Geometric separator theorems and applications - DRAFT

Warren D. Smith*

Nicholas C. Wormald†

wds@research.NJ.NEC.COM, nick@ms.unimelb.edu.au

Abstract —

We find a large number of "geometric separator theorems" such as: I: Given N disjoint iso-oriented squares in the plane, there exists a rectangle with $\le 2N/3$ squares inside, $\le 2N/3$ squares outside, and $\le (4+o(1))\sqrt{N}$ partly in & out. II: There exists a rectangle that is crossed by the minimal spanning tree of N sites in the plane at $\le (4 \cdot 3^{1/4} + o(1))\sqrt{N}$ points, having $\le 2N/3$ sites inside and outside. These theorems yield a large number of applications, such as subexponential algorithms for traveling salesman tour and rectilinear Steiner minimal tree, new point location algorithms, and new upper and lower bound proofs for "planar separator theorems." We also survey graph separator theorems and study the problem of covering convex bodies with smaller copies.

Keywords — Separator theorems, Steiner minimal tree, Rectilinear Steiner minimal tree, traveling salesman tour, minimum spanning tree, spanners, banyans, minimum matching, centerpoints, sandwich theorems, cube tiling, Hadwiger hypothesis, pigeonhole principle, torus graphs, planar graphs, point location, computational geometry, VC dimension, covering problems, discrete isoperimetric theorems, Delaunay triangulation, Gabriel graph, universal graphs, nested dissection, squared rectangles, obstacle avoiding shortest paths.

Contents

*NECI, 4 Independence Way, Princeton NJ 08540

†Dept of Mathematics and Statistics, University of Melbourne, Parkville VIC 3052, Australia. Research supported by the ARC.

1 Introduction

THIS PAPER is remarkable because it contains around 10^3 or 10^4 theorems. This is possible because of the "chinese menu effect:" if you combine theorem components from columns A, B, C, and D, the result is a very large number of possible theorems.

"Columns A,B,C,D" are respectively:

A. Theorems about covering d dimensional objects by smaller versions of themselves. The simplest example: a d-box may be divided into two smaller d-boxes.

B. Separator theorems about geometrical objects. One of our simplest examples (and it is related to the example in 'A' above) is: Given N interior-disjoint squares in the plane, there exists a rectangle (both the squares and the rectangle have sides oriented parallel to the coordinate axes) such that $\le 2N/3$ squares's interiors are entirely inside it, $\le 2N/3$ are entirely outside, and $\le (4+o(1))\sqrt{N}$ are partly inside and partly outside. In this theorem "2/3" is best possible.

C. Separator theorems about geometrical graphs. A simple example: given N sites in the plane, there exists a rectangle R (with sides at angle 45° to the coordinate axes) such that the E-edge rectilinear Steiner minimal tree (RSMT) of the sites, has $\le 2E/3$ of its edges entirely inside R, $\le 2E/3$ of its edges entirely outside, and $\le (4+o(1))\sqrt{E}$ edges cross the boundary of R. This result follows from the example in 'B' above once you know the "diamond property" of RSMT edges e: the squares with diagonal e, are interior disjoint.

D. Applications of separator theorems. The example in 'C' above leads to a fairly simple algorithm for finding the RSMT of N sites in the plane, in subexponential worst case time $N^{O(\sqrt{N})}$.

Actually not all the four menu choices are always completely independent; the examples above, in fact, each depend on the previous one. Nevertheless, there is enough freedom of choice at each stage to still get a very large number of results.

We now discuss each of these four menu columns in more detail.

A. Covering bodies with smaller versions of themselves (§3). There are a lot of choices because we may consider different kinds of "bodies," e.g. iso-oriented (or not) boxes, cubes, simplices, spheres, L_p balls, general convex bodies. Additional freedom comes from the fact that there are three interesting versions of "versions:" scaled translates, scaled copies with both rotations and translations allowed, and volume reducing affine transformations.

B. Separator theorems about geometrical objects (§6). The example result with squares (§4.2) is, roughly speaking, proven by a three-step argument:

1. "Sup trick" allows us to bound the number of squares outside the separating rectangle.
2. "Pigeonhole principle" allows us to bound the number of squares inside the separating rectangle. (This step depends on the covering results in "column A.")
3. Randomizing argument allows us to bound the number of squares partly inside and partly outside.

Each of the three steps in this argument is highly generalizable, and consequently we get a large number of variations (§4.2.1) of the basic theorem, including:

1. The squares could be other things: cubes, simplices, octahedra, general convex body of CV aspect ratio (definition 4) bounded by B. In the latter case, "$(4+o(1))\sqrt{N}$" changes to "$(4+o(1))B\sqrt{N}$."
2. The separator could be other things than an iso-oriented rectangular box: a cube, simplex, octahedron, sphere, or general convex body of bounded aspect ratio. (The constants may change.)
3. The objects need not be interior-disjoint; it will suffice if at most a constant number κ of them cover any point ("κ-thick"). In some variants, even weaker conditions ("κ-overloaded," "(λ,κ)-thick;" see §1.1.4) will suffice. In such cases the bound "$4B\sqrt{N}$" becomes "$4B\sqrt{N\kappa}$."
4. We need not stay in the plane; it is also possible to go to higher dimensions d. The $O(B\sqrt{N\kappa})$ bound is replaced[1] by $O(B\kappa^{1/d}N^{1-1/d})$ with an implied constant depending on d and the specifics of the theorem. In some versions the 1/3-2/3 split weakens so that the "1/3" is replaced by a decreasing function of d.
5. Instead of splitting the *objects* 1/3-2/3 at worst, we can instead split some arbitrary measure on the plane, which need not have anything to do with the objects.
6. We also have linear time algorithmic versions of our theorems (some with weaker constants).

In contrast, a separator theorem proved by Miller and Thurston depends on special properties of the "inversive group" and thus works only for separating *spheres* with *spheres*. Their proof also cannot be made to work for κ-overloaded and (λ,κ)-thick spheres. The present paper makes it clear that the Miller-Thurston result is a special case of a much more general phenomenon. But the Miller & Thurston result makes up for its specialness with better performance – that is, better constants (with better dependence on the dimension d). Unfortunately, these constants were not determined, nor even crudely estimated, by Miller et al. [132], so we redo their proof, this time working them out explicitly. After we did this, we learned much of our analysis duplicated work by Spielman and Teng [167]. However, we can go beyond [167] by getting better balance ratios than Miller et al.'s $(d+1):1$; e.g. we can get $2:1$ using ellipsoids whose principal axes vary by at most a constant factor.

We also have (§4.1) a simple separator theorem for iso-oriented d-boxes (now *not* required to have bounded aspect ratios).

C. Separator theorems about geometrical graphs (§5, §6). We use our separator theorems about geometrical objects as tools to obtain these.

Thus in (§5) we get new proofs of the famed "planar separator theorem" of Lipton and Tarjan. In some cases we can get new record upper and lower bounds on the constants, and new kinds of separators – for example, we get a Jordan curve separator theorem for Torus graphs.

The RSMT example theorem is in §6. Variations include: d dimensions, other separators than boxes, (e.g. circles) and other graphs than RSMT, e.g.: Steiner Minimal tree (SMT), optimal traveling salesman tour (TST), Minimum spanning tree (MST), all-nearest neighbor graph (ANN), minimum matching (MM), "spanners," and "banyans," and L_1-norm versions of all of these. The Gabriel graph (GG) and Delaunay triangualtion (DT) are examples of graphs without geometric separators (§6.8).

The proof idea is to somehow associate an object (or objects) with each SMT edge, and prove that these objects have bounded aspect ratios and are disjoint (or merely "κ-thick," "κ-overloaded," or "(λ,κ)-thick;" these are weaker versions of the word "disjoint" defined in §1.1.4). Then by use of our previous separator theorems about objects, one gets a separator theorem about SMT (or TST, MM; whatever) edges.

D. Finally, applications of our theorems, including new algorithms and data structures, are discussed in §7.

1. New algorithm to compute optimal traveling salesman tour of N sites in d-space. Runs in $2^{d^{O(d)}}N^{dN^{1-1/d}}$ steps and consumes $O(Nd)$ space. This simplifies and generalizes a previous 2D algorithm by Smith [162].
2. New algorithm to compute optimal rectilinear Steiner tree of N sites in d-space. Runs in $2^{d^{O(d)}}N^{d^3N^{1-1/d}}$ steps and consumes $O(Nd)$ space.

[1] In this paper, all O's are valid uniformly as all the quantities inside them vary independently over their full allowed ranges – in this case, as $\kappa \geq 1$, $B \geq 1$, $d \geq 2$, and $N \geq 2$ vary. In the few cases in which we want to disallow variation we will indicate so in the text, or write, e.g., $O_d(N)$, meaning that d is to be regarded as fixed while N varies unboundedly, and the implied constant factor depends on d.

This simplifies and generalizes a previous 2D algorithm by Smith [164].

3. New algorithm to compute $1 + O(N^{-p})$ times optimal length Euclidean Steiner tree for N sites in d-space. Runs in $(N^{1+p}d^d)^{O(d^{5/2}N^{1-1/d})}$ time.

4. New data structure for point location among N disjoint (or κ-thick) d-objects with aspect ratios bounded by B. A data structure requiring $d^{O(d)}N$ storage is constructed in $d^{O(d)}N \log N$ preprocessing time, which supports point location queries in $O(\kappa B^d O(d)^d T + \log N)$ time per query, where it takes time T to test if a point is in one given object.

   Also, the data structure may be "dynamized" to allow insertion and deletion of objects (but respecting κ-thickness). In this case, the (amortized) time bound for a query increases by a factor of $\log N$, the space requirement is affected by only a constant factor, and the insertion and deletion times are $d^{O(d)} \log N$.

5. New algorithm for point location among N disjoint (or κ-thick) iso-oriented d-boxes.

6. New algorithm for obstacle avoiding short paths among boundedly thick obstacles of bounded aspect ratio in the plane. E.g. this may be solved in preprocessing time $O(N^{3/2} \log N)$ with storage $O(N \log N)$ and query time $O(\log N)$ to report a path 3.01 times longer than optimal (in the L_1 obstacle avoiding metric) between any two given points.

7. Algorithm to determine all intersection relationships among N κ-thick d-objects with aspect ratios bounded by B. The runtime is $(\log N + \kappa B^d C)d^{O(d)}N$ and the memory requirement is $d^{O(d)}N$. Here C is the amount of time to test if a point is inside just one specified object.

8. Separator theorems for intersection graphs of κ-thick d-objects with aspect ratios bounded by B, and consequently the ability to c-color such graphs in $(c-1)^{O(d\kappa^{1/d}BN^{1-1/d})}$ time (or prove impossibility), solve sparse linear systems with such graph structure in $O([d\kappa^{1/d}BN^{1-1/d}]^\omega)$ time ($\omega \in (2, 3]$ is the exponent in the runtime of $n \times n$ matrix multiplication; [53] showed $\omega \leq 2.376$), and find "universal" sparse graphs containing all such graphs.

Some of these algorithms and data structures look highly practical.

1.1 Notation and basic definitions

We employ the useful abbreviation WLOG, meaning "without loss of generality." Logarithms: $\ln x = \log_e x$, $\lg x = \log_2 x$, and $\log x$ intentionally leaves the base of the logarithm unspecified, although using e will work in all our uses. For definitions of geometrical graphs, see §6.1. We sometimes denote the boundary of B by ∂B.

Definition 1 *An "iso-oriented d-box" is a cartesian product of d 1-dimensional real intervals.*

1.1.1 Aspect ratio

Roughly speaking, the aspect ratio of an object is the ratio of its longest dimension to its shortest.

Definition 2 *The "diameter" of a compact subset S of some metric space is the supremum of the distance between any pair of points in S. If S lies in a euclidean space, its "width" is the infemal separation between two parallel hyperplanes it lies between.*

Definition 3 *The "DW aspect ratio" of a compact set in $\mathbf{R}^d$ is the ratio of its Diameter to its Width.*

The following alternative definition, based on Cube Volume, is better for many purposes of this paper.

Definition 4 *The "CV aspect ratio" of a compact set in $\mathbf{R}^d$ is the dth root of the ratio of the d-volume of the smallest enclosing iso-oriented d-cube, to the set's own d-volume.*

Thus the CV aspect ratio of an iso-cube is 1, and everything else has a CV aspect ratio > 1. The DW aspect ratio of a ball is 1, and everything else has a DW aspect ratio ≥ 1. (If the dimension d is fixed, then all reasonable definitions of aspect ratio are equivalent to within constant factors.)

1.1.2 Volumes of various d-dimensional objects

Definition 5 *Let $\newcommand{\fb}{\bullet}\fb_d = \pi^{d/2}/(d/2)!$ be the d-volume of a d-ball of radius 1 [161]. Also let $\circ_d = d\bullet_d$ be the $(d-1)$-surface of a d-ball of radius 1.*

Remark. If d is odd, the factorial function will be applied to half-integral argument, which is no problem if you know that $(-1/2)! = \sqrt{\pi}$, $0! = 1$, and $n! = n \cdot (n-1)!$.

Definition 6

$$\triangle_d = \frac{\sqrt{1+d}}{d!}(1 + \frac{1}{d})^{d/2} \tag{1}$$

is the d-volume of a d-dimensional regular simplex inscribed in a d-ball of radius 1.

1.1.3 Separation notions

Definition 7 *To "a-separate" N things, means to partition them into disjoint sets of cardinality $\leq a$. To "a-weight separate" them means that the sets have weight $\leq a$, for some notion of a weight function.*

1.1.4 Generalizations of the notion of "disjoint" sets

Definition 8 *A collection of point sets are "κ-thick" if no point is common to more than κ sets. (In particular, 1-thickness is equivalent to disjointness.)*

Definition 9 *For $\lambda > 1$ define "(λ, κ)-thick:" no point is covered by more than κ of the objects whose linear dimensions vary by a factor at most λ.*

Definition 10 *A collection of compact measurable sets are "κ-overloaded" if for any subcollection of the objects, the sum of their individual measures is at most a constant κ times the measure of their union.*

Note that (λ, κ)-thick and κ-overloaded sets can be infinitely thick. See figure 1.

Figure 1: Some collections of objects.

1.2 Plan of the rest of the paper

In §2*[2], we survey graph separator theorems.

In §3, we survey (and contribute new results to) the area of covering convex bodies by smaller versions of themselves.

In §4, we reach the core of the paper – separator theorems about geometrical objects. There are 3 main classes of such theorems: for iso-oriented d-boxes, objects of bounded aspect ratio (such as d-cubes), and for spheres (Miller & Thurston's theorem); we also survey (and present new) separator results of a topological nature. We also prove lower bounds to show that some of our theorems (in particular, the 1/3-2/3 split) are best possible. The results of §3 are important as tools to get the results in §4, but the methods used in §3 to derive those tools, do not need to be understood to understand §4.

In §2.7*, we discuss separator theorems for planar graphs (also, Torus and Klein bottle graphs) and show some beautiful connections to geometry. Some of this was already known [167] [163]. We get some new upper and lower bounds, as well as new understanding. It's probably possible to get separator theorems for graphs of bounded genus in a similar manner, but we leave this as an open problem (§8).

In §6, we obtain separator theorems about geometrical graphs such as 2-optimal traveling salesman tours, Steiner minimal trees, minimum spanning trees, etc. Although we devised ad hoc arguments for each of these particular graphs, our results (§6.7) about "spanners" and "banyans" show that a very wide class of approximately minimal-length graphs have separator theorems.

Finally, §7 shows how to apply the separator theorems of §4 and §6 to get new algorithms, data structures, and theorems.

§8 lists some interesting open problems arising from or highlit by this work.

1.3 Mathematical tools used

A surprising number of mathematical techniques come into play:

- Topological arguments (§4.3) including theorems of Borsuk about spheres and balls, Brouwer's fixed point theorem, and "ham sandwich" and "Rado point" theorems;
- Arguments based on "VC dimension" (§4.7.4);
- Graph geometrizations via "squared rectangles" (§5.2.1) and Koebe circles (§5.2.2);
- Minkowski's "polar map" mapping an oriented half-space $\{\vec{x} \in \mathbf{R}^d \mid \vec{a} \cdot \vec{x} \le 1\}$ to the vector $\vec{a}$. For example, each point on the surface of a convex d-body has a (possibly non-unique) oriented support hyperplane, and applying the "polar map" to these hyperplanes yields the boundary of the "polar body" – another[3] d-dimensional convex body. If we follow the application of the polar map by a renormalization to unit length, we map each point on the surface of a convex d-body to a point $\vec{a}/|\vec{a}|$ (or a spherically convex set with more than one point) on the surface of a d-ball of radius 1.
- New cube tilings (§4.6).
- And of course "divide and conquer," (which is what separator theorems are all about).

But perhaps overshadowing all of these is the humble "pigeonhole principle" (which, in its simplest form, states that if N pigeonholes contain P pigeons, then a pigeonhole with $\ge \lceil P/N \rceil$ pigeons must exist!) whose power, in this paper, is impressive.

2 Survey of graph separator theorems (1978-1997)

Definition 11 *An "$(\alpha, \beta(V))$-separator theorem" for a specified class of graphs asserts that for each graph G in the class having V vertices, there exists a subset C, $|C| \le \beta(V)$, called the "separator," of the vertices whose removal (along with the incident edges) divides the graph into parts of weight $\le \alpha$. (For some notion of positive "weight," where the weight of the whole graph is 1. The default notion is just the fractional number of vertices.)*

Similarly an *edge separator* is a small set of *edges* whose removal divides the graph into parts of weight less than α.

Edge separator theorems are better to have than vertex separator theorems because an edge separator instantly implies a vertex separator of the same cardinality.

A short survey covering graph separator theorems before 1978 may be found in [125].

The present paper is primarily concerned with geometric (as opposed to graphical) separator theorems, but the two are related (we'll see in §2.7 and §7.10 how the latter can sometimes be derived from the former).

As one example (before we start) of a geometric separator theorem, we mention

[2] In this sketch, "*" denotes sections peripheral to the "main line" of argument, which may be skipped by those uninterested in them.

[3] The polar body to B may also be defined slickly as $\{\vec{x} \in \mathbf{R}^d \mid \vec{x} \cdot \vec{y} \le 1 \;\; \forall \vec{y} \in B\}$.

Theorem 12 (Atallah & Chen [8]) *Given N interior disjoint iso-oriented rectangles in the plane, there exists an iso-oriented monotonic "staircase" made of $\leq N$ line segments, each vertical or horizontal, such that at most $\lceil N/2 \rceil$ of the rectangles lie on either side of the staircase, and the staircase avoids all the rectangle interiors. There is an algorithm to find this staircase in $O(N \log N)$ time.*

2.1 Trees and polygons

Any V-vertex tree has a single vertex (its "centroid") whose removal cuts it into pieces with $< V/2$ vertices.

Indeed, if any positive real weights are assigned to the vertices and edges, then each piece WLOG has $< 1/2$ of the total weight.

If the tree is binary and has positive weights on its vertices, there is a single edge (the "splitting edge") whose removal cuts it into pieces with $\leq 2/3$ of the weight. The splitting edge is always incident on the centroid. (For trees with all non-leaf valencies $\leq b$, $b \geq 2$ – or forests – an edge exists whose removal leaves pieces with $\leq (b-1)V/b$ vertices.)

These may be found algorithmically in $O(V)$ time. One may see [103] for these facts, although they apparently date at least back to C.Jordan (1838-1921) who used them in a paper on counting trees [101].

Chazelle [36] observed that a triangulated simple polygon has triangles whose planar dual graph is a binary tree, hence any such polygon has a "splitting diagonal" allowing splitting a simple N-gon into two polygons, each with $\leq 2 + 2N/3$ vertices.

Lingas [124] gave some separator theorems valid for polygons P with k polygonal holes. For example, given a triangulation of such a beast with N vertices, one may in $O(N)$ time find $k+1$ diagonals splitting P into two simple polygons, each of weight (for any nonnegative weights on vertices summing to 1, but not counting the weights on the splitting diagonals as part of the "weight" of the two smaller polygons) $\leq 2/3$.

Guibas et al. [91], in an appendix, showed that the complete recursive decomposition of a V-vertex binary tree via the splitting-edge theorem above could be found in $O(V)$ time. H.Booth [23] gave a simpler and clearer algorithm to accomplish a similar task. Neither source gave an algorithm to find the recursive centroid decomposition for general trees in $O(V)$ time, but this appears to be possible with the aid of the "link cut tree' data structure in [169] and the fact that runtime recurrences of the form $T(N) = T(2N/3) + T(N/3) + f(N)$ have solution $T(N) = O(N)$ if $c > 1 \Rightarrow \sum_{t \geq 0} c^{-t} f(c^t) < \infty$.

Bhatt & Leighton [14] showed that one could remove $O(k \log V)$ vertices from a V-vertex tree in which each vertex has one of k colors, to leave fragments which could be grouped into two sets, each set having an exactly equal number of vertices of each color. If instead of exact equality, we only require approximation to within a factor of 2, then only $O(k)$ vertices need to be removed.

2.2 k-Outerplanar graphs

The subgraph of a planar graph induced by only using the vertices on faces within distance $k-1$ (in the dual planar graph) of a given face, is "k-outerplanar."

The fact that a planar graph may be decomposed into "layers," each a k-outerplanar graph, was used by Brenda Baker in a famous paper [11]. It is also true that k-outerplanar graphs have $2k$-vertex separators which split the graph approximately 50-50 (with arbitrary nonnegative vertex weights).

2.3 Chordal graphs

Chordal graphs are graphs in which every cycle of length at least 4 has a chord. A chordal graph with V vertices and E edges can be cut in half by removing a set of $O(\sqrt{E})$ vertices [83] that may found in $O(E+V)$ time. A similar result holds if the vertices have non-negative weights and we want to bisect the graph by weight.

2.4 Series Parallel graphs

Series parallel graphs are graphs with no K_4 minor [175]. (See also 6.10 of [152].) They include 1-outerplanar graphs. (They arise when, e.g., analyzing the flow of control in a program.) They have a 1/3-2/3 separator, consisting of only 2 vertices [95], which may be found in $O(V)$ time.

2.5 Algebraic graphs

For fixed integers a, b, Ming Li [120] showed that the V vertex graph in which vertex i is joined to vertex j if $j = ai + b \bmod V$ has a 1/2-1/2 separator of cardinality $O(V/\sqrt{\log_a V})$. M.Klawe [107] showed a similar result.

The n-dimensional hypercube graph with $V = 2^n$ vertices and $E = n2^{n-1}$ edges has a 1/2-1/2 separator of size $\binom{n}{\lfloor n/2 \rfloor}$, i.e. of order $V/\sqrt{\log V}$.

These are mainly useful as negative results showing that these graphs cannot be "expanders."

2.6 Connection to eigenvalues; cutting manifolds

Cheeger [42] showed that

$$h(M)^2 \leq 4|\lambda_1(M)| \tag{2}$$

where $\lambda_1(M)$ is the second greatest eigenvalue (the greatest eigenvalue is $\lambda_0 = 0$, achieved by the constant eigenfunction) of the Laplacian operator in a compact Riemannian manifold M with "Neumann" boundary conditions at boundaries (if there are any), and

$$h(M) = \inf_S \frac{\text{area}(S)}{\min[\text{vol}(A), \text{vol}(B)]}, \tag{3}$$

where S is a hypersurface dividing M into two parts A and B. In other words, very "bass" manifolds have good separating cuts.

Buser showed that for each compact manifold, there exists a Riemannian metric for it, which causes (EQ 2) to become sharp. He [32] also showed a reversed version of (EQ 2)

$$|\lambda_1(M)| \;\le\; 2\sqrt{(d-1)|C|}\,h(M) \;+\; 10h(M)^2 \tag{4}$$

assuming the manifold is d-dimensional and that $C \le 0$ is a lower bound on its Ricci curvature.

Alon [3] found graph theoretic analogues of these inequalities. A "(V,k,c)-magnifier" is a V-vertex graph with maximal valence k such that every subset S, $|S| \le V/2$ of its vertices, has $\ge c|S|$ neighbors not in S. Let the "discrete laplacian" of a graph be the $V \times V$ matrix L with L_{ii} being the negated valence of vertex i, and $L_{ij} = 1$ if vertices i and j are adjacent, otherwise $L_{Ij} = 0$. L has all eigenvalues real (since it is symmetric) and maximum eigenvalue $\lambda_0 = 0$. Let $\lambda_1 \le 0$ be the second largest eigenvalue. ($\lambda_1 < 0$ iff the graph is connected.) Then

Fact 13 (Alon [3]) *If a V-vertex graph of maximal valence k has $|\lambda_1| = z \ge 0$, then it is a (V,k,c)-magnifier with*

$$\frac{2z}{k+2z} \le c \tag{5}$$

Conversely, if the graph is a (V,k,c)-magnifier, then

$$\frac{c^2}{4+2c^2} \le z. \tag{6}$$

In other words, graphs have small separators if and only if they are "bass."

2.7 Planar graphs

A deeper discussion of this (including proofs, geometrizations of the problem, and inter-relations among them) is in §5.

Lipton and Tarjan [125] showed a $(2/3, \sqrt{8V})$-separator theorem for planar graphs. This was improved by Djidjev [60] to $(2/3, \sqrt{6V})$. These partitions may be found algorithmically in $O(V)$ time.

Of course, these $(2/3, O(\sqrt{V}))$-separator theorems may be improved to $(1/2, O(\sqrt{V}))$, and all the weighted versions below may be improved to split any constant number of different weight measures simulataneously, at the cost of increasing the size of the separator by a constant factor. (This is easily accomplished by re-using the separator theorem on the parts.)

Miller [131] showed that if the planar graph were 2-connected, then C in the Lipton-Tarjan theorem could WLOG be taken to be a simple cycle, and the two parts would then be its interior and exterior. More generally, Miller allowed the assignment of arbitrary non-negative weights to the vertices, edges, and faces of the graph (but with no individual face weight $> 2/3$ of the total weight) and then found A, B, C as above but such that the total weight of A (or of B) was $\le 2/3$ of the total weight of the original graph.

H.Gazit and G.L.Miller [80] gave an algorithm yielding an *edge*-separator for planar graphs consisting of $\le 1.58\sqrt{\sum_i \nu_i^2}$ *edges* whose removal splits the graph into parts with $\le 2V/3$ vertices, and where ν_i is the valence of vertex i. The 1.58 has been brought down to 1, at the cost of changing 2/3 to 3/4, by Spielman and Teng [167]. (This result cannot be improved by more than a constant factor, as is shown by the "spoked wheel" graphs.)

Djidjev [61] showed that in $O(V)$ time one could find a 1/3-2/3 weight separator of *weight* $O(\sqrt{\sum_v W_v^2})$, where W_v are non-negative vertex weights, and this is optimal up to the constant factor.

Applications of the Lipton-Tarjan theorem are exhibited in [127], [126], [40], [45], [48], [14], [150], [174].

Improvements of the constants in the planar separator theorem have been found by a large number of authors and this will probably continue for some time into the future.

Unfortunately there is no total ordering one can easily use to distinguish a weaker theorem from a stronger one. Spielman and Teng [167] suggested using as a "figure of merit" (smaller is better) for an $(\alpha, \beta\sqrt{V})$-separator theorem

$$F_{\text{Nested diss.}} = \frac{\beta^3}{1-\alpha^{3/2}-(1-\alpha)^{3/2}} \tag{7}$$

which is proportional to an upper bound on the runtime for one common separator application – the "nested dissection" algorithm[4] for solving sparse systems of linear equations [127]. But different applications lead to different figures of merit[5]. For example, we may count k-colorings of a planar V-vertex graph in a time proportional to $T(V)$ where $T(V) = (k-1)^{\beta\sqrt{V}}[T(\alpha V) + T((1-\alpha)V)]$. Lipton and Tarjan [125] observed that an $(\alpha, \beta V^p)$-separator theorem with $\alpha > 1/2$ and $0 < p < 1$ could be converted to a $(1/2, \beta/(1-\alpha^p))$-separator theorem[6] by repeatedly separating the largest component and then rearranging the pieces. Both of these argue for the following figure of merit:

$$F_{50-50\text{ split}} = \frac{\beta}{1-\alpha^p}. \tag{8}$$

One of the most recent improvements is by Alon, Seymour, and Thomas [4], who got

Theorem 14 (Alon, Seymour, Thomas) *A maximal planar graph G with V vertices has a simple cycle separator with $\le \sqrt{4.5V}$ vertices, such that the total weight inside (or outside) the cycle, plus 1/2 the weight on the*

[4]Actually, they use this to compare two nonalgorithmic results, which seems ridiculous to us since the nested dissection application depends on having an *algorithmic* separator theorem.

[5] Even the nested dissection application would have a different figure of merit if $O(N^\omega)$-time "fast matrix multiplication" were used in the combining step: in this case we would find that N linear equations with planar graph sparsity structure could be solved in $O(N^{\omega/2})$ steps. ($\omega = 2.376$ [53] leads to $O(N^{1.188})$ runtime.) See §7.9. The 3's in (EQ 7) would have to be replaced by ω's.

[6]For ways to get 50-50 planar separator theorems with better constants, see [48] [49].

cycle, is $\leq 2/3$, for any assignment of positive weights, summing to 1, to the edges and vertices of G.

However, unlike [125], [60], and [131], Alon et al. [4] do not provide an $O(V)$-time algorithm to find the separator. They merely give an existence proof. Djidjev's $(2/3, \sqrt{6V})$ result still seems to be the best one that comes with an $O(V)$ algorithm.

Goodrich [87] showed that the complete recursive binary tree-like decomposition of a planar graph via a planar separator theorem, could be computed in $O(N)$ time by a complicated algorithm involving fairly sophisticated dynamic data structures.

2.8 Computational complexity of finding optimal and near optimal cuts and edge separators

A "cut" in a graph is a set of edges whose removal divides the graph into two disconnected subgraphs. Suppose the graph has non-negative real edge costs and vertex weights. The "quotient" of a cut is then the sum of the costs of the edges in the cut, divided by the minimum of the two weight sums (for the vertices in the two subgraphs). A quotient cut is "optimal" if it has minimal quotient.

If we had a polynomial time algorithm for approximating the value of the optimal quotient cut to within a factor of $F(V)$ (V is the number of vertices in the graph), then for any ϵ, $0 < \epsilon \leq 1/6$, we could also find, in polynomial time, an edge 2/3 separator whose cost was within a factor

$$(\frac{1}{\epsilon} + \ln\frac{2-3\epsilon}{3\epsilon})F(V) \tag{9}$$

times the cost of the optimal edge $(2/3-\epsilon)$-fraction separator. This arises by repeatedly chopping off pieces using approximately optimal quotient cuts.

Unfortunately, finding the optimal quotient cut, is, in general, NP-complete [31] [78], even if the edge costs and vertex weights are expressed in unary. But Leighton & Rao [119] showed how, in any V-vertex graph, to find a cut whose quotient was only $O(\log V)$ times optimal, in polynomial time, by solving multicommodity flow problems. Via (EQ 9) this leads to "pseudo-approximation algorithms" for optimal edge separators.

Park and Philips [143], building on earlier work of Rao, showed how, in polynomial time, to find an *optimal* "quotient cut" in a planar graph with positive real edge costs and positive integer vertex weights, provided the vertex weights are expressed in *unary.* This plus some further ideas led [79] to a polynomial time algorithm for finding edge 2/3-weight separators in these graphs within a factor of 2 of optimal cost.

On the other hand, [143] showed that with vertex weights and edge costs expressed as binary integers, finding minimal quotient cuts is NP-hard even for 2-outerplanar and series parallel graphs, and finding optimal edge separators is NP-hard even for graphs of "tree width" 2.

2.9 Graphs of low genus, or with excluded minors

Theorem 15 (Djidjev [59]) *For any V-vertex graph with genus g, there exists a partition of its vertices into three sets A, B, C with $|A| \leq 2V/3$, $|B| \leq 2V/3$, $|C| \leq \sqrt{(12g+6)V}$, such that there are no edges between A and B. Furthermore, a partition with $|A| \leq 2V/3$, $|B| \leq 2V/3$, $|C| \leq \sqrt{(21g+15)V}$ may be found by an algorithm running in $O(V)$ time. This algorithm does not need to know the value of g nor an embedding of the graph, and the constant in the "O" does not depend upon g.*

Djidjev even implemented this algorithm.

