm.

Lemma 4.2.5 *Let G be any 3-connected graph with any node r specified as its root, and let T be a depth-first search spanning tree in G. Let $ij \in E \setminus E(T)$, and let P be a path in T*

starting at R containing both endnodes of ij. Let i be he first endnode of ij P encounters. Let P' be the subpath of P between r and i and P'', the subpath between i and j. Choose ij so that P' is maximal (with respect to inclusion). Among these edges, choose one so that P'' is minimal. Then the cycle C formed by P'' and ij is non-separating.

Proof. Observe that $i \neq r$; indeed $i = r$ could only happen if all edges in $E \setminus E(T)$ would be incident with r, but any endnode of T different from r is incident with at least two edges in $E \setminus E(T)$, one of which does not go to r.

Suppose that C has a chord. This is an edge in $E \setminus T$, and either it connects two points of P'' different from i (which contradicts the maximality of P'), or it connects i to a node of P'' different from j (which contradicts the minimality of P'').

Second, suppose that $G \setminus V(C)$ is disconnected. Then $V(C)$ separates some node x from r. Walk from x along the tree to r, and let k be the first node of C we encounter. let T' be the subtree of T separated (in T) from r by x.

There is a path Q from x to r avoiding the pair $\{i, k\}$ (by 3-connectivity). Traversing this path from x, let v be its first node not in T', and uv its edge before v. By the depth-first search property of T, this node must be on P between k and r, and we know it is different from i and k. This also implies that $uv \in E \setminus E(T)$.

Now if v is between i and r, then walking from v to r along P we get a path from x to r avoiding C, which is impossible. If v is between i and k, then the edge uv contradicts the maximality of P'. □

Now if G is planar, then the cycle C in Lemma 4.2.5 is a face by Lemma 1.1.5, and so we can use it as the starting face in Tutte's method.

4.3 Rubber bands and connectivity

The idea of rubber bands can be related to graph connectivity, and can be used to give a test for k-connectivity of a graph (Linial, Lovász and Wigderson [126]).

4.3.1 Convex embeddings

Let $G = (V, E)$ be a graph and $S \subset V$, a fixed set of its nodes. A *convex representation of G* (in dimension d, with boundary S) is an mapping of V to $\mathbb{R}^d$ such that every node in $V \setminus S$ is in the convex hull of its neighbors. The representation is *in general position* if any $d + 1$ representing points are affine independent.

The following trivial fact is nevertheless useful to state:

Proposition 4.3.1 *If x is a convex representation of a graph $G = (V, E)$ with boundary S, then $x(V) \subseteq \mathrm{conv}(x(S))$.*

Our discussion when introducing rubber bands implies the following:

Proposition 4.3.2 *The rubber band representation extending any map from $S \subseteq V$ to $\mathbb{R}^d$ is convex with boundary S.*

(Not every convex representation is constructible this way.)

Convex representations in $\mathbb{R}^1$ with a two-element boundary set $\{s, t\}$ can be defined combinatorially, and they constitute an important tool for graph algorithms.

Let $G = (V, E)$ be a graph on n nodes and $s, t \in V$, $s \neq t$. A labeling $\ell : V \to \{1, \ldots, n\}$ is called an $s-t$-*numbering* if $\ell(s) = 1$, $\ell(t) = n$, and every node $i \in V \setminus \{s, t\}$ has a neighbor j with $\ell(j) < \ell(i)$ and another neighbor j with $\ell(j') > \ell(i)$.

We quote the following well known graph-theoretic fact without proof.

Proposition 4.3.3 *If $G = (V, E)$ is a 2-connected graph, then it has an $s-t$-numbering for every $s, t \in V$, $s \neq t$.*

Exercise 4.6 (a) Prove proposition 4.3.3.

(b) Show that instead of the 2-connectivity of G, it suffices to assume that deleting any node, either the rest is connected or it has two components, one containing s and one containing t.

(c) Consider the harmonic extension f of any function $f_0 : \{s, t\} \to \mathbb{R}$. Prove that if the values of f are distinct, then the ordering of the values gives an $s - t$-numbering.

4.3.2 Degeneracy: essential and non-essential

We start with a discussion of what causes degeneracy in rubber band embeddings. Consider the two graphs in Figure 4.5. It is clear that if we nail the nodes on the convex hull, and then let the rest find its equilibrium, then there will be a degeneracy: the grey nodes will all movetothe same position. However, the reasons for this degeneracy are different: In the first case, it is due to symmetry; in the second, it is due to the node that separates the grey nodes from the rest, and thereby pulls them onto itself.

One can distinguish the two kinds of degeneracy as follows: In the first graph, the strengths of the rubber bands must be strictly equal; varying these strengths it is easy to break the symmetry and thereby get rid of the degeneracy. However, in the second graph, no matter how we change the strengths of the rubber bands (as long as they remain positive), the grey nodes will always be pulled together into one point.

Figure 4.6 illustrates a bit more delicate degeneracy. In all three pictures, the grey points end up collinear in the rubber band embedding. In the first graph, the reason is symmetry again. In the second, there is a lot of symmetry, but it does not explain why the three grey nodes are collinear in the equilibrium. (It is not hard to argue though that they are collinear: a good exercise!) In the third graph (which is not drawn in its equilibrium position, but before it) there are two nodes separating the grey nodes from the nailed nodes, and the grey

nodes will end up on the segment connecting these two nodes. In the first two cases, changing the strength of the rubber bands will pull the grey nodes off the line; in the third, this does not happen.

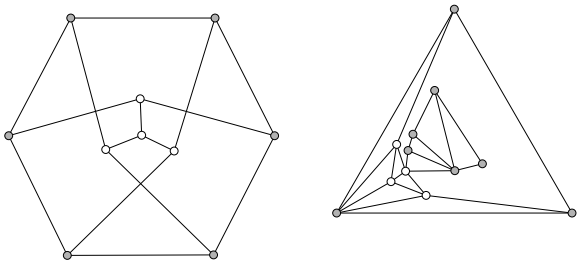

Figure 4.5: Two reasons why two nodes and up on top of each other: symmetry, or a separating node

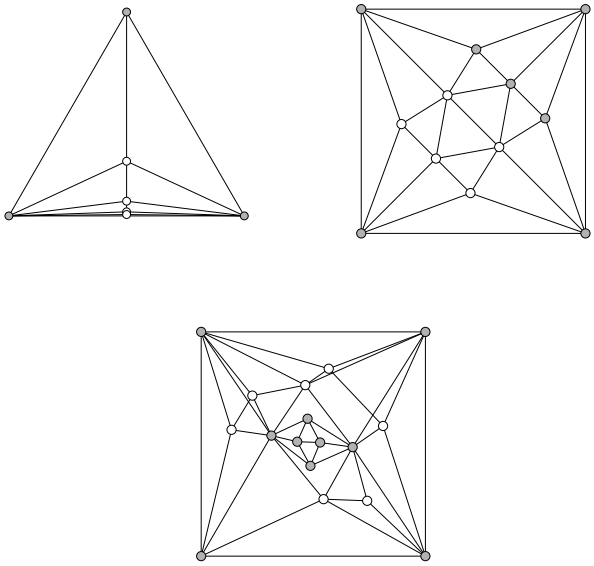

Figure 4.6: Three reasons why three nodes can end up collinear: symmetry, a separating pair of nodes, or just accident

4.3.3 Connectivity and degeneracy

For $X, Y \subseteq V$, we denote by $\kappa(X, Y)$ the maximum number of vertex disjoint paths from X to Y (if X and Y are not disjoint, some of these paths may be single nodes). By Menger's

Theorem, $\kappa(X, Y)$ is the minimum number of nodes that cover all $X - Y$ paths in G. We say that X and Y are *fully connected* if $|X| = |Y| = \kappa(X, Y)$. The graph G is k-connected if and only if $|V| > k$ and any two k-subsets are fully connected. The largest k for which this holds is the *vertex-connectivity* of G, denoted $\kappa(G)$. By convention, the complete graph K_n is $(n-1)$-connected but not n-connected.

Lemma 4.3.4 *Let $S, T \subseteq V$. Then for every convex representation x of G with boundary S, $\mathrm{rk}(x(T)) \leq \kappa(S, T)$.*

Proof. There is a subset $U \subseteq V$ with $|U| = \kappa(S, T)$ such that $V \setminus U$ contains no (S, T)-paths. Let W be the union of connected components of $G \setminus U$ containing a vertex from T. Then x, restricted to W, gives a convex representation of $G[W]$ with boundary U. Hence by Proposition 4.3.1, $x(W) \subseteq \mathrm{conv}(x(U))$, and so

$$\mathrm{rk}(x(T)) \leq \mathrm{rk}(x(W)) = \mathrm{rk}(x(U)) \leq |U| = \kappa(S, T).$$

□

The Lemma gives a lower bound on the connectivity between two sets S and T. The following theorem asserts that if we take the best convex representation, this lower bound is tight:

Theorem 4.3.5 *Let $G = (V, E)$ be a graph and $S, T \subseteq V$ with $\kappa(S, T) = d + 1$. Then G has a convex representation in $\mathbb{R}^d$ with boundary S such that $\mathrm{rk}(x_T) = d + 1$.*

Corollary 4.3.6 *Let $G = (V, E)$ be a graph, $d \geq 1$ and $S \subseteq V$. Then G has a general position convex representation in $\mathbb{R}^d$ with boundary S if and only if no node of G can be separated from S by fewer than $d + 1$ nodes.*

Corollary 4.3.7 *A graph G is k-connected if and only if for every $S \subseteq V$ with $|S| = k$, G has a general position convex representation with boundary S.*

To prove Theorem 4.3.5, we have to construct a convex representation, and we use 4.3.2. However, a rubber band representation will not automatically satisfy this: a subset of nodes can be affine dependent (say, three nodes can fall on one line) because of symmetry, or just by coincidence. We'll have to choose "generic" edge weights. We are going to prove:

Theorem 4.3.8 *Let $G = (V, E)$ be a graph and $S, T \subseteq V$ with $\kappa(S, T) \geq d + 1$. Choose a random edge weight $c_{ij} \in (0, 1]$ for every edge ij, independently and uniformly. Map the nodes of S into $\mathbb{R}^d$ so that the images of any $d + 1$ nodes are affine independent. Then with probability 1, the rubber band extension of this map is a convex representation of G in $\mathbb{R}^d$ with boundary S such that $\mathrm{rk}(x_T) = d + 1$.*

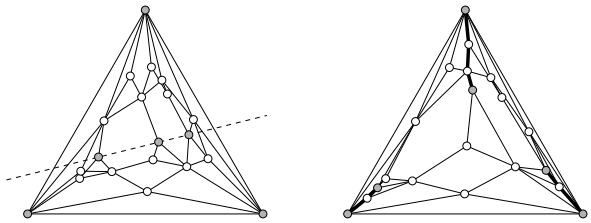

Figure 4.7: Three nodes accidentally on a line. Making the edges along the three paths shown very strong pulls them apart.

Proof. The proof will consist of two steps: first, we show that there is *some* choice of the edgeweights for which the conclusion holds; then we use this to prove that the conclusion holds for *almost all* choices of edge-weights.

For the first step, we use that by Menger's Theorem, there are $d+1$ disjoint paths $P_0, \ldots, P_d$ such that each P_i connects a node $s_i \in S$ with a node $t_i \in T$. The idea is to make the rubber bands on these paths very strong (while keeping the strength of the other edges fixed). Then these paths will pull each node t_i very close s_i. Since the positions of $s_0, \ldots s_d$ are affine independent, so are the positions of the nodes $s_0, \ldots s_d$ (Figure 4.7).

To make this precise, let D be the diameter of the set $\{x_i : i \in S\}$, let $E' = \cup_i E(P_i)$, fix any $R > 0$, and define a weighting $w = w_R$ of the edges by

$$w_e = \begin{cases} R, & \text{if } e \in E', \\ 1, & \text{otherwise.} \end{cases}$$

Let x be the rubber band extension of the given mapping of the nodes of S with these strengths.

Recall that f minimizes the potential $\mathcal{E}_w$ over all representations of G with the given nodes nailed. Let y be the representation with $y_j = x_j$ if $j \in S$, $y_j = s_i$ if $j \in P_i$ and (say) $f'(v) = s_0$ for any node not in S or any P_i. In y the edges with strength are have 0 length, and so

$$\mathcal{E}_w(x) \le \mathcal{E}_w(y) \le D^2|E|.$$

On the other hand,

$$\mathcal{E}_w(x) \ge \sum_{uv \in E'} R\|x_u - x_v\|^2.$$

By the Cauchy-Schwartz inequality we get that

$$\|x_{s_i} - x_{t_i}\| \le \sum_{uv \in E(P_i)} \|x_u - x_v\| \le \left(|E(P_i)| \sum_{uv \in E(P_i)} \|x_u - x_v\|^2 \right)^{1/2}$$
$$\le \left(\frac{n}{R} \mathcal{E}_w(x) \right)^{1/2} \le \frac{\sqrt{n|E|}D}{\sqrt{R}}.$$

Thus $x_{t_i} \to x_{s_i}$ if $R \to \infty$.

Since $x_{s_0}, \dots, x_{s_d}$ are affine independent, the columns of the matrix $M = \begin{pmatrix} 1 & \dots & 1 \\ x_{s_0} & \dots & x_{s_d} \end{pmatrix}$ are linearly independent, and so the matrix $M^{\mathsf{T}}M = (1 + x_{s_i}^{\mathsf{T}} x_{s_j})_{i,j=0}^{d}$ is nonsingular. So $\det(M^{\mathsf{T}}M) \neq 0$. But $1 + x_{t_i}^{\mathsf{T}} x_{t_j} \to 1 + x_{s_i}^{\mathsf{T}} x_{s_j}$, so the matrix $(1 + x_{t_i}^{\mathsf{T}} x_{t_j})_{i,j=0}^{d}$ is nonsingular if R is large enough. This proves that $x_{t_0}, \dots, x_{t_d}$ are affine independent.

This completes the first step. Now we argue that this holds for almost all edge-weightings. To prove this, we only need some general considerations. The embedding minimizing the energy is unique, and so it can be computed from the equations (4.5) (say, by Cramer's Rule). What is important from this is that the x_i can be expressed as rational functions of the edgeweights. Furthermore, the value $\det((1 + x_{t_i}^{\mathsf{T}} x_{t_j})_{i,j=0}^{d})$ is a polynomial in the coordinates of the x_i, and so it is also a rational function of the edgeweights. We know that this rational function is not identically 0; hence it follows that it is 0 only on a set of weight-vectors of measure 0. □

4.3.4 Turning the method to an algorithm

Computing the rubber band representation

Let $G = (V, E)$ be a connected graph on n nodes and $S \subseteq V$, $|S| = k$. Given a map $x : S \to \mathbb{R}^d$, we can compute its rubber band extension by solving the system of linear equations

$$\sum_{j \in N(i)} c_{ij}(x_i - x_j) = 0 \qquad (i \in V \setminus S). \tag{4.11}$$

This system has $(n-k)d$ unknowns and the same number of equations, and we know that it has a unique solution, since this is where the gradient of a strictly convex function (which tends to ∞ at ∞) vanishes.

At the first sight, solving (4.11) takes inverting an $(n-k)d \times (n-k)d$ matrix. However, we can immediately see that the coordinates can be computed independently, and since they satisfy the same equations except for the right hand side, it suffices to invert the matrix of the system once.

Below, we shall have to compute the rubber band representations of the *same* graph, changing only the nailed set S. Can we make use of some of the computation done for one of these representations when computing the others?

The answer is yes. First, we do something which seems to make things worse: We create new "equilibrium" equations for the nailed nodes, introducing new variables for the force that acts on the nail. So we can write (4.11) as

$$L_c x = \begin{pmatrix} y \\ 0 \end{pmatrix}, \tag{4.12}$$

where $y \in \mathbb{R}^S$, and L_c is the matrix

$$(L_c)_{ij} = \begin{cases} -c_{ij} & \text{if } ij \in E, \\ \sum_{k \in N(i)} c_{ik}, & \text{if } i = j, \\ 0, & \text{otherwise.} \end{cases}$$

Now we use the same trick as we did for harmonic functrions: L_c is "almost" invertible in the sense that the only vector in its nullspace in $\mathbf{1}$. Hence $L_c + J$ is invertible; let $M = (L_c + J)^{-1}$. Since $(L_c + J)J = J^2 = nJ$, it follows that $MJ = \frac{1}{n}J$. Furthermore, from $M(L_c + J) = I$ it follows that $ML_c = I - MJ = I - \frac{1}{n}J$. Hence multiplying (4.12) by M from the left, we get

$$\Big(I - \frac{1}{n}J\Big)x = M\begin{pmatrix} y \\ 0 \end{pmatrix},$$

or

$$x = z\mathbf{1} + M\begin{pmatrix} y \\ 0 \end{pmatrix}, \tag{4.13}$$

where $z = \frac{1}{n}\sum_{i \in v} x_i$.

First, let us look at the equations in (4.13) that correspond to nodes in S. Here the left hand side is given, so we get k equations in the variables z and y.***

Nondegeneracy and connectivity testing

Rubber band representations yield a (randomized) graph connectivity algorithm with good running time.

We start with describing a test checking whether or not two given k-tuples X and Y of nodes are fully linked. For this, it suffices to map the nodes in X into $\mathbb{R}^{k-1}$ in general position, say one node is mapped onto the origin and the others to the basis vectors e_i. Then we compute the rubber band representation extending this map, and check whether the points of Y are in general position.

If we want to apply a linkedness test for connectivity testing, it seems that we have to apply it for all pairs of k-sets, which is too much. The following lemma shows how to get around this.

Lemma 4.3.9 *For every vertex $v \in V$ we select an arbitrary k-subset $S(v)$ of $N(v)$. Then G is k-connected iff $S(u)$ and $S(v)$ are linked for every u and v.*

Proof. The "only if" part follows from the well-known property of k-connected graphs that any two k-subsets are linked. The "if" part follows from the observation that if $S(u)$ and $S(v)$ are linked then u and v are connected by k openly disjoint paths. □

Thus the linkedness subroutine needs be called at most $O(n^2)$ times. But we do not even have to check this for every pair (u, v), if we use the following simple lemma.

Lemma 4.3.10 *Let G be any graph and H, a k-connected graph with $V(H) = V(G)$. Then G is k-connected iff u and v are connected by k openly disjoint paths in G for every edge $uv \in E(H)$.*

Proof. The "only if" part is trivial. To prove the "if" part, suppose that G has a cutset S with $|S| < k$. Let G_1 be a connected component of G. Since S cannot be a cutset in H, there is an edge $uv \in E(H)$ connecting $V(G_1)$ to $V(G) \setminus S \setminus V(G_1)$. But the S separates u from v in G, a contradiction. □

This lemma implies that it suffices to check that $S(u)$ and $S(v)$ are linked for every edge $uv \in E(H)$. This means $O(nk)$ calls on a linkedness subroutine rather than $O(n^2)$. (If we allow randomization then we can do even better; we refer to [126] for details.)

Numerical issues

(See Figure 4.8.)

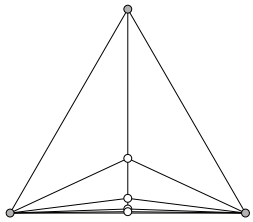

Figure 4.8: The rubber band embedding gives nodes that are exponentially close.

The computation of the rubber band representation requires solving a system of linear equations. We have seen that for a graph with n nodes, the positions of the nodes can get exponentially close in a rubber band representation, which means that we might have to compute with exponentially small numbers (in n), which means that we have to compute with linearly many digits, which gives an extra factor of n in the running time. For computational purposes it makes sense to solve the system in a finite field rather than in $\mathbb{R}$. Of course, this "modular" embedding has no physical or geometrical meaning any more, but the algebraic structure remains!

Let $G(V,E)$, $|V| = n$, $X, Y \subseteq V$, $|X| = |Y| = d+1$, p a prime and $c \in \mathbb{F}_p^E$. Let $X = \{0, 1, \ldots, d\}$. A *modular rubber band representation* of G (with respect to X, p and c) is defined as any assignment $V \to \mathbb{F}_p^d$ satisfying (4.6) such that $x_0 = 0$ and $x_i = e_i$ for $i \in X \setminus \{0\}$.

Lemma 4.3.11 *Let G, X and Y be as above, and let $N > 0$ be an integer. Choose uniformly a random a prime $p < N$ and a random vector $c \in \mathbb{F}_p^E$. Then*

(a) *With probability at least $1-n^2/N$, there is a unique modular rubber band representation of G with respect to X, p and c.*

(a) *With probability at least $1 - n^2/N$, a modular rubber band representation x_c satisfies* $\mathrm{rk}(f(Y)) = \kappa(X,Y)$.

Proof. The determinant of the system 4.6 is a polynomial of degree $\leq n^2$ in the strengths c_{ij}. A standard application of Schwartz's Lemma gives a bound on the probability that this determinant vanishes, which gives (a). The proof of (b) is similar. □

4.4 Repulsive springs and approximating maximum cut

A *cut* in a graph $G = (V,E)$ is the set of edges connecting a set $S \subseteq V$ to $V \setminus S$, where $\emptyset \subset S \subset V$. The *Max Cut Problem* is to find a cut with maximum cardinality. We denote by $\mathcal{C}_{\max}$ this maximum.

(More generally, we can be given a weighting $w : V \to \mathbb{R}_+$, and we could be looking for a cut with maximum total weight. To keep things simple, however, we restrict our introductory discussions to the unweighted case.)

The Max Cut Problem is NP-hard; one natural approach is to find an "approximately" maximum cut. Formulated in different terms, Erdős in 1967 described the following simple heuristic algorithm for the Max Cut Problem: for an arbitrary ordering $(v_1, \ldots, v_n)$ of the nodes, we color $v_1, v_2, \ldots, v_n$ successively red or blue. For each i, v_i is colored blue iff the number of edges connecting v_i to blue nodes among $v_1, \ldots, v_{i-1}$ is less than the number of edges connecting v_i to red nodes in this set. Then the cut formed by the edges between red and blue nodes contains at least half of all edges. In particular, we get a cut that is at least half as large as the maximum cut.

There is an even easier randomized algorithm to achieve this approximation, at least in expected value. Let us 2-color the nodes of G randomly, so that each node is colored red or blue independently, with probability 1/2. Then the probability that an edge belongs to the cut between red and blue is 1/2, and expected number of edges in this cut is $|E|/2$.

Both of these algorithms show that the maximum cut can be approximated from below in polynomial time with a multiplicative error of at most 1/2. Can we do better? The following

strong negative result of Hastad [88] (improving results of Arora et al. [17] and Bellare et al. [23]) shows that we cannot get arbitrarily close to the optimum:

Proposition 4.4.1 *It is NP-hard to find a cut with more than* $(16/17)\mathcal{C}_{\max} \approx .94\mathcal{C}_{\max}$ *edges.*

Building on results of Delorme, Poljak and Rendl [49, 161], Goemans and Williamson [79] give a polynomial time algorithm that approximates the maximum cut in a graph with a relative error of about 13%:

Theorem 4.4.2 *One can find in polynomial time a cut with at least* $.878\mathcal{C}_{\max}$ *edges.*

The algorithm of Goemans and Williamson makes use of the following geometric construction. We want to find an embedding $i \mapsto u_i$ $(i \in V)$ of the nodes of the graph in the unit sphere in $\mathbb{R}^d$ so that the following "energy" is maximized:

$$\mathcal{E}(u) = \sum_{ij \in E} \frac{1}{4}(u_i - u_j)^2 = \sum_{ij \in E} \frac{1 - u_i^{\mathsf{T}} u_j}{2}.$$

We can think of replacing the rubber bands by (strange) repulsive strings, which push their endpoints apart with a force that increases proportionally with the length.

If we work in $\mathbb{R}^1$, then the problem is equivalent to the Maximum Cut problem: each node is represented by either 1 or -1, and the edges between differently labeled nodes contribute 1 to the energy, the other edges contribute 0. Hence the maximum energy $\mathcal{E}_{\max}$ is an upper bound on the maximum size $\mathcal{C}_{\max}$ of a cut.

Unfortunately, the argument above also implies that for $d = 1$, the optimal embedding is NP-hard to find. While I am not aware of a proof of this, it is probably NP-hard for $d = 2$ and more generally, for any fixed d. The surprising fact is that for $d = n$, such an embedding can be found in polynomial time using semidefinite optimization (cf. Chapter 13.3).

Let X denote the $V \times V$ matrix defined by $X_{ij} = u_i^{\mathsf{T}} u_j$. The X satisfies the constraints:

$$X \succeq 0, \tag{4.14}$$

$$X_{ii} = 1, \tag{4.15}$$

and the energy $\mathcal{E}(u)$ can be expressed as

$$\sum_{ij \in E} \left(1 - \frac{1}{2} X_{ij}\right). \tag{4.16}$$

Conversely, if X is a $V \times V$ matrix satisfying (4.14) and (4.15), then we can write it as a Gram matrix of vectors in $\mathbb{R}^n$, these vectors will have unit length, and (4.16) gives the energy.

The semidefinite optimization problem of maximizing (4.16), subject to (4.14) and (4.15) can be solved in polynomial time (with an arbitrarily small relative error). So $\mathcal{E}_{\max}$ is a polynomial time computable upper bound on the size of the maximum cut.

How good is this bound? And how to construct an approximately optimum cut from this representation? Here is the simple but powerful trick: *take a random hyperplane H through the origin in* $\mathbb{R}^n$ (Figure 4.9). The partition of R^d given by H yields a cut in our graph. Since the construction pushes adjacent points apart, one expects that the random cut will intersect many edges.

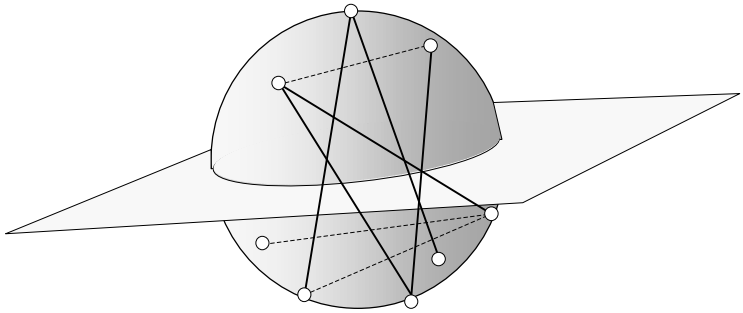

Figure 4.9: A cut in the graph given by a random hyperplane

To be more precise, let $ij \in E$ and let $u_i, u_j \in S^{n-1}$ be the corresponding vectors in the embedding constructed above. It is easy to see that the probability that a random hyperplane H through 0 separates u_i and u_j is α_{ij}/π, where $\alpha_{ij} = \arccos u_i^{\mathsf{T}} u_j$ is the angle between u_i and u_j. It is not difficult to verify that if $-1 \le t \le 1$, then $\arccos t \ge 1.38005(1-t)$. Thus the expected number of edges intersected by H is

$$\sum_{ij\in E} \frac{\arccos u_i^{\mathsf{T}} u_j}{\pi} \ge \sum_{ij\in E} 1.38005 \frac{1 - u_i^{\mathsf{T}} u_j}{\pi} = \frac{1.38005}{\pi} 2\mathcal{E}_{\max} \ge .878\mathcal{C}_{\max}.$$

(One objection to the above algorithm could be that it uses random numbers. In fact, the algorithm can be *derandomized* by well established but non-trivial techniques. We do not consider this issue in these notes; see e.g. [15], Chapter 15 for a survey of derandomization methods.)

Chapter 5

Rigidity

We can replace the edges of a graph by rigid bars, instead of rubber bands. This way we obtain a physical model, which is related to rubber bands, but is richer, and more important from the point of view of applications. We have to restrict ourselves to a few basic results of this rich theory; see e.g. Recski [164] or Whiteley [206] for more.

5.1 Stresses

Let $G = (V, E)$ be a graph and $x : V \to \mathbb{R}^d$, a geometric representation of G. A function $\sigma : E \to \mathbb{R}$ is called a *stress* if for every node i,

$$\sum_{j \in N(i)} \sigma_{ij}(x_j - x_i) = 0. \tag{5.1}$$

We can rewrite this as follows. Let M_x denote the $m \times dn$ matrix, in which the rows are indexed by edges $e \in E$, the columns are indexed by pairs it ($i \in V$, $1 \le t \le d$), and

$$(M_x)_{e,it} = \begin{cases} x_{it} - x_{jt} & \text{if } i \text{ is an endpoint of } e = ij, \\ 0 & \text{otherwise.} \end{cases} \tag{5.2}$$

Then σ is a stress means that $M_x^{\mathsf{T}}\sigma = 0$, i.e., σ is in the left nullspace of M_x. We denote this space of stresses by S_x.

In a rubber band representation, the strengths of the edges "almost" form a stress: (5.1) holds for all nodes that are not nailed. Lemma 4.3 said in this language the following:

Corollary 5.1.1 *A rubber band representation of a simple connected planar graph in the plane with a triangular face nailed has a stress that is* 1 *on the internal edges and negative on the boundary edges.*

An important result about stresses is a classic indeed: it was proved by Cauchy in 1822.

Theorem 5.1.2 (Cauchy's theorem) *The skeleton of a 3-polytope has no nonzero stress.*

Proof. Suppose that a convex polytope P carries a nonzero stress σ. Let G' be the subgraph of its skeleton formed by those edges which have a nonzero stress, together with the vertices they are incident with. Color an edge ij of G' red if $\sigma_{ij} > 0$, and blue otherwise. The graph G' is planar, and in fact it comes embedded in the surface of P, which is homeomorphic to the sphere. By lemma 1.1.8, G' has a node i that such that the red edges (and then also the blue edges) incident with v are consecutive in the given embedding of G'. This implies that we can find a plane through x_i which separates (in their embedding on the surface of P) the red and blue edges incident with i. Let e be the normal vector of this plane pointing in the halfspace which contains the red edges. Then for every red edge ij we have $e^{\mathsf{T}}(x_j - x_i) > 0$, and for every blue edge ij we have $e^{\mathsf{T}}(x_j - x_i) < 0$. By the definition of the coloring, this means that we have $\sigma_{ij}e^{\mathsf{T}}(x_j - x_i) > 0$ for every edge ij of G'. Also by definition, we have $\sigma_{ij}e^{\mathsf{T}}(x_j - x_i) = 0$ for those edges ij of G_P that are not edges of G. Thus

$$\sum_{j \in N(i)} \sigma_{ij} e^{\mathsf{T}}(x_j - x_i) > 0.$$

But

$$\sum_{j \in N(i)} \sigma_{ij} e^{\mathsf{T}}(x_j - x_i) = e^{\mathsf{T}}\Big(\sum_{j \in N(i)} \sigma_{ij}(x_j - x_i)\Big) = 0$$

by the definition of a stress, a contradiction. □

The most interesting consequence of Cauchy's Theorem is that is we make a convex polyhedron out of cardboard, then it will be rigid (see Section 5.2). This is not true if instead of convexity we only assume that the surface of the polytope is homeomorphic to the sphere, as shown by a tricky counterexample of Connelly [47].

Example 5.1.3 (Connelly's flexible polyhedron)

5.2 Rigidity

Let $G = (V, E)$ be a graph and $x : V \to \mathbb{R}^d$, a geometric representation of G. We think of the edges of G as rigid bars, and of the nodes, as flexible joints. We are interested in the rigidity of such a structure, or more generally, in its possible motions.

5.2.1 Infinitesimal motions

Suppose that the nodes of the graph move smoothly in d-space; let $x_i(t)$ denote the position of node i at time t. The fact that the bars are rigid says that for every edge $ij \in E(G)$,

$$\big(x_i(t) - x_j(t)\big) \cdot \big(x_i(t) - x_j(t)\big) = \text{const.}$$

Differentiating, we get

$$\big(x_i(t) - x_j(t)\big) \cdot \big(\dot{x}_i(t) - \dot{x}_j(t)\big) = 0.$$

This motivates the following definition. Given a geometric representation $x : V \to \mathbb{R}^d$ of G, a map $v : V \to \mathbb{R}^d$ is called an *infinitesimal motion* if the equation

$$(x_i - x_j) \cdot (v_i - v_j) = 0 \tag{5.3}$$

holds for every edge $ij \in E$). The vector v_i can be thought of as the *velocity* of node i.

There is another way of describing this. Suppose that we move each node i from x_i to $x_i + \varepsilon v_i$ for some vectors $v_i \mathbb{R}^d$. Then the squared length of an edge ij changes by

$$\Delta_{ij}(\varepsilon) = |(x_i + \varepsilon v_i) - (x_j + \varepsilon v_j)|^2 - |x_i - x_j|^2 = 2\varepsilon(x_i - x_j) \cdot (v_i - v_j) + \varepsilon^2 |v_i - v_j|^2.$$

Hence we can read off that if $(v_i : \ i \in V)$ is an infinitesimal motion, then $\Delta_{ij}(\varepsilon) = O(\varepsilon^2)$ $(\varepsilon \to 0)$; else, $\Delta_{ij}(\varepsilon) = \Theta(\varepsilon^2)$.

There are some trivial infinitesimal motions.

(a) If v_i is the same vector v for every $i \in V$, then (5.3) is trivially satisfied. This solution corresponds to the parallel translation of the structure with velocity v.

(b) In the plane, let R denote the rotation of the plane by $90°$ in the positive direction. Then $v_i = Rx_i$ $(i \in V)$ defines an infinitesimal motion; this motion corresponds to the rotation of the structure about the origin.

(c) In 3-space, $v_i = w \times x_i$ gives a solution for every fixed vector w, corresponding to rotation with axis w.

In general, every rigid motion of the whole space gives an infinitesimal motion. These trivial infinitesimal motions are characterized by the property that (5.3) holds for every pair of nodes $i \neq j$ (not just adjacent pairs).

The system (5.3) can be viewed as a system of linear equations for the coordinates of the velocity vectors v_i. In fact, it involves the same matrix M as defined by (5.2). If $v \in \mathbb{R}^{dn}$ denotes the vector obtained by concatenating the vectors v_i $(i \in V)$ (so ots coordinates are indexed by pairs i, j where $i \in V$ and $1 \leq j \leq d$), then we can write (5.3) as

$$Mv = 0. \tag{5.4}$$

The solutions of this equation form a linear space V_x, and the trivial solutions mentioned above form a subspace U_x of this of dimansion $\binom{d+1}{2}$ (if the vectors x_i are not contained in an affine hyperplane).

We say that the structure is *infinitesimally rigid*, if every infinitesimal motion is trivial. Whether or not a structure is infinitesimally rigid can be decided by simple linear algebra.

There are (at least) two other versions of the notion of rigidity. We can ask whether there is a finite motion in the sense one can specify a continuous orbit $x_i(t)$ for each node i so that

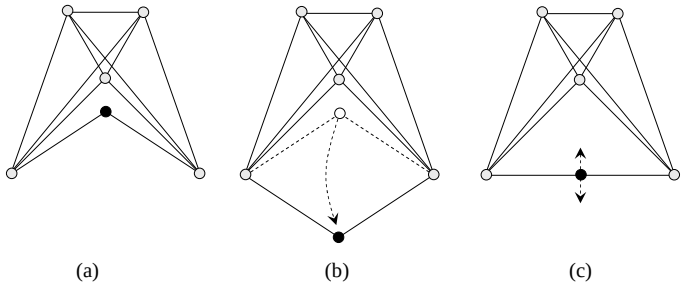

Figure 5.1: Structure (a) is infinitesimally rigid and has no finite motion in the plane, but has a different realization (b). Structure (c) has no finite motion and no other realization, but it is not infinitesimally rigid.

$x_i(0) = x_i$, $|x_i - x_j|$ is constant for every pair of adjacent nodes, but not for all pairs of nodes. Or we can ask whether there is a different realization of the same structure in the form of another representation $y : V \to \mathbb{R}^d$ such that $|x_i - x_j| = |y_i - y_j|$ for every pair of adjacent nodes, but not for all pairs of nodes. (In this case, it is possible that the representation y cannot be obtained from the representation x by a continuous motion preserving the lengths of the edges.) Figure 5.1 shows two examples of structures rigid in one sense but not in the other. It is in general difficult to decide whether the structure is rigid in these senses, and we'll only be concerned with infinitesimal rigidity.

Let us conclude this section with a little counting. The system of equations 5.3 has dn unknowns (the coordinates of the velocities) and m equations. We also know that it has a solution space of dimension at least $\binom{d+1}{2}$.

How many of these equations are linearly independent? Let λ_{ij} $(ij \in E)$ be multipliers such that combining the equations with them we get 0. This means that

$$\sum_{j \in N(i)} \lambda_{ij}(x_j - x_i) = 0$$

for every node i; in other words, λ is a stress!

This discussion implies that

$$\dim(V_x) - \dim(S_x) = dn - m. \tag{5.5}$$

An interesting special case is obtained when $m = dn - \binom{d+1}{2}$. The structure is infinitesimally rigid if and only $\dim(V_x) = \dim(U_x) = \binom{d+1}{2}$, which in this case is equivalent to $\dim(S_x) = 0$, which means that there is no nonzero stress.

Example 5.2.1 A representation of the 3-prism in the plane has an infinitesimal motion (equivalently a stress) if and only if it is drawn so that the lines of the three thick edges pass through one point (Figure 5.2). This is also the case when it has a nonzero stress.

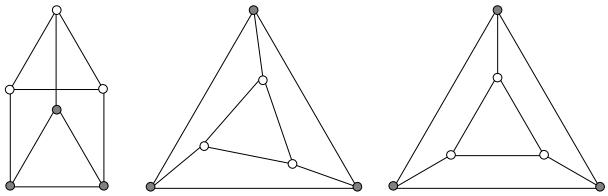

Figure 5.2: Rigid and nonrigid representations of the 3-prism in the plane. In the first, fixing the black nodes, the white triangle can "circle" the black triangle. The second is infinitesimally rigid and stress-free; the third has an infinitesimal motion (equivalently, it has a stress) but no finite motion.

5.2.2 Rigidity of convex 3-polytopes

For example, let G be the skeleton of a convex 3-polytope P, with the representation given by the positions of the vertices. Cauchy's Theorem 5.1.2 tells us that the space of stresses has dimension 0, so we get that $d_{\text{imot}} = 3n - m$. We know that $d_{\text{imot}} \geq \binom{4}{2} = 6$, and hence $m \leq 3n - 6$. Of course, this inequality we know already (Corollary 1.1.2). It also follows that if equality holds, i.e., when all faces of P are triangles, then the space of infinitesimal motions is 6-dimensional; in other words,

Corollary 5.2.2 *The structure consisting of the vertices and edges of a convex polytope with triangular faces is rigid.*

If not all faces of the polytope P are triangles, then it follows by the same computation that the skeleton is not infinitesimally rigid. However, if the faces are "made out of cardboard", which means that they are forced to preserve their shape, then the polytope will be rigid. To prove this, one can put a pyramid over each face (flat enough to preserve convexity), and apply Corollary 5.2.2 to this new polytope P' with triangular faces. Every nontrivial infinitesimal motion of P' would extend to a nontrivial infinitesimal motion of P, but we know already that P has no such motion.

5.2.3 Generic rigidity

We say that a representation $x : V \to \mathbb{R}^d$ of a graph is *generic*, if all the coordinates of the vectors x_i are algebraically independent transcendentals. This assumption is generally an overkill, and we could always replace it by "the coordinates of the representing vectors do not satisfy any algebraic relation that they don't have to, and which is important to avoid in the proof". Another way of handling the genericity assumption is to choose independent random values for the coordinates from any distribution that is absolutely continuous with respect

to the Lebesgue measure (e.g. Gaussian, or uniform from a bounded domain). We call this a *random representation* of G. We can then state the results as probabilistic statements holding with probability 1.

We say that a graph is *generically stress-free in* R^d, if every generic representation of it in R^d is stress-free; similarly, we say that G is *generically rigid in* R^d, if every generic representation of it in R^d is infinitesimally rigid.

Proposition 5.2.3 *Let G be a graph and $d \geq 1$. If G has a representation in $\mathbb{R}^d$ that is infinitesimally rigid [stress-free], then every generic representation of G in $\mathbb{R}^d$ is infinitesimally rigid [stress-free].*

Proof. Both infinitesimal rigidity and stress-freeness can be expressed in terms of the non-vanishing of certain determinants composed of the differences of the coordinates of the representation. If this holds for some choice of the coordinates, then it holds for the algebraically independent choice. □

Another way of stating this observation:

Proposition 5.2.4 *If G is generically rigid [stress-free] in R^d, then a random representation of G is infinitesimally rigid [stress-free] in R^d with probability* 1.

Generic rigidity and finite motion

In a generic position, there is no difference between infinitesimal non-rigidity and finite motion, as proved by Asimov and Roth [18]:

Theorem 5.2.5 *Suppose that a generic representation of a graph in $\mathbb{R}^d$ admits a nonzero infinitesimal motion. Then it admits a finite motion.*

Proof. Let x be a generic representation of G in $\mathbb{R}^d$, and let B be a basis for the columns of the matrix M_x. If c is any column not in B, then c can be expressed as a linear combination of the columns in B. From Cramer's Rule it follows that the coefficients in this linear combination are rational functions of the entries of M_x with integral coefficients. Multiplying by the common denominator, we get a vector $v = v(c, x)$ such that $M_x v = 0$, the entries of v are polynomials of the coordinates of the x_i with integer coefficients, and the nonzero entries are all contained in $B \cup \{c\}$.

It follows by elementary linear algebra that the vectors $v(c, x)$, where c ranges through all columns of M_x outside B, generate the space V_x, and since $V_x \neq U_x$ by hypothesis, there is a column c such that $v(c, x) \notin U_x$. Let us fix such a c.

Note that $M_x v(c, x) = 0$ is an algebraic equation in the entries $x_{i,t}$, and so by the assumption that these entries are algebraically independent, it follows that $M_y v(c, y) = 0$

holds identically for every representation $y : V \to \mathbb{R}^d$. In other words, $v(c, y) \in V_y$ for every y. If we choose y in a small neighborhood of x, then $v(c, y) \notin U_y$ must also hold.

Now consider the differential equations

$$\dot{y}(t) = v(c, y(t)), \qquad y(0) = x.$$

Since the left hand side is a continuous function of y, this system has a solution in a small interval $0 \le t \le T$. This defines a non-rigid motion of the graph. Indeed, for every edge ij,

$$\frac{d}{dt}|y_i(t)-y_j(t)|^2 = (y_i(t)-y_j(t))\cdot(\dot{y}_i(t)-\dot{y}_j(t)) = (y_i(t)-y_j(t))\cdot(v(c,y_i(t))-v(c,y_j(t))) = 0,$$

since $v(c, y(t)) \in V_y$. □

Generic rigidity in the plane

Laman [119] gave a characterization of generically stress-free graphs in the plane. To formulate the theorem, we need a couple of definitions. A graph $G = (V, E)$ is called a *Laman graph*, if if every subset $S \subseteq V$ with $|S| \ge 2$ spans at most $2|S| - 3$ edges. Note that the Laman condition implies that the graph has no multiple edges.

To motivate this notion, we note that every planar representation of a graph with n nodes and more than $2n - 3$ edges will admit a nonzero stress, by (5.5). If a graph has exactly $2n - 3$ edges, then a representation of it will be rigid if and only if it is stress-free.

The *Henneberg construction* increases the number of nodes of a graph G in one of two ways: (H1) create a new node and connect it to at most two old nodes, (H2) subdivide an edge and connect the new node to a third node. If a graph can be obtained by iterated Henneberg construction, starting with a single edge, then we call it a *Henneberg graph*.

Theorem 5.2.6 (Laman's Theorem) *For a graph $G = (V, E)$ the following are equivalent:*

(a) *G is generically stress-free in the plane;*

(b) *G is a Laman graph;*

(c) *G is a Henneberg graph.*

Proof. (a)⇒(b): Suppose that V has a subset S with $|S| \ge 2$ spanning more than $2|S| - 3$ edges. Then the equations corresponding to these edges are dependent, so there is a stress on the graph.

(b)⇒(c): We prove by induction on the number of nodes that a Laman graph G can be built up by the Henneberg construction. Since the average degree G is $2|E|/|V| \le (4|V| - 6)/|V| < 4$, there is a node i of degree at most 3. If i has degree 2, then we can delete i and proceed by induction and (H1). So we may assume that i has degree 3. Let a, b, c be its neighbors.

Call a set $S \subseteq V$ with $|S| \geq 2$ spanning exactly $2|S| - 3$ edges of G a *tight set*. The key observation is:

Claim 5.2.7 *If two tight sets have more than one node in common, then their union is also tight.*

Indeed, suppose that S_1 and S_2 are tight, and let E_1 and E_2 be the sets of edges they span. Then $E_1 \cap E_2$ is the set of edges spanned by $S_1 \cap S_2$, and so $|E_1 \cap E_2| \geq 2|S_1 \cap S_2| - 3$ (using the assumption that $|S_1 \cap S_2| \geq 2$). The set $E_1 \cup E_2$ is a subset of the set of edges spanned by $S_1 \cup S_2$, and so the number of edges spanned by $S_1 \cup S_2$ is at least

$$|E_1 \cup E_2| = |E_1| + |E_2| - |E_1 \cap E_2| \geq (2|S_1| - 3) + (2|S_2| - 3) - (2|S_1 \cap S_2| - 3) = 2|S_1 \cup S_2| - 3.$$

Since G is a Laman graph, we must have equality here, implying that $S_1 \cup S_2$ is tight.

Claim 5.2.8 *There is no tight set containing $\{a, b, c\}$ but not i.*

Indeed, adding i to such a set the Laman property would be violated.

We want to prove that we can delete i and create a new edge between two of its neighbors to get a Laman graph G'; then G arises from G' by Henneberg construction (H2). Let us try to add the edge ab to $G - i$; if the resulting graph G' is not a Laman graph, then there is a set $S \subseteq V \setminus \{i\}$ with $|S| \geq 2$ spanning more than $2|S| - 3$ edges of G'. Since S spans at most $2|S| - 3$ edges of G, this implies that S is tight, and $a, b \in S$. (If a and b are adjacent in G, then $\{a, b\}$ is such a tight set.) If there are several sets S with this property, then Claim 5.2.7 implies that their union is also tight. We denote by S_{ab} this union. Similarly we get the tight sets S_{bc} and S_{ca}.

Claim 5.2.8 implies that the sets S_{ab}, S_{bc} and S_{ca} are distinct, and hence Claim 5.2.7 implies that any two of them have one node in common. Then the set $S = S_{ab} \cup S_{bc} \cup S_{ca} \cup \{i\}$ spans at least

$$(2|S_{ab}| - 3) + (2|S_{bc}| - 3) + (2|S_{ca}| - 3) + 3 = 2|S| - 2$$

edges, a contradiction.

(c)⇒(a): Let G arise from G' by the Henneberg construction, and let σ be a nonzero stress on G. We want to construct a nonzero stress on G' (which would complete the proof by induction).

First suppose that we get G by step (H1), and let i be the new node. It suffices to argue that σ is 0 on the new edges, so that σ gives a stress on G'. If i has degree 0 or 1, then this is trivial. Suppose that i has two neighbors a and b. Then by the assumption that the representation is generic, we know that the points x_i, $x_a$and x_b are not collinear, so from the stress condition

$$\sigma_{ia}(x_i - x_a) + \sigma_{ib}(x_i - x_b) = 0$$

it follows that $\sigma_{ia} = \sigma_{ib} = 0$.