Gilbert, Hutchinson, and Tarjan [82] showed that V-vertex genus-g graphs exist, for every $g < V$, which do not have $o(\sqrt{gV})$ separator theorems, i.e. theorem 15 is tight to within a constant factor. (Counterexample graphs may be constructed by subdividing the triangles of an embedded complete graph into meshes.)

The fact that Djidjev's algorithm does not depend on having an embedding or knowing g makes it far more impressive and useful than the algorithm of [82], since the GRAPH GENUS problem is NP-complete [171] [172]. Indeed [43], it is NP-hard even to embed the graph in a surface of genus $g + O(V^{1-\epsilon})$ where g is the true genus and ϵ is any fixed positive real. (Note [59]: There exist $0 < c_1 < c_2 < c_3$ such that almost all V-vertex, c_1V-edge graphs have genus g obeying $c_2V < g < c_3V$ as $V \to \infty$.)

Nevertheless, finding an embedding of a graph on a genus-g surface (or producing a proof of impossibility) is claimed by Mohar [139] to be doable in c_gV time for some c_g growing very rapidly with g.

Aleksandrov and Djidjev [2] showed that indeed a complete decomposition of an *embedded* V-vertex, E-edge genus-g graph into pieces of size $\leq \epsilon V$ (arising by removing $O(\sqrt{(g+1/\epsilon)V})$ vertices) could be found in $O(V+E)$ time. This result is optimal. This generalizes the planar result of Goodrich [87]. But Goodrich also finds a recursive binary tree structure, which is not provided by [2], which in a sense only provides something like the leaves of such a tree.

Conjecture 16 (Alon, Seymour, Thomas) *Given a graph G whose V vertices have positive real weights summing to 1. Suppose G contains no instance of some "minor" with h vertices. Then G may be separated into disconnected components by the removal of $O(h\sqrt{V})$ vertices in such a way that each component's total weight is $\leq 2/3$.*

(E.g., the planar separator theorem would follow from this via Kuratowski's theorem that planar graphs exclude K_5 and $K_{3,3}$. The genus-g results would also follow via formulas of Ringel and Youngs [151] showing that the genus of K_h is of order h^2.)

This remains unsolved, but Alon, Seymour, Thomas [5] showed this with $h^{3/2}\sqrt{V}$ while Plotkin, Rao, Smith [146] showed this with $O(h\sqrt{V \log V})$. Both these papers come with polynomial time algorithms too, although not very fast ones.

One might further conjecture that there should be an $O(V)$ time algorithm, or even that Goodrich's complete recursive decomposition should be extensible to handle graphs with excluded minors.

Djidjev [59] remarked that his separator algorithm for graphs of genus g would, if applied to a graph G of possibly huge genus (he says genus and edge count of order $V^{5/3}$ are possible) and with the "safety cutoff" in step 1 of his algorithm (preventing anything from happening if $E > 4V$) disabled, *still* in $O(V+E)$ time find a separator with $\leq \sqrt{15V}$ vertices, provided G excluded $K_{3,3}$ as a minor.

All of the results above were proved by graph theoretic or combinatorial arguments. However, recently it has been realized that it is also possible to prove the planar separator theorem (and indeed to get some of the best available constants in it) by using *geometric* arguments. We'll discuss this in §2.7, where we have some new results, e.g. involving squares not circles, and allowing Torus and Klein bottle graphs. For geometric techniques concerning graphs of higher genus, see §8.1.

2.10 Other

Ming Li [121] gave and used a separator theorem for graphs arising from state transitions in computers.

3 Covering convex bodies with smaller copies

Definition 17 *A "convex d-body" is a compact convex set in $\mathbf{R}^d$, $1 \leq d < \infty$, that has interior points. It is "strictly convex" if its boundary contains no line segment; equivalently, if every support hyperplane has exactly one point of contact with the body.*

Definition 18 *A "smooth point" on the surface of a convex body is one with a unique tangent hyperplane. A convex body will be called "smooth" if every point on its surface is smooth.*

3.1 Smaller scaled translated copies

Definition 19 *A "homothet" of a set is a scaled and translated copy of it. For an "s-homothet," the scaling factor is s, and if s is left undefined or unspecified, we will take it to mean "for some s, $0 < s < 1$."*

3.1.1 Spheres

At least $d+1$ smaller balls are needed to cover a d-ball, as we'll see later.

Theorem 20 *The minimum value of r so that it is possible to cover a d-ball of radius 1 by $d+1$ balls of radius r, is $r = \sqrt{1-d^{-2}}$.*

Proof. One may verify (e.g. via explicit coordinates [55]) that if the balls are the circumballs of the $(d-1)$-faces of a regular d-simplex inscribed in the d-ball, we get a covering with this r value.

We will now prove that an r value at least this large is required, even merely to cover the *surface* of the unit ball. In fact, consider covering this surface with $d+1$ equal spherical caps of angular radius $\theta = \arcsin r$. Since $\theta < \pi$, the centers of the caps cannot all lie in a hemisphere, hence their convex hull is a simplex enclosing the origin. Also their "dual convex hull" (that is, the intersection of the halfspaces corresponding to the hyperplanes tangent to the sphere at the cap centers) is another simplex S enclosing the origin. Now it is known ([106] and 2.50-2.52 on page 506 of [137]) that the ratio ρ of the radius of the smallest ball enclosing a d-simplex, to the radius of the largest ball it encloses, is $\geq d$. Hence at least one vertex of S cannot be inside the ball of radius d concentric with the unit ball. This vertex corresponds to a "deep hole" in the covering and implies that to have a covering we must have $\cos\theta \leq 1/R \leq 1/d$, which implies $r \geq \sqrt{1-d^{-2}}$.

Indeed, since it is known that $\rho = d$ if and only if the simplex is regular, the optimal covering is unique. □

Remark. This theorem was previously known only when $d \leq 3$. Indeed, the minimal scaling factors for covering a circle by N congruent circles are known [130] when $N \leq 9$. If $90° \leq \theta < 180^circ$ the problem of covering the sphere with caps of angular diameter θ is the same problem as covering the ball with balls (make the boundary of a covering cap be the equator of a covering ball). The minimal-θ coverings of a sphere in 3D by N caps are known when $N \leq 7$ [130].

Alternatively, one may cover a d-ball of radius 1 with $2d$ balls each of radius $\sqrt{(d-1)/d}$. These balls are the circumballs of the faces of a regular d-cube inscribed in the unit ball. But this scaling factor is not known to be best possible except when $d = 2$.

Of course, ellipsoids are affinely equivalent to spheres and so everything we've said about spheres, one can say about ellipsoid homothets.

3.1.2 d-Simplices

WLOG (by an affine transform) the simplex is regular. A d-simplex may be covered by $d+1$ s-homothets, where the scaling factor is $s = d/(d+1)$. The homothets are crammed into the $d+1$ corners. (With any scaling factor below $d/(d+1)$, the center of the simplex would be uncovered.) This is easy to see if you consider the $d+1$ hyperplanes corresponding to the un-crammed faces of the smaller simplices, which are parallel to the far face and go through the center. You can't use fewer than $d+1$ copies without leaving one of the $d+1$ corners uncovered.

3.1.3 Regular d-octahedra (L_1-balls)

May be covered by $2d$ s-homothets with $s = (d-1)/d$. The copies are crammed into the $2d$ corners. Note, each face of the octahedron, which is a regular $(d-1)$-simplex, is covered according to the previous discussion, and with any smaller scaling factor, the face centers would be uncovered. Since when $d \geq 2$ the dihedral angles of the octahedron are $\geq 90°$ (in fact they are $\arccos((2-d)/d)$) and since the scaling factor is $\geq 1/2$, everything inside is covered once you see that the outside is. It is impossible to use $< 2d$ s-homothets for any $s < 1$ because a corner would be uncovered (consider L_1 distance).

3.1.4 d-Cubes

(Or by an affine transformation, any parallelipiped) may be covered by 2^d s-homothets with $s = 1/2$. It is impossible to use fewer for any scaling factor in $[1/2, 1)$ because one of the corners would be left uncovered (consider L_∞ distance). Scaling factor $s = 1/2$ is optimal by considering d-volume.

3.1.5 d objects do not suffice, in general

Fact 21 *It is impossible to cover any smooth convex body with only d s-homothets.*

Proof. This is because the surface is then going to be partitioned into d sets (possibly with overlap). By the famous "Borsuk antipodes theorem[7]" [26] from topology (cf. §4.3) at least one of these d sets must contain 2 "antipodal" (that is, with parallel tangent hyperplanes) points – preventing covering and leading to a contradiction. □

3.1.6 Any convex body, $d = 2$

Lassak [116] used a result of Zindler that a parallelogram with any two given diagonal directions may be inscribed in a 2D convex body, to show that any convex 2-body could be covered by 4 homothets, each scaled by $1/\sqrt{2}$. Here "4" is best possible (square) and with 4, "$1/\sqrt{2}$" is also (disc). Also, 4 are required only if the body is a parallelogram. Krotoszynski [114] showed that with 5 translated copies, a scaling factor of $1/2$ suffices.

3.1.7 Hadwiger hypothesis

H.Hadwiger conjectured in 1957 that the d-cube was the worst convex body, i.e. that $< 2^d$ translated smaller scaled copies of itself would suffice to cover any convex d-body – except for bodies affinely equivalent to a d-cube, which as we've seen require exactly 2^d.

This frustrating conjecture stands unsolved!

Boltyanskii [18] [19] showed the number of copies needed was always the same as the number of light sources needed to illuminate the surface of the body, where a light source at p is said to "illuminate" a surface point q if the ray $\vec{pq}$ intersects the body's interior, but only after it reaches q.

Another equivalent formulation arises from applying the "polar map" (§1.3). The number of copies needed is the same as the number of open halfspaces through the origin needed to cover the surface of the polar body, and in such a way that each (closed) face of the polar body is entirely contained in at least one of the halfspaces. Thus, the polar equivalent to Hadwiger's conjecture is: "The surface of any convex d-body B containing the origin may be covered by 2^d open halfspaces through the origin in such a way that every (closed) face of B is entirely contained in at least one of the halfspaces."

A related geometric separator conjecture is

Conjecture 22 *For any convex d-polytope with F faces, there exists a hyperplane with at least $\lfloor c_d F \rfloor$ faces entirely to each side of it. (Further conjecture: $c_d = 2^{-O(d)}$ suffices.)*

Chakerian and Stein [33] showed that for any convex d-body K, there exists a parallipiped P so that $P/d \subset K \subset P$, where "P/d" denotes a parallipiped concentric with P with $1/d$ times the linear dimensions. From this it instantly follows that d^d copies suffice. Lassak [117] improved this by about 13% to reduce the bound to

$$(d+1)d^{d-1} - (d-1)(d-2)^{d-1} \sim (1-e^{-2})d^d. \quad (10)$$

3.1.8 General "smooth" convex objects

Fact 23 *Any smooth convex body may be covered by $d+1$ homothets.*

Proof. Define the 1-1 correspondence[8] between points on the surface of the body and the surface of a ball: points correspond if they have same oriented tangent hyperplanes. Now, a $(1-\delta)$-homothet translated by ϵ will cover (in the limit $\epsilon \to 0+$, where a sufficiently tiny δ is chosen for each ϵ) the points on the surface of our body which correspond to a spherical cap of angular radius slightly below $\pi/2$. Now considering our previous discussion on "spheres" proves the claim. □

Also, a more careful argument [19] will show that $d+1$ homothets suffice even if the body's surface contains up to d non-smooth points. (If $d = 3$, up to 4 non-smooth points are permissible [13].)

Also, if the body is so smooth that a radius-r ball inside it can touch every point of its boundary, and the body has diameter D, then one may see that the scaling factor $1 - 2rd^{-2}D^{-1}/(1+\sqrt{1-d^{-2}})$ suffices. (This is tight for a d-ball.)

3.1.9 L_p-balls, $1 < p < \infty$

[Note: this section still under construction]

In fact, for the purpose of getting an explicit bound on the scaling factor (in the above proof that $d+1$ homothets suffice for smooth convex bodies) it is not necessary that we have a nonzero lower bound on r, the radius of curvature. Bounds on the "modulus of smoothness" sometimes suffice. These can exist even for cases such as $y \ge |x|^{1.1}$ near $x = 0$, in which no positive lower bound on r exists.

The "modulus of convexity" $\delta(\epsilon) > 0$ of a convex d-body $|\vec{x}|_B \le 1$ for some norm B is

$$\delta_B(\epsilon) = \sup_{|\vec{x}|_B, |\vec{y}|_B \le 1, \; |\vec{x}-\vec{y}|_B \ge \epsilon} 1 - \left|\frac{\vec{x}+\vec{y}}{2}\right|_B. \quad (11)$$

[7] We've also seen attributions of this theorem to Lyusternik and Shnirel'man. Although we daresay there are reasons for that, it seems that L & S's paper was in 1947, whereas Borsuk's was in 1933, and L & S's main concern was proving that any Riemannian metric, homeomorphic to the surface of a 3-ball, must have 3 non-self-intersecting closed geodesics. See [110].

[8] Actually, many points on the smooth body can correspond to 1 point on the sphere (consider a cylinder capped by two hemispheres), but this does not affect the validity of our argument.

The "modulus of smoothness" $\rho(\tau) > 0$ is

$$\rho_B(\tau) = \sup_{|\vec{x}|_B \le 1,\ |\vec{y}|_B \le \tau} \frac{|\vec{x}+\vec{y}|_B + |\vec{x}-\vec{y}|_B - 2}{2}. \quad (12)$$

As we have given them [50], these definitions only work for centrally symmetric convex bodies (i.e., those definable via norms). That will suffice for our present purposes[9]: L_p balls

$$\sum_{i=1}^{d} |x_i|^p \le 1, \quad 1 < p < \infty. \quad (13)$$

Known facts about these functions include the following: If we let B^* denote the polar body of B, then [122]

$$\rho_{B^*}(\tau) = \sup_{\epsilon > 0}[\frac{\tau\epsilon}{2} - \delta_B(\epsilon)], \quad \delta_{B^*}(\epsilon) = \sup_{0 \le \tau \le 2}[\frac{\tau\epsilon}{2} - \rho_B(\tau)]. \quad (14)$$

A body is smooth if and only if $\lim_{\tau \to 0+} \rho(\tau)/\tau = 0$. A body is strictly convex if and only if $\delta(\epsilon) > 0$ for all $\epsilon > 0$, which happens if and only if its polar body is smooth.

Exact expressions for $\delta(\epsilon)$ for the L_p balls (EQ 13) have been worked out by Hanner [93]. From these, (EQ 14), and the fact that the polar body to an L_p ball (EQ 13) is an L_q ball with $1/p + 1/q = 1$, one may derive expressions for $\rho(\tau)$. For our purposes it will suffice to know the asymptotic behaviors [123] when $\epsilon, \tau \to 0+$, which are

$$\delta(\epsilon) = \begin{pmatrix} (p-1)\epsilon^2/8 + o(\epsilon^2) \text{ if } 1 < p < 2 \\ (\epsilon/2)^p/p + o(\epsilon^p) \text{ if } 2 \le p \end{pmatrix} \quad (15)$$

$$\rho(\tau) = \begin{pmatrix} \tau^p/p + o(\tau^p) \text{ if } 1 < p < 2 \\ (p-1)\tau^2/2 + o(\tau^2) \text{ if } 2 \le p. \end{pmatrix} \quad (16)$$

3.1.10 Symmetric bodies

Schramm [159] showed that for any *strictly* convex d-body invariant under a group of reflections acting *irreducibly*[10], $d+1$ s-homothets suffice.

Rogers [155] had realized many years before, but only published in 1997, the fact that for centrally symmetric convex bodies,

$$(1 + s^{-1})^d (\ln d + \ln\ln d + 5)d \quad (17)$$

s-homothets suffice. This bound with $s = 1$ is valid for s-homothets with s infinitesimally below 1.

3.1.11 Bounds by Rogers and Zong

Rogers and Zong [155] also found that for general convex bodies, $\binom{2d}{d}(\ln d + \ln\ln d + 5)d$ s-homothets suffice. (This is better than Lassak's bound EQ 10 when $d \ge 6$.) Also, a convex body K may be covered by $2^d(\ln d + \ln\ln d + 5)d$ translated copies of its negation $-K$. Both these and (EQ 17) were consequences of the following theorem

[9] The whole theory has been recast in a more general setting by Klee [108].

[10] That is, in such a way that only the body's centroid, or the entire body – but no other subset of the body – is preserved by every group action.

Theorem 24 (Rogers and Zong's homothet covering theorem) *To cover a convex d-body K with translated copies of a convex d-body H, it suffices to use*

$$\frac{\text{vol}(K - H)}{\text{vol}(H)} \zeta_d(H) \quad (18)$$

copies, where $\zeta_d(H)$ is the minimal covering density of d-space by translates of H, and where "$K - H$" denotes the set $\{\vec{x} - \vec{y} \mid \vec{x} \in K,\ \vec{y} \in H\}$.

The consequences mentioned then follow from Rogers's earlier [153] bound $\zeta_d(H) \le (\ln d + \ln\ln d + 5)d$, and the inequalities

$$2^d \le \frac{\text{vol}(K - K)}{\text{vol}(K)} \le \binom{2d}{d} \quad (19)$$

which are respectively a special case of the "Brunn-Minkowski theorem" (EQ 20) and the upper bound from [154] [34]. These are tight respectively for a centrosymmetric d-body and a d-simplex.

The Brunn-Minkowski inequality [22] states that if A and B are convex d-bodies,

$$\text{vol}(A)^{1/d} + \text{vol}(B)^{1/d} \le \text{vol}(A + B)^{1/d}. \quad (20)$$

A reverse form of this inequality due to V.D.Milman [133] [144] [145] states that there exists an absolute constant C such that for each A, B there exists a determinant-1 affine transformation u such that

$$\text{vol}(A + uB)^{1/d} \le [\text{vol}(A)^{1/d} + \text{vol}(uB)^{1/d}]C_d, \quad C_d \le C. \quad (21)$$

The minimal values of C_d seem to be totally unknown, aside from the fact that $1 \le C_d \le C < \infty$. However,

Theorem 25 (Bound on C_d) *For all $d \ge 1$, $C_d < \sqrt{2\pi d/e}$.*

Proof. Let E_X denote the Löwner-John min-volume ellipsoid [88] enclosing X.

$$\begin{aligned} \text{vol}(A + uB)^{1/d} &\le \text{vol}(E_A + uE_B)^{1/d} = \\ &= \text{vol}(E_A)^{1/d} + \text{vol}(E_B)^{1/d} \\ &\le C_d[\text{vol}(A)^{1/d} + \text{vol}(uB)^{1/d}] \end{aligned} \quad (22)$$

if

$$C_d = (\bullet_d / \triangle_d)^{1/d}. \quad (23)$$

The final inequality in (EQ 22) is a consequence of [176], from which it follows that for any convex d-body K of unit volume, its Löwner-John min-volume enclosing ellipsoid E_K has d-volume at most C_d^d times larger (and this is tight if K is a simplex). The equality in (EQ 22) is because by picking the affine u appropriately, we can make E_A and E_B both be balls. In view of known formulas (§1.1.2) for $\bullet_d$ and $\triangle_d$ we have $C \le C_d$ where

$$C_d = (\frac{d!}{(d/2)!\sqrt{1+d}})^{1/d}(\frac{\pi d}{d+1})^{1/2} \le \sqrt{2d\pi/e}. \quad (24)$$

□

Consequently (see also [113])

Theorem 26 (Ultra general homothet covering theorem) *There exists an absolute constant C so that: For any convex d-bodies A and B, there exists a determinant-1 affine transformation u such that A may be covered by*

$$(\ln d + \ln\ln d + 5) d C_d^d \frac{[\mathrm{vol}(A)^{1/d} + \mathrm{vol}(B)^{1/d}]^d}{\mathrm{vol}(B)} \quad (25)$$

translates of uB, where either $C_d \geq 1$ is defined by (EQ 24), or $C_d = C$ (either is valid).

3.1.12 Bodies of constant width

Schramm [159] showed that for convex bodies of constant width, $5(3/2)^{d/2} d^{3/2}(4 + \ln d)$ s-homothets suffice. This bound is below 2^d when $d \geq 16$. Weissbach [180] showed that Schramm's bound may be strengthened to 6 when $d = 3$; Chakerian & Sallee [35] strengthened it to 3 when $d = 2$ (indeed showing that any scaling factor $s \geq .9101$ suffices).

3.1.13 Zonotopes & Zonoids

"Zonotopes" are Minkowski sums of a finite set of line segments. (Equivalently, they are the d-polytopes each of whose faces is a centrally symmetric $(d-1)$-polytope.) "Zonoids" are limits of zonotopes. Martini [128] showed that for non-parallelepiped d-zonotopes, $3 \cdot 2^{d-2}$ copies suffice, and this was extended to all zonoids in [20].

3.1.14 Any symmetric convex body, $d = 3$

For centrally symmetric convex 3-bodies, Lassak [115] showed that 8 s-homothets suffice. Another proof of the same thing was [165]. Bezdek [12] extended this to work for 3D bodies invariant under any affine symmetry.

3.2 Rotations allowed?

Definition 27 *A "rotothet" of a set is a scaled and possibly rotated and translated, copy of it. For an "s-rotothet," the scaling factor is s, and if s is left undefined or unspecified, we will take it to mean "for some s, $0 < s < 1$."*

How many s-rotothets of a convex d-body are required to cover it, if the copies may each be independently rotated and translated?

The answer can be smaller than in the case where rotations are not allowed (although not for a regular simplex or a ball, in which cases $d + 1$ are always required). For example[11], a 45-45-90 right triangle is tiled by only 2 copies of itself, each scaled by $1/\sqrt{2}$.

In d-space for all sufficiently large d, at least $1.203^{\sqrt{d}}$ s-rotothets are needed to cover the convex d-bodies constructed by Kahn and Kalai [102] [141].

When $d = 2$, we have

Theorem 28 *Any 2D convex body may be covered by 3 smaller rotothets – and 3 are necessary for the disk and for the equilateral triangle.*

[11] Conjecturally, this example is unique – see §3.2.1.

Proof. Let the body be called Q. We may assume Q is not a body of constant width, since that case was already proved by [35] (§3.1.12). In the below we'll implicitly use the fact that a convex body's surface is continuous and (in 2D) "differentiable to the left" (and right). It is also differentiable almost everywhere[12].

So assume Q has width equal to its diameter (which is the maximum possible width) only over some fraction γ, $0 \leq \gamma < 1$, of the possible rotation angles.

Rotate the object randomly within a set, of measure μ, $0 < \mu < 1 - \gamma$, in which we get East-West widths bounded above by some bound below the diameter of Q. Translate 2 copies, shrunk a sufficiently tiny amount, a sufficiently tiny amount North and South. Now only 2 small bits will be left uncovered on the East and West (unless, e.g. the westmost point is non-unique, which won't happen with probability 1 because of the random rotation). Now the maximum distance between any pair of points in these 2 small bits is less than the diameter of the object. Hence these two small regions may both be covered by a third rotothet. □

3.2.1 Simplices divisible into two congruent scaled versions of themselves

Conjecture 29 *The only d-simplex divisible into two congruent scaled versions of itself, in any dimension $d \geq 2$, is the 45-45-90 right triangle in 2D.*

Fact 30 *Conjecture 29 is true in 2D.*

Proof. In 2D there are only 2 possible ways the triangle could split (see figure 2, top 2 illustrations). In the lefthand illustration, 45-45-90 is clearly forced. In the righthand illustration, the strict triangle inequality is violated, forcing the triangle to be degenerate and hence of no interest as a 2D example of anything. □.

For further discussion relevant to conjecture 29, see [166].

Figure 2: Some simplices in 2D and 3D.

Fact 31 *Conjecture 29 is true in 3D, if attention is restricted to tetrahedra which are sliced by a plane of mirror symmetry.*

Proof. The bottom illustration in figure 2 depicts the top half (after the slicing) of the tetrahedron, where the $\triangle abs$ defines the slicing plane, and the dihedral angle on side s (for "slice") is being bisected, whereas the a and b edges used to be in the middle of a face and were created by the cut.

Consider the dihedral angles at the 6 edges. The cut has introduced 3 new angles at a,b, and s_{new}. But it has also gotten rid of 3 old angles at s_{old}, c, and d. Hence these two 3-element sets must be identical.

[12] In fact, convex d-bodies are 2-time differentiable almost everywhere, as was proved by A.D.Aleksandrov in 1936. Some results of this kind are surveyed in [89].

We now restrict attention to the mirror case where the slice plane is a plane of mirror symmetry. In that case a and b have 90° dihedrals, and

I. $s_{\text{old}} \neq 90°$, so c and d are both 90°. But then we gain s_{new} during the cut and lose s_{old}, with no way to regain s_{old} – impossible.

II. $s_{\text{old}} = 90°$, $s_{\text{new}} = 45°$.

   A. If d and c are both 90° then all of the original dihedrals were 90° except possibly at edge h. We claim such a tetrahedron cannot exist since two of its faces must lie in parallel planes.

   B. So only one of $\{d, c\}$ is 90°. So $a = b = c = s_{\text{old}} = 90°$, $s_{\text{new}} = d = 45°$ and h is unspecified. But this is degenerate: at the dsb vertex, since $45 + 45 + 90 = 180$, the spherical triangle at this vertex has 0 solid angle.

□

Remark. Any tetrahedron that could be split into two congruent half-volume scaled copies of itself, would automatically tile 3-space. We have checked[13] the known families [85] of space-tiling tetrahedra, and none of them are splittable.

Finally, we sketch an argument, based on counting degrees of freedom, which makes it plausible that no such simplex exists in dimensions $d > 3$.

The slice plane must have exactly 1 vertex above it and 1 vertex below it (so that the 2 components that result, are simplices) hence $d - 1$ vertices lie on it, and the remaining (extra) dth vertex must be due to the cut plane slicing an edge (1-flat).

WLOG let some vertex of the original simplex, which also is on the cut plane, be at the origin.

Consider the $d \times d$ matrix A whose rows are the altitude vectors of the d-simplex to all vertices except 0. When we cut the simplex we change 2 rows of this matrix, getting a new matrix $\tilde{A}$.

We want that $\tilde{A}$ is merely a rotation (or anyhow, $d \times d$ orthogonal transformation) of A/s, where $s = 2^{1/d}$. That is: $Q = s\tilde{A}A^{-1}$ is orthogonal, that is $QQ^T = I$. This is equivalent to

$$s^2(\tilde{A}^T\tilde{A}) = (A^T A). \tag{26}$$

where $\tilde{A}$ is the same as A except that two rows have been changed. Note: this must also be true of $\hat{A}$, the matrix corresponding to the simplex on the other side of the cut, and the same two rows have to change (although the changes may be different) with in fact one of these 2 rows for $\tilde{A}$ being the same, up to a scaling factor, as the other of these 2 rows for $\hat{A}$.

The condition that A actually corresponds to a genuine d-simplex, is that if you define B to be A, except that its rows have been scaled so that their lengths are the reciprocals of what they were, then

$$B \text{ must be invertible.} \tag{27}$$

[13] This should be rechecked using a computer.

(In fact B^{-1} will give coordinates of the non-$\vec{0}$ vertices, up to some overall scaling factor.)

Finally, $\tilde{A}$ is not just any arbitrary change to 2 rows of A. In fact, it must correspond to a cut plane through $\vec{0}$, which has $d-1$ degrees of freedom, whereas, any arbitrary change to 2 rows of A (satisfying the determinant of (EQ 26)) would have had $2d - 1$ degrees of freedom. So there must be some d additional constraints.

However, forget them. Just (EQ 26) alone, is going to be enough to make it plausible that no such d-simplices exist if $d > 5$.

(EQ 26) constitutes $d(d + 1)/2$ equations (since the matrices are symmetric) for $\tilde{A}$ plus $d(d+1)/2$ for $\hat{A}$ and we have got (for the 2-row changes, plus assuming A is upper triangular and with unit determinant WLOG) $3d + d(d + 1)/2$ degrees of freedom. If $d > 5$, then the number of constraining equations exceeds the number of degrees of freedom, hence it is plausible that no solution can exist if $d > 5$. (And even if we drop the 50-50 volume split demand.)

If one puts in the ignored additional constraints, it becomes plausible there is no solution if $d > 3$.

So probably only the case $d = 3$ is of interest, and solutions there are probably going to have to be ruled out by inequalities (as in the proof above for the mirror symmetric case) not just degree of freedom arguments (assuming these could ever be made rigorous).

3.3 Affinities allowed? Answer: 2?

Definition 32 *An "affinity" of a set is an affine (i.e. general linear) transformation of it. For an "s-affinity," the volume scaling factor (|determinant|) is s^d, and if s is left undefined or unspecified, we will take it to mean "for some s, $0 < s < 1$."*

How many s-affinities (where the affine transformations are chosen independently for each copy) are needed to cover a convex d-body?

We may tile d-boxes and d-simplices with only 2 affine versions of themselves, each with $s^d = 1/2$, i.e. half volume. A unit volume d-ball may be covered by two d-ellipsoids, each of volume

$$s^d = (\frac{d}{d+1})^{(d+1)/2}(\frac{d}{d-1})^{(d-1)/2} < \exp(\frac{-1}{2d}), \tag{28}$$

and this is tight[14].

These facts suggest

[14] This fact underlies the "ellipsoid algorithm" for convex programming, cf. lemma 3.1.34 of [88]. If the principal axes of the ellipsoid are $L_1 \leq L_2 \leq \dots \leq L_d$, then we may slice it with a hyperplane through its center and perpendicular to the L_d axis, and then cover each of the two resulting hemiellipsoids with ellipsoids with axes $dL_i/\sqrt{d^2-1}$ for $1 \leq i \leq d-1$, and $dL_d/(d+1)$. Let the DW aspect ratio of the original ellipsoid be $A = L_d/L_1$. Notice that if $A \geq \sqrt{(d+1)/(d-1)}$, then the two smaller ellipsoids will have DW aspect ratios $\leq A$, while if $A \leq \sqrt{(d+1)/(d-1)}$, then the two smaller ellipsoids will have DW aspect ratios $\leq \sqrt{(d+1)/(d-1)}$ also.

Conjecture 33 (Two affines?) *Two s-affinities of a convex d-body always suffice to cover it*[15].

In theorem 35 we prove conjecture 33 for smooth convex bodies, and also for general 2D convex bodies.

We will need a preparatory lemma.

Lemma 34 *The min-volume affine version of a convex body B, containing a convex body A, touches A's boundary at at least $d+1$ points. There exist support hyperplanes of A at these points forming a finite d-simplex (i.e. as opposed to something infinite) surrounding A. There exist support hyperplanes of B at these points forming a finite d-simplex surrounding B.*

We only assert this lemma under the assumption that either: B is smooth, or we are in 2D.

Proof. I. If they only touched at $\le d$ points, then we could shrink the volume of B by contracting in a direction orthogonal to the subspace defined by the $\le d$ points.

II. If all outward unit normals to the support hyperplanes of B at points of $\partial A \cap \partial B$ *necessarily* lay on the same closed hemisphere of a unit sphere (the lemma asserts this is impossible) then

1. Expand B by factor $1+\epsilon$ in all directions, increasing its d-volume by $(1+\epsilon)^d = 1 + d\epsilon + O(\epsilon^2)$.

2. Dilate A, in the 1 "vertical" direction defined by the hemisphere's center only, by $1+\delta$ where δ is much larger than ϵ. For example, if B's boundary were 2-time differentiable, with bounded second derivative at points at which the normals to the support hyperplanes lie on the hemisphere's "equator," then at least order $\sqrt{\epsilon}$ would be possible. Unfortunately, our definition 18 of smoothness does not imply 2-time differentiability everywhere – despite footnote 12. However ([123] and cf. §3.1.9) it turns out that a convex body is smooth[16] if and only if its "modulus of smoothness" $\rho(\tau)$ obeys $\lim_{\tau\to 0+} \rho(\tau)/\tau = 0$, in other words, if and only if we may find a suitable $\delta(\epsilon)$ with $\delta/\epsilon \to \infty$ as $\epsilon \to 0+$, precisely what we need. In 2D, where we have not assumed smoothness, we still know that convex functions are differentiable to the left and right everywhere. This alone, however, does not suffice to show that a suitable δ exists[17]. But, all we need, even in the footnote's scenario with dense corners, is that there be some neighborhood of the two antipodal equatorial points[18], with radius δ, over which the horizontal displacement h away from the vertical, obeys $2h < \delta$ for all sufficiently small δ. An averaging argument shows that this must exist. Hence this step is permissible under either of the assumptions of the lemma.