Suppose that G arises by Henneberg construction (H2), by subdividing the edge $ab \in E(G')$ and connecting the new node i to $c \in V(G') \setminus \{a, b\}$. Let us modify the representation x as follows:

$$x'_j = \begin{cases} \frac{1}{2}(x_a + x_b) & \text{if } j = i, \\ x_j & \text{otherwise.} \end{cases}$$

This representation for G is not generic any more, but it is generic if we restrict it to G'.

If the generic representation x of G admits a nonzero stress, then by Proposition 5.2.3 so does every other representation, in particular x' admits a nonzero stress σ'. Consider the stress condition for i:

$$\sigma'_{ia}(x'_i - x'_a) + \sigma'_{ib}(x'_i - x'_b) + \sigma'_{ic}(x'_i - x'_c) = 0.$$

Here $x'_i - x'_a = \frac{1}{2}(x_b - x_a)$ and $x'_i - x'_b = \frac{1}{2}(x_a - x_b)$ are parallel but $x'_i - x'_c$ is not parallel to them. So it follows that $\sigma'_{ic} = 0$ and $\sigma'_{ia} = \sigma'_{ib}$. Defining $\sigma'_{ab} = \sigma'_{ia}$, we get a nonzero stress on G', a contradiction. □

Corollary 5.2.9 *Let G be a graph with n nodes and $2n - 3$ edges. Then G is generically rigid in the plane if and only if it is a Laman graph.*

Laman's Theorem can be used to prove the following characterization of generically rigid graphs in the plane.

Theorem 5.2.10 *A graph $G = (V, E)$ is generically rigid in the plane if and only if for every decomposition $G = G_1 \cup \cdots \cup G_k$ we have $\sum_i (2|V(G_i)| - 3) \geq 2|V| - 3$.*

Proof.

□

It is worth while to state the following consequence of Theorem 5.2.10:

Corollary 5.2.11 *Every 6-connected graph is generically rigid in the plane.*

Proof. Suppose that $G = (V, E)$ is 6-connected but not generically rigid, and consider such a graph with a minimum number of nodes. Then by Theorem 5.2.10 there exist subgraphs $G_i = (V_i, E_i)$ $(i = 1, \ldots, k)$ such that $G = G_1 \cup \cdots \cup G_k$ and

$$\sum_{i=1}^{k} (2|V_i| - 3) < 2|V| - 3. \tag{5.6}$$

We may assume that every G_i is complete, since adding edges inside a set V_i does not change (5.6). Furthermore, we may assume that every node belong to at least two of the V_i: indeed,

if $v \in V_i$ is not contained in any other V_j, then deleting it preserves (5.6) and also preserves 6-connectivity (deleting a node from a graph whose neighbors form a complete graph does not decrease the connectivity of this graph).

We claim that

$$\sum_{V_i \ni v} \Big(2 - \frac{3}{|V_i|}\Big) \geq 2 \tag{5.7}$$

for each node v. Let (say) $V_1, \ldots, V_r$ be those V_i containing v, $|V_1| \leq |V_2| \leq \ldots$ Each term on the left hand side is at least $1/2$, so if $r \geq 4$, we are done. Furthermore, the graph is 6-connected, hence v has degree at least 6, and so

$$\sum_{i=1}^{r} (|V_i| - 1) \geq 6. \tag{5.8}$$

If $r = 3$, then $|V_3| \geq 3$, and so the left hand side of (5.7) is at least $1/2 + 1/2 + 1 = 2$. If $r = 2$ and $|V_1| \geq 3$, then the left hand side of (5.7) is at least $1 + 1 = 2$. Finally, if $r = 2$ and $|V_1| = 2$, then $|V_2| \geq 6$ by (5.8), and so the left hand side of (5.7) is at least $1/2 + 3/2 = 2$.

Now summing (5.7) over all nodes v, we get

$$\sum_{v \in V} \sum_{V_i \ni v} \Big(2 - \frac{3}{|V_i|}\Big) = \sum_{i=1}^{k} \Big(2 - \frac{3}{|V_i|}\Big) \sum_{v \in V_i} 1 = \sum_{i=1}^{k} (2|V_i| - 3)$$

on the left hand side and $2|V|$ on the right hand side, which contradicts (5.6). □

Exercise 5.1 Construct a 5-connected graph that is not generically rigid in the plane.

A 3-dimensional analogue of Laman's Theorem, or Theorem 5.2.10 is not known. It is not even known whether there is a positive integer k such that every k-connected graph is generically rigid in 3-space.

Chapter 6

Representing graphs by touching domains

6.1 Coin representation

6.1.1 Koebe's theorem

We prove Koebe's important theorem on representing a planar graph by touching circles [115], and its extension to a Steinitz representation, the Cage Theorem.

Theorem 6.1.1 (Koebe's Theorem) *Let G be a 3-connected planar graph. Then one can assign to each node i a circle C_i in the plane so that their interiors are disjoint, and two nodes are adjacent if and only if the corresponding circles are tangent.*

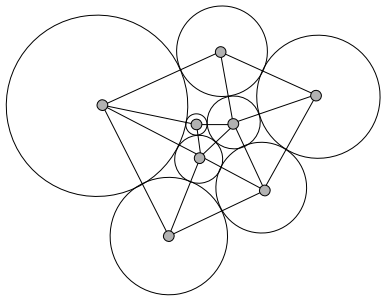

Figure 6.1: The coin representation of a planar graph

If we represent each of these circles by their center, and connect two of these centers by a segment if the corresponding circles touch, we get a planar map, which we call the *tangency graph* of the family of circles. Koebe's Theorem says that every planar graph is the tangency graph of a family of openly disjoint circular discs.

The following theorem, due to Andre'ev [16], gives a strengthening of Koebe's theorem in terms of a simultaneous representation of a 3-connected planar graph and of its dual by touching circles. To be precise, we define a *double circle representation in the plane* of a planar map G as two families of circles, $(C_i : i \in V)$ and $D_j : j \in V^*)$ in the plane, so that for every edge ij, bordering faces a and b, the following holds: the circles C_i and C_j are tangent at a point x; the circles D_a and D_b are tangent at the same point x; and the circles D_a and D_b intersect the circles C_i and C_b at this point. Furthermore, the circles C_i are disjoint and so are the circles D_j, except that the circle D_{j_0} representing the outer face contains all the other circles in its interior.

Theorem 6.1.2 *Every 3-connected planar map G has a double circle representation in the plane.*

6.1.2 Formulation in the space

Theorem 6.1.3 (The Cage Theorem) *Every 3-connected planar graph is isomorphic to the 1-skeleton of a convex 3-polytope such that every edge of the polytope touches a given sphere.*

It is worth while to describe this in another way. Let $G = (V, E)$ be a 3-connected planar graph and $G^* = V^*, E^*)$ its dual. There is a bijection $\eta : E \to E^*$ from duality. It will be convenient to define η so that if pq is oriented, then $ij = \eta(pq)$ is also oriented, and so that pq crosses ij from left to right.

We define the *diamond graph* $G^\diamondsuit = (V^\diamondsuit, E^\diamondsuit)$ of G follows. We let $V^\diamondsuit = V \cup V^*$, and we connect $p \in V$ to $i \in V^*$ if p is a node of face i. We do not connect two nodes in V nor in V^*, so $G^\diamondsuit$ is a bipartite graph. It is also clear that $G^\diamondsuit$ is planar. The double cycle representation of G gives a representation of $G^\diamondsuit$ by circles such that adjacent nodes correspond to orthogonal circles.

Figure 6.3(a) shows this double circle representation of the simplest 3-connected planar graph, namely K_4. For one of the circles, the exterior domain should be considered as the disk it bounds. The other picture shows part of the double cycle representation of a larger planar graph.

6.1.3 Preparation for the proof

We fix a triangular face i_0 in G or G^* (say, G) as the outer face; let a, b, c be the nodes of i_0. For every node $p \in V$, let $F(p)$ denote the set of bounded faces containing p, and for every face i, let $V(i)$ denote the set of nodes contained in i. Let $U = V \cup V^* \setminus \{i_0\}$, and let J denote the set of pairs ip with $i \in V^* \setminus \{i_0\}$ and $p \in V(i)$.

Let us start with assigning a positive real number r_u to every node $u \in U$. Think of this as a guess for the radii of the circles we seek (we don't guess the radius for the circle

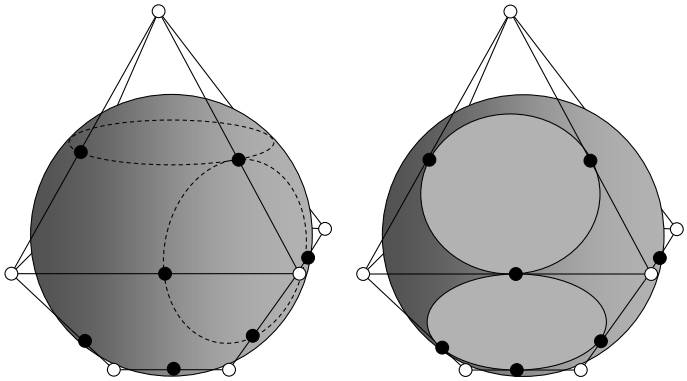

Figure 6.2: A convex polytope in which every edge is tangent to a sphere creates two families of circles on the sphere.

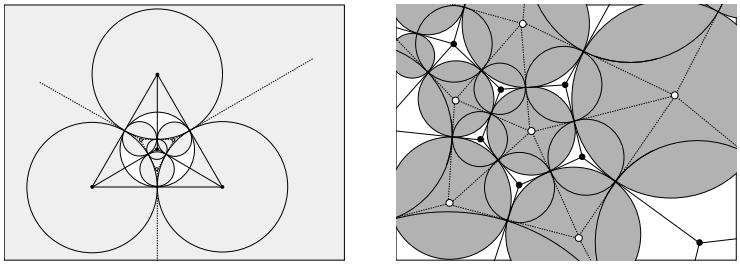

Figure 6.3: Two sets of circles, representing (a) K_4 and its dual (which is another K_4); (b) a planar graph and its dual.

representing i_0; this will be easy to add at the end). For every $ip \in J$, we define

$$\alpha_{ip} = \arctan \frac{r_p}{r_i}.$$

and

$$\alpha_{pi} = \arctan \frac{r_i}{r_p} = \frac{\pi}{2} - \alpha_{ip}.$$

Suppose that p is an internal node. If the radii correspond to a correct double circle representation, then $2\alpha_{pi}$ is the angle between the two edges of the face i at p (Figure 6.4). Thus in this case we have

$$\sum_{i \in F(p)} 2\alpha_{pi} = 2\pi.$$

Of course, we cannot expect this to hold for an arbitrary choice of the r_u; let as call the value

$$\delta_p = \sum_{i \in F(p)} \alpha_{pi} - \pi$$

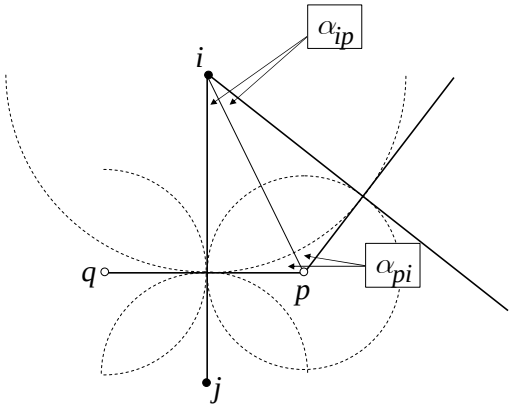

Figure 6.4: Notation.

the *defect* of the node p (for the given choice of radii; we divided by 2 for convenience).

For a boundary node $p \in V(i_0)$, we have to modify this definition: In the ideal situation, we would have

$$\sum_{i \in F(p)} 2\alpha_{pi} = \frac{\pi}{3},$$

and so we define the defect by

$$\delta_p = \sum_{i \in F(p)} \alpha_{pi} - \frac{\pi}{6}.$$

If i is a bounded face, then for the right choice of radii we should have

$$\sum_{p \in V(i)} 2\alpha_{ip} = 2\pi.$$

Accordingly, we define the defect of bounded faces:

$$\delta_i = \sum_{i \in V(p)} \alpha_{ip} - \pi.$$

(Note that all these defects may be positive or negative.) This way we define a mapping $\Phi : \mathbb{R}^U \to \mathbb{R}^U$ by

$$\Phi : (r_u : u \in U) \mapsto (\delta_u : u \in U). \tag{6.1}$$

The crucial observation is the following.

Lemma 6.1.4 *If radii $r_u > 0$ $(u \in U)$ are chosen so that we have $\delta_u = 0$ for all $u \in U$, then there is a double circuit representation with these radii.*

Proof. Let us construct two right triangles with sides r_i and r_p for every $ip \in J$, one with each orientation, and glue these two triangles together along their hypotenuse to get a *kite* K_{ip}. Now starting from a node p_1 of i_0, we put down all kites $K_{p_1 i}$ in the order of the corresponding faces in a planar embedding of G. By $\delta_{p_1} = 0$, these will fill an angle of $\pi/3$ at p_1. No proceed to a bounded face i_1 containing p_1, and put down all the remaining kites $K_{i_1 p}$ in the order in which the nodes of i_1 follow each other on the boundary of i_1. By $\delta_{i_1} = 0$, these triangles will cover a neighborhood of p_1. We proceed similarly to the other bounded faces containing p_1, then to the other nodes of i_1, etc. The condition that $\delta = 0$ will guarantee that we tile a regular triangle.

Let C_p be the circle with radius r_p about the position of node p constructed above, and define D_i similarly. We still need to define D_{i_0}. It is clear that i_0 is drawn as a regular triangle, and hence we necessarily have $r_a = r_b = r_c$. We define D_{p_0} as the inscribed circle of the regular triangle *abc*. □

6.1.4 The range of defects

The defect vector $\delta(r)$ depends on the ratio of the r_u only; hence we may restrict our attention to positive radii satisfying

$$\sum_{u \in U} r_u = 1.$$

Then the domain of the map δ is the interior of the $(n+f-2)$-dimensional simplex Σ defined by

$$x_u \geq 0, \qquad \sum_{u \in U} x_u = 1.$$

Our next goal is to determine the range. By Lemma 6.1.4, all we need to show is that this contains the origin.

While of course an arbitrary choice of the radii r_u will not guarantee that $\delta_u = 0$ for all $u \in U$, the following lemma shows that this is true at least "one the average":

Lemma 6.1.5 *For every assignment of radii, we have*

$$\sum_{u \in U} \delta_u = 0.$$

Proof. From the definition,

$$\sum_{u \in U} \delta_u = \sum_{i \in V \setminus \{a,b,c\}} \left(\sum_{p \in F(i)} \alpha_{ip} - \pi \right) + \sum_{i \in \{a,b,c\}} \left(\sum_{p \in F(i)} \alpha_{ip} - \frac{\pi}{6} \right) + \sum_{p \in V^* \setminus \{p_0\}} \left(\sum_{i \in V(p)} \alpha_{pi} - \pi \right).$$

Every pair $ip \in J$ contributes $\alpha_{ip} + \alpha_{pi} = \pi/2$. Since $|J| = 2m - 3$, we get

$$(2m-3)\frac{\pi}{2} - (n-3)\pi - 3\frac{\pi}{6} - (f-1)\pi = (m-n-f+2)\pi$$

By Euler's formula, this proves the lemma. □

Let us examine the range of the mapping δ in more detail. We already know that it lies in the hyperplane defined by $\sum_u \delta_u = 0$. We can derive some inequalities too.

For $S \subseteq U$, define $J[S] = \{ip \in J : \ i, p \in S\}$ and

$$f(S) = \frac{1}{2}|J[S]| - |S| + \frac{5}{6}|S \cap \{a, b, c\}|.$$

Lemma 6.1.6 (a) $f(\emptyset) = f(U) = 0$;

(b) $f(S) < 0$ *for every set* $\emptyset \subset S \subset U$.

Proof. The proof of (a) is left to the reader as Exercise 6.1.

We could prove (b) by combinatorial arguments (e.g., induction), but let us give a geometric proof whose elements we can use later again.

Consider any straight line embedding of the graph (say, the Tutte rubber band embedding), with i_0 forming a regular triangle and bounding the unbounded face. For $ip \in J$ let β_{pi} denote the angle of the polygon i at the vertex p, and let $\beta_{ip} = \pi - \beta_{pi}$. These numbers obviously have the following properties:

$$0 < \beta_{ip} < \pi, \qquad 0 < \beta_{pi} < \pi. \tag{6.2}$$

Furthermore,

$$\sum_{i \in F(p)} \beta_{pi} = 2\pi \tag{6.3}$$

for every node $p \neq a, b, c$,

$$\sum_{i \in F(p)} \beta_{pi} = \frac{\pi}{3} \tag{6.4}$$

for $i \in \{a, b, c\}$, and

$$\sum_{p \in V(i)} \beta_{ip} = 2\pi \tag{6.5}$$

for every bounded face p. We can do the following computation using (6.3), (6.4) and (6.5):

$$\begin{aligned}
\pi|J(S)| &= \sum_{\substack{ip \in J \\ i,p \in S}} (\beta_{pi} + \beta_{ip}) \\
&< \sum_{i \in S \cap V^*} \sum_{p \in V(i)} \beta_{ip} + \sum_{p \in S \cap V} \sum_{i \in F(p)} \beta_{pi} \\
&= 2\pi|S| - \frac{5}{3}\pi|S \cap \{a, b, c\}|.
\end{aligned}$$

(The strict inequality comes from the fact that there is at least one pair ip where exactly one element belongs to S, and here we omitted a positive term β_{ip} or β_{pi}.) This proves the Lemma. □

Lemma 6.1.7 *For every set $\emptyset \subset S \subset U$,*

$$\sum_{u\in S} \delta_u > f(S)\pi.$$

Proof. We have

$$\begin{aligned}\sum_{u\in S} \delta_u &= \sum_{p\in S\cap V} \sum_{i\in F(i)} \alpha_{pi} + \sum_{i\in S\cap V^*} \sum_{p\in V(i)} \alpha_{ip} \\ &\quad - |I(S)|\pi - |S\cap(V\setminus\{a,b,c\}|\pi - |S\cap\{a,b,c\}|\frac{\pi}{6} \\ &= \sum_{\substack{ip\in J\\ i\in S, p\notin S}} \alpha_{ip} + \sum_{\substack{ip\in J\\ i\notin S, p\in S}} \alpha_{pi} + \sum_{ip\in J[S]} (\alpha_{ip}+\alpha_{pi}) \qquad (6.6)\\ &\quad - |I(S)|\pi - |S\cap(V\setminus\{a,b,c\}|\pi - |S\cap\{a,b,c\}|\frac{\pi}{6} \\ &= \sum_{\substack{ip\in J\\ i\in S, p\notin S}} \alpha_{ip} + \sum_{\substack{ip\in J\\ i\notin S, p\in S}} \alpha_{pi} + f(S)\pi.\end{aligned}$$

Since the first two terms are nonnegative, and (as in the previous proof) at least one is positive, the lemma follows. □

Now we are prepared to describe the range of defects. Let P denote the polyhedron in $\mathbb{R}^U$ defined by

$$\sum_{u\in U} x_u = 0, \qquad (6.7)$$

$$\sum_{u\in S} x_u \geq d(S)\pi \quad (\emptyset \subset S \subset U). \qquad (6.8)$$

Clearly P is bounded. Our lemmas imply the following properties of P and the map Φ defined in (6.1):

Corollary 6.1.8 *The origin is in the relative interior of P. The range of Φ is contained in P.*

The first statement follows by Lemma 6.1.7. For a vector in the range of Φ, equation (6.7) follows by Lemma 6.1.5, while inequality (6.8) follows by Lemma 6.1.6.

Lemma 6.1.9 *The map Φ is injective.*

Proof. Consider two different choices r and r' of radii. Let S be the set of elements of U for which r'_u/r_u is maximum, then S is a nonempty proper subset of U. There is a pair $ip \in J$ ($i \in V^*$, $p \in V$) such that exactly one of i and p is in S; say, for example, that $i \in S$ and $p \notin S$. Then for every pair iq, $q \in V(i)$, we have $r'_q/r'_i \leq r_q/r_i$, and strict inequality holds if $q = p$. Thus

$$\sum_{q \in V(i)} \arctan \frac{r'_q}{r'_i} < \sum_{q \in V(i)} \arctan \frac{r_q}{r_i},$$

showing that $\delta_i(r') < \delta_i(r)$, which proves the lemma. □

The next (technical) lemma is needed whenever we construct some assignment of radii as a limit of other such assignments.

Lemma 6.1.10 *Let $r^{(1)}, r^{(2)}, \cdots \in \mathbb{R}^U$ be a sequence of assignments of radii, and let $\delta^{(k)} = \Phi(r^{(k)})$. Suppose that for each $u \in U$, $r_u^{(k)} \to \rho_u$ as $k \to \infty$. Let $S = \{u : \rho_u > 0\}$. Then*

$$\sum_{u \in S} \delta_u^{(k)} \to d(S)\pi \qquad (k \to \infty).$$

(We note that the right hand side is negative by Lemma 6.1.6.)

Proof. Recall the computation (6.6). The terms that result in strict inequality are α_{pi} with $p \in S$, $i \notin S$, and α_{ip} with $i \in S$, $p \notin S$. These terms tend to 0, so the slack in (6.7) tends to 0. □

Now we are able to prove the main result in this section.

Theorem 6.1.11 *The range of Φ is exactly the interior of P.*

Proof. Suppose not, and let y_1 be an interior point in P not in the range of Δ. Let y_2 be any point in the range. The segment connecting y_1 and y_2 contains a point y which is an interior point of P and on the boundary of the range of Δ. Consider a sequence $(r^1, r^2, \dots)$ with $\Delta(r^k) \to y$. We may assume (by compactness) that (r^k) is convergent. If (r^k) tends to an interior point r of Σ, then the fact that Δ is continuous, injective, and $\dim(P) = \dim(\Sigma)$ imply that the image of a neighborhood of r covers a neighborhood of y, a contradiction. If (r^k) tends to a boundary point of Σ, then by Lemma 6.1.10, $\delta(r_k)$ tends to the boundary of Σ, a contradiction. □

Since the origin is in the interior of P by Corollary 6.1.8, this also completes the proof of Theorem 6.1.2.

The proof we saw is just an existence proof, however. In the next section we describe an algorithmic proof. We'll need most of the lemmas proved above in the analysis of the algorithm.

Exercise 6.1 Prove part (a) of Lemma 6.1.6

6.1.5 An algorithmic proof

We measure the "badness" of the assignment of the radii by error

$$\mathcal{E} = \sum_{u \in U} \delta_u^2.$$

We want to modify the radii so that we reduce the error. The key observation is the following. Let i be a face with $\delta_i > 0$. Suppose that we increase the radius r_i, while keep the other radii fixed. Then α_{ip} decreases for every $p \in V(i)$, and correspondingly α_{pi} increases, but nothing else changes. Hence δ_i decreases, δ_p increases for every $p \in V(i)$, and all the other defects remain unchanged. Since the total defect remains the same by Lemma 6.1.5, we can describe this as follows: if we increase a radius, some of the defect of that node is distributed to its neighbors (in the $G^\diamondsuit$ graph). Note, however, that it is more difficult to describe in what proportion this defect is distributed (and we'll try to avoid to have to describe this).

How much of the defect of i can be distributed this way? If $r_i \to \infty$, then $\alpha_{ip} \to 0$ for every $p \in V(i)$, and so $\delta_i \to -\pi$. This means that we "overshoot". So we can distribute at least the positive part of the defect this way.

The same argument applies to the defects of nodes, and we can distribute negative defects similarly by decreasing the appropriate radius. Let us immediately renormalize the radii, so that we maintain that $\sum_u r_u = 1$. Recall that this does not change the defects.

There are many schemes that can be based on this observation: we can try to distribute the largest (positive) defect, or a positive defect adjacent to a node with negative defect etc. Brightwell and Scheinerman [26] prove that if we repeatedly pick any element $u \in U$ with positive defect and distribute all its defect, then the process will converge to an assignment of radii with no defect. There is a technical hurdle to overcome: since the process is infinite, one must argue that no radius tends to 0 or ∞. We give a somewhat different argument.

Consider a subset $\emptyset \subset S \subset U$, and multiply each radius r_u, $u \notin S$, by the same factor $0 < \lambda < 1$. Then α_{ip} is unchanged if both i and p are in S or outside S; if $i \in S$ and $p \notin S$ then α_{ip} decreases while α_{pi} increases; and similarly the other way around. Hence δ_u does not increase if $u \in S$ and strictly decreases if $u \in S$ has a neighbor outside S. Similarly, δ_u does not decrease if $u \in U \setminus S$ and strictly increases if $u \in U \setminus S$ has a neighbor in S.

Let $\emptyset \subset S \subset U$ be a set such that $\min_{u \in S} \delta_u > \max_{U \in U \setminus S} \delta_u$. We claim that we can decrease the radii in $U \setminus S$ until one of the δ_u, $u \in S$, becomes equal to a δ_v, $v \notin S$. If not, then (after renormalization) the radii in $U \setminus S$ would tend to 0 while still any defect in S would be larger than any defect in $U \setminus S$. But it follows by Lemmas 6.1.6 and 6.1.10) that in this case the total defect in S tends to a negative value, and so the total defect in $U \setminus S$ tends to a positive value. So there is an element of S with negative defect and an element of $U \setminus S$ with positive defect, which is a contradiction.

Let t be this common value, and let δ'_u be the new defects. Then the change in the error

is

$$\mathcal{E} - \mathcal{E}' = \sum_{u \in U} \delta_u^2 - \sum_{u \in U} {\delta'_u}^2$$

Using Lemma 6.1.5, we can write this in the form

$$\mathcal{E} - \mathcal{E}' = \sum_{u \in U} (\delta_u - \delta'_u)^2 + 2 \sum_{u \in U} (t - \delta'_u)(\delta'_u - \delta_u).$$

By the choice of t, we have $t \le \delta'_u$ and $\delta_u \ge \delta'_u$ for $u \in S$ and $t \ge \delta'_u$ and $\delta_u \le \delta'_u$ for $u \notin S$. Hence the second sum in $\mathcal{E} - \mathcal{E}'$ is nonnegative, while the first is positive. So the error decreases; in fact, it decreases by at least $(\delta_u - t)^2 + (\delta_v - t)^2 \ge (\delta_u - \delta_v)^2/4$ for some $u \in S$ and $v \in U \setminus S$. If we choose the largest gap in the sequence of the δ_u ordered decreasingly, then this gain is at least

$$\left(\frac{1}{m}(\max \delta_u - \min \delta_u)\right)^2 \ge \left(\frac{1}{m}\sqrt{\frac{\mathcal{E}}{m}}\right)^2 = \frac{\mathcal{E}}{m^3}.$$

Thus we have

$$\mathcal{E}(r') \le \left(1 - \frac{1}{m^3}\right) \mathcal{E}(r). \tag{6.9}$$

If we iterate this procedure, we get a sequence of vectors $r^1, r^2, \cdots \in \Sigma$ for which $\mathcal{E}(r^k) \to 0$. No subsequence of the r^k can tend to a boundary point of Σ. Indeed, by Lemma 6.1.10, for such a sequence $\delta(r^{(k)})$ would tend to the boundary of P, and so by Corollary 6.1.8, the error would stay bounded away from 0. Similarly, if a subsequence tends to an interior point $r \in \Sigma$, then $\delta(r) = 0$, and by Claim 6.1.9, this limit is unique. It follows that there is a (unique) point r in the interior of Σ with $\delta(r) = 0$, and the sequence r^k tends to this point.

6.1.6 *Another algorithmic proof

We present a second proof of Theorem 6.1.2, due to Colin de Verdière, which is shorter, but uses a less transparent (more tricky) optimization argument. One advantage is that we can use an "off the shelf" optimization algorithm for smooth convex functions to compute the representation.

Define

$$\phi(x) = 2 \int_{-\infty}^{x} \arctan(e^t)\, dt.$$

It is easy to verify that ϕ is monotone increasing, convex, and

$$\phi(x) = \max\{0, \pi x\} + O(1). \tag{6.10}$$

Let $x \in \mathbb{R}^V$, $y \in \mathbb{R}^{\mathcal{F}}$. Using the numbers β_{ip} introduced above in the proof of Lemma 6.1.6, consider the function

$$F(x,y) = \sum_{i,p:\, p\in F(i)} \Big(\phi(y_p - x_i) - \beta_{ip}(y_p - x_i)\Big).$$

Claim. If $|x| + |y| \to \infty$ while (say) $x_1 = 0$, then $F(x,y) \to \infty$.

We need to fix one of the x_i, since if we add the came value to each x_i and y_p, then the value of F does not change.

To prove the claim, we use (6.10):

$$\begin{aligned} F(x,y) &= \sum_{i,p:\, p\in F(i)} \Big(\phi(y_p - x_i) - \beta_{ip}(y_p - x_i)\Big) \\ &= \sum_{i,p:\, p\in F(i)} \Big(\max\{0, \pi(y_p - x_i)\} - \beta_{ip}(y_p - x_i)\Big) + O(1) \\ &= \sum_{i,p:\, p\in F(i)} \Big(\max\{-\beta_{ip}(y_p - x_i), (\pi - \beta_{ip})(y_p - x_i)\}\Big) + O(1). \end{aligned}$$

Since $-\beta_{ip}$ is negative but $\pi - \beta_{ip}$ is positive, each term here is nonnegative, and a given term tends to infinity if $|x_i - y_p| \to \infty$. If x_1 remains 0 but $|x| + |y| \to infty$, then at least one difference $|x_i - y_p|$ must tend to infinity. This proves the Claim.

It follows from this Claim that F has a minimum at some point (x,y). Let i be an internal node, then

$$\frac{\partial}{\partial x_i} F(x,y) = -\sum_{p\in F(i)} \phi'(y_p - x_i) + \sum_{p\in F(i)} \beta_{ip} = -2\sum_{p\in F(i)} \arctan(e^{y_p - x_i}) + 2\pi$$

(using (6.3)), and so

$$2\sum_{p\in F(i)} \arctan(e^{y_p - x_i}) = 2\pi. \tag{6.11}$$

It follows by a similar computation that

$$2\sum_{p\in F(i)} \arctan(e^{y_p - x_i}) = \frac{\pi}{3} \tag{6.12}$$

for the three boundary nodes i, and

$$2\sum_{i\in V(p)} \arctan(e^{y_p - x_i}) = (d_p - 2)\pi$$

for every bounded face p. We can rewrite this condition as

$$2\sum_{i\in V(p)} \arctan(e^{x_i - y_p}) = 2\pi, \tag{6.13}$$

since $\arctan(e^{x_i - y_p}) + \arctan(e^{y_p - x_i}) = \pi/2$. This completes the proof.

6.1.7 *From rubber bands to touching circles

It is not hard to prove (Exercise 6.2) that every double circuit representation of a planar graph is in fact a rubber band representation. Can we go the other way?

It is not true that every rubber band representation gives rise to a representation by touching circles. Consider a graph G that is a triangulation of the plane with outer triangle abc. When constructing a rubber band representation, we have the freedom of choosing the rubber band strengths arbitrarily. Not all of these lead to different drawings; we have $2(n-3)$ equations

$$\sum_{j\in N(i)} c_{ij}(x_i - x_j) = 0 \qquad (i \neq a, b, c), \tag{6.14}$$

which we now consider as equations in the unknowns c_{ij}, for a fixed position of the nodes. It can be shown that these equations are independent (Exercise 6.3), and hence an $(m-2n+6)$-dimensional linear space of the c_{ij} lead to one and the same drawing. But this still leaves us with a manifold of drawings with dimension $m - (m - 2n + 6) = 2n - 6$, while the representation with touching circles is uniquely determined.

Let us fix the rubber band strengths arbitrarily, and consider the resulting drawing. Recall that the positions x_i of the nodes can be obtained by minimizing the quadratic function

$$\sum_{ij\in E(G)} c_{ij}|x_i - x_j|^2, \tag{6.15}$$

or equivalently, by solving the linear equations

$$\sum_{j\in N(i)} c_{ij}(x_i - x_j) = 0 \qquad (i \neq a, b, c) \tag{6.16}$$

(while keeping x_a, x_b and x_c fixed).

When will this come from a touching circle representation? Clearly a necessary and sufficient condition is that the system of equations

$$r_i + r_j = |x_i - x_j| \tag{6.17}$$

has a solution in the r_i such that all of them are positive. This system consists of $3n - 6$ equations with n unknowns, so it is very much overdetermined. We can compute the best approximate solution by minimizing the quadratic function

$$\sum_{ij\in E(G)} (r_i + r_j - |x_i - x_j|)^2 \tag{6.18}$$

in the variables r_i. The function is strictly convex (since G is nonbipartite), hence the optimum solution is unique, and satisfies

$$\sum_{j\in N(i)} (r_i + r_j - |x_i - x_j|) = 0. \tag{6.19}$$

for all i. This means that these r_i satisfy (6.17) "on the average" at each node, but of course not for particular edges.

Intuitively, if $r_i + r_j < |x_i - x_j|$ for an edge, then it is too long, so we want to increase its strength, while in the other case, we want to decrease it.

[A natural update is

$$c'_{ij} = \frac{|x_i - x_j|}{r_i + r_j} c_{ij}.$$

With these new strengths, we get new positions x'_i for the nodes by solving (6.16), and then new approximate radii r'_i by solving (6.19).]

Set $\Delta x_i = x_i - x'_i$, $\Delta r_i = r_i - r'_i$, $\Delta c_{ij} = c_{ij} - c'_{ij}$ and $\Delta d_{ij} = d_{ij} - d'_{ij}$, where $d_{ij} = |x_i - x_j|$. Then

$$\begin{aligned}\Delta\mathcal{E} &= \sum_{ij \in E(G)} (r_i + r_j - |x_i - x_j|)^2 - \sum_{ij \in E(G)} (r'_i + r'_j - |x'_i - x'_j|)^2 \\ &= \sum_{ij \in E(G)} (\Delta r_i + \Delta r_j - \Delta d_{ij})(r_i + r'_i + r_j + r'_j - d_{ij} - d'_{ij}).\end{aligned}$$

Here all terms containing Δr_i sum to 0 by (6.19), and similarly for all terms containing Δr_i. Thus

$$\Delta\mathcal{E} = -\sum_{ij \in E(G)} \Delta d_{ij}(r_i + r'_i + r_j + r'_j - d_{ij} - d'_{ij}).$$

Exercise 6.2 Prove that every Andre'ev representation of a planar graph is a rubber band representation with appropriate rubber band strengths.

Exercise 6.3 Show that equations 6.14 are independent.

6.1.8 Conformal transformations

The double circuit representation of a planar graph is uniquely determined, once a triangular infinite face is chosen and the circuits representing the nodes of this triangular face are fixed. This follows from Lemma 6.1.9. Similar assertion is true for double circuit representations in the sphere. However, in the sphere there is no difference between faces, and we may not want to "normalize" by fixing a face. Often it is more useful to apply a circle-preserving transformation that distributes the circles on the sphere in a "uniform" way. The following Lemma shows that this is possible with various notions of uniformity.

Lemma 6.1.12 *Let $F : (B^3)^n \to B^3$ be a continuous map with the property that whenever $n-2$ of the vectors u_i are equal to $u \in S^2$ then $u^{\mathsf{T}} F(u_1, \ldots, u_n) \geq 0$. Let $(C_1, \ldots, C_n)$ be a family of openly disjoint caps on the sphere. Then there is a circle preserving transformation A of the sphere such that $F(v_1, \ldots, v_n) = 0$, where $v_i = v(AC_i)$ is the center of AC_i.*

Examples of functions F to which this lemma are the center of gravity of $u_1, \ldots, u_n$, or the center of gravity of their convex hull, or the center of the inscribed ball.

Proof. For every interior point p of the unit ball, we define a circle-preserving transformation A_p of the unit sphere as follows. Take a tangent plane T at $p_0 = p/\|p\|$, and project the sphere stereographically on T; blow up the plane from center p_0 by a factor $\lambda = 1/(1-\|p\|)$; and project it back stereographically to the sphere. Let $v_i(p)$ denote the center of the cap $A_p(C_i)$. (Warning: this is not the image of the center of C_i in general! Circle-preserving transformations don't preserve the centers of circles.)

We want to show that the range of F contains the origin. Suppose not. Let $0 < \varepsilon_1, \varepsilon_2, \cdots < 1$ be a sequence tending to 0. Let B_k be the ball with radius $1 - \varepsilon_k$ about the origin. For $x \in B_k$, define

$$f_k(x) = (1-\varepsilon_k)F(v(A_xC_1), \ldots, v(A_xC_n))/|F(v(A_xC_1), \ldots, v(A_xC_n))|.$$

Then f_k is a continuous map from B_k to B_k, and so by Brouwer's Fixed Point Theorem, it has a fixed point p_k. Clearly $|p_k| = 1 - \varepsilon_k$. We may assume by compactness that $p_k \to p \in S^2$ and also that $v(A_{p_k}C_i) \to w_i \in S^2$. By the continuity of F, we have $F(v(A_{p_k}C_1), \ldots, v(A_{p_k}C_n)) \to F(w_1, \ldots, w_n)$. Furthermore,

$$\frac{F(w_1, \ldots, w_n)}{|F(w_1, \ldots, w_n)|} = \lim_{k\to\infty} \frac{F(v(A_{p_k}C_1), \ldots, v(A_{p_k}C_n))}{|F(v(A_{p_k}C_1), \ldots, v(A_{p_k}C_n))} = \lim_{k\to\infty} \frac{1}{1-\varepsilon_k} f_k(p_k) = \lim_{k\to\infty} p_k = p.$$

By elementary geometry, for every point $x \in S^2$, $x \neq p$, the images $A_{p_k}x \to -p$. Hence it follows that if C_i does not contain p, then $w_i = -p$. So at most two of $\{w_1, \ldots, w_n\}$ are different from the South pole, and hence $F(w_1, \ldots, w_n)/|F(w_1, \ldots, w_n)|$ is contained in the Southern hemisphere. This is a contradiction. □

6.1.9 Applications of circle packing

Planar separators

Koebe's Theorem has several important applications. We start with a simple proof by Miller and Thurston [151] of the Planar Separator Theorem 1.2.1 of Lipton and Tarjan [127] (we present the proof with a weaker bound of $3n/4$ on the sizes of the components instead of $2n/3$; see [187] for an improved analysis of the method).

We need the notion of the "statistical center", which is important in many other studies in geometry. Before defining it, we prove a simple lemma.

Lemma 6.1.13 *For every set $S \subseteq \mathbb{R}^d$ of n points there is a point $c \in \mathbb{R}^n$ such that every closed halfspace containing c contains at least $n/(d+1)$ elements of S.*

Proof. Let $\mathcal{H}$ be the family of all closed halfspaces that contain more than $dn/(d+1)$ points of S. The intersection of any $(d+1)$ of these still contains an element of S, so in particular it is nonempty. Thus by Helly's Theorem, the intersection of all of them is nonempty. We claim that any $c \in \cap\mathcal{H}$ satisfies the conclusion of the Lemma.

If H be any open halfspace containing c, then $\mathbb{R}^d \setminus H \notin \mathcal{H}$, which means that H contains at least $n/(d+1)$ points of S. If H is a closed halfspace containing c, then it is contained in an open halfspace H' that intersects S in exactly the same set, and applying the previous argument to H' we are done. □

A point c as in Lemma 6.1.13 is sometimes called a "statistical center" of the set S. To make this point well-defined, we call the center of gravity of all points c satisfying the conclusion of Lemma 6.1.13 *the statistical center* of the set (note: points c satisfying the conclusion of Lemma form a convex set, whose center of gravity is well defined).

Proof of Theorem 1.2.1. Let $(C_i : i \in V)$ be a Koebe representation of G on the unit sphere, and let u_i be the center of C_i on the sphere, and ρ_i, the spherical radius of C_i. By Lemma 6.1.12, we may assume that the statistical center of the points u_i is the origin.

Take any plane H through 0. Let S denote the set of nodes i for which C_i intersects H, and let S_1 and S_2 denote the sets of nodes for which C_i lies on one side and the other of H. Clearly there is no edge between S_1 and S_2, and so the subgraphs G_1 and G_2 are disjoint and their union is $G \setminus S$. Since 0 is a statistical center of the u_i, it follows that $|S_1|, |S_2| \le 3n/4$.

It remains to make sure that S is small. To this end, we choose H at random, and estimate the expected size of S.

What is the probability that H intersects C_i? If $\rho_i \ge \pi/2$, then this probability is 1, but there is at most one such node, so we can safely ignore it, and suppose that $\rho_i < \pi/2$ for every i. By symmetry, instead of fixing C_i and choosing H at random, we can fix H and choose the center of C_i at random. Think of H as the plane of the equator. Then C_i will intersect H if and only if it is center is at a latitude at most ρ_i (North or South). The area of this belt around the equator is, by elementary geometry, $4\pi \sin\rho_i$, and so the probability that the center of C_i falls into here is $2\sin\rho_i$. It follows that the expected number of caps C_i intersected by H is $\sum_{i\in V} 2\sin\rho_i$.

To get an upper bound on this quantity, we use the surface area of the cap C_i is $2\pi(1-\cos\rho_i) = 4\pi\sin^2(\rho_i/2)$, and since these are disjoint, we have

$$\sum_{i\in V} \left(\sin\frac{\rho_i}{2}\right)^2 < 1. \tag{6.20}$$

Using that $\sin\rho_i \le 2\sin\frac{\rho_i}{2}$, we get by Cauchy-Schwartz

$$\sum_{i\in V} 2\sin\rho_i \le 2\sqrt{n}\left(\sum_{i\in V}(\sin\rho_i)^2\right)^{1/2} \le 4\sqrt{n}\left(\sum_{i\in V}\left(\sin\frac{\rho_i}{2}\right)^2\right)^{1/2} < 4\sqrt{n}.$$

So the expected size of S is less than $4\sqrt{n}$, and so there is at least one choice of H for which $|S| < 4\sqrt{n}$. □

Laplacians of planar graphs

The Planar Separator theorem was first proved by direct graph-theoretic arguments (and other elegant graph-theoretic proofs are available [14]). For the following theorem on the eigenvalue gap of the Laplacian of planar graphs by Spielman and Teng [186] there is no proof known avoiding Koebe's theorem.

Theorem 6.1.14 *For every connected planar graph $G = (V, E)$ on n nodes and maximum degree D, the second smallest eigenvalue of L_G is at most $8D/n$.*

Proof. Let $C_i : i \in V$ be a Koebe representation of G on the unit sphere, and let u_i be the center of C_i, and ρ_i, the spherical radius of C_i. By Lemma 6.1.12 may assume that $\sum_i u_i = 0$.

The second smallest eigenvalue of L_G is given by

$$\lambda_2 = \min_{\substack{x \neq 0 \\ \sum_i x_i = 0}} \frac{\sum_{ij \in E}(x_i - x_j)^2}{\sum_{i \in V} x_i^2},$$

Let $u_i = (u_{i1}, u_{i2}, u_{i3})$, then this implies that

$$\sum_{ij \in E} (u_{ik} - u_{jk})^2 \geq \lambda_2 \sum_{i \in V} u_{ik}^2$$

holds for every coordinate k, and summing over k, we get

$$\sum_{ij \in E} \|u_i - u_j\|^2 \geq \lambda_2 \sum_{i \in V} \|u_i\|^2 = \lambda_2 n. \tag{6.21}$$

On the other hand, we have

$$\begin{aligned} \|u_i - u_j\|^2 = 4\left(\sin\frac{\rho_i + \rho_j}{2}\right)^2 &= 4\left(\sin\frac{\rho_i}{2}\cos\frac{\rho_j}{2} + \sin\frac{\rho_j}{2}\cos\frac{\rho_i}{2}\right)^2 \\ &\leq 4\left(\sin\frac{\rho_i}{2} + \sin\frac{\rho_j}{2}\right)^2 \leq 8\left(\sin\frac{\rho_i}{2}\right)^2 + 8\left(\sin\frac{\rho_j}{2}\right)^2, \end{aligned}$$

and so by (6.20)

$$\sum_{ij \in E} \|u_i - u_j\|^2 \leq 8D \sum_{i \in V} \left(\sin\frac{\rho_i}{2}\right)^2 \leq 8D.$$

Comparison with (6.21) proves the theorem. □

This theorem says that planar graphs are very bad expanders. The result does not translate directly to eigenvalues of the adjacency matrix or the transition matrix of the random walk on G, but for graphs with bounded degree it does imply the following:

Corollary 6.1.15 *Let G be a connected planar graph on n nodes with maximum degree D. Then the second largest eigenvalue of the transition matrix is at least $1 - 8D/n$, and the mixing time of the random walk on G is at least $\Omega(n/D)$.*

Among other applications, we mention a bound on the cover time of the random walk on a planar graph by Jonasson and Schramm [107].

Exercise 6.4 Show by an example that the bound in Lemma 6.1.13 is sharp.

6.1.10 Circle packing and the Riemann Mapping Theorem

Koebe's Circle Packing Theorem and the Riemann Mapping Theorem in complex analysis are closely related. More exactly, we consider the following generalization of the Riemann Mapping Theorem.

Theorem 6.1.16 (The Koebe-Poincaré Uniformization Theorem) *Every open domain in the sphere whose complement has a finite number of connected components is conformally equivalent to a domain obtained from the sphere by removing a finite number of disjoint disks and points.*

In fact, the Circle Packing Theorem and the Uniformization Theorem are mutually limiting cases of each other (Koebe [115], Rodin and Sullivan [173]). The exact proof of this fact has substantial technical difficulties, but it is not hard to describe the idea.

1. To see that the Uniformization Theorem implies the Circle Packing Theorem, let G be a planar map and G^* its dual. We may assume that G and G^* are 3-connected, and that G^* has straight edges (these assumptions are not essential, just convenient). Let $\varepsilon > 0$, and let U denote the ε-neighborhood of G^*. By Theorem 6.1.16, there is a conformal map of U onto a domain $D' \subseteq S^2$ which is obtained by removing a finite number of disjoint caps and points from the plane (which we consider as degenerate caps). If ε is small enough, then these caps are in one-to-one correspondence with the nodes of G. We normalize using Lemma 6.1.12 and assume that the center of gravity of the cap centers is 0.

Letting $\varepsilon \to 0$, we may assume that the cap representing any given node $v \in V(G)$ converges to a cap C_v. It is not hard to argue that these caps are non-degenerate, caps representing different nodes tend to openly disjoint caps, and caps representing adjacent nodes tend to caps that are touching.

2. In the other direction, let $U = S^2 \setminus K_1 \setminus \cdots \setminus K_n$, where $K_1, \ldots, K_n$ are disjoint closed connected sets which don't separate the sphere. Let $\varepsilon > 0$. It is not hard to construct a family $\mathcal{C}(\varepsilon)$ of openly disjoint caps such that the radius of each cap is less than ε and their tangency graph G is a triangulation of the sphere.

Let H_i denote the subgraph of G consisting of those edges intersecting K_i'. If ε is small enough, then the subgraphs H_i are node-disjoint, and each H_i is nonempty except possibly if K_i is a singleton. It is also easy to see that the subgraphs H_i are connected.