And this (after performing the right affine to get A back to normal shape, and assuming we chose ϵ sufficiently small) also yields a contradiction with minimality.

III. If the outward unit normals to the support hyperplanes of A all necessarily lay on the same closed hemisphere of a unit sphere then so would B's. □

Theorem 35 *Any convex d-body B may be covered by 2 affine versions of itself, each with smaller d-volume.*

We only assert this under the assumption that either: B is smooth, or we are in 2D.

Proof. Choose a direction $\vec{x}$ such that the "equator" E on B's surface (points on the surface with support hyperplanes tangent to $\vec{x}$) consists entirely of points "smooth in the direction $\vec{x}$," i.e. for each equatorial point, it is impossible to rotate its support hyperplane except in ways which preserve the fact that the hyperplane contains the direction $\vec{x}$.

Under the assumptions of the theorem, this obviously is always possible[19].

Cut the body into 2 pieces B_1 and B_2, the convex hulls of the stuff above/on and below/on the equator. These two pieces cover all of B; they can overlap.

Now, apply the lemma to see that the obvious affine version of B (namely, B itself) surrounding B_1 is not minimal volume. Ditto for B_2. Alter each to get minimal volume, and we are done. □

Remark. Theorem 35 also holds if the body is permitted to have 1 non-smooth point.

4 Separator theorems about geometrical objects

4.1 A separator theorem for d-boxes

We originally proved the following theorem in a form which was roughly a factor of 2 weaker, and which depended on a randomizing argument. But then we saw [6] had proved the theorem below at full strength, and deterministically, except only in the special case $\kappa = 1$ of disjoint boxes. It was then an easy matter to get the best of both worlds by generalizing the proof of [6] to work for arbitrary κ.

Theorem 36 (Separator hyperplane for κ-thick iso-oriented d-boxes.) *Given N iso-oriented d-boxes whose interiors are κ-thick: There exists a hyperplane, orthogonal to a coordinate axis, such that at least*

$$\lfloor (N+1-\kappa)/(2d) \rfloor \tag{29}$$

of the d-box interiors lie to each side of the hyperplane.

[15] One might further conjecture that the scaling factor on the left hand side of (EQ 28) is the worst possible, i.e. the ball is the worst convex object.

[16] Actually, [123] only claimed this for centrally symmetric convex bodies, which were all they were interested in, but central symmetry would seem to have nothing to do with anything.

[17] Consider a "nasty" convex curve with a corner at every rational arclength p/q, having bending angle proportional to q^{-4}.

[18] If there are more than two, our job is made easier.

[19] But we have no characterization of the convex bodies for which this is possible.

Proof. When $d = 1$, the result is easy. So suppose $d \geq 2$.

Find the leftmost hyperplane with at least $\lfloor (N+1-\kappa)/(2d) \rfloor$ boxes entirely to its left. (Or any hyperplane between this and its right-facing alter ego.) If there are $\lfloor (N+1-\kappa)/(2d) \rfloor$ boxes to its right, we are done. Otherwise, we have a $(d-1)$-dimensional problem inside the hyperplane, and by induction on d, by solving this problem we will get a hyperplane with at least $\lfloor (M+1-\kappa)/(2d-2) \rfloor$ boxes to each side of it, where $M+1 > N+1-2\lfloor (N+1-\kappa)/(2d) \rfloor$, i.e. $M+1 > (1-1/d)(N+1)+\kappa/d$. Hence, the number of boxes on each side will be $\geq \lfloor (1-1/d)(N+1-\kappa)/(2d-2) \rfloor = \lfloor (N+1-\kappa)/(2d) \rfloor$. □

Remark. Examples in [6] show that (EQ 29) is best possible for an infinite number of N for each $d \geq 1$, when $\kappa = 1$. Also, see the examples in figure 3.

Figure 3: Examples of near tightness in theorem ??. (a) $N = 12$, $d = 2$, $\kappa = 1$. If the dotted lines are ignored we have $N = 4$, $d = 2$, $\kappa = 1$. If we then project down onto a line, we have $N = 4$, $d = 1$, $\kappa = 2$. The picture generalizes to give an example in d dimensions with $\kappa = 1$, $N = 2md$ in which no iso-oriented hyperplane cuts off more than m boxes. (b) $N = 6$, $d = 2$, $\kappa = 2$. The picture generalizes to give an example in d dimensions with $\kappa = 2$, $N = 2 + 2d$ in which no iso-oriented hyperplane cuts off more than 1 box.

A related result, which is best possible for all N and d, is

Fact 37 *Given any N disjoint iso-oriented d-boxes whose interiors are $(N-1)$-thick, there always exists a hyperplane perpendicular to some coordinate axis that separates 2 of them.*

Proof. Let P_i denote the hyperplane perpendicular to the x_i coordinate axis, among the hyperplanes corresponding to the minimal-x_i faces of our boxes, with maximal x_i.

Suppose P_1 does not work, i.e. no box interior lies entirely on the lesser-x_1 side of it. In that case, P_1 could be pushed infinitesimally in the x_1 direction (getting P_1') and then all N boxes would have to be split by P_1. Then suppose P_2 also does not work. In that case all the box interiors would intersect P_2', similarly. This would force all the boxes in fact to intersect $P_1' \cap P_2'$, by consideration of the $(d-1)$-dimensional version of the problem projected into P_1'. (Note, the $(N-1)$-thickness property will be preserved under this particular projection in our particular case.) Continuing, we conclude that if no P_i works, the point $\cap_{i=1}^{d} P_i'$ would have to be inside all N of the boxes, contradicting our assumption of $(N-1)$-thickness. □

Lemma 38 (Algorithmic version) *The hyperplane of fact 37 or theorem 36 may be found in $O(Nd)$ steps. Also, if the $2d$ lists of the lower and upper endpoints of the intervals defining the boxes's ith coordinates are pre-sorted, then the* best *such hyperplane (according to a wide variety of optimality measures) may be found in $O(Nd)$ steps.*

Proof. In theorem 36, the algorithm given in the previous proof [6] works.

In fact 37, if the coordinate direction is x_i, then this hyperplane WLOG is the hyperplane of the minimal-x_i face of a box, in fact the hyperplane (among these) with maximal x_i. Hence if i were known, a suitable x_i coordinate of the hyperplane could be found in $O(N)$ steps. Since i is not known, we try all d possibilities, which takes $O(Nd)$ steps. Now: which of the d candidate hyperplanes actually works? Well, one can check if a candidate hyperplane works in $O(N)$ steps (see if some box lies below and some box above it), so the total runtime is $O(dN)$.

To find the best hyperplane, one may use the presorted lists to compute the crossing counts and the above/below counts for every one of the $O(Nd)$ combinatorially distinct kinds of hyperplanes. This works with any notion of "best" depending only on these three counts. □

Remark. One may also show: Given a κ-thick set of N convex d-objects, such that among any 2 disjoint such objects, there exists a separating hyperplane selected from one of C possible orientations. Then there exists a C-oriented hyperplane such that at least $(N+1-\kappa)/O(C)$ of the objects lie entirely to each side of it.

4.2 A separator theorem for d-cubes

Theorem 39 (Cube separator theorem) *Let there be a set $\mathcal{S}$ of N iso-oriented d-cubes in a euclidean d-space, whose interiors are κ-thick, or more generally κ-overloaded (definition 10) or (λ,κ)-thick (definition 9). Then there exists an iso-oriented d-box (with maximal sidelength at most 2 times the minimal sidelength) with at most $2N/3$ cube interiors entirely inside it, at most $2N/3$ cube interiors entirely outside (where the restrictions that "cardinality $\leq 2N/3$" may be generalized to be "weight $\leq 2/3$" where "weight" is defined via any of a wide class of mass-1 measures on d-space, which may or may not bear any relation to $\mathcal{S}$), and moreover satisfying the following condition.*
(a) If the interiors are κ-thick, the number of cube interiors partly inside and partly outside the box is

$$\leq c_d(\epsilon)\,\kappa^{1/d}\,N^{1-1/d} + \left(\frac{2}{\epsilon}+2\right)^d \kappa \tag{30}$$

where

$$c_d(\epsilon) = \left(\frac{[1+2^d(H_{2d}-H_d-\frac{1}{2}+O(\epsilon d))]\,(2d)!}{d!}\right)^{1/d} \tag{31}$$

provided ϵd is sufficiently small, and $H_m = \sum_{j=1}^{m} j^{-1}$.

(b) If the interiors are κ-overloaded, then for $K > 0$, the number of cubes which have side less than K times the maximal sidelength of the box and whose interiors are partly inside and partly outside the box is $O\left(Kd\kappa^{1/d}N^{1-1/d}\right)$.

(c) If the interiors are (λ, κ)-thick, then given any $K > 0$, the number of cubes which have side less than K times the maximal sidelength of the box and whose interiors are partly inside and partly outside the box is $O\left(\frac{\lambda}{1-\lambda^{-(d-1)}}(d^d K^d \kappa + 1)N^{1-1/d}\right)$.

Remark. Take κ fixed. When $N \to \infty$ with d fixed, best results in (a) are obtained if ϵ is of order $N^{-\frac{d-1}{(d+1)d}}$, in which case (EQ 30) is $\le c_d(0)\kappa^{1/d}N^{1-1/d}(1+o(1))$. Note that $H_{2d} - H_d < \ln 2$, with equality in the limit as $d \to \infty$.

Remark. When $d = 2$ and $d = 3$ the main term in the right hand side of (EQ 30) (ignoring ϵ's) becomes respectively $4\sqrt{\kappa N}$ and $(232\kappa)^{1/3}N^{2/3}$. When $d \to \infty$, it is asymptotic to

$$\frac{8}{e}d\kappa^{1/d}N^{1-1/d} \tag{32}$$

where $e \approx 2.71828$.

Proof. Unfortunately, the full proof has become rather long due to a considerable number of ϵs and δs and their ilk. But a quick summary of the proof idea may be found in remark iii below.

Part (a) is treated first, so assume the cube interiors are κ-thick. Let B be a brick (iso-oriented d-box) with side length ratios $d+1$:$d+2$:...:$2d$ (not necessarily in that order) having maximal d-volume subject to the constraint that every brick congruent to it has at least $N/3$ cube interiors (or at least $1/3$ of some arbitrary[20] measure on d-space) entirely outside.

WLOG (by scaling) B_0's dimensions are in fact exactly $(d+1) \times (d+2) \times \ldots \times (2d)$. Picking in advance some δ, $0 < \delta < 1/d$, we now define [21] B_0 to be a slightly exapnded copy of B (namely, the dimensions of B are $1 - \delta/3$ times those of B_0) with $< N/3$ cube interiors ($< 1/3$ weight) lying entirely outside it. If δ is sufficiently small, such a B_0 must exist.

Cut B_0 in half by bisecting its largest dimension to get two subbricks B_i and B_{ii} with side lengths $d \times (d+1) \times \cdots \times (2d-1)$. By the *pigeonhole principle* at least one of these contains or intersects at least $N/3$ cubes of $\mathcal{S}$ ($\ge 1/3$ weight).

WLOG that subbrick is B_i. Also WLOG (by a translation) B_i's "lower left" corner (i.e. the one with minimal coordinates in every direction) lies at the origin.

[20] The measure defining the notion of "weight" has to be such that this maximum can exist, or at least the corresponding supremum, in which case B's size is chosen arbitrarily close to the supremum.

[21] We will eventually take the limit as $\delta \to 0$ and hence the proof is really unaffected by the presence of δ. In fact in an earlier version of this proof we were merely instructing the reader to "infinitesimally expand" B, but later decided that we had to put in δ in order to be convincing that its absence did not matter!

Note that if every sidelength of B_i were expanded by adding $1 - \delta$, the result would be a brick B' contained in a brick congruent to B. So consider in fact the family $\mathcal{F}$ of translates of B' whose lower left corner coordinates are $(-t, -t, -t, \ldots, -t)$ for $0 < t < 1 - \delta d$.

Every member of $\mathcal{F}$ contains B_i and hence contains or intersects at least $N/3$ cubes of $\mathcal{S}$. Also, every member of $\mathcal{F}$, like B_0, is contained in a brick congruent to B and hence (by the definition of B) must have at least $N/3$ cube interiors entirely outside it. So in order to complete the proof, we need only to show that some member (in fact, a random member) of $\mathcal{F}$ cuts sufficiently few cubes of $\mathcal{S}$.

See figure 4 for a picture in 2 dimensions, where for convenience we take the limiting picture as $\delta \to 0$.

Figure 4: Sliding boxes in $\mathcal{F}$.

Let $A_0 = (2d)!/d!$ be the total d-volume (i.e. area, in the case $d = 2$) of B_0. Let the d-volume swept out by the surfaces of the boxes in $\mathcal{F}$ as t goes from 0 to $1-\delta d$ be $A_1 + A_2$, where A_k is the portion swept through k times by the box surface. (Notice that $k = 1$ and $k = 2$ are possible, but $k \ge 3$ is not, due to convexity.) For instance, in figure 4, A_2 is represented by the two triangular areas.

Write "$a \approx b$" to mean "$a = (1 + O(\delta))b$ for fixed d when δ is sufficiently small." By considering the volume of the annulus swept out by the brick's surface,

$$A_1 + 2A_2 \approx 1 \cdot S_0 \tag{33}$$

where $S_0 = 2A_0 \sum_{j=d+1}^{2d} j^{-1} = 2(H_{2d} - H_d)A_0$ is the surface $(d-1)$-area of B_0. Meanwhile

$$A_1 + A_2 \approx \frac{S_0 + A_0}{2} \tag{34}$$

is the volume swept out by the "top" (i.e. with higher coordinate values) faces of the brick, plus the relevant volume inside the bottommost brick in $\mathcal{F}$ (which is $\approx A_0/2$ since the "hole" in the middle is half the volume). Solving, we find $A_1 \approx A_0$ and $A_2 \approx (S_0 - A_0)/2$.

Armed with the above d-volume formulae, we're ready to proceed. First, choose $\epsilon > 0$, and obtain $\overline{\mathcal{S}}$ from $\mathcal{S}$ by throwing away all cubes of sidelength greater than ϵd which intersect the annulus described above. Put $\overline{N} = |\overline{\mathcal{S}}|$. Note that by the κ-thick property, the number of cubes thrown away is $N - \overline{N} \le (2/\epsilon + 2)^d\kappa$ (i.e., a constant independent of N). Give the space inside A_k monetary value k^d dollars per unit d-volume ($k = 1$ and 2). Let t_i be the probability that a random member of $\mathcal{F}$ cuts the ith d-cube in $\overline{\mathcal{S}}$. If this cube has side length s_i then $t_i = ks_i$, and the space it occupies is worth $k^d s_i^d = t_i^d$ dollars, if it is wholly inside A_k. It is easily checked that cubes which are partly in A_1 and partly in A_2 also have value at least t_i^d dollars. Cubes in $\overline{\mathcal{S}}$ which are partly outside $A_1 \cup A_2$ will also have this property if we give value 2^d per unit volume to all the space outside

$A_1 \cup A_2$ which can be covered by cubes of side length ϵd which intersect $A_1 \cup A_2$. Note this space includes part of the "hole", but nevertheless its volume is $O(\epsilon d A_0)$.

If C is the expected number of cubes of $\overline{\mathcal{S}}$ cut by a random member of $\mathcal{F}$, then

$$C = \sum_{i=1}^{\overline{N}} t_i. \tag{35}$$

But any point in space can only be covered by at most κ cubes in $\overline{\mathcal{S}}$, which gives the total-dollars constraint

$$\sum_{i=1}^{\overline{N}} t_i^d \leq (A_1 + 2^d A_2 + O(2^d \epsilon d A_0))\kappa. \tag{36}$$

Maximizing C subject to this constraint, convexity implies that the maximum occurs when all t_i are equal, and we finally get

$$\begin{aligned} C &\leq \overline{N}^{1-1/d}(A_1 + 2^d A_2 + O(2^d \epsilon d A_0))^{1/d}\kappa^{1/d} \\ &\approx \overline{N}^{1-1/d} A_0^{1/d}\kappa^{1/d} \\ &\quad \times (1 + 2^d[H_{2d} - H_d - 1/2 + O(\epsilon d)])^{1/d}. \end{aligned}$$

This is equivalent to (EQ 30), when we take $\delta \to 0$. Since some member of $\mathcal{F}$ must cut the expected number (or fewer) cubes of $\overline{\mathcal{S}}$, we are done. (The asymptotic behavior in (EQ 32) comes from Stirling's formula $x!^{1/x} \sim x/e$.) This proves part (a) of the theorem.

For part (b), assume $\mathcal{S}$ is κ-overloaded. Follow the proof of part (a), but write $t_i = 2s_i$ for an upper bound on the probability that the ith d-cube is cut, and instead of giving different parts of space different values, note only that the total volume which can be covered by cubes with sidelength at most K times any of the sidelengths of the d-box is $O(Kd)^d$. By the κ-overloaded property, the total volume of all these cubes is at most $\kappa \cdot O(Kd)^d$. With this upper bound on $\sum s_i^d$, maximising $\sum 2s_i$ yields the required bound.

For part (c), assume $\mathcal{S}$ is (λ, κ)-thick. Again, follow the proof of part (a), but this time define $\overline{\mathcal{S}}$ by throwing away all cubes of sidelength $\geq 2Kd$. The expected number of intersections C, which must be maximised, is (EQ 35) as before, but the constraint (EQ 36) does not necessarily hold since there is only a bounded number of overlaps on cubes which are roughly the same size. Write the summation (EQ 35) as $S_1 + S_2$ where S_1 is the contribution from all $t_i < N^{-1/d}$ and S_2 is the rest (in which $t_i \leq 2Kd$. Immediately, $S_1 < N^{1-1/d}$. To bound S_2, note by the definition of (λ, κ)-thickness that for all x, $\sum_{x \leq t_i \leq \min\{\lambda x, 2Kd\}} t_i^d \leq A\kappa$ for some constant $A = O(Kd)^d$ related to the volume of the annulus, cf. (EQ 36).

For $j \geq 0$ let b_j denote the number of i for which $\lambda^j \leq t_i N^{1/d} \leq \min\{\lambda^{j+1}, 2Kd\}$. By the constraint above, $\lambda^{dj} b_j / N \leq A\kappa$ for each j, i.e.

$$b_j \leq \lambda^{-dj} A\kappa N. \tag{37}$$

Hence

$$S_2 < \sum_{j \geq 0} \lambda^{j+1} b_j N^{-1/d} \leq \frac{\lambda}{1 - \lambda^{-(d-1)}} A\kappa N^{1-1/d} \tag{38}$$

if $d \geq 2$. Hence

$$C = S_1 + S_2 \leq \frac{\lambda}{1 - \lambda^{-(d-1)}}(A\kappa + 1)N^{1-1/d}, \tag{39}$$

as required for (c).

□

Remarks:
(i) Theorem 39 and proof also hold inside a "cubical[22] d-torus" (iso-oriented d-box with opposite faces identified to give "periodic boundary conditions") by just substituting this for "d-space" everywhere. B_0 (or the members of $\mathcal{F}$) may "wrap" around and overlap itself, but this does not cause any difficulty if we agree to count the d-volume inside B_0 according to multiplicity, but count cube interiors (or weight) normally. Cubes crossing the boundary of a box in $\mathcal{F}$ are (to be conservative) counted according to their multiplicity – although we'll assume that none of the cubes in $\mathcal{S}$ is self-overlapping! The "inside" of a box is determined by the local picture near the box's corners as usual.

(ii) If, instead of iso-oriented d-cubes, $\mathcal{S}$ consists of any other assortment of objects with the property that their "CV aspect ratio" (definition 4) is bounded above by τ, then the theorem still holds, but with κ in (EQ 30) replaced by $\kappa\tau^d$.

(iii) It is also possible to demand that the separator shape be something other than an iso-oriented d-box. Let us recall the proof technique. We began by finding a maximal d-volume separator shape B_0 with at least a constant fraction p of the weight outside it. We then subdivided this shape into 2 smaller sets – more generally we may cover, rather[23] than subdivide, the shape by $m \geq 2$ sets. By the pigeonhole principle, at least one of these sets, call it Q, must contain $\geq (1-p)/m$ of the weight. (The best value of p to pick is $1/(m+1)$; we are going to prove a $(m/[m+1], O(N^{1-1/d}))$ separation result.) We then argued that a random separator shape congruent to B_0 and containing Q would, in expectation, cross no more than $cN^{1-1/d}\kappa^{1/d}$ objects from $\mathcal{S}$. Here the value of c will depend upon the separator shape, the shape and size of the covering objects, and the probability distribution you use.

4.2.1 Variations on the cube separator theme

In the previous proof, one may replace the trivial theorem that a d-box of bounded CV aspect ratio may be tiled by two d-boxes of bounded CV aspect ratio, with theorem 24 showing that any convex d-body Q may be covered by $2^{O(d)}$ half-scaled translated copies of itself. Or other theorems from §3, so long as they involve some scaling s and some covering cardinality m.

[22] Or rectangular – provided the torus has bounded ratio of maximum to minimum sidelength.
[23] And this is the point of §3.

Also one may replace the randomizing argument involving a "diagonally shifted" 1-parameter family $\mathcal{F}$ of boxes, by, e.g., instead, a 1-parameter family of scaled Q's with common incenter. A Q with scaling factor x is chosen with probability proportional to x.

One possible very general separator theorem one may get in this way, is theorem 40.

In the following theorem, "equivalent" copies (or "versions") of a convex body can be defined to be rotothets or to be homothets.

Theorem 40 (Highly general separator theorem) *Let there be a κ-thick set of N convex objects in d-space, each with bounded aspect ratio (see definition of $V_{\rm obj}$ below). Let Q be a convex shape that may be covered by m smaller copies of itself, each copy having s^d times smaller volume, and let $0 < \epsilon < s/2$. Then there exists a convex body T equivalent to Q such that $\leq mN/(m+1)$ of the objects are completely inside T, $\leq mN/(m+1)$ are completely outside, and at most*

$$(1+O(\epsilon))\frac{2A_{\rm sep}}{1-s^2}(\frac{\kappa V_{\rm sep} d\ln\frac{1}{s}}{V_{\rm obj}})^{1/d}N^{1-1/d}+O(1+1/\epsilon)^d\kappa$$

are partly inside and partly outside. Here $V_{\rm obj}$ is such that the volume of any object of diameter L in the given set is at least $V_{\rm obj}L^d$, and the volume of the separator T is at most $V_{\rm sep}r^d$ if its inradius is r. Also, $A_{\rm sep}$ is any upper bound on the ratio of the maximum to the minimum separation between the incenter of Q and a support plane supported by a smooth point on Q's boundary.

Note The theorem is also true for "equivalent copies" defined to mean any affinities, in the following sense. The hypothesis about Q is changed to: Let $\mathcal{Q}$ be a family of convex shapes such that each $Q \in \mathcal{Q}$ may be covered by m shapes in the family, each such shape having s^d times smaller volume than Q. Here s is a constant universal for $\mathcal{Q}$ and independent of Q. We require also that the bounds $A_{\rm sep}$ and $V_{\rm sep}$ are fixed bounds which apply to all shapes in the family $\mathcal{Q}$. In the conclusion, "T equivalent to Q" becomes "T equivalent to one of the shapes in Q".

Proof sketch. We begin by finding a maximal d-volume version of Q such that *any* version of Q with this same volume has at least $N/(m+1)$ objects wholy outside it. Then, as in the proof of theorem 39, find a version Q_0 with infinitesimally larger volume, having less than $N/(m+1)$ objects wholy outside it.

Cover Q_0 with m versions of Q, each of volume $\leq s^d\,\mathrm{vol}(Q_0)$.

One of these m coverers, by the pigeonhole principle, must contain bits of at least $N/(m+1)$ of the objects. Call it Q_1. WLOG (by a scaling) Q_1 has inradius s.

Now consider a scaled version T of Q_1, having the same incenter (center of largest inscribed ball) as Q_1, but scaled to have inradius t, $s<t<1-\delta$. (As in the proof of theorem 39, δ will go to 0 and thereby have no effect in the result obtained. Its role is purely to ensure that the family of scaled copies of Q_1 all have volume less than $(1-\delta)\,\mathrm{vol}(Q_0)$. So henceforth in this sketch we argue as if $\delta=0$.) Choose t at random with probability density $f(t)\,\mathrm{d}t$ where

$$f(t) \;=\; \frac{2t}{1-s^2}. \tag{40}$$

Clearly T will contain bits of at least $N/(m+1)$ objects inside (since $Q_1 \subset S$) and at least $N/(m+1)$ objects lie wholy outside it (by the definition of Q_0 — this is where we recall that δ is actually non-zero). We now want to bound the expected number of objects intersecting ∂T. For each i, let t_i denote the minimal value of t such that the scaling of Q_0 of inradius t intersects object i. Let the diameter of object i be L_i. Define $\overline{S}$ to contain all the objects which are very small; in this case, with $L_i<\epsilon$, and assume WLOG these are the objects are $1,2,\dots,\overline{N}$. Again define C to be the expected number of objects in $\overline{S}$ cut by T. Then, noting that $L_iA_{\rm sep}$ is an upper bound on the difference $|t_a-t_b|$ for which a t_a-scaled and t_b-scaled version of Q_1 have two points of distance at most L_i apart, we obtain

$$C \leq \sum_{i=1}^{\overline{N}} L_iA_{\rm sep}f(t_i+L_i), \tag{41}$$

or

$$C \leq \frac{2(1+\epsilon)A_{\rm sep}}{1-s^2}\sum_{i=1}^{N}\overline{N}L_it_i. \tag{42}$$

Give weight $t^{-d}\mathrm{d}V$ to a volume element $\mathrm{d}V$ at distance t from the incenter of Q_1. By adding up weighted volumes we have

$$\sum_{i=1}^{\overline{N}} V_{\rm obj}L_i^dt_i^{-d} \leq \kappa W \tag{43}$$

where

$$W=\int_{s-\epsilon}^{1+\epsilon}t^{-d}\frac{\mathrm{d}(V_{\rm sep}t^d)}{\mathrm{d}t}\,\mathrm{d}t = d\cdot V_{\rm sep}\ln\frac{1+\epsilon}{s-\epsilon}. \tag{44}$$

Thus

$$\sum_{i=1}^{N} L_i^dt_i^{-d} \leq \frac{\kappa V_{\rm sep}d}{V_{\rm obj}}\ln\frac{1+\epsilon}{s-\epsilon}. \tag{45}$$

Of course we need only concern ourselves with objects such that $t_i<1$, i.e. $t<1$. The largest possible value of (EQ 42) such that the (L_i/t_i) ratios satisfy (EQ 45) is

$$C \leq \frac{2(1+\epsilon)A_{\rm sep}}{1-s^2}\left(\frac{\kappa V_{\rm sep}d\ln\frac{1+\epsilon}{s-\epsilon}}{V_{\rm obj}}\right)^{1/d}\overline{N}^{1-1/d}. \tag{46}$$

The theorem follows on noting that the number of objects with $L_i\geq\epsilon$ is $O(1+1/\epsilon)^d\kappa$. □

4.3 Topological separation results

Definition 41 *A "Rado point" of a unit mass measure μ in d-space is a point such that any hyperplane through it will split the the measure so that $\leq d/(d+1)$ mass will lie in either open halfspace. A "tight Rado point" is a Rado point such that there exists a hyperplane through it for which this inequality is tight.*

R.Rado [148] showed that any measure in d-space has a Rado point, indeed a tight one[24], as a consequence of Helly's [97] theorem that if every $d+1$ among M convex sets intersect, then the intersection of all the sets is nonempty. (See also [66].)

This may also be derived from "Brouwer's fixed point theorem" that continuous self mappings of a d-ball have a fixed point [30] [134].

For *general* measures, it is impossible to get a "super-Rado" point in which "$d/(d+1)$" may be replaced by some smaller value [66]. However,

Theorem 42 ($1/e$ split for convex bodies) *Given a convex compact d-body B of unit volume. The smaller-volume side of any hyperplane cut through B's centroid has volume $\geq [d/(d+1)]^d > e^{-1} \approx .367879$.*

Proof. Given a cutting hyperplane H through B's center of mass, we may WLOG apply "Steiner reflection symmetrizations" to B about any "mirror hyperplane" M that also goes through B and is orthogonal to H. (Steiner reflection symmetrizations are discussed in §9, pages 76-77, of [22].) Thus WLOG B is a convex body of revolution and H is orthogonal to its axis and goes through its center of mass.

Now, given this, it is easy to see that the "worst" B (minimizing the volume fraction cut off) is uniquely a cone with H parallel to its base and lying at height $h/(d+1)$ above it, where h is the height (distance from base to apex) of the cone. The result follows. □

Remark. This theorem was probably known before; so call it a folk theorem, since we can't find a proof in the literature. A related interesting theorem may be found in [28].

Also, the present proof generalizes to show:

Given any probability density Ψ in d-space which is unimodal on lines (that is, has a unique maximum on any line; it decreases monotonically [non-strictly] as we move along the line away from its maximum) the smaller-mass side of any hyperplane cut through Ψ's center of mass has mass $\geq [d/(d+1)]^d > e^{-1} \approx .367879$.

Another famous fact is the "ham sandwich theorem," (see [168]) which states that given d measures in d-space, there exists a hyperplane whose removal leaves two pieces, each with $\leq 50\%$ of each of the measures. (Thus, in 3D, a single knife cut can bisect the bread, ham, and cheese!) This is a consequence of Borsuk's theorem[25] [26] that for any continuous map f from a sphere (the surface of a $(d+1)$-ball) to a flat d-space, there must exist two antipodes with a common image: $f(\vec{x}) = f(-\vec{x})$.

For another application of Borsuk's theorem, see fact 21.

As a (new) application of the Ham Sandwich theorem, we mention

Theorem 43 *Let there be k measures on the real line. Then there exist $\leq k$ points such that if the line is cut at those points, and the odd numbered pieces are collected together and called S, evens are T, then S and T each have $\leq 1/2$ of each of the k measures, simultaneously.*

Proof. Map the line into k-space according to

$$x \to (x, x^2, \ldots, x^{k-1}, x^k). \tag{47}$$

Now find a "ham sandwich hyperplane" in k-space. Such a hyperplane can only cross this curve k times at most (a polynomial of degree k has at most k roots). □

The entirety of [84] is thereby compressed to 2 paragraphs, demonstrating the power of ham sandwich. Another easy result of this nature is

Theorem 44 *Given 5 measures in the plane, there exists a conic curve which bisects them all simultaneously. That is, once the curve is removed, each piece of the plane that remains (there hereby are two, where the quadratic form is >0 and <0) has $\leq 1/2$ of the mass of each measure.*

Proof. Map the plane into 5-space via

$$(x, y) \to (x, y, x^2, y^2, xy) \tag{48}$$

then apply the ham sandwich theorem. □

The most general of this ilk is

Theorem 45 *Given $\binom{k+d}{d}-1$ measures in d-space, there exists an algebraic surface of degree k which bisects them all (that is, once the surface is removed, two sets – where the degree-k polynomial is >0 and <0 – remain, and each set contains $\leq 1/2$ of the mass of each measure) simultaneously.*

Rado's theorem and the ham sandwich theorem were recently unified by Zivaljevic and Vrecica [182], who point out that they are both merely special cases (when $k=0$ and $k=d-1$, respectively) of their amazing new "center-traversal theorem:"

Theorem 46 (Zivaljevic & Vrecica)
Let $\mu_0, \mu_1, \ldots, \mu_k$ be $k+1$ measures in $\mathbf{R}^d$. Then there exists a k-dimensional affine subspace $S \subset \mathbf{R}^d$ (indeed even a tight one) such that every hyperplane containing S simultaneously $(d-k)/(d-k+1)$-separates all the measures.

Incidentally, one may even generalize this slightly further by demanding that μ_i be tightly λ_i-separated, where $\lambda_0, \lambda_1, \ldots, \lambda_k$ are any specified numbers in $[(d-k)/(d-k+1), 1)$. There is also a corresponding[26] grand generalization of Helly's theorem, due to Dol'nikov [62] (also see [64]).

[24] In fact, we can force there to be d different tight hyperplanes through the Rado point, and in the argument later even $d+1$, but this does not seem to help.

[25] Sometimes called the "Borsuk-Ulam theorem," since it had been conjectured by Ulam.

[26] Another Helly-like result: Katchalski & Lewis [104] showed that if for any 3 among N convex bodies in the plane, there exists a line stabbing all 3 of them, then there exists a line stabbing all N of them – *except* for possibly 603 exceptions. Conjecturally the "603" may be reduced to "2;" H.Tverberg has shown that any counterexample to that conjecture WLOG has $N \leq 49$, reducing its confirmation or disproof to a finite (though enormous) computation.

4.4 *Strengthenings of Miller & Thurston's separator theorem for d-spheres*

Theorem 47 (Strengthened Miller Thurston sphere separator theorem I) *Given N balls in d-space, whose interiors are κ-thick, there exists a sphere S such that $\le (d+1)N/(d+2)$ of the balls lie entirely inside S, $\le (d+1)N/(d+2)$ of the balls lie entirely outside S, and $\le c_d \kappa^{1/d} N^{1-1/d}$ of the balls are cut by S, where $c_1 = 1$, $c_2 = 2$, $c_3 < 2.135$, $c_4 < 2.280$, $c_5 < 2.421$, and more generally if $d \ge 2$,*

$$c_d \le 2d^{1/d}(\frac{\bigcirc_d}{\bigcirc_{d+1}})^{1-1/d} = (\frac{2d}{\pi})^{1/2} \cdot [1 + O(\frac{1}{\log d})]. \quad (49)$$

(See definition 5.) This also works with any of a wide class of weight functions where instead of demanding that $\le N/(d+2)$ of the balls lie entirely inside and outside, we demand that $\le 1/(d+2)$ of the weight lie inside or outside.

Proof. This theorem was the central achievement in [132]. However for an inexplicable reason the results there were stated in weak forms involving "big-O" notation with undetermined constants (which in fact could depend on d in a totally unspecified way), forcing us to re-prove it here. Considering the circumstances, though, we'll work fast and omit some details; also most of what we say was done by Spielman and Teng [167].

In the special case $d = 1$, the claim is trivially seen to be true, so assume $d \ge 2$.

Some key lemmas of [132] are that all "inversive transformations" are conformal and sphere-preserving, and in particular the "stereographic projection" from a d-space onto the surface of a sphere in $(d+1)$-space is conformal and sphere-preserving. (Hyperplanes are regarded as spheres of infinite radius.)[27] Also, given N points on a sphere in $(d+1)$-space (or lying on a flat d-space) there exists an inversive transformation which will (of course) leave these points co-spherical, and such that the center of the sphere will become a Rado point, indeed a tight one. (This is because one may always apply an infinitesimal inversive transformation which "shoves" the Rado point in any desired direction, while leaving the sphere invariant.)

So place 1 "representative" point somewhere in the interior of each of our N balls (arbitrarily) and stereographically map everything onto a sphere Q in $(d+1)$-space in such a way that the sphere center is a Rado point. Now consider a random hyperplane through the sphere center (which upon un-transforming, will be the separating sphere S in d-space). The Rado point property forces the claims about "$\le N/(d+2)$," so the only thing left to prove is the claim about "$\le c_d \kappa^{1/d} N^{1-1/d}$."

[27] The "inversive group" of transformations of $\mathbf{R}^d$ consists of translations $\vec{x} \to \vec{x} + \vec{a}$, scalings $\vec{x} \to c\vec{x}$, and inversions $\vec{x} \to \vec{x}/|\vec{x}|^2$, and everything one may generate by composing these. A famous theorem of J.Liouville states that the elements of this group are the only conformal maps of $\mathbf{R}^d \cup \infty$ when $d \ge 3$.

Now it may be shown[28] that if the Euclidean radii of the N spheres on Q are $r_1, r_2, \ldots, r_N$, then the expected number of them cut by a random hyperplane through the center of Q is

$$\frac{1}{F_{d+1}(1)} \sum_{i=1}^{N} F_{d+1}(r_i), \quad (50)$$

$$F_d(x) = \int_0^x (1 - x^2)^{(d-3)/2} \mathrm{d}x. \quad (51)$$

This is an incomplete beta function. But we will only use the trivial facts that

- $F_3(x) = x$
- if $d > 3$ then $0 < F_{d+1}(x) < x$.
- $F_{d+1}(1) = \frac{(d+1) \bullet_{d+1}}{2d \bullet_d} = \frac{\bigcirc_{d+1}}{2 \bigcirc_d}$.

Hence the expected number C of balls cut by S obeys

$$C \le \frac{2 \bigcirc_d}{\bigcirc_{d+1}} \sum_{i=1}^{N} r_i. \quad (52)$$

Now by considering d-area, we know that

$$\bullet_d \sum_{i=1}^{N} r_i^d \ \le\ \kappa \bigcirc_{d+1}. \quad (53)$$

The combination of (EQ 52) and (EQ 53) with Hölder's inequality [94] leads to

$$C \le \frac{2 \bigcirc_d}{\bigcirc_{d+1}} N^{1-1/d} \kappa^{1/d} (\frac{\bigcirc_{d+1}}{\bullet_d})^{1/d} = 2d^{1/d} (\frac{\bigcirc_d}{\bigcirc_{d+1}})^{1-1/d} N^{1-1/d} \quad (54)$$

which is the result (EQ 49) claimed (the asymptotic series arises from "Stirling's formula").

Theorem 48 (Additional strengthening of Miller by constant factors) *The bound $c_d \kappa^{1/d} N^{1-1/d}$ of the previous theorem may be decreased as follows.*

1. *If $\rho(d, \kappa) \in (0, 1]$ is the maximal possible average euclidean radius achievable by a κ-thick set of N spherical caps on a unit sphere in $(d+1)$-space,* expressed as a fraction *of the euclidean radius of a spherical cap of area $\bigcirc_d \kappa / N$, then we may multiply c_d by*

$$\rho(d, \kappa)\,(1 + o(1)). \quad (55)$$

For all sufficiently large d, $0.5 \le \rho(d, 1) <$ 0.660185901 if N is large enough.

2. *For the unweighted case only, we may (also) multiply c_d by*

$$\frac{1 + (d+1)^{1-1/d}}{(4 + 2d)^{1-1/d}}\,(1 + o(1)), \quad (56)$$

which when d is large is approximately $1/2$.

[28] By realizing that the unit vector orthogonal to the hyperplane lies in a "great belt" of width $2r_k$ if and only if the hyperplane cuts a spherical cap of euclidean radius r_k.

Here the "$o(1)$" terms apply in the limit where $\kappa_d N^{d-1} \to \infty$ with d fixed.

Remark. When $d = 2$, we had $c_2 = 2$. Applying item 1 only yields $c_2' < 1.90463$, while applying both yields $c_2''' < 1.83973$.

Remark. When $d = 3$, we had $c_3 \approx 2.135$. Applying item 1 alone yields (conjecturally) $c_3' \approx 1.48853$; applying item 2 alone yields $c_3'' \approx 1.61876$; and applying both yields (conjecturally) $c_3''' \approx 1.48634$. See remark (i) below.

Proof. To prove (EQ 56) in the unweighted case, we make the center of the sphere a *tight* Rado point (i.e., an intentionally bad one!) so that, ignoring $o(N)$ terms, $N/(d+2)$ cap-representatives lie on some hemisphere and $(d+1)N/(d+2)$ on the other. This allows us to multiply c_d by a factor $\frac{1+(d+1)^{1-1/d}}{(4+2d)^{1-1/d}}$, which is approximately $1/2$ when d is large.

Also, our previous argument for bounding c_d had argued that the sum of some disc areas could not exceed κ times the area of the region they were packed into. But, going further, we realize that there is always going to be some "inefficiency" in our packing, since spheres, unlike cubes, cannot tile. The situation is complicated by the facts that

1. It might be best (in order to maximize the average radius) to use discs whose sizes are slightly unequal, and
2. We are not asking for a packing (except when $\kappa = 1$), we are asking for a κ-thick configuration.

Anyhow, if $0 < \rho(d, \kappa, N) < 1$ is the maximum possible average euclidean radius of κ-thick spherical caps on the surface d-area of a sphere in $(d + 1)$-space, as a fraction of the euclidean radius of a spherical cap of area $\bigcirc_d \kappa/N$, then we are allowed to improve our bound by multiplying c_d by $\rho(d, \kappa, N) < 1$. And in fact, since in the theorem statement we are neglecting $o(N)$ terms, we are even allowed to use $\rho(d, \kappa) \equiv \limsup_{N\to\infty} \rho(d, \kappa, N)$.

The fact that $.5 < \rho(d, 1) < 0.660185901$ arises from the randomizing argument lower bound $2^{-(1+o(1))d}$ and the Kabatiansky & Levenshtein [52] upper bound $0.660185900765^{(1+o(1))d}$ on the density of packings of spherical caps of equal angular diameters θ, $0 < \theta < \pi/2$. To explain how an upper bound on packing density for *equal* balls can lead to an upper bound on the *average* radius for disjoint *unequal* balls: To define a convenient length scale, suppose the radius of N equal spherical caps, under the false assumption that they could tile a sphere's surface perfectly without wasting any area, would be 1. Kabatiansky & Levenshtein then says that in fact, the radius of equal spherical caps would have to be $\le K$, $K = 0.660185900765...$, in high dimensions. With unequal caps, the cardinality fraction with radius $\ge (1 + \epsilon)K$ cannot be more than $(1 + \epsilon)^{-d}$. This will cause the average radius to be

$$\le K + K\frac{\int_0^\infty x(1+x)^{-d-1}\mathrm{d}x}{\int_0^\infty (1+x)^{-d-1}\mathrm{d}x} \tag{57}$$

$$nonumber \tag{58}$$

$$= K + \frac{K}{d-1} = \frac{Kd}{d-1}. \tag{59}$$

If d is sufficiently large, $Kd/(d-1) < 0.660185901$ (note the upward rounding) as claimed. □

Remarks.

(i) The ρ idea is useless when $\kappa \to \infty$, because (as one may easily show by using a random placement of equal spherical caps, with "excess removed") then $\rho \to 1-$. But for any fixed finite κ, improvement is possible.

Table 1 gives the known values of the maximal area density achievable by κ-thick unit discs in the plane, whose centers are constrained to lie on a lattice. (When $\kappa = 1$, this density is optimal even without the lattice constraint [52]. The latest result – $\kappa = 9$ – is due to Temesvari [170]; earlier cases were by Heppes [98] and Blundon [17], among others.) In the cases with $\kappa \le 4$, the number is simply the same as just κ copies of the penny packing, but when $\kappa > 5$, higher density is possible. One might then conjecture (with less and less confidence for higher values of κ) that $\rho(2, \kappa, N)^d$ is upper bounded by κ^{-1} times the entry in the κth row of this table. In fact, for $\kappa = 1$, this conjecture was proved by Spielman and Teng [167] but, more simply, follows immediately from our argument involving lemma **??** and Fejes Toth's [67] [68] bound

$$r^2 \le 3 - \cot\left(\frac{\pi N}{6N - 12}\right)^2 = \frac{8\pi}{\sqrt{3}N} + O(N^{-2}) \tag{60}$$

on the euclidean radius r of $N \ge 3$ disjoint equal circles packed on a sphere.

κ	density
1	$\pi/\sqrt{12} \approx 0.9069$
2	$\pi/\sqrt{3} \approx 1.8138$
3	$\sqrt{3}\pi/2 \approx 2.7207$
4	$2\pi/\sqrt{3} \approx 3.6276$
5	$4\pi/\sqrt{7} \approx 4.7496$
6	$35\pi/\sqrt{384} \approx 5.6112$
7	$8\pi/\sqrt{15} \approx 6.4892$
8	$\frac{3969\pi}{4\sqrt{220-2\sqrt{193}}\sqrt{449+32\sqrt{193}}} \approx 7.5217$
9	$29\pi/\sqrt{84} \approx 8.5694$

Table 1: Maximum lattice densities for κ-thick unit balls in 2D.

Similarly table 2 gives the known values of the maximal density achievable by κ-thick unit balls in 3-space, whose centers are constrained to lie on a lattice [69]. W-Y. Hsiang claims [100] to have finally established "Kepler's conjecture" that when $\kappa = 1$, this density is optimal even without the lattice constraint, but the following experts on sphere packing {K.Bezdek, J.Conway, G.Fejes Toth, T.Hales, D.Muder, N.Sloane} reject that claim[29]. The best upper bound on density for 3D nonlattice packings of equal balls currently available to those disputing Hsiang's proof is Muder's [140] 0.773055 . . ..

[29] But everybody nevertheless believes the conjecture is true.

κ	density
1	$\pi/\sqrt{18} \approx 0.74048$
2	$4\sqrt{12}\pi/27 \approx 1.61226$

Table 2: Maximum lattice densities for κ-thick unit balls in 3D.

With $\kappa = 1$, this is due to C.F.Gauss in 1831 [52]; the result with $\kappa = 2$ is from [70].

It is known [25] that the maximum achievable volumetric density of 120 equal spherical caps on the surface of a 4-ball is

$$12 - \frac{60}{\pi} \sin \frac{\pi}{5} \approx 0.77413, \tag{61}$$

(achieved by the 120 in-balls of the dodecahedral faces of the "120-cell" regular 4D polytope, projected onto a sphere; In this packing, each ball touches 12 neighbors and the points of contact are the 12 vertices of a regular icosahedron) and it seems extremely likely [56] [68] that

$$\rho(3,1)^3 \leq \max_N \rho(3,1,N)^3 = \rho(3,1,120)^3 < 0.77413 \tag{62}$$

This is the conjecture we've used to get $c_3' \approx 1.48853$ in the remark. The stronger conjectural bound $\rho(3,1)^3 \leq \pi/\sqrt{18} \approx 0.74048$ is also plausible and would have led to $c_3' \approx 1.42383$.

(ii) The theorem is also valid if we are interested in having *weight* $\leq 1/(d+2)$ on either side of the separating sphere, for any of a wide class of unit-mass weight measures. However, it is not valid to use the improvement (EQ 56) resulting from the "tight Rado point trick" in this case. The improvements arising from bounds on packing efficiency (i.e. $\rho < 1$) are still legitimate, when you can get them.

(iii) We had hopes that even better constants could be obtained, both here and in the planar separator theorem (§2.7), by use of the "second moment method" from probabilistic combinatorics.

However, this hope is squashed by the following: Consider the usual "penny packing" (where the centers of the pennies lie at the vertices of the equilateral triangle lattice). Consider a random line; how many pennies does it hit? The answer is that the infinite strip of width 1 centered on the line contains within it an asymptotic density of penny centers which does not depend on the line (if we exclude special "rational" lines occuring with probability 0); and by use of this fact it is fairly easy to construct sets of N pennies (equal spherical caps) on the sphere with asymptotically maximal density, and such that random great circles intersect a number of pennies with a maximal deviation from the mean, which is negligible in comparison to that mean itself.

Theorem 49 (Additional strengthening of Miller for equal balls) *If we have N equal radius d-balls which are κ-thick, then there exists a hyperplane separator with $\leq dN/(d+1)$ of the balls entirely on one side (or the other side), and such that at most $O(N^{1-1/d}\kappa^{1/d})$ balls intersect the hyperplane.*

Proof. Choose representative points in each ball. Find a Radon point P of the representatives. Consider a random hyperplane through P. Thanks to the Radon point property, the $\leq dN/(d+1)$ split bound follows. But now we may project the balls radially onto a unit sphere centered at P and then apply the same analysis (on this sphere) as in the previous proofs. The sum of the radii of the spherical caps is $\leq \kappa d \sum_{m=1}^{O(N/\kappa)^{1/d}} (1/m) m^(d-1)$ roughly (the maximum possible value is achieved when the N balls are crammed into as small a ball, centered at P, as possible; the radius of this ball is the upper limit on the sum), which is $O(\kappa^{1/d} N^{1-1/d})$. Hence the result follows by (EQ 52). □

Remark. The assumption of equal balls has improved things in every way: The separation constants are better, we get a slightly better split, and we get to use a hyperplane, not a sphere.

4.5 Two thirds is best possible

Theorem 50 *For each $N > 0$, there exists a set of N disjoint disks in the plane, such that any circle C, cutting $O(\sqrt{N})$ of the disks, and with at most fN disks entirely inside and at most fN entirely outside, necessarily has $f \geq 2/3 - o(1)$.*

Proof. Make 3 small-diameter clusters of $N/3$ disks located near the vertices of an equilateral triangle. Each cluster will be an "exponential spiral" of disks whose radius increases by a constant factor every turn of the spiral, and with a constant number of disks per turn, and such that each disk touches its neighbors in the spiral ordering. For example, one could use the "loxodromic progression" of circles (pictured in figure 9 page 114 of [54]) each of radius $g + \sqrt{g} \approx 2.89005$ (where $g = (1 + \sqrt{5})/2 \approx 1.61803$ is the "golden ratio") in which any 4 consecutive disks are mutually tangent.

In order to get $f < 2/3$, C would have to cut at least 1 of the 3 clusters into two parts, each of cardinality of order N. If C's radius is large compared to the diameter of a cluster, then this is impossible without cutting order N disks. If C's radius is comparable to or smaller than the diameter of a cluster, then both the other clusters will lie outside C. □

Remarks.

(i) The same 2D example, but with d-balls instead of 2-balls, shows that $f < 2/3$ is also impossible for sphere separators for balls in d-space, for any $d \geq 2$.

(ii) This same counterexample also works in a large number of other scenarios, for example if the separator is a square, ellipse of bounded eccentricity, or equilateral triangle, rather than a circle. Or if the objects are squares rather than discs. In all these cases also, 2/3 is best possible[30].

[30] One nice proof for squares arises from clusters that are subsets

We also claim

Theorem 51 *For each $N > 0$, there exists a set of N disjoint disks in the plane, such that any circle C with at most fN disks entirely inside (or entirely outside) it (for some constant f, $1/2 < f < 1$) must cut at least $(1 - o(1))\sqrt{2\sqrt{3}\pi(1-f)f}$ circles.*

Proof sketch. This comes, via a stereographic projection, from spherical caps whose centers are from the point set in the proof of theorem 60. □

4.6 The dependency on d, κ, and N is best possible up to a factor ~ 4.5

In theorem 39, the number of d-cubes partly inside and partly outside the separating box was

$$O(d\kappa^{1/d}N^{1-1/d}) \tag{63}$$

as $N \to \infty$.

Theorem 52 (Construction showing best possible) *For κ-thick objects, this bound (EQ 63) is optimal up to the implied constant factor in the O, assuming we are living in toroidal d-space as in remark (i) of theorem 39.*

Proof. Let $0 < X < 1$. When $d \geq 1$ there exists a lattice tiling of d-space by cubes of side 1 and side X, where each side-X cube is adjacent to $2d$ side-1 cubes (1 per face). Each side-1 cube is adjacent to $2d$ side-X cubes (1 per face) as well as to some side 1-cubes.

When $d = 1$, this tiling is trivial. When $d \geq 2$, the lower-coordinate corners of the unit cubes lie at the points of the lattice consisting of the integer linear combinations of the rows of the following $d \times d$ Toeplitz[31] matrix M:

```
 1  0  0  ...  0  0  X
-X  1  0  ...  0  0  0
 0 -X  1  0  ...  0  0
 0  0 -X  1  0  ...  0
          ...
 0  0  0  ...  0 -X  1 .
```

One may verify that $\det(M) = 1 + X^d$.

Any nonzero integer linear combination of M's rows must have L_∞ distance ≥ 1 from $\vec{0}$. Hence unit cubes whose lower-coordinate corners lie at the lattice points, won't overlap. A cube of side X whose lower-coordinate corner is at $(1, 0, ..., 0)$ also will not overlap any of the unit cubes. To see the above two non-overlap claims, consider the coordinates in left to right order; to get a small value in each coordinate the next coefficient needs to be at least as large to compensate, which can't keep happening since d is finite. An alternative demonstration of the first nonoverlap claim, pointed out by L.Gurvits, is to realize that it is implied by the fact that M^{-1} has L_∞ norm < 1, and this fact in turn is easy to see, once you realize that $(1 + X^d)M^{-1}$ is also Toeplitz, with diagonal values $-X$, $-X^2$, ..., $-X^{d-1}$, 1, X, X^2, ..., X^{d-1} in order from top right to bottom left.

This tiling is of independent interest. When $d = 2$ it was known to the ancients (cf. figure 2.4.2g of [90]), but already when $d = 3$ it's apparently new. In the limit $X \to 0$ it degenerates to the usual cube tiling from the lattice $\mathbf{Z}^d$; while when $X \to 1$ we get the "checkerboard lattice" D_d [52].

If X is irrational, then every point of every hyperplane (except for the union of a measure-0 set with a set of L_∞ diameter $\leq 1 + X$) lies inside a cubical tile. Thus in a strong sense, there are no (perfect) "separating hyperplanes" if $d \geq 2$. Also, no two cubes share a (mutually complete) $(d-1)$-face[32], or for that matter, apparently any k-face for any $k > 0$.

If $X = p/q$ is rational we have a tiling of a cubical d-torus of side $q + p$ by cubes of sides 1 and X, such that every point of every hyperplane (except for the union of a measure-0 set with a set of L_∞ diameter $\leq 1 + X$) lies inside a cubical tile.

Tile a d-torus universe with a roughly $N^{1/d} \times N^{1/d} \times ... \times N^{1/d}$ grid where each cube has side 1 or side α, where α is arbitrarily close to 1, as above.

Now, consider a separating d-box. This box cannot be a 1/3-2/3 separator unless its "inside" and "outside" both have d-volumes between $.33N$ and $.67N$.

Now, the box's surface area S (by an isoperimetric theorem for d-boxes which is readily proven by Steiner symmetrization) must then obey $S \geq 2d(.33N)^{(d-1)/d}$. Then, for any d-box with surface area S, we claim at least $(1 - o(1))S$ of our cubes intersect its surface[33].

To conclude, we've shown the existence of an example with $\kappa = 1$ such that any 1/3-2/3 separating box must cross at least

$$2d\kappa^{1/d}(.33N)^{1-1/d} \tag{64}$$

of our cubes. The same result for any $\kappa > 1$ arises by superimposing κ copies of the example with $\kappa = 1$. □

This is within an asymptotic factor of 4.5 of the upper bound (EQ 32) of theorem 39, which, for d large, is

$$\sim 2.95d\kappa^{1/d}N^{1-1/d}. \tag{65}$$

Remark. Other proofs of optimality up to a constant factor (but with worse constants) may be based on tilings of the plane by dominos with no separating line (leading to a proof for all $d \geq 2$), or on a clever tiling of

of the well known "golden spiral" tiling of the plane by squares, where each successive square's sidelength is g times larger. This is best made by starting with a golden rectangle ($g \times 1$), dividing it into a square and a golden rectangle, and continuing on. (Also see figure 2.4.9 of [90].)

[31] A matrix is "Toeplitz" if it is constant on diagonals.

[32] The small cubes do share their entire face with part of the face of a large neighbor, though.

[33] Worries about overcounting cubes which intersect more than one face of the box may be avoided by realizing that they are asymptotically negligible if every sidelength of the d-box is $\gg 1$. On the other hand, if any sidelength were only $O(1)$, then the box's surface area would be far larger than our bound on S, easily overcoming any $2d$ overcounting factor if $N = d^{\Omega(d)}$.

$\mathbf{R}^3$ by unit cubes, with no separating plane, found by Peter Shor: The cubes have centers at $(0,0,0)$, $(0,1,x)$, $(1,x,0)$, $(x,0,1)$, $(0,1,\bar{x})$, $(1,\bar{x},0)$, $(\bar{x},0,1)$, and $(1,1,1)$ mod 2, where $0 < x < 1$ and $x+\bar{x} = 2$ (leading to a proof for all $d \geq 3$).

4.7 Algorithmic versions

For this subsection, assume the input consists of iso-oriented d-cubes with disjoint interiors. Extensions to the κ-thick case are quite easy. In $O(N^{2d+1})$ steps, one could consider all possible inequivalent rectangle shapes and thus find the best rectangle separator by brute force. However, this approach is inefficient.

A better approach is based on the idea of a "separating d-annulus." This is two concentric d-boxes of bounded CV aspect ratios and with a ratio of linear dimensions bounded below by some constant greater than 1, such that at least a constant fraction of the objects's boundaries lie inside the inner box, and at least constant a fraction lie outside the outer box.

If we can find a separating annulus, then it immediately follows from the assumption that the interiors of the objects are disjoint – by a randomizing argument similar to the one in the proof of theorem 39 involving "$\mathcal{F}$" – that a random d-box containing the inner box and contained in the outer box, will cut an expected number of $O(N^{1-1/d})$ of the objects. Since a nonnegative random variable lies at or below twice its expectation value with probability $\geq 1/2$, we may, then, simply guess a box and then confirm in $O(Nd)$ steps that it works (with an expected number of 2 guesses being required before succeeding).

It is also possible to find the best (that is, cutting the fewest objects) separating box concentric with the inner annulus box, deterministically in $O(N \log N)$ time by sorting the min-radius and max-radius points in the objects, then performing a linear time scan over the resulting 1D intervals.

Also, a not necessarily best, but nevertheless good enough (i.e. within a constant factor of best), separating box concentric with the inner annulus box, may be found by a "bucket sort" of these radius values into $2N$ equally spaced "buckets," and then find an upper bound on the number of 1D intervals overlapping each bucket. One may prove that at least one bucket must have few enough overlaps, because otherwise the total d-volume of the objects would be too large. This approach runs on $O(N)$ time on a RAM featuring a unit time $\lfloor x \rfloor$ operation.

We know of two efficient ways to find a separating d-annulus; one is deterministic and the other is randomized.

4.7.1 Deterministic method

For simplicity, we will describe the method in 2D assuming the objects we are separating are iso-oriented squares. We'll then describe its generalization to d dimensions. We will find a separating d-annulus made of two concentric d-cubes with sidelength ratios $1:3$. Unfortunately, this method does not achieve the optimal constant "2/3" in the split in theorem 39.

1. We assume WLOG that the N input squares are in general position, in particular the coordinates of all their corners are distinct. The reason we may assume this is because when we input them we could preshrink them by random factors selected from the range $(1-\epsilon, 1)$, where $\epsilon > 0$ is infinitesimal.

2. For any $\epsilon > 0$, we can find[34] in linear time numbers x_1, x_2, x_3 such that the four intervals $(-\infty, x_1]$, $(x_1, x_2]$, $(x_2, x_3]$, (x_3, ∞) each contain 1/4 (plus or minus ϵ) of the x-coordinates of the left hand sides of the squares.

3. Choose whichever of the two finite intervals I is shorter – WLOG $I = (x_2, x_3]$. Let W denote the set of squares whose left sides fall in $(x_2, x_3]$ and now do the same trick with y-coordinates of tops of squares in W – giving intervals $(-\infty, y_1]$, $(y_1, y_2]$, $(y_2, y_3]$, (y_3, ∞) – let J be the shorter of the two finite ones. Then $1/16 \pm 2\epsilon$ of all squares have their top left corners in the rectangle $R = I \times J$. Choose whichever of I and J is longer, let the inner square S_0 of the annulus just contain R, and let the outer square S_1 have the same center but with side *almost* three times longer.

4. Conclusion: Immediately $R \subseteq S_0$ contains bits of at least $N/16$ squares. Also, S_1 does not reach across the other of the two inner intervals in the long direction (e.g. if the long direction of R is x-direction, it does not reach x-coordinate x_1), and so at least $N/4$ of the squares are not wholly inside S_1.

5. To optimize this, instead of 1/4-1/4-1/4-1/4 split in step 2, use .1-.4-.4-.1. This will force at least $.1N$ squares to be at least partly inside S_0 and at least $.1N$ to be at least partly outside. S_0 and S_1 will still be concentric with length ratio $3:1$.

This algorithm generalizes to d dimensions. The best splits are A_i-B_i-B_i-A_i in the ith dimension, where A_i and B_i may be generated backwards starting from $A_d = B_d = 1/4$ by means of the recurrence $2A_i = 1 - 2B_i = A_{i+1}/(1 + A_{i+1})$. The solution of this recurrence is $2A_{d-k} = 1/(2^{k+2} - 2^k - 1)$.

The outer cube S_1 will be 3 times the sidelength of the inner cube S_0, and at least $A_1 N$ cubes have borders at least partly inside S_0 and at least $A_1 N$ cubes will not be wholly inside S_1, where $A_1 = 1/(2^{d+2} - 2^d - 2)$.

4.7.2 Randomized method

The randomized method is as follows. Pick a random subset q of the N objects, and find their optimal separator (or just find any separating d-annulus) by brute force

[34] By using a linear time selection algorithm [16] [72].

in (e.g., if the objects are iso-oriented d-cubes) $O(|q|)^{d+1}$ steps. (Anyhow, some function of $|q|$ and d steps, more generally.)

Now we claim that, if $|q|$ is a sufficiently large constant, in fact it will suffice if $|q| = \Omega(\epsilon^{-2}d^3 \ln(d/\epsilon))$, then the resulting separator will in fact *be* a separator with a split balance only $1+\epsilon$ times worse than best possible, with constant success probability. Success may be verified in $O(Nd)$ steps.

This may be proven using the VC dimension techniques explained in §4.7.4.

4.7.3 Comparison

To compare the two algorithmic approaches: The approach of §4.7.1 ran in $O(dN)$ time but produced a possibly exponentially poorly balanced – $\Omega(2^d)$:1 – split. Also, although perhaps it may be generalized to separate objects other than d-cubes by separators other than boxes, such generalizations seem to be ad hoc and unreliable. The randomized approach (§4.7.2) produces a well balanced split – $O(1)$:1 at worst, and within $(1+\epsilon)$ of the optimal balance, with high probability – but consumes $(d/\epsilon)^{O(d)}N$ (expected) runtime. This approach is trivial to generalize maximally.

This same approach also works to algorithmicize theorems 47 and 48.

4.7.4 VC dimension in a nutshell

Suppose you have a set of allowed "ranges" R over some space S.

Definition 53 *The "VC-dimension" of R in S is the maximum v such that for any v-point subset Q of S: for each of the 2^v subsets q of Q, there exists an $r \in R$ such that $r \cap Q = q$, i.e. Q is "shattered" by R.*

Example: If S is d-dimensional Euclidean space and R is halfspaces, then $v = d+1$.

The two key theorems about VC dimension [177] [63] [96] [112] are as follows.

Theorem 54 (ϵ-approximation theorem) . *Given a finite set Q of N points in S, and a set R of ranges (subsets of Q)*[35] *with finite VC dimension v, there exists a subset q of these N points which is an "ϵ-approximation for R." Meaning: for all $r \in R$,*

$$\left|\frac{|q\cap r|}{|q|} - \frac{|Q\cap r|}{|Q|}\right| \le \epsilon. \tag{66}$$

Furthermore, not only does such an ϵ-approximation exist, in fact a random subset of cardinality

$$\frac{c}{\epsilon^2}\left(v \ln\frac{v}{\epsilon} + \log\frac{1}{\delta}\right) \tag{67}$$

for some absolute constant c (and $c = 24$ suffices) will work, with success probability at least $1-\delta$.

[35]Instead of requiring Q to be finite, we may instead demand that the indicator function of R over Q be *measurable* over the combined space of both Q and R.

It seems to be unknown what the best possible constant c is, in theorem 54. Indeed the best possible "constant" may tend to zero, since when $\epsilon \to 0+$ with v held fixed, better asymptotic behavior is available [129].

However (EQ 68) in the theorem below has the best possible constant [112].

Theorem 55 (ϵ-net theorem) *Given a finite set Q of N points in S, and a set R of ranges with finite VC dimension v, there exists a subset q of these N points, of cardinality*

$$(1+o(1))\frac{v}{\epsilon}\ln\frac{v}{\epsilon} \tag{68}$$

which is an "ϵ-net for R." Meaning: for all $r \in R$,

$$\frac{|Q\cap r|}{|Q|} \ge \epsilon \quad\Rightarrow\quad q \cap r \ne \emptyset. \tag{69}$$

In fact, if "v" in (EQ 68) is weakened to "$8v$" then a random subset works, with constant probability.

Now, to prove the algorithmic claim in §4.7.2, we make the "ranges" be cubical "annuli," which we can see[36] have VC dimension $v \le 2d+2$ in d-space, and ask for ϵ/d approximation.