Let us contract each nonempty connected H_i to a single node w_i. If K_i is a singleton set and H_i is empty, we add a new node w_i to G in the triangle containing K_i', and connect it to the nodes of this triangle. The spherical map G' obtained this way can be represented as the tangency graph of a family of caps $\mathcal{D} = \{D_u : u \in V(G')\}$. We can normalize so that the center of gravity of the centers of $D_{w_1}, \ldots, D_{w_n}$ is the origin.

Now let $\varepsilon \to 0$. We may assume that each $D_{w_i} = D_{w_i}(\varepsilon)$ tends to a cap $D_{w_i}(0)$. Furthermore, we have a map f_ε that assigns to each node u of G_ε the center of the corresponding cap D_u. One can prove (but this is nontrivial) that these maps f_ε, in the limit as $\varepsilon \to 0$, give a conformal map of U onto $S^2 \setminus D_{w_1}(0) \setminus \cdots \setminus D_{w_n}(0)$.

6.2 *Extensions

There are extensions by Andre'ev [16] and Thurston [195] to circles meeting at other angles.

6.2.1 Orthogonal circles

One of the many extensions of Koebe's Theorem characterizes triangulations of the plane that have a representation by orthogonal circles: more exactly, circles representing adjacent nodes must intersect at 90°, other pairs, at $> 90°$ (i.e., their centers must be farther apart) [16, 195, 118] (Figure 6.5.

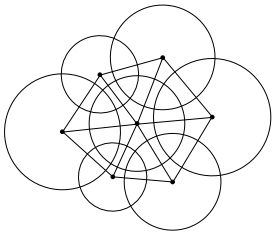

Figure 6.5: Representing a planar graph by orthogonal circles

Such a representation, if it exists, can be projected to a representation by orthogonal circles on the unit sphere; with a little care, one can do the projection so that each disk bounded by one of the circles is mapped onto a "cap" which covers less than half of the sphere. Then each cap has a unique *pole*: the point in space from which the part of the sphere you see is exactly the given cap. The key observation is that *two circles are orthogonal if and only if the corresponding poles have inner product* 1 (Figure 6.6). This translates a representation with orthogonal circles into a representation by vectors of length larger than

1, where adjacent nodes are represented by vectors with inner product 1, non-adjacent nodes by vectors with inner product less than 1.

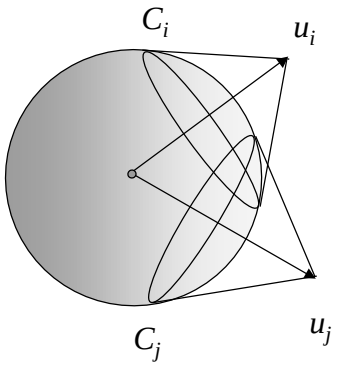

Figure 6.6: Poles of circles

This in turn can be translated into semidefinite matrices. We only state the final result of these transformations. Consider the following two sets of semidefinite constraints:

$$\begin{aligned} Y &\succeq 0 \\ Y_{ij} &= 1 \qquad \forall\, ij \in E, \\ Y_{ij} &< 1 \qquad \forall\, ij \notin E,\ i \neq j, \\ Y_{ii} &> 1 \end{aligned} \tag{6.22}$$

and the weaker set of constraints

$$\begin{aligned} Y &\succeq 0 \\ Y_{ij} &= 1 \qquad \forall\, ij \in E, \\ Y_{ij} &< 1 \qquad \forall\, ij \notin E,\ i \neq j, \end{aligned} \tag{6.23}$$

$$\tag{6.24}$$

To formulate the theorem, we need two simple definitions. A cycle C in a graph G is called *separating*, if $G \setminus V(C)$ has at least two connected components, where any chord of C is counted as a connected component here. The cycle C is called *strongly separating*, if $G \setminus V(C)$ has at least two connected components, each of which has at least 2 nodes. If G is a 3-connected planar map, then its non-separating cycles are exactly the boundaries of the faces.

The following was proved in [118]:

Theorem 6.2.1 *Let G be a 3-connected planar graph.*

(a) *If (6.23) has a solution of rank 3, then G is planar.*

(b) *Assume that G is a maximal planar graph. Then (6.22) has a solution of rank 3 if and only if G has no separating 3- and 4-cycles.*

(c) *Assume that G is a maximal planar graph. Then (6.23) has a solution with rank 3 if and only if G has no strongly separating 3- and 4-cycles.*

6.2.2 Tangency graphs of general convex domains

Let $\mathcal{H}$ be a family of closed convex domains in the plane, such that their interiors are disjoint. If we represent each of these domains by a node, and connect two of these nodes if the corresponding domains touch, we get a $G_{\mathcal{H}}$, which we call the *tangency graph* of the family.

Exercise 6.5 Let $\mathcal{H}$ be a family of convex domains in the plane with smooth boundaries and disjoint interiors. Then the tangency graph of $\mathcal{H}$ is planar.

In what follows, we restrict ourselves to the case when the members of $\mathcal{H}$ are all homothetical copies of a centrally symmetric convex domain. It is natural in this case to represent each domain by its center.

Exercise 6.6 Let $\mathcal{H}$ be a family of homothetical centrally symmetric convex domains in the plane with smooth boundaries and disjoint interiors. Then the centers of the bodies give a straight line embedding in the plane of its tangency graph (Figure 6.7).

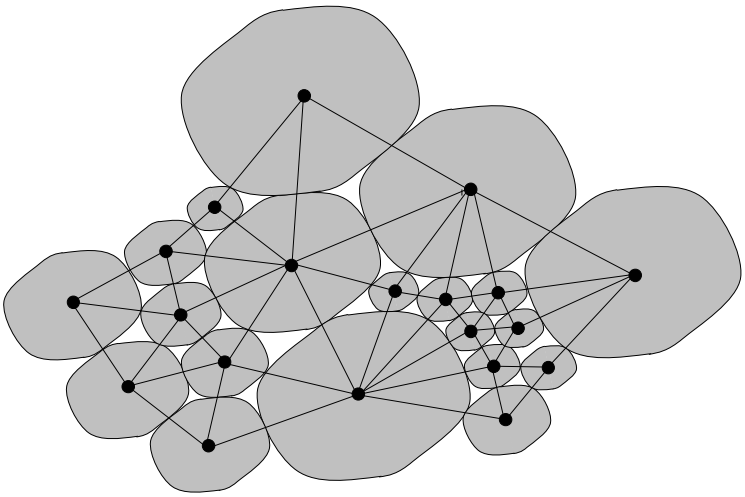

Figure 6.7: Straight line embedding of a planar graph from touching convex figures.

Schramm [179] proved the following deep converse to this fact.

Theorem 6.2.2 *For a smooth strictly convex domain D, every planar graph can be represented as the tangency graph of a family of homothetical copies of D.*

Schramm [180] also proved the following very general extension of Theorem 6.1.3:

Theorem 6.2.3 (Caging the Egg) *Given a smooth strictly convex body C in $\mathbb{R}^3$, every 3-connected planar graph has a Steinitz representation such that all its edges touch C.*

The smoothness of the domain is a condition that cannot be dropped. For example, K_4 cannot be represented by touching squares. A strong result about representation by touching squares, also due to Schramm, will be stated precisely and proved in section 6.3.

6.3 Square tilings

We can also represent planar graphs by squares, rather then circles, in the plane (with some mild restrictions). There are in fact two quite different ways of doing this: the squares can correspond to the edges (a classic result of Brooks, Smith, Stone and Tutte), or the squares can correspond to the nodes (a quite recent result of Schramm).

6.3.1 Current flow through a rectangle

A beautiful connection between square tilings and harmonic functions was described in the classic paper of Brooks, Smith, Stone and Tutte [35]. They considered tilings of squares by smaller squares, and used a physical model of current flows to show that such tilings can be obtained from any connect planar graph. Their ultimate goal was to construct tilings of a square with squares whose edge-lengths are all different; this will not be our concern; we'll allow squares that are equal and also the domain being tiled can be a rectangle, not necessarily a square.

Consider tiling $\mathcal{T}$ of a rectangle R with a finite number of squares, whose sides are parallel to the coordinate axes. We associate a planar map with this tiling as follows. Represent any maximal horizontal segment composed of edges of the squares by a single node (say, positioned at the midpoint of the segment). Each square "connects" two horizontal segments, and we can represent it by an edge connecting the two corresponding nodes, directed top-down. We get a directed graph $G_{\mathcal{T}}$ (Figure 6.8), with a single source s (representing the upper edge of the rectangle) and a single sink t (representing the upper edge). It is not hard to see that $G_{\mathcal{T}}$ is planar.

A little attention must be paid to points where four squares meet. Suppose that A, B, C, D share a corner p, where A is the upper left, and B, C, D follow clockwise. In this case, we may consider the lower edges of A and B to belong to a single horizontal segment, or to belong to different horizontal segments. In the latter case, we may or may not imagine that there is an infinitesimally small square sitting at p. What this means is that we have to declare if the four edges of $G_{\mathcal{T}}$ corresponding to A, B, C and D form two pairs of parallel edges, an empty

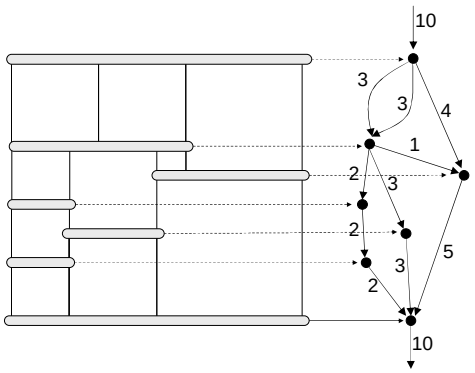

Figure 6.8: The Brooks–Smith–Stone–Tutte construction

quadrilateral, or a quadrilateral with a horizontal diagonal. We can orient this horizontal edge arbitrarily (Figure 6.9).

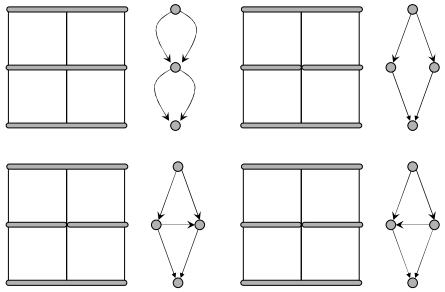

Figure 6.9: Possible declarations about four squares meeting at a point

If we assign the edge length of each square to the corresponding edge, we get a flow f from s to t: If a node v represents a segment I, then the total flow into v is the sum of edge length of squares attached to I from the top, while the total flow out of v is the sum of edge length of squares attached to I from the bottom. Both of these sums are equal to the length of I.

Let $h(v)$ denote the distance of node v from the upper edge of R. Since the edge-length of a square is also the difference between the y-coordinates of its upper and lower edges, the function h is harmonic:

$$h(i) = \frac{1}{d_i} \sum_{j \in N(i)} h(j)$$

for every node different from s and t (Figure 6.8).

Theorem 6.3.1 *For every connected planar map G with two specified nodes s and t on the unbounded face, there is a unique tiling $\mathcal{T}$ of a rectangle such that G is $G_{\mathcal{T}}$.*

6.3.2 Tangency graphs of square tilings

Let R be a rectangle in the plane, and consider a tiling of R by squares. Let us add four further squares attached to each edge of R from the outside, sharing the edge with R. We want to look at the tangency graph of this family of squares.

Since the squares do not have a smooth boundary, the assertion of Exercise 6.5 does not apply, and the tangency graph of the squares may not be planar. Let us try to draw the tangency graph in the plane by representing each square by its center and connecting the centers of touching squares by a straight line segment. We get into trouble when four squares share a vertex, in which case two edges will cross at this vertex. In this case we specify arbitrarily one diametrically opposite pair as "infinitesimally overlapping", and connect the centers of these two but not the other two centers. We call this a *resolved tangency graph.*

Every resolved tangency graph is planar, and it is easy to see that it has exactly one face that is a quadrangle (namely, the infinite face), and its other faces are triangles; briefly, it is a triangulation of a quadrilateral (Figure 6.10).

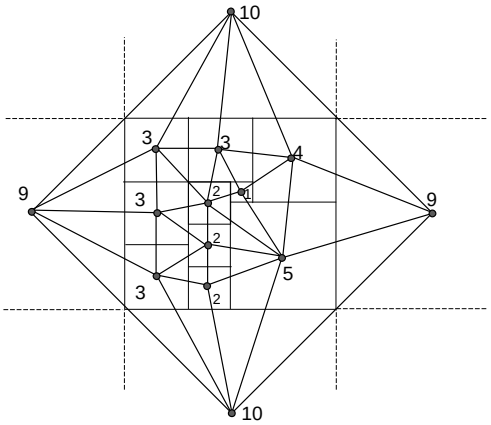

Figure 6.10: The resolved tangency graph of a tiling of a rectangle by squares

Under some mild conditions, this fact has a converse, due to Schramm [181].

Theorem 6.3.2 *Every planar triangulation of a quadrilateral such that no 3-cycle or 4-cycle contains a point in the interior can be represented as a resolved tangency graph of a square tiling of a rectangle.*

Before proving this theorem, we need some lemmas from combinatorial optimization.

Let G be a planar triangulation of a quadrilateral $q_1q_2q_3q_4$. We assign a real variable x_i to each node i (this will eventually mean the side length of the square representing i, but at

the moment, it is just a variable). Consider the following conditions:

$$x_i \geq 0 \qquad \text{for all nodes } i, \tag{6.25}$$

$$\sum_{i \in P} x_i \geq 1 \qquad \text{for all } q_1\text{–}q_3 \text{ paths } P \tag{6.26}$$

(In this proof, when we talk about paths we mean their node sets.) Let $P \subseteq \mathbb{R}^V$ denote the solution set of these inequalities. It is clear that P is an ascending polyhedron.

Lemma 6.3.3 *The vertices of P are the incidence vectors of q_2–q_4 paths. The blocker of P is defined by the inequalities*

$$x_i \geq 0 \qquad \textit{for all nodes } i, \tag{6.27}$$

$$\sum_{i \in Q} x_i \geq 1 \qquad \textit{for all } q_2\text{–}q_4 \textit{ paths } Q. \tag{6.28}$$

The vertices of P^{bl} are the incidence vectors of q_1–q_3 paths.

These facts can be derived from the Max-Flow-Min-Cut Theorem (or can be proved directly using the uncrossing method (see Claim 6.3.6 below).

Proof of Theorem 6.3.2. Consider the solution $\overline{x}$ of (6.25)–(6.26) minimizing the objective function $\sum_i x_i^2$, and let R^2 be the minimum value. By Theorem 13.7.12, $\overline{y} = (1/R^2)\overline{x}$ minimizes the same objective function over P^{bl}. Let us rescale these vectors to get $a = \frac{1}{R}\overline{x} = R\overline{y}$. Then we have

$$\sum_{i \in P} a_i \geq \frac{1}{R} \qquad \text{for all } q_1\text{–}q_3 \text{ paths } P \tag{6.29}$$

$$\sum_{i \in Q} a_i \geq R \qquad \text{for all } q_2\text{–}q_4 \text{ paths } Q \tag{6.30}$$

$$\sum_{i \in V} a_i^2 = 1. \tag{6.31}$$

Furthermore,

$$a^{\mathsf{T}} a = \frac{1}{R^2} \sum_i x_i^2 = 1. \tag{6.32}$$

For each path $P = (v_1, \ldots, v_k)$, we define is *a-length* by

$$\ell_a(P) = \frac{1}{2} a_{v_1} + a_{v_2} + \cdots + a_{v_{k-1}} + \frac{1}{2} a_{v_k}.$$

Note that if we cut a path into two, then their a-lengths add up to the a-length of the original path. For any two nodes, their *a-distance* as the minimum a-length of a path connecting them.

We know that $\overline{x}$ is in the polyhedron defined by (6.25)–(6.26), and so it can be written as a convex combination of vertices, which are incidence vectors of node sets of q_2–q_4 paths. Hence a can be written as

$$a = \sum_{P \in \mathcal{P}} \lambda_P \mathbf{1}_P, \tag{6.33}$$

where $\mathcal{P}$ is a set of q_2–q_4 paths, $\lambda_i > 0$ and $\sum_i \lambda_i = \frac{1}{R}$. (Note that we don't include paths with weight 0.) Similarly, we have a decomposition

$$a = \sum_{Q \in \mathcal{Q}} \mu_Q \mathbf{1}_Q, \tag{6.34}$$

where the $\mathcal{Q}$ is a set of q_1–q_3 paths, $\mu_j > 0$ and $\sum_j \mu_j = R$. Trivially a node v has $a_v > 0$ if and only if it is contained in one of the paths in $\mathcal{P}$ (equivalently, in one of the paths in $\mathcal{Q}$).

Claim 6.3.4 *$|P \cap Q| = 1$ for all $P \in \mathcal{P}, Q \in \mathcal{Q}$.*

It is trivial that $|P \cap Q| \geq 1$. By (6.32),

$$\begin{aligned} 1 = a^{\mathsf{T}} a &= \Big(\sum_P \lambda_P \mathbf{1}_P\Big)\Big(\sum_Q \mu_Q \mathbf{1}_Q\Big) = \sum_{P,Q} \lambda_P \mu_Q |P \cap Q| \\ &\geq \sum_{P,Q} \lambda_P \mu_Q = \Big(\sum_P \lambda_P\Big)\Big(\sum_Q \mu_Q\Big) = 1. \end{aligned}$$

We must have equality here, which proves the Claim.

Claim 6.3.5 *All paths in $P \in \mathcal{P}$ have $a(P) = R$, and all paths $Q \in \mathcal{Q}$ have $a(Q) = 1/R$.*

Let (say) $Q \in \mathcal{Q}$, then

$$\sum_{i \in Q} a_i = a^{\mathsf{T}} \mathbf{1}_Q = \sum_{P \in \mathcal{P}} \lambda_P \mathbf{1}_P^{\mathsf{T}} \mathbf{1}_Q = \sum_{P \in \mathcal{P}} \lambda_P |P \cap Q| = \sum_{P \in \mathcal{P}} \lambda_P = \frac{1}{R}.$$

One consequence of this Claim is that the paths in $\mathcal{P}$ and $\mathcal{Q}$ are chordless, since if (say) $P \in \mathcal{P}$ had a cord uv, then bypassing the part of P between u and v would decrease its total weight, which would contradict (6.29).

Next, we simplify the decompositions (6.33) and (6.34). In the plane embedding of G with $F = q_1q_2q_3q_4$ as the infinite face, every q_1–q_3 path Q cuts the interior of F into two parts (one of them may be empty for the paths $q_1q_2q_3$ and $q_1q_4q_3$). We think of F as a diamond standing on its vertex q_3, so we can call the two parts the "left side" and "right side" of Q. For two q_1–q_3 paths we say that Q' is to the right of Q, if every node of it is either on Q or on the right side of Q. This defines a partial order on the set of q_1–q_3 paths. We say that a family $\mathcal{Q}$ of q_1–q_3 paths is *laminar*, if it is totally ordered. Similarly, every q_2–q_4 path P has an "upper side" and a "lower side", and we can define when a family of q_2–q_4 paths is laminar.

The following step is known in combinatorial optimization as "uncrossing".

Claim 6.3.6 *In* (6.33) *and* (6.34) *the families* $\mathcal{P}$ *and* $\mathcal{Q}$ *can be chosen so that they are laminar.*

Consider any two paths $Q, Q' \in \mathcal{Q}$, and suppose that neither of them is to the right of the other. Then they must intersect. Let i be a common node of them, then their parts between q_1 and i must have the same ℓ_a-length (else, we could interchange these parts and get a path contradicting (6.29). Hence their common points are in the same order on both of them, and interchanging their parts between consecutive common points arbitrarily results in two paths of the same length. This way we can define the path $Q \wedge Q'$, which is obtained by taking always the parts that are to the left, and the path $Q \vee Q'$, which is obtained by taking always the parts that are to the right.

Now assume that $\mu \leq \mu'$, and replace $\mu_Q \mathbf{1}_Q + \mu_{Q'} \mathbf{1}_{Q'}$ with $\mu_Q \mathbf{1}_{Q \wedge Q'} + \mu_Q \mathbf{1}_{Q \vee Q'} Q + (\lambda_{Q'} - \lambda_Q)\mathbf{1}_{Q'}$ in (6.33). It is easy to check that we get another valid decomposition, and that $Q \wedge Q'$, $Q \wedge Q'$ and Q' are laminar. We call this procedure *uncrossing* Q *and* Q'.

What needs to be argued is that repeating this uncrossing, in a finite number of steps we arrive at a laminar family. This is possible but a bit tedious; instead, we take a shortcut as follows. Let $L(Q)$ denote the set of nodes to the left of the q_1–q_3 path Q. Then $L(Q \wedge Q') = L(Q) \cap L(Q')$ and $L(Q \vee Q') = L(Q) \cup L(Q')$, from which it follows that uncrossing increases the sum

$$\sum_{Q \in \mathcal{Q}} \lambda_Q |L(Q)|^2. \tag{6.35}$$

So if we start with a decomposition (6.33) with (6.35) maximal (which clearly exists), then $\mathcal{Q}$ is laminar.

From now on, we assume that $\mathcal{P}$ and $\mathcal{Q}$ are laminar families.

Claim 6.3.7 *Every node* i *has* $a_i > 0$.

Suppose that $a_i = 0$. Then clearly no path in $\mathcal{P} \cup \mathcal{Q}$ goes through i. Let Q_1 [Q_2] be the rightmost [leftmost] path in $\mathcal{Q}$ to the left [right] of i, and similarly, let P_1 [P_2] be the lowest [highest] path in $\mathcal{P}$ above [below] i. Each P_t intersects each Q_s in exactly one node u_{ts}; let P'_t be the part of P_t connecting the intersection points u_{t1} and u_{t2} (this may be just a single node if $u_{t1} = u_{t2}$). We define Q'_s analogously.

The path P'_t cannot have more than one edge. Indeed, suppose that it has an interior node j. Then $a_j > 0$ (since $j \in P_t$), and so there is path $Q \in \mathcal{Q}$ through j. But then Q is between Q_1 and Q_2 in the ordering of $\mathcal{Q}$, which contradicts the choice of Q_1 and Q_2. Similarly, the paths Q'_s have at most one edge.

Thus the cycle $C = P'_1 \cup P'_2 \cup Q'_1 \cup Q'_2$ is either a triangle or a quadrilateral, and it is separating since i is in its interior. This is a contradiction, which proves the Claim.

Claim 6.3.8 *Let $i, j \in V$. Then there is a path $P \in \mathcal{P}$ going through both i and j if and only if $|d_a(i, q_1) - d_a(j, q_1)| < \frac{1}{2}a_i + \frac{1}{2}a_j$.*

Let $\mathcal{P}_i$ be the set of those paths in $\mathcal{P}$ that go through i, and $\mathcal{P}'_i$, the set of those paths in $\mathcal{P}$ that go above i. Define $\mathcal{P}_j$ and $\mathcal{P}'_j$ similarly. We have $a_i = \lambda(\mathcal{P}_i)$. Furthermore, if Q' is the part of a path $Q \in \mathcal{P}$ through i between q_1 and i, then every path in $\mathcal{P}'_j$ intersects Q' in exactly one point, and so $d_a(q_1, i) = \lambda(\mathcal{P}'_i) + \frac{1}{2}\lambda(\mathcal{P}_i) + \frac{1}{2}a_{q_1}$. Similarly, $a_j = \lambda(\mathcal{P}_j)$ and $d_a(q_1, j) = \lambda(\mathcal{P}'_j) + \frac{1}{2}\lambda(\mathcal{P}_j) + \frac{1}{2}a_{q_1}$.

Suppose $\mathcal{P}_i$ and $\mathcal{P}_j$ are disjoint, then by the assumption that $\mathcal{P}$ is laminar, one of them is above the other. Say $\mathcal{P}_i$ is above $\mathcal{P}_j$, then $\mathcal{P}'_i \cup \mathcal{P}_i \subseteq \mathcal{P}'_j$ and hence $d_a(i, q_1) < d_a(j, q_1)$. Thus

$$\begin{aligned}|d_a(i, q_1) - d_a(j, q_1)| &= \lambda(\mathcal{P}'_j) + \frac{1}{2}\lambda(\mathcal{P}_j) - \lambda(\mathcal{P}'_i) - \frac{1}{2}\lambda(\mathcal{P}_i) \\ &\geq \lambda(\mathcal{P}'_i) + \lambda(\mathcal{P}_i) + \frac{1}{2}\lambda(\mathcal{P}_j) - \lambda(\mathcal{P}'_i) - \frac{1}{2}\lambda(\mathcal{P}_i) = \frac{1}{2}a_i + \frac{1}{2}a_j.\end{aligned}$$

On the other hand, if $\mathcal{P}_i$ and $\mathcal{P}_j$ have a path in common, then we have $\mathcal{P}'_j \subset \mathcal{P}'_i \cup \mathcal{P}_i$ (strict containment!), and so

$$\begin{aligned}d_a(j, q_1) - d_a(i, q_1) &= \lambda(\mathcal{P}'_j) + \frac{1}{2}\lambda(\mathcal{P}_j) - \lambda(\mathcal{P}'_i) - \frac{1}{2}\lambda(\mathcal{P}_i) \\ &< \lambda(\mathcal{P}'_i) + \lambda(\mathcal{P}_i) + \frac{1}{2}\lambda(\mathcal{P}_j) - \lambda(\mathcal{P}'_i) - \frac{1}{2}\lambda(\mathcal{P}_i) = \frac{1}{2}a_i + \frac{1}{2}a_j.\end{aligned}$$

The same bound for the difference $d_a(i, q_1) - d_a(j, q_1)$ follows similarly.

Note the following consequence:

Claim 6.3.9 *The inequalities $|d_a(i, q_1) - d_a(j, q_1)| < \frac{1}{2}a_i + \frac{1}{2}a_j$ and $|d_a(i, q_2) - d_a(j, q_2)| < \frac{1}{2}a_i + \frac{1}{2}a_j$ cannot simultaneously hold.*

Indeed, by Claim 6.3.8 this would imply that ij is contained in one of the paths in $\mathcal{P}$ as well as in one of the paths in $\mathcal{Q}$, contradicting Claim 6.3.4.

Let ij be any edge. Then clearly $|d_a(i, q_1) - d_a(j, q_1)| \leq \frac{1}{2}a_i + \frac{1}{2}a_j$ and similarly $|d_a(i, q_2) - d_a(j, q_2)| \leq \frac{1}{2}a_i + \frac{1}{2}a_j$. We know by Claim 6.3.9 that at least one of these inequalities holds with equality. Using Claim 6.3.8, we see that there will be three types of edges:

(i) $|d_a(i, q_1) - d_a(j, q_1)| = \frac{1}{2}a_i + \frac{1}{2}a_j$, but $|d_a(i, q_2) - d_a(j, q_2)| < \frac{1}{2}a_i + \frac{1}{2}a_j$; in this case ij lies on one of the paths in $\mathcal{Q}$.

(ii) $|d_a(i, q_2) - d_a(j, q_2)| = \frac{1}{2}a_i + \frac{1}{2}a_j$, but $|d_a(i, q_1) - d_a(j, q_1)| < \frac{1}{2}a_i + \frac{1}{2}a_j$; in this case ij lies on one of the paths in $\mathcal{P}$.

(iii) $|d_a(i, q_2) - d_a(j, q_2)| = |d_a(i, q_1) - d_a(j, q_1)| = \frac{1}{2}a_i + \frac{1}{2}a_j$, and no path in $\mathcal{P}$ or $\mathcal{Q}$ goes through ij.

For each node i, consider the point $p_i = (d_a(q_1, i), d_a(q_2, i))$ in the plane, and let S_i denote the square with center p_i and side a_i.

Claim 6.3.10 *The squares S_i ($i \in V$) have no interior point in common.*

Indeed, for S_i and S_j to intersect, we would need $|d_a(i,q_1) - d_a(j,q_1)| < \frac{1}{2}a_i + \frac{1}{2}a_j$ and $|d_a(i,q_2) - d_a(j,q_2)| < \frac{1}{2}a_i + \frac{1}{2}a_j$, which is impossible.

Claim 6.3.11 *If $ij \in E$, then S_i and S_j have a boundary point in common.*

This is trivial in all three alternatives (i)–(iii) above: in cases (i) and (ii), the two squares have a (horizontal, resp. vertical) common boundary segment, in case (iii), they have a common vertex.

It follows that the four squares S_{q_i} are attached from the outside to the four edges of a rectangle R, and all the other squares S_i are contained in R. We show that they tile the rectangle.

We can redraw G in the plane so that node i is at position p_i, and every edge is a straight segment. We know from Claims 6.3.10 and 6.3.11 that every edge ij will be covered by the squares S_i and S_j.

Every point of R is contained in a finite face F of G, which is a triangle abc. The squares S_a, S_b and S_c are centered at the nodes, and each edge is covered by the two squares centered at its endpoints. Elementary geometry implies that these three squares cover the whole triangle.

Finally, we show that an appropriately resolved tangency graph of the squares S_i is equal to G. By the above, it contains G (where for edges of type (iii), the 4-corner is resolved so as to get the edge of G). But since G is a triangulation of the quadrilateral $q_1q_2q_3q_4$, the only additional edges of the tangency graph could be q_1q_3 or q_2q_4, which are not edges of the tangency graph.

Thus G is a subgraph of the (resolved) tangency graph of the squares. Since G is a triangulation, it cannot be a proper subgraph, so it is the tangency graph. □

Exercise 6.7 Prove that if G is a resolved tangency graph of a square tiling of a rectangle, then every triangle in G is a face.

Exercise 6.8 Construct a resolved tangency graph of a square tiling of a rectangle, which contains a quadrilateral with a further node in its interior.

Chapter 7

Analytic functions on graphs

7.1 Circulations and homology

Let S be a closed orientable compact surface, and consider a *map* on S, i.e., a graph $G = (V, E)$ embedded in S so that each face is a disc. We can describe the map as a triple $G = (V, E, \mathcal{F})$, where V is the set of nodes, E is the set of edges, and $\mathcal{F}$ is the set of faces of G. We fix a reference orientation of G; then each edge $e \in E$ has a *tail* $t(e) \in V$, a *head* $h(e) \in V$, a *right shore* $r(e) \in \mathcal{F}$, and a *left shore* $l(e) \in \mathcal{F}$.

The embedding of G defines a *dual map* G^*. Combinatorially, we can think of G^* as the triple $(\mathcal{F}, E, V)$, where the meaning of "node" and "face", "head" and "right shore", and "tail" and "left shore" is interchanged.

Let G be a finite graph with a reference orientation. For each node v, let $\delta v \in \mathbb{R}^E$ denote the coboundary of v:

$$(\delta v)_e = \begin{cases} 1, & \text{if } h(e) = v, \\ -1, & \text{if } t(e) = v, \\ 0, & \text{otherwise.} \end{cases}$$

Thus $|\delta v|^2 = d_v$ is the degree of v.

For every face $F \in \mathcal{F}$, we denote by $\partial F \in \mathbb{R}^E$ the boundary of F:

$$(\partial F)_e = \begin{cases} 1, & \text{if } r(e) = F, \\ -1, & \text{if } l(e) = F, \\ 0, & \text{otherwise.} \end{cases}$$

Then $d_F = |\partial F|^2$ is the length of the cycle bounding F.

A vector $\varphi \in \mathbb{R}^E$ is a *circulation* if

$$\varphi \cdot \delta v = \sum_{e:\ h(e)=v} \varphi(e) - \sum_{e:\ t(e)=v} \varphi(e) = 0 \qquad (\forall v \in V).$$

Each vector ∂F is a circulation; circulations that are linear combinations of vectors ∂F are called *null-homologous*. Two circulations φ and φ' are called *homologous* if $\varphi - \varphi'$ is null-homologous.

A vector $\varphi \in \mathbb{R}^E$ is *rotation-free* if for every face $F \in \mathcal{F}$, we have

$$\varphi \cdot \partial F = 0.$$

This is equivalent to saying that φ is a circulation on the dual map G^*.

Rotation-free circulations can be considered as *discrete holomorphic 1-forms* and they are related to *analytic functions*. These functions were introduced for the case of the square grid a long time ago [71, 54, 55]. For the case of a general planar graph, the notion is implicit in [35]. For a detailed treatment see [148].

To explain the connection, let φ be a rotation-free circulation on a graph G embedded in a surface. The φ can be thought of as a discrete analogue of a holomorphic 1-form. Consider a planar piece of the surface. Then on the set $\mathcal{F}'$ of faces contained in this planar piece, we have a function $\sigma : \mathcal{F}' \to \mathbb{R}$ such that $\partial\sigma = \varphi$, i.e., $\varphi(e) = \sigma(r(e)) - \sigma(l(e))$ for every edge e. Similarly, we have a function $\pi : V' \to \mathbb{R}$ (where V' is the set of nodes in this planar piece), such that $\delta\pi = \varphi$, i.e., $\varphi(e) = \pi(t(e)) - \pi(h(e))$ for every edge e. We can think of π and σ as the real and imaginary parts of a (discrete) analytic function, which satisfy

$$\delta\pi = \partial\sigma = \varphi, \tag{7.1}$$

which can be thought of as a discrete analogue of the Cauchy–Riemann equations.

Thus we have the two orthogonal linear subspaces: $\mathcal{A} \subseteq \mathbb{R}^E$ generated by the vectors δv ($v \in V$) and $\mathcal{B} \subseteq \mathbb{R}^E$ generated by the vectors ∂F ($F \in \mathcal{F}$). Vectors in $\mathcal{B}$ are 0-homologous circulations. The orthogonal complement $\mathcal{A}^\perp$ is the space of all circulations, and $\mathcal{B}^\perp$ is the space of circulations on the dual graph. The intersection $\mathcal{C} = \mathcal{A}^\perp \cap \mathcal{B}^\perp$ is the space of rotation-free circulations. So $\mathbb{R}^E = \mathcal{A} \oplus \mathcal{B} \oplus \mathcal{C}$. From this picture we conclude the following.

Lemma 7.1.1 *Every circulation is homologous to a unique rotation-free circulation.*

It also follows that $\mathcal{C}$ is isomorphic to the first homology group of S (over the reals), and hence we get the following:

Theorem 7.1.2 *The dimension of the space $\mathcal{C}$ of rotation-free circulations is $2g$.*

7.2 Discrete holomorphic forms from harmonic functions

We can use harmonic functions to give a more explicit description of rotation-free circulations in a special case. For any edge e of G, let η_e be the orthogonal projection of $\mathbf{1}_e$ onto $\mathcal{C}$.

Lemma 7.2.1 *Let $a, b \in E$ be two edges of G. Then $(\eta_a)_b$ is given by*

$$(\pi_b)_{h(a)} - (\pi_b)_{t(a)} + (\pi_b^*)_{r(a)} - (\pi_b^*)_{l(a)} + 1,$$

if $a = b$, and by

$$(\pi_b)_{h(a)} - (\pi_b)_{t(a)} + (\pi_b^*)_{r(a)} - (\pi_b^*)_{l(a)},$$

if $a \neq b$.

Proof. Let x_1, x_2 and x_3 be the projections of $\mathbf{1}_b$ on the linear subspaces $\mathcal{A}$, $\mathcal{B}$ and $\mathcal{C}$, respectively. The vector x_1 can be expressed as a linear combination of the vectors δv ($v \in V$), which means that there is a vector $y \in \mathbb{R}^V$ so that $x_1 = My$. Similarly, we can write $x_2 = Nz$. Together with $x = x_3$, these vectors satisfy the following system of linear equations:

$$\begin{cases} x + My + Nz = \mathbf{1}_b \\ M^\mathsf{T} x = 0 \\ N^\mathsf{T} x = 0 \end{cases} \tag{7.2}$$

Multiplying the first m equations by the matrix M^T, and using the second equation and the fact that $M^\mathsf{T} N = 0$, we get

$$M^\mathsf{T} M y = M^\mathsf{T} \mathbf{1}_b, \tag{7.3}$$

and similarly,

$$N^\mathsf{T} N z = N^\mathsf{T} \mathbf{1}_b. \tag{7.4}$$

Here $M^\mathsf{T} M$ is the Laplacian of G and $N^\mathsf{T} N$ is the Laplacian of G^*, and so (7.3) implies that $y = \pi_b + c\mathbf{1}$ for some scalar c. Similarly, $z = \pi_b^* + c^*\mathbf{1}$ for some scalar c'. Thus

$$\begin{aligned} x &= \mathbf{1}_b - M^\mathsf{T}(\pi_b + c\mathbf{1}) - N^\mathsf{T}(\pi_b^* + c^*\mathbf{1}) \\ &= \mathbf{1}_b - M^\mathsf{T}\pi_b - N^\mathsf{T}\pi_b^*, \end{aligned}$$

which is just the formula in the lemma, written in matrix form. □

The case $a = b$ of the previous formula has the following formulation:

Corollary 7.2.2 *For an edge a of a map G, let R_a denote the effective resistance between the endpoints of a, and let R_a^* denote the effective resistance of the dual map between the endpoints of the edge dual to a. Then*

$$(\eta_a)_a = 1 - R_a - R_a^*.$$

7.3 Operations

7.3.1 Integration

Let f and g be two functions on the nodes of a discrete weighted map in the plane. Integration is easiest to define along a path $P = (v_0, v_1, \ldots, v_k)$ in the diamond graph $G_\diamond$ (this has the advantage that it is symmetric with respect to G and G^*). We define

$$\int_P f\, dg = \sum_{i=0}^{k-1} \frac{1}{2}(f(v_{i+1}) + f(v_i))(g(v_{i+1}) - g(v_i)).$$

The nice fact about this integral is that for analytic functions, it is independent of the path P, depends on the endpoints only. More precisely, let P and P' be two paths on $G_\diamond$ with the same beginning node and endnode. Then

$$\int_P f\, dg = \int_{P'} f\, dg. \tag{7.5}$$

This is equivalent to saying that

$$\int_P f\, dg = 0 \tag{7.6}$$

if P is a closed path. It suffices to verify this for the boundary of a face of $G_\diamond$, which only takes a straightforward computation. It follows that we can write

$$\int_u^v f\, dg$$

as long as the homotopy type of the path from u to v is determined (or understood).

Similarly, it is also easy to check the rule of integration by parts: If P is a path connecting $u, v \in V \cup V^*$, then

$$\int_P f\, dg = f(v)g(v) - f(u)g(u) - \int_P g\, df. \tag{7.7}$$

Let P be a closed path in $G_\diamond$ that bounds a disk D. Let f be an analytic function and g an arbitrary function. Define $\hat{g}(e) = g(h_e) - g(t_e) - i(g(l_e) - g(r_e))$ (the "analycity defect" of g on edge e. Then it is not hard to verify the following generalization of (7.6):

$$\int_P f\, dg = \sum_{e \subset D} (f(h_e) - f(t_e))\hat{g}(e). \tag{7.8}$$

This can be viewed as a discrete version of the Residue Theorem. For further versions, see [148].

Take two analytic function f and g, and construct the polygons $f(u)P_u$ (multiplication by the complex number $f(u)$ corresponds to blowing up and rotating) as in section 7.5.3. The resulting polygons will not meet at the appropriate vertices any more, but we can try to translate them so that they do. Now equation (7.6) tells us that we can do that (Figure 7.1(b)). Conversely, every "deformation" of the picture such that the polygons P_u remain similar to themselves defines an analytic function on G.

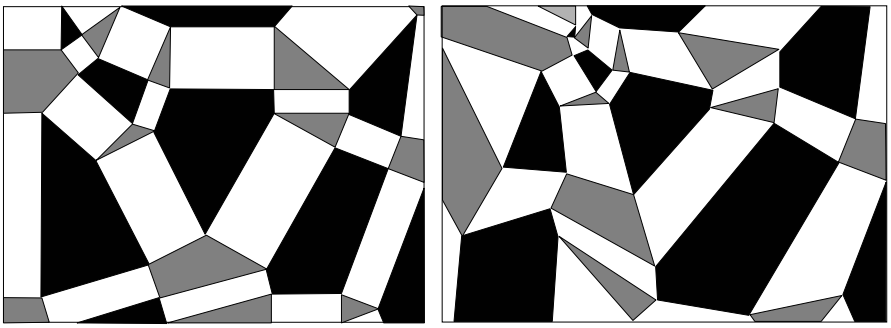

Figure 7.1: The deformation of the touching polygon representation given by another analytic function.

7.3.2 Critical analytic functions

These have been the good news. Now the bad part: for a fixed starting node u, the function

$$F(v) = \int_u^v f\, dg$$

is uniquely determined, but it is *not* analytic in general. In fact, a simple computation shows that for any edge e,

$$\begin{aligned}\hat{F}(e) &= \frac{F(h_e) - F(t_e)}{\ell_e} - i\frac{F(l_e) - F(r_e)}{\ell_{e^*}} \\ &= i\frac{f(h_e) - f(t_e)}{\ell_e}\Big[g(t_e) + g(h_e) - g(r_e) - g(l_e)\Big]. \end{aligned} \tag{7.9}$$

So it would be nice to have an analytic function g such that the factor in brackets in (7.9) is 0 for every edge:

$$g(t_e) + g(h_e) = g(r_e) + g(l_e) \tag{7.10}$$

Let us call such an analytic function *critical*. What we found above is that $\int_u^v f\, dg$ is an analytic function of v for every analytic function f if and only if g is critical.

This notion was introduced in a somewhat different setting by Duffin [55] under the name of *rhombic lattice*. Mercat [148] defined *critical maps*: these are maps which admit a critical analytic function.

Geometrically, this condition means the following. Consider the function g as a mapping of $G \cup G^*$ into the complex plane $\mathbb{C}$. This defines embeddings of G, G^* and $G\diamondsuit$ in the plane with following (equivalent) properties:

(a) The faces of $G^\diamondsuit$ are rhomboids.

(b) Every edge of $G_\diamondsuit$ has the same length.

(c) Every face of G is inscribed in a unit circle.

(d) Every face of G^* is inscribed in a unit circle.

Criticality can be expressed in terms of holomorphic forms as well. Let φ be a (complex valued) holomorphic form on a weighted map G. We say that φ is *critical* if the following condition holds: Let $e = xy$ and $f = yz$ be two edges of G bounding a corner at y, with (say) directed so that the corner is on their left, then

$$\ell_e \varphi(e) + \ell_f \varphi(f) = \ell_{e^*} \varphi(e^*) - \ell_{f^*} \varphi(f^*). \tag{7.11}$$

Note that both f and f^* are directed into h_f, which explains the negative sign on the right hand side. To digest this condition, consider a plane piece of the map and a primitive function g of ψ. Then (7.11) means that

$$g(y') - g(y) = g(q) - g(q'),$$

which we can rewrite in the following form:

$$g(x) + g(y) - g(p) - g(q) = g(x) + g(y') - g(p) - g(q').$$

This means that $g(h_e) + g(t_e) - g(l_e) - g(r_e)$ is the same for every edge e, and since we are free to add a constant to the value of g at every node in V^* (say), we can choose the primitive function g so that g is critical.

Whether or not a weighted map in the plane has a critical holomorphic form depends on the weighting. Which maps can be weighted this way? Kenyon and Schlenker [113] answer this question. Consider any face F_0 of the diamond graph $G_\diamond$, and a face F_1 incident with it. This is a quadrilateral, so there is a well-defined face F_2 so that F_0 and F_2 are attached to F_1 along opposite edges. Repeating this, we get a sequence of faces $(F_0, F_1, F_2 \dots)$. Using the face attached to F_0 on the opposite side to F_1, we can extend this to a two-way infinite sequence $(\dots, F_{-1}, F_0, F_1, \dots)$. We call such a sequence a *track*.

Theorem 7.3.1 *A planar map has a rhomboidal embedding in the plane if and only if every track consists of different faces and any two tracks have at most one face in common.*

7.3.3 Polynomials, exponentials and approximation

Once we can integrate, we can define polynomials. More exactly, let G be a map in the plane, and let us select any node to be called 0. Let Z denote a critical analytic function on G such that $Z(0) = 0$. Then we have

$$\int_0^x 1 \, dZ = Z(x).$$

Now we can define higher powers of Z by repeated integration:

$$Z^{:n:}(x) = n \int_0^x Z^{:n-1:} dZ.$$

We can define a *discrete polynomial* of degree n as any linear combination of $1, Z, Z^{:2:}, \ldots, Z^{:n:}$. The powers of Z of course depend on the choice of the origin, and the formulas describing how it is transformed by shifting the origin are more complicated than in the classical case. However, the space of polynomials of degree see is invariant under shifting the origin [149]).

Further, we can define the exponential function $exp(x)$ as a discrete analytic function $\mathrm{Exp}(x)$ on $V \cap V^*$ satisfying

$$d\mathrm{Exp}(x) = \mathrm{Exp}(x)dZ.$$

More generally, it is worth while to define a 2-variable function $\mathrm{Exp}(x, \lambda)$ as the solution of the difference equation

$$d\mathrm{Exp}(\lambda, x) = \lambda \mathrm{Exp}(\lambda, x)dZ.$$

It can be shown that there is a unique such function, and there are various more explicit formulas, including

$$\mathrm{Exp}(\lambda, x) = \sum_{n=0}^{\infty} \frac{Z^{:n:}}{n!},$$

(at least as long as the series on the right hand side is absolute convergent).

Mercat [149, 150] uses these tools to show that exponentials form a basis for all discrete analytic functions, and to generalize results of Duffin [54] about approximability of analytic functions by discrete analytic functions.

7.4 Nondegeneracy properties of rotation-free circulations

We state and prove two key properties of rotation-free circulations: one, that the projection of a basis vector to the space of rotation-free circulations is non-zero, and two, that rotation-free circulations are spread out essentially over the whole graph in the sense that every connected piece of the graph where a non-zero rotation-free circulation vanishes can be isolated from the rest by a small number of points. These results were proved in [24, 25].

We start with a simple lemma about maps. For every face F, let a_F denote the number of times the orientation changes if we move along the the boundary of F. For every node v, let b_v denote the number of times the orientation changes in their cyclic order as they emanate from v.

Lemma 7.4.1 *Let $G = (V, E, \mathcal{F})$ be any digraph embedded on an orientable surface S of genus g. Then*

$$\sum_{F \in \mathcal{F}} (a_F - 2) + \sum_{v \in V} (b_v - 2) = 4g - 4.$$

Proof. Clearly

$$\sum_F a_F = \sum_v (d_v - b_v),$$

and so, using Euler's formula,

$$\sum_F a_F + \sum_v b_v = \sum_v d_v = 2m = 2n + 2f + 4g - 4.$$

Rearranging and dividing by 2, we get the equality in the lemma. □

If G is planar, then $R_a + R_a^* = 1$, a well known fact. For any other underlying surface, we have

$$(\eta_a)_a = \eta_a \cdot \mathbf{1}_a = |\eta_a|^2$$

(since η_a is a projection of $\mathbf{1}_a$), and so it follows that $R_a + R_a^* \le 1$. It follows from theorem 7.4.2 below that strict inequality holds here.