4.8 Counterexamples to putative strengthenings

The Miller-Thurston theorem 47 cannot be made to hold for separating, e.g., squares, at least not by means of some simple trick such as circumscribing circles about the squares. This is because configurations of disjoint squares exist whose circumballs are infinitely thick, cf. figure 10.

Also, it is impossible to get a separator theorem like theorem 39 or 47 but using a *hyperplane* as the separator. This is because "exponential spiral" configurations of disjoint squares exist such that any line with a constant fraction of the squares lying to each side, must cut at least a constant fraction of the squares (§6.2 and footnote 30).

There exists a set of N 2-thick convex objects in d-space, for any $d \ge 3$, $N \ge 2$, such that every pair of objects overlap. (To see this, consider N mutually neighboring Voronoi polytopes as in [160], shrink each of them slightly to make them all disjoint, and then "inflate" each face of each polytope into a low "pyramid" so that every polytope now intersects every other polytope with which it used to share a $(d-1)$-face.) Hence there is no way to separate even the elements of a single pair by use of a hyperplane, or sphere, or anything else.

A simple 2D version of the counterexample above is the fact than N line segments can (and generally will, if they are long enough) be mutually crossing. Hence slightly

[36] Even the full set of all doubly convex d-annuli, that is, convex bodies with a hole which is an enclosed convex body, has VC dimension $\le 2d+2$, because the following $(2d+3)$-point set $S = S_1 \cup S_2 \cup P$ cannot be shattered: S_1 is $d+1$ points forming the vertices of a d-simplex; S_2 is another $d+1$ forming an enclosed d-simplex, and P is a final point enclosed by both simplices. It is impossible for all of S_2 but none of $S_1 \cup P$ to be in a doubly convex d-annulus.

"thickening" the line segments yields a 2-thick set of N convex objects, all mutually overlapping. But this counterexample is not as satisfying as the (≥ 3)-dimensional counterexample above, because our 2D convex objects are no longer "pseudodiscs."

"Pseudodiscs" in the plane are the interior of Jordan curves such that any two pseudodiscs intersect in a manner topologically equivalent to the way two circular discs intersect. (E.g. with boundaries crossing each other at ≤ 2 points.) Let $N(\kappa)$ denote the maximum number of κ-thick pairwise mutually overlapping pseudodiscs. $N(1) = 1$.

Fact 56 $N(2) = 4$. *Consequently, among any 5 convex 2-thick pseudodiscs in the plane, there exists a line separating at least one of them from at least one other among them.*

Proof. The 2-thick regions may be thought of as the edges of a planar graph whose N vertices correspond to the N pseudodiscs. (Planar: because no two 2-thick regions may intersect, due to 2-thickness.) Kuratowski's theorem on the non-planarity of K_5 then shows $N(2) \leq 4$. Figure 1 shows $N(2) \geq 4$. The existence of the line follows from the fact that any two disjoint convex sets, are separable by a hyperplane. □

We do not even know whether $N(3)$ is finite.

5 Planar graph separator theorems: four proofs and two lower bounds

In §5.1 we present simplified versions of older proofs of the planar separator theorem. Our geometric results lead to new proofs (§5.2): a "squares" proof which also works for torus and Klein bottle graphs, and a "circles" proof yielding a record value [167] for a planar separator constant.

5.1 Two simplified combinatorial proofs

Lipton & Tarjan [125]'s original proof may be greatly simplified if we weaken their constant to 4 instead of $\sqrt{8}$. We may assume the graph is maximal planar WLOG since the separator of a graph with an edge superset can only be worse, and since the algorithm of [44] will suffice to embed the graph in the plane (and then triangulating it by adding diagonals in linear time is trivial) in linear time.

Theorem 57 (Simple planar separator) *The V vertices of any maximal planar graph G with non-negative vertex weights W_v summing to 1 may be partitioned into 3 sets A, B, C so that $\sum_{v\in A} W_v \leq 2/3$, $\sum_{v\in B} W_v \leq 2/3$, and $|C| \leq 4\sqrt{V}$. This partition may be found algorithmically in $O(V)$ time.*

Algorithmic Proof. Do a breadth first search of G starting from some vertex r. This BFS forms a tree T of the shortest paths from r. Find the least D such that there is weight $\leq 1/2$ among the vertices at distances $\geq D$ from r. Find two distances D_1 and D_2 so that $\max\{0, D - \sqrt{V}/2\} \leq D_1 \leq D < D_2 \leq D + \sqrt{V}/2$ and such that the numbers of vertices at distances D_1 and D_2 from r total $\leq 2\sqrt{V}$. (These must exist.)

Consider cutting G into ≤ 3 pieces by removing the vertices at distances D_1 and D_2 from r. If the middle piece has weight $\leq 2/3$ we are done, so suppose not. In that case (by contracting all the vertices at distances $\geq D_2$ from r into a supervertex q, then contracting all the vertices at distances $\leq D_1$ from r into r; weights on supervertices are the sums) we may WLOG assume that all vertices of G are at distances $\leq \sqrt{V}$ from r.

Now the "outside" of T may be thought of as a triangulated simple polygon (drawn on a sphere) with $\leq 2V-2$ sides (the $\leq V-1$ edges of T, going both ways) and whose diagonals are the edges of $G-T$. By Chazelle's (weighted) polygon cutting theorem (§2.1), this polygon has a "splitting diagonal" with weight at most 2/3 to either side of it. But the addition of this edge to T results in a graph with exactly one cycle, of circumference $\leq 2\sqrt{V}$.

Let C consist of the vertices of this cycle, together with the two levels at distances D_1 and D_2 from r in the original G. C has total cardinality $\leq 4\sqrt{V}$ and must (by its removal) split G into pieces of weight $\leq 2/3$. □

A more elaborate version [60] of the argument above reduces the bound "$4\sqrt{V}$" to "$\sqrt{6V}$," still with an $O(V)$-time algorithm. This is the best known linear time algorithmic planar separator result, as measured by either of the figures of merit (EQ 7), (EQ 8).

Perhaps the simplest proof of the above planar separator theorem (but now purely an existence proof, without any $O(V)$-time algorithm) arises by simplifying the proof of [4].

Ultrasimple graph-theoretic proof. Consider the shortest simple cycle containing weight $\geq 2/3$ on or "inside," but $\leq 2/3$ inside. If this cycle has $\leq 4\sqrt{V}$ edges we are done, so assume not and we'll show a contradiction. View the cycle as a "square with $> \sqrt{V}$ dots on each edge." Remove everything outside the cycle. Disjoint paths from the ith vertex on the East side of the square to the ith on the West must exist, and similarly North-South. Hence there is a "grid" forcing there to be $> V$ vertices inside/on the cycle, an impossibility. Why must all these disjoint paths exist? Well, if a path existed shorter than $\sqrt{V}$ we could contradict minimality; if two paths conjoined we would contradict minimality also, via "Menger's theorem," (a special case of the max-flow min-cut theorem) which states in this case that the $\sqrt{V}$ North-Souths had better exist disjointly (for "flow$\geq \sqrt{V}$"), otherwise an East-West path ("cut") will exist shorter than $\sqrt{V}$. □

A much more elaborate version [5] of the above proof reduces the bound "$4\sqrt{V}$" to "$\sqrt{4.5V}$." This is the best known weighted (and unweighted) planar separator result, as measured by (EQ 8).

5.2 *Two geometric proofs*

But it is also possible to prove our planar separator theorems geometrically. We now prove an *edge* separator theorem for torus graphs.

5.2.1 *Geometric proof via squares – also works for torus graphs.*

There is a beautiful 1-1 correspondence, first discovered by Brooks, Smith, Stone, Tutte [29], between

- polyhedral (that is, 3-connected) planar graphs G with E edges, one of them distinguished, and
- "squared rectangles" R (that is, a tiling of a rectangle by squares, not necessarily of equal sizes) with $E-1$ tiles.

We'll now describe and prove this.

Each vertex of G will correspond to a maximal horizontal line segment in R (note: such a line segment could be the concatenation of more than one tile side). Each edge of G will correspond to a square tile in R, except for the distinguished edge, which corresponds to the exterior of R. Finally, each face of G corresponds to a maximal vertical line segment in R.

Figure 5: Squared square with 13 tiles. (Bouwkamp 1946.)

Figure 6: Equivalent electrical network with currents shown. Dashed edge is the "distinguished" edge containing the battery; other edges are 1Ω resistors.

Now view each edge of G as a 1-ohm resistor, except for the distinguished edge, which is a battery. The current through each edge will correspond to the width of the corresponding square tile, and the voltage across each resistor will correspond to the height of the corresponding square tile. The currents that will then arise satisfy "Kirchoff's laws," which state that

- **charge conservation** The current flowing into a vertex equals the current flowing out – this corresponds to the fact that the sum of the sidelengths of the squares above some horizontal line segment must equal the sum of the sidelengths of the squares below.
- **no flow cycles** The voltage differences around any cycle (in particular a face) of G, excluding cycles which involve the battery, is 0 – this corresponds to the fact that the sum of the sidelengths of the squares to the left of some vertical line segment must equal the sum of those to the right. (Or equivalently, this is charge conservation in the planar dual graph G' using planar dual currents; planar duality corresponds in the rectangle world to turning your head 90 degrees.)
- **Ohm law** The current across an 1-ohm resistor is the same as the voltage across it. This corresponds to the fact that the height of a square is the same as its width, i.e. the tiles are indeed squares.

Since any such electrical circuit has a valid current flow, which is unique, through it, and as we've seen the equations of current flow are exactly the conditions for validity of a squared rectangle, we see that for each G with a distinguished edge there exists a unique squared rectangle. On the other hand, for each squared rectangle there clearly exists a unique corresponding G.

Now applying theorem 39 instantly shows that for any embedded 3-connected planar graph G with E edges, there exists a Jordan curve (which may travel through faces, or encounter vertices, or cross or traverse edges, of G) crossing or traversing $\le (4+o(1))\sqrt{E}$ edges and having a total number of encounters with vertices, plus having a total number of trips through faces, totalling $\le (4+o(1))\sqrt{E}$, and such that this Jordan curve will split any of a variety of weight measures 1/3-2/3 or better.

We conclude

Theorem 58 *For any polyhedral graph with E edges, there exists a smooth closed curve traversing and/or crossing $\le (4+o(1))\sqrt{E}$ edges in total, whose removal will subdivide the graph into 3 parts, the two main ones each having weight $\le 2/3$ (for any of a large variety of permissible non-negative weight functions summing to 1, e.g. weights on edges).*

Remarks:

(i) The proof above leads to a polynomial time algorithm, because the square sizes may be found by solving a system of V sparse linear equations embodying Kirchoff's laws. But this algorithm is not particularly fast – using the scheme of [127] improved as in footnote 5, we get a solution in $O(V^{1.188})$ arithmetic steps.

(ii) The squares proof above was actually only for 3-connected planar graphs ("planar nets") but generalizes to show that 2-connected torus graphs, each of whose valencies and face sizes are ≥ 3, have a 1/3-2/3 separator theorem with $\le (4+o(1))\sqrt{E}$ vertices in the separator via "squared rectangular toruses" (a "rectangular torus" is a rectangle with opposite edges identified to get "periodic boundary conditions," cf. figure 7).

Squared tori arise as follows. Take a 2-connected embedded torus graph G (each of whose valencies and face sizes are ≥ 3, and whose edges are to be regarded as 1-ohm resistors) and draw a smooth closed non-self-intersecting curve J which cannot be shrunk to a point (due to the fact that it winds exactly once around the "handle" or the "hole" of the torus) and which cuts edges at most once per edge. Wherever J crosses an edge, put a battery in series with the 1-ohm resistor corresponding to that edge. All the batteries have the same voltage and all their polarities are oriented "outward" with respect to J.

(iii) This is not the record smallest constant in a torus graph separator theorem; e.g. by combining [1] and [4] one may get an (unweighted) 2/3-separator for torus graphs with $\sqrt{12.5V}$ vertices. However, our separator has simple structure, since it corresponds to a single closed curve; the competing theorems so far have only achieved better constants at the cost of structural complexity. Consideration of the $\sqrt{V} \times \sqrt{V}$ toroidal grid graph indicates that it would be impossible to shrink "$\sqrt{12.5}$" below "2."

(iv) One can also make "squared cylinders" [9], which arise very similarly to squared rectangles (which are a special case of them) except that instead of having a distinguished *edge*, we have a distinguished pair of vertices. Of course, these still are merely a special case of squared tori.

(v) "Squared Moebius strips" and "squared Klein bottles" can also be done (J in the torus construction now becomes a cut across the Moebius strip, respectively across the orientation reversing side of the Klein bottle square, rather than a closed curve).

(vi) This method does not work to produce abundances of "squared projective planes" because the current is supposed to flow "down" and there is no down. Colloquially, "you can comb the hair on a rectangle, cylinder, Moebius strip, torus, and Klein bottle, but you can't comb the hair on a projective plane." (See figure 7.) The question of whether it is possible to "square" the projective plane depends on the definition of a "squaring." Since L_∞ balls can include digons (as in figure 7) and even stranger shapes (figure 8), it seems unnatural to define "squares" to be "L_∞ balls." This problem arises because the projective plane, as we have drawn it in figure 7, is not a manifold at all. However it may be realized as a hemisphere with antipodal equatorial points identified (this is just noneuclidean "elliptic" geometry) in which case we do have a genuine manifold. Then every point would be locally euclidean, but only in *infinitesimal* neighborhoods, since there is constant positive Gaussian curvature. If we require a "square" to be a 4-gon, then this curvature frustrates squaring attempts: It is impossible to tile elliptic geometry with topological 4-gons with 90° corners[37]! Nevertheless by consideration of a cube drawn (by central projection) on the surface of a sphere (and by identifying antipodal points, the sphere's surface may then be converted to elliptic geometry), we see it is possible to "square" the projective plane, if squares with 120° corners, and equal geodesics for edges, are allowed.

From these ruminations we conclude that the only "natural" way to define a "squared projective plane" is to require the "squares" to be simultaneously L_∞ balls, *and* locally convex 4-gons, i.e. to disallow squares with a noneuclidean point of the manifold inside. This same convention allows one to consider, e.g. "squaring the cube" by regarding the surface of a cube, using the usual surface metric, as a manifold with 6 special noneuclidean points. Indeed, it becomes possible to consider "squaring" any other manifold with a finite number of special points and divisible into regions isometric with euclidean squares. See §8.2.

Figure 7: Some 2-dimensional manifolds. Pairs of points on opposite sides of the square are identified in the orientations indicated by the arrows – for sides with arrows. If there are no arrows, that square side is just a boundary. All manifolds shown have exactly the same metric as does the Euclidean plane, in all sufficiently small finite neighborhoods of any non-boundary point, *except* for the projective plane, whose "corners" are locally not euclidean – as may be seen by considering the small 2-gon shown, which has two 90° turns, both "inward." Thus as we have drawn it, the projective plane is not really a manifold at all.

Figure 8: An L_∞ ball in the projective plane (center shown)

5.2.2 *Geometric proof via circles*

It is a theorem first shown by Koebe in 1936 [111], using deep techniques from complex analysis, and later by Andreev and Thurston using simpler techniques from hyperbolic geometry, that:

Theorem 59 (Koebe) *Every V-vertex planar graph may be realized as the contact graph of V interior-disjoint circles in the plane. In fact, every V-vertex planar graph G may be realized as a circle configuration in this way while simultaneously realizing its planar dual G' and with the contact points for the dual circles (corresponding to the edges of G') being identical to the contact points for the primal circles (corresponding to the edges of G);* and *with the primal and dual circles being orthogonal at these contact points.* And *this realization is unique up to inversive transformations.*

Smith [163] extended this theorem (and reproved it) by showing that suitable center coordinates, and the radii, of the circles, could be found to D decimal places in time polynomial in V and D. Smith's proof worked by reducing the problem to a "convex minimization program" in which the function being minimized corresponded to a certain hyperbolic volume. (Also Mohar [138] independently found a polynomial time algorithm.)

[37] Consider the graph of a "squared" manifold with no boundary (where the vertices of the graph are the points where at least three squares meet and edges are the sections of boundaries of squares joining them). With F squares there are $4F$ "corners," occurring two at each vertex of degree 3 and four at each vertex of degree 4. So if V_3 and V_4 are the numbers of such vertices respectively we have $4F = 2V_3 + 4V_4$, but also $2E = 3V_3 + 4V_4$ and $V = V_3 + V_4$ where E and V are the numbers of edges and vertices respectively. From these equations we get $E = V + F$, so the Euler characteristic of the surface is 0. Thus a squared torus or Klein bottle is allowed, but a squared projective plane is not.

One may argue about whether the Koebe, Andreev, Thurston geometrization via circles is more or less beautiful than the Brooks, Smith, Stone, Tutte geometrization via squares, but anyhow, applying theorem 47 instantly yields a $(3/4, 1.83973\sqrt{V})$ separator theorem for planar graphs [167], and a $(3/4, 1.90463\sqrt{V})$ weighted separator theorem with arbitrary weights (summing to 1) on each vertex, edge, and/or face of a polyhedral planar graph. If the graph is maximal planar, the separator produced will automatically be a simple cycle.

These are the best known weighted and unweighted planar separator results, as measured by (EQ 7).

5.3 *Two lower bounds (one new) for planar separator constants*

A lower bound by Djidjev [60], whose proof we will simplify greatly, is

Theorem 60 *There exist V-vertex planar graphs such that any subset C of the vertices whose removal splits the graph into A, B with $|A| \le xV$, $|B| \le (1-x)V$, $x \le 1/2$, must have cardinality at least*

$$(1+o(1))\sqrt{2\sqrt{3}\pi x(1-x)V}. \tag{70}$$

Proof. There are configurations of V points on the unit sphere such that their convex hull graph has every edge length (in angular measure) $< \pi\sqrt{2/(\sqrt{3}V)}$, and such that every spherically convex region of area $4\pi x$, $0 < x < 1/2$, x fixed, contains $\le xV\pi/\sqrt{12} + o(V)$ points. (Such configurations may be constructed with, e.g., the techniques of [178]; one approximates the sphere by a sequence of developable surfaces on which are drawn equilateral triangle grids.) Now the result follows from the isoperimetric theorem on the sphere [71]. □

However, as we will now see, Djidjev's lower bound is improvable.

Smith [162] observed that the "R-refined icosahedra" maximal planar graphs I_R (obtained by starting with an icosahedron I_0 and getting I_{R+1} from I_R by dividing every triangle into 4 triangles by adding extra vertices at the edge midpoints) has $20 \cdot 4^R$ triangles and $10 \cdot 4^R + 2$ vertices. All the vertices have valence 6 except for 12 vertices of valence 5. These graphs lead to lower bounds which are slightly stronger than Djidjev's in the range $0.08112 \le x \le 0.35759$.

For example when $x = 1/3$, we need at least $\sqrt{5V/2} > 1.5811\sqrt{V}$ vertices in the separator, whereas Djidjev's bound was ≈ 1.5551; when $x = 1/4$, we need at least $\sqrt{2.1V} > 1.4491\sqrt{V}$, versus Djidjev's ≈ 1.4284. Specifically:

Theorem 61 ("Icosahedral" lower bound for planar separator theorem) *Let C be a cycle of I_R with at least $V/2$ vertices outside it ($V = 10 \cdot 4^R + 2$, R sufficiently large) and at least xV inside. Let $|C|$ denote its length (number of edges).*

- *If $.325 \le x \le 1/2$, then $|C|$ is at least $\sim 5\sqrt{V/10}$. (This is superior to Djidjev's lower bound in the range $0.325 \le x \le .35759$.)*
- *If $.3 \le x \le .325$, then $|C|$ is at least $\sim \sqrt{(4x+1.2)V}$. (Superior to Djidjev for all x.)*
- *If $\frac{1}{6} \le x \le .3$, then $|C|$ is at least $\sim \sqrt{(6x+.6)V}$. (Superior to Djidjev for all x.)*
- *If $\frac{2}{15} \le x \le \frac{1}{6}$, then $|C|$ is at least $\sim \sqrt{(8x+\frac{4}{15})V}$. (Superior to Djidjev for all x.)*
- *If $0 \le x \le \frac{2}{15}$, at least $\sim \sqrt{10xV}$ edges are required. (Superior to Djidjev's bound when $0.08112 \le x \le \frac{2}{15}$.)*

These cases arise respectively from the following.

Theorem 62 (Isoperimetric theorem for I_R) *For simple cycles in I_R containing exactly $4^R t$ triangles, $0 \le t \le 20$:*

- *With $6.5 \le t \le 13.5$: The length of the cycle is $\ge 5 \cdot 2^R - O(1)$. (Achieved by "geodesics" centered at icosahedron vertices.)*
- *With $6 \le t \le 6.5$: If the length of the cycle is $2u \cdot 2^R$, then the number of triangles inside is at most $(2u^2-6+o(1))4^R$ (for $6 \le 2u^2-6 \le 6.5$). (Achieved by the vertices lying at path distance approximately $(u-2)2^R$ from two adjacent icosahedron faces.)*
- *With $\frac{10}{3} \le t \le 6$: If the length of the cycle is $3u \cdot 2^R$, then the number of triangles inside is at most $(3u^2-2+o(1))4^R$ (for $\frac{10}{3} \le 3u^2-2 \le 6$). (Achieved by the vertices lying at path distance approximately $(u-1)2^R$ from an icosahedron face.)*
- *With $\frac{8}{3} \le t \le \frac{10}{3}$: If the length of the cycle is $4u \cdot 2^R$, then the number of triangles inside is at most $(4u^2-\frac{2}{3}+o(1))4^R$ (for $\frac{8}{3} \le 4u^2-\frac{2}{3} \le \frac{10}{3}$). (Achieved by the vertices lying at path distance approximately $(u-\frac{1}{3})2^R$ from the middle one-third section of an icosahedron edge.)*
- *With $0 \le t \le \frac{8}{3}$: If the length of the cycle is $5u2^R$, then the number of triangles inside is at most $(5u^2+o(1))4^R$ (for $0 \le 5u \le \frac{8}{3}$). (Achieved by "regular pentagons" of vertices of distance approximately $u \cdot 2^R$ from icosahedron vertices.)*

Proof. It is straightforward to confirm the following claims about simple cycles C in I_R of bounded length and containing the maximum number of triangles inside.

1. In an anticlockwise traversal of C (so that the interior is on the left), at each vertex we must either "go straight" or make a "left turn." In this statement, we may define "go straight" to mean that there are three triangles interior to C meeting the vertex, whilst at a left turn there are at most two.

2. It is impossible for C to have 3 consecutive vertices at which we respectively go straight, turn left by an acuter than minimal bend, and go straight, unless C has length at most 9.

3. If three consecutive left turns are at u, v and w then the number of steps from u to v and the number of steps from v to w differ by at most 2 (otherwise changing the number of steps between turns by 1 or 2 creates another cycle with the same length and a larger number of triangles inside),

4. The number of left turns is equal to 6 minus the number of 5-valent original icosahedral vertices which are interior to C.

If u is the (approximate) number of steps between left turns, then the length of C is $ku + O(1)$ where k is the number of turns, $k \geq 1$. The number of 5-valent vertices inside is $6 - k$.

If $k = 6$ we get a hexagon with length $6u + O(1)$ and $6u^2 + O(u)$ triangles inside. This is beaten by the $k = 5$ case.

If $k = 5$, there is exactly one 5-valent vertex interior; call it v. Cut I_R along one of the original icosahedron edges incident with v, and lay it out onto the planar grid of equilateral triangles (usually called the hexagonal grid). The cycle C becomes part of an almost regular hexagon in this grid, traversing five of the triangular faces of I_R, and the distance of the cut ends from v must both be the same. The almost regularity of the hexagon now forces one of the left turns to be within $O(1)$ steps of the cut. Hence the centre of the hexagon is within $O(1)$ edge-lengths of v, and the number of triangles inside it but not in the cut-out part, and thus inside C, is $5u^2+O(u)$. Here C has length $5u + O(1)$.

If $k = 4$ there are exactly two 5-valent vertices, U and V, inside, and C has length $4u + O(1)$. They must be adjacent in the original icosahedron. We can argue in a similar fashion, cutting I_R along the edges from U to V and also along one more of the original icosahedral edges incident with U not sharing a triangle with UV. Again, C must form part of a nearly regular hexagon, and the constraint that the distances from the cut ends to the appropriate vertices of the hexagonal grid are equal implies that this distance is $u - 2 \cdot 2^R/3 + O(1)$. It can now be calculated that the number of triangles inside C is $4u^2 - \frac{2}{3}4^R + O(u)$.

If $k = 3$ there are three 5-valent vertices inside. These must lie in a triangle of the original icosahedron, since otherwise a simple convexity argument applied to the shape of C shows that a fourth must be included. (Or it is very nearly missed, due to the slight inequality in the lengths of straight paths of C, but this is very nearly the case $k = 4$ so can be ignored.) The cutting argument as above now shows that the cut ends of C have distance $u - 2^R + O(1)$ from the 5-valent vertex. For $k = 2$ there are four inside, which again by convexity must form two triangles, and cutting and flattening gives that C has distance $u - 2 \cdot 2^R + O(1)$ from the double triangle. For $k = 1$ there are five vertices inside and convexity rules this out altogether (or it's very nearly equal to the $k = 0$ case). If $k = 0$ we clearly have one of the geodesics mentioned above, with $t \geq 6.5$.

The claims in the theorem follow from these results after checking which of the five cases is minimal for which ranges of t. (Also, it is not hard to verify that using the union of two or more cycles C, will not allow enclosing more triangles with the same circumferential length, essentially because the number of enclosed triangles is a concave-∪ function of u.) Note that in each range, the required values of r can indeed be realized. (If they couldn't, that would lead to an improvement in the bound anyway.) □

Probably these bounds could be improved further by use of more complicated families of graphs than I_R. For example, using refinements of the regular dodecahedron with each pentagonal face subdivided by 5 "spokes" might extend the region of superiority versus Djidjev's bound further toward $x = 1/2$. But great perseverance would be required to take on the requisite more complicated case analyses.

6 Separator theorems for geometric graphs

Joe Ganley's "Steiner tree web page"
`http://www.cs.virginia.edu/ jlg8k/steiner/`
contained

Conjecture 63 (Ganley, false)[38] *Given N sites in the plane, their rectilinear Steiner minimal tree (RSMT) always has the property that there exists a horizontal or vertical line, which cuts the RSMT at $O(\sqrt{N})$ places and which separates the sites into 2 sets, each of cardinality $< cN$ for some constant c, $0 < c < 1$.*

We managed to disprove this conjecture. However, it turns out that some separator theorems quite similar to Ganley's conjecture are true, and for several other geometrical graphs besides just RSMTs.

As one application of these theorems, we'll obtain new algorithms for finding optimal RSMTs.

6.1 Definitions of geometrical graphs

Given N point sites in Euclidean space, a "geometrical graph" is a set of line segments ("edges") whose endpoints include the sites. In the below we'll assume the sites are in general position so that we may avoid worrying about, e.g., nonunique minimum spaning trees.

The optimum *traveling salesman tour (TST)* is the shortest cyclic path that visits every site. A cyclic path, visiting every site (a "tour"), whose length may not be decreased by removing k edges and substituting k others (while preserving the property of being a tour) is "k-optimal."

A *Steiner tree* is a network which connects all the sites. If the network consists solely of line segments parallel to

[38] Ganley now blames this conjecture partly on Jeff Salowe.

the coordinate axes, then it is a *rectilinear Steiner tree* (RST).

The *Steiner minimal tree* (SMT) is the shortest Steiner tree, and the RSMT is the shortest rectilinear Steiner tree.

A *spanning tree* is a network made of site-to-site line segments *only*, which connects all the sites. The *minimum spanning tree* (MST) is the shortest spanning tree.

A *matching* is a partitioning of the N sites into $N/2$ pairs (for this, we assume N is even). The *minimum matching* (MM) is the matching such that the sum of the lengths (i.e. distances between the 2 members of) of the pairs is minimum.

The *all nearest neighbor* digraph (ANND) is the digraph in which every site is joined to its nearest neighbor; ANN is the undirected version of this graph (with duplicate edges uniquified).

The *Gabriel graph* (GG) is the graph of intersite line segments such that the circumballs of the line segments are empty of sites.

The *Delaunay triangulation* (DT) is the set of d-simplices, whose vertices are sites, such that the circumballs of the simplices are empty of sites. It is known that

$$ANND \subseteq MST \subset GG \subset DT. \tag{71}$$

6.2 Disproof of the Ganley conjecture

Theorem 64 *The Ganley conjecture 63 is false. The analogous conjectures for RMST (rectilinear minimum spanning tree), SMT (Euclidean Steiner minimal tree), MST (Euclidean minimum spanning tree), RANN (rectilinear all nearest neighbor graph), ANN (Euclidean all nearest neighbor graph), and TST (optimal traveling salesman tour) are also false.*

Proof. A suitable point set for all the rectilinear graphs (aside from TST) is a "squared off exponential spiral." Specifically, view the plane as the complex z plane so that we may write (x, y) coordinates with only one complex number $z = x + iy$. Let the nth point z_n in an infinite sequence z_0, z_1, ... of points be defined by

$$z_n = z_{n-1} + i^{\lfloor n/k \rfloor} 2^{n/k}. \tag{72}$$

For all the Euclidean graphs (other than TST), a "round" exponential spiral such as

$$z_n = 20^{n/200} (\cos \frac{2\pi n}{200} + i \sin \frac{2\pi n}{200}) \tag{73}$$

works.

For TST, one can make a similar example using a double spiral (to provide a return path for the tour).

In the above examples, the proof is not yet complete because we have not yet demonstrated that the SMT, MST, etc, really are the graphs which they "obviously" are. In the cases of MST, RMST, ANN, and RANN, it is easily shown (for example by considering "Kruskal's algorithm" for constructing MSTs) that the spiral paths and the graphs are identical.

We now sketch how this may also be accomplished for RSMT, SMT, and TST.

To argue that the obvious squared-spiral RMST path of (EQ 72) with $k = 11$ (see figure 9) is in fact the RSMT, we may use the (apparently new) "method of electrical shorts."

Figure 9: Disproof of the Ganley conjecture. The fact that the spiral path shown is the RSMT may be shown inductively.

Recall that the RSMT is the shortest network of "wires" required to electrically interconnect all the "cities," subject to the constraint that all the wires have to be horizontally or vertically oriented.

The fundamental lemma underlying this method is

Lemma 65 (Electrical shorting) *Let A and B be finite point sets in a metric space, and let Q be a "metal plate" (closed polygon) with A interior and B exterior. Then the length of* $\mathrm{SMT}(A \cup B)$ *is at least as great as the sum of the lengths of* $\mathrm{SMT}(A \cup \{Q\})$ *and* $\mathrm{SMT}(\{Q\} \cup B)$. *This lemma also holds if "TST" is substituted for "SMT" everywhere.*

Note: By a Steiner tree on $A \cup \{Q\})$ we mean a tree T containing all the points in A and at least one point in Q, and not crossing Q; its length is computed for the purpose of this lemma as the length of T minus the length of $T \cap Q$). Hence the name "shorting": connections along Q cost nothing.

We will let Q be the dashed polygon shown in the figure, A are the sites inside and on the boundary of Q, and B are the sites outside it.

Assume inductively that the shortest RSMT of the sites A within Q, even if the entire exterior P_B of Q is assumed to be "metal" so that the cost of electrical interconnections within P_B is zero is the spiral path shown. Now by a small exhaustive search, we realize that the shortest RSMT electrically interconnecting the next k sites on the spiral path, and Q (where now the interior of Q is regarded as being made of metal, and a $2\times$ larger scaled version of Q, rotated $90°$, enclosing these next k sites in a fashion precisely similar to the way Q encloses the previous k sites [partially shown dot-dashed in the figure], is also regarded as having a metallic *exterior*) is the spiral path shown. Now applying lemma 65 (or, actually, a generalization involving two polygons) shows that the inductive assumption on N sites follows from itself with $N - k$ sites. Hence, after turning our head $90°$ and adjusting our spectacles so that the length scale is 50%, we may now continue the induction with k more sites. The base case of the induction is also handled using a small exhaustive search[39].

[39]Incidentally, if exhaustive searches to find RSMTs on 11-site sets with two metallic polygons are regarded as too painful, we remark that we chose the parameters in this example mainly to get a nice looking picture, as opposed to trying to reduce the size of

Similarly the fact that the Euclidean MST and SMT are identical for the round spiral point set (EQ 73) may be shown by a similar, but easier, induction instead involving the "shorting" of roughly circular regions, and increasing by only 1 site per inductive step, not $k = 11$.

For TST (or RTST), we argue that the least cost to tour all the k outer sites, given that the $N - k$ inner sites are all "shorted," is (by exhaustive search) the obvious double spiral portion. This again leads to an inductive proof. □

6.3 Minimum spanning trees, Steiner trees, and All nearest neighbor graph

We'll now argue that MST, SMT, and ANN in the Euclidean metric enjoy geometric separator theorems. This can be (and will be) shown in two ways:

1. We construct a convex object (called a "diamond") of bounded DW aspect ratio enclosing each MST edge, and show that all the diamonds are interior disjoint. Then we apply a previous geometric separator theorem for interior disjoint convex objects of bounded DW aspect ratio.

2. We show that the circumballs of the MST edges, although not necessarily disjoint, are "$2^{O(d)}$-thick" (definition 8). Then we apply the Miller-Thurston separator theorems 47, 48 for κ-thick balls.

As stated, these two approaches only prove a separator theorem for MST, but similar results immediately follow for SMT (since an SMT is the MST of its N sites and its $\le N - 2$ Steiner points) and ANN (since ANN is a subgraph of MST, and hence any κ-thickness and disjointness properties for objects associated with MST edges, are even more true for ANN edges).

The first method of proof is conceptually simpler and shorter; and also it yields the best constants in low dimensions. The second method seems to lead to better constants (by a factor of order d^2) in high dimensions, and is the progenitor of the more powerful techniques we will need later.

Definition 66 *The "diamond" of an MST edge AB denotes [81] the intersection of two cones of half angle* $30°$ *with respective apices at* A *and* B*, each with axial line* AB*.*

For example in 2D this is a 60-120-60-120 rhombus whose $60°$ *corners are at* A *and* B*. In 3D, the diamonds are the bodies of revolution obtained by rotation of the rhombi.*

Lemma 67 (Diamond property for euclidean MSTs) *The diamonds of the MST edges in any Euclidean d-space, are interior disjoint.*

the computation. By using growth factors larger than 2, the computations may be rendered trivial and much larger "safety factors" may be incorporated.

Proof. In 2D, this is shown in §8.6 of Gilbert & Pollak [81]. On page 22 of the same paper, a proof in dimensions $d \ge 3$ is presented, and attributed to R.L.Graham and J.H.Van Lint. □

Remark. Lemma 67 is clearly best possible in the sense that these diamonds cannot be increased by adding any protuberance (at least, if that addition is made in ignorance of the rest of the MST).

By theorem 39 (in the form of remark (ii) after the proof) applied to $60°$ rhombi[40] in 2D, we get

Theorem 68 (2D MST separator theorem) *Consider the MST of* N *sites in the Euclidean plane. There exists a rectangle* R *such that at most* $2N/3$ *of the sites (or the MST edges, or at most* $2/3$ *of any of a wide class of "weights") are wholy inside* R*, at most* $2N/3$ *are wholy outside, and* $\le (4 \cdot 3^{1/4} + o(1))\sqrt{N}$ *MST edges are partly inside and partly outside* R*.*

In general dimension d, we also get a 2/3-separating rectangle, but the bound on the number of edge crossings obtained by this technique is $O(d^{3/2})N^{1-1/d}$, which is not as good a bound as we will obtain in theorem 73 below (albeit at the cost of weakening the split balance from $2:1$ to to $d+1:1$).

Lemma 69 (MSTs are short) *The MST of* N *sites in (or on the surface of) a unit ball in d-space has length* $< \frac{2d}{d-1} N^{1-1/d}(1 + N^{-1/d})$.

Proof. Rescale the unit ball to have radius $(1 + N^{-1/d})^{-1}$. By considering their d-volume, we see that it is impossible for balls centered at the sites and of radius $N^{-1/d}$ all to be disjoint. Hence, at least one site has a neighbor within distance $2N^{-1/d}$. Draw the edge between these two closest neighbor sites and then mentally remove one of them. Now continue recursively to draw a spanning tree of the remaining $N-1$ sites. The total length of the spanning tree we construct will be $\le 2\sum_{m=1}^{N} m^{-1/d}$, which is $< \frac{2d}{d-1} N^{1-1/d}$. Rescaling by $1 + N^{-1/d}$, we get the stated bound. □

Definition 70 *Let* τ_d *be the "kissing number" in d-space, that is, the maximum number of interior-disjoint unit d-balls which can touch one.*

It is known [52] that $\tau_1 = 2$, $\tau_2 = 6$, $\tau_3 = 12$, $\tau_8 = 240$, and $\tau_{24} = 196560$; and $\tau_d \le 2^{0.401d+o(d)}$.

Lemma 71 (3D ball lemma) *We claim that if two 3D balls* B_1 *and* B_2 *exist, both containing the origin, and with the radius of* B_2 *being at least 2 times as large as the radius of* B_1*, and if two antipodal points of* B_1 *both lie outside (or on the surface of)* B_2*, then the angle subtended by the centers of* B_1 *and* B_2 *at the origin is at least* $60°$*.*

[40] Whose area is $1/\sqrt{3}$ for a unit length SMT edge, so we may take $\tau = \sqrt{3}$.

Proof. Let B_1 be of unit radius WLOG.

This is equivalent to the statement that if we have a tetrahedron 0ABC where we call the midpoint of AB, M, then if $AB = 2$, $2 \le 0C \le \min(AC, BC)$, $0M = 1$, then $\angle M0C \ge 60°$.

The minimal angle occurs when $0C = AC = BC = 2$ and $C0M$ is a 30-60-90 triangle. A and B may be rotated anywhere on the circle of radius 1 at which the two spheres intersect, and the angle $\angle M0C$ will remain 60°. On the other hand if AB is rotated out of this plane then clearly one of $\{A, B\}$ will penetrate the big sphere, which is illegal. This proves the configuration mentioned locally minimizes the angle. In fact it does so globally because plainly the angle minimizing configuration must have both A and B on the large sphere's surface (otherwise we could push one closer) and if so, the two spheres must intersect at a great circle of the smaller sphere, leading to our configuration if the radius of the larger sphere is $r = 2$. If $r > 2$ the angle is larger. □

Theorem 72 (MST circumballs aren't too thick) *The circumballs of the edges of an MST in d-space are M-thick with*

$$M \le 1 + 2\left(\frac{4d}{d-1}\right)^d(\tau_d - 1) \le 2^{2.401d+o(d)}. \tag{74}$$

Proof. First, we remark that MST edges obey the "empty lune property" that the intersection ("lune") of the two d-balls of radius L centered at A and B (for some MST edge AB of length L) is empty of sites. The circumball of AB is entirely contained in the lune of AB, hence all such circumballs are empty. (Proof: if a site C were in the lune of AB, then C is connected by an MST path to either A or B (first), say WLOG A, and then removing AB and substituting BC would make the MST shorter but still connected, a contradiction.)

Now suppose some point of d-space (WLOG, the origin) is contained in M circumballs. WLOG let the smallest of these balls have diameter 1 and indeed let the diameter of the ith smallest ball be $Q[i]$.

All the $2M$ endpoints of the MST edges defining the balls must lie in the ball of radius $Q[M]$ centered at the origin. These MST edges have total length L with $L \ge M$. However, the length of an MST of $2M$ points in a d-ball of radius $Q[M]$ is $L < \frac{2^{3-1/d}dQ[M]}{d-1}M^{1-1/d}$. This leads to a contradiction if $Q[M] \le 2^{-2+1/d}\frac{d-1}{d}M^{1/d}$. Indeed, we have a contradiction if $Q[k]/Q[j] \le 2^{-2+1/d}\frac{d-1}{d}(k-j)^{1/d}$ for any j, k, $1 \le j < k \le M$.

Therefore, it must be the case that $Q[j] \ge 2Q[j - 2(\frac{4d}{d-1})^d]$, i.e. the balls at least double their diameter after every $2(\frac{4d}{d-1})^d$ balls.

Now, consider two balls B_1 and B_2, the second one having a factor of 2 larger radius. We will be interested in the centers of these 2 balls, the MST endpoints which form a diameter of B_1, and the origin. These 5 points lie in a 3-dimensional subspace. (Generically 5 points define a 4-space, but here the 3 points on the MST edge are collinear, so we get a 3-space.) Hence by lemma 71, their centers must subtend an angle $\ge 60°$ at the origin.

Since the maximum number of vectors, all $\ge 60°$ apart, which can exist is τ_d, we conclude finally that $M \le 1 + 2(\frac{4d}{d-1})^d(\tau_d - 1)$. □

Remark. Of course, the theorem above is also true of SMT edges (since SMTs are MSTs of their vertices) and ANN edges (since ANN⊆MST).

By applying separator theorem 48 (theorems 39 and 47 could also be applied) we conclude

Theorem 73 (MST, SMT, ANN separator theorem) *Let MST, SMT, and ANN be the minimum spanning tree, minimum Steiner tree, and all nearest neighbor graphs, respectively, of N sites in d-space, $N > d \ge 1$. Then there exists an d-sphere such that at most $(d+1)N/(d+2)$ of the sites are inside it; at most $(d+1)N/(d+2)$ are outside; and at most $cd^{1/2}N^{1-1/d}(1+o(1))$ of the graph edges cross the sphere. Here $c = 2^{1.401}\sqrt{2/\pi} \approx 2.11$ suffices. Here the "o" applies when both d and N go to ∞; if only one of them does, it should be replaced by an "O."*

6.4 *Rectilinear minimum spanning trees, Steiner trees, and All nearest neighbor graph*

In this section we'll show that *rectilinear* MSTs, SMTs, and ANN graphs enjoy geometric separator theorems. One might initially think this demonstration would just be a straightforward modification of the analogous arguments for the non-rectilinear graphs. In fact, some of these analogies can be carried through, but others do not work.

The easiest example of an analogy that does work, is for 2D RMSTs, which satisfy a "diamond property" analogous to lemma 67.

Lemma 74 (RMST diamond property in 2D) *If each (horizontal or vertical) RMST line segment in the plane is regarded as the diagonal of a (45° tilted) square, then: all these squares are interior disjoint.*

Proof. WLOG CD is horizontal. If AB is vertical the result is immediate because CD's diamond is empty of A and of B. (We've previously seen it's impossible for any other site to be inside an RMST edge's L_1 circumball.) So AB is horizontal too; say A is left of B, C is left of D WLOG. Let AB be shorter than CD WLOG. (Equal lengths are possible too, but we ignore that case by a random infinitesimal pre-perturbation.) If the diamonds corresponding to AB and CD overlap, where AB is shorter than CD, then we claim $\ell(CA) < \ell(CD)$ and $\ell(DB) < \ell(CD)$, a contradiction with the claim that AB and CD are both RMST edges. To see that, start at the corner of CD's diamond contained in AB's, WLOG this is the uppermost corner of CD's, CD is horizontal, and AB is horizontal. This corner is closer to both of $\{A, B\}$ than it is to either of $\{C, D\}$. Now walking left to A only makes you get closer to C. Similarly walking right to B only makes you get closer to D. □

Remark. Of course, RSMT and RANN also have interior disjoint diamonds in 2D.

Remark. However, it is possible for the L_2-circumballs of RMST edges to be unboundedly thick, even in 2D, see figure 10.

Figure 10: Example of unboundedly thick L_2-circumballs for RSMT edges.

Definition 75 *The "θ-diamond" of a line segment AB means the intersection of the cone with apex A and axial ray $\vec{AB}$ with the cone with apex B and axial ray $\vec{BA}$, where both cones have halfangle θ.*

Unfortunately,

Lemma 76 *RMST and RANN do not enjoy any disjoint θ-diamond property, even in 3D and even for arbitrarily small $\theta > 0$.*

Proof. In 2D draw two length-1 edges crossing one another at right angles at their midpoints. Now join the Eastmost and Southmost vertices by a long chain of tiny RMST edges that stay far away from the crossing. Now lift the "crossing" out of the plane by any $\epsilon > 0$, and we claim the result (figure 11),

Figure 11: A 3D rectilinear all nearest neighbor graph with only infinitesimally skinny disjoint diamonds.

is a valid RMST, and even with care a valid rectilinear all nearest neighbor graph, in 3D. □

But RSMT still does enjoy a diamond property, lemma 80 below.

Definition 77 *An "L_1-ball" of radius r, centered at the origin in d-space, is the convex hull of the $2d$ points with coordinates that are permutations of $(\pm r, 0, 0, ..., 0)$.*

Lemma 78 (Empty octahedron property) *The L_1 ball whose diameter is an RMST (axis-aligned) edge, is empty of sites.*

Proof. Same as proof of the empty lune property mentioned in the proof of theorem 72, only now with L_1 distances. □

Lemma 79 (RSMT diamond property) *If $d \geq 3$ and $\theta < \operatorname{arccot}(2\sqrt{d-1})$, then the θ-diamonds of the line segments of an RSMT in d-space are interior disjoint.*

Actually, rather than proving lemma 79, we will prove the following strictly stronger lemma 80, which involves stretched d-octahedral, rather than biconical, "diamonds." The new diamonds include the old ones, but are a lot larger in some directions, e.g. the angle of the edges to the main axis is $\arctan \frac{1}{2} \approx 26.6°$.

Lemma 80 (Stronger RSMT diamond property) *For an RSMT edge XY form a "diamond" as follows: for any point R on XY, include all points Z such that $D_1(R, Z) \leq (1/2)\min\{D_1(R, X), D_1(R, Y)\}$, where D_1 denotes L_1 distance. All these diamonds are interior disjoint.*

Proof. Suppose the diamonds of RSMT edges AB and XY intersect. Let P be a point in the intersection. WLOG let AB be the further of the two edges from P and of length 1. Let Q and R be the closest points to P on AB and XY respectively. Then WLOG $D_1(Q, A) \leq D_1(Q, B)$ and $D_1(P, Q) < D_1(Q, A)/2$. Hence $D_1(R, Q) < D_1(Q, A) \leq D_1(Q, B)$, contradicting the definition of the diamond. □

Remark. The "diamonds" in lemma 80 *still* are not optimal in any dimension $d \geq 3$. (With $d = 2$ this was due to lemma 74 since $45° > 26.6°$.) This is because the intersection of our diamond with a hyperplane perpendicular to XY is an L_1-ball in $d - 1$ dimensions, which may be "inflated" somewhat due to the fact that if two skew diamonds just touch, they contact only on their edges (1-skeletons) and not on higher dimensional facets. Hence, if we leave the 1-skeleton invariant but "inflate" the higher dimensional facets (possibly into curved surfaces), we can still guarantee disjointness. We do not know what optimal diamonds are.

When $d = 2$, we may combine theorem 39 and lemma 74 to get

Theorem 81 (RMST, RSMT, RANN separator theorem when $d = 2$) *Let MST, SMT, and ANN be the minimum spanning tree, minimum Steiner tree, and all nearest neighbor graphs, respectively, of N sites in 2-space, $N > 2$. Suppose the number of edges in such a graph is E. Then there exists a rectangle, rotated 45° to the coordinate axes, such that at most $2N/3$ of the sites are inside it; at most $2N/3$ are outside; and at most $4(1+o(1))\sqrt{E}$ of the graph edges cross the box boundary.*

Remark. It seems plausible that the 45°-tilted squares whose diagonals are RMST edges, cannot tile the plane and indeed only cover some[41] fraction ρ of the plane, $0 < \rho < 1$ (assuming we ignore RMST edges longer than some constant).

Also by combining lemma 80 with theorem 39 we find that[42]

Theorem 82 (RSMT separator theorem) *Let G be the RSMT of N sites in d-space, $N > d \geq 1$. Let if have E line segments. Then there exists an iso-oriented d-box such that at most $2N/3$ of the sites are inside it; at most $2N/3$ are outside; and at most $32e^{-2}d^2(1 + o(1))E^{1-1/d}$*

[41]Perhaps $\rho = 1/2$ is best possible? Cf. [46]. If so, the "4" in the theorem will be improvable to $4\sqrt{\rho}$.

[42]The $32e^{-2}d^2$ factor arises as follows. A factor of $8d/e$ comes from theorem 39, and the remaining factor of $4d/e$ is really $(4^d d!)^{1/d}$, arising via remark (ii) of theorem 39 from the fact that the d-volume of our stretched d-octehedral diamond, for a unit-length RSMT edge, is $8 \cdot 4^{-d}/d!$.

of the RSMT edges cross the box boundary. Here the "o" applies when both d and N go to ∞; if only one of them does, it should be replaced by an "O."

Because of the unfortunate lemma 76, theorem 82 does not hold for RMST and RANN. However, it is possible to generalize the thickness argument (lemma 72) to show:

Theorem 83 (RMST circumballs aren't too thick) *The L_1-circumballs of the edges of an RMST in d-space are M-thick with*

$$M = d^{O(d)}. \tag{75}$$

This is done with the aid of the following lemmas, whose proofs are omitted since they are analogous to previous proofs from §6.3.

Lemma 84 (RMSTs are short) *The RMST of N sites in (or on the surface of) a unit L_1-ball in d-space has length $< \frac{2d}{d-1}N^{1-1/d}(1+N^{-1/d})$.*

Lemma 85 (3D octahedron lemma) *We claim that if two 3D L_1-balls B_1 and B_2 exist, both containing the origin, and with the radius of B_2 being at least 2 times as large as the radius of B_1, and if two antipodal points of B_1 both lie outside (or on the surface of) B_2, then the center of B_1 also lies outside (or on the surface of) B_2. Hence the angle subtended by the centers of B_1 and B_2 at the origin is at least* $\operatorname{arccot}(\sqrt{d})$.

We conclude similarly to in §6.3 that

Theorem 86 (RMST, RANN separator theorem) *Let RMST, RANN be the minimum spanning tree, and all nearest neighbor graphs, respectively, of N sites in d-space, $N > d \geq 1$. Let the graph under discussion have E edges, $1 \leq E < N$. Then there exists an iso-oriented d-box such that at most $2N/3$ of the sites are inside it; at most $2N/3$ are outside; and at most $O(d^2)E^{1-1/d}$ of the graph edges cross the box boundary.*

6.5 Optimal traveling salesman tours

Lemma 87 *Let $\lambda > 1$. Then there exists k_λ such that the set of circumballs of the edges of every minimal TST is (λ, k_λ)-thick. (See definition 9.) Indeed it will suffice if*

$$k_\lambda = 16^d \log_2 \lambda, \tag{76}$$

and it will also suffice if the TST is merely "2-optimal" (i.e. can't be shortened by removing 2 edges and substituting 2 others).

Proof. Let $\epsilon > 0$ be fixed. We can choose k large enough that if some point is covered by more than k circumballs whose sizes vary by factor at most λ, then by the *pigeonhole principle*, at least three of these circumballs come from edges aa', bb' and cc' whose lengths (after scaling) are all within ϵ of 1, and such that a, b and c lie in the same ball of radius ϵ, as do a', b' and c'.

The minimal TST, outside these three edges, joins them in a cyclic fashion, and clearly at least one of these joins must be from a, b or c to a', b' or c'. Assume wlog that it is a to b'. Then replacing the tour edges aa' and bb' by ab and $a'b'$ shortens the tour provided ϵ was chosen sufficiently small – a contradiction.

In fact it will suffice in the first paragraph if

$$k = \frac{\log \lambda}{\log([1+\epsilon]/[1-\epsilon])}\left(\frac{1+\epsilon}{\epsilon}\right)^{2d} \tag{77}$$

and it will suffice in the second paragraph if $\epsilon = 1/3$, which leads to (EQ 76). □

To apply part (c) of theorem 39, we also need to know that at most a constant number of TST edges larger than the rectangular "separating annulus" in the proof of theorem 39, can intersect it. (A very large number of TST circumballs could intersect the rectangle, but we ignore them unless the TST edge itself does.)

Lemma 88 *At most a constant number ($2^{O(d)}$ suffices) of TST edges of length ≥ 1 can intersect a unit cube or ball. (It will suffice if the TST is 2-optimal.)*

Proof. Suppose not. Then among the unboundedly large number of such TST edges, we may by the *pigeonhole principle* find 3 all at angles within ϵ of each other, and "nearly overlapping," i.e. a length of at least 1/2, say, of one is very close to a similar length of the other. But then the "minimal" TST could be shortened. □

We may now combine lemmas 87 and 88 and part (c) of theorem 39 to get

Theorem 89 (TST separator theorem) *Let TST be any 2-optimal traveling salesman tour of N sites in d-space, $N > d \geq 1$. Then there exists an iso-oriented d-box such that at most $2N/3$ of the sites are inside it; at most $2N/3$ are outside; and at most $2^{O(d)}N^{1-1/d}$ of the TST edges cross the boundary of the box.*

6.6 Minimum matching

An argument very similar to (but easier than) the TST argument in §6.5 shows that

Theorem 90 (MM separator theorem) *Let MM be the min-length matching of N sites in d-space, $N > d \geq 1$, N even. Then there exists an iso-oriented d-box such that at most $2N/3$ of the sites are inside it; at most $2N/3$ are outside; and at most $O(d)N^{1-1/d} + 2^{O(d)}$ of the MM edges cross the boundary of the box.*

Remark. It is also possible to prove a result like this by showing that the θ-diamonds of MM edges are $2^{O(d)}$-overloaded, if θ is a sufficiently small constant.

Remark. No matter how small $\theta > 0$ is, the θ-diamonds of MM edges can be unboundedly thick as $N \to \infty$ with θ and d fixed. Therefore, no argument based on κ-thickness could have achieved anything for MM (and presumably the same is true for TST).

6.7 "Spanners" and "Banyans"

Definition 91 *A "$(1+\epsilon)$-spanner" of a set of points is a subgraph of the complete Euclidean graph where for any u and v the length of the shortest path from u to v is at most $(1+\epsilon)$ times the Euclidean distance between u and v.*

Arya et al. [7] showed (building on earlier work by [57]) – and their work was redone more quantitatively by in an appendix of [149] that for N sites in d-space, and any $\epsilon > 0$, $(1+\epsilon)$-spanners exist

1. Whose total length is only $d^{O(d)}$ times longer than the Steiner minimum tree of the N sites
2. With maximum valency $(d/\epsilon)^{O(d)}$,
3. Which may be constructed in $(d/\epsilon)^{O(d)}N + O(dN \log N)$ time.

We would now like to indicate here that furthermore, WLOG, these spanners also obey a (λ, κ)-thickness property. Specifically,

Theorem 92 *WLOG, for some $\theta = \Omega(\epsilon)$, the θ-diamonds (definition 75) of the Arya et al. spanner edges are (λ, κ)-thick, for some $\kappa = (d/\epsilon)^{O(d)}$ and with $\lambda = (d/\epsilon)^{O(1)}$.*

Proof sketch. Follow the argument, sketched in Arya et al. [7]'s section 5 and again more quantitatively in the appendix of [149], which proves (what [149] call) "(κ, c)-isolation."

This argument remains valid, at the present level of precision, if, instead of small cylinders of height and radius $c\ell$ (where ℓ is the length of the spanner edge), which are κ-thick, we instead use skinny diamonds, of width $c\ell$. □

A similar result necessarily holds for the "banyans" of [149].

Lemma 93 *At most a constant number ($\epsilon^{-O(d)}$ suffices) of Arya spanner edges of length ≥ 1 can intersect a unit cube or ball.*

Proof sketch. Suppose not. Then among the unboundedly large number of such spanner edges, we may by the pigeonhole principle find 2 at angles within ϵ of each other, and "nearly overlapping," i.e. a length of at least 1/2, say, of one is very close to a similar length of the other. But this would contradict the spanner "shortest edge in ϵ-cone" properties enjoyed by the Arya et al [7] construction. □

We may now apply part (c) of theorem 39 to get

Theorem 94 (Spanner & Banyan separator theorem) *Let B be the $(1+\epsilon)$-spanner or $(1+\epsilon)$-banyan graph used in [149]. Then there exists an iso-oriented d-box such that at most $2N/3$ of the sites are inside it; at most $2N/3$ are outside; and of the E edges in the spanner (or banyan), at most $(d/\epsilon)^{O(d)}E^{1-1/d}$ of the edges cross the boundary of the box.*

Theorem 94 has tremendous applicability. For example, it immediately proves separator theorems for the $(1+\epsilon)$-approximately optimal SMTs, MSTs, TSTs, and MMs arising by replacing all edges in the optimal versions, by paths in the spanner or banyan.

6.8 Delaunay triangulations do not have separating circles

Since we've just demonstrated that a large number of well known geometric graphs have geometric separator theorems, it seems only fair to mention a graph without one – the "Delaunay triangulation."

For N sites in 3D, the Delaunay triangulation edge-graph can be the complete graph K_N [160] so clearly there is no separator, even a graph-theoretic one.

For N sites in 2D, the Delaunay triangulation edge-graph is planar and hence always has a *graph-theoretic* $(2/3, \sqrt{8N})$ separator (cf. §2.7), but we argue that it need not have any separating *circle*. To do so, we exploit the fact [66] that the Delaunay triangulation of N sites in the plane, is the same as the convex hull of those N sites, stereographically projected onto a sphere in 3D. The separating circle is then going to correspond to a plane cutting the sphere and the convex hull.

Theorem 95 *Let $c > 0$. Then there exist N point sites on a sphere in $\mathbf{R}^3$ such that any plane with at least cN sites on each side of it, must cross at least $\Omega(N)$ edges of the convex hull of the sites.*

Proof sketch. Consider some large but constant number K of roughly equal spherical caps packed on the surface of the sphere. Space $(N/K - 1)$ sites uniformly around the perimeter of each cap, and one site at each cap center. Notice the convex hull graph will include as subgraphs all the "spoked wheels" corresponding to each of these caps. If a plane cuts a constant fraction of the perimeter off any cap, then it must cross at least $\Omega(N/K)$ convex hull edges, proving the theorem – unless it cuts *precisely* through the cap center. However, if the caps are in "general position" then it is impossible for a plane to cut precisely through the center of more than 3 caps. Also, if the caps are packed reasonably densely and K is large enough as a function of c (It will suffice if $K = \Omega(1/c)$) then it is impossible to avoid cutting at least a constant fraction off at least a constant number (in fact $\Omega(\sqrt{K})$) of caps[43]. □

Remark. The same construction also shows that the 2D Gabriel graph GG need not have a $o(N)$-separator.

[43] This may be seen by considering cutting the usual hexagonal penny packing by a line – no matter what line one uses, a constant fraction will be cut off some penny.

7 Applications: Algorithms and data structures

7.1 Point location in iso-oriented d-boxes

The following problem is a classic task in computer science, with efficient solutions in low dimensions tracing back to G.Lueker and D.Willard in the late 1970s.
Given: A set $\mathcal{S}$ of iso-oriented d-boxes, $d \geq 2$.
Task: To preprocess them to create a data structure to support "point location queries."
These queries are: "Name the boxes containing the point $\vec{x}$."

Three possible generalizations are as follows.

1. Box generalization: instead of a query *point* we have an *iso-oriented d-box* and then we are to name the boxes intersecting the query box.

2. Semigroup generalization: each of the boxes have "weights" which are elements of some semigroup with an operation '+'; and instead of naming the boxes, we are required to state the value of the 'sum' of the weights. (For example, we could find the maximum-weight box by using *max* as our '+'.)

3. Group generalization: same as above, but '+' is a group operation, i.e. there is a '-'. (For example, we could count the boxes by using unit weights and '+' is integer addition.)

Chazelle [38] [37], building on work by a large number of other authors, solved the box generalization of our problem, with query time $(K + \log N) \cdot O(\log^{d-2} N)$ (where $K - 1$ is the size of the output) on a pointer machine, using storage $O(N[\lg N/\lg\lg N]^{d-1})$, and with $O(N \lg^{d-1} N)$ time required to build the data structure. Chazelle [39] then showed, for the box generalization of our problem, the optimality of this storage bound, on a pointer machine, in any fixed dimension d. (Apparently Chazelle's lower and upper bounds differ by a factor $d^{O(d)}$.) Chazelle [41] showed a lower bound on the query time (amortized over N queries), in the semigroup and box double generalization of our problem, of $O(N[\lg N/\lg\lg N]^{d-1})$.

Edelsbrunner, Haring, and Hilbert [65] proposed a solution to our original problem, but provided the boxes were *interior disjoint*, requiring: $O(Nd)$ space to store the data structure, which required $N \cdot O(d + \log N)$ time to build, and for which a point location query could be carried out in time[44] $O(\lg^{d-1} N)$.

In [162] it was observed that the disjointness requirement of [65] could be replaced by a κ-thickness requirement, at the cost of increasing the query time bound to $O(\kappa + \lg^{d-1} N)$, but with no other cost.

Smith (section III.C.8.2.10 of [162]) further generalized the problem by allowing the "boxes" to instead be any mixture of coordinatewise scaled L_p balls ($1 \leq p \leq \infty$), and achieved $O(\kappa + \lg^d N)$ query time and $O(N \lg^{d-1} N)$ space, with $O(N \lg^d N)$ time required to build the data structure. Probably some of the space reduction techniques of [38] also could be applied here.

None of the solutions mentioned above utilized any sort of separator theorem, and instead depended mainly on properties of ranks as binary numbers.

In fact, our separator theorem 36 does not seem to allow improving the previous results of [65] [162]. However, it does allow us to get approximately the same results using a different method.

Theorem 96 *If the boxes are κ-thick, the task defined at the beginning of this section may be solved using Storage: $O(Nd)$. Build time: $O(d^2 N \ln(N+2d-\kappa))$. Query time: $\leq \binom{2d\ln(N+2d-\kappa)}{d-2} O(d+\kappa+\log N)$.*

Proof. If $d \leq 2$, one may solve the problem by Chazelle's method in optimal ($O(N \log N)$ preprocessing time, $O(N)$ storage, $O(\log N)$ query time.

Otherwise, use the separator theorem 36 to decompose the boxes recursively into a ternary tree. The 3 child subtrees of a node correspond to the boxes entirely (1) left of, (2) right of, (3) crossing the separator hyperplane.

Note, subtrees of type (3) are $(d-1)$-dimensional, because we may project all the boxes in the crossing set orthogonally down onto the hyperplane (also the query point) while preserving κ-thickness.

Because each of the 3 sets is guaranteed to make $N + 2d - \kappa$ become a factor $1 - 1/(2d)$ times smaller, the tree has depth $\leq \ln(N + 2d - \kappa)/\ln[2d/(2d-1)] < 2d\ln(N+2d-\kappa)$. Because each non-leaf node of the tree has outvalency 3 and requires $O(1)$ storage (3 pointers, a number in $\{1, 2, \ldots, d\}$, and a coordinate) and the leaf nodes store disjoint sets of objects, the total storage required is $O(Nd)$.

To locate a point: If the root is childless, determine the set of boxes the point is in by brute force examination of all the boxes (of which there will be $< 2d + \kappa$).

Otherwise, see if the point is to the left of, right of, or on the separating hyperplane for the root node. If right of, explore subtree (2) recursively. If left of, explore subtree (1) recursively. Always explore subtree (3) using the query point projected into the hyperplane, and then, from the set of boxes returned, prune out the boxes not containing the query point.

During a query, we will only explore paths in the ternary tree which include $\leq d-2$ type-3 parent-child edges. The number of such paths is $\leq \binom{2d\ln(N+2d-\kappa)}{d-2}$.

Since the work at a leaf node is $O(d + \kappa + \log(\kappa N))$, and the work during "prunings" can be made negligible, the time per query is $\leq \binom{2d\ln(N+2d-\kappa)}{d-2} O(d+\kappa+\log N)$. $\square$

Remark. Note, if d is fixed, our results are the same, up to constant factors, as the previous solution by [65]. However, the asymptotic behavior of our worst case query time when d becomes large is a factor roughly $(2e\ln 2)^d$ slower than [65]'s. On the other hand, in the

[44] Actually they said $O(\lg^d N)$, but noted that $O(\lg^{d-1} N)$ was possible by using an optimal 2D point location structure [105] [157] at the bottom of their recursion.

best case in which our separators, by luck, happen to yield 50-50 splits, our typical query time will be roughly $d!$ times faster than [65]'s. (No "luck" is required if the boxes are very small and far apart, since in this case 50-50 splitting is normal. And recall that we may use the "best" hyperplane as in lemma 38 every time – because sorted list "unmerging" after a split is possible in $O(N)$ time – with only a constant factor increase in the build time.) Hence, it is not clear which method is better in practice.