If $g = 0$, then there is no nonzero rotation-free circulation by Theorem 7.1.2, and hence $\eta_e = 0$ for every edge e. But for $g > 0$ we have:

Theorem 7.4.2 *If $g > 0$, then $\eta_e \neq 0$ for every edge e.*

Proof. Suppose that $\eta_e = 0$ for some edge e. Then by Lemma 7.2.1, there are vectors $\pi = \pi(e) \in \mathbb{R}^V$ and $\pi^* = \pi^*(e) \in \mathbb{R}^F$ such that

$$\pi_{h(a)} - \pi_{t(a)} = \pi^*_{r(a)} - \pi^*_{l(a)} \tag{7.12}$$

for every edge $a \neq e$, but

$$\pi_{h(e)} - \pi_{t(e)} = 1 + \pi^*_{r(e)} - \pi^*_{l(a)}. \tag{7.13}$$

We define a convenient orientation of G. Let $E(G) = E_1 \cup E_2$, where E_1 consists of edges a with $\varphi(h(a)) \not/ \varphi(t(a))$, and E_2 is the rest. Every edge $a \in E_1$ is oriented so that $\pi h(a) > \pi t(a)$. Consider any connected component C of the subgraph formed by edges in E_2. Let $u_1, \ldots, u_k$ be the nodes of C that are incident with edges in E_1. Add a new node v to C and connect it to $u_1, \ldots, u_k$ to get a graph C'. Clearly C' is 2-connected, so it has an acyclic orientation such that every node is contained in a path from v to u_1. The corresponding orientation of C is acyclic and every has the property that it has no source or sink other than possibly $u_1, \ldots, u_k$.

Carrying this out for every connected component of G', we get an orientation of G. We claim this orientation is acyclic. Indeed, if we had a directed cycle, then walking around it π would never decrease, so it would have to stay constant. But then all edges of the cycle would belong to E_2, contradicting the way these edges were oriented.

We also claim this orientation has only one source and one sink. Indeed, if a node $v \neq h(e), t(e)$ is incident with an edge of E_1, then it has at least one edge of E_1 entering it and at least one leaving it. If v is not incident with any edge of E_1, then it is an internal node of a component C, and so it is not a source or sink by the construction of the orientation of C.

Take the union of G and the dual graph G^*. This gives a graph H embedded in S. Clearly H inherits an orientation from G and from the corresponding orientation of G^*.

We are going to apply Lemma 7.4.1. Every face of H will $a_F = 2$ (this just follows from the way how the orientation of G^* was defined). Those nodes of H which arise as the intersection of an edge of G with an edge of G^* will have $b_v = 2$.

Consider a node v of G. If $v = h(a)$ then clearly all edges are directed toward v, so $b_{h(a)} = 0$. Similarly, we have $b_{t(v)} = 0$. We claim that $b_v = 2$ for every other node. Since obviously v is not a source or a sink, we have $b_v \geq 2$. Suppose that $b_v > 2$. Then we have for edges e_1, e_2, e_3, e_4 incident with v in this cyclic order, so that e_1 and e_2 form a corner of a face F, e_3 and e_4 form a corner of a face F', $h(e_1) = h(e_3) = v$ and $t(e_2) = t(e_3) = v$.

Consider π^* of the faces incident with v. We may assume that $\pi^*(F) \leq \pi^*(F')$. From the orientation of the edges e_1 and e_2 it follows that $\pi^*(F)$ is larger than π^* of its neighbors. Let $\mathcal{F}$ be the union of all faces F'' with $\pi^*(F'') \geq \pi^*(F)$. The boundary of $\mathcal{F}$ is an eulerian subgraph, and so it can be decomposed into edge-disjoint cycles $D_1, \ldots, D_t$. Since the boundary goes through v twice (once along e_1 and e_2, once along two other edges with the corner of F' on the left hand side), we have $t \geq 2$, and so one of these cycles, say D_1, does not contain e. But then by the definition of the orientation and by (7.12), D_1 is a directed cycle, which is a contradiction.

A similar argument shows that if v is a node corresponding to a face not incident with e, then $b_v = 0$; while if v comes from $r(e)$ or from $l(e)$, then $b_v = 2$.

So substituting in Lemma 7.4.1, only two terms on the left hand side will be non-zero, yielding $-4 = 4g - 4$, or $g = 0$. □

Corollary 7.4.3 *If $g > 0$, then there exists a rotation-free circulation that does not vanish on any edge.*

Corollary 7.4.4 *If $g > 0$, then for every edge e, $(\eta_e)_e \geq n^{-n} f^{-f}$.*

Indeed, combining with the remark after Corollary 7.2.2, we see that $(\eta_e)_e > 0$ if $g > 0$. But $(\eta_e)_e = 1 - R_e - R_e^*$ is a rational number, and it is easy to see that its denominator is not larger than $n^n f^f$.

Theorem 7.4.5 *Let G be a graph embedded in an orientable surface S of genus $g > 0$ so that all faces are discs. Let φ be a non-zero rotation-free circulation on G and let G' be the*

subgraph of G on which φ does not vanish. Suppose that φ vanishes on all edges incident with a connected subgraph U of G. Then U can be separated from G' by at most $4g-3$ points.

The assumption that the connectivity between U and the rest of the graph must be linear in g is sharp in the following sense. Suppose X is a connected induced subgraph of G separated from the rest of G by $\leq 2g$ nodes, and suppose (for simplicity) that X is embedded in a subset of S that is topologically a disc. Contract X to a single point x, and erase the resulting multiplicities of edges. We get a graph G' still embedded in S so that each face is a disc. Thus this graph has a $(2g)$-dimensional space of circulations, and hence there is a non-zero rotation-free circulation ψ vanishing on $2g-1$ of the edges incident with x. Since this is a circulation, it must vanish on all the edges incident with x. Uncontracting X, and extending ψ with 0-s to the edges of X, it is not hard to check that we get a rotation-free circulation.

Proof. Let W be the connected component of $G \setminus V(G')$ containing U, and let Y denote the set of nodes in $V(G) \setminus V(W)$ adjacent to W.

Consider an edge e with $\varphi(e) = 0$. If e is not a loop, then we can contract e and get a map on the same surface with a rotation-free flow on it. If $G - e$ is still a map, i.e., every face is a disc, then φ is a rotation-free flow on it. If $G - e$ is not a map, then both sides of e must be the same face. So we can eliminate edges with $\varphi(e) = 0$ unless $h(e) = t(e)$ and $r(e) = l(e)$ (we call these edges strange loops). In this latter case, we can change $\varphi(e)$ to any non-zero value and still have a rotation-free flow.

Applying this reduction procedure, we may assume that $W = \{w\}$ consists of a single node, and the only edges with $\varphi = 0$ are the edges between w and Y, or between two nodes of Y. We cannot try to contract edges between nodes in Y (we don't want to reduce the size of Y), but we can try to delete them; if this does not work, then every such edge must have $r(e) = l(e)$.

Also, if more than one edge remains between w and a node $y \in Y$, then each of them has $r(e) = l(e)$ (else, one of them could be deleted). Note that we may have some strange loops attached at w. Let D be the number of edges between w and Y.

Re-orient each edge with $\varphi \neq 0$ in the direction of the flow φ, and orient the edges between w and Y alternatingly in an out from w. Orient the edges with $\varphi = 0$ between two nodes of Y arbitrarily. We get a digraph G_1.

It is easy to check that G_1 has no sources or sinks, so $b_v \geq 2$ for every node v, and of course $b_w \geq |Y| - 1$. Furthermore, every face either has an edge with $\varphi > 0$ on its boundary, or an edge with $r(e) = l(e)$. If a face has at least one edge with $\varphi > 0$, then it cannot be bounded by a directed cycle, since φ would add up to a positive number on its boundary. If a face boundary goes through an edge with $r(e) = l(e)$, then it goes through it twice in different directions, so again it is not directed. So we have $a_F \geq 2$ for every face.

Substituting in Lemma 7.4.1, we get that $|Y| - 1 \leq 4g - 4$, or $|Y| \leq d_w \leq 4g - 3$. Since Y separates U from G', this proves the theorem. □

7.4.1 Rubber bands and analytic functions

We apply the method to toroidal graphs.

Let G be a toroidal map. We consider the torus as $\mathbb{R}^2/\mathbb{Z}^2$, endowed with the metric coming from the euclidean metric on $\mathbb{R}^2$. Let us replace each edge by a rubber band, and let the system find its equilibrium. Topology prevents the map from collapsing to a single point. In mathematical terms, we are minimizing

$$\sum_{ij \in E(G)} \ell(ij)^2, \tag{7.14}$$

where the length $\ell(ij)$ of the edge ij is measured in the given metric, and we are minimizing over all continuous mappings of the graph into the torus homomorphic to the original embedding.

It is not hard to see that the minimum is attained, and the minimizing mapping is unique up to isometries of the torus. We call it the *rubber band mapping*. Clearly, the edges are mapped onto geodesic curves. A nontrivial fact is that if G is a simple 3-connected toroidal map, then the rubber band mapping is an embedding.

We can lift this optimizing embedding to the universal cover space, to get a planar map which is doubly periodic, and the edges are straight line segments. Moreover, *every node is at the center of gravity of its neighbors.* This follows simply from the minimality of (7.14). This means that both coordinate functions are harmonic and periodic, and so their coboundaries are rotation-free circulations on the original graph. Since the dimension of the space $\mathcal{C}$ of rotation-free circulations on a toroidal map is 2, this construction gives us the whole space $\mathcal{C}$.

This last remark also implies that if G is a simple 3-connected toroidal map, then selecting any basis φ_1, φ_2 in $\mathcal{C}$, the primitive functions of φ_1 and φ_2 give a doubly periodic straight-line embedding of the universal cover map in the plane.

7.5 Geometric representations and discrete analytic functions

7.5.1 Square tilings and analytic functions

A beautiful connection between square tilings and rotation-free flows was described in the classic paper of Brooks, Smith, Stone and Tutte [35]. For our purposes, periodic tilings of the whole plane are more convenient to use.

Consider tiling of the plane with squares, whose sides are parallel to the coordinate axes. Assume that the tiling is discrete, i.e., every bounded region contains only a finite number of squares. We associate a map in the plane with this tiling as follows. As in Section 6.3, we represent each maximal horizontal segment composed of edges of the squares by a single node at the midpoint of the segment. Each square "connects" two horizontal segments, and we can represent it by an edge connecting the two corresponding nodes, directed top-down. (We handle points where four squares meet just as in the finite case.) We get an (infinite) directed graph G (Figure 7.2).

It is not hard to see that G is planar. If we assign the edge length of each square to the corresponding edge, we get a circulation: If a node v represents a segment I, then the total flow into v is the sum of edge length of squares attached to I from the top, while the total flow out of v is the sum of edge length of squares attached to I from the bottom. Both of these sums are equal to the length of I (let's ignore the possibility that I is infinite for the moment). Furthermore, since the edge-length of a square is also the difference between the y-coordinates of its upper and lower edges, this flow is rotation-free.

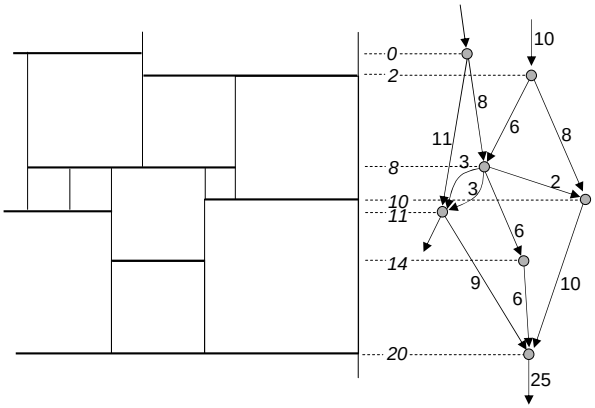

Figure 7.2: The Brooks–Smith–Stone–Tutte construction

Now suppose that the tiling is double periodic with period vectors $a, b \in \mathbb{R}^2$ (i.e., we consider a square tiling of the torus). Then so will be the graph G, and so factoring out the period, we get a map on the torus. Since the tiling is discrete, we get a finite graph. This also fixes the problem with the infinite segment I: it will become a closed curve on the torus, and so we can argue with its length on the torus, which is finite now. The flow we constructed will also be periodic, so we get a rotation-free circulation on the torus.

We can repeat the same construction using the vertical edges of the squares. It is not hard to see this gives the dual graph, with the dual rotation-free circulation on it.

This construction can be reversed. Take a toroidal map G^* and any rotation-free circulation on it. Then this circulation can be obtained from a doubly periodic tiling of the plane

by squares, where the edge-length of a square is the flow through the corresponding edge. (We suppress details.)

If an edge has 0 flow, then the corresponding square will degenerate to s single point. Luckily, we know (Corollary 7.4.3) that for a simple 3-connected toroidal map, there is always a nowhere-zero rotation-free circulation, so these graphs can be represented by a square tiling with no degenerate squares.

7.5.2 Circle packings and analytic functions

We can also relate to discrete holomorphic forms are circle representations.

It is again best to go to the universal cover map $\hat{G}$. Then the result says that for every 3-connected toroidal graph G we can construct two (infinite, but discrete) families $\mathcal{F}$ and $\mathcal{F}^*$ of circles in the plane so that they are double periodic modulo a lattice $L = \mathbb{Z}a + \mathbb{Z}b$, $\mathcal{F}$ (mod L) corresponds to the nodes of G, $\mathcal{F}^*$ (mod L) corresponds to the faces of G, and for ever edge e, there are two circles C, C' representing h_e and t_e, and two circles D and D' representing r_e and l_e so that C, C' are tangent at a point p, D, D' are tangent at the same point p, and C, D are orthogonal.

If we consider the centers the circles in $\mathcal{F}$ as nodes, and connect two centers by a straight line segment if the circles touch each other, then we get a straight line embedding of the universal cover map in the plane (appropriately periodic modulo L). Let $f(i)$ denote the point representing node i of the universal cover map. or of its dual.

To get a holomorphic form out of this representation, consider the plane as the complex plane, and define $\varphi(ij) = \rho(j) - \rho(i)$ for every edge of $\hat{G}$ or $\hat{G}^*$. Clearly φ is invariant under L, so it can be considered as a function on $E(G)$. By the orthogonality property of the circle representation, $\varphi(e)/\varphi(e^*)$ is a positive multiple of i. In other words,

$$\frac{\varphi(e)}{|\varphi(e)|} = i\frac{\varphi(e^*)}{|\varphi(e^*)|}$$

It follows that if we consider the map G with weights

$$\ell_e = |\varphi(e)|, \qquad \ell_{e^*} = |\varphi(e^*)|,$$

then φ is a discrete holomorphic form on this weighted map.

It would be nice to be able to turn this construction around, and construct a circle representation using discrete holomorphic forms.

7.5.3 Touching polygon representations

Kenyon's ideas in [112] give a another geometric interpretation of analytic functions. Let G be a map in the plane and let f be an analytic function on G. Let us assume that there is a straight-line embedding of G in the plane with convex faces, and similarly, a straight-line

embedding of G^* with convex faces. Let P_u denote the convex polygon representing the face of G (or G^*) corresponding to $u \in V^*$ (or $u \in V$)). Shrink each F_u from the point $f(u)$ by a factor of 2. Then we get a system of convex polygons where for every edge $uv \in G_\diamond$, the two polygons P_u and P_v share a vertex at the point $(f(u) + f(v))/2$ (Figure 7.1(a)). There are two kinds of polygons (corresponding to the nodes in V and V^*, respectively). It can be shown that the interiors of the polygons P_u will be disjoint (the point $f(u)$ is not necessarily in the interior of P_u). The white areas between the polygons correspond to the edges of G. They are rectangles, and the sides of the rectangle corresponding to edge e are $f(h_e) - f(t_e)$ and $f(l_e) - f(r_e)$.

7.6 Novikov's discrete analytic functions

A similar, but different theory of discrete analytic functions was developed by Dynnikov and Novikov [57].

7.7 Discrete analytic functions from circle packings

There is a book on this subject by Kenneth Stephenson [190].

Chapter 8

The Colin de Verdière Number

In 1990, Colin de Verdière [44] introduced a spectral invariant $\mu(G)$ of a graph G (for a survey, see [102]).

8.1 The definition

8.1.1 Motivation

Let G be a connected graph. We know that (by the Perron–Frobenius Theorem) the largest eigenvalue of the adjacency matrix has multiplicity 1. What about the second largest?

The eigenvalues of most graphs are all different, so the multiplicity of any of them gives any information about the graph in rare cases only. Therefore, it makes sense to try to maximize the multiplicity of the second largest eigenvalue by weighting the edges by positive numbers. The diagonal entries of the adjacency matrix don't carry any information, so we allow to put there any real numbers. The off-diagonal matrix entries that correspond to nonadjacent nodes remain 0.

There is a technical restriction, called the Strong Arnold Property, which excludes very degenerate choices of edgeweights and diagonal entries. We'll discuss this condition later.

If we maximize the multiplicity of the largest eigenvalue this way, we get the number of connected components of the graph. So for the second largest, we expect to get some parameter related to connectivity. However, the situation is more complex (and more interesting).

Since we don't put any restriction on the diagonal entries, we may add a constant to the diagonal entries to shift the spectrum so that the second largest eigenvalue becomes 0, without changing its multiplicity. The multiplicity in question is then the corank of the matrix (the dimension of its nullspace).

Finally, we multiply the matrix by -1, to follow convention.

This discussion hopefully makes most of the formal definition in the next section clear.

8.1.2 Formal definition

For a connected graph $G = (V, E)$, we consider a matrix $M \in \mathbb{R}^{V \times V}$ with the following properties:

(M1) $M_{ij} \begin{cases} < 0, & \text{if } ij \in E, \\ = 0, & \text{if } ij \notin E,\ i \neq j; \end{cases}$

(M2) M has exactly one negative eigenvalue (of multiplicity 1).

(M3) [Strong Arnold Property] If X is a symmetric $n \times n$ matrix such that $X_{ij} = 0$ whenever $i = j$ or $ij \in E$, and $MX = 0$, then $X = 0$.

The *Colin de Verdiére number* $\mu(G)$ of the graph is defined as the maximum corank of such a matrix.

8.1.3 The Strong Arnold Property

Condition (M3) requires some explanation. Let $\mathbb{R}^{V^{[2]}}$ denote the space of all symmetric $V \times V$ matrices. Consider the manifold $\mathcal{R}_k$ of all matrices in $\mathbb{R}^{V^{[2]}}$ with rank at most $k := \mathrm{rk}(M)$, and the linear space $\mathcal{O}_G$ of all symmetric $n \times n$ matrices A in $\mathbb{R}^{V^{[2]}}$ such that $A_{ij} = 0$ for all $ij \in \overline{E}$.

We need the following lemma describing the tangent space of $\mathcal{R}_k$ and its orthogonal complement.

Lemma 8.1.1 *Let $M \in \mathcal{R}^k$, and let S denote the matrix of the orthogonal projection onto the nullspace of M.*

(a) *The tangent space $\mathcal{T}(M)$ of $\mathcal{R}_k$ at M consists of all matrices of the form $MU + U^{\mathsf{T}}M$, where U is a (not necessarily symmetric) $V \times V$ matrix.*

(b) *The normal space $\mathcal{N}(M)$ of $\mathcal{R}_k$ at M consists of all symmetric $V \times V$ matrices X such that $MX = 0$.*

Proof. (a) Suppose that $Y = MU + U^{\mathsf{T}}M$. Consider the family

$$M(t) = (I + tU)^{\mathsf{T}} M (I + tU).$$

Clearly $\mathrm{rk}(M(t)) \leq \mathrm{rk}(M) \leq k$ and equality holds if $|t|$ is small enough. Furthermore, $M'(0) = MU + U^{\mathsf{T}}M = Y$. Hence $Y \in \mathcal{T}(M)$.

Conversely, let $Y \in \mathcal{T}(M)$. Then there is a one-parameter differentiable family $M(t)$ of symmetric matrices, defined in a neighborhood of $t = 0$, so that $M(t) \in \mathcal{R}_k$, $M(0) = M$ and $M'(0) = Y$. Let $S(t)$ denote the matrix of the orthogonal projection onto the nullspace of $M(t)$, and set $S = S(0)$. By definition, we have $M(t)S(t) = 0$, and hence by differentiation,

we get $M'(0)S(0) + M(0)S'(0) = 0$, or $YS + MS'(0) = 0$. Multiplying by S from the left, we get $SYS = 0$. So we can write

$$Y = \frac{1}{2}((I-S)Y(I+S) + (I+S)Y(I-S)).$$

Notice that $I - S$ is the orthogonal projection onto the range of M, and hence we can write $I - S = MV$ with some matrix V. By transposition, we have $I - S = V^\mathsf{T} M$. Then

$$Y = \frac{1}{2}MVY(I+S) + \frac{1}{2}(I+S)YV^\mathsf{T}M = MU + U^\mathsf{T}M,$$

where $U = \frac{1}{2}VY(I+S)$.

(b) We have $X \in \mathcal{N}(M)$ iff $X \cdot Y = 0$ for every $Y \in TT(M)$. By (a), this means that $X \cdot (MU + U^\mathsf{T}M) = 0$ for every U. This can be written as

$$0 = \mathrm{Tr}(X(MU+U^\mathsf{T}M)) = \mathrm{Tr}(XMU)+\mathrm{Tr}(XU^\mathsf{T}M) = \mathrm{Tr}(XMU)+\mathrm{Tr}(MUX) = 2\mathrm{Tr}(XMU).$$

This holds for every U if and only if $XM = 0$. □

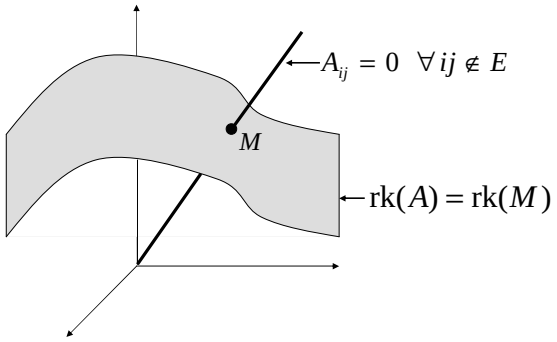

Figure 8.1: The Strong Arnold Property

Lemma 8.1.2 *For a connected graph G and a matrix M satisfying* (M1) *and* (M2)*, the following are equivalent.*

(a) (M3) *holds.*

(b) *The manifolds $\mathcal{R}_k$ and $\mathcal{O}_G$ intersect transversally at M.*

(c) *For every symmetric matrix B there exists a matrix $A \in \mathcal{O}_G$ such that $x^\mathsf{T}(A-B)x = 0$ for all vectors x in the nullspace of M.*

Proof. Let $\mathcal{T}_M$ denote the tangent space of $\mathcal{R}_k$ at M.

Then $X_{ij} = 0$ for all $ij \in E \cup \Delta$ means that X is orthogonal to $\mathcal{O}_G$. Moreover, using some elementary linear algebra and differential geometry one can easily show that $MX = 0$ is equivalent to saying that X is orthogonal to $\mathcal{T}_M$, the tangent space of $\mathcal{R}_k$ at M.

Hence condition (M3) is equivalent to requiring that $\mathcal{T}_M$ and $\mathcal{O}_G$ span the space of all symmetric $n \times n$ matrices. □

Exercise 8.1 Let $f(G)$ denote the maximum multiplicity or the largest eigenvalue of any matrix obtained from the adjacency matrix of a graph G, where 1's are replaced by arbitrary positive numbers and the diagonal entries are changed arbitrarily. Prove that $f(G)$ is the number of connected components of the graph G.

Exercise 8.2

8.2 Basic properties

Theorem 8.2.1 *The Colin de Verdière number is minor-monotone.*

Proof.

□

Theorem 8.2.2 *If $\mu(G) > 2$, then $\mu(G)$ is invariant under subdivision.*

Proof.

□

The following result was proved by Bacher and Colin de Verdière [19].

Theorem 8.2.3 *If $\mu(G) > 3$, then $\mu(G)$ is invariant under $\Delta - Y$ transformation.*

Proof.

□

8.3 Small values

Graphs with Colin de Verdière number up to 4 are characterized.

Theorem 8.3.1 *Let G be a connected graph.*

(a) *$\mu(G) \leq 1$ if and only if G is a path;*

(b) *$\mu(G) \leq 2$ if and only if G is outerplanar;*

(c) *$\mu(G) \leq 3$ if and only if G is planar;*

(d) *$\mu(G) \leq 4$ if and only if G is linklessly embedable.*

Statements (a)–(c) were proved by Colin de Verdière [43]; an elementary proof is due to Hein van der Holst [95]); (d) is due to Lovász and Schrijver [138].

Proof.

□

8.4 Nullspace representation

Every weighted adjacency matrix M with a d-dimensional nullspace gives rise to an embedding of the graph in d-space: we take a basis $v_1, \dots, v_d$ of the null space, write them down as column vectors. Each node i of the graph corresponds to a row u_i of the matrix obtained, which is a vector in $\mathbb{R}^d$, and so we get the *nullspace representation* of the graph. Note that this representation is uniquely determined up to a linear transformation of $\mathbb{R}^d$.

Saying this in a fancier way, each coordinate function is a linear functional on the nullspace, and so the nullspace representation is in fact a representation in the dual space of the nullspace.

The nullspace representation u satisfies the equations

$$\sum_{j \in V} M_{ij} u_j = 0 \qquad \text{for all } i \in V. \tag{8.1}$$

In the case when M is a Colin der Verdière matrix of the graph G, we have $M_{ij} < 0$ for every edge ij, and so we can write this as

$$\sum_{j \in N(i)} (-M_{ij}) u_j = M_{ii} u_i \qquad \text{for all } i \in V,$$

and so each vector u_i is in the cone spanned by its neighbors, or in the negative of this cone (depending on the sign of M_{ii}). So from this matrix we get a representation of the graph in $\mu(G)$-space.

8.4.1 A nodal theorem for Colin de Verdière matrices

Van der Holst [95] proved a key lemma about Colin de Verdière matrices, which was subsequently generalized [98, 139]. We state a generalization i a geometric form.

Lemma 8.4.1 *Let G be a connected graph with $\mu(G) = \mu$, let M be a Colin de Verdière matrix of G, and let u be a nullspace representation belonging to M. Let H be a closed halfspace in $\mathbb{R}^d$ containing the origin. Then the subgraph G' of G induced by nodes in the interior of H is nonempty, and either*

— G' is connected, or

— 0 is on the boundary of H, and there is a linear subspace $S \subseteq H$ of dimension $\mu - 2$ and $r \geq 3$ half-hyperplanes $T_1, \dots, T_r$ with boundary S, so that each node of G lies on $\cup_i T_i$, and

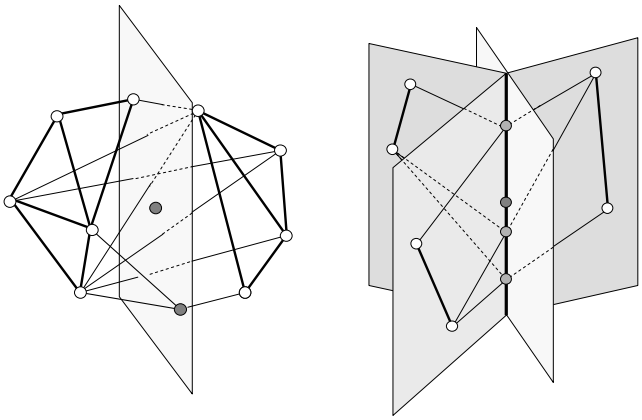

Figure 8.2: Two possibilities for the nullspace representation of a connected graph: for every hyperplane through the origin, either both sides are connected, or...

each edge of G is contained in one of the T_i. Furthermore, each node of G on S is connected to nodes on every T_i, or else only to nodes on S. The halfspace H contains at least two of the half-hyperplanes T_i.

(See Figure 8.2.)

Proof.

□

Corollary 8.4.2 *Let G be a connected graph with $\mu(G) = \mu$, let M be a Colin de Verdière matrix of G, and let u be a nullspace representation belonging to M.*

(a) (Van der Holst's Lemma) *Every $(\mu-1)$-dimensional subspace of $\mathbb{R}^\mu$ spanned by $\mu-1$ vectors u_i separates G into two non-empty connected subgraphs.*

(b) *Every $(\mu-1)$-dimensional subspace of $\mathbb{R}^\mu$ not containing any of the vectors u_i separates G into two non-empty connected subgraphs.*

(c) *Every open halfspace of $\mathbb{R}^\mu$ containing the origin contains a non-empty connected subgraph of G.*

Sometimes it is more useful to set this in algebraic terms.

Corollary 8.4.3 *(a) Let $x \in \ker M$ have minimal support. Then both $\mathrm{supp}^+(x)$ and $\mathrm{supp}^-(x)$ are nonempty and induce connected subgraphs of G.*

(b) *Let $x \in \ker M$ have $\mathrm{supp}(x) = V$. Then both $\mathrm{supp}^+(x)$ and $\mathrm{supp}^-(x)$ are nonempty and induce connected subgraphs of G.*

In fact, one can prove a stronger lemma in algebraic terms, which we state as an exercise.

Exercise 8.3 Let G be a connected graph with $\mu(G) = \mu$, let M be a Colin de Verdière matrix of G, and let $x \in \mathbb{R}^V$ be any vector with $Mx \leq 0$. Then $\text{supp}_+(x)$ is nonempty, and either

— $\text{supp}_+(x)$ spans a connected subgraph, or

— G has a cutset S with the following properties: Let $G_1, \ldots, G_r$ ($r \geq 2$) be the connected components of $G - S$. There exist non-zero, non-negative vectors $x_1, \ldots, x_r$ and y in $\mathbb{R}^V$ such that $\text{supp}(x_i) = V(G_i)$ and $\text{supp}(y) \subseteq S$, $Mx_1 = Mx_2 = \cdots = Mx_r = -y$, and x is a linear combination $x = \sum_i \alpha_i x_i$, where $\sum_i \alpha_i \geq 0$, and at least two α_i are positive.

8.4.2 Steinitz representations and Colin de Verdière matrices

For 3-connected planar graphs, the nullspace representation gives a Steinitz representation [139, 134]; in fact, Steinitz representations naturally correspond to Colin de Verdière matrices.

8.5 Gram representation

Every Colin de Verdiére matrix of a graph gives rise to another geometric representation of the graph, this time in the $(n - \mu(G) - 1)$-dimensional space [118]. This is related to the Koebe–Andre'ev representation if the complement of the graph is a maximal planar graph.

8.6 Related graph parameters

Van der Holst's lemma motivates this related graph parameter, which is again related to planarity and other geometric representations. This was studied by Van der Holst, Laurent and Schrijver [97].

Chapter 9

Orthogonal representations

9.1 Orthogonal representations: definition

Let $G = (V, E)$ be a simple graph. We will denote by $\overline{G} = (V, \overline{E})$ its complement.

An *orthogonal representation* of a graph $G = (V, E)$ in $\mathbb{R}^d$ assigns to each $i \in V$ a nonzero vector $u_i \in \mathbb{R}^d$ such that $u_i^{\mathsf{T}} u_j = 0$ whenever $ij \in \overline{E}$. An *orthonormal representation* is an orthogonal representation in which all the representing vectors have unit length. Clearly we can always scale the vectors forming an orthogonal representation this way, and usually this does not change any substantial feature of the problem.

Sometimes we specify a further unit vector $c \in \mathbb{R}^d$, and call it the *handle* of the orthogonal representation. (For the origin of the name, see Example 9.1.4 below).

Note that we did not insist that different nodes are mapped onto different vectors, nor that adjacent nodes are mapped on non-orthogonal vectors. If these conditions also hold, we call the orthogonal representation *faithful*.

Example 9.1.1 Every graph has a trivial orthogonal (in fact, orthonormal) representation in $\mathbb{R}^V$, in which node i is represented by the standard basis vector e_i.

Of course, we are interested in "nontrivial" orthogonal representations, which are more "economical" than the trivial one.

Example 9.1.2 Figure 9.1 below shows that for the graph obtained by adding a diagonal to the pentagon a simple orthogonal representation in 2 dimensions can be constructed.

This example can be generalized as follows.

Example 9.1.3 Let $k = \overline{\chi}(G)$, and let $\{B_1, \ldots, B_k\}$ be a family of disjoint complete subgraphs covering all the nodes. Let $\{e_1, \ldots, e_k\}$ be the standard basis of $\mathbb{R}^k$. Then mapping every $x \in B_i$ to e_i is an orthonormal representation.

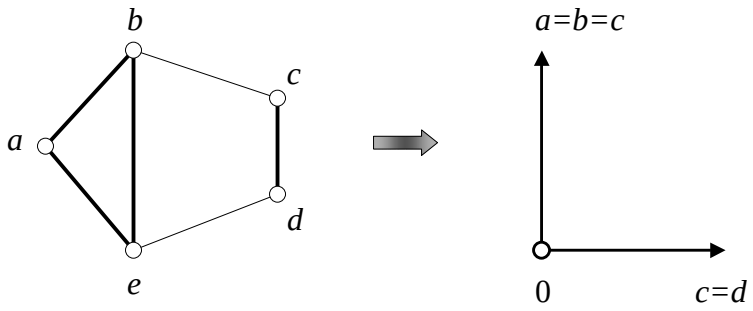

Figure 9.1: An (almost) trivial orthogonal representation

A more "geometric" orthogonal representation is described by the following example.

Example 9.1.4 Consider an "umbrella" in $\mathbb{R}^3$ with a unit vector c as its handle, and 5 ribs of unit length (Figure 9.2). Open it up to the point when non-consecutive ribs are orthogonal. This way we get 5 unit vectors u_0, u_1, u_2, u_3, u_4, assigned to the nodes of C_5 so that each u_i forms the same angle with c and any two non-adjacent nodes are labeled with orthogonal vectors. These vectors give an orthogonal representation of C_5 in 3-space.

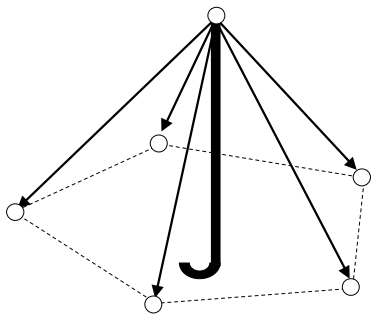

Figure 9.2: An orthogonal representation of C_5.

9.2 Smallest cone and the theta function

When looking for "economic" orthogonal representations, we can define "economic" in several ways. For example, we may want to find an orthogonal representation in a dimension as low as possible (even though this particular way of phrasing the question does not seem to be the most fruitful). Other definitions of "economic" are also often related to interesting graph properties.

We start with the problem from information theory that motivated th introduction of orthogonal representations [131].

9.2.1 Shannon capacity

Consider a noisy channel through which we are sending messages over a finite alphabet V. The noise may blur some letters so that certain pairs can be confused. We want to select as many words of length k as possible so that no two can possibly be confused. As we shall see, the number of words we can select grows as Θ^k for some $\Theta \geq 1$, which is called the *Shannon zero-error capacity* of the channel.

In terms of graphs, we can model the problem as follows. We consider V as the set of nodes of a graph, and connect two of them by an edge if they can be confused. This way we obtain a graph $G = (V, E)$. We denote by $\alpha(G)$ the maximum number of independent points (the maximum size of a stable set) in the graph G. If $k = 1$, then the maximum number of non-confusable messages is $\alpha(G)$.

To describe longer messages, we define the *strong product* $G \boxtimes H$ of two graphs $G = (V, E)$ and $H = (W, F)$ as the graph with $V(G \boxtimes H) = V \times W$, with $(i, u)(j, v) \in E(G \boxtimes H)$ iff $ij \in E$ and $uv \in F$, or $ij \in E$ and $u = v$, or $i = j$ and $uv \in F$. (If, for the purposes of this problem, we define two nodes to be adjacent if they are either equal or connected by and edge, then we can say that (i, u) and (j, v) are adjacent in $G \boxtimes H$ if and only if i is adjacent to j in G and u is adjacent to v in H.)

It is easy to see that this multiplication is associative and commutative (if we don't distinguish isomorphic graphs). The product of k copies of G is denoted by G^k.

Then $\alpha(G^k)$ is the maximum number of words of length k, composed of elements of V, so that for every two words there is at least one i $(1 \leq i \leq k)$ such that the i-th letters are different and non-adjacent in G, i.e., non-confusable. It is easy to see that

$$\alpha(G \boxtimes H) \geq \alpha(G)\alpha(H). \tag{9.1}$$

This implies that

$$\alpha(G^{k+l}) \geq \alpha(G^k)\alpha(G^l), \tag{9.2}$$

and

$$\alpha(G^k) \geq \alpha(G)^k. \tag{9.3}$$

The *Shannon capacity* of a graph G is the value

$$\Theta(G) = \lim_{k\to\infty} \alpha(G^k)^{1/k}$$

Inequality (9.2) implies that the limit exists, and inequality (9.3) implies that

$$\Theta(G) \geq \alpha(G). \tag{9.4}$$

It is not known whether $\Theta(G)$ can be computed for all graphs by any algorithm (polynomial or not), although there are several special classes of graphs for which this is not hard.

Example 9.2.1 Let C_4 denote a 4-cycle with nodes (a, b, c, d). By (9.4), we have $\Theta(G) \geq 2$. On the other hand, if we use a word, then all the 2^k words obtained from it by replacing a and b by each other, as well as c and d by each other, are excluded. Hence $\alpha(C_4^k) \leq 4^k/2^k = 2^k$, which implies that $\Theta(C_4) = 2$.

The argument for bounding Θ from above in this last example can be generalized as follows. Let $\overline{\chi}(G)$ denote the minimum number of complete subgraphs covering the nodes of G. (This is the same as the chromatic number of the complementary graph.) Trivially

$$\alpha(G) \leq \overline{\chi}(G), \tag{9.5}$$

and it is easy to see that

$$\overline{\chi}(G \boxtimes H) \leq \overline{\chi}(G)\overline{\chi}(H). \tag{9.6}$$

Hence

$$\alpha(G^k) \leq \overline{\chi}(G^k) \leq \overline{\chi}(G)^k,$$

and thus

$$\Theta(G) \leq \overline{\chi}(G). \tag{9.7}$$

It follows that if $\alpha(G) = \overline{\chi}(G)$, then $\Theta(G) = \alpha(G)$; for such graphs, nothing better can be done than reducing the alphabet to the largest mutually non-confusable subset.

Example 9.2.2 The smallest graph for which $\Theta(G)$ cannot be computed by these means is the pentagon C_5. If we set $V(C_5) = \{0, 1, 2, 3, 4\}$ with $E(C_5) = \{01, 12, 23, 34, 40\}$, then C_5^2 contains the stable set $\{(0, 0), (1, 2), (2, 4), (3, 1), (4, 3)\}$. So $\alpha(C_5)^{2k} = \alpha(C_5^2)^k \geq 5^k$, and hence $\Theta(C_5) \geq \sqrt{5}$.

We will show that equality holds here [131], but first we need some definitions.

9.2.2 Definition and basic properties of the theta function

The smallest angle of a rotational cone (in arbitrary dimension) which contains all vectors in an orthogonal representation of the graph gives rise to the theta-function of the graph [131]. Formally, we define

$$\vartheta(G) = \min_{(u_i),c} \max_{i \in V} \frac{1}{(c^{\mathsf{T}} u_i)^2},$$

where the minimum is taken over all orthonormal representations $(u_i : i \in V)$ of G and all unit vectors c. (Of course, we could fix c, but this is not always convenient.)

From the trivial orthogonal representation (Example 9.1.1) we get that

$$\vartheta(G) \le |V|.$$

Tighter inequalities can be proved:

Theorem 9.2.3 *For every graph G,*

$$\alpha(G) \le \vartheta(G) \le \overline{\chi}(G)).$$

Proof. First, let $S \subseteq V$ be a maximum independent set of nodes in G. Then in every orthonormal representation (u_i), the vectors $\{u_i : i \in S\}$ are mutually orthogonal unit vectors. Hence

$$1 = c^{\mathsf{T}}c \ge \sum_{i \in S} (c^{\mathsf{T}}u_i)^2 \ge |S| \min_i (c^{\mathsf{T}}u_i)^2,$$

and so

$$\max_{i \in V} \frac{1}{(c^{\mathsf{T}}u_i)^2} \ge |S| = \alpha(G).$$

This implies the first inequality.

The second inequality follows from Example 9.1.3, using $c = \frac{1}{k}(e_1 + \ldots e_k)$ as the handle. □

From Example 9.1.4 we get, using elementary trigonometry that

$$\vartheta(C_5) \le \sqrt{5}. \tag{9.8}$$

We'll see that equality holds here.

Lemma 9.2.4 *For any two graphs G and H,*

$$\vartheta(G \boxtimes H) \le \vartheta(G)\vartheta(H).$$

We will prove later (Corollary 9.2.17) that equality holds here.

Proof. The *tensor product* of two vectors $(u_1, \ldots, u_n) \in \mathbb{R}^n$ and $(v_1, \ldots, v_m) \in \mathbb{R}^m$ is the vector

$$u \circ v = (u_1v_1, \ldots, u_1v_m, u_2v_1, \ldots, u_2v_m, \ldots, u_nv_1, \ldots, u_nv_m) \in \mathbb{R}^{nm}.$$

The inner product of two tensor products can be expressed easily: if $u, x \in \mathbb{R}^n$ and $v, y \in \mathbb{R}^m$, then

$$(u \circ v)^{\mathsf{T}}(x \circ y) = (u^{\mathsf{T}}x)(v^{\mathsf{T}}y). \tag{9.9}$$

Now let $(u_i : i \in V)$ be an optimal orthogonal representation of G with handle c $(u_i, c \in \mathbb{R}^n)$, and let $(v_j : j \in V(H))$ be an optimal orthogonal representation of H with handle d $(v_j, d \in \mathbb{R}^m)$. It is easy to check, using (9.9), that the vectors $u_i \circ v_j$ $((i,j) \in V(G) \times V(H))$ form an orthogonal representation of $G \boxtimes H$. Furthermore, taking $c \circ d$ as its handle, we have by (9.9) again that

$$\big((c \circ d)^{\mathsf{T}}(u_i \circ v_j)\big)^2 = (c^{\mathsf{T}}u_i)^2(d \circ v_j)^2 \geq \frac{1}{\vartheta(G)} \cdot \frac{1}{\vartheta(H)},$$

and hence

$$\vartheta(G \boxtimes H) \leq \max_{i,j} \frac{1}{\big((c \circ d)^{\mathsf{T}}(u_i \circ v_j)\big)^2} \leq \vartheta(G)\vartheta(H).$$

□

This inequality has some important corollaries. First, we have

$$\alpha(G^k) \leq \vartheta(G^k) \leq \vartheta(G)^k,$$

which implies

Corollary 9.2.5 *For every graph G,*

$$\Theta(G) \leq \vartheta(G).$$

Second, the product graph $G \boxtimes \overline{G}$ has independent points in the diagonal, implying that

$$\vartheta(G \boxtimes \overline{G}) \geq \alpha(G \boxtimes \overline{G}) \geq |V|.$$

Hence

Corollary 9.2.6 *For every graph G,*

$$\vartheta(G)\vartheta(\overline{G}) \geq |V|.$$

Since $\overline{C_5} \cong C_5$, Corollary 9.2.6 implies that $\vartheta(C_5) \geq \sqrt{5}$, and together with (9.8), we get that $\vartheta(C_5) = \sqrt{5}$. Equality does not hold in general in Corollary 9.2.6, but it does when G has a node-transitive automorphism group. We postpone the proof of this fact until some further formulas for ϑ will be developed (Corollary 9.2.14).

Exercise 9.1 If $\vartheta(\overline{G}) = 2$, then G is bipartite.

Exercise 9.2 (a) If H is an induced subgraph of G, then $\vartheta(H) \leq \vartheta(G)$. (b) If H is a spanning subgraph of G, then $\vartheta(H) \geq \vartheta(G)$.

Exercise 9.3 Let G be a graph and $v \in V$. (a) $\vartheta(G - v) \geq \vartheta(G) - 1$. (b) If v is an isolated node, then equality holds. (c) If v is adjacent to all other nodes, then $\vartheta(G - v) = \vartheta(G)$.

Exercise 9.4 Let $G = (V, E)$ be a graph and let $V = S_1 \cup \cdots \cup S_k$ e a partition of V. (a) Then $\vartheta(G) \leq \sum_i \vartheta(G[S_i])$. (b) If no edge connects nodes in different sets S_i, then equality holds. (c) Suppose that any two nodes in different sets S_i are adjacent. How can $\vartheta(G)$ be expressed in terms of the $\vartheta(G[S_i])$?

9.2.3 More expressions for ϑ

We prove a number of formulas for $\vartheta(G)$, which together have a lot of implications. These will be proved together.

For every graph G, let $\mathcal{M}(G)$ denote the set of symmetric matrices $A \in \mathbb{R}^{V \times V}$,such that $A_{ij} = -1$ if $i = j$ or $ij \in \overline{E}$.

Proposition 9.2.7 (Minimizing the largest eigenvalue) *For every graph $G = (V, E)$,*

$$\vartheta(G) = \min\{\lambda_{\max}(A) : A \in \mathcal{M}(G)\}.$$

Consider the following semidefinite programs:

$$\begin{array}{ll} \text{minimize} & t \\ \text{subject to} & Y \succeq 0 \\ & Y_{ij} = -1 \quad (\forall\ ij \in \overline{E}) \\ & Y_{ii} = t - 1 \end{array} \tag{9.10}$$

and

$$\begin{array}{ll} \text{maximize} & \displaystyle\sum_{i,j \in V} Z_{ij} \\ \text{subject to} & Z \succeq 0 \\ & Z_{ij} = 0 \quad (\forall\ ij \in E) \\ & \mathrm{tr}(Z) = 1 \end{array} \tag{9.11}$$

It is not hard to check that these are duals.

Proposition 9.2.8 (Semidefinite programs) *For every graph G, $\vartheta(G)$ is the optimum of either one of the programs* (9.10), (9.11).

Finally, we use orthonormal representations of the complementary graph:

Proposition 9.2.9 (Complementary graph) *For every graph G,*

$$\vartheta(G) = \max \sum_{i \in V} (d^{\mathsf{T}} v_i)^2,$$

where the maximum extends over all orthonormal representations $(v_i : \ i \in V)$ of the complementary graph $\overline{G}$ and all unit vectors d.

Proof. We prove Propositions 9.2.7, 9.2.8 and 9.2.9 together. First we show:

$$\vartheta(G) \leq \min\{\lambda_{\max}(A) : \ A \in \mathcal{M}(G)\}. \tag{9.12}$$

Let A be a matrix attaining the minimum on the right hand side, and let $t = \lambda_{\max}(A)$. Then the matrix $tI - A$ is positive semidefinite, and so it can be written as a Gram matrix, i.e., there are vectors $x_i \in \mathbb{R}^n$ such that

$$x_i^{\mathsf{T}} x_j = (tI - A)_{ij} = \begin{cases} -1 & \text{if } ij \in \overline{E}, \\ t - 1 & \text{if } i = j \end{cases}$$

(no condition if $ij \in E$). Let c be a vector orthogonal to all the x_i (we increase the dimension of the space if necessary), and set

$$u_i = \frac{1}{\sqrt{t}}(c + x_i).$$

Then $u_i^{\mathsf{T}} u_i = 1$ and $u_i^{\mathsf{T}} u_j = 0$ for $ij \in \overline{E}$, so (u_i) is an orthonormal representation of G. Furthermore, with handle c we have $c^{\mathsf{T}} u_i = 1/\sqrt{t}$, which implies that $\vartheta(G) \leq t$.

Next, we claim that

$$\min\{\lambda_{\max}(A) : \ A \in \mathcal{M}(G)\} \leq \min\{t : \ (t, Y) \text{ satisfies (9.10)}\}. \tag{9.13}$$

Indeed, if (t, Y) is an optimum solution of (9.10), then $A = tI - Y$ is a matrix satisfying the conditions in Proposition 9.2.7, and $\lambda_{\max}(A) \leq t$.

Our next step is:

$$\min\{t : \ (t, Y) \text{ satisfies (9.10)}\} = \max\{ \sum_{i,j \in V} Z_{ij} : \ Z \text{ satisfies (9.11)}\}. \tag{9.14}$$

Indeed, the second program is the dual of the first. The first program has a strictly feasible solution (just choose t large enough), so by the Duality Theorem of semidefinite programming, the two programs have the same objective value.

Now we get to orthonormal representations of the complementary graph.

$$\max\{ \sum_{i,j \in V} Z_{ij} : \ Z \text{ satisfies (9.11)}\} \leq \max\Big\{\sum_{i \in V} (d^{\mathsf{T}} v_i)^2 : \ (v_i, d) \text{ ONR of } \overline{G}\Big\}. \tag{9.15}$$

Indeed, let Z be an optimum solution of (9.11) with objective function value t. We can write Z is a Gram matrix: $Z_{ij} = z_i^{\mathsf{T}} z_j$ where $z_i \in \mathbb{R}^k$ for some $k \geq 1$. By the properties of Z, the vectors z_i form an orthogonal representation of $\overline{G}$. We have

$$t = \sum_{i,j} z_i^{\mathsf{T}} z_j = \Big(\sum_i z_i\Big)^{\mathsf{T}} \Big(\sum_i z_i\Big) = \Big|\sum_i z_i\Big|^2, \tag{9.16}$$

and

$$1 = \mathrm{tr}(Z) = \sum_i z_i^{\mathsf{T}} z_i = \sum_i |z_i|^2.$$

Let us rescale the vectors z_i to get the orthonormal representation $v_i = z_i/|z_i|$ (if $z_i = 0$ then we take a unit vector orthogonal to everything else as v_i). We define $d = (\sum_i z_i)/\sqrt{t}$ (which is a unit vector by (9.16)). Using Chauchy–Schwarz,

$$\begin{aligned}\sum_i (d^{\mathsf{T}} v_i)^2 &= \Big(\sum_i |z_i|^2\Big)\Big(\sum_i (d^{\mathsf{T}} v_i)^2\Big) \\ &\geq \Big(\sum_i |z_i| d^{\mathsf{T}} v_i\Big)^2 = \Big(\sum_i d^{\mathsf{T}} z_i\Big)^2 = \Big(d^{\mathsf{T}} \sum_i z_i\Big)^2 = t.\end{aligned}$$

This proves that the maximum of the right hand side of (9.15) is at least t.

The last step is to prove:

$$\max\Big\{\sum_{i\in V} (d^{\mathsf{T}} v_i)^2 : \ (v_i, d) \text{ ONR of } \overline{G}\Big\} \leq \vartheta(G). \tag{9.17}$$

Using the definition of ϑ, it suffices to prove that if $(u_i : \ i \in V)$ is an orthonormal representation of G in $\mathbb{R}^n$ with handle c, and $(v_i : \ i \in V)$ is an orthonormal representation of $\overline{G}$ in $\mathbb{R}^m$ with handle d, then

$$\sum_{i\in V} (d^{\mathsf{T}} v_i)^2 \leq \max_{i\in V} \frac{1}{(c^{\mathsf{T}} u_i)^2}. \tag{9.18}$$

By the same computation as in the proof of Lemma 9.2.4, we get that the vectors $u_i \circ v_i$ $(i \in V)$ are mutually orthogonal unit vectors, and hence

$$\sum_i (c^{\mathsf{T}} u_i)^2 (d^{\mathsf{T}} v_j)^2 = \sum_i \big((c \circ d)^{\mathsf{T}} (u_i \circ v_i)\big)^2 \leq 1.$$

On the other hand,

$$\sum_i (c^{\mathsf{T}} u_i)^2 (d^{\mathsf{T}} v_j)^2 \geq \min_i (c^{\mathsf{T}} u_i)^2 \sum_i (d^{\mathsf{T}} v_j)^2,$$

which implies that

$$\sum_i (d^{\mathsf{T}} v_j)^2 \geq \frac{1}{\min_i (c^{\mathsf{T}} u_i)^2} = \max_i \frac{1}{(c^{\mathsf{T}} u_i)^2}.$$

This proves (9.18). □

Exercise 9.5 Prove that $\vartheta(G)$ is the maximum of the largest eigenvalues of matrices $(v_i^{\mathsf{T}} v_j)$, taken over all orthonormal representations (v_i) of $\overline{G}$.

Exercise 9.6 (Alon and Kahale) Let G be a graph, $v \in V$, and let H be obtained from G by removing v and all its neighbors. Prove that

$$\vartheta(G) \leq 1 + \sqrt{|V(H)|\vartheta(H)}.$$

9.2.4 More properties

Perhaps the most important consequence of the formulas proved in the preceding section is that the value of $\vartheta(G)$ is polynomial time computable [80]. More precisely,

Theorem 9.2.10 *There is a polynomial time algorithm that computes, for every graph G and every $\varepsilon > 0$, a real number t such that*

$$|\vartheta(G) - t| < \varepsilon.$$

The significance of this fact is underlined if we combine it with Theorem 9.2.3: The two important graph parameters $\alpha(G)$ and $\chi(\overline{G})$ are both NP-hard, but they have a polynomial time computable quantity sandwiched between them.

Algorithms proving this theorem can be based on almost any of our formulas for ϑ. The simplest is to refer to Proposition 9.2.8, and the polynomial time solvability of semidefinite programs (see Section 13.8.3 in the Appendix).

We can write Proposition 9.2.9 as

$$\min_{(u_i),c} \max_{i \in V} \frac{1}{(c^{\mathsf{T}} u_i)^2} = \max_{(v_i),d} \sum_{i \in V} (d^{\mathsf{T}} v_i)^2 \tag{9.19}$$

(where (u_i) ranges through orthonormal representations of G, (v_i), through orthonormal representations of $\overline{G}$, and c, d are unit vectors in the corresponding euclidean spaces). From the fact that equality holds here, it follows that equality holds in the arguments above. In particular, we obtain the following corollaries:

Proposition 9.2.11 *Every graph G has an orthonormal representation (u_i) with handle c such that for every node i,*

$$c^{\mathsf{T}} u_i = \frac{1}{\sqrt{\vartheta(G)}}.$$

Proof. The representation constructed in the proof of (9.12) must be optimal with $t = \vartheta(G)$, and it has the property in the Corollary. □

This corollary can be used to give another definition of ϑ, discovered independently by Karger, Motwani and Sudan. Consider the vectors

$$v_i = \sqrt{\frac{\vartheta(G)}{\vartheta(G)-1}}\Big(u_i - \frac{1}{\sqrt{\vartheta(G)}}c\Big).$$

It is easy to check that these are unit vectors and satisfy

$$v_i^{\mathsf{T}} v_j = \frac{-1}{\vartheta(G)-1}$$

for every pair i, j of nonadjacent nodes. This computation can be reversed, and we get

Proposition 9.2.12 *For every graph $G = (V, E)$, $\vartheta(G)$ can be defined as the smallest real number $t \geq 0$ for which there exists a representation $i \mapsto v_i$ $(i \in V)$ such that $|v_i| = 1$ and $v_i^{\mathsf{T}} v_j = -1/(t-1)$ for all $ij \in \overline{E}$.*

For optimal orthogonal representations in the dual definition of $\vartheta(G)$ we cannot claim such a nice "homogeneity" as in Corollary 9.2.11, but we get something from the symmetries of the graph.

Proposition 9.2.13 *Every graph G has an orthonormal representation (u_i, c) in $\mathbb{R}^n$ with $c^{\mathsf{T}} u_i = 1/\sqrt{\vartheta(G)}$, and its complement has an orthonormal representation (v_i, d) in $\mathbb{R}^n$ with $\sum_i (d^{\mathsf{T}} v_i)^2 = \vartheta(G)$, such that every automorphism of G lifts to an orthogonal transformation of $\mathbb{R}^n$ that leaves both representations invariant.*

Proof. We give the proof for the orthonormal representation of the complement. The set of optimum solutions of the semidefinite program (9.11) form a bounded convex set, which is invariant under the transformations $Z \mapsto PZP$, where P is the permutation matrix of an automorphism of G. The center of gravity of this convex set is a matrix Z which is fixed by these transformations, i.e., it satisfies $PZP = Z$ for all automorphisms P. The construction of an orthonormal representation of $\overline{G}$ in the proof of (9.15) can be done in a canonical way (e.g., choosing the rows of $Z^{1/2}$ as the vectors z_i), and so the obtained optimal orthonormal representation will be invariant under the automorphism group of G (acting as permutation matrices). □

Corollary 9.2.14 *If G has a node-transitive automorphism group, then*

$$\vartheta(G)\vartheta(\overline{G}) = |V|.$$

Proof. It follows from Proposition 9.2.13 that $\overline{G}$ has an orthonormal representation (v_i, d) in $\mathbb{R}^n$ such that $\sum_i (d^{\mathsf{T}} v_i)^2 = \vartheta(G)$, and $d^{\mathsf{T}} v_i$ is independent of i. So $(d^{\mathsf{T}} v_i)^2 = \vartheta(G)/|V|$ for

all nodes i, and

$$\vartheta(\overline{G}) \le \max_i \frac{1}{(d^{\mathsf{T}} v_i)^2} = \frac{|V|}{\vartheta(G)}.$$

Since we already know the reverse inequality (9.2.6), this proves the Corollary. □

Corollary 9.2.15 *If G is a self-complementary graph with a node-transitive automorphism group, then*

$$\Theta(G) = \vartheta(G) = \sqrt{|V|}.$$

Proof. The diagonal in $G \boxtimes \overline{G}$ is independent, so $\alpha(G \boxtimes G) = \alpha(G \boxtimes \overline{G}) \ge |V|$, and hence $\Theta(G) \ge \sqrt{|V|}$. On the other hand, $\vartheta(G) \le \sqrt{|V|}$ follows by Corollary 9.2.14. □

An example to which this corollary applies is any *Paley graph*: for a prime $p \equiv 1 \pmod 4$, we take the $\{0, 1, \ldots, p-1\}$ as nodes, and connect two of them iff their difference is a quadratic residue. Thus we get an infinite family for which the Shannon capacity is non-trivial (i.e., $\Theta > \alpha$), and can be determined exactly.

The Paley graphs are quite similar to random graphs, and indeed, for random graphs ϑ behaves similarly (Juhász [108]). (It is not known however how large the Shannon capacity of a random graph is.)

Theorem 9.2.16 *If G is a random graph on n nodes with edge density $1/2$, then $\sqrt{n} < \vartheta(G) < 2\sqrt{n}$ with probability $1 + o(1)$.*

As a further application of the duality established in Section 9.2.3, we prove:

Proposition 9.2.17 *For any two graphs G and H,*

$$\vartheta(G \boxtimes H) = \vartheta(G)\vartheta(H).$$

Proof. Let (v_i, d) be an orthonormal representation of $\overline{G}$ which is optimal in the sense that $\sum_i (d^{\mathsf{T}} v_i)^2 = \vartheta(G)$, and let (w_j, e) be an orthonormal representation of $\overline{H}$ such that $\sum_i (e^{\mathsf{T}} w_j)^2 = \vartheta(H)$. It is easy to check that the vectors $v_i \circ w_j$ form an orthonormal representation of $\overline{G \boxtimes H}$, and so using handle $d \circ e$ we get

$$\vartheta(G \boxtimes H) \ge \sum_{i,j} \big((d \circ e)^{\mathsf{T}} (v_i \circ w_j)\big)^2 = \sum_{i,j} (d^{\mathsf{T}} v_i)^2 (e^{\mathsf{T}} w_j)^2 = \vartheta(G)\vartheta(H).$$

We already know the reverse inequality, which completes the proof. □

The *fractional chromatic number* $\chi^*(G)$ is defined as the least t for which there exists a family $(A_j : j = 1, \ldots, p)$ of stable sets in G, and nonnegative weights $(\tau_j : j = 1, \ldots, p)$

such that $\sum\{\tau_j : A_j \ni i\} \geq 1$ for all $i \in V$ and $\sum_j \tau_j = t$. Note that the definition χ^* can be considered as a linear program. By linear programming duality, $\chi^*(G)$ is equal to the largest s for which there exist weights $(\sigma_i : i \in V)$ such that $\sum_{i \in A} \sigma_i \leq 1$ for every stable set A and $\sum_i \sigma_i = s$.

Clearly $\omega(G) \leq \chi^*(G) \leq \chi(G)$. The following is easy to prove:

Proposition 9.2.18 *$\vartheta(G) \leq \chi^*(\overline{G})$.*

Exercise 9.7 (a) Show that any stable set S provides a feasible solution of (9.11). (b) Show that any k-coloring of $\overline{G}$ provides a feasible solution of (9.10). (c) Give a new proof of the Sandwich Theorem 9.2.3 based on (a) and (b).

Exercise 9.8 Prove Proposition 9.2.18.

9.2.5 How good is ϑ as an approximation?

How good an approximation does ϑ provide for α? Unfortunately, it can be quite bad. First, consider the case when α is very small. Koniagin [116] constructed a graph that has $\alpha(G) = 2$ and $\vartheta(G) = \Omega(n^{1/3})$. This is the largest $\vartheta(G)$ can be; in fact, Alon and Kahale [11], improving results of Kashin and Koniagin [111], proved that there is a constant c such that

$$\vartheta(G) \leq 9n^{1-2/(\alpha(G)+1)}. \tag{9.20}$$

Once α is unbounded, this inequality does not say much, and in fact very little is true. Feige [71] showed that there are graphs for which $\alpha(G) = n^{o(1)}$ and $\vartheta(G) = n^{1-o(1)}$; in other words, ϑ/α can be larger than $n^{1-\varepsilon}$ for every $\varepsilon > 0$. (The existence of such graphs also follows from the results of Håstad [89] showing that it is NP-hard to determine $\alpha(G)$ with a relative error less than $n^{1-\varepsilon}$, where $n = |V|$.) By results of Szegedy [192], this also implies that $\vartheta(\overline{G})$ does not approximate the chromatic number within a factor of $n^{1-\varepsilon}$.

Let us consider the other end of the scale, when $\vartheta(\overline{G})$ is small. The following theorem was proved by Karger, Motwani and Sudan [110]:

Theorem 9.2.19 *For every graph G,*

$$\alpha(G) \geq \frac{n^{3/(\vartheta(\overline{G})+1)}}{3\ln n}.$$

Note that we have $\vartheta(G) \geq n/\vartheta(\overline{G})$ by Corollary 9.2.6.

Proof. Let $t = \vartheta(\overline{G})$. The case when the maximum degree Δ of the graph satisfies $\Delta > n^{t/(t+1)}$ can be settled by a simple induction: Let v be a node with degree Δ, and let G' be the

subgraph induced by the neighbors v. It is easy to see that $\vartheta(G') \le t-1$, and so (by induction on n) we can find in G' a stable set of size at least $\Delta^{3/t}/(2\sqrt{\ln \Delta}) > n^{3/(t+1)}/(2\sqrt{\ln n})$.

So suppose that $\Delta \le n^{t/(t+1)}$. By Corollary 9.2.12, there are unit vectors $u_i \in \mathbb{R}^d$ for some d such that $u_i^{\mathsf{T}} u_j = -1/(t-1)$ whenever $ij \in E$. The case $t = 2$ is easy (cf. Exercise 9.1), so we assume that $t > 2$.

Let w be a random point in $\mathbb{R}^d$ whose coordinates are independent standard Gaussian random variables. Fix an $s > 0$, and consider the set $S = \{i : \ w^{\mathsf{T}} u_i \ge c\}$. The inner product $w^{\mathsf{T}} u_i$ has standard Gaussian distribution, and hence the probability that a given node belongs to S is

$$Q(s) = \frac{1}{\sqrt{2\pi}} \int_s^\infty e^{-x^2/2}\, dx.$$

Thus the expected size of S is $\mathsf{E}|S| = Q(s)n$. Furthermore, let $ij \in E$, then $|u_i + u_j|^2 = (2t-4)/(t-1)$, and so the probability that both nodes u_i and u_j belong to S can be estimated as follows:

$$\mathcal{P}(w^{\mathsf{T}} u_i \ge s,\ w^{\mathsf{T}} u_j \ge s) \le \mathcal{P}(w^{\mathsf{T}}(u_i + u_j) \ge 2s) = Q\Big(\sqrt{\frac{2t-2}{t-2}}\, s\Big).$$

Hence the expected number of edges of E spanned by S satisfies

$$\mathsf{E}|E[S]| \le Q\Big(\sqrt{\frac{2t-2}{t-2}}\, s\Big)|E| \le Q\Big(\sqrt{\frac{2t-2}{t-2}}\, s\Big)\frac{n\Delta}{2}.$$

We can delete at most $|E[S]|$ nodes from S to get a clique K. The expected size of K is at least

$$\mathsf{E}|K| \ge \mathsf{E}(|S| - |E[S]|) = \mathsf{E}|S| - \mathsf{E}|E[S]| \ge Q(s)n - Q\Big(\sqrt{\frac{2t-2}{t-2}}\, s\Big)\frac{n\Delta}{2}.$$

Using also the bound on Δ, it follows that

$$\omega(G) \ge Q(s)n - Q\Big(\sqrt{\frac{2t-2}{t-2}}\, s\Big)\frac{n^{2t+1}t + 1}{2}.$$

We choose s so that it maximizes the right hand side. By elementary computation we get that

$$s = \sqrt{\frac{2t-4}{t+1}\ln n}$$

is an approximately optimal choice, and it gives

$$\alpha(G) \ge \frac{n^{2/t}}{3\sqrt{\ln n}}.$$

□

The previous theorem has an important application to a coloring problem. Suppose that somebody gives a graph and guarantees that the graph is 3-colorable, without telling us its 3-coloring. Can we find this 3-coloring? (This may sound artificial, but this kind of situation does arise in cryptography and other data security applications; one can think of the hidden 3-coloring as a "watermark" that can be verified if we know where to look.)

It is easy to argue that knowing that the graph is 3-colorable does not help: it is still NP-hard to find the 3-coloration. But suppose that we would be satisfied with finding a 4-coloration, or 5-coloration, or $(\log n)$-coloration; is this easier? It is known that to find a 4-coloration is still NP-hard, but little is known above this. Improving earlier results, Karger, Motwani and Sudan [110] gave a polynomial time algorithm that, given a 3-colorable graph, computes a coloring with $O(n^{1/4}(\ln n)^{3/2})$ colors. More recently, this was improved by Blum and Karger [33] to $O(n^{3/14})$.

The algorithm of Karger, Motwani and Sudan starts with computing $\vartheta(\overline{G})$, which is at most 3 by Theorem 9.2.3. Using Theorem 9.2.19, they find a stable set of size $\Omega(n^{3/4}/\sqrt{\ln n})$. Deleting this set from G and iterating, they get a coloring of G with $O(n^{1/4}(\ln n)^{3/2})$ colors.

9.2.6 Perfect graphs

Recall that for a graph G, we denote by $\omega(G)$ the size of its largest clique, and by $\chi(G)$, its chromatic number. A graph G is called *perfect*, if for every induced subgraph G' of G, we have $\omega(G') = \chi(G')$. To be perfect is a rather strong structural property; nevertheless, many interesting classes of graphs are perfect (bipartite graphs, their complements and their linegraphs; interval graphs; comparability and incomparability graphs of posets; chordal graphs). We do not discuss perfect graphs, only to the extent needed to show their connection with orthogonal representations.

The following deep characterization perfect graphs was conjectured by Berge in 1961 and proved by Chudnovsky, Robertson, Seymour and Thomas [41].

Theorem 9.2.20 (The Strong Perfect Graph Theorem [41]) *A graph is perfect if and only if neither the graph nor its complement contains a chordless odd cycle longer than* 3*.*

As a corollary we can formulate the "The Weak Perfect Graph Theorem" proved much earlier [128]:

Proposition 9.2.21 *The complement of a perfect graph is perfect.*

From this corollary it follows that in the definition of perfect graphs we could replace the equation $\omega(G') = \chi(G')$ by $\alpha(G') = \chi(\overline{G'})$. In particular, if G is a perfect graph, then $\alpha(G) = \chi(\overline{G})$, and so by Theorem 9.2.3,

Proposition 9.2.22 *For every perfect graph* G*,* $\Theta(G) = \vartheta(G) = \alpha(G) = \chi(\overline{G})$*.*

Corollary 9.2.23 *The independence number and the chromatic number of a perfect graph are polynomial time computable.*

9.2.7 The TSTAB body and weighted θ-function

For every orthonormal representation $(v_i : i \in V)$ of $\overline{G}$, we consider the linear constraint

$$\sum_{i \in V} (e_1^{\mathsf{T}} v_i)^2 x_i \leq 1. \tag{9.21}$$

It is easy to see that these inequalities are valid for STAB(G); we call them *orthogonality constraints*. The solution set of non-negativity and orthogonality constraints is denoted by TSTAB(G). It is clear that TSTAB is a closed, convex set. The incidence vector of any stable set A satisfies (9.21). Indeed, it then says that

$$\sum_{i \in A} (e_1^{\mathsf{T}} v_i)^2 \leq 1.$$

Since the v_i $(i \in A)$ are mutually orthogonal, the left hand side is just the squared length projection of e_1 onto the subspace spanned by these e_i, and the length of this projection is at most the length of e_1, which is 1.

Furthermore, every clique constraint is an orthogonality constraint. Indeed,

$$\sum_{i \in B} x_i \leq 1$$

is the constraint derived from the orthogonal representation

$$i \mapsto \begin{cases} e_1, & i \in A, \\ e_i, & \text{otherwise.} \end{cases}$$

Hence we have

$$\text{STAB}(G) \subseteq \text{TSTAB}(G) \subseteq \text{QSTAB}(G)$$

for every graph G.

There is a dual characterization of TSTAB [82]. For every orthonormal representation $(u_i : i \in V)$, consider the vector $x[u] = (e_1^{\mathsf{T}} u_i)^2 : i \in V) \in \mathbb{R}^V$.

Theorem 9.2.24 TSTAB(G) $= \{x[u] : u$ is an orthonormal representation of $G\}$.

Proof. This can be derived from semidefinite duality. □

Not every orthogonality constraint is a clique constraint; in fact, the number of essential orthogonality constraints is infinite in general:

Theorem 9.2.25 TSTAB(G) *is polyhedral if and only if the graph is perfect. In this case* TSTAB = STAB = QSTAB.

While TSTAB is a rather complicated set, in many respects it behaves much better than, say, STAB. For example, it has a very nice connection with graph complementation:

Theorem 9.2.26 TSTAB($\overline{G}$) *is the antiblocker of* TSTAB(G).

Maximizing a linear function over STAB(G) or QSTAB(G) is NP-hard; but, surprisingly, TSTAB behaves much better:

Theorem 9.2.27 *Every linear objective function can be maximized over* TSTAB(G) *(with arbitrarily small error) in polynomial time.*

The maximum of $\sum_i x_i$ over TSTAB(G) is the familiar function $\vartheta(G)$.

See [83, 114] for more detail.

9.3 Minimum dimension

Perhaps the most natural way to be "economic" in constructing an orthogonal representation is to minimize the dimension. We can say only a little about the minimum dimension of all orthogonal representations, but we get interesting results if we impose some "non-degeneracy" conditions. We will study three nondegeneracy conditions: general position, faithfulness, and the strong Arnold property.

9.3.1 Minimum dimension with no restrictions

Let $d_{\min}(G)$ denote the minimum dimension in which D has an orthonormal representation. The following are easy to prove, using Theorem 2.4.3 for (b).

Proposition 9.3.1 (a) $\vartheta(G) \le d_{\min}(G)$.

(b) $\frac{1}{2}\log\chi(\overline{G}) \le d_{\min}(G)$.

(c) $d_{\min}(G \boxtimes H) \le d_{\min}(G)d_{\min}(H)$.

This Proposition shows that we can use $d_{\min}(G)$ as an upper bound on $\Theta(G)$; however, it would not be better than $\vartheta(G)$. On the other hand, if we consider orthogonal representations over fields of finite characteristic, the dimension may be a better bound on the Shannon capacity than ϑ [87, 9].

9.3.2 General position orthogonal representations

The first non-degeneracy condition we study is *general position*: we assume that any d of the representing vectors in $\mathbb{R}^d$ are linearly independent.

A result of Lovász, Saks and Schrijver [137] (see [13] for an application in quantum computing) finds an exact condition for this type of geometric representability.

Theorem 9.3.2 *A graph with n nodes has a general position orthogonal representation in $\mathbb{R}^d$ if and only if it is $(n-d)$-connected.*

The condition that the given set of representing vectors is in general position is not easy to check (it is NP-hard). A weaker, but very useful condition will be that the vectors representing the nodes non-adjacent to any node v are linearly independent. We say that such a representation is in *locally general position*. To exclude some trivial complications, we assume in this case that all the representing vectors are nonzero.

The following Theorem is due to Lovász, Saks and Schrijver [137].

Theorem 9.3.3 *If G is a graph with n nodes then the following are equivalent:*

(i) *G is $(n-d)$-connected;*

(ii) *G has a general position orthogonal representation in $\mathbb{R}^d$;*

(iii) *G has a locally general position orthogonal representation in $\mathbb{R}^d$.*

Proof. We start with proving (iii)⇒(i), which illustrates the connection with connectivity. Suppose that G is not $(n-d)$-connected, then there is a partition $V = V_0 \cup V_1 \cup V_2$ such that $|V_0| \leq n-d-1$ and no edge connects V_1 and V_2.

Suppose that G has an orthonormal representation in $\mathbb{R}^d$ as in (iii). Since the nodes in V_1 are non-adjacent to any node of V_2, the vectors representing V_1 are linearly independent. Similarly, the vectors representing V_2 are linearly independent. Since the vectors representing V_1 and V_2 are mutually orthogonal, all vectors representing $V_1 \cup V_2$ are linearly independent. But $|V_1 \cup V_2| = n - |V_0| \geq d+1$, a contradiction.

The implication (ii)⇒(iii) is trivial.

The difficult part is (i)⇒(ii), i.e., the construction of a general position orthogonal (or orthonormal) representation for $(n-d)$-connected graphs. As a matter of fact, the following construction is almost trivial, the difficulty is the proof of its validity.

Let $\sigma = (v_1, ..., v_n)$ be any ordering of the nodes of $G = (V, E)$. Let us choose $f(v_1), f(v_2), \ldots$ consecutively as follows. $f(v_1)$ is any vector of unit length. Suppose that $f(v_i)$ $(1 \leq i \leq j)$ are already chosen. Then we choose $f(v_j + 1)$ randomly, subject to the constraints that it has to be orthogonal to certain previous vectors $f(v_i)$. These orthogonality constraints restrict $f(v_{j+1})$ to a linear subspace L_{j+1}, and we choose it from the uniform

distribution over the unit sphere of L_{j+1}. Note that if G is $(n-d)$-connected then every node of it has degree at least $n-d$ and hence

$$\dim L \geq d - \#\{i : \ i \leq j, \ v_i v_j \notin E\} \geq d - (d-1) = 1,$$

and so $f(v_{j+1})$ can always be chosen.

This way we get a random mapping $f : \ V \to \mathbb{R}^d$, i.e., a probability distribution over $(S^d)^V$, which we denote by μ_σ (it may depend on the initial ordering of the nodes). We call f the *random sequential orthogonal representation* of G (associated with the ordering $(v_1, ..., v_n)$). Theorem 9.3.3 will then follow from Theorem 9.3.4 below. □

Theorem 9.3.4 *Let G be any graph and fix any ordering of its nodes. Let f be the random sequential orthogonal representation of G. Then*

(a) *If G is not $(n-d)$-connected then f is not in locally general position;*

(b) *If G is $(n-d)$-connected then with probability 1, f is in general position.*

As we pointed out, the distribution of a random sequential orthogonal representation may depend on the initial ordering. The key to the proof will be that this dependence is not too strong. We say that two probability measures μ and ν on the same probability space S are *mutually absolute continuous*, if for any measurable subset A of S, $\mu(A) = 0$ if and only if $\nu(A) = 0$.

Lemma 9.3.5 *For any two orderings σ and τ of V, the distributions μ_σ and μ_τ are mutually absolute continuous.*

Before proving this lemma, we have to state and prove a simple technical fact. We say that two linear subspaces $A, B \subseteq \mathbb{R}^d$ are *orthogonal*, if every $a \in A$ is orthogonal to every $b \in B$. A different relation is that A, B are *weakly orthogonal*, which means that the orthogonal projection of A onto B is $A \cap B$. This is equivalent to saying that $A \cap B$ and $A \cap B^\perp$ generate A. It is easy to see that this weak orthogonality relation is symmetric. If A and B are weakly orthogonal, then $A \cap B$, $A \cap B^\perp$ and $A^\perp \cap B$ are orthogonal.

Lemma 9.3.6 *Let A and B be weakly orthogonal linear subspaces of $\mathbb{R}^d$ with $\dim(A \cap B) \geq 2$. Select a unit vector a_1 uniformly from A, and then select a unit vector b_1 uniformly from $B \cap a_1^\perp$. Also, select a unit vector b_2 uniformly from B, and then select a unit vector a_2 uniformly from $A \cap b_2^\perp$. Then the distributions of (a_1, b_1) and (a_2, b_2) are mutually absolute continuous.*

Proof. The special case when $A \subseteq B$ is trivial. Suppose that $A \not\subseteq B$ and also $B \not\subseteq A$.

Observe that if Y and Z are orthogonal spaces, a unit vector a in $Y \oplus Z$ can be written uniquely in the form $y \cos\theta + z \sin\theta$ where y is a unit vector in Y, z is a unit vector in Z,

and $\theta \in [0, \pi/2]$. Uniform selection of a means independent uniform selection of y and z, and an independent selection of θ from a distribution $\zeta_{s,t}$ that depends only on $s = \dim(Y)$ and $t = \dim(Z)$. If $t = 0$ then $\theta = 0$ with probability 1. If $s, t \geq 1$, the only thing we need about $\zeta_{s,t}$ is that it is mutually absolute continuous with respect to the uniform distribution on the interval $[0, \pi/2]$.

So (a_1, b_1) can be generated through five independent choices: a uniform unit vector x_1 from $A \cap B^\perp$, a uniform unit vector z_1 from $A^\perp \cap B$, a pair of orthogonal unit vectors (y_1, y_2) selected from $A \cap B$ (uniformly over all such pairs: this is possible since $\dim(A \cap B) \geq 2$), and two numbers θ_1 selected according to ζ_{c_0,c_1} and θ_2 is selected according to ζ_{c_0-1,c_2}. The distribution of (a_2, b_2) is described similarly except that θ_2 is selected according to ζ_{c_0,c_2} and θ_1 is selected according to ζ_{c_0-1,c_1}.

Since $c_0 \geq 2$, ζ_{c_0,c_2} and ζ_{c_0-1,c_2} are mutually absolute continuous and ζ_{c_0,c_1} and ζ_{c_0-1,c_1} are mutually absolute continuous, from which we deduce that the distributions of (a_1, b_1) and (a_2, b_2) are mutually absolute continuous. □

Proof of Lemma 9.3.5. It suffices to prove that if τ is obtained from σ by swapping $v_{\sigma(j)}$ and $v_{\sigma(j+1)}$ $(1 \leq j \leq n-1)$ then $\mu(\sigma)$ and $\mu(\tau)$ are mutually absolute continuous. We prove this by induction on j.

Fix $j \geq 1$. We may assume, without loss of generality, that $\sigma = (v_1, \ldots, v_n)$. For $1 \leq i \leq n$, let $V_i = \{v_1, \ldots, v_i\}$.

We want to compare the distributions of $f(v_1), \ldots, f(v_n)$ and $g(v_1), \ldots, g(v_n)$. It suffices to prove that the distributions of $f(v_1), \ldots, f(v_{j+1})$ and $g(v_1), \ldots, g(v_{j+1})$ are mutually absolute continuous, since conditioned on any given assignment of vectors to $v_1, \ldots, v_{j+1}$, the distributions μ_σ and μ_τ are identical.

Also, note that the distributions of $f(v_1), \ldots, f(v_{j-1})$ and $g(v_1), \ldots, g(v_{j-1})$ are identical. Let $x_1, \ldots, x_{j-1}$ any selection of vectors for the first $j-1$ nodes. It suffices to show that the distributions of $f(v_j), f(v_{j+1})$ and $g(v_j), g(v_{j+1})$, conditioned on $f(v_i) = g(v_i) = x_i$ $(i = 1, \ldots, j-1)$, are mutually absolute continuous.

Case 1. v_j and v_{j+1} are adjacent. When conditioned on $f(v_1), \ldots, f(v_{j-1})$, the vectors $f(v_j)$ and $f(v_{j+1})$ are independent, so it does not mater in which order they are selected.

Case 2. v_j and v_{j+1} are not adjacent, but they are joined by a path that lies entirely in V_{j+1}. Let P be a shortest such path and t be its length (number of edges), so $2 \leq t \leq j$. For fixed j, we argue by induction on t. Let v_i be any internal node of P. We transform σ to τ by the following steps:

(1) Swap v_i and v_j to get σ^1. Since this can be obtained by successive adjacent swaps among the first j elements, μ_σ and μ_{σ^1} are mutually absolute continuous by the induction hypothesis on j.

(2) Swap v_i and v_{j+1}, to get σ^2. By the induction hypothesis on t (or by Case 1), μ_{σ^2} and μ_{σ^1} are mutually absolute continuous.

(3) Swap v_{j+1} and v_j, to get σ^3. As in (1), μ_{σ^3} and μ_{σ^2} are mutually absolute continuous.

(4) Swap v_j and v_i, to get σ^4. As in (2), μ_{σ^4} and μ_{σ^3} are mutually absolute continuous.

(5) Swap v_{j+1} and v_i, to get τ. As in (1), μ_τ and μ_{σ^4} are mutually absolute continuous.

Thus μ_σ and μ_τ are mutually absolute continuous.

Case 3. There is no path connecting v_j to v_{j+1} in V_{j+1}. Let U_0 (resp. U_1) denote the set of nodes of V_{j-1} that are adjacent to v_j (resp. v_{j+1}). Then V_{j-1} has a partition $W_0 \cup W_1$ so that there is no edge between W_0 and W_1, $U_0 \subseteq W_0$ and $U_1 \subseteq W_1$. Furthermore, it follows that $V \setminus V_{j+1}$ is a cutset, whence $|V \setminus V_{j+1}| \geq n - d$ and so $j \leq d - 1$.

For $S \subseteq V_{j-1}$, let $L(S)$ denote the linear subspace of $\mathbb{R}^d$ generated by the vectors x_i, $i \in S$. Let $L_i = L(V_{j-1} \setminus U_i)$. Then the condition on $f(v_j)$ is that it must be in $L_0^\perp$, and $f(v_{j+1})$ must be in $L_1^\perp \cap f(v_j)^\perp$.

We claim that $L_0^\perp$ and $L_1^\perp$ are weakly orthogonal. It suffices to check that L_0 and L_1 are weakly orthogonal. But this is easy, since a vector $x \in L_0$ can be written as $x' + x''$, where $x' \in L(W_1)$ and $x'' \in L(W_0 \setminus U_0$. The projection of this to L_1 is obtained by keeping x'' and projecting x' onto $L(W_1 \setminus U_1$, which is in $L_0 \cap L_1$.

Furthermore, we claim that $\dim(L_0^\perp \cap L_1^\perp) \geq 2$. Indeed, $L_0^\perp \cap L_1^\perp = (L_0 + L_1)^\perp$, and so

$$\dim(L_0^\perp \cap L_1^\perp) = d - \dim(L_0 + L_1) \geq d - (j - 1) \geq 2.$$

So Lemma 9.3.6 applies, which completes the proof. □

9.3.3 Faithful orthogonal representations

An orthogonal representation is *faithful* if different nodes are represented by non-parallel vectors and adjacent nodes are represented by non-orthogonal vectors.

We do not know how to determine the minimum dimension of a *faithful* orthogonal representation. It was proved by Maehara (1987) that if the maximum degree of the complementary graph $\overline{G}$ of a graph G is D, then G has a faithful orthogonal representation in D^3 dimensions. He conjectured that the bound on the dimension can be improved to $D + 1$. In other words,

Conjecture 9.3.7 If every node of G has degree at least $n - d$ (where n is the number of nodes), then G has a faithful orthogonal representation in $\mathbb{R}^d$.

We derive from the results in Section 9.3.2 that Maehara's conjecture is true if we strengthen its assumption by requiring that G is $(n - D - 1)$-connected.

Theorem 9.3.8 *Every $(n-d)$-connected graph on n nodes has a faithful general position orthogonal representation in $\mathbb{R}^d$.*

Proof. It suffices to show that in a random sequential orthogonal representation, the probability of the event that two nodes are represented by parallel vectors, or two adjacent nodes are represented by orthogonal vectors, is 0. By the Lemma 9.3.5, it suffices to prove this for the representation obtained from an ordering starting with these two nodes. But then the assertion is obvious. □

This implies the following result of Rödl (1987), improving the bound of Maehara:

Corollary 9.3.9 *If the maximum degree of the complementary graph $\overline{G}$ of a graph G is D, then G has a faithful orthogonal representation in $2D+1$ dimensions.*

Indeed, a graph with minimum degree $n-D-1$ is at least $(n-2D)$-connected.

9.3.4 Orthogonal representations with the Strong Arnold Property

The Strong Arnold Property. Consider an orthogonal representation $i \mapsto v_i \in \mathbb{R}^d$ of a graph G. We can view this as a point in the $\mathbb{R}^{d|V|}$, satisfying the quadratic equations

$$v_i^{\mathsf{T}} v_j = 0 \qquad (ij \in \overline{E}). \tag{9.22}$$

Each of these equation defines a hypersurface in $\mathbb{R}^{d|V|}$. We say that the orthogonal representation $i \mapsto v_i$ has the *Strong Arnold Property* if the surfaces (9.22) intersect transversally at this point. This means that their normal vectors are linearly independent.

This can be rephrased in more explicit terms as follows. For each nonadjacent pair $i, j \in V$, form the $d \times V$ matrix $V^{i,j}$ in which the i-th column is v_j, the j-th column is v_i, and the rest if 0. Then the Strong Arnold Property says that the matrices $V^{i,j}$ are linearly independent.

Another way of saying this is that there is no symmetric $V \times V$ matrix $X \neq 0$ such that $X_{ij} = 0$ if $i = j$ or $ij \in E$, and

$$\sum_j X_{ij} v_j = 0 \tag{9.23}$$

for every node i.

Algebraic width: definition. Colin de Verdière [45] introduced an interesting graph invariant related to connectivity. Let d be the smallest dimension in which G has a faithful orthogonal representation with the Strong Arnold Property, and define $a(G) = n - d$. We call $a(G)$ the *algebraic width* of the graph.

This can be rephrased in terms of matrices, making it more similar to the other Colin de Verdiére number. We consider a matrix $N \in \mathbb{R}^{V \times V}$ with the following properties:

(N1) $N_{ij} \begin{cases} = 0 & \text{if } ij \notin E,\ i \neq j;, \\ \neq 0 & \text{if } ij \in E. \end{cases}$

(N2) N is positive semidefinite;

(N3) [Strong Arnold Property] If X is a symmetric $n \times n$ matrix such that $X_{ij} = 0$ whenever $i = j$ or $ij \in E$, and $NX = 0$, then $X = 0$.

Lemma 9.3.10 *The algebraic width of a graph G is the maximum corank of a matrix with properties* (N1)–(N3).

The proof is quite straightforward, and is left to the reader.

Example 9.3.11 (Complete graphs) The Strong Arnold Property is void, and every representation is orthogonal, so $a(K_n) = n - 1$.

Example 9.3.12 (Paths)

Basic properties. The main advantage of this nondegeneracy condition is that it implies the following.

Lemma 9.3.13 *The graph parameter $a(G)$ is minor-monotone.*

The parameter has other nice properties, of which the following will be relevant:

Lemma 9.3.14 *Let G be a graph and B a complete subgraph. Let $G_1, \ldots, G_k$ be the connected components of $G \setminus B$, and let H_i be the subgraph induced by $V(G_i) \cup B$. Then*

$$a(G) = \max_i a(H_i).$$

Algebraic width, tree-width and connectivity. Tree-width is a parameter related to connectivity, introduced by Robertson and Seymour [168] as an important element in their graph minor theory (see [136] for a survey). Colin de Verdière [45] defines the *tree-width* of a graph G as the smallest r for which G is a minor of a Cartesian sum $K_r \oplus T$, where T is a tree. (This is not quite the same as the more standard notion of tree-width introduced by Robertson and Seymour, but the difference is at most 1, as shown by van der Holst [96]). The parameter $\mathsf{tw}(G)$ is minor-monotone.

The *monotone connectivity* $\kappa_{\mathrm{mon}}(G)$ of a graph G is defined as the maximum connectivity of any minor of G. It is easy to see that $\kappa_{\mathrm{mon}}(G) \leq \mathsf{tw}(G)$. The algebraic width is sandwiched between these two parameters:

Theorem 9.3.15 *For every graph,*

$$\kappa_{\mathrm{mon}}(G) \leq a(G) \leq \mathsf{tw}(G).$$

The upper bound was proved by Colin de Verdière [45], while the lower bound follows easily from the results in Section 9.3.2.

Proof. Since every general position orthogonal representation has the Strong Arnold Property, we get by Theorem 9.3.8 that $a(G) \geq \kappa(G)$. Since $a(G)$ is minor-monotone, this implies that $a(G) \geq \kappa_{\mathrm{mon}}$.

To prove the second inequality, we use again the minor-monotonicity of $a(G)$ to notice that it suffices to prove that

$$a(K_r \oplus T) \leq r \tag{9.24}$$

for any $r \geq 1$ and tree T. Note that $a(K_r \oplus T) \geq a(K_r) = r - 1$ follows by the minor monotonicity of r.

We use induction on $|V(T)|$. If T is a single node, then $K_r \oplus T$ is a complete r-graph and the assertion is trivial.

The most complicated case is when T has two nodes. We have to prove then that $K_2 \oplus T$ has no orthogonal representation in $\mathbb{R}^{r-1}$ with the Strong Arnold Property. Suppose that $i \mapsto v_i$ is such a representation.

Let $\{a_1, \ldots, a_r\}$ and $\{b_1, \ldots, b_r\}$ be the two r-cliques corresponding to the two nodes of T, where a_i is adjacent to b_i. For $S \subseteq [r]$, define $U_S = \{v_{a_i} : i \in S\}$ and $W_S = \{v_{b_i} : i \in S\}$.

If there is a subset $S \subseteq [r]$ such that both U_S and $W_{[r]\setminus S}$ are linearly independent, then (since these sets are mutually orthogonal) we get

$$\mathrm{rk}(U_S \cup W_{[r]\setminus S}) = \mathrm{rk}(U_S) + \mathrm{rk}(W_{[r]\setminus S}) = |U_S| + |W_{[r]\setminus S}| = r,$$

which is a contradiction since $U_S \cup W_{[r]\setminus S} \subseteq \mathbb{R}^{d-1}$.

So we may assume that no such subset S exists. Then by the Matroid Sum Theorem, there is a subset $T \subseteq [q]$ such that $\mathrm{rk}(U_T) + \mathrm{rk}(W_T) < |T|$. We may assume that $T = [r]$, since we can delete the nodes a_j and b_j with $j \notin T$. We may also assume (by scaling the v_{a_i}) that there is a vector $y \in \mathrm{lin}(U_T)$ such that $y^{\mathsf{T}} v_{a_i} = 1$ nd similarly there is a vector $z \in \mathrm{lin}(W_T)$ such that $z^{\mathsf{T}} v_{b_i} = 1$.

Consider the matrices V_{a_i,b_j} defined above, for $i, j \in T$. These matrices must be linearly independent. On the other hand, the r columns corresponding to the nodes a_i are all from a space with dimension $\mathrm{rk}(W_T)$, and the other r columns are from a space with dimension $\mathrm{rk}(U_T)$. So all these matrices belong to a linear space with dimension $r\mathrm{rk}(W_T) + r\mathrm{rk}(U_T) \leq r(r-1)$. Since the number of matrices V_{a_i,b_j} $(i, j \in T)$ is exactly $r(r-1)$, the must span the space. However, this is impossible, since the matrix $Y = (z, \ldots, z, -y, \ldots, -y)$ also belongs to this space and

$$Y \cdot V_{i,j} = z^{\mathsf{T}} v_{b_j} - y^{\mathsf{T}} v_{a_i} = 1 - 1 = 0.$$

This concludes the proof of the case $|V(T)| = 2$.

Finally, if T has more than two nodes, then it has an internal node v, and the clique corresponding to this node separates the graph. So 9.24 follows by Lemma 9.3.14 and induction. □

Colin de Verdière conjectured that equality holds in the upper bound in Theorem 9.3.15. This was proved by Van der Holst [96] and Kotlov [117] for $a(G) \leq 2$, but the following example of Kotlov shows that it is false in general.

Example 9.3.16 The k-cube Q^k has $a(Q^k) = O(2^{k/2})$ but $\mathsf{tw}(Q^k) = \Theta(2^k)$.

It is not known whether $\mathsf{tw}(G)$ can be larger than $a(G)^2$.

9.4 The variety of orthogonal representations

Let G be a k-connected graph with n nodes and set $d = n - k$. Then we know that G has a general position orthogonal representation in $\mathbb{R}^d$. One may suspect that more is true: every orthogonal representation in $\mathbb{R}^d$ is the limit of general position orthogonal representations, i.e., the set $\mathcal{OR}_d(G)$ of all orthogonal representations is the closure of the set $\mathcal{GOR}_d(G)$ of general position orthogonal representations of G. The following example shows that this is not true in general (but it can be proved under additional hypotheses about the graph G).

Example 9.4.1 Let Q denote the graph of the (ordinary) 3-dimensional cube, and let $G = \overline{Q}$. Note that Q is bipartite; let $U = \{u_1, u_2, u_3, u_4\}$ and $W = \{w_1, w_2, w_3, w_4\}$ be its color classes. The indices are chosen so that u_i is adjacent to v_i in G.

Since G is 4-connected, it has a general position orthogonal representation in $\mathbb{R}^4$. This is in fact easy to construct. Choose any four linearly independent vectors $x_1, \ldots, x_4$ to represent $u_1, \ldots, u_4$, then the vector y_i representing w_i is uniquely determined (up to scaling) by the condition that it has to be orthogonal to the three vectors x_j, $j \neq i$.

There is another type of orthogonal representation in which the elements of U are represented by vectors $x_1, \ldots, x_4$ in a 2-dimensional subspace L of $\mathbb{R}^4$ and the elements of W are represented by vectors $y_1, \ldots, y_4$ in the orthogonal complement $L^\perp$ of L. Clearly, this is an orthogonal representation for any choice of the representing vectors in these subspaces.