7.2 Point location in κ-thick d-objects with bounded aspect ratios

Given: A set $\mathcal{S}$ of κ-thick objects with aspect ratios (definition 4) bounded by B.
Task: To preprocess them to create a data structure to support "point location queries."
These queries are: "Name the objects containing the point $\vec{x}$."

Theorem 97 *This problem may be solved with a data structure that may be stored in $d^{O(d)}N$ memory locations. The expected time required to build it (the build algorithm is randomized) is $d^{O(d)}N \log N$. Point location queries require $O(\kappa B^d O(d)^d C + \log N)$ steps. Here C denotes the time to determine if a point is in one given object.*

The data structure may be "dynamized" to allow also the insertion and deletion of objects. In this case, the (amortized) time bound for a query increases by a factor of $\log N$, the space requirement is affected by only a constant factor, and the insertion and deletion times are $O(1/N)$ times the build time of a static data structure for N objects.

Proof.
If $N \le \kappa B^d O(d)^d$ do not build a data structure – we'll just do point location by brute force.

Otherwise, use the appropriate variant of the separator theorem 39 (theorems 47 and 48 could also be used, in some cases) to decompose the boxes recursively into a binary tree. The two child subtrees of a node correspond to the objects (1) inside of or overlapping, (2) outside of or overlapping the separator.

Because each of the 2 sets is guaranteed to be a constant factor smaller, the tree has depth $O(\log N)$ (for theorem 39; $O(d \log N)$ for theorem 47). The 2 children of a node may in total correspond to more objects than were known to its parent, due to duplication of objects overlapping the separator. However, it is easily verified[45] that the total storage amplification factor remains bounded by a constant independent of N, in fact $d^{O(d)}$. Then because each non-leaf node of the tree has outvalency 2 and requires $O(1)$ storage (2 pointers, a few numbers describing the separator shape) and the leaf nodes store disjoint sets of objects, the total storage required is only a $d^{O(d)}$ factor larger (at most) that the storage required to write down the input objects in the first place.

To locate a point: If the root is childless, determine the set of boxes the point is in by brute force examination of all the boxes (of which there will be $\kappa B^d O(d)^d$).

Otherwise, see if the point is inside, outside, or on the separator boundary for the root node. If inside or on, explore subtree (1) recursively. Otherwise, explore subtree (2) recursively.

The total query time will be $O(\kappa B^d O(d)^d C + \log N)$ (with theorem 39; with theorem 47, where the objects must be balls, it is $O(\kappa 2^{O(d)} C + d \log N)$).

To build the data structure, employ the randomized algorithmic version of theorem 39 in §4.7.2. The expected time to build the data structure is then $d^{O(d)} N \log N$.

Finally, the dynamization results are a standard application of the ideas in [142]. In short, one maintains $O(\log N)$ different static structures, each a factor ≈ 2 larger than the preceding one. Queries are performed by searching all $\log N$ structures. Insertions are performed by destroying and rebuilding the smallest structure – although if the resulting structure would be too large, then a "carry" is performed, and a destruction and rebuild of the next larger structure is required. Deletions are done "lazily" by simply marking deleted objects as "nonexistent." Once 50% of the objects become nonexistent, a global rebuild is performed to get rid of them. □

7.3 Finding all intersections among N κ-thick d-objects with bounded aspect ratios

Given: A set $\mathcal{S}$ of κ-thick objects with aspect ratios (definition 4) bounded by B.
Task: To find all intersection relationships among pairs of objects.

Theorem 98 *The problem above may be solved in time $(\log N + \kappa B^d T) d^{O(d)} N$ while consuming $d^{O(d)} N$ memory locations. Here it takes time T to determine whether two given objects intersect.*

Proof. In the point-location tree structure of the previous section, objects can only intersect other objects if they are in the same leaf of the tree. Since each tree leaf corresponds to a set of objects of cardinality $\kappa B^d O(d)^d$ at most, and the total number of objects in leaves, counting all duplicated objects multiply, is $O(N)$, the total runtime (assuming the tree has already been constructed)[46] is $\kappa O(d)^d NT$. □

7.4 Finding the optimal traveling salesman tour of N sites in d-space

Given: N sites in d-space.
Task: To find their optimal traveling salesman tour.

[45] Since $(1 + dN^{-1/d})^{d \log N}$ is bounded by $d^{O(d)}$.

[46] Note, because we counted all duplicated objects multiply, we have correctly taken account of the fact that an object could be in more than one leaf of the tree.

Theorem 99 *This*[47] *may be accomplished in time*

$$2^{d^{O(d)}} N^{O(dN^{1-1/d})}. \tag{78}$$

The algorithm will be to guess the separating box of theorem 89 among $\lesssim [(N+1)N]^d$ inequivalent possibilities. Then the optimal tour must cross this box at $O(dN^{1-1/d}) + 2^{O(d)}$ places out of $\leq (N-1)N$ possibilities[48]. Guess them too. Finally, for each such guess, guess also the "boundary conditions"' at the crossing points, that is, how the tour joins the crossing points on each side of the separator surface. For C crossing points, this is $C!$ possible kinds of boundary conditions. (In the plane, there are only $2^{O(C)}$ possibilities [162].)

Finally, solve the two smaller "traveling salesman tour plus boundary conditions" problems inside and on, and outside and on, the separator, recursively. (If $N < d^{O(d)}$, solve the problem by brute force in $2^{O(N)}$ steps.)

Needless to say, all "guessing" must be done exhaustively.

The runtime $T(N)$ obeys the recurrence

$$T(N) \leq [(N+1)N]^d \cdot N^{O(dN^{1-1/d})+2^{O(d)}} [T(2N/3)+T(N/3)]. \tag{79}$$

The solution is

$$T(N) = 2^{d^{O(d)}} N^{O(dN^{1-1/d})}. \tag{80}$$

The space needs can be kept linear [162]. □

7.5 *Finding the rectilinear Steiner minimal tree of N sites in d-space*

Given: N sites in d-space.
Task: To find their optimal rectilinear Steiner tree.

Theorem 100 *This may be accomplished in*

$$2^{d^{O(d)}} N^{O(d^3 N^{1-1/d})} \tag{81}$$

steps.

A "rectilinear Steiner tree" (RST) is a collection of line segments parallel to the coordinate axes, which interconnect N given sites. The rectilinear Steiner *minimal* tree (RSMT) is the shortest such network.

Some properties of RSMTs:

1. They are trees. (If a cycle existed, one could shorten the RSMT by removing the longest edge in the cycle, contradicting minimality.)
2. The angle formed by 2 coterminous line segments is $\geq 90°$.
3. Each vertex has valency $\leq 2d$.
4. The total number of "Steiner points" (points of valency ≥ 3 which are not sites) is $\leq N-1$.
5. The total number of line segments is $< 2dN$.

A RST "topology" is a specification of the tree structure of the RST, indicating where each site lies, but ignoring the locations of 2-valent non-site corners on paths, and ignoring geometric information (such as lengths, coordinates) generally.

A fundamental theorem about RSMTs was shown by Hanan [92] when $d = 2$, but it seems to be only a "folk theorem" for general d (meaning: we were unable to find a proof in previous literature!). This is

Theorem 101 *The shortest RST with a given "topology" WLOG is one in which all line segments are edges (1-flats) in the arrangement of dN hyperplanes orthogonal to the coordinate axes going through the sites.*

Proof. In the shortest RST, the coordinates of a Steiner point are determined by the coordinates of its $\leq 2d$ neighbors as follows: the ith coordinate is the median of the ith coordinates of all the neighbors. (The "median" of an even number of values is non-unique, and any value lying at or between the bimedians may be used without affecting the length.)

Hence, the complete system of equations for the optimal coordinates of all Steiner points in a given RST topology are a system of linear equations, in which every equation is of the form $a = b$. Hence, every Steiner point's ith coordinate is the same, WLOG, as some site's ith coordinate. □

This theorem implies that WLOG there are only N^d possible locations for Steiner points (lying on an $N \times N \times \ldots \times N$ "grid") and only $\binom{N}{d-1} d! < dN^{d-1}$ possible lines that RSMT line segments could be subsegments of. Hence we've reduced the problem of finding the RSMT to a finite search problem.

Our algorithm will be to guess the separating d-box of theorem 82 from among $\lesssim N^{O(d)}$ inequivalent possibilities. In 2D, it is probably best to use a d-box rotated 45° due to theorem 81, but in high dimensions we just use a box aligned with the coordinate directions.

Then the optimal RSMT must cross the separator boundary at $O(d^2 N^{1-1/d})$ places out of $< 2dN^{d-1}$ possibilities. Guess them too. Finally, for each such guess, guess also the "boundary conditions"' at the crossing points, that is, how the RSMT partitions the joinings of the crossing points induced by the RSMT on each side of the separator surface.

For C crossing points, this is certainly $< 4^C C!$ possible kinds of boundary conditions. (In the plane, there are only $2^{O(C)}$ possibilities [162].)

Finally, solve the two smaller "RSMT plus boundary conditions" problems inside and on, and outside and on, the separator, recursively. (If $N < d^{O(d)}$, solve the problem by brute force in $2^{O(N)}$ steps.)

[47] We assume we are using a real RAM [147] or similar model of computation, so that difficulties arising from having to determine which of two sums of square roots of integers is greater [76] may be ignored.

[48] Defined by the $(N-1)N/2$ site pairs, and using the fact that a line segment can cross a box boundary in ≤ 2 places because boxes are convex. Actually we could eliminate the 2-time crossings from consideration and hence reduce the "$(N-1)N$" to "$(N-1)N/2$."

Needless to say, all "guessing" must be done exhaustively.

The runtime $T(N)$ obeys the recurrence

$$T(N) \leq N^{O(d)}(8dN^{d-1})^{d^2 N^{1-1/d}}[T(2N/3) + T(N/3)]. \tag{82}$$

The solution is (we assume $N > d$)

$$T(N) = 2^{d^{O(d)}} N^{O(d^3 N^{1-1/d})} \tag{83}$$

The space needs are linear. $\square$

7.6 Euclidean Steiner minimal trees in the plane

Can the exact SMT on N sites in the plane be found in $N^{O(\sqrt{N})}$, or anyhow $2^{o(N)}$, time?[49]

Warme [179] and Winter and Zachariasen [181] concentrated on the practical – as opposed to theoretical – problem of how to find 2D SMTs, and advocated a 2-phase strategy.

1. Generate a superset of all the possible "full Steiner trees[50]" (FSTs) that the SMT could be made of.
2. In phase 2, one finds the best way to combine (an appropriate subset of) the FSTs from phase 1, to get the shortest possible SMT.

The most naive possible strategy for phase 1 would simply be to consider all the 2^N possible subsets of the N sites, and all the possible ordered binary tree topologies on each subset, and then find the best coordinates for the Steiner points for each one. However, the vast majority of the resulting FSTs would obviously be impossible, and could be ruled out by simple geometrical tests such as the requirement that no two edges cross, or the "empty lune condition" mentioned in the proof of theorem 72. Winter and Zachariasen [181] worked long and hard to devise an FST generation algorithm that incorporated many such geometrical tests. Empirically, for N *random* sites in a unit square, $N \leq 150$, their program seems to run in $\approx 0.02N^{2.2}$ seconds and generates $.75N + .001N^2$ FSTs. Furthermore, FSTs involving more than 6 sites seem to be extremely rare.

But, it seems possible to devise *artificial* N-site sets which will cause [181]'s FST generator – or anything like it – to generate c^N FSTs for some $c > 1$. Of course we are leaving this intentionally vague (what does "anything like it" mean?), but here is the vague idea. All we need is a single k-point set which has ≥ 2 "valid" FSTs sufficiently nearby in angle. We then eliminate one of the k sites ("amputating an arm" of the SMT) and replace it with a "hand" which is a sufficiently tiny scaled copy of the original set (also with a site eliminated). The result is a $2k - 2$ site set with ≥ 4 valid FSTs. Continuing on with similar surgical operations using smaller and smaller scaled copies, m in all, we get sets with $k+(k-2)m$ sites with $\geq 2^m$ FSTs.

Warme [179] recognized the combining problem in phase 2 to be one of finding the "minimum spanning tree in a hypergraph" and showed how to solve it via "branch and cut" methods from integer programming. The resulting 2 phase algorithm was, empirically, astonishingly effective – allowing the determination of the exact SMT for a 2000-site set [179]! However, because the MST in hypergraph problem is NP-complete[51] it is unlikely that such good performance persists forever, or in hard cases.

We thus have the paradox that the best known algorithm – empirically – for solving SMT and RSMT in the plane, would appear to have a worst case runtime behavior growing at least doubly exponentially, like A^{B^N} for some $A, B > 1$; *and* this algorithm is capable of solving real-world problems with $N = 2000$... thus proving the inadequacy of theoretical computer science.

We will now partially ameliorate this embarrassing situation by showing that

1. Phase 2 may be accomplished in $M^{O(\sqrt{N})}$ steps, where M is the number of edges in the FSTs from phase 1. In other words, the MST in a hypergraph problem is soluble in subexponential time, if the hypergraph arises from FSTs from N sites in the euclidean plane.
2. If phase 1 is only permitted to generate FSTs of at most k sites, $2 \leq k \leq N$, then it (even using brute force instead of [181]) is easily implemented to run in $O(N)^k$ steps.
3. The combination of phase 2 and truncated phase 1 as above, will therefore run in $N^{O(k\sqrt{N})}$ steps. If, e.g., $k = O(N^{0.49})$, this is subexponential time. The SMT it finds will be optimal if the true SMT happens to have no FSTs with more than k sites. Otherwise, although it will be non-optimal, still [24] shows that its length will be at most $1 + 1/\lg(k/2)$ times longer than the optimal SMT[52]. (Cf. §8.3.)
4. Also, by a different algorithm running in $(N^{1+p}d^d)^{O(d^{5/2}N^{1-1/d})}$ time, we can find an approximate SMT of N sites in d-space, which is guaranteed to be at most $1+N^{-p}$ times longer than the optimal SMT.

Item 2 is easy.

Item 1 is by a "guess the separator" algorithm similar to the ones in §7.4 and §7.5. We know from theorem 68 that the SMT has a separating rectangle with at most $2N/3$ sites on each side, and crossing at most $(4 \cdot 3^{1/4} + o(1))\sqrt{N}$ of the SMT edges. Hence if we knew

[49] A positive solution to this problem – albeit dependent on a conjecture – was given in [164].

[50] An N-site SMT is "full" if it has exactly $N - 2$ Steiner points. All SMTs are unions of FSTs on site subsets.

[51] As Warme [179] mentioned, giving a proof he ascribed to M.Queyranne. Also, NP-completeness follows because in the constructions of [76] [77] showing the NP-hardness of computing the SMT or RSMT of sites in the plane, all FSTs are on ≤ 4 sites. Hence "phase 1" is polynomial time, and only "phase 2" (the MST in hypergraph problem) can be NP-hard.

[52] We would conjecture $1 + O(k^{-p})$ for some constant $p > 0$.

this rectangle, we could then guess the places where the SMT crosses it from the $\le 2M$ possible candidate locations. We could then also guess the "boundary conditions" specifying the topologies of the joinings of these crossing points. If we knew all this, we could then solve the two resulting subproblems. Because there are only $(2M)^4$ possible inequivalent rectangles at most, we can do our "guessing" exhaustively. Letting $T(N)$ denote the time required to solve our problem on N sites, and performing all guessing exhaustively, we get a time recurrence

$$T(N) \le (2M)^4 \binom{M}{O(\sqrt{N})} 2^{O(\sqrt{N})}[T(N/3) + T(2N/3)] \tag{84}$$

which solves to $T(N) = M^{O(\sqrt{N})}$, assuming WLOG that $M \ge N \ge 4$.

Of course, one might quibble that although our separator theorem 68 holds for SMTs, it does *not* necessarily hold for the minimal trees T_k that are unions of SMTs on $(\le k)$-site subsets. The answer to that quibble is that T_k still necessarily obeys the empty lune property. Hence the proof of theorem 72 still goes through (appropriately modified) and we still get a separator theorem analogous to theorem 73 for T_k for any $k \ge 2$.

Finally, we must prove item 4.

Draw the complete graph on all the N sites, then define a circumball for each edge in this graph, then fill each circumball, expanded about its center by a factor of $d^{O(d)}N^p$, with a grid of points, with grid spacing $N^{-p}d^{-O(d)}L$, where L is the length of the corresponding edge of the complete graph. The result is a set of points of cardinality

$$N^{2dp+2}d^{O(d^2)}. \tag{85}$$

Finally, consider the complete graph G drawn on these points (as well as the original N sites), which has

$$N^{4dp+4}d^{O(d^2)} \tag{86}$$

line segment edges.

It is shown in section 4 of [149] that G is a "$(1+N^{-p})$-banyan;" that is, some interconnecting network lying entirely inside G exists, which is $\le 1 + N^{-p}$ times longer than the SMT.

Now, apply our usual "guess the separating sphere" trick to find a good approximate SMT; everything is as before, except that the $O(d^{-1/2}N^{1-1/d})$ points where the SMT edges cross the separator are going to be selected from the points where G crosses it. There are $(N^{1+p}d^d)^{O(d^2)}$ inequivalent spheres. After guessing the separating d-sphere, and these crossing points, and the "boundary conditions," we solve the two smaller subproblems recursively. The runtime recurrence is

$$T(N) \le N^{O(d^2)}(N^{1+p}d^d)^{O(d^{3/2}N^{1-1/d})} \times [T((d+1)N/(d+2)) + T(N/(d+2))] \tag{87}$$

which solves (assuming $2 \le d < N$ WLOG) to

$$T(N) = (N^{1+p}d^d)^{O(d^{5/2}N^{1-1/d})}. \tag{88}$$

Of course, one might quibble that our separator theorem 73 held for SMTs, and *not* necessarily for the minimal trees that lie inside our "banyan." The answer to that quibble is these trees are still the MSTs of their vertices, and hence our theorem 73 does apply.

Theorem 102 *Given N sites in d-space, $2 \le d < N$, and any fixed real $p > 0$. An algorithm running in time $T(N)$, defined in (EQ 88), exists to find a Steiner tree at most $1 + N^{-p}$ times longer than the Steiner minimal tree.*

7.7 *Approximate obstacle avoiding shortest paths in the plane*

Given: A set $\mathcal{P}$ of M convex polygonal obstacles in the plane. We will suppose them κ-thick (note: this allows us to make nonconvex obstacles) and they each have DW aspect ratio $\le B$, and they have the property that the points where the boundaries of a given pair of them intersect, or the points where a given object's boundary intersects a line, may be computed in $O(1)$ time. (This computational property is true for, e.g., triangular and rectangular obstacles.)
Task: To preprocess these polygons to create a data structure to support "approximate shortest path queries."
These queries are: Given two arbitrary points s and t, find an approximately shortest path (avoiding the obstacles) between s and t (or report that there is no such path) and/or just report the length of such a path.

Theorem 103 *Let $\epsilon > 0$. This task may be solved in preprocessing time $O(B\kappa^{1/2}\epsilon^{-1}N^{3/2}\log N)$ with storage $O(B^2\kappa\epsilon^{-2}N\log N)$ (this must be multiplied by $\min(\epsilon^{-1/2}, \log N)$ for L_2 metric) and query time $O(B^4\kappa^2\epsilon^{-2} + \log N + P)$ (the $\log N$ term may be eliminated if s and t are obstacle vertices) and approximation factor $3+\epsilon$, where $N = O(\kappa BM)$ is the number of boundary segments in the arrangement of the M obstacles, and paths are measured in the L_1 metric. P is the size of the output, i.e. $P = O(1)$ if only the distance is reported, but if the entire path is reported, $P = O(N)$ is the number of segments in that path. In the path reporting case the storage requirement grows to $S(N) = O(B\kappa^{1/2}\epsilon^{-1}N^{3/2})$.*

Proof. We'll concentrate on demonstrating this when the paths are to be found in the L_1 metric (i.e., the path is required to be made of vertical and horizontal line segments, or curve segments along an obstacle boundary measured in the L_1 norm) where only the path length is to be reported. Side notes will show how to alter things to make it work in the L_2 metric, and the alterations to allow path reporting as well as distance reporting are straightforward. The fact that $N = O(\kappa BM)$ arises by realizing that an obstacle can intersect at most $O(\kappa B)$ obstacles of larger area.

We first find a 45° rotated square $\diamondsuit$ which separates the convex obstacles, and which $O(B\sqrt{\kappa N})$ cross. (Theorem 39 (a).)

On each of the $O(B\sqrt{\kappa N})$ subsegments of $\partial\diamondsuit$ between obstacles, we place $2\epsilon^{-1}$ uniformly spaced artificial extra vertices. Then we cut away all the parts of all the obstacles outside $\diamondsuit$, and find the L_1 Voronoi diagram [135][53] (in the L_1 obstacle avoiding metric) of 4 artificial sites at (supposing $\diamondsuit$ has corners at $(0, \pm 1)$ and $(\pm 1, 0)$ WLOG) $(\pm 1, \pm 1)$. This L_1 Voronoi diagram (after preprocessing [157] to support fast "point location") enables the determination of the L_1 obstacle avoiding shortest path distance from any point inside $\diamondsuit$ to its boundary in $O(\log N)$ time. In fact with 4 times the work we could determine the shortest path distance, among paths which never go outside $\diamondsuit$, from any point inside $\diamondsuit$ to any desired one of the 4 sides of $\diamondsuit$.

Similarly, by cutting away all the parts of all the obstacles inside $\diamondsuit$, and finding the Voronoi diagram of an artificial site at $(0, 0)$, we then may quickly determine the shortest path distance from any point outside $\diamondsuit$ to its boundary. In fact by use of artificial obstacles inside $\diamondsuit$ and 4 times the work, we could determine the shortest path distance from any point outside $\diamondsuit$ to any desired one of the 4 sides of $\diamondsuit$, among paths never going inside[54] $\diamondsuit$.

Note: if we wish to use the L_2 metric instead of L_1, then the unpublished paper of Hershberger & Suri [99], giving an $O(N \log N)$ time and space algorithm to find L_2 voronoi diagrams in obstacle avoiding metrics in the plane, must be used in place of [135]. This paper works via a "wavefront propagation" paradigm and must be initialized with a wavefront shaped like the boundary of $\diamondsuit$.

Recursively proceed to separate the objects overlapping (or outside) $\diamondsuit$ recursively (the recursions stop once there are $O(B^2\kappa)$ objects left), creating a well balanced "separator tree."

To find the approximate shortest path distance from s to t, we begin by locating s and t with respect to the 45° rotated squares defining the nodes of this separator tree, and then by finding the "youngest common ancestor" of s and t (all of which requires time $O(\log N)$) we may find a separator square "$\diamondsuit$" separating s and t. (If s and t are vertices of the arrangement induced by the obstacles, then this location task may be performed in $O(1)$ steps by lookup in a precomputed table, and the deepest ancestor identification is also possible in $O(1)$ time [158].)

Assuming s and t are respectively inside and outside $\diamondsuit$, we shall use, as the approximately shortest path from s to t, the shortest path from s to $\partial\diamondsuit$ (say it hits $\partial\diamondsuit$ at a point P_s), the shortest path from t to $\partial\diamondsuit$, (say it hits at P_t) and the shortest path from P_s to P_t. The union of these 3 paths is easily seen to be no more than 3 times as long as the shortest path from s to t. But actually, instead of using the true shortest path from P_s to P_t, we shall use an $(1 + \epsilon/2)$ approximately shortest path. Specifically:

1. If P_s and P_t are on the same segment of the same side of $\partial\diamondsuit$ (i.e. with no obstacles between) then use the direct line segment P_sP_t, which is the true shortest path.

2. If P_s and P_t have obstacles between, then walk from P_s along $\partial\diamondsuit$ toward P_t until you hit an obstacle or one of the $2\epsilon^{-1}$ artificial vertices on the current segment (call this point ${P_s}'$). Similarly define ${P_t}'$. Then retrieve the precomputed true shortest path from ${P_s}'$ to ${P_t}'$.

3. The above two ideas could generate a factor of 2 too long an L_1 path from P_s to P_t in the very exceptional case where P_s and P_t are on adjacent but different sides of $\diamondsuit$ with no artificial vertices nor obstacles between. But even then it seems that we may redefine P_s and P_t to force them to lie on the same side (among the 4 sides of) $\diamondsuit$ (since we may try all 4 sides and take the shortest of the 4 final paths we wind up with, in addition to trying the preceding two ideas), and thus assure $(1+\epsilon/2)$ approximation.

We have left unmentioned the possibility that both s and t are inside (or both outside) $\diamondsuit$, which can only happen if they are both in the same "leaf" of the separator tree. But in this case an approximately shortest st path may be found by brute force in time $O(B^4\kappa^2/\epsilon^2)$ with the aid of the precomputed distances among the $O(B^2\kappa/\epsilon)$ points on the boundary of the separating square defining that leaf.

The storage requirement in our algorithm obeys a recurrence like $S(N) = B^2\kappa\epsilon^{-2}N + S(2N/3) + S(N/3)$ with solution $S(N) = O(B^2\kappa\epsilon^{-2}N \log N)$ with the dominant cost being the tables of precomputed shortest path distances among the $O(B\kappa^{1/2}\epsilon^{-1}\sqrt{N})$ points on the boundary of the separator squares. Note, the costs to store the L_1 Voronoi diagrams are $O(N)$; and note the essential fact that since we are only considering shortest paths to $\partial\diamondsuit$, the objects on the other side of $\diamondsuit$ are irrelevant and may be discarded when computing the Voronoi diagram. The preprocessing time in our algorithm obeys a recurrence like $T(N) = B\kappa^{1/2}\epsilon^{-1}N^{3/2}\log N + T(2N/3) + T(N/3)$ with solution $T(N) = O(B\kappa^{1/2}\epsilon^{-1}N^{3/2}\log N)$. Here the dominant cost is computing the exact shortest path distances to put in the tables. Incidentally, all shortest path distances among U of the V vertices of an E-edge graph with positive real edge weights may be found in $UO(E + V \log V)$ time and stored in $O(U^2)$ space [74], and the paths themselves may be stored in $O(UV)$ space. So we may compute a vertical and horizontal "visibility graph" with $O(N)$ edges and vertices and use it when computing these tables (or proceed directly by using the algorithm of [135]). In the L_2 case we could proceed directly by using [99], or we could use

[53] Also perhaps relevant are [136] [173]. Mitchell's algorithms in theorems 2 and 3 of [135], required $O(N \log^2 N)$ time and $O(N \log N)$ space. However, Mitchell in a private communication has informed us that he has unpublished improvements of both of these bounds by $\log N$ factors.

[54] This restriction may be enforced by placing $O(B\sqrt{\kappa N})$ artificial barriers into the problem used to define the Voronoi diagram.

a $O(\epsilon^{-1/2})$-directional visibility graph with $O(N)$ vertices and $O(N\epsilon^{-1/2})$ edges (this approximation of the L_2 metric by a metric based on a regular $O(\epsilon^{-1/2})$-gon instead of a circular norm, causes a path length increase by a factor of $1+\epsilon$). In the case where paths as well as distances need to be reported, these visibility graph approaches incur a space usage recurrence of the form $S(N) = B\kappa^{1/2}\epsilon^{-1}N^{3/2} + S(2N/3) + S(N/3)$ with solution $S(N) = O(B\kappa^{1/2}\epsilon^{-1}N^{3/2})$. □

Remark. We claim without proof that one may reduce the length approximation factor to $1+\epsilon$, provided that the storage bound is increased to $O(B\epsilon^{-1}\kappa^{1/2}N^{3/2} + B^2\kappa\epsilon^{-2}N\log N)$ and the query time to $O(B^4\kappa^2\epsilon^{-2} + \log N + B\epsilon^{-1}\kappa^{1/2}N^{1/2} + P)$.

It seems likely that our separator theorem may not have been needed here, because the plain Lipton-Tarjan theorem could have been made to suffice. It would be more complicated, though. Another virtue of our approach is its possible use in higher dimensions in the future.

7.8 Coloring, independent sets, and counting problems

For optimally coloring, or finding maximum independent sets in, or finding maximum cliques in (or counting cliques, or maximal cliques, in), intersection graphs of d-objects with bounded aspect ratio and bounded thickness in *fixed* dimension d, $2^{O(N^{1-1/d})}$ time suffices [150]. Counting k-colorings may be accomplished in $(k-1)^{O(N^{1-1/d})}$ time.

7.9 Gaussian elimination for systems with the graph structure of an intersection graph of d-objects of bounded aspect ratio and thickness

In a V vertex graph family with a $(\alpha V, \beta V^p)$ separator theorem (definition 11), one may perform "Gaussian elimination" to solve a system of linear equations in time $T(V)$, where $T(V)$ obeys the recurrence

$$T(V) \le T(\alpha V) + T((1-\alpha)V) + (\beta V^p)^\omega. \qquad (89)$$

Here ω, $2 < \omega \le 3$, is the exponent for the runtime of dense matrix multiplication. If $\omega p > 1$, this solves to

$$T(V) = O(\frac{\beta^\omega}{1-\alpha^{p\omega}-(1-\alpha)^{p\omega}} V^{p\omega}). \qquad (90)$$

If we use $\omega \le 2.376$ [53] and $p = 1 - 1/d$, we see that we have subquadratic time when $d \le 6$. This idea dates to [127].

In contrast, the "conjugate gradient method" [86] may be used to solve any system of V linear equations, where there are E nonzero terms in the matrix, in $O(VE)$ arithmetic operations. For a sparse graph – with $E = O(V)$ – this is quadratic time.

7.10 Universal graphs

The following results[55] are typical ones about "universal graphs."

1. For each $V \ge 1$ there exists a graph with V vertices and $O(V^{3/2})$ edges containing all planar V-vertex graphs [10].
2. If η is fixed, then there exists a graph with $O_\eta(V)$ vertices and $O_\eta(V\log V)$ edges containing every V-vertex planar graph with maximum valence η [15].
3. For each $V \ge 1$ there exists a graph with V vertices and $O(V\log V)$ edges containing all V-vertex trees [47] and indeed all V-vertex outerplanar graphs [15]. This is optimal up to a constant factor [47].
4. If η is fixed, then there exists [15] a graph with V vertices and maximum valence $F(\eta)$, containing every V-vertex planar graph with maximum valence η.

Many of these results arose from separator theorems.

Theorem 104 *There exists a V-vertex graph with $O(d\kappa^{1/d}BV^{2-1/d}) + [O(d)^d\kappa B^d]^2$ edges containing all intersection graphs of V convex d-objects with aspect ratio bounded by B and thickness κ.*

Proof sketch. The graph is constructed recursively by placing $O(d\kappa^{1/d}BV^{1-1/d})$ vertices in the "middle" and joining each of these middle vertices to every vertex. Then two disjoint sets of $V/2$ vertices each are chosen and one recursively constructs universal graphs on these two subsets. The fact that this graph is universal arises from the fact that our intersection graphs have 1/2-1/2 separators of size $O(d\kappa^{1/d}BV^{1-1/d})$, which arises from our 1/3-2/3 geometric separator theorem as in the discussion near (EQ 8). The recursions stop, and one uses a complete graph, when $V < O(d)^d\kappa B^d$. □

Theorem 105 *There exists a universal V-vertex graph with $O(d\kappa^{1/d}BV^{2-2/d})$ edges containing all bounded valence intersection graphs of V convex d-objects with aspect ratio bounded by B and thickness bounded by κ.*

Proof. Follows from our separator theorem 39 (a) and [15]. □

8 Open problems

8.1 Graphs of genus g

It should be possible to prove a $K_{h,c}\sqrt{V}$ separator theorem for V-vertex graphs embeddable in 2-dimensional manifolds with h handles and c crosscaps, by geometric

[55] Bondy [21] posed as an open problem: "How many edges are required in a V-vertex graph containing every n-cycle with $3 \le n \le V$?" He stated without proof that $V + \lg(V-1) + \log^* V + O(1)$ edges suffice.

means. This may allow either better constants, separators of simpler structure, or better understanding.

This is because essentially any graph embeddable on any "orbifold" is realizable as a contact graph of an infinite periodic arrangement of interior-disjoint discs. Here "periodic" means symmetric under an infinite symmetry group corresponding to the orbifold; this often requires using hyperbolic or spherical plane geometries instead of the usual Euclidean one[56]. In other words, just as patterns drawn on tori may be "unwrapped" to get infinite periodic patterns in the plane, general orbifolds may also be unwrapped.