We claim that such a representation is the limit of general position orthogonal representations only if the cross ratio $(x_1 : x_2 : x_3 : x_4)$ is equal to the cross ratio of $(y_1 : y_2 : y_3 : y_4)$. (The cross ratio $(x_1 : x_2 : x_3 : x_4)$ of four vectors x_1, x_2, x_3, x_4 in a 2-dimensional subspace can be defined as follows: we write $x_3 = \lambda_1 x_1 + \lambda_2 x_2$ and $x_4 = \mu_1 x_1 + \mu_2 x_2$, and take $(x_1 : x_2 : x_3 : x_4) = (\lambda_1 \mu_2)/(\lambda_2 \mu_1)$. This number is invariant under linear transformations.)

To prove this claim, we consider any general position orthogonal representation g of G. Let M be the linear subspace spanned by the vectors $g(u_1)$ and $g(u_2)$. Then its orthogonal complement $M^\perp$ is spanned by $g(v_3)$ and $g(v_4)$. Let b_i be the orthogonal projection of $g(u_i)$

onto M $(i = 1, 2, 3, 4)$ and c_i, the orthogonal projection of $g(v_i)$ onto $M^\perp$ $(i = 1, 2, 3, 4)$. If we show that $(b_1 : b_2 : b_3 : b_4) = (c_1 : c_2 : c_3 : c_4)$, then the claim follows since $b_i \to x_i$ and $c_i \to y_i$.

The proof of $(b_1 : b_2 : b_3 : b_4) = (c_1 : c_2 : c_3 : c_4)$ is an exercise in linear algebra (or projective geometry) and is left to the reader.

One reason for asking the question whether $\mathcal{GOR}_d(G)$ is dense in $\mathcal{OR}_d(G)$ is the following. The set $\mathcal{OR}_d(G)$ is an algebraic variety in $\mathbb{C}^d$, and it is a natural question whether it is irreducible. (A set $A \subseteq \mathbb{C}^N$ is *irreducible* if whenever the product $p \cdot q$ of two polynomials in N variables vanishes on A, then either p or q vanishes on A; equivalently, the polynomial ideal $\{p : \ p \text{ vanishes on } A\}$ is a prime ideal.)

Let us begin with the question of irreducibility of the set $\mathcal{GOR}_d(G)$ of general position orthogonal representations of G. This can be settled quite easily.

Theorem 9.4.2 *Let G be any graph and $d \geq 1$. Then $\mathcal{GOR}_d(G)$ is irreducible.*

Proof. Let G have n nodes. We may assume that G is $(n - d)$-connected, else $\mathcal{GOR}_d(G)$ is empty and the assertion is vacuously true.

First we show that there exist vectors $\mathbf{x}_v = \mathbf{x}_v(X) \in \mathbb{R}[x]^d$ $(v \in V)$ whose entries are multivariate polynomials with real coefficients in variables $X_1, X_2, \ldots$ (the number of these variables does not matter) such that whenever u and v are non-adjacent then $\mathbf{x}_u \cdot \mathbf{x}_v$ is identically 0, and such that every general position representation of G arises from $\mathbf{x}$ by substitution for the variables X_i. We do this by induction on n.

Let $v \in V$. Suppose that the vectors of polynomials $\mathbf{x}_u(X')$ of length d exist for all $u \in V - \{v\}$ satisfying the requirements above for the graph $G - v$ (since $G - v$ has $n - 1$ nodes and is $(n - d - 1)$-connected, this is indeed the right induction hyposthesis). Let $\mathbf{x}' = (\mathbf{x}_u : \ u \in V - v)$. Let $u_1, \ldots, u_m$ be the nodes in $V - \{v\}$ non-adjacent to v; clearly $m \leq d - 1$. Let $x_1, \ldots, x_d - 1 - m$ be vectors of length d composed of new variables and let y be another new variable; X will consist of X' and these new variables. Consider the $d \times (d - 1)$ matrix

$$F = (\mathbf{x}_{u_1}, \ldots, \mathbf{x}_{u_m}, x_1, \ldots, x_{d-1-m}),$$

and let p_j be the determinant of the submatrix obtained by dropping the j-th row. Then we define $\mathbf{x}_v = y(p_1, \ldots, p_d)^T$.

It is obvious from the construction and elementary linear algebra that $\mathbf{x}_v$ is orthogonal to every vector $\mathbf{x}_u$ for which u and v are non-adjacent. We show that every general position orthogonal representation of G can be obtained from $\mathbf{x}$ by substitution. In fact, let f be a general position orthogonal representation of G. Then $f' = f|_{V-v}$ is a general position orthogonal representation of $G - v$ in $\mathbb{R}^d$, and hence by the induction hypothesis, f' can

be obtained from $\mathbf{x}'$ by substituting for the variables X'. The vectors $f(u_1), \ldots, f(u_m)$ are linearly independent and orthogonal to $f(v)$; let $a_1, \ldots, a_{d-1-m}$ be vectors completing the system $f(u_1), \ldots, f(u_m)$ to a basis of $f(v)^\perp$. Substitute $x_i = a_i$. Then the vector $(p_1, \ldots, p_d)^T$ will become a non-zero vector parallel to $f(v)$ and hence y can be chosen so that $\mathbf{x}_v$ will be equal to $f(v)$.

Note that from the fact that $\mathbf{x}(X)$ is in general position for some substitution it follows that the set of substitutions for which it is in general position is everywhere dense.

From here the proof is quite easy. Let p and q be two polynomials such that $p \cdot q$ vanishes on $\mathcal{GOR}_d(G)$. Then $p(\mathbf{x}(X)) \cdot q(\mathbf{x}(X))$ vanishes on an everywhere dense set of substitutions for X, and hence it vanishes identically. So either $p(\mathbf{x}(X))$ or $q(\mathbf{x}(X))$ vanishes identically; say the first occurs. Since every general position orthogonal representation of G arises from $\mathbf{x}$ by substitution, it follows that p vanishes on $\mathcal{GOR}_d(G)$. □

The proof in fact gives more. Let $\mathcal{LGOR}_d(G)$ denote the set of locally general position orthogonal representations of G in $\mathbb{R}^d$.

Corollary 9.4.3 *Every locally general position orthogonal representation of G can be obtained from $\mathbf{x}$ by substitution. Hence $\mathcal{LGOR}_d(G)$ is irreducible, and $\mathcal{GOR}_d(G)$ is dense in $\mathcal{LGOR}_d(G)$.*

Now back to the (perhaps) more natural question of the irreducibility of $\mathcal{OR}_d(G)$.

Proposition 9.4.4 *If G is not $(n-d)$-connected, then $\mathcal{OR}_d(G)$ is not irreducible.*

Proof. If G is not $(n-d)$-connected, then V has a partition $V_0 \cup V_1 \cup V_2$ such that $|V_0| = n-d-1$ and there is no edge between V_1 and V_2. Let $d_r = |V_r|$, then $d_1 + d_2 = d+1$.

Let $i \mapsto v_i$ be any orthogonal representation of G, and let A_r be the Gram matrix of the vectors $\{v_i : i \in V_r\}$. As in the proof of Theorem 9.3.3, either $\{v_i : i \in V_1\}$ or $\{v_i : i \in V_2\}$ must be linearly dependent, and hence $\det(A_1)\det(A_2) = 0$ for every orthogonal representation in $\mathbb{R}^d$.

On the other hand, $\det(A_1)$ is not zero for all orthogonal representations. Indeed, selecting mutually orthogonal unit vectors in $\mathbb{R}^d$ for $\{v_i : i \in V_1\}$ (which is possible as $d_1 \le d$, and $v_i = 0$ for $i \in V \setminus V_1$, we get an orthogonal representation with $\det(A_1) = 1$. Similarly, $\det(A_2)$ is not always zero on $\mathcal{OR}_d(G)$. □

In the other direction, theorem 2.1 implies the following.

Lemma 9.4.5 *If $\mathcal{GOR}_d(G)$ is dense in $\mathcal{OR}_d(G)$ then $\mathcal{OR}_d(G)$ is irreducible.*

We know that $\mathcal{GOR}_d(G)$ is nonempty if and only of G is $(n-d)$-connected. For $d \le 3$, more is true:

Theorem 9.4.6 *Let G be an $(n-d)$-connected graph with n nodes.*

(a) If $d \leq 3$, then $\mathcal{GOR}_d(G)$ is everywhere dense in $\mathcal{OR}_d(G)$.

(b) If $d = 4$, and the complement of G is connected and non-bipartite, then $\mathcal{GOR}_d(G)$ is everywhere dense in $\mathcal{OR}_d(G)$.

On the other hand, for $d = 4$, not every $(n-d)$-connected graph gives an irreducible variety, but those graphs whose complement do. The general description of all graphs with $\mathcal{OR}_d(G)$ irreducible is open.

Let us say that an orthogonal representation of a graph G in $\mathbb{R}^d$ is *special* if it does not belong to the topological closure of $\mathcal{GOR}_d(G)$. A graph G is *d-critical* if it has a special orthogonal representation in $\mathbb{R}^d$, but no induced subgraph of it does so. (Observe that if a graph has no special representation in $\mathbb{R}^d$, then neither do its induced subgraphs: a special orthogonal representation of an induced subgraph can be extended with 0 vectors to a special orthogonal representation of G.)

The following lemma shows that special orthogonal representations are just the opposite of being in locally general position.

Lemma 9.4.7 *For every special orthogonal representation of a d-critical graph in $\mathbb{R}^d$ the non-neighbors of every node are linearly dependent.*

Proof. Call a node *good* if its non-neighbors are represented by linearly independent vectors. Let $i \mapsto v_i$ be a special orthogonal representation of G that has a good node i. We construct a general position orthogonal representation of G in an arbitrarily small neighborhood of v. Let $m = n - |N(i)| - 1 \leq d-1$, and extend $\{v_j : j \notin N(i)\}$ by arbitrary vectors $a_1, \ldots, a_{d-1-m}$ to a basis B of $\mathbb{R}^d$. Let, say,

$$\det(B) = 1.$$

Now v', obtained by restricting v to $V(G-i)$, is an orthogonal representation of $G-i$, and hence by the criticality of G, there exists a general position orthogonal representation u of $G-i$ in $\mathbb{R}^d$ in an arbitrarily small neighborhood of v'. We extend u to an orthogonal representation u^* of G as follows. Clearly if u is close enough to v', then the vectors $\{u_j : j \in V \setminus N(i) \setminus \{i\}\} \cup \{a_1, \ldots, a_{d-1-m}\}$ are linearly independent and hence they uniquely determine a vector u_i^* orthogonal to all of them such that the basis $\{u_j : j \in \overline{N}(i)\} \cup \{a_1, \ldots, a_{d-1-m}, u_i^*\}$ has determinant 1.

It is obvious that u^* is an orthogonal representation of G and that if u is close enough to v', then u^* is arbitrarily close to v. Unfortunately, it does not follow in general that this extended u^* is in general position; but at least every "good" node remains "good" (if v' and u are close enough). Moreover, we know that any d vectors representing nodes different from i are linearly independent; in particular, every node adjacent to i is now "good".

Now if j is any other "good" node, then we can repeat the same argument and find an orthogonal representation u^{**} very close to u^* in which every node previously good remains good, and in addition all the neighbors of j become good. Since G is connected, repeating this argument at most n times we obtain an orthogonal representation w of G arbitrarily close to v in which every node is "good", i.e., which is is locally general position. By the remark above, such a representation is in the closure of $\mathcal{GOR}_d(G)$, and hence we find arbitrarily close to it a general position orthogonal representation f^* of G. □

If f is a special representation of G in $\mathbb{R}^d$ then, by definition, there exists an $\varepsilon > 0$ such that if g is another orthogonal representation of G in $\mathbb{R}^d$ and $|f - g| < \varepsilon$ then g is also special. There must be linear dependencies among the vectors $f(v)$; if ε is small enough then there will be no new dependencies among the vectors $g(v)$. We say that an orthogonal representation f is *locally free-est* if there exists an $\varepsilon > 0$ such that for every orthogonal representation g with $|f - g| < \varepsilon$, and every subset $U \subseteq V$, $f(U)$ is linearly dependent iff $g(U)$ is. Clearly every graph having a special representation in $\mathbb{R}^d$ also has a locally free-est special representation.

Lemma 9.4.8 *In a locally free-est special representation of a d-critical graph, any two representing vectors are linearly independent.*

Proof. Let f be a locally free-est special orthogonal representation of the given graph G. First assume that $f(v)$ is the 0 vector for some v. Since the number of non-neighbors of v is at most $d - 1$, we can replace $f(v)$ by an arbitrarily short non-zero vector orthogonal to all vectors $f(u)$ where u is a non-neighbor of v, and shorter then ε. This is clearly another special orthogonal representation in $\mathbb{R}^d$ with fewer 0 vectors, a contradiction.

Second, let v and w be two nodes with $f(v)$ and $f(w)$ parallel. By Lemma 9.4.7, the set of vectors $f(u)$, where u is a non-neighbor of v, is linearly dependent, and hence these vectors span a linear subspace L of dimension at most $d - 2$. Thus there exists a vector $a \in L^\perp$ not parallel to $f(v)$. We can replace $f(v)$ by $f(v) + \delta a$, and obtain another orthogonal representation of G. If δ is small enough, then this new representation is also special and it has fewer pairs of parallel representing vectors than f, a contradiction again. □

Corollary 9.4.9 *If G is a d-critical graph with n nodes then every node has degree at most $n - 4$.*

Proof. Let f be a locally free-est special orthogonal representation of G in $\mathbb{R}^d$. If a node v has degree $n - 3$ then let u_1 and u_2 be the two non-neighbors of v. If $f(u_1)$ and $f(u_2)$ are parallel then we have a contradiction at Lemma 9.4.8. If they are not parallel, we have a contradiction at Lemma 9.4.8. If the degree of v is larger than $n-3$, the argument is similar. □

Let f be an orthogonal representation of G and $v \in V$. Let A_v be the linear span of the non-neighbors of v and B_v, its orthogonal complement. So $f(v) \in B_v$.

Lemma 9.4.10 *Let G be a d-critical graph and f, a locally free-est special representation of G. Let $v \in V$ and u, a non-neighbor of v such that $f(u)$ is linearly dependent on the vectors representing the other non-neighbors. Then $B_u \subseteq A_v$.*

Proof. Suppose not, then B_u contains a vector b arbitrarily close to $f(u)$ but not in A_v. Then replacing $f(u)$ by b we obtain another orthogonal representation f' of G. Moreover, b does not depend linearly on the vectors representing the other non-neighbors of v, which contradicts the definition of locally free-est special representations. □

Next we turn to the case $d = 4$.

Theorem 9.4.11 *If G is a 4-critical graph then $\overline{G}$ is 3-regular and bipartite.*

Proof. Let G have n nodes. Then Corollary 9.4.9 implies that it is regular of degree $n - 4$, i.e., $\overline{G}$ is 3-regular. Consider the subspaces A_v and B_v defined above. Lemma 9.4.7 implies that $\dim A_v \leq 2$ and Lemma 9.4.8 implies that $\dim A_v \geq 2$, so $\dim A_v = 2$, and hence also $\dim B_v = 2$. Thus Lemma 9.4.10 implies that for any two non-adjacent nodes u and v, $A_u = B_v$. So, fixing any node v, the rest of the nodes fall into two classes: those with $A_u = A_v$ and those with $A_u = B_v$. Moreover, any edge in $\overline{G}$ connects nodes in different classes. Hence $\overline{G}$ is bipartite as claimed. □

We do not know if there is any other 4-critical graph. An analysis of the cases $d \geq 5$ seems even harder.

9.5 Related representations

Thresholds etc.

Chapter 10

Graph independence to linear independence

10.1 Cover-critical and independence-critical graphs

As a useful tool in the study of graphs critical with respect to stability number, Lovász [130] considered vector representations with the property that *every set of nodes that covers all the edges, spans the space.* We call this *cover-preserving.*

One can dualize this notion (see section 13.2), to get the following condition on vector representations: *every independent set of nodes is represented by linearly independent vectors.* Such representations are called *independence-preserving.*

This looks like a play with words, but in fact such representations turn out very useful. Obviously, every orthogonal representation has this property.

In this case, the dimension problem is trivial: an independence preserving representation exists in dimension $\alpha(G)$ and higher, and a cover preserving representation exists in dimension $\tau(G)$ and lower. For example, we can consider the *generic representation* in the given dimension d, in which every d vectors are linearly independent.

But these representations become interesting in conjunction with criticality. We say that a cover-preserving representation is *cover-critical*, if deleting any edge from the graph, the representation does not remain cover-preserving.

Lemma 10.1.1 (a) *Let* $(v_i : i \in V(G))$ *be a cover preserving representation of a graph* G. *For some* $j \in V(G)$, *let* v'_j *be a generic point in* $\mathrm{lin}\{v_i : i \in N(j)\}$. *Then replacing* v_j *by* v'_j *we get a cover preserving representation.*

(b) *If, in addition,* $(v_i : i \in V(G))$ *is cover-critical and* v_j *is a generic point of the space, then replacing* v_j *by* v'_j *we get a cover-critical representation.*

Proof.

□

Lemma 10.1.2 (a) *The projection of a cover preserving representation onto any subspace is cover preserving.*

(b) *If the kernel of the projection is a vector v_i that is contained in* $\text{lin}\{v_j : j \in N(i)\}$, *then* $(v_j : j \in V(G) \setminus \{i\})$ *is a cover-critical representation of* $G - i$.

Proof.

□

Lemma 10.1.3 *In a cover-critical representation, the neighbors of any node are represented by linearly independent vectors.*

Proof. Let $(v_i : i \in V(G))$ be a cover-critical representation of a graph G n dimension d. Consider any node a and a neighbor $b \in N(i)$. The graph $G \setminus ab$ has a node-cover T that does not span the space R^d. The vectors v_a and v_b cannot belong to the linear span of T, since otherwise $T \cup \{a, b\}$ would be a nonspanning node-cover of G. In particular, T does not contain a, and hence it must contain all neighbors of a other than b. So v_b is not spanned by the vectors v_j, $j \in N(a)$. Since b is an arbitrary element of $N(a)$, it follows that the vectors representing $N(a)$ are linearly independent. □

Corollary 10.1.4 *If G has a cover-critical representation in dimension d, then every node has degree at most d.*

Theorem 10.1.5 *If G has a cover-critical representation in dimension d, then* $|E(G)| \leq \binom{d+1}{2}$.

Proof. Let $(v_i : i \in V(G))$ be a cover-critical representation of a graph G n dimension d. Consider the matrices $M_{ij} = v_i \wedge v_j = v_i v_j^{\mathsf{T}} + v_j v_i^{\mathsf{T}}$ $(ij \in E)$.

We claim that these matrices are linearly independent. Suppose that there is a linear dependence

$$\sum_{ij \in E(G)} \lambda_{ij} M_{ij} = 0, \tag{10.1}$$

where $\lambda_{ab} \neq 0$ for some edge ab. The graph $G \setminus ab$ has a node-cover T that does not span the space R^d; let $u \in \mathbb{R}^d$ be orthogonal to the span of T. Clearly $u^{\mathsf{T}} v_a \neq 0$ and $u^{\mathsf{T}} v_b \neq 0$, else $\{j \in V(G) : u^{\mathsf{T}} v_j\} = 0$ would cover all edges of G, and the representations would not be cover preserving. Then

$$u^{\mathsf{T}} M_{ij} u = u^{\mathsf{T}} v_i v_j^{\mathsf{T}} u + u^{\mathsf{T}} v_j v_i^{\mathsf{T}} u = 2(u^{\mathsf{T}} v_i)(u^{\mathsf{T}} v_j) = 0$$

for every edge $ij \neq ab$ (since one of i and j must be in T), but

$$u^{\mathsf{T}} M_{ab} u = 2(u^{\mathsf{T}} v_a)(u^{\mathsf{T}} v_b) \neq 0.$$

This contradicts (10.1).

Thus the matrices M_{ij} are linearly independent. Hence their number cannot be larger than the dimension of the space of symmetric $d \times d$ matrices, which is $\binom{d+1}{2}$. □

Remark 10.1.6 Representing the nodes of a complete graph on $d+1$ nodes by vectors in $\mathbb{R}^d$ in general position, we get a cover-critical representation, showing that the bound in Theorem 10.1.5 is tight.

What about the number of nodes? Let $(v_i : i \in V(G))$ be a cover-critical representation of a graph G in dimension d. An isolated node plays no role, and can be omitted. (A node represented by the 0 vector is necessarily isolated.) After deleting isolated nodes, at most $d(d+1)$ nodes remain by Theorem 10.1.5.

One cannot say more in general, as the following construction shows. Let us split each node i of G into $d(i)$ nodes of degree 1, to get a graph G' consisting of independent edges. Keep the vector representing i for each of the corresponding new node (or replace it by a parallel vector). Then we get another cover-critical representation of G'. (Conversely, if we identify nodes that are represented by parallel vectors in a cover-critical representation, we get a cover-critical representation of the resulting graph.)

Chapter 11

Metric embeddings

Given a graph, we would like to embed it in a euclidean space so that the distances between nodes in the graph should be the same, or at least close to, the geometric distance of the representing vectors. It is not hard to see that one will necessarily have some distortion in non-trivial cases. For example, the "claw" $K_{1,3}$ cannot be embedded without distortion in any dimension. Furthermore, the complete k-graph can be embedded without distortion in dimension $k-1$ or more. So we are interested in two parameters: the dimension and the distortion.

These results are best stated in the generality of finite metric spaces. Recall that a metric space is a set V endowed with a distance function $d: V \times V$ such that $d(u,v) = 0$ if and only if $u = v$, $d(v,u) = d(u,v)$, and $d(u,w) \le d(u,v) + d(v,w)$ for all $u, v, w \in V$.

Let $F: V_1 \to V_2$ be a mapping of the metric space (V_1, d_1) into the metric space (V_2, d_2). We define the *distortion* of F as

$$\max_{u,v \in V_1} \frac{d_2(F(u),F(v))}{d_1(u,v)} \Big/ \min_{u,v} \frac{d_2(F(u),F(v))}{d_1(u,v)}.$$

Note that to have finite distortion, the map F must be injective. The distortion does not change if all distances in one of the metric spaces are scaled by the same factor. So if we are looking for embeddings in a Banach space, then we may consider embeddings that are contractive, i.e., $d_2(F(u),F(v)) \le d_1(u,v)$ for all $u, v \in V_1$.

11.1 Embeddings in low dimension

Often, the dimension problem is easy to handle, due to a fundamental lemma [105]:

Lemma 11.1.1 (Johnson–Lindenstrauss) *For every $0 < \varepsilon < 1$, every n-point set $S \subset \mathbb{R}^n$ can be mapped into $\mathbb{R}^d$ with $d < 60(\ln n)/\varepsilon^2)$ with distortion $1+\varepsilon$.*

Proof. Orthogonal projection onto a random d-dimensional subspace does the job.

First, let us see what happens to the distance of a fixed pair of points $x, y \in S$. We may assume that $y = 0$ and $|x| = 1$. Instead of projecting a fixed unit vector on a random subspace, we project a random unit vector X on a fixed subspace, say on the subspace of $\mathbb{R}^n$ in which the last $n - d$ coordinates are 0.

We can generate a random unit vector by generating n independent standard Gaussian variables $X_1, \ldots, X_n$, and normalizing:

$$X = \frac{1}{|X_1 + \cdots + X_n|}(X_1, \ldots, X_n).$$

The squared length of the projection is

$$|X'|^2 = \frac{X_1^2 + \cdots + X_d^2}{X_1^2 + \cdots + X_n^2}.$$

Here the numerator is concentrated around d, and the denominator is concentrated around n. Hence the length of X' is concentrated around $\sqrt{d/n}$. In fact, both the numerator and the denominator are from a chi-squared distribution with parameters d and n, and standard arguments give that

$$\mathsf{P}(||X_1|^2 + \cdots + |X_n|^2 - n| > \varepsilon n) \le 2e^{-\varepsilon^2 n/3},$$

and similarly

$$\mathsf{P}(||X_1|^2 + \cdots + |X_d|^2 - d| > \varepsilon d) \le 2e^{-\varepsilon^2 d/3}.$$

So with probability at least $1 - 4e^{-\varepsilon^2 d/3}$, we have

$$\frac{1-\varepsilon}{1+\varepsilon} \cdot \frac{d}{n} \le \frac{|X_1|^2 + \cdots + |X_d|^2}{|X_1|^2 + \cdots + |X_n|^2} \le \frac{1+\varepsilon}{1-\varepsilon} \cdot \frac{d}{n}.$$

Going back to the original setup, we get that with probability at least $1 - 4e^{-\varepsilon^2 d/3}$, we have

$$\sqrt{\frac{1-\varepsilon}{1+\varepsilon}}\sqrt{\frac{d}{n}} \le \frac{|F(x) - F(y)|}{|x - y|} \le \sqrt{\frac{1+\varepsilon}{1-\varepsilon}}\sqrt{\frac{d}{n}}.$$

It follows that with probability at least $1 - \binom{n}{2} 4e^{-\varepsilon^2 d/3}$, this holds for all pairs $x, y \in S$, and then the distortion of the projection is at most

$$\frac{1+\varepsilon}{1-\varepsilon} < 1 + 3\varepsilon.$$

Replacing ε by $\varepsilon/3$ and choosing d as in the Theorem, this probability is positive. □

11.2 Embeddings with small distortion

Bourgain [29] proved the following theorem:

Theorem 11.2.1 *Every metric space with n elements can be embedded in an $O(\log n)$-dimensional euclidean space with $O(\log n)$ distortion.*

Proof. First we ignore the bound on the dimension.

Let $m = \lceil \log n \rceil$, and choose an integer $k \in [m]$ uniformly at random. Choose a random subset $A \subseteq V$, putting every element $i \in V$ into A with probability 2^{-k}. Let $d(i, A)$ denote the distance of i from the set A (if $A = \emptyset$, we set $d(i, A) = 0$). By the triangle inequality

$$|d(i, A) - d(j, A)| \le d(i, j). \tag{11.1}$$

The key to the proof is the following reverse inequality:

$$\mathsf{E}|d(i, A) - d(j, A)| \ge \frac{1}{4m} d(i, j). \tag{11.2}$$

To prove this inequality, let $D(v, t)$ denote the set of 2^t nearest points to $v \in V$, and let $\rho_v(t) = \max\{d(v, x) : \ x \in D(v, t)\}$, and $\rho(t) = \max\{\rho_i(t), \rho_j(t)\}$. Let h be the largest integer for which $\rho(h) < d(i, j)/2$. Clearly $h \le m$ and $D(i, h) \cap D(j, h) = \emptyset$. Modify the definition of $D(i, h + 1)$ by deleting from it all points in $D(j, h)$, and vice versa, and also set $\rho(h + 1) = d(i, j)/2$.

We claim that for a fixed choice of $1 \le k \le h + 1$,

$$\mathsf{E}_A|d(i, A) - d(j, A)| \ge \frac{1}{10}(\rho(k) - \rho(k - 1)). \tag{11.3}$$

We may assume that $\rho_j(k) \le \rho_i(k)$, so that $\rho(k) = \rho_i(k)$. If A intersects $D(j, k-1)$ but does not intersect $D(i, k)$, then

$$d(i, A) - d(j, A) \ge \rho_i(k) - \rho_j(k - 1) \ge \rho(k) - \rho(k - 1)$$

(it is easy to check that this holds also for $k = h+1$). Furthermore, the probability that this happens can be bounded using that $D(i, k) \cap D(j, k - 1) = \emptyset$:

$$\begin{aligned}
\mathsf{P}(A \cap D(j, k - 1) \ne \emptyset, \ A \cap D(i, k) = \emptyset) &= \Big(1 - \Big(1 - \frac{1}{2^k}\Big)^{|D(j,k-1)|}\Big)\Big(1 - \frac{1}{2^k}\Big)^{|D(i,k)|} \\
&\le \Big(1 - \Big(1 - \frac{1}{2^k}\Big)^{2^{k-1}}\Big)\Big(1 - \frac{1}{2^k}\Big)^{2^k} \\
&\approx \Big(1 - \frac{1}{\sqrt{e}}\Big)\frac{1}{e} > \frac{1}{10}.
\end{aligned}$$

This proves (11.3).

Averaging over the random choice of k, we get

$$\begin{aligned}
\mathsf{E}|d(i, A) - d(j, A)| &= \sum_{k=1}^{m} \frac{1}{m} \mathsf{E}_A|d(i, A) - d(j, A)| \ge \sum_{k=1}^{h+1} \frac{1}{10m}(\rho(k) - \rho(k - 1)) \\
&= \frac{1}{10m}(\rho(h + 1) - \rho(0)) = \frac{1}{20m} d(i, j),
\end{aligned}$$

which proves (11.2).

Now let $A_1, \dots, A_N$ be independently generated as A above, and consider the embedding

$$F: \ i \mapsto \Big(\frac{1}{\sqrt{N}} d(i, A_1), \dots, \frac{1}{\sqrt{N}} d(i, A_N)\Big).$$

Then by (11.1)

$$|F(i) - F(j)| = \sqrt{\frac{1}{N} \sum_{k=1}^{N} (d(i, A_k) - d(j, A_k))^2} \le d(i, j),$$

so F is contractive. On the other hand, by the inequality between quadratic and arithmetic means,

$$|F(i) - F(j)| = \sqrt{\sum_{k=1}^{N} \frac{1}{N} (d(i, A_k) - d(j, A_k))^2} \ge \frac{1}{N} \sum_{k=1}^{N} |d(i, A_k) - d(j, A_k)|.$$

Here by the Law of Large Numbers and (11.2),

$$\frac{1}{N} \sum_{k=1}^{N} |d(i, A_k) - d(j, A_k)| \approx \mathsf{E}|d(i, A) - d(j, A)| \ge \frac{1}{20m} d(i, j),$$

where the relative error at the $\approx$ sign is less than 2 with probability at least $1 - n^{-2}$ if N is large enough. Thus with positive probability,

$$\frac{1}{40m} d(i, j) \le |F(i) - F(j)| \le d(i, j)$$

holds for every i and j, and so F has distortion at most $40m$.

The condition on the dimension can be satisfied by an application of the Johnson–Lindenstrauss Lemma. □

We note that essentially the same embedding works for any norm ℓ_p, $1 \le p \le \infty$.

Linial, London and Rabinovitch [125] showed how to construct an embedding satisfying the conditions in Theorem 11.2.1 algorithmically, and gave the application described in the next section. Matoušek [147] showed that for an expander graph this is best possible.

11.3 Application to multicommodity flows

A fundamental result in the theory of multicommodity flows is the theorem of Leighton and Rao [124]. Stated in a larger generality, as proved by Linial, London and Rabinovich [125], it says the following.

Suppose that we have a multicommodity flow problem on a graph on n nodes. Obvious cut-conditions provide a system of necessary conditions for the problem to be feasible; but (unlike for the case of a single commodity), these conditions are not sufficient in general. The theorem asserts that *if the cut-conditions are satisfied, then relaxing the capacities by a factor of $O(\log n)$, the problem becomes feasible.*

11.4 Volume-respecting embeddings

A very interesting extension and application of Bourgain's method was given by Feige [70]. Let $F: V \to \mathbb{R}^n$ be a mapping of a finite metric space (V, d) in to the euclidean space $\mathbb{R}^n$. We want an embedding that is

(a) contractive,

(b) volume respecting, which means that every set of at most s nodes spans a simplex whose volume is almost as large as possible.

Obviously, (b) needs an explanation. Consider any set $S \subseteq V$ with k elements. Let T be a shortest spanning tree on S with respect to the distance d, and

$$\mathsf{treevol}(S) = \frac{1}{k!} \prod_{uv \in E(T)} d(u, v).$$

Lemma 11.4.1 *For every contractive map $F: V \to \mathbb{R}^n$, the volume of the simplex spanned by $F(S)$ is at most* $\mathsf{treevol}(S)$.

Proof.

□

Feige proves the following theorem.

Theorem 11.4.2 *Every finite metric space (V, d) with n elements has a contractive map $F: V \to \mathbb{R}^d$ with $d = O((\log n)^3)$, such that for each set S with $|S| \leq \log n$, the volume of the simplex spanned by $F(S)$ is at least* $\mathsf{treevol}(S)/(\log n)^{2|S|}$.

Proof.

□

As an application of this result, we describe an algorithm that finds a polylogarithmic approximation of the bandwidth of a graph in polynomial time. The ordering of the nodes which approximates the bandwidth is now obtained through a random projection of the representation to the line, in a fashion similar to the Goemans–Williamson algorithm:

Theorem 11.4.3 *Let $G = (V, E)$ be a graph, and let $d(u, v)$ denote the distance of nodes u and v in the graph. Let $F: V \to \mathbb{R}^d$ be a map with $d = O((\log n)^3)$, such that for each set S with $|S| \leq \log n$, the volume of the simplex spanned by $F(S)$ is at least* $\mathsf{treevol}(S)/(\log n)^{2|S|}$. *Let $P: \mathbb{R}^d \to L$ be the orthogonal projection onto a random line L through the origin, and let $(i_1, \ldots, i_n)$ be the ordering of V according to the order of the projected nodes on the line. Then $(i_1, \ldots, i_n)$ approximates the bandwidth within a factor of $O((\log n)^5)$.*

Proof.

□

Chapter 12

Adjacency matrix and regularity partitions

12.1 Similarity metric

We can equip every graph $G = (V, E)$ with a metric as follows. Let A be the adjacency matrix of G. We define the *similarity distance* of two nodes $i, j \in V$ as the ℓ_1 distance of the corresponding rows of A^2 (squaring the matrix seems unnatural, but it is crucial; it turns out to get rid of random fluctuations).

12.2 Regularity partitions

The following was proved (in somewhat different form) in [140].

Theorem 12.2.1 *Let G be a graph and let $\mathcal{P} = \{V_1, \dots, V_k\}$ be a partition of V.*

(a) *If $d_\square(G, G_\mathcal{P}) = \varepsilon$, then there is a set $S \subseteq V$ with $|S| \le 8\sqrt{\varepsilon}|V|$ such that for each partition class, $V_i \setminus S$ has diameter at most $8\sqrt{\varepsilon}$ in the d_2 metric.*

(b) *If there is a set $S \subseteq V$ with $|S| \le \delta|V|$ such that for each partition class, $V_i \setminus S$ has diameter at most δ in the d_2 metric, then $d_\square(G, G_\mathcal{P}) \le 24\delta$.*

Theorem 12.2.1 suggests to define the *dimension* of a family $\mathcal{G}$ of graphs as the infimum of real numbers $d > 0$ for which the following holds: for every $\varepsilon > 0$ and $G \in \mathcal{G}$ the node set of G can be partitioned into a set of at most $\varepsilon|V(G)|$ nodes and into at most ε^{-d} sets of diameter at most ε. (This number can be infinite.) In the cases when the graphs have a natural dimensionality, this dimension tends to give the right value. For example, let G be obtained by selecting n random points on the d-dimensional unit sphere, and connecting two of these points x and y with a probability $W(x, y)$, which is a continuous function of x and y. With probability 1, this sequence has dimension $\Theta(d)$.

Chapter 13

Some general issues

Is there a way to fit the many forms of geometric representations discussed in these notes into a single theory? Perhaps not, considering the variety of possibilities how the graph structure can be reflected by the geometry. Nevertheless, there are some general ideas that can be pointed out.

13.1 Non-degeneracy

A common theme in connection with various representations is that imposing non-degeneracy conditions on the representation often makes it easier to analyze and therefore more useful (basically, by eliminating the possibility of numerical coincidence). There are at least 3 types of non-degeneracy conditions that are used in various geometric representation; we illustrate the different possibilities by formulating them in the case of unit distance representations in $\mathbb{R}^d$. All three are easily extended to other kinds of representations.

The most natural non-degeneracy condition is *faithfulness*: we want to represent the graph so that adjacent nodes *and only those* are at unit distance. This is usually not strong enough. *General position* means that no $d+1$ of the points are contained in a hyperplane.

Perhaps the deepest non-degeneracy notion is the following. We write the condition on the representation as a system of algebraic equations. For example, for unit distance representations we write

$$\|u_i - u_j\|^2 = 1 \qquad (ij \in E).$$

We have nd unknowns (the coordinates of the u_i). Each of these equations defines a hypersurface in $\mathbb{R}^{nd}$, and a representation corresponds to a point where these hypersurfaces intersect. Now we say that this representation has the *Strong Arnold Property* if the hypersurfaces intersect transversally, i.e., their normal vectors at this point are linearly independent. This condition means that the intersection point is not just accidental, but is forced by some more fundamental structure; for example, if the representation has the Strong Arnold Property,

and we change by a small amount each constant 1 on the right hand sides of the defining equations, we get another solvable system.

13.2 Duality

The following notion of *duality* is known under many aliases: dual chain group in matroid theory, dual code in coding theory, Gale diagram in the theory of hyperplane arrangements, etc. Let $u_1, \dots, u_n \in \mathbb{R}^d$. Write down these vectors as column vectors, and let L be the row space of the resulting matrix. Pick any basis in the orthogonal complement $L^\perp$ of L, write them down as row vectors, and let $v_1, \dots, v_n \in \mathbb{R}^{n-d}$ be the columns of the resulting matrix. One of the main properties of this construction is that a set of the u_i forms a basis of $\mathbb{R}^d$ if and only if the complementary set of the v_i forms a basis of $\mathbb{R}^{n-d}$.

We can carry out this construction for any vector representation of a graph G, to get a *dual vector representation*. In some cases, this gives interesting constructions; for example, from cover-preserving representations we get independence-preserving representations. But note that (at least in the definition above) the dual is only determined up to an affine transformation; for geometric representations with metric properties (which is the majority), dualization does not seem to make sense. Yet it seem that in some cases more than the basic linear structure is dualized, and we don't have a general explanation for this. Let us briefly mention two examples.

In [82], a duality for orthogonal representations of a graph and its complement has been described. One of the consequences is that every graph G has an orthogonal representation whose dual (in the sense described above) becomes an orthogonal representation of the complementary graph $\overline{G}$, if an appropriate single row is added. This result is connected to the duality theory of semidefinite programming.

In [118], it was pointed out that there seems to be a duality between the Colin de Verdère numbers of planar graphs and their complements. Again (up to a single row) the nullspace representation and the Gram representation derived from a Colin de Verdère matrix of a graph are dual to each other; but while the Gram representation has strong metric properties, it is unclear how to impose those on the nullspace representation.

13.3 Algorithms

To represent a graph geometrically is a natural goal in itself, but in addition it is an important tool in the study of various graph properties, including their algorithmic aspects. There are several levels of this interplay between algorithms and geometry.

— Often the aim is to find a way to represent a graph in a "good" way. We refer to Kuratowski's characterization of planar graphs, its more recent extensions most notably by Robertson and Seymour, and to Steinitz's theorem representing 3-connected planar graphs

by 3-dimensional polyhedra. Many difficult algorithmic problems in connection with these representations have been studied.

— In other cases, graphs come together with a geometric representation, and the issue is to test certain properties, or compute some parameters, that connect the combinatorial and geometric structure. A typical question in this class is rigidity of bar-and-joint frameworks, an area whose study goes back to the work of Cauchy and Maxwell.

— Most interesting are the cases when a good geometric representation of a graph leads to algorithmic solutions of purely graph-theoretic questions that, at least on the surface, do not seem to have anything to do with geometry. Our discussions contained several examples of this (but the list will be far from complete): graph connectivity, graph coloring, finding maximum cliques in perfect graphs, giving capacity bounds in information theory, approximating the maximum cut and the bandwidth, planarity, linkless embedability, and rigidity of frameworks.

Appendix: Background material

13.4 Background from linear algebra

13.4.1 Basic facts about eigenvalues

Let A be an $n \times n$ real matrix. An *eigenvector* of A is a vector such that Ax is parallel to x; in other words, $Ax = \lambda x$ for some real or complex number λ. This number λ is called the *eigenvalue* of A belonging to eigenvector v. Clearly λ is an eigenvalue iff the matrix $A - \lambda I$ is singular, equivalently, iff $det(A - \lambda I) = 0$. This is an algebraic equation of degree n for λ, and hence has n roots (with multiplicity).

The *trace* of the square matrix $A = (A_{ij})$ is defined as

$$\mathrm{tr}(A) = \sum_{i=1}^{n} A_{ii}.$$

The trace of A is the sum of the eigenvalues of A, each taken with the same multiplicity as it occurs among the roots of the equation $det(A - \lambda I) = 0$.

If the matrix A is symmetric, then its eigenvalues and eigenvectors are particularly well behaved. All the eigenvalues are real. Furthermore, there is an orthogonal basis $v_1, \ldots, v_n$ of the space consisting of eigenvectors of A, so that the corresponding eigenvalues $\lambda_1, \ldots, \lambda_n$ are precisely the roots of $det(A - \lambda I) = 0$. We may assume that $|v_1| = \cdots = |v_n| = 1$; then A can be written as

$$A = \sum_{i+1}^{n} \lambda_i v_i v_i^{\mathsf{T}}.$$

Another way of saying this is that every symmetric matrix can be written as $U^{\mathsf{T}} D U$, where U is an orthogonal matrix and D is a diagonal matrix. The eigenvalues of A are just the diagonal entries of D.

To state a further important property of eigenvalues of symmetric matrices, we need the following definition. A *symmetric minor* of A is a submatrix B obtained by deleting some rows and the *corresponding* columns.

Theorem 13.4.1 (Interlacing eigenvalues) *Let A be an $n \times n$ symmetric matrix with eigenvalues $\lambda_1 \geq \cdots \geq \lambda_n$. Let B be an $(n-k) \times (n-k)$ symmetric minor of A with*

eigenvalues $\mu_1 \geq \cdots \geq \mu_{n-k}$. *Then*

$$\lambda_i \leq \mu_i \leq \lambda_{i+k}.$$

We conclude this little overview with a further basic fact about nonnegative matrices.

Theorem 13.4.2 (Perron-Frobenius) *If an* $n \times n$ *matrix has nonnegative entries then it has a nonnegative real eigenvalue* λ *which has maximum absolute value among all eigenvalues. This eigenvalue* λ *has a nonnegative real eigenvector. If, in addition, the matrix has no block-triangular decomposition (i.e., it does not contain a* $k \times (n-k)$ *block of* 0*-s disjoint from the diagonal), then* λ *has multiplicity* 1 *and the corresponding eigenvector is positive.*

13.4.2 Semidefinite matrices

A symmetric $n \times n$ matrix A is called *positive semidefinite*, if all of its eigenvalues are nonnegative. This property is denoted by $A \succeq 0$. The matrix is *positive definite*, if all of its eigenvalues are positive.

There are many equivalent ways of defining positive semidefinite matrices, some of which are summarized in the Proposition below.

Proposition 13.4.3 *For a real symmetric* $n \times n$ *matrix* A*, the following are equivalent:*

(i) A *is positive semidefinite;*

(ii) *the quadratic form* $x^T Ax$ *is nonnegative for every* $x \in \mathbb{R}^n$*;*

(iii) A *can be written as the Gram matrix of* n *vectors* $u_1, ..., u_n \in \mathbb{R}^m$ *for some* m*; this means that* $a_{ij} = u_i^{\mathsf{T}} u_j$*. Equivalently,* $A = U^{\mathsf{T}} U$ *for some matrix* U*;*

(iv) A *is a nonnegative linear combination of matrices of the type* xx^{T}*;*

(v) *The determinant of every symmetric minor of* A *is nonnegative.*

Let me add some comments. The least m for which a representation as in (iii) is possible is equal to the rank of A. It follows e.g. from (ii) that the diagonal entries of any positive semidefinite matrix are nonnegative, and it is not hard to work out the case of equality: all entries in a row or column with a 0 diagonal entry are 0 as well. In particular, the trace of a positive semidefinite matrix A is nonnegative, and $\mathrm{tr}(A) = 0$ if and only if $A = 0$.

The sum of two positive semidefinite matrices is again positive semidefinite (this follows e.g. from (ii) again). The simplest positive semidefinite matrices are of the form aa^{T} for some vector a (by (ii): we have $x^{\mathsf{T}}(aa^{\mathsf{T}})x = (a^{\mathsf{T}}x)^2 \geq 0$ for every vector x). These matrices are precisely the positive semidefinite matrices of rank 1. Property (iv) above shows that every positive semidefinite matrix can be written as the sum of rank-1 positive semidefinite matrices.

The product of two positive semidefinite matrices A and B is not even symmetric in general (and so it is not positive semidefinite); but the following can still be claimed about the product:

Proposition 13.4.4 *If A and B are positive semidefinite matrices, then $\mathrm{tr}(AB) \geq 0$, and equality holds iff $AB = 0$.*

Property (v) provides a way to check whether a given matrix is positive semidefinite. This works well for small matrices, but it becomes inefficient very soon, since there are many symmetric minors to check. An efficient method to test if a symmetric matrix A is positive semidefinite is the following algorithm. Carry out Gaussian elimination on A, pivoting always on diagonal entries. If you ever find a negative diagonal entry, or a 0 diagonal entry whose row contains a non-zero, stop: the matrix is not positive semidefinite. If you obtain an all-zero matrix (or eliminate the whole matrix), stop: the matrix is positive semidefinite.

If this simple algorithm finds that A is not positive semidefinite, it also provides a certificate in the form of a vector v with $v^{\mathsf{T}}Av < 0$. Assume that the i-th diagonal entry of the matrix $A^{(k)}$ after k steps is negative. Write $A^{(k)} = E_k^{\mathsf{T}} \dots E_1^{\mathsf{T}} A E_1 \dots E_k$, where E_i are elementary matrices. Then we can take the vector $v = E_1 \dots E_k e_i$. The case when there is a 0 diagonal entry whose row contains a non-zero is similar.

It will be important to think of $n \times n$ matrices as vectors with n^2 coordinates. In this space, the usual inner product is written as $A \cdot B$. This should not be confused with the matrix product AB. However, we can express the inner product of two $n \times n$ matrices A and B as follows:

$$A \cdot B = \sum_{i=1}^{n} \sum_{j=1}^{n} A_{ij} B_{ij} = \mathrm{tr}(A^{\mathsf{T}} B).$$

Positive semidefinite matrices have some important properties in terms of the geometry of this space. To state these, we need two definitions. A *convex cone* in $\mathbb{R}^n$ is a set of vectors which along with any vector, also contains any positive scalar multiple of it, and along with any two vectors, also contains their sum. Any system of homogeneous linear inequalities

$$a_1^{\mathsf{T}} x \geq 0, \quad \dots \quad a_m^{\mathsf{T}} x \geq 0$$

defines a convex cone; convex cones defined by such (finite) systems are called *polyhedral*.

For every convex cone C, we can form its *polar cone* C^*, defined by

$$C^* = \{x \in \mathbb{R}^n : x^{\mathsf{T}} y \geq 0 \ \forall y \in C\}.$$

This is again a convex cone. If C is closed (in the topological sense), then we have $(C^*)^* = C$.

The fact that the sum of two such matrices is again positive semidefinite (together with the trivial fact that every positive scalar multiple of a positive semidefinite matrix is positive

semidefinite), translates into the geometric statement that *the set of all positive semidefinite matrices forms a convex closed cone* $\mathcal{P}_n$ *in* $\mathbb{R}^{n\times n}$ *with vertex* 0. This cone $\mathcal{P}_n$ is important, but its structure is quite non-trivial. In particular, it is non-polyhedral for $n \geq 2$; for $n = 2$ it is a nice rotational cone (Figure 13.1; the fourth coordinate x_{21}, which is always equal to x_{12} by symmetry, is suppressed). For $n \geq 3$ the situation becomes more complicated, because $\mathcal{P}_n$ is neither polyhedral nor smooth: any matrix of rank less than $n-1$ is on the boundary, but the boundary is not differentiable at that point.

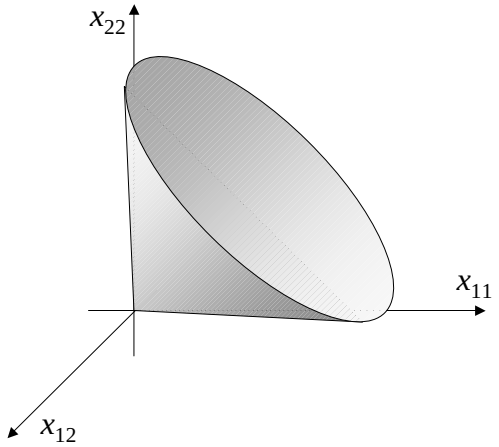

Figure 13.1: The semidefinite cone for $n = 2$.

The polar cone of $\mathcal{P}$ is itself; in other words,

Proposition 13.4.5 *A matrix* A *is positive semidefinite iff* $A \cdot B \geq 0$ *for every positive semidefinite matrix* B*.*

13.4.3 Cross product

This construction is probably familiar from physics. For $a, b \in \mathbb{R}^3$, we define their *cross product* as the vector

$$a \times b = |a| \cdot |b| \cdot \sin\phi \cdot u, \tag{13.1}$$

where ϕ is the angle between a and b ($0 \leq \phi \leq \pi$), and u is a unit vector in $\mathbb{R}^3$ orthogonal to the plane of a and b, so that the triple (a, b, u) is right-handed (positively oriented). The definition of u is ambiguous if a and b are parallel, but then $\sin\phi = 0$, so the cross product is 0 anyway. The length of the cross product gives the area of the parallelogram spanned by a and b.

The cross product is distributive with respect to linear combination of vectors, it is anticommutative: $a \times b = -b \times a$, and $a \times b = 0$ if and only if a and b are parallel. The cross

product is not associative; instead, it satisfies the *Expansion Identity*

$$(a \times b) \times c = (a \cdot c)b - (b \cdot c)a, \tag{13.2}$$

which implies the *Jacobi Identity*

$$(a \times b) \times c + (b \times c) \times a + (c \times a) \times b = 0. \tag{13.3}$$

Another useful replacement for the associativity is the following.

$$(a \times b) \cdot c = a \cdot (b \times c) = \det(a, b, c) \tag{13.4}$$

(here (a, b, c) is the 3×3 matrix with columns a, b and c.

We often use the cross product in the special case when the vectors lie in a fixed plane Π. Let k be a unit vector normal to Π, then $a \times b$ is Ak, where A is the signed area of the parallelogram spanned by a and b (this means that T is positive iff a positive rotation takes the direction of a to the direction of b, when viewed from the direction of k). Thus in this case all the information about $a \times b$ is contained in this scalar A, which in tensor algebra would be denoted by $a \wedge b$. But not to complicate notation, we'll use the cross product in this case as well.

13.5 Graph theory

13.5.1 Basics

13.5.2 Szemerédi partitions

13.6 Eigenvalues of graphs

13.6.1 Matrices associated with graphs

We introduce the adjacency matrix, the Laplacian and the transition matrix of the random walk, and their eigenvalues.

Let G be a (finite, undirected, simple) graph with node set $V(G) = \{1, \ldots, n\}$. The *adjacency matrix* of G is be defined as the $n \times n$ matrix $A_G = (A_{ij})$ in which

$$A_{ij} = \begin{cases} 1, & \text{if } i \text{ and } j \text{ are adjacent,} \\ 0, & \text{otherwise.} \end{cases}$$

We can extend this definition to the case when G has multiple edges: we just let A_{ij} be the number of edges connecting i and j. We can also have weights on the edges, in which case we let A_{ij} be the weight of the edges. We could also allow loops and include this information in the diagonal, but we don't need this in this course.

The *Laplacian* of the graph is defined as the $n \times n$ matrix $L_G = (L_{ij})$ in which

$$L_{ij} = \begin{cases} d_i, & \text{if } i = j, \\ -A_{ij}, & \text{if } i \neq j. \end{cases}$$

Here d_i denotes the degree of node i. In the case of weighted graphs, we define $d_i = \sum_j A_{ij}$. So $L_G = D_G - A_G$, where D_G is the diagonal matrix of the degrees of G.

The *transition matrix of the random walk on* G is defined as the $n \times n$ matrix $P_G = (P_{ij})$ in which

$$P_{ij} = \frac{1}{d_i} A_{ij}.$$

So $P_G = D_G^{-1} A$.

The matrices A_G and L_G are symmetric, so their eigenvalues are real. The matrix P_G is not symmetric, but it is conjugate to a symmetric matrix. Let

$$N_G = D_G^{-1/2} A_G D_G^{-1/2},$$

then N_G is symmetric, and

$$P_G = D_G^{-1/2} N_G D_G^{1/2}.$$

The matrices A_G and L_G and N_G are symmetric, so their eigenvalues are real. The matrices P_G and N_G have the same eigenvalues, and so all eigenvalues of P_G are real. We denote these eigenvalues as follows:

$$\begin{aligned} A_G: &\ \lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n, \\ L_G: &\ \mu_1 \leq \mu_2 \leq \cdots \leq \mu_n, \\ A_G: &\ \nu_1 \geq \nu_2 \geq \cdots \geq \nu_n, \end{aligned}$$

Exercise 13.1 Compute the spectrum of complete graphs, cubes, stars, paths.

We'll often use the (generally non-square) *incidence matrix* of G. This notion comes in two flavors. Let $V(G) = \{1, \ldots, n\}$ and $E(G) = \{e_1, \ldots, e_m$, and let B_G denote the $n \times m$ matrix for which

$$(B_G)_{ij} = \begin{cases} 1 & \text{if } i \text{ is and endpoint of } e_j, \\ 0 & \text{otherwise.} \end{cases}$$

Often, however, the following matrix is more useful: Let us fix an orientation of each edge, to get an oriented graph $\overrightarrow{G}$. Then let $B_{\overrightarrow{G}}$ denote the $n \times m$ matrix for which

$$(B_{\overrightarrow{G}})_{ij} = \begin{cases} 1 & \text{if } i \text{ is the head of } e_j, \\ -1 & \text{if } i \text{ is the tail of } e_j, \\ 0 & \text{otherwise.} \end{cases}$$

Changing the orientation only means scaling some columns by -1, which often does not matter much. For example, it is easy to check that independently of the orientation,

$$L_G = B_{\overrightarrow{G}} B_{\overrightarrow{G}}^{\mathsf{T}}. \tag{13.5}$$

It is worth while to express this equation in terms of quadratic forms:

$$x^{\mathsf{T}} L_G x = \sum_{ij \in E(G)} (x_i - x_j)^2. \tag{13.6}$$

13.6.2 The largest eigenvalue

Adjacency matrix

The Perron–Frobenius Theorem implies immediately that if G is connected, then the largest eigenvalue $\lambda_{\max}$ of A_G of A_G has multiplicity 1. This eigenvalue is relatively uninteresting, it is a kind of "average degree". More precisely, let $d_{\min}$ denote the minimum degree of G, let $\overline{d}$ be the average degree, and let $d_{\max}$ be the maximum degree.

Proposition 13.6.1 *For every graph G,*

$$\max\{\overline{d}, \sqrt{d_{\max}}\} \le \lambda_{\max} \le d_{\max}.$$

Proof.

□

Exercise 13.2 Compute the largest eigenvalue of a star.

Laplacian

For the Laplacian L_G, this corresponds to the smallest eigenvalue, which is really uninteresting, since it is 0:

Proposition 13.6.2 *The Laplacian L_G is singular and positive semidefinite.*

Proof. The proof follows immediately from (13.5) or (13.6), which show that L_G is positive semidefinite. Since $\mathbf{1} = (1, \dots, 1)^{\mathsf{T}}$ is in the null space of L_G, it is singular. □

If G is connected, then 0, as an eigenvalue of L_G, has multiplicity 1; we get this by applying the Perron–Frobenius Theorem to $cI - L_G$, where c is a large real number. The eigenvector belonging to this eigenvalue is $\mathbf{1} = (1, \dots, 1)^{\mathsf{T}}$ (and its scalar multiples).

We note that for a general graph, the multiplicity of the 0 eigenvalue of the Laplacian is equal to the number of connected components. Similar statement is not true for the adjacency matrix (if the largest eigenvalues of the connected components of G are different, then the largest eigenvalue of the whole graph has multiplicity 1). This illustrates the phenomenon that the Laplacian is often better behaved algebraically than the adjacency matrix.

Transition matrix

The largest eigenvalue of P_G is 1, and it has multiplicity 1 for connected graphs. It is straightforward to check that the right eigenvector belonging to it is $\mathbf{1}$, and the left eigenvector is given by $\pi_i = d_i/(2m)$ (where m is the number of edges). This vector π describes the *stationary distribution* of a random walk, and it is very important in the theory of random walks (see later).

13.6.3 The smallest eigenvalue

Proposition 13.6.3 (a) *A graph is bipartite if and only if its spectrum is symmetric about the origin.*

(b) *A connected graph G is bipartite if and only if $\lambda_{\min}(G) = -\lambda_{\max}(G)$.*

Proof.

□

The "only if" part of Proposition 13.6.3 can be generalized: The ratio between the largest and smallest eigenvalue can be used to estimate the chromatic number (Hoffman [101]).

Theorem 13.6.4

$$\chi(G) \geq 1 + \frac{\lambda_{\min}}{\lambda_{\max}}.$$

Proof. Let $k = \chi(G)$, then A_G can be partitioned as

$$\begin{pmatrix} 0 & M_{12} & \dots & M_{1k} \\ M_{21} & 0 & & M_{2k} \\ \vdots & \vdots & \ddots & \\ M_{k1} & M_{k2} & & 0, \end{pmatrix}$$

where M_{ij} is an $m_i \times m_j$ matrix (where m_i is the number of points with color i).

Let $\mathbf{v}$ be an eigenvector belonging to λ_1. Let us break $\mathbf{v}$ into pieces $\mathbf{v}_1, \dots, \mathbf{v}_k$ of length $m_1, \dots, m_k$, respectively. Set

$$\mathbf{w}_i = \begin{pmatrix} |\mathbf{v}_i| \\ 0 \\ \vdots \\ 0 \end{pmatrix} \in \mathbb{R}^{m_i} \quad \mathbf{w} = \begin{pmatrix} \mathbf{w}_1 \\ \vdots \\ \mathbf{w}_k \end{pmatrix}.$$

Let B_i be any orthogonal matrix such that

$$B_i \mathbf{w}_i = \mathbf{v}_i \qquad (i = 1, \dots, k),$$

and

$$B = \begin{pmatrix} B_1 & & & 0 \\ & B_2 & & \\ 0 & & \ddots & \\ & & & B_k \end{pmatrix}.$$

Then $B\mathbf{w} = \mathbf{v}$ and

$$B^{-1}AB\mathbf{w} = B^{-1}A\mathbf{v} = \lambda_1 B^{-1}\mathbf{v} = \lambda_1 \mathbf{w}$$

so $\mathbf{w}$ is an eigenvector of $B^{-1}AB$. Moreover, $B^{-1}AB$ has the form

$$\begin{pmatrix} 0 & B_1^{-1}A_{12}B_2 & \dots & B_1^{-1}A_{1k}B_k \\ B_2^{-1}A_{21}B_1 & 0 & & B_2^{-1}A_{2k}B_k \\ \vdots & & \ddots & \vdots \\ B_k^{-1}A_{k1}B_1 & B_k^{-1}A_{k2}B_2 & \dots & 0 \end{pmatrix}.$$

Pick the entry in the upper left corner of each of the k^2 submatrices $B_i^{-1}A_{ij}B_j$ $(A_{ii} = 0)$, these form a $k \times k$ submatrix D. Observe that

$$\mathbf{u} = \begin{pmatrix} |\mathbf{v}_1| \\ \vdots \\ |\mathbf{v}_k| \end{pmatrix}$$

is an eigenvector of D; for $\mathbf{w}$ is an eigenvector of $B^{-1}AB$ and has 0 entries on places corresponding to those rows and columns of $B^{-1}AB$, which are to be deleted to get D. Moreover, the eigenvalue belonging to $\mathbf{u}$ is λ_1.

Let $\alpha_1 \geq \cdots \geq \alpha_k$ be the eigenvalues of D. Since D has 0's in its main diagonal,

$$\alpha_1 + \cdots + \alpha_k = 0.$$

On the other hand, λ_1 is an eigenvalue of D and so

$$\lambda_1 \leq \alpha_1,$$

while by the Interlacing Eigenvalue Theorem

$$\lambda_n \leq \alpha_k, \dots, \lambda_{n-k+2} \leq \alpha_2.$$

Thus

$$\lambda_n + \cdots + \lambda_{n-k+2} \leq \alpha_k + \cdots + \alpha_2 = -\alpha_1 \leq -\lambda_1.$$

□

Remark 13.6.5 The proof did not use that the edges were represented by the number 1, only that the non-edges and diagonal entries were 0. So if we want to get the strongest possible lower bound on the chromatic number that this method provides, we can try to find a way of choosing the entries in A corresponding to edges of G in such a way that the right hand side is minimized. This can be done efficiently.

The smallest eigenvalue is closely related to the characterization of linegraphs. The correspondence is not perfect though. To state the result, we need some definitions. Let G be a simple graph. A *pending star* in G is a maximal set of edges which are incident with the same node and whose other endpoints have degree 1. The *linegraph* $L(G)$ of G is defined on $V(L(G)) = E(G)$, where to edges of G are adjacent in $L(G)$ if and only if they have a node in common. A graph H is called a *modified linegraph* of G if it is obtained from $L(G)$ by deleting a set of disjoint edges from each clique corresponding to a pending star of G.

Part (a) of the following theorem is due to Hoffman [100], part (b), to Cameron, Goethals, Seidel and Shult [37].

Proposition 13.6.6 (a) *Let H be the generalized linegraph of G. Then $\lambda_{\min}(H) \geq -2$; if $|E(G)| > |V(G)|$, then $\lambda_{\min}(H) = -2$.*

(b) *Let H be a simple graph such that $\lambda_{\min}(H) \geq -2$. Assume that $|V(H)| \geq 37$. Then G is a modified linegraph.*

Proof. We only give the proof for part (a), and only in the case when $H = L(G)$. It is easy to check that we have

$$A_{L(G)} = B_G^{\mathsf{T}} B_G - 2I.$$

Since $B_G^T B_G$ is positive semidefinite, all of its eigenvalues are non-negative. Hence, the eigenvalues of $A_{L(G)}$ are ≥ -2. Moreover, if $|V(G)| < |E(G)|$, then

$$r(B^T B) = r(B) \leq |V(G)| < |E(G)|$$

($r(X)$ is the rank of the matrix X). So, $B^T B$ has at least one 0 eigenvalue, i.e. $A_{L(G)}$ has at least one -2 eigenvalue. □

Exercise 13.3 Modify the proof above to get (a) in general.

13.6.4 The eigenvalue gap

The gap between the second and the first eigenvalues is an extremely important parameter in many branches of mathematics.

If the graph is connected, then the largest eigenvalue of the adjacency matrix as well as the smallest eigenvalue of the Laplacian have multiplicity 1. We can expect that the gap between

this and the nearest eigenvalue is related to some kind of connectivity measure of the graph. Indeed, fundamental results due to Alon–Milman [12], Alon [8] and Jerrum–Sinclair [106] relate the eigenvalue gap to expansion (isoperimetric) properties of graphs. These results can be considered as discrete analogues of Cheeger's inequality in differential geometry.

There are many related (but not equivalent) versions of these results. We illustrate this connection by two versions that are of special interest: a spectral characterization of expanders and a bound on the mixing time of random walks on graphs. For this, we discuss very briefly expanders and also random walks and their connections with eigenvalues (see [4] and [143] for more).

The multiplicity of the second largest eigenvalue will be discussed in connection with the Colin de Verdière number.

Expanders

An *expander* is a regular graph with small degree in which the number of neighbors of any set containing at most half of the nodes is at least a constant factor of its size. To be precise, an ε-expander is a graph $G = (V, E)$ in which for every set $S \subset V$ with $|S| \leq |V|/2$, the number of nodes in $V \setminus S$ adjacent to some node in S is at least $\varepsilon|S|$.

Expanders play an important role in many applications of graph theory, in particular in computer science. The most important expanders are d-regular expanders, where $d \geq 3$ is a small constant. Such graphs are not easy to construct. One method is to do a random construction: for example, we can pick d random perfect matchings on $2n$ nodes (independently, uniformly over all perfect matchings), and let G be the union of them. Then a moderately complicated computation shows that G is an ε-expander with positive probability for a sufficiently small ε. Deterministic constructions are much more difficult to obtain; the first construction was found by Margulis [146]; see also [144]. Most of these constructions are based on deep algebraic facts.

Our goal here is to state and prove a spectral characterization of expanders, due to Alon [8], which plays an important role in analyzing some of the above mentioned algebraic constructions. note that since we are considering only regular graphs, the adjacency matrix, the Laplacian and the transition matrix are easily expressed, and so we shall only consider the adjacency matrix.

Theorem 13.6.7 *Let G be a d-regular graph.*

(a) *If $d - \lambda_2 \geq 2\varepsilon d$, then G is an ε-expander.*

(b) *If G is an ε-expander, then $d - \lambda_2 \geq \varepsilon^2/5$.*

Proof. The proof is similar to the proof of Theorem 13.6.8 below. □

Edge expansion (conductance)

We study the connection of the eigenvalue gap of the transition matrix with a quantity that can be viewed as an edge-counting version of the expansion. Let $1 = \lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_n$ be the eigenvalues of P_G.

The *conductance* of a graph $G = (V, E)$ is defined as follows. For two sets $S_1, S_2 \subseteq V$, let $e_G(S_1, S_2)$ denote the number of edges ij with $i \in S_1, j \in S_2$. For every subset $S \subseteq V$, let $d(S) = \sum_{i \in S} d_i$, and define

$$\Phi(G) = \min_{\emptyset \subset S \subset V} \frac{2me_G(S, V \setminus S)}{d(S) \cdot d(V \setminus S)}.$$

For a d-regular graph, this can be written as

$$\Phi(G) = \min_{\emptyset \subset S \subset V} \frac{n}{d} \frac{e_G(S, V \setminus S)}{|S| \cdot |V \setminus S|}.$$

The following basic inequality was proved by Jerrum and Sinclair [106]:

Theorem 13.6.8 *For every graph G,*

$$\frac{\Phi(G)^2}{16} \leq 1 - \lambda_2 \leq \Phi(G)$$

We start with a lemma expressing the eigenvalue gap of P_G in a manner similar to the Rayleigh quotient.

Lemma 13.6.9 *For every graph G we have*

$$1 - \lambda_2 = \min \sum_{(i,j) \in E} (x_i - x_j)^2,$$

where the minimum is taken over all vectors $x \in \mathbb{R}^V$ such that

$$\sum_{i \in V} d_i x_i = 0, \qquad \sum_{i \in V} d_i x_i^2 = 1. \tag{13.7}$$

Proof. As remarked before, the symmetrized matrix $N_G = D_G^{1/2} P_G D_G^{-1/2}$ has the same eigenvalues as P_G. For a symmetric matrix, the second largest eigenvalue can be obtained as

$$\lambda_2 = \max y^{\mathsf{T}} N_G y,$$

where y ranges over all vectors of unit length orthogonal to the eigenvector belonging to the largest eigenvalue. This latter eigenvector is given (up to scaling) by $v_i = \sqrt{d_i}$, so the conditions on y are

$$\sum_{i \in V} \sqrt{d_i} y_i = 0, \qquad \sum_{i \in V} y_i^2 = 1. \tag{13.8}$$

Let us write $y_i = x_i\sqrt{d_i}$, then the conditions (13.8) on y translate into conditions (13.7) on x. Furthermore,

$$\begin{aligned}\sum_{(i,j)\in E}(x_i - x_j)^2 &= 2m\sum_{(i,j)\in E}(\frac{y_i}{\sqrt{d_i}} - \frac{y_j}{\sqrt{d_j}})^2\\ &= \sum_i d_i\frac{y_i^2}{d_i} - 2\sum_{(i,j)\in E}(\frac{y_iy_j}{\sqrt{d_i}\sqrt{d_j}}\\ &= 1 - y^{\mathsf{T}}N_Gy.\end{aligned}$$

The minimum of the left hand side subject to (13.7) is equal to the minimum of the right hand side subject to (13.8), which proves the Lemma. □

Now we can prove the theorem.

Proof. The upper bound is easy: let $\emptyset \neq S \subset V$ be a set with

$$\frac{2me_G(S, V\setminus S)}{d(S)\cdot d(V\setminus S)} = \Phi(G).$$

Let x be a vector on the nodes defined by

$$x_i = \begin{cases}\sqrt{\frac{d(V\setminus S)}{2md(S)}} & \text{if } i\in S,\\ -\sqrt{\frac{d(S)}{2md(V\setminus S)}} & \text{if } i\in V\setminus S.\end{cases}$$

It is easy to check that

$$\sum_{i\in V} d_ix_i = 0, \qquad \sum_{i\in V} d_ix_i^2 = 1.$$

Thus by Lemma 13.6.9,

$$\begin{aligned}1-\lambda_2 \geq \sum_{ij\in E}(x_i - x_j)^2 &= e_G(S, V\setminus S)\left(\sqrt{\frac{d(V\setminus S)}{2md(S)}} + \sqrt{\frac{d(S)}{2md(V\setminus S)}}\right)^2\\ &= \frac{2me_G(S, V\setminus S)}{d(S)d(V\setminus S)} = \Phi(G).\end{aligned}$$

It is easy to see that the statement giving the lower bound can be written as follows: let $y\in\mathbb{R}^V$ and let $\hat{y} = (1/2m)\sum_i d_iy_i$. Then we have

$$\sum_{(i,j)\in E}(y_i - y_j)^2 \geq \frac{\Phi^2}{16}\sum_i(y_i - \hat{y})^2. \tag{13.9}$$

To prove this, we need a lemma that can be thought of as a linear version of (13.9). For every real vector $y = (y_1, \ldots, y_n)$, we define its *median* (relative to the degree sequence d_i) as a the member y_M of the sequence for which

$$\sum_{k:\, y_k\leq y_M} d_k \leq m, \qquad \sum_{k:\, y_k > y_M} d_k < m.$$

Lemma 13.6.10 *Let $G = (V, E)$ be a graph with conductance $\Phi(G)$. Let $y \in \mathbb{R}^V$, and let y_M be the median of y. Then*

$$\sum_{(i,j)\in E} |y_i - y_j| \geq \frac{\Phi}{2} \sum_i d_i |y_i - y_M|.$$

Proof. [of the Lemma] We may label the nodes so that $y_1 \leq y_2 \leq \ldots \leq y_n$. We also may assume that $y_M = 0$ (the assertion of the Lemma is invariant under shifting the entries of y). Substituting

$$y_j - y_i = (y_{i+1} - y_i) + \cdots + (y_j - y_{j-1}),$$

we have

$$\sum_{(i,j)\in E} |y_i - y_j| = \sum_{k=1}^{n-1} e(\leq k, > k)(y_{k+1} - y_k).$$

By the definition of Φ, this implies

$$\begin{aligned}
\sum_{(i,j)\in E} |y_i - y_j| &\geq \frac{\Phi}{2m} \sum_{k=1}^{n-1} d(\leq k) d(> k)(y_{k+1} - y_k) \\
&\geq \frac{\Phi}{2m} \sum_{k<M} d(\leq k) m (y_{k+1} - y_k) + \frac{\Phi}{2m} \sum_{k\geq M} m d(> k)(y_{k+1} - y_k) \\
&= \frac{\Phi}{2} \sum_{i\leq M} d_i y_i - \frac{\Phi}{2} \sum_{i>M} d_i y_i \\
&= \frac{\Phi}{2} \sum_i d_i |y_i|.
\end{aligned}$$

□

Now we return to the proof of the lower bound in Theorem 13.6.8. Let x be a unit length eigenvector belonging to λ_2. We may assume that the nodes are labeled so that $x_1 \geq x_2 \geq \ldots \geq x_n$. Let x_M be the median of x. Note that the average $(1/(2m)) \sum_i d_i x_i = 0$. Set $z_i = \big(\max\{0, x_i - x_M\}\big)$ and $u_i = \big(\max\{0, x_M - x_i\}\big)$. Then

$$\sum_i d_i z_i^2 + \sum_i d_i u_i^2 = \sum_i d_i (x_i - x_M)^2 = \sum_i x_i^2 + 2m x_M^2 \geq \sum_i d_i x_i^2 = 1,$$

and so we may assume (replacing x by $-x$ if necessary) that

$$\sum_i d_i z_i^2 \geq \frac{1}{2}.$$

By Lemma 13.6.10

$$\sum_{(i,j)\in E} |z_i^2 - z_j^2| \geq \frac{\Phi}{2} \sum_i d_i z_i^2.$$

On the other hand, using the Cauchy-Schwartz inequality,

$$\sum_{(i,j)\in E} |z_i^2 - z_j^2| = \sum_{(i,j)\in E} |z_i - z_j| \cdot |z_i + z_j|$$
$$\leq \left(\sum_{(i,j)\in E} (z_i - z_j)^2 \right)^{1/2} \left(\sum_{(i,j)\in E} (z_i + z_j)^2 \right)^{1/2}.$$

Here the second factor can be estimated as follows:

$$\sum_{(i,j)\in E} (z_i + z_j)^2 \leq 2 \sum_{(i,j)\in E} (z_i^2 + z_j^2) = 2 \sum_i d_i z_i^2.$$

Combining these inequalities, we obtain

$$\sum_{(i,j)\in E} (z_i - z_j)^2 \geq \left(\sum_{(i,j)\in E} |z_i^2 - z_j^2| \right)^2 \Big/ \sum_{(i,j)\in E} (z_i + z_j)^2$$
$$\geq \frac{\Phi^2}{4} \left(\sum_i d_i z_i^2 \right)^2 \Big/ 2 \sum_i d_i z_i^2 = \frac{\Phi^2}{8} \sum_i d_i z_i^2 \geq \frac{\Phi^2}{16}.$$

Since

$$\sum_{(i,j)\in E} (x_i - x_j)^2 \geq \sum_{(i,j)\in E} (z_i - z_j)^2,$$

from here we can conclude by Lemma 13.6.9. □

The quantity $\Phi(G)$ is NP-complete to compute. An important theorem of Leighton and Rao gives an approximate min-max theorem for it, which also yields a polynomial time approximation algorithm, all with an error factor of $O(\log n)$.

Random walks

A *random walk* on a graph G is a random sequence $(v^0, v^1, \dots)$ of nodes constructed as follows: We pick a starting point v^0 from a specified initial distribution σ, we select a neighbor v^1 of it at random (each neighbor is selected with the same probability $1/d(v^0)$), then we select a neighbor v^2 of this node v^1 at random, etc. We denote by σ^k the distribution of v^k.

In the language of probability theory, a random walk is a finite time-reversible Markov chain. (There is not much difference between the theory of random walks on graphs and the theory of finite Markov chains; every Markov chain can be viewed as random walk on a directed graph, if we allow weighted edges, and every time-reversible Markov chain can be viewed as random walks on an edge-weighted undirected graph.)

Let π denote the probability distribution in which the probability of a node is proportional to its degree:

$$\pi(v) = \frac{d(v)}{2m}.$$

This distribution is called the *stationary distribution* of the random walk. It is easy to check that if v^0 is selected from π, then after any number of steps, v^k will have the same distribution π. This explains the name of π. Algebraically, this means that π is a left eigenvector of P_G with eigenvalue 1:

$$\pi^T P_G = \pi^T.$$

Theorem 13.6.11 *If G is a connected nonbipartite graph, then $\sigma^k \to \pi$ for every starting distribution σ.*

It is clear that the conditions are necessary.

Before proving this theorem, let us make some remarks on one of its important applications, namely *sampling*. Suppose that we want to pick a random element uniformly from some finite set. We can then construct a connected nonbipartite regular graph on this set, and start a random walk on this graph. A node of the random walk after sufficiently many steps is therefore essentially uniformly distributed.

(It is perhaps surprising that there is any need for a non-trivial way of generating an element from such a simple distribution as the uniform. But think of the first application of random walk techniques in real world, namely shuffling a deck of cards, as generating a random permutation of 52 elements from the uniform distribution over all permutations. The problem is that the set we want a random element from is exponentially large. In many applications, it has in addition a complicated structure; say, we consider the set of lattice points in a convex body or the set of linear extensions of a partial order. Very often this random walk sampling is the only known method.)

With this application in mind, we see that not only the fact of convergence matters, but also the rate of this convergence, called the *mixing rate*. The proof below will show how this relates to the eigenvalue gap. In fact, we prove:

Theorem 13.6.12 *Let $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$ be the eigenvalues of P_G, and let $\mu = \max\{\lambda_2, \lambda_n\}$. Then for every starting node i and any node j, and every $t \geq 0$, we have*

$$|\Pr(v^t = j) - \pi(j)| \leq \sqrt{\frac{\pi(j)}{\pi(i)}}\mu^t.$$

More generally, for every set $A \subseteq V$,

$$|\Pr(v^t \in A) - \pi(A)| \leq \sqrt{\frac{\pi(A)}{\pi(i)}}\mu^t.$$

Proof. We prove the first inequality; the second is left to the reader as an exercise. We know that the matrix N_G has the same eigenvalues as P_G, and it is symmetric, so we can

write it as

$$N_G = \sum_{k=1}^{n} \lambda_k v_k v_k^{\mathsf{T}},$$

where $v_1, \dots, v_n$ are mutually orthogonal eigenvectors. It is easy to check that we can choose

$$v_{1i} = \sqrt{\pi_i}$$

(we don't know anything special about the other eigenvectors). Hence we get

$$\begin{aligned}\Pr(v^t = j) = (P^t)_{ij} = e_i^{\mathsf{T}} D^{-1/2} N^t d^{1/2} e_j = \sum_{k=1}^{n} \lambda_k^t (e_i^{\mathsf{T}} D^{-1/2} v_k)(e_j^{\mathsf{T}} D^{1/2} v_k) \\ = \sum_{k=1}^{n} \lambda_k^t \frac{1}{\sqrt{\pi(i)}} v_{ki} \sqrt{\pi(j)} v_{kj} = \pi(j) + \sqrt{\frac{\pi(j)}{\pi(i)}} \sum_{k=2}^{n} \lambda_k^t v_{ki} v_{kj}.\end{aligned}$$

Here the first term is the limit; we need to estimate the second. We have

$$\Big|\sum_{k=2}^{n} \lambda_k^t v_{ki} v_{kj}\Big| \le \mu^t \sum_{k=2}^{n} \Big|v_{ki} v_{kj}\Big| \le \mu^t \sum_{k=1}^{n} \Big|v_{ki} v_{kj}\Big| \le \mu^t \left(\sum_{k=1}^{n} v_{ki}^2\right)^{1/2} \big(v_{kj}^2\big)^{1/2} = \mu^t.$$

This proves the inequality. □

If we want to find a bound on the number of steps we need before, say,

$$|\Pr(v^k \in A) - \pi(A)| < \varepsilon$$

holds for every j, then it suffices to find a k for which

$$\mu^k < \varepsilon \sqrt{\pi_i}.$$

Writing $mu = 1 - \gamma$, and using that $1 - \gamma < e^{-\gamma}$, it suffices to have

$$e^{-\gamma k} < \varepsilon \sqrt{\pi_i},$$

and expressing k,

$$k > \frac{1}{\gamma}\Big(\ln \frac{1}{\varepsilon} + \frac{1}{2} \ln \frac{1}{\pi_i}\Big)$$

So we see that (up to logarithmic factors), it is the reciprocal of the eigenvalue gap that governs the mixing time.

In applications, the appearance of the smallest eigenvalue λ_n is usually not important, and what we need to work on is bounding the *eigenvalue gap* $1-\lambda_2$. The trick is the following: If the smallest eigenvalue is too small, then we can modify the walk as follows. At each step, we flip a coin and move with probability 1/2 and stay where we are with probability 1/2.

The stationary distribution of this modified walk is the same, and the transition matrix P_G is replaced by $\frac{1}{2}(P_G + I)$. For this modified walk, all eigenvalues are nonnegative, and the eigenvalue gap is half of the original. So applying the theorem to this, we only use a factor of 2.

Explanation of conductance: In a stationary random walk on G, we cross every edge in every direction with the same frequency, once in every $2m$ steps on the average. So $Q(S, V \setminus S)$ is the frequency with which we step out from S. If instead we consider a sequence of independent samples from π, the frequency with which we step out from S is $\pi(S)\pi(V \setminus S)$. The ratio of these two frequencies is one of many possible ways comparing a random walk with a sequence of independent samples.

Exercise 13.4 Let $G = (V, E)$ be a simple graph, and define

$$\rho(G) = \min_{\emptyset \subset S \subset V} \frac{e_G(S, V \setminus S)}{|S| \cdot |V \setminus S|}.$$

Let λ_2 denote the second smallest eigenvalue of the Laplacian L_G of a graph G. Then

$$\lambda_2 \leq n\rho(G) \leq \sqrt{\lambda_2 d_{\max}}.$$

13.6.5 The number of different eigenvalues

Multiplicity of eigenvalues usually corresponds to symmetries in the graph (although the correspondence is not exact). We prove two results in this direction. The following theorem was proved by Mowshowitz [155] and Sachs [174]:

Theorem 13.6.13 *If all eigenvalues of A are different, then every automorphism of A has order 1 or 2.*

Proof. Every automorphism of G can be described by a permutation matrix P such that $AP = PA$. Let u be an eigenvector of A with eigenvalue λ. Then

$$A(Pu) = PAu = P(\lambda u) = \lambda(Pu),$$

so Pu is also an eigenvector of A with the same eigenvalue. Since Pu has the same length as u, it follows that $Pu = \pm u$ and hence $P^2u = u$. This holds for every eigenvector u of A, and since there is a basis consisting of eigenvectors, it follows that $P^2 = I$. □

A graph G is called *strongly regular*, if it is regular, and there are two nonnegative integers a and b such that for every pair i, j of nodes the number of common neighbors of i and j is

$$\begin{cases} a, & \text{if } a \text{ and } b \text{ are adjacent,} \\ b, & \text{if } a \text{ and } b \text{ are nonadjacent.} \end{cases}$$

Example 13.6.14 Compute the spectrum of the Petersen graph, Paley graphs, incidence graphs of finite projective planes.

The following characterization of strongly regular graphs is easy to prove:

Theorem 13.6.15 *A connected graph G is strongly regular if and only if it is regular and A_G has at most 3 different eigenvalues.*

Proof. The adjacency matrix of a strongly regular graph satisfies

$$A^2 = aA + b(J - A - I) + dI. \tag{13.10}$$

The largest eigenvalue is d, all the others are roots of the equation

$$\lambda^2 - (a-b)\lambda - (d-b), \tag{13.11}$$

Thus there are at most three distinct eigenvalues.

Conversely, suppose that G is d-regular and has at most three different eigenvalues. One of these is d, with eigenvector $\mathbf{1}$. Let λ_1 and λ_2 be the other two (I suppose there are two more—the case when there is at most one other is easy). Then

$$B = A^2 - (\lambda_1 + \lambda_2)A + \lambda_1\lambda_2 I$$

is a matrix for which $Bu = 0$ for every eigenvector of A except $\mathbf{1}$ (and its scalar multiples). Furthermore, $B\mathbf{1} = c\mathbf{1}$, where $c = (d - \lambda_1)(d - \lambda_2)$. Hence $B = (c/n)J$, and so

$$A^2 = (\lambda_1 + \lambda_2)A - \lambda_1\lambda_2 I + (c/n)J.$$

This means that $(A^2)_{ij}$ $(i \neq j)$ depends only on whether i and j are adjacent, proving that G is strongly regular. □

We can get more out of equation (13.11). We can solve it:

$$\lambda_{1,2} = \frac{a - b \pm \sqrt{(a-b)^2 + 4(d-b)}}{2}. \tag{13.12}$$

Counting induced paths of length 2, we also get the equation

$$(d - a - 1)d = (n - d - 1)b. \tag{13.13}$$

Let m_1 and m_2 be the multiplicities of the eigenvalues λ_1 and λ_2. Clearly

$$m_1 + m_2 = n - 1 \tag{13.14}$$

Taking the trace of A, we get

$$d + m_1\lambda_1 + m_2\lambda_2 = 0,$$

or

$$2d + (n-1)(a-b) + (m_1 - m_2)\sqrt{(a-b)^2 + 4(d-b)} = 0. \tag{13.15}$$

If the square root is irrational, the only solution is $d = (n-1)/2$, $b = (n-1)/4$, $a = b - 1$. There are many solutions where the square root is an integer.

A nice application of these formulas is the "Friendship Theorem":

Theorem 13.6.16 *If G is a graph in which every two nodes have exactly one common neighbor, then it has a node adjacent to every other node.*

Proof. First we show that two non-adjacent nodes must have the same degree. Suppose that there are two non-adjacent nodes u, v of different degree. For every neighbor w of u there is a common neighbor w' of w and v. For different neighbors w_1 and w_2 of u, the nodes w_1' and w_2' must be different, else $w - 1$ and $w - 2$ would have two common neighbors. So v has at least as many neighbors as u. By a symmetric reasoning, we get $d_u = d_v$.

If G has a node v whose degree occurs only once, then by the above, v must be connected to every other node, and we are done. So suppose that no such node exists.

If G has two nodes u and v of different degree, then it contains two other nodes x and y such that $d_u = d_x$ and $d_v = d_y$. But then both x and u are common neighbors of v and y, contradicting the assumption.

Now if G is regular, then it is strongly regular, and $a = b = 1$. From (13.15),

$$d + (m_1 - m_2)\sqrt{d-1} = 0.$$

The square root must be integral, hence $d = k^2 + 1$. But then $k \mid k^2 + 1$, whence $k = 1$, $d = 2$, and the graph is a triangle, which is not a counterexample. □

Exercise 13.5 Prove that every graph with only two different eigenvalues is complete.

Exercise 13.6 Describe all disconnected strongly regular graphs. Show that there are disconnected graphs with only 3 distinct eigenvalues that are not strongly regular.

13.6.6 Spectra of graphs and optimization

There are many useful connections between the eigenvalues of a graph and its combinatorial properties. The first of these follows easily from interlacing eigenvalues.

Proposition 13.6.17 *The maximum size $\omega(G)$ of a clique in G is at most $\lambda_1 + 1$. This bound remains valid even if we replace the non-diagonal 0's in the adjacency matrix by arbitrary real numbers.*

The following bound on the chromatic number is due to Hoffman.

Proposition 13.6.18 *The chromatic number $\chi(G)$ of G is at least $1-(\lambda_1/\lambda_n)$. This bound remains valid even if we replace the 1's in the adjacency matrix by arbitrary real numbers.*

The following bound on the maximum size of a cut is due to Delorme and Poljak [50, 49, 153, 161], and was the basis for the Goemans-Williamson algorithm discussed in the introduction.

Proposition 13.6.19 *The maximum size $\gamma(G)$ of a cut in G is at most $|E|/2-(n/4)\lambda_n$. This bound remains valid even if we replace the diagonal 0's in the adjacency matrix by arbitrary real numbers.*

Observation: to determine the best choice of the "free" entries in 13.6.17, 13.6.18 and 13.6.19 takes a semidefinite program. Consider 13.6.17 for example: we fix the diagonal entries at 0, the entries corresponding to edges at 1, but are free to choose the entries corresponding to non-adjacent pairs of vertices (replacing the off-diagonal 1's in the adjacency matrix). We want to minimize the largest eigenvalue. This can be written as a semidefinite program:

$$\begin{aligned} \text{minimize} \quad & t \\ \text{subject to} \quad & tI - X \succeq 0, \\ & X_{ii} = 0 \quad (\forall i \in V), \\ & X_{ij} = 1 \quad (\forall ij \in E). \end{aligned}$$

It turns out that the semidefinite program constructed for 13.6.18 is just the dual of this, and their common optimum value is the parameter $\vartheta(G)$ introduced before. The program for 13.6.19 gives the approximation used by Goemans and Williamson (for the case when all weights are 1, from which it is easily extended).

13.7 Convex polytopes

13.7.1 Polytopes and polyhedra

The convex hull of a finite set of points in $\mathbb{R}^d$ is called a (convex) *polytope*. The intersection of a finite number of halfspaces in $\mathbb{R}^d$ is called a (convex) *polyhedron*.

Proposition 13.7.1 *Every polytope is a polyhedron. A polyhedron is a polytope if and only if it is bounded.*

For every polytope, there is a unique smallest affine subspace that contains it, called its *affine hull*. The *dimension* of a polytope is the dimension of it affine hull. A polytope in $\mathbb{R}^d$ that has dimension d (equivalently, that has an interior point) is called a *d-polytope*.

A hyperplane H is said to *support* the polytope if it has a point in common with the polytope and the polytope is contained in one of the closed halfspaces with boundary H. A *face* of a polytope is its intersection with a supporting hyperplane. A face of a polytope that has dimension one less than the dimension of the polytope is called a *facet*. A face of dimension 0 (i.e., a single point) is called a *vertex*.

Proposition 13.7.2 *Every face of a polytope is a polytope. Every vertex of a face is a vertex of the polytope. Every polytope has a finite number of faces.*

Proposition 13.7.3 *Every polytope is the convex hull of its facets. The set of vertices is the unique minimal finite set of points whose convex hull is the polytope.*

Let P be a d-polytope. Then every facet F of P spans a (unique) supporting hyperplane, and the hyperplane is the boundary of a uniquely determined halfspace that contains the polytope. We'll call this halfspace the *halfspace of F*.

Proposition 13.7.4 *Every polytope is the intersection of the halfspaces of its facets.*

13.7.2 The skeleton of a polytope

The vertices and edges of a polytope P form a simple graph G_P, which we call the *skeleton* of the polytope.

Proposition 13.7.5 *Let P be a polytope in $\mathbb{R}^d$, let $a \in \mathbb{R}^d$, and let u be a vertex of P. Suppose that there P has a vertex V such that $a^\mathsf{T}u < a^\mathsf{T}v$. Then P has a vertex w such that uw is an edge and $a^\mathsf{T}u < a^\mathsf{T}w$.*

Another way of formulating this is that if we consider the linear objective function $a^\mathsf{T}x$ on a polytope P, then from any vertex we can walk on the skeleton to a vertex that maximizes the objective function so that the value of the objective function increases at every step. This important fact is the basis for the *Simplex Method*.

For our purposes, however, the following corollaries of Proposition 13.7.5 will be important:

Corollary 13.7.6 *The skeleton of any polytope is a connected graph.*

Corollary 13.7.7 *Let G be the skeleton of a d-polytope, and let H be an (open or closed) halfspace containing an interior point of the polytope. Then the subgraph of G_P induced by those vertices of P that are contained in this halfspace is connected.*

From Corollary 13.7.7, it is not hard to derive the following (see Theorem 1.3.2 for $d = 3$):

Theorem 13.7.8 *The skeleton of a d-dimensional polytope is d-connected.*

13.7.3 Polar, blocker and antiblocker

Let P be a convex polytope containing the origin as an interior point. Then the *polar* of P is defined as

$$P^* = \{x \in \mathbb{R}^d : \ x^\mathsf{T} y \leq 1 \forall y \in P\}$$

Proposition 13.7.9 *(a) The polar of a polytope is a polytope. For every polytope P we have $(P^*)^* = P$.*

(b) Let $v_0, \ldots, v_m$ be the vertices of a k-dimensional face F of P. Then

$$F^\perp = \{x \in P^* : \ v_0^\mathsf{T} x = 1, \ldots, v_m^\mathsf{T} x = 1\}$$

defines a $d - k - 1$-dimensional face of P^*. Furthermore, $(F^\perp)^\perp = F$.

In particular, every vertex v of P corresponds to a facet $v^\perp$ of P^* and vice versa. The vector v is a normal vector of the facet $v^\perp$.

There are two constructions similar to polarity that concern polyhedra that do not contain the origin in their interior; rather, they are contained in the nonnegative orthant.

A polyhedron P in $\mathbb{R}^d$ is called *ascending*, if $P \subseteq \mathbb{R}^d_+$ and whenever $x \in \P$, $y \in \mathbb{R}^d$ and $y \geq x$ then $y \in P$.

The *blocker* of an ascending polyhedron is defined by

$$P^{\mathrm{bl}} = \{x \in \mathbb{R}^d_+ : \ x^\mathsf{T} y \leq 1 \forall y \in P\}.$$

Proposition 13.7.10 *The blocker of an ascending polyhedron is an ascending polyhedron. For every ascending polyhedron P we have $(P^{\mathrm{bl}})^{\mathrm{bl}} = P$.*

The correspondence between faces of P and P^{bl} is a bit more complicated than for polarity, and we describe the relationship between vertices and facets only. Every vertex v of P gives rise to a facet $v\perp$, which corresponds to the halfspace $v^\mathsf{T} x \geq 1$. This construction gives all the facets of P^{bl}, except possibly those corresponding to the nonnegativity constraints $x_i \geq 0$, which may or may not define facets.

A d-polytope P is called a *corner polytope*, if $P \subseteq \mathbb{R}^d_+$ and whenever $x \in \P$, $y \in \mathbb{R}^d$ and $0 \leq y \leq x$ then $y \in P$.

The *antiblocker* of a corner polytope is defined by

$$P^{\mathrm{abl}} = \{x \in \mathbb{R}^d_+ : \ x^\mathsf{T} y \leq 1 \forall y \in P\}.$$

Proposition 13.7.11 *The antiblocker of a corner polytope is a corner polytope. For every corner polytope P we have $(P^{\mathrm{abl}})^{\mathrm{abl}} = P$.*

The correspondence between faces of P and P^{abl} is more complicated than for the blocking polyhedra. The nonnegativity constraints $x_i \geq 0$ always define facets, and they don't correspond to vertices in the antiblocker. All other facets of P correspond to vertices of P^{abl}. Not every vertex of P defines a facet in P^{abl}. The origin is a trivial exceptional vertex, but there may be further exceptional vertices. We call a vertex v *dominated*, if there is another vertex w such that $v \leq w$. Now a vertex of P defines a facet of P^* if and only if it is not dominated.

13.7.4 Optimization

Let $P \subseteq \mathbb{R}^n$ be an ascending polyhedron. It is easy to see that P has a unique point which is closest to the origin. We'll see later on that this point has combinatorial significance in some cases. Right now, we state the following simple theorem that relates this point for the analogous point in the blocker.

Theorem 13.7.12 *Let $P \subseteq \mathbb{R}^n$ be an ascending polyhedron, and let $a \in P$ minimize the objective function $\sum_i x_i^2$. Let $\alpha = \sum_i a_i^2$ be the minimum value. Then $b = (1/\alpha)a$ is in the blocker P^{bl}, and it minimizes the objective function $\sum_i x_i^2$ over P^{bl}.*

Proof.

□

13.8 Semidefinite optimization

Linear programming has been one of the most fundamental and successful tools in optimization and discrete mathematics. Its applications include exact and approximation algorithms, as well as structural results and estimates. The key point is that linear programs are very efficiently solvable, and have a powerful duality theory.