The main obstacle seems to lie in bounding the aspect ratio of the fundamental "unit cell" of the infinite periodic arrangement.

8.2 "Squared" projective plane?

In §5.2 we observed that rectangles, cylinders, Moebius strips, toruses, and Klein bottles could be "squared," e.g. tiled by unequal squares. Can any interesting examples of "squared" projective planes (see figure 7) be produced? [57] Related problems are discussed in [75].

8.3 k-Steiner ratio

The "k-Steiner ratio" is the supremal value of the ratio of the length of the shortest interconnecting network that is a union of Steiner minimal trees on ($\leq k$)-element subsets, to the length of the Steiner minimal tree, for some point set in some metric space. (Cf. §7.6.)

In the case when the metric space is the Euclidean plane, we conjecture that the k-Steiner ratio is $1+O(k^{-p})$ for some fixed constant $p>0$; but the best known upper bound at present [24] is $1+O(1/\log k)$.

8.4 Hadwiger hypothesis and related covering problems

The "Hadwiger hypothesis" conjectures that any convex d-body may be covered by 2^d smaller scaled translated copies – and less than 2^d should suffice unless the body is a parallelipiped. We've also conjectured that any convex d-body may be covered by 2 smaller-volume affine versions – and that spheres require the largest volume scaling factor (EQ 28). The question of how many smaller copies one needs if rotations (but not affinities) are allowed, seems comparatively very difficult. See §3 for our survey of all this. Some open questions that ought to be comparatively easy to resolve:

1. What is the minimal s, $0<s<1$, such that any 2D convex body may be covered by 3 rotated and translated versions of itself, each scaled by s?

2. Find the minimal scaling factor s for covering a unit L_p ball in $\mathbf{R}^d$ by $d+1$ such balls of radius s.

Re the $1/e$ volume splitting theorem 42, another open problem is how badly balanced the split of *surface area* can be.

Conjecture 22, about the existence of a good cut plane for splitting face count, would immediately yield a data structure of size $N^{F(d)}$ for solving the "post-office problem" of locating the closest city (among N cities) to a query point, and the polytope containment problem of deciding whether is query point is inside some fixed d-polytope (with N faces), in $O_d(\log N)$ time per query. This would not be as good as [51], although still interesting because of its simplicity and the prospect of typical performance much better than the worst case bounds.

8.5 Practicality of algorithms; more applications

We've presented new algorithms with either more simplicity, or better asymptotic runtime bounds (or both), for finding rectilinear steiner minimal trees and optimal traveling salesman tours for N sites in d-space. Can these algorithms be actually be implemented to yield champion computer programs for these tasks?

Deneen et al. [58] had previously realized that our RSMT separator theorem, and in fact Ganley's conjecture 63, hold (with high probability) under the assumption that the sites are randomly selected from the uniform distribution on $[0,1]^2$. They implemented a $N^{O(\sqrt{N})}$ algorithm depending on that assumption. But, their program performed poorly in comparison to competing programs [156] [73] [179] only featuring $2^{O(N)}$ runtime bounds – or worse. This experience suggests that either our RSMT algorithmic framework is not practical, or that making it practical depends on hybridizing our subexponential framework with the ideas in [156] [73] [179]. Similar remarks apply even more strongly to Euclidean TST [118].

It seems quite likely that our point location (§7.2) and intersection graph (§7.3) algorithms are of practical interest.

[56] These noneuclidean geometries are no problem for us, since, e.g., the circle separator theorems 47 48 remain valid – since the stereographic projection of spherical geometry onto euclidean geometry [and the analogous "conformal disk model" of the hyperbolic plane inside a euclidean disk] is circle preserving.

[57] To enforce nontriviality we would prefer a "simple perfect" example (in the terminology of [27]), i.e. one not containing a squared rectangle or Moebius strip, and in which all the square tiles have unequal sizes. Also, one must define a "square" and a "squaring" as in the discussion of §5.2, remark (vi).

References

[1] M.O.Albertson & J.P.Hutchinson: On the independence ratio of a graph, J. Graph Theory 2 (1978) 1-8.

[2] L.Aleksandrov & H.N.Djidjev: Linear algorithms for partitioning embedded graphs of bounded genus, SIAM Journal of Discrete Mathematics 9 (1996), 129-150.

[3] N.Alon: Eigenvalues and expanders, Combinatorica 6,2 (1986) 83-96.

[4] N.Alon, P.Seymour, R.Thomas: Planar separators, SIAM J. Discr. Math. 7,2 (1994) 184-193.

[5] Noga Alon, Paul Seymour, & Robin Thomas: A separator theorem for nonplanar graphs, J. Amer. Math. Soc. 3,4 (Oct. 1990) 801-808

[6] F. d'Amore & P.G.Franciosa: Separating sets of hyperrectangles, Int'l. J. Comput. Geom. Applic. 3,2 (1993) 155-165; and see missing figure: 3,3 (1993) 345.

[7] S.Arya, G.Das, D.M.Mount, J.S.Salowe, M.Smid: Euclidean spanners: short, thin, and lanky, Symp. on Theory of Computing (STOC) 27 (1995) 489-498.

[8] M.J. Atallah & D.Z. Chen: Applications of a Numbering Scheme for Polygonal Obstacles in the Plane, pp. 1-24 in Springer Lecture Notes in Computer Science #1178, (Proc. International Symposium on Algorithms and Computation (ISAAC) 7 (Osaka Japan 1996).

[9] A.Augusteijn & A.J.W.Duijvestijn: Simple perfect square-cylinders of low order, J. Combinatorial Theory (JCT) B 35 (1983) 333-337.

[10] L. Babai, F. R. K. Chung, P. Erdös, R. L. Graham, J.H.Spencer: On graphs which contain all sparse graphs, Ann. Discr. Math. 12 (1982) 21-26.

[11] Brenda S. Baker: Approximation Algorithms for NP-Complete Problems on Planar Graphs, J. ACM 41,1 (1984) 153-180.

[12] K.Bezdek: The problem of illumination of the boundary of a convex body by affine subspaces, Mathematika 38 (1991) 362-375.

[13] K.Bezdek: Hadwiger-Levi's covering problem revisited, pp. 199-233 in J.Pach (ed.): New trends in discrete and computational geometry, Springer 1993.

[14] S.N.Bhatt & F.T.Leighton: A framework for solving VLSI graph layout problems, J. Comput. and Systems Sci. (JCSS) 28,2 (1984) 300-343.

[15] S.N.Bhatt, F.R.K.Chung, F.T.Leighton, A.Rosenberg: Universal graphs for bounded degree trees and planar graphs, SIAM J. Discr. Math. 2 (1989) 145-155.

[16] M.Blum, R.W.Floyd, V.Pratt, R.L.Rivest, R.E.Tarjan: Time bounds for selection, JCSS 7 (1973) 448-461.

[17] W.J.Blundon: Multiple packing of circles in the plane, J.London Math. Soc. 38 (1963) 176-182.

[18] V.G.Boltyanskii: Problem of illumination of the boundary of a convex body (in Russian), Izvestiya Moldovskog Filiala Akad. Nauk.SSSR 10, 76 (1960) 79-86.

[19] V.G.Boltjansky & I.Ts.Gohberg: Results and problems in combinatorial geometry, Cambridge Univ. Press 1985.

[20] V.G.Boltyanskii & P.S.Soltan: A solution of Hadwiger's problem for zonoids, Combinatorica 12,4 (1992) 381-388.

[21] J. A. Bondy: Pancyclic graphs I, J. Combin. Theory Ser. B 11 (1971) 80-84.

[22] T.Bonnesen & W.Fenchel: Theory of convex bodies, BCS associates, Moscow Idaho USA 1987.

[23] Heather Donnell Booth: Some fast algorithms on graphs and trees, PhD thesis, Computer Science dept. Princeton University, January 1991.

[24] A.Borchers & D-Z. Du: The k-Steiner ratio in graphs, SIAM J. Comput. 26,3 (1997) 857-869.

[25] K. Böröczky: Packing of spheres in spaces of constant curvature, Acta Math. Acad. Sci. Hungaricae 32 (1978) 243-261.

[26] K.Borsuk: Drei Sätze über die n-dimensionalen euklidische Sphäre, Fundamenta Math. 20 (1933) 177-190.

[27] C.J.Bouwkamp: On the dissection of rectangles into squares, I: Koninkl. Nederl. Akad. Wetensch. Proc. Ser. A 49 (1946) 1172-1178; II: 50 (1947) 58-71; III: 50 (1947) 72-78.

[28] H.Brönniman, B.Chazelle, J.Pach: How Hard Is Halfspace Range Searching? D&CG 10 (1993) 143-155.

[29] R. L. Brooks, C. A. B. Smith, A. H. Stone & W. T. Tutte: The dissection of rectangles into squares, Duke Math. J., 7 (1940) 312-340.

[30] L.E.J.Brouwer: Über abbildung von Mannigfaltigkeiten, Math. Annalen 71 (1912) 97-115.

[31] T.N.Bui & C.Jones: Finding good approximate vertex and edge partitions is NP-hard, Info. Proc. Lett. (IPL) 42,3 (1992) 153-159.

[32] P.Buser: A note on the isoperimetric constant, Ann. Sci. Ecole Normale Sup. 15 (1982) 213-230.

[33] G.D.Chakerian & S.K.Stein: Some intersection properties of convex bodies, Proc. Amer. Math. Soc. 18 (1967) 109-112.

[34] G.D.Chakerian: Inequalities for the difference body of a convex body, Proc. Amer. Math. Soc. 18 (1967) 879-884.

[35] G.D.Chakerian & G.T.Sallee: An intersection theorem for sets of constant width, Duke Math. J. 36 (1969) 165-170.

[36] B.M.Chazelle: A theorem on polygon cutting with applications, Symp. on Foundations of Computer Sci. (SFOCS) 23 (1982) 339-349

[37] B.Chazelle: Filtering search, a new approach to query answering, SIAM J. Comput. 15,3 (1986) 703-724.

[38] B.M.Chazelle: A functional approach to data structures and its use in multidimensional searching, SIAM J. Comput. 17,3 (1988) 427-462

[39] B.M.Chazelle: Lower bounds for orthogonal range searching, I: reporting case: J.Assoc.Comput.Mach. 37 (1990) 200-212; II: arithmetic model: 439-463.

[40] B.Chazelle: Triangulating a simple polygon in linear time, Discr. & Comput. Geometry (D&CG) 6,5 (1991) 485-524.

[41] B.Chazelle: Lower bounds for off-line range searching, STOC 27 (1995) 733-740.

[42] J.Cheeger: A lower bound for the smallest eigenvalue of the laplacian, p.195-199 in R.Gunning (ed.): Problems in Analysis, Princeton Univ. Press 1970

[43] J.Chen, S.P.Kanchi, & A.Kanevsky: A note on approximating graph genus, IPL 61 (1997) 317-322.

[44] N.Chiba, T.Nishizeki, T.Ozawa: A linear algorithm for embedding planar graphs using PQ-trees, JCSS 30 (1985) 54-76.

[45] N.Chiba, T.Nishizeki, N.Saito: Applications of the Lipton and Tarjan planar separator theorem, J.Information Proc. 4,4 (1981) 203-207.

[46] F.R.K.Chung & R.L.Graham: On Steiner trees for bounded point sets, Geometriae Dedicata 11 (1981) 353-361.

[47] F. R. K. Chung and R. L. Graham: On universal graphs for spanning trees, JLMS (ser. 2) 27,2 (1983) 203-211.

[48] F.R.K.Chung: A survey of separator theorems, pp. 17-34 in B.Korte et al. (eds.): Paths, flows, and VLSI layouts, Springer 1990.

[49] F.R.K.Chung: Improved separators for planar graphs, pp. 265-270 in: Graph theory, combinatorics, and applications, (Y.Alavi et al. eds.) J.Wiley 1991.

[50] James A. Clarkson: Uniformly convex spaces, Trans. AMS 40 (1936) 394-414.

[51] Kenneth L. Clarkson: A probabilistic algorithm for the post office problem, STOC 17 (1985) 175-184.

[52] J.H.Conway & N.J.A.Sloane: Sphere packings, lattices, and groups, Springer 1988.

[53] Don Coppersmith & Shmuel Winograd: Matrix Multiplication via Arithmetic Progressions, J. Symbolic Computation 9,3 (1990) 251-280.

[54] H.S.M. Coxeter: Loxodromic sequences of tangent spheres, Aequationes Math. 1 (1968) 104-121.

[55] H.S.M.Coxeter: Regular polytopes, 3d ed., Dover Publications New York 1973.

[56] H.S.M.Coxeter: A packing of 840 balls of radius $9°0'19''$ on the 3-sphere, pp. 127-137 in Intuitive Geometry (Siofok 1985; eds. K.Böröczky & G.Fejes Toth; CMSJB 48) North-Holland 1987.

[57] G. Das, G. Narasimhan, J. Salowe: A new way to weigh malnourished Euclidean graphs, Symp. on Discr. Algorithms 6 (1995) 215-222.

[58] L. L. Deneen, G. M. Shute, C. D. Thomborson: A probably fast, provably optimal algorithm for rectilinear Steiner trees, Random Structures and Algorithms 5 (1994) 535-557.

[59] H.N.Djidjev: A linear algorithm / separator theorem for graphs of bounded genus, Serdica Bulgaricae Math. Publications 11 (1985) 319-329 / 369-387.

[60] H.N.Djidjev: On the problem of partitioning planar graphs, SIAM J. Algeb. Discr. Meth. 3,2 (1982) 229-240.

[61] Hristo Djidjev: Partitioning Graphs with Costs and Weights on Vertices: Algorithms and Applications, Technical Report TR96-260, Department of Computer Science, Rice University, 1996; to appear Europ. Symp. Algorithms 1997.

[62] V.L. Dol'nikov: Generalized transversals of families of sets in R^n and connections with the Helly and Borsuk theorems, Soviet Math. Dokl. 36, 3 (1988) 519-522.

[63] R.M.Dudley: Central limit theorems for empirical processes, Annals Probab. 6,6, (1978) 899-929.

[64] J.Eckhoff: Helly, Radon, and Caratheodory type theorems, pp. 389-448 in P.M.Gruber & J.M.Wills (eds.): Handbook of convex geometry, North-Holland 1993.

[65] H.Edelsbrunner, G.Haring, D.Hilbert: Rectilinear point location in d dimensions with applications, Computer J. 29,1 (1986) 76-82.

[66] H.Edelsbrunner: Algorithms in combinatorial geometry, Springer 1987.

[67] L. Fejes Toth: Packing spherical caps, Amer. Math. Monthly 56 (1949) 330-331.

[68] L. Fejes Toth: Lagerungen in der Ebene, auf der Kugel, und im Raum, (2nd ed.) Springer 1972 (GMW #65).

[69] L.Few: Multiple packing of spheres, a survey, pages 88-93 in W.Fenchel (ed.): Proc. Colloq. Convexity (Copenhagen 1965) Kobenhavn Univ. Mat. Inst. 1967.

[70] L.Few & P.Kanagasabapathy: The double packing of spheres, J. London Math. Soc. 44 (1969) 141-146.

[71] T.Figiel, J.Lindenstrauss, V.D.Milman: The dimension of almost spherical sections of convex bodies, Acta Math. 139,1-2 (1977) 53-94.

[72] R.W.Floyd & R.Rivest: Expected time bounds for selection, Commun. ACM 18,3 (March 1975) 165-173.

[73] U.Fössmeier & M.Kaufmann: Solving rectilinear Steiner tree problems exactly in theory and practice, Proceedings of the European Symposium on Algorithms 1997.

[74] M.L. Fredman & R.E. Tarjan: Fibonacci Heaps and their uses in improved network optimization algorithms, JACM 34,3 (1987) 596-615.

[75] D. Gale: Mathematical entertainments column, Mathematical Intelligencer 15,1 (1993) 48-52. I. Stewart: Mathematical recreations column, Scientific American, July 1997 and March 1998.

[76] M.R. Garey, R.L. Graham, D.S. Johnson: The complexity of computing Steiner minimal trees, SIAM J. Appl. Math. 32, 4 (1977) 835-859.

[77] M.R.Garey & D.S.Johnson: The rectilinear Steiner problem is NP-complete, SIAM J. Appl. Math. 32, 4 (1977) 826-834.

[78] M.R.Garey, D.S.Johnson, L.J.Stockmeyer: Some simplified NP-complete graph problems, Theor. Comput. Sci. 1 (1976) 237-267.

[79] Naveen Garg, Huzur Saran, & Vijay V. Vazirani: Finding separator cuts in planar graphs within twice the optimal, SFOCS (Symposium on Foundations of Computer Science) 35 (1994) 14-23.

[80] H.Gazit & G.L.Miller: Planar separators and the euclidean norm, SIGAL 90 (Info. Proc. Soc. of Japan; Springer Lecture notes in CS #450) 1990, pp. 338-347.

[81] E.N. Gilbert & H.O.Pollak: Steiner minimal trees, SIAM J. Appl. Math. 16,1 (1968) 1-29.

[82] John R. Gilbert, Joan P. Hutchinson, & Robert Endre Tarjan: A separator theorem for graphs of bounded genus, J. Algorithms, 5,3 (Sept. 1984) 391-407.

[83] John R. Gilbert, Donald J. Rose, Anders Edenbrandt: A Separator Theorem for Chordal Graphs, SIAM J. Algeb. Discr. Methods 5,3 (1984) 306-313.

[84] C.H.Goldberg & D.B.West: Bisection of circle colorings, SIAM J. Algeb. & Discr. Meth. 6 (1985) 93-106.

[85] M.Goldberg: Three infinite families of tetrahedral space fillers, JCT A 16 (1974) 348-354.

[86] G.H.Golub & C.F.Van Loan: Matrix computations, Johns Hopkins University Press 1983.

[87] M.T.Goodrich: Planar separators and parallel polygon triangulation, JCSS 51 (1995) 374-389.

[88] M.Grötschel, L.Lovasz, A.Schrivjer: Geometric algorithms and combinatorial optimization, Springer 1990.

[89] P.Gruber (ed.): Handbook of convex geometry, Elsevier 1993.

[90] Branko Grunbaum, G.C. Shephard: Tilings and patterns, W.H. Freeman, 1989.

[91] L.Guibas, J.Hershberger, D.Leven, M.Sharir, R.Tarjan: Linear time algorithms for visibility and shortest path problems inside triangulated simple polygons, Algorithmica 2 (1987) 209-233.

[92] M.Hanan: On Steiner's problem with rectilinear distance, SIAM Journal on Applied Math. 14 (1966) 255-265.

[93] Olaf Hanner: On the uniform convexity of L^p and ℓ^p, Arkiv för Matematik 3,19 (1955) 239-244.

[94] G.H.Hardy, J.E.Littlewood, G.Polya: Inequalities, Cambridge Univ. Press 1952 (2nd ed.).

[95] R.Hassin & A.Tamir: Efficient algorithms for optimization and selection in series parallel graphs, SIAM J. Algeb. Discr. Methods 7,3 (1986) 379-389.

[96] D.Hausler & E.Welzl: ϵ-Nets and simplex range queries, D&CG 2 (1987) 127-151.

[97] E.Helly: Ueber mengen Konvexer Körper mit gemeinschaftlichen Punkten, Jahresbericht der Deutschen Math. Vereinigung 32 (1923) 175-176.

[98] A.Heppes: Mehrfache gitterförmige Kreislagerunger in der ebene, Acta Math. Sci. Acad. Hungar. 10 (1959) 141-148.

[99] J.Hershberger & S.Suri: Efficient computation of euclidean shortest paths in the plane (abbreviated paper), SFOCS 34 (1993) 508-517. Improved and longer version to appear in SIAM J. Comput. $\approx$ 1998.

[100] W-Y. Hsiang: On the sphere packing problem and the proof of Kepler's conjecture, Internat. J. Math. 4 (1993) 739-831.

[101] Camille Jordan: Sur les assemblages de lignes. Journal für die reine und angewandte Mathematik 70 (1869) 185-190.

[102] J.Kahn & G.Kalai: A counterexample to Borsuk's conjecture, Bull. AMS 29 (1993) 60-62.

[103] A.C.Kang & D.A.Ault: Some properties of the centroid of a free tree, IPL 4,1 (1975) 18-20.

[104] M.Katchalski & T.Lewis: Cutting complexes and Convex sets, Proc. Amer. Math. Soc. 79 (1980) 457-461.

[105] D.Kirkpatrick: Optimal search in planar subdivisions, SIAM J. Comput. 12 (1983) 28-35.

[106] M.S.Klamkin: Solution of problem 85-26, inequality for a simplex, SIAM Rev. 28,4 (1986) 579-580.

Maria Klawe: Limitations on explicit constructions of expanding graphs, SIAM J. Comput. 13,1 (1984) 156-166.

[107] Maria Klawe: Limitations on explicit constructions of expanding graphs, SIAM J. Comput. 13,1 (1984) 156-166.

[108] V.L.Klee: Some new results on smoothness and rotundity in normed linear spaces, Math. Annalen 139 (1959) 51-63.

[109] T.Kloks: Treewidth computations and approximations, Springer 1994.

[110] W.Klingenberg: Lectures on closed geodesics, Springer 1978.

[111] P.Koebe: Kontaktprobleme der konformen Abbildung, Berichte über die verhandlungen der Sächsischen Akademie der Wissenschaften, Liepzig, Mathematische-Physische Klasse 88 (1936) 141-164.

[112] J.Komlós, J.Pach, G.Woeginger: Almost tight bounds for epsilon nets, D&CG 7 (1992) 163-173.

[113] H.Konig & V.D.Milman: On the covering numbers of convex bodies, pp. 82-95 in Springer Lecture Notes in Math. #1267 (1987).

[114] S.Krotoszynski: Covering a plane convex body with 5 smaller homothetical copies, Beiträge Algebra Geom. 25 (1987) 171-176.

[115] M.Lassak: Solution of Hadwiger's covering problem for centrally symmetric convex bodies in E^3, J.London Math.Soc. (ser. 2) 30 (1984) 501-511.

[116] M.Lassak: Covering a plane convex body by 4 homothetical copies with the smallest positive ratio, Geometriae Dedicata 21 (1986) 157-167.

[117] M.Lassak: Covering the boundary of a convex set by tiles, Proc. Amer. Math. Soc. 104 (1988) 269-272.

[118] E.L. Lawler, J.K. Lenstra, A.H.G. Rinnooy Kan, D.B. Shmoys (eds.): The traveling salesman problem, J.Wiley 1985.

[119] F.T.Leighton & S.B.Rao: An approximate max-flow min-cut theorem for uniform multicommodity flow problems with applications to approximation algorithms, SFOCS 29 (1988) 422-431.

[120] Ming Li: A separator theorem for 1-dimensional graphs under linear mapping, IPL 27,1 (1988) 9-11.

[121] Ming Li: Simulating two pushdown stores by one tape in $O(N^{1.5}\sqrt{\log N})$ time, JCSS 37,1 (1988) 101-116.

[122] J.Lindenstrauss: On the modulus of smoothness and divergent series in Banach spaces, Michigan Math J. 10 (1963) 241-252.

[123] J.Lindenstrauss & L.Tzafriri: Classical Banach spaces II, Springer 1979 (EMG #97).

[124] A.Lingas: On partitioning polygons, ACM Sympos. Computational Geom. (1985) 288-295.

[125] R.J.Lipton & R.E.Tarjan: A separator theorem for planar graphs, SIAM J. Appl. Math. 36 (1979) 177-189.

[126] R.J.Lipton & R.E.Tarjan: Applications of a planar separator theorem, SIAM J. Computing 9,3 (1980) 615-627. Also an earlier and in some ways more complete version was in SFOCS 18 (1977) 162-170.

[127] R.J.Lipton, D.J. Rose, R.E. Tarjan: Generalized nested dissection, SIAM J. Numer. Anal. 16 (1979) 346-358.

[128] H.Martini: Some results and problems around zonotopes, pp. 383-418 in Intuitive Geometry (Siofok 1985; eds. K.Böröczky & G.Fejes Toth; CMSJB 48) North-Holland 1987.

[129] J.Matousek, E.Welzl, L.Wernisch: Disrepancy and ϵ-approximations for bounded VC dimension, Combinatorica 1992. Also earlier version was in SFOCS 32 (1991) 424-430.

[130] J. B. M. Melissen: Packing and Coverings with Circles, PhD thesis, 1997 (Universiteit Utrecht). To be published as book by Kluwer.

[131] G.L.Miller: Finding small simple cycle separators for 2-connected planar graphs, JCSS 32 (1986) 265-279.

[132] Gary L. Miller, Shang-Hua Teng, William Thurston, & Stephen A. Vavasis: Separators for sphere-packings and nearest neighbor graphs, J. ACM, 44, 1 (Jan 1997) 1-29.

[133] V.D.Milman: Inegalite de Brunn-Minkowski inverse et applications a la theorie locale des espaces normes [An inverse form of the Brunn-Minkowski inequality with applications to the local theory of normed spaces], C.R.Acad.Sci.Paris I Math. 302,1 (1986) 25-28. MR87f:52018.

[134] John Milnor: Analytic proofs of the 'Hairy ball theorem' and the Brouwer fixed point theorem, Amer.Math.Monthly 85 (1978) 521-524.

[135] J.S.B. Mitchell: L_1 Shortest paths among polygonal obstacles in the plane, Algorithmica 8 (1992) 55-88.

[136] J.S.B. Mitchell: Shortest paths among obstacles in the plane, Internat. J. Comput. Geom. Applic. 6,3 (1996) 309-332.

[137] D.S. Mitrinovic: J.E. Pecaric V. Volenec: Recent advances in geometric inequalities, Kluwer 1989.

[138] Bojan Mohar: A polynomial time circle packing algorithm, Discrete Math. 117 (1993) 257-263.

[139] Bojan Mohar: Embedding graphs in an arbitrary surface in linear time, STOC 28 (1996) 392-397. Later version submitted to SIAM J. Discr. Math. 1997.

[140] D.J.Muder: A new bound on the local density of sphere packings, D&CG 10 (1993) 351-375.

[141] A.Nilli: On Borsuk's problem, Contemp. Math. 178 (1994) 209-210.

[142] M.H.Overmars: The design of dynamic data structures, Springer 1983 (Lecture notes in Computer Sci. #156).

[143] James K. Park & Cynthia A. Phillips: Finding minimum-quotient cuts in planar graphs (extended abstract), STOC 25 (1993) 766-775.

[144] Gilles Pisier: The volume of convex bodies and Banach space geometry, Cambridge University Press 1989.

[145] Gilles Pisier: J.Reine & Angew. Math. 393 (1989) 115-131. MR90a:46038.

[146] S.Plotkin, S.Rao, & W.D.Smith: Shallow Excluded Minors and Improved Graph Decompositions. Symp. on Discrete Algorithms 5 (1994) 462-470.

[147] F.P.Preparata & M.I.Shamos: Computational geometry, an introduction, Springer-Verlag 1988 (corrected & expanded 2nd printing).

S.Plotkin, S.Rao, & W.D.Smith: Shallow Excluded Minors and Improved Graph Decompositions. Symp. on Discrete Algorithms 5 (1994) 462-470.

[148] R.Rado: A theorem on general measure, J.London Math. Soc. 21 (1946) 291-300.

[149] Satish B. Rao & Warren D. Smith: Improved approximation schemes for geometrical graphs via "spanners" and "banyans," NECI Tech. Report, December 1997. A shortened version of this has been accepted by STOC 30 (1998).

[150] S.S.Ravi, Harry B. Hunt III: An Application of the Planar Separator Theorem to Counting Problems, IPL 25,5 (July 1987) 317-322.

[151] G.Ringel & J.W.T.Youngs: Solution of the Heawood map coloring problem, Proc. Nat. Acad. Sci. USA 60 (1968) 438-445.

[152] J.Riordan: Introduction to combinatorial analysis, Wiley 1958.

[153] C.A.Rogers: A note on coverings, Mathematika 4,7 (1957) 1-6.

[154] C.A.Rogers & G.C.Shephard: The difference body of a convex body, Archiv der Math. 8 (1957) 220-233.

[155] C.A.Rogers & C.Zong: Covering convex bodies by translates of convex bodies, Mathematika 44 (1997) 215-218.

[156] J. S. Salowe & D. M. Warme: 35-point rectilinear Steiner minimal trees in a day. Networks 25 (1995) 69-87.

[157] Neil Sarnak, Robert E. Tarjan: Planar point location using persistent search trees, Commun. Assoc. Comput. Mach. 29, 7 (1986) 669-679.

[158] Baruch Schieber & Uzi Vishkin: On finding lowest common ancestors: simplification and parallelization, SIAM J. Comput. 17,6 (1988) 1253-1262.

[159] Oded Schramm: Illuminating sets of constant width, Mathematika 35 (1988) 180-189.

[160] Raimund Seidel: Exact upper bounds for the number of faces in Voronoi diagrams, DIMACS ser. Discr. Math. & Theor. Comput. Sci. 4 (1991) 517-529. AMS 1991.

[161] D.M.Y.Sommerville: An introduction to the geometry of N dimensions, Methuen, London, 1929.

[162] W.D.Smith: Studies in computational geometry motivated by mesh generation, (485 pp.), PhD. thesis, Princeton University applied math dept., 1988. Available from University Microfilms International, order number 9002713.

[163] W.D.Smith: Accurate Circle Configurations and Numerical Conformal Mapping in Polynomial Time, NECI Tech. Report, December 1991. (Available electronically on `http://www.neci.nj.nec.com/homepages/wds/works`

[164] W.D.Smith: How to find Steiner minimal trees in d-space, Algorithmica 7 (1992) 137-177.

[165] P.S.Soltan & V.P.Soltan: On the X-raying of convex bodies (in Russian), Dokl. Akad. Nauk SSSR 286 (1986) 50-53.

[166] A. Soifer: How to cut up a triangle, ?????

[167] Daniel A. Spielman and Shang-Hua Teng: Disk Packings and Planar Separators, 12th Annual ACM Symposium on Computational Geometry (1996) 349-358.

[168] A.H.Stone & J.W.Tukey: Generalized 'sandwich' theorems, Duke Math. J. 9 (1942) 356-359.

[169] R.E.Tarjan: Data structures and network algorithms, SIAM 1983.

[170] A.Temesvari: Die dichteste gitterförmige 9-fache Kreispackung, Rad. Hrvatske Akad. Znan. Umj. Mat. 11 (1994) 95-110.

[171] C.Thomassen: The graph genus problem is NP-complete, J.Algorithms 10 (1989) 568-576.

[172] C.Thomassen: Triangulating a surface with a prescribed graph, JCT B 57 (1993) 196-206.

[173] Y.H. Tsin & Cao-An Wang: Geodesic Voronoi diagrams in the presence of rectilinear barriers, Nordic J. Computing, 3,1 (spring 1996) 1-26.

[174] G.Turan: Lower bounds for synchronous arcs and planar circuits, IPL 30,1 (1989) 37-40.

[175] J.Valdes, R.E.Tarjan, E.Lawler: The recognition of series parallel digraphs, SIAM J. Comput. 11 (1982) 298-313.

[176] D.L.Vandev: A minimal volume ellipsoid around a simplex, C.R.Acad.Bulgare Sci. 45,6 (1992) 37-40.

[177] V.N.Vapnik & A. Ya. Chervonenkis: On the uniform convergence of events to their probabilities, Theory of Probab. & Applications 16 (1971) 264-280.

[178] B.L. van der Waerden: Punkte auf der Kugel. Drei Zusätze, Math. Annalen 123 (1952) 213-222. [Math Reviews 14, 401.]

[179] David M. Warme: A New Exact Algorithm for Rectilinear Steiner Trees, Corrected version (last updated 09/02/97) of paper presented at the International Symposium on Mathematical Programming, Lausanne, Switzerland, August 24-29 1997.

[180] Benulf Weißbach: Invariante Beleuchtung konvexer Körper, Beiträge zur Algebra und Geometrie 37,1 (1996) 9-15.

[181] P. Winter & M. Zachariasen: Euclidean Steiner minimal trees, An improved exact algorithm, Networks 30,3 (1997) 149-166.

[182] R.T.Zivaljevic & Sinisa T. Vrecica: An extension of the ham sandwich theorem, Bull. London Math. Soc. 22,2 (1990) 183-186.