Linear programs are special cases of convex programs; *semidefinite programs* are more general but still convex programs, to which many of the useful properties of linear programs extend. Recently, semidefinite programming arose as a generalization of linear programming with substantial novel applications. Again, it can be used both in proofs and in the design of exact and approximation algorithms. It turns out that various combinatorial optimization problems have semidefinite (rather than linear) relaxations which are still efficiently computable, but approximate the optimum much better. This fact has lead to a real breakthrough in approximation algorithms.

Semidefinite programs arise in a variety of ways: as certain geometric extremal problems, as relaxations (stronger than linear relaxations) of combinatorial optimization problems, in optimizing eigenvalue bounds in graph theory, as stability problems in engineering, etc.

For more comprehensive studies of issues concerning semidefinite optimization, see [209, 135].

13.8.1 Semidefinite programs

A semidefinite program is an optimization problem of the following form:

$$\begin{aligned} &\text{minimize } c^{\mathsf{T}} x \\ &\text{subject to } x_1 A_1 + \ldots x_n A_n - B \succeq 0 \end{aligned} \tag{13.16}$$

Here $A_1, \ldots, A_n, B$ are given symmetric $m \times m$ matrices, and $c \in \mathbb{R}^n$ is a given vector. We can think of $X = x_1 A_1 + \ldots x_n A_n - B$ as a matrix whose entries are linear functions of the variables.

As usual, any choice of the values x_i that satisfies the given constraint is called a *feasible solution*. A solution is *strictly feasible*, if the matrix X is positive definite. We denote by v_{primal} the supremum of the objective function.

The special case when $A_1, \ldots, A_n, B$ are diagonal matrices is just a "generic" linear program, and it is very fruitful to think of semidefinite programs as generalizations of linear programs. But there are important technical differences. The following example shows that, unlike in the case of linear programs, the supremum may be finite but not a maximum, i.e., not attained by any feasible solution.

Example 13.8.1 Consider the semidefinite program

$$\begin{aligned} &\text{minimize } x_1 \\ &\text{subject to } \begin{pmatrix} x_1 & 1 \\ 1 & x_2 \end{pmatrix} \succeq 0 \end{aligned}$$

The semidefiniteness condition boils down to the inequalities $x_1, x_2 \geq 0$ and $x_1 x_2 \geq 1$, so the possible values of the objective function are all negative real numbers. Thus $v_{\text{primal}} = 0$, but the supremum is not assumed.

As in the theory of linear programs, there are a large number of equivalent formulations of a semidefinite program. Of course, we could consider minimization instead of maximization. We could stipulate that the x_i are nonnegative, or more generally, we could allow additional linear constraints on the variables x_i (inequalities and/or equations). These could be incorporated into the form above by extending the A_i and B with new diagonal entries.

We could introduce the entries of A as variables, in which case the fact that they are linear functions of the original variables translates into linear relations between them. Straightforward linear algebra transforms (13.16) into an optimization problem of the form

$$\begin{aligned} &\text{maximize } C \cdot X \\ &\text{subject to } X \succeq 0 \\ &\qquad D_1 \cdot X = d_1 \\ &\qquad \vdots \\ &\qquad D_k \cdot X = d_k, \end{aligned} \tag{13.17}$$

where $C, D_1, \ldots, D_k$ are symmetric $m \times m$ matrices and $d_1, \ldots, d_k \in \mathbb{R}$. Note that $C \cdot X$ is the general form of a linear combination of entries of X, and so $D_i \cdot X = d_i$ is the general form of a linear equation in the entries of X.

It is easy to see that we would not get any substantially more general problem if we allowed linear inequalities in the entries of X in addition to the equations.

13.8.2 Fundamental properties of semidefinite programs

We begin with the semidefinite version of the Farkas Lemma:

Lemma 13.8.2 [Homogeneous version] *Let $A_1, \ldots, A_n$ be symmetric $m \times m$ matrices. The system*

$$x_1 A_1 + \cdots + x_n A_n \succ 0$$

has no solution in $x_1, \ldots, x_n$ if and only if there exists a symmetric matrix $Y \neq 0$ such that

$$\begin{aligned} A_1 \cdot Y &= 0 \\ A_2 \cdot Y &= 0 \\ &\vdots \\ A_n \cdot Y &= 0 \\ Y &\succeq 0. \end{aligned}$$

There is an inhomogeneous version of this lemma.

Lemma 13.8.3 [Inhomogeneous version] *Let $A_1, \ldots, A_n, B$ be symmetric $m \times m$ matrices. The system*

$$x_1 A_1 + \ldots x_n A_n - B \succ 0$$

has no solution in $x_1, \ldots, x_n$ if and only if there exists a symmetric matrix $Y \neq 0$ such that

$$\begin{aligned} A_1 \cdot Y &= 0 \\ A_2 \cdot Y &= 0 \\ &\vdots \\ A_n \cdot Y &= 0 \\ B \cdot Y &\geq 0 \\ Y &\succeq 0. \end{aligned}$$

Given a semidefinite program (13.16), one can formulate the *dual program*:

$$\begin{aligned} &\text{maximize } B \cdot Y \\ &\text{subject to } A_1 \cdot Y = c_1 \\ &\qquad\qquad A_2 \cdot Y = c_2 \\ &\qquad\qquad \vdots \\ &\qquad\qquad A_n \cdot Y = c_m \\ &\qquad\qquad Y \succeq 0. \end{aligned} \tag{13.18}$$

Note that this too is a semidefinite program in the general sense. We denote by v_{dual} the infimum of the objective function.

With this notion of duality, the Duality Theorem holds in the following sense (see e.g. [207, 202, 203]):

Theorem 13.8.4 *Assume that both the primal and the dual semidefinite programs have feasible solutions. Then $v_{\text{primal}} \leq v_{\text{dual}}$. If, in addition, the primal program (say) has a strictly feasible solution, then the dual optimum is attained and $v_{\text{primal}} = v_{\text{dual}}$. In particular, if both programs have strictly feasible solutions, then the supremum resp. infimum of the objective functions are attained.*

The following *complementary slackness conditions* also follow from this argument.

Proposition 13.8.5 *Let x be a feasible solution of the primal program and Y, a feasible solution of the dual program. Then $v_{\text{primal}} = v_{\text{dual}}$ and both x and Y are optimal solutions if and only if $Y(\sum_i x_i A_i - B) = 0$.*

The following example shows that the somewhat awkward conditions about the strictly feasible solvability of the primal and dual programs cannot be omitted (see [163] for a detailed discussion of conditions for exact duality).

Example 13.8.6 Consider the semidefinite program

$$\begin{aligned} &\text{minimize} x_1 \\ &\text{subject to} \begin{pmatrix} 0 & x_1 & 0 \\ x_1 & x_2 & 0 \\ 0 & 0 & x_1+1 \end{pmatrix} \succeq 0 \end{aligned}$$

The feasible solutions are $x_1 = 0$, $x_2 \geq 0$. Hence v_{primal} is assumed and is equal to 0. The dual program is

$$\begin{aligned} &\text{maximize} -Y_{33} \\ &\text{subject to} Y_{12} + Y_{21} + Y_{33} = 1 \\ &\qquad\qquad Y_{22} = 0 \\ &\qquad\qquad Y \succeq 0. \end{aligned}$$

The feasible solutions are all matrices of the form

$$\begin{pmatrix} a & 0 & b \\ 0 & 0 & 0 \\ b & 0 & 1 \end{pmatrix}$$

where $a \geq b^2$. Hence $v_{\text{dual}} = -1$.

13.8.3 Algorithms for semidefinite programs

There are two essentially different algorithms known that solve semidefinite programs in polynomial time: the *ellipsoid method* and *interior point/barrier methods*. Both of these have many variants, and the exact technical descriptions are quite complicated; so we restrict ourselves to describing the general principles underlying these algorithms, and to some comments on their usefulness. We ignore numerical problems, arising from the fact that the optimum solutions may be irrational and the feasible regions may be very small; we refer to [162, 163] for discussions of these problems.

The first polynomial time algorithm to solve semidefinite optimization problems in polynomial time was the ellipsoid method. Let K be a convex body (closed, compact, convex, full-dimensional set) in $\mathbb{R}^N$. We set $S(K,t) = \{x \in \mathbb{R}^N : \ d(x,K) \leq t\}$, where d denotes euclidean distance. Thus $S(0,t)$ is the ball with radius t about 0.

A *(weak) separation oracle* for a convex body $K \subseteq \mathbb{R}^N$ is an oracle whose input is a rational vector $x \in \mathbb{R}^N$ and a rational $\varepsilon > 0$; the oracle either asserts that $x \in S(K,\varepsilon)$ or returns an "almost separating hyperplane" in the form of a vector $0 \neq y \in \mathbb{R}^N$ such that $y^{\mathsf{T}}x > y^{\mathsf{T}}z - \varepsilon|y|$ for all $z \in K$.

If we have a weak separation oracle for a convex body (in practice, any subroutine the realizes this oracle) then we can use the ellipsoid method to optimize any linear objective function over K [83]:

Theorem 13.8.7 *Let K be a convex body in $\mathbb{R}^n$ and assume that we know two real numbers $R > r > 0$ such that $S(0,r) \subseteq K \subseteq S(0,R)$. Assume further that we have a weak separation oracle for K. Let a (rational) vector $c \in \mathbb{R}^n$ and an error bound $0 < \varepsilon < 1$ be also given. Then we can compute a (rational) vector $x \in \mathbb{R}^n$ such that $x \in K$ and $c^{\mathsf{T}}x \geq c^T z - \varepsilon$ for every $y \in K$. The number of calls on the oracle and the number of arithmetic operations in the algorithm are polynomial in $\log(R/r) + \log(1/\varepsilon) + n$.*

This method can be applied to solve semidefinite programs in polynomial time, modulo some technical conditions. (Note that some complications arise already from the fact that the optimum value is not necessarily a rational number, even if all parameters are rational. A further warning is example 13.8.6.)

Assume that we are given a semidefinite program (13.16) with rational coefficients and a rational error bound $\varepsilon > 0$. Also assume that we know a rational, strictly feasible solution $\tilde{x}$, and a bound $R > 0$ for the coordinates of an optimal solution. Then the set K of feasible solutions is a closed, convex, bounded, full-dimensional set in $\mathbb{R}^n$. It is easy to compute a small ball around x_0 that is contained in K.

The key step is to design a separation oracle for K. Given a vector x, we need only check whether $x \in K$ and if not, find a separating hyperplane. Ignoring numerical problems, we can use Gaussian elimination to check whether the matrix $Y = \sum_i x_i A_i - B$ is positive semidefinite. If it is, then $x \in K$. If not, the algorithm also returns a vector $v \in \mathbb{R}^m$ such that $v^{\mathsf{T}}Yv < 0$. Then $\sum_i x_i v^{\mathsf{T}} A_i v = v^{\mathsf{T}} B v$ is a separating hyperplane. (Because of numerical problems, the error bound in the definition of the weak separation oracle is needed.)

Thus using the ellipsoid method we can compute, in time polynomial in $\log(1/\varepsilon)$ and in the number of digits in the coefficients and in x_0, a feasible solution x such that the value of the objective function is at most $v_{\text{primal}} + \varepsilon$.

Unfortunately, the above argument gives an algorithm which is polynomial, but hopelessly slow, and practically useless. Still, the flexibility of the ellipsoid method makes it an inevitable tool in proving the *existence* (and not much more) of a polynomial time algorithm for many optimization problems.

Semidefinite programs can be solved in polynomial time and also *practically efficiently* by interior point methods [158, 5, 6]. The key to this method is the following property of the determinant of positive semidefinite matrices.

Lemma 13.8.8 *The function F defined by*

$$F(Y) = -\log\det\,(Y)$$

is convex and analytic in the interior of the semidefinite cone $\mathcal{P}_n$, and tends to ∞ at the boundary.

The algorithm can be described very informally as follows. The feasible domain of our semidefinite optimization problem is of the form $K = \mathcal{P}_n \cap A$, where A is an affine subspace of symmetric matrices. We want to minimize a linear function $C \cdot X$ over $X \in K$. The good news is that K is convex. The bad news is that the minimum will be attained on the boundary of K, and this boundary can have a very complicated structure; it is neither smooth nor polyhedral. Therefore, neither gradient-type methods nor the methods of linear programming can be used to minimize $C \cdot X$.

The main idea of barrier methods is that instead of minimizing $C^{\mathsf{T}}X$, we minimize the function $F_C(X) = F(X) + \lambda C^{\mathsf{T}}X$ for some $\lambda > 0$. Since F_λ tends to infinity on the boundary of K, the minimum will be attained in the interior. Since F_λ is convex and analytic in the interior, the minimum can be very efficiently computed by a variety of numerical methods (conjugate gradient etc.)

Of course, the point we obtain this way is not what we want, but if λ is large it will be close. If we don't like it, we can increase λ and use the minimizing point for the old F_λ as the starting point for a new gradient type algorithm. (In practice, we can increase λ after each iteration of this gradient algorithm.)

One can show that (under some technical assumptions about the feasible domain) this algorithm gives an approximation of the optimum with relative error ε in time polynomial in $\log(1/\varepsilon)$ and the size of the presentation of the program. The proof of this depends on a further rather technical property of the determinant, called "self-concordance". We don't go into the details, but refer to the articles [6, 202, 203] and the book [157].

Bibliography

[1] B.M. Ábrego, S. Fernández-Merchant: A lower bound for the rectilinear crossing number. Manuscript, 2003.

[2] O. Aichholzer, F. Aurenhammer and H. Krasser: On the crossing number of complete graphs, in: *Proc.* 18^{th} *Ann. ACM Symp. Comp. Geom., Barcelona, Spain* (2002), pp. 19–24.

[3] M. Ajtai, Chvtal, V.; Newborn, M.; Szemerdi, E. (1982). "Crossing-free subgraphs". Theory and Practice of Combinatorics: 912, North-Holland Mathematics Studies, vol. 60.

[4] D. J. Aldous and J. Fill: *Reversible Markov Chains and Random Walks on Graphs* (book in preparation); URL for draft at http://www.stat.Berkeley.edu/users/aldous

[5] F. Alizadeh: Combinatorial optimization with semi-definite matrices, in: *Integer Programming and Combinatorial Optimization* (Proceedings of IPCO '92), (eds. E. Balas, G. Cornuéjols and R. Kannan), Carnegie Mellon University Printing (1992), 385–405.

[6] F. Alizadeh, Interior point methods in semidefinite programming with applications to combinatorial optimization, *SIAM J. Optim.* **5** (1995), 13–51.

[7] F. Alizadeh, J.-P. Haeberly, and M. Overton: Complementarity and nondegeneracy in semidefinite programming, in: *Semidefinite Programming*, *Math. Programming* Ser. B, **77** (1997), 111–128.

[8] N. Alon: Eigenvalues and expanders, *Combinatorica* **6**(1986), 83–96.

[9] N. Alon, The Shannon capacity of a union, *Combinatorica* **18** (1998), 301–310.

[10] N. Alon and E. Győri: The number of small semispaces of a finite set of points in the plane, *J. Combin. Theory Ser. A* **41** (1986), 154–157.

[11] N. Alon and N. Kahale: Approximating the independence number via the ϑ-function, *Math. Programming* **80** (1998), Ser. A, 253–264.

[12] N. Alon and V. D. Milman: λ_1, isoperimetric inequalities for graphs and superconcentrators, *J. Combinatorial Theory B* **38**(1985), 73–88.

[13] N. Alon, L. Lovász: Unextendible product bases, *J. Combin. Theory Ser. A* **95** (2001), 169–179.

[14] N. Alon, P.D. Seymour and R. Thomas: Planar separators, *SIAM J. Disc. Math.* **7** (1994), 184–193.

[15] N. Alon and J.H. Spencer: *The Probabilistic Method*, Wiley, New York, 1992.

[16] E. Andre'ev: On convex polyhedra in Lobachevsky spaces, *Mat. Sbornik*, Nov. Ser. **81** (1970), 445–478.

[17] S. Arora, C. Lund, R. Motwani, M. Sudan, M. Szegedy: Proof verification and hardness of approximation problems *Proc. 33rd FOCS* (1992), 14–23.

[18] L. Asimov and B. Roth: The rigidity of graphs, *Trans. Amer. Math. Soc.*, **245** (1978), 279–289.

[19] R. Bacher and Y. Colin de Verdière: Multiplicités des valeurs propres et transformations étoile-triangle des graphes, *Bull. Soc. Math. France* **123** (1995), 101-117.

[20] E.G. Bajmóczy, I. Bárány, On a common generalization of Borsuk's and Radon's theorem, *Acta Mathematica Academiae Scientiarum Hungaricae* 34 (1979) 347–350.

[21] A.I. Barvinok: Problems of distance geometry and convex properties of quadratic maps, *Discrete and Comp. Geometry* **13** (1995), 189–202.

[22] A.I. Barvinok: A remark on the rank of positive semidefinite matrices subject to affine constraints, *Discrete and Comp. Geometry* (to appear)

[23] M. Bellare, O. Goldreich, and M. Sudan: Free bits, PCP and non-approximability—towards tight results, *SIAM J. on Computing* **27** (1998), 804–915.

[24] I. Benjamini and L. Lovász: Global Information from Local Observation, *Proc. 43rd Ann. Symp. on Found. of Comp. Sci.* (2002) 701–710.

[25] I. Benjamini and L. Lovász: Harmonic and analytic frunctions on graphs, *Journal of Geometry* **76** (2003), 3–15.

[26] I. Benjamini and O. Schramm: Harmonic functions on planar and almost planar graphs and manifolds, via circle packings. *Invent. Math.* **126** (1996), 565–587.

[27] I. Benjamini and O. Schramm: Random walks and harmonic functions on infinite planar graphs using square tilings, *Ann. Probab.* **24** (1996), 1219–1238.

[28] T. Böhme, On spatial representations of graphs, in: *Contemporary Methods in Graph Theory* (R. Bodendieck, ed.), BI-Wiss.-Verl. Mannheim, Wien/Zurich, 1990, pp. 151–167.

[29] J. Bourgain: On Lipschitz embedding of finite metric spaces in Hilbert space, *Israel J. Math.* **52** (1985), 46–52.

[30] P. Brass: On the maximum number of unit distances among n points in dimension four, *Intuitive geometry* Bolyai Soc. Math. Stud. **6**, J. Bolyai Math. Soc., Budapest (1997), 277–290.

[31] P. Brass: Erdős distance problems in normed spaces, *Comput. Geom.* **6** (1996), 195–214.

[32] P. Brass: The maximum number of second smallest distances in finite planar sets. *Discrete Comput. Geom.* **7** (1992), 371–379.

[33] P. Brass, G. Károlyi: On an extremal theory of convex geometric graphs (manuscript)

[34] G.R. Brightwell, E. Scheinerman: Representations of planar graphs, *SIAM J. Discrete Math.* **6** (1993) 214–229.

[35] R.L. Brooks, C.A.B. Smith, A.H. Stone and W.T. Tutte, The dissection of rectangles into squares. Duke Math. J. 7, (1940). 312–340.

[36] A.E. Brouwer and W.H. Haemers: Association Schemes, in *Handbook of Combinatorics* (ed. R. Graham, M. Grötschel, L. Lovász), Elsevier Science B.V. (1995), 747–771.

[37] P. Cameron, J.M. Goethals, J.J. Seidel and E.E. Shult, Line graphs, root systems and elliptic geometry, *J. Algebra* **43** (1976), 305-327.

[38] A.K. Chandra, P. Raghavan, W.L. Ruzzo, R. Smolensky and P. Tiwari: The electrical resistance of a graph captures its commute and cover times, *Proc. 21st ACM STOC* (1989), 574–586.

[39] S.Y. Cheng, Eigenfunctions and nodal sets, *Commentarii Mathematici Helvetici* 51 (1976) 43–55.

[40] J. Cheriyan, J.H. Reif: Directed *s-t* numberings, rubber bands, and testing digraph k-vertex connectivity, *Proceedings of the Third Annual ACM-SIAM Symposium on Discrete Algorithms*, ACM, New York (1992), 335–344.

[41] M. Chudnovsky, N. Robertson, P. Seymour and R. Thomas: The strong perfect graph theorem, *Annals of Math.*, **164** (2006), 51-229.

[42] Y. Colin de Verdière: Sur la multiplicité de la première valeur propre non nulle du laplacien, *Comment. Math. Helv.* **61** (1986), 254–270.

[43] Y. Colin de Verdière: Sur un nouvel invariant des graphes et un critère de planarité, *Journal of Combinatorial Theory, Series B* **50** (1990), 11–21. English translation: On a new graph invariant and a criterion for planarity, in: *Graph Structure Theory* (Robertson and P. D. Seymour, eds.), Contemporary Mathematics, Amer. Math. Soc., Providence, RI (1993), 137–147.

[44] Y. Colin de Verdière: Une principe variationnel pour les empilements de cercles, *Inventiones Math.* **104** (1991) 655–669.

[45] Y. Colin de Verdière: Multiplicities of eigenvalues and tree-width of graphs, *J. Combin. Theory Ser. B* **74** (1998), 121–146.

[46] Y. Colin de Verdière: *Spectres de graphes*, Cours Spécialisés **4**, Société Mathématique de France, Paris, 1998.

[47] R. Connelly: A counterexample to the rigidity conjecture for polyhedra, *Publications Mathmatiques de l'IHS* **47** (1977), 333–338.

[48] G. Csizmadia: The multiplicity of the two smallest distances among points, *Discrete Math.* **194** (1999), 67–86.

[49] C. Delorme and S. Poljak: Laplacian eigenvalues and the maximum cut problem, *Math. Programming* **62** (1993) 557–574.

[50] C. Delorme and S. Poljak: Combinatorial properties and the complexity of max-cut approximations, *Europ. J. Combin.* **14** (1993), 313–333.

[51] T.L. Dey: Improved bounds for planar k-sets and related problems, *Discrete and Computational Geometry* **19** (1998), 373–382.

[52] M. Deza and M. Laurent: *Geometry of Cuts and Metrics*, Springer Verlag, 1997.

[53] P.D. Doyle and J.L. Snell: *Random walks and electric networks,* Math. Assoc. of Amer., Washington, D.C. 1984.

[54] R.J. Duffin: Basic properties of discrete analytic functions, *Duke Math. J.* **23** (1956), 335–363.

[55] R.J. Duffin: Potential theory on the rhombic lattice, *J. Comb. Theory* **5** (1968) 258–272.

[56] R.J. Duffin and E.L. Peterson: The discrete analogue of a class of entire functions, *J. Math. Anal. Appl.* **21** (1968) 619–642.

[57] I.A. Dynnikov and S.P. Novikov: Geometry of the Triangle Equation on Two-Manifolds, *Mosc. Math. J.* **3** (2003), 419–438.

[58] H. Edelsbrunner, P. Hajnal: A lower bound on the number of unit distances between the vertices of a convex polygon, *J. Combin. Theory Ser. A* **56** (1991), 312–316.

[59] H.G. Eggleston: Covering a three-dimensional set with sets of smaller diameter, *J. London Math. Soc.* **30** (1955), 11–24.

[60] P. Erdős, L. Lovász, G. Simmons, and E. Strauss: Dissection graphs of planar point sets, in: *A Survey of Combinatorial Theory*, North Holland, Amsterdam, (1973), pp. 139–149.

[61] P. Erdős, On sets of distances of n points, *Amer. Math. Monthly* **53** (1946), 248–250.

[62] P. Erdős: On sets of distances of n points in Euclidean space, *Publ. Math. Inst. Hung. Acad. Sci.* **5** (1960), 165–169.

[63] P. Erdős: Gráfok páros körüljárású részgráfjairól (On bipartite subgraphs of graphs, in Hungarian), *Mat. Lapok* **18** (1967), 283–288.

[64] P. Erdős: On some applications of graph theory to geometry, *Canad. J. Math.* **19** (1967), 968–971.

[65] P. Erdős and N.G. de Bruijn: A colour problem for infinite graphs and a problem in the theory of relations, *Nederl. Akad. Wetensch. Proc. – Indagationes Math.* **13** (1951), 369–373.

[66] P. Erdős, F. Harary and W.T. Tutte, On the dimension of a graph *Mathematika* **12** (1965), 118–122.

[67] P. Erdős, L. Lovász, K. Vesztergombi: On the graph of large distances, *Discrete and Computational Geometry* **4** (1989) 541–549.

[68] P. Erdős, L. Lovász, K. Vesztergombi: The chromatic number of the graph of large distances, in: *Combinatorics*, Proc. Coll. Eger 1987, Coll. Math. Soc. J. Bolyai **52**, North-Holland, 547–551.

[69] P. Erdős and M. Simonovits: On the chromatic number of geometric graphs, *Ars Combin.* **9** (1980), 229–246.

[70] U. Feige: Approximating the Bandwidth via Volume Respecting Embeddings, Tech. Report CS98-03, Weizmann Institute (1998).

[71] J. Ferrand, Fonctions préharmoniques et fonctions préholomorphes. (French) *Bull. Sci. Math.* **68**, (1944). 152–180.

[72] C.M. Fiduccia, E.R. Scheinerman, A. Trenk, J.S. Zito: Dot product representations of graphs, *Discrete Math.* **181** (1998), 113–138.

[73] H. de Fraysseix, J. Pach, R. Pollack: How to draw a planar graph on a grid, *Combinatorica* **10** (1990), 41–51.

[74] P. Frankl, H. Maehara: The Johnson-Lindenstrauss lemma and the sphericity of some graphs, *J. Combin. Theory Ser. B* **44** (1988), 355–362.

[75] P. Frankl, H. Maehara: On the contact dimension of graphs, *Disc. Comput. Geom.* **3** (1988) 89–96.

[76] P. Frankl, R.M. Wilson: Intersection theorems with geometric consequences, *Combinatorica* **1** (1981), 357–368.

[77] S. Friedland and R. Loewy, Subspaces of symmetric matrices containing matrices with multiple first eigenvalue, *Pacific J. Math.* **62** (1976), 389–399.

[78] Z. Füredi: The maximum number of unit distances in a convex n-gon, *J. Combin. Theory Ser. A* **55** (1990), 316–320.

[79] M. X. Goemans and D. P. Williamson: Improved approximation algorithms for maximum cut and satisfiability problems using semidefinite programming, *J. Assoc. Comput. Mach.* **42** (1995), 1115–1145.

[80] M. Grötschel, L. Lovász and A. Schrijver: The ellipsoid method and its consequences in combinatorial optimization, *Combinatorica* **1** (1981), 169-197.

[81] M. Grötschel, L. Lovász and A. Schrijver: Polynomial algorithms for perfect graphs, *Annals of Discrete Math.* **21** (1984), 325-256.

[82] M. Grötschel, L. Lovász, A. Schrijver: Relaxations of vertex packing, *J. Combin. Theory B* **40** (1986), 330-343.

[83] M. Grötschel, L. Lovász, A. Schrijver: Geometric Algorithms and Combinatorial Optimization, Springer (1988).

[84] B. Grünbaum: A proof of Vázsonyi's conjecture, *Bull. Res. Council Israel, section A* **6** (1956), 77–78.

[85] Richard K. Guy: A combinatorial problem, *Bull. Malayan Math. Soc.* **7** (1967), 68–72.

[86] H. Hadwiger: Ein Überdeckungssatz für den Euklidischen Raum, *Portugaliae Math. 4* (1944), 140–144.

[87] W. Haemers: On some problems of Lovász concerning the Shannon capacity of a graph, *IEEE Trans. Inform. Theory* **25** (1979), 231–232.

[88] J. Håstad: Some optimal in-approximability results, *Proc. 29th ACM Symp. on Theory of Comp.*, 1997, 1–10.

[89] J. Håstad: Clique is hard to approximate within a factor of $n^{1-\varepsilon}$, *Acta Math.* **182** (1999), 105–142.

[90] Z.-X. He, O. Schramm: Hyperbolic and parabolic packings, *Discrete Comput. Geom.* **14** (1995), 123–149.

[91] Z.-X. He, O. Schramm: On the convergence of circle packings to the Riemann map, *Invent. Math.* **125** (1996), 285–305.

[92] A. Heppes: Beweis einer Vermutung von A. Vázsonyi, *Acta Math. Acad. Sci. Hung.* **7** (1956), 463–466.

[93] A. Heppes, P. Révész: A splitting problem of Borsuk. (Hungarian) *Mat. Lapok* **7** (1956), 108–111.

[94] A.J. Hoffman, Some recent results on spectral properties of graphs, in: *Beiträge zur Graphentheorie,* Teubner, Leipzig (1968) 75–80.

[95] H. van der Holst: A short proof of the planarity characterization of Colin de Verdière, *Journal of Combinatorial Theory, Series B* 65 (1995) 269–272.

[96] H. van der Holst: *Topological and Spectral Graph Characterizations*, Ph.D. Thesis, University of Amsterdam, Amsterdam, 1996.

[97] H. van der Holst, M. Laurent, A. Schrijver, On a minor-monotone graph invariant, *Journal of Combinatorial Theory, Series B* 65 (1995) 291–304.

[98] H. van der Holst, L. Lovász and A. Schrijver: On the invariance of Colin de Verdière's graph parameter under clique sums, *Linear Algebra and its Applications*, **226–228** (1995), 509–518.

[99] H. van der Holst, L. Lovász, A. Schrijver: The Colin de Verdière graph parameter, in: *Graph Theory and Combinatorial Biology*, Bolyai Soc. Math. Stud. **7**, János Bolyai Math. Soc., Budapest (1999), 29–85.

[100] A.J. Hoffman: On graphs whose least eigenvalue exceeds $-1-\sqrt{2}$, *Linear Algebra and Appl.* **16** (1977), 153–165.

[101] A.J. Hoffman: On eigenvalues and colorings of graphs. *Graph Theory and its Applications*, Academic Press, New York (1970), 79–91.

[102] H. van der Holst, L. Lovász, A. Schrijver: The Colin de Verdière graph parameter, in: *Graph Theory and Combinatorial Biology*, Bolyai Soc. Math. Stud. **7**, János Bolyai Math. Soc., Budapest (1999), 29–85.

[103] H. Hopf, E. Pannwitz: *Jahresbericht der Deutsch. Math. Vereinigung* **43** (1934), 2 Abt. **114**, Problem 167.

[104] R. Isaacs, Monodiffric functions, *Natl. Bureau Standards App. Math. Series* **18** (1952), 257–266.

[105] W.B. Johnson, J. Lindenstrauss: Extensions of Lipschitz mappings into a Hilbert space. *Conference in modern analysis and probability* (New Haven, Conn., 1982), 189–206, Contemp. Math., 26, Amer. Math. Soc., Providence, R.I., 1984.

[106] M.R. Jerrum and A. Sinclair: Approximating the permanent, *SIAM J. Comput.* **18**(1989), 1149–1178.

[107] J. Jonasson and O. Schramm: On the cover time of planar graphs *Electron. Comm. Probab.* **5** (2000), paper no. 10, 85–90.

[108] F. Juhász: The asymptotic behaviour of Lovász' ϑ function for random graphs, *Combinatorica* **2** (1982) 153–155.

[109] J. Kahn, G, Kalai: A counterexample to Borsuk's conjecture. *Bull. Amer. Math. Soc.* **29** (1993), 60–62.

[110] D. Karger, R. Motwani, M. Sudan: Approximate graph coloring by semidefinite programming, *Proc. 35th FOCS* (1994), 2–13.

[111] B. S. Kashin and S. V. Konyagin: On systems of vectors in Hilbert spaces, *Trudy Mat. Inst. V.A.Steklova* **157** (1981), 64–67; English translation: *Proc. of the Steklov Inst. of Math.* (AMS 1983), 67–70.

[112] R. Kenyon: The Laplacian and Dirac operators on critical planar graphs, *Inventiones Math* **150** (2002), 409-439.

[113] R. Kenyon and J.-M. Schlenker: Rhombic embeddings of planar graphs with faces of degree 4 (math-ph/0305057).

[114] D. E. Knuth: The sandwich theorem, *The Electronic Journal of Combinatorics* **1** (1994), 48 pp.

[115] P. Koebe: Kontaktprobleme der konformen Abbildung, *Berichte über die Verhandlungen d. Sächs. Akad. d. Wiss.*, Math.–Phys. Klasse, **88** (1936) 141–164.

[116] V. S. Konyagin, Systems of vectors in Euclidean space and an extremal problem for polynomials, *Mat. Zametky* **29** (1981), 63–74. English translation: *Math. Notes of the Academy USSR* **29** (1981), 33–39.

[117] A. Kotlov: A spectral characterization of tree-width-two graphs, *Combinatorica* **20** (2000)

[118] A. Kotlov, L. Lovász, S.!Vempala, The Colin de Verdière number and sphere representations of a graph, *Combinatorica* **17** (1997) 483–521.

[119] G. Laman (1970): On graphs and rigidity of plane skeletal structures, *J. Engrg. Math.* **4**, 331–340.

[120] P. Lancaster, M. Tismenetsky, *The Theory of Matrices, Second Edition, with Applications*, Academic Press, Orlando, 1985.

[121] D.G. Larman, C.A. Rogers: The realization of distances within sets in Euclidean space. *Mathematika* **19** (1972), 1–24.

[122] M. Laurent and S. Poljak: On the facial structure of the set of correlation matrices, *SIAM J. on Matrix Analysis and Applications* **17** (1996), 530–547.

[123] F.T. Leighton: *Complexity Issues in VLSI*, Foundations of Computing Series, MIT Press, Cambridge, MA (1983).

[124] F.T. Leighton and S. Rao, An approximate max-flow min-cut theorem for uniform multicommodity flow problems with applications to approximation algorithms, *Proc. 29th Annual Symp. on Found. of Computer Science,* IEEE Computer Soc., (1988), 422-431.

[125] N. Linial, E. London, Y. Rabinovich: The geometry of graphs and some of its algorithmic applications, *Combinatorica* **15** (1995), 215–245.

[126] N. Linial, L. Lovász, A. Wigderson: Rubber bands, convex embeddings, and graph connectivity, *Combinatorica* **8** (1988) 91-102.

[127] R.J. Lipton, R.E. Tarjan: A Separator Theorem for Planar Graphs, *SIAM J. on Applied Math.*, **36** (1979), 177–189.

[128] L. Lovász: Normal hypergraphs and the perfect graph conjecture, *Discrete Math.* **2** (1972), 253-267.

[129] L. Lovász: Combinatorial Problems and Exercises, Second Edition, Akadémiai Kiadó - North Holland, Budapest, 1991

[130] L. Lovász: Flats in matroids and geometric graphs, in: *Combinatorial Surveys*, Proc. 6th British Comb. Conf., Academic Press (1977), 45-86.

[131] L. Lovász: On the Shannon capacity of graphs, *IEEE Trans. Inform. Theory* **25** (1979), 1–7.

[132] L. Lovász: Singular spaces of matrices and their applications in combinatorics, *Bol. Soc. Braz. Mat.* **20** (1989), 87–99.

[133] L. Lovász: Random walks on graphs: a survey, in: *Combinatorics, Paul Erdős is Eighty*, Vol. 2 (ed. D. Miklos, V. T. Sos, T. Szonyi), János Bolyai Mathematical Society, Budapest, 1996, 353–398.

[134] L. Lovász: Steinitz representations of polyhedra and the Colin de Verdière number, *J. Comb. Theory B* **82** (2001), 223–236.

[135] L. Lovász: Semidefinite programs and combinatorial optimization, in: *Recent Advances in Algorithms and Combinatorics,* CMS Books Math./Ouvrages Math. SMC *11*, Springer, New York (2003), 137-194.

[136] L. Lovász: Graph minor theory, *Bull. Amer. Math. Soc.* **43** (2006), 75–86.

[137] L. Lovász, M. Saks, A. Schrijver: Orthogonal representations and connectivity of graphs, *Linear Alg. Appl.* **114/115** (1989), 439–454; A Correction: Orthogonal representations and connectivity of graphs, *Linear Algebra Appl.* **313** (2000) 101–105.

[138] L. Lovász, A. Schrijver: A Borsuk theorem for antipodal links and a spectral characterization of linklessly embedable graphs, *Proceedings of the American Mathematical Society* **126** (1998) 1275–1285.

[139] L. Lovász, A. Schrijver: On the null space of a Colin de Verdière matrix, *Annales de l'Institute Fourier* **49**, 1017–1026.

[140] L. Lovász and B. Szegedy: Szemerédi's Lemma for the analyst, *Geom. Func. Anal.* **17** (2007), 252–270.

[141] L. Lovász, K. Vesztergombi: Geometric representations of graphs, to appear in *Paul Erdos and his Mathematics,* (ed. G. Halász, L. Lovász, M. Simonovits, V.T. Sós), Bolyai Society–Springer Verlag.

[142] L. Lovász, K. Vesztergombi, U. Wagner, E. Welzl: Convex quadrilaterals and k-sets, in: *Towards a Theory of Geometric Graphs*, (J. Pach, Ed.), AMS Contemporary Mathematics **342** (2004), 139–148.

[143] L. Lovász and P. Winkler: Mixing times, in: *Microsurveys in Discrete Probability* (ed. D. Aldous and J. Propp), DIMACS Series in Discr. Math. and Theor. Comp. Sci., **41** (1998), 85–133.

[144] Lubotzky, Phillips, Sarnak

[145] P. Mani, Automorphismen von polyedrischen Graphen, *Math. Annalen* **192** (1971), 279–303.

[146] Margulis [expander]

[147] J. Matoušek: On embedding expanders into l^p spaces. *Israel J. Math.* **102** (1997), 189–197.

[148] C. Mercat, Discrete Riemann surfaces and the Ising model. *Comm. Math. Phys.* **218** (2001), no. 1, 177–216.

[149] C. Mercat: Discrete polynomials and discrete holomorphic approximation (2002)

[150] C. Mercat: Exponentials form a basis for discrete holomorphic functions (2002)

[151] G.L. Miller, W. Thurston: Separators in two and three dimensions, *Proc. 22th ACM STOC* (1990), 300–309.

[152] G.L. Miller, S.-H. Teng, W. Thurston, S.A. Vavasis: Separators for sphere-packings and nearest neighbor graphs. *J. ACM* **44** (1997), 1–29.

[153] B. Mohar and S. Poljak: Eigenvalues and the max-cut problem, *Czechoslovak Mathematical Journal* **40** (1990), 343–352.

[154] B. Mohar and C. Thomassen: Graphs on surfaces. Johns Hopkins Studies in the Mathematical Sciences. Johns Hopkins University Press, Baltimore, MD, 2001.

[155] A. Mowshowitz: Graphs, groups and matrices, in: *Proc. 25th Summer Meeting Canad. Math. Congress, Congr. Numer.* **4** (1971), Utilitas Mathematica, Winnipeg, 509-522.

[156] C. St.J. A. Nash-Williams, Random walks and electric currents in networks, *Proc. Cambridge Phil. Soc.* **55** (1959), 181–194.

[157] Yu. E. Nesterov and A. Nemirovsky: *Interior-point polynomial methods in convex programming,* Studies in Appl. Math. **13**, SIAM, Philadelphia, 1994.

[158] M. L. Overton: On minimizing the maximum eigenvalue of a symmetric matrix, *SIAM J. on Matrix Analysis and Appl.* **9** (1988), 256–268.

[159] M. L. Overton and R. Womersley: On the sum of the largest eigenvalues of a symmetric matrix, *SIAM J. on Matrix Analysis and Appl.* **13** (1992), 41–45.

[160] G. Pataki: On the rank of extreme matrices in semidefinite programs and the multiplicity of optimal eigenvalues, *Math. of Oper. Res.* **23** (1998), 339–358.

[161] S. Poljak and F. Rendl: Nonpolyhedral relaxations of graph-bisection problems, DIMACS Tech. Report 92-55 (1992).

[162] L. Porkoláb and L. Khachiyan: On the complexity of semidefinite programs, *J. Global Optim.* **10** (1997), 351–365.

[163] M. Ramana: An exact duality theory for semidefinite programming and its complexity implications, in: *Semidefinite programming. Math. Programming* Ser. B, **77** (1997), 129–162.

[164] A. Recski: *Matroid theory and its applications in electric network theory and in statics*, Algorithms and Combinatorics, 6. Springer-Verlag, Berlin; Akadémiai Kiadó, Budapest, 1989.

[165] J. Reiterman, V. Rödl, E. Šinajová: Embeddings of graphs in Euclidean spaces, *Discr. Comput. Geom.* **4** (1989), 349–364.

[166] J. Reiterman, V. Rödl and E. Šinajová: On embedding of graphs into Euclidean spaces of small dimension, *J. Combin. Theory Ser. B* **56** (1992), 1–8.

[167] J. Richter-Gebert: *Realization spaces of polytopes*, Lecture Notes in Mathematics **1643**, Springer-Verlag, Berlin, 1996.

[168] N. Robertson, P.D. Seymour: Graph minors III: Planar tree-width, *J. Combin. Theory B* **36** (1984), 49–64.

[169] N. Robertson, P.D. Seymour, Graph minors XX: Wagner's conjecture (preprint, 1988).

[170] N. Robertson, P. Seymour, R. Thomas: A survey of linkless embeddings, in: *Graph Structure Theory* (N. Robertson, P. Seymour, eds.), Contemporary Mathematics, American Mathematical Society, Providence, Rhode Island, 1993, pp. 125–136.

[171] N. Robertson, P. Seymour, R. Thomas: Hadwiger's conjecture for K_6-free graphs, *Combinatorica* **13** (1993) 279–361.

[172] N. Robertson, P. Seymour, R. Thomas: Sachs' linkless embedding conjecture, *Journal of Combinatorial Theory, Series B* **64** (1995) 185–227.

[173] B. Rodin and D. Sullivan: The convergence of circle packings to the Riemann mapping, *J. Diff. Geom.* **26** (1987), 349–360.

[174] H. Sachs,

[175] H. Sachs, Coin graphs, polyhedra, and conformal mapping, *Discrete Math.* **134** (1994), 133–138.

[176] F. Saliola and W. Whiteley: Constraining plane configurations in CAD: Circles, lines and angles in the plane *SIAM J. Discrete Math.* **18** (2004), 246–271.

[177] H. Saran: *Constructive Results in Graph Minors: Linkless Embeddings*, Ph.D. Thesis, University of California at Berkeley, 1989.

[178] W. Schnyder: Embedding planar graphs on the grid, *Proc. First Annual ACM-SIAM Symp. on Discrete Algorithms*, SIAM, Philadelphia (1990), 138–148;

[179] O. Schramm: Existence and uniqueness of packings with specified combinatorics, *Israel Jour. Math.* **73** (1991), 321–341.

[180] O. Schramm: How to cage an egg, *Invent. Math.* **107** (1992), 543–560.

[181] O. Schramm: Square tilings with prescribed combinatorics, *Israel Jour. Math.* **84** (1993), 97–118.

[182] A. Schrijver, Minor-monotone graph invariants, in: *Combinatorics 1997* (P. Cameron, ed.), Cambridge University Press, Cambridge, 1997, to appear.

[183] D. Singer: Rectilinear crossing numbers, Manuscript (1971), available at `http://www.cwru.edu/artsci/math/singer/home.html`.

[184] P.M. Soardi: *Potential Theory on Infinite Networks*, Lecture notes in Math. **1590**, Springer-Verlag, Berlin–Heidelberg, 1994.

[185] J. Spencer, E. Szemerédi and W.T. Trotter: Unit distances in the Euclidean plane, *Graph Theory and Combinatorics*, Academic Press, London–New York (1984), 293–303.

[186] D.A. Spielman and S.-H. Teng: Disk Packings and Planar Separators, *12th Annual ACM Symposium on Computational Geometry* (1996), 349–358.

[187] D.A. Spielman and S.-H. Teng: Spectral Partitioning Works: Planar graphs and finite element meshes, *Proceedings of the 37th Annual IEEE Conference on Foundations of Comp. Sci.* (1996), 96–105.

[188] E. Steinitz: Polyeder und Raumabteilungen, in: *Encyclopädie der Math. Wissenschaften* **3** (Geometrie) 3AB12, 1–139, 1922.

[189] J.W. Sutherland: *Jahresbericht der Deutsch. Math. Vereinigung* **45** (1935), 2 Abt. 33.

[190] K. Stephenson: *Introduction to Circle Packing: The Theory of Discrete Analytic Functions*, Cambridge University Press, 2005.

[191] S. Straszewicz: Sur un problème geometrique de Erdős, *Bull. Acad. Polon. Sci. CI.III* **5** (1957), 39–40.

[192] M. Szegedy: A note on the θ number of Lovász and the generalized Delsarte bound, *Proc. 35th FOCS* (1994), 36–39.

[193] L.A. Székely: Measurable chromatic number of geometric graphs and sets without some distances in Euclidean space, *Combinatorica* **4** (1984), 213–218.

[194] L.A. Székely: Crossing numbers and hard Erdős problems in discrete geometry. *Combin. Probab. Comput.* **6** (1997), 353–358.

[195] W.P. Thurston: *Three-dimensional Geometry and Topology*, Princeton Mathematical Series **35**, Princeton University Press, Princeton, NJ, 1997.

[196] G. Tóth: Point sets with many k-sets, *Discr. Comput. Geom.* **26** (2001), 187–194.

[197] W.T. Tutte: How to draw a graph, *Proc. London Math. Soc.* **13** (1963), 743–768.

[198] K. Vesztergombi: On the distribution of distances in finite sets in the plane, *Discrete Math.*, 57, [1985], 129–145.

[199] K. Vesztergombi: Bounds on the number of small distances in a finite planar set, *Stu. Sci. Acad. Hung.*, 22 [1987], 95–101.

[200] K. Vesztergombi: On the two largest distances in finite planar sets, *Discrete Math.* **150** (1996) 379–386.

[201] W. Whiteley, Infinitesimally rigid polyhedra, *Trans. Amer. Math. Soc.* **285** (1984), 431–465.

[202] L. Vandeberghe and S. Boyd: Semidefinite programming, in: *Math. Programming: State of the Art* (ed. J. R. Birge and K. G. Murty), Univ. of Michigan, 1994.

[203] L. Vandeberghe and S. Boyd: Semidefinite programming. SIAM Rev. 38 (1996), no. 1, 49–95.

[204] R.J. Vanderbei and B. Yang: The simplest semidefinite programs are trivial, *Math. of Oper. Res.* **20** (1995), no. 3, 590–596.

[205] E. Welzl: Entering and leaving j-facets, *Discr. Comput. Geom.* **25** (2001), 351–364.

[206] W. Whiteley: ***

[207] H. Wolkowitz: Some applications of optimization in matrix theory, *Linear Algebra and its Applications* **40** (1981), 101–118.

[208] H. Wolkowicz: Explicit solutions for interval semidefinite linear programs, *Linear Algebra Appl.* **236** (1996), 95–104.

[209] H. Wolkowicz, R. Saigal and L. Vandenberghe: *Handbook of semidefinite programming. Theory, algorithms, and applications.* Int. Ser. Oper. Res. & Man. Sci., **27** (2000) Kluwer Academic Publishers, Boston, MA.

[210] D. Zeilberger: A new basis for discrete analytic polynomials, *J. Austral. Math. Soc.* Ser. A **23** (1977), 95-104.

[211] D. Zeilberger: A new approach to the theory of discrete analytic functions, *J. Math. Anal. Appl.* **57** (1977), 350-367.

[212] D. Zeilberger and H. Dym: Further properties of discrete analytic functions, *J. Math. Anal. Appl.* **58** (1977), 405-418.