GAFA, Geom. funct. anal.
Special Volume – GAFA2000, 359 – 382
1016-443X/00/S10359-24 $ 1.50+0.20/0

GAFA Geometric And Functional Analysis

DISCRETE AND CONTINUOUS: TWO SIDES OF THE SAME?

László Lovász

Abstract

How deep is the dividing line between discrete and continuous mathematics? Basic structures and methods of both sides of our science are quite different. But on a deeper level, there is a more solid connection than meets the eye.

1 Introduction

There is a rather clear dividing line between discrete and continuous mathematics. Continuous mathematics is classical, well established, with a rich variety of applications. Discrete mathematics grew out of puzzles and then is often identified with specific application areas like computer science or operations research. Basic structures and methods of both sides of our science are quite different (continuous mathematicians use limits; discrete mathematicians use induction).

This state of mathematics at this particular time, the turn of the millennium, is a product of strong intrinsic logic, but also of historical coincidence.

The main external source of mathematical problems is science, in particular physics. The traditional view is that space and time are continuous, and that the laws of nature are described by differential equations. There are, of course, exceptions: chemistry, statistical mechanics, and quantum physics are based on at least a certain degree of discreteness. But these discrete objects (molecules, atoms, elementary particles, Feynman graphs) live in the continuum of space and time. String theory tries to explain these discrete objects as singularities of a higher dimensional continuum.

Accordingly, the mathematics used in most applications is analysis, the real hard core of our science. But especially in newly developing sciences, discrete models are becoming more and more important.

One might observe that there is a finite number of events in any finite domain of space-time. Is there a physical meaning of the rest of a four (or ten) dimensional manifold? Does "the point in time half way between two consecutive interactions of an elementary particle" make sense? Should

the answer to this question be yes or no, discrete features of subatomic world are undeniable. How far would a combinatorial description of the world take us? Or could it happen that the descriptions of the world as a continuum, or as a discrete (but terribly huge) structure, are equivalent, emphasizing two sides of the same reality?

But let me escape from these unanswerable questions and look around elsewhere in science. Biology tries to understand the genetic code: a gigantic task, which is the key to understanding life and, ultimately, ourselves. The genetic code is discrete: simple basic questions like finding matching patterns, or tracing consequences of flipping over substrings, sound more familiar to the graph theorist than to the researcher of differential equations. Questions about the information content, redundancy, or stability of the code may sound too vague to a classical mathematician but a theoretical computer scientist will immediately see at least some tools to formalize them (even if to find the answer may be too difficult at the moment).

Economics is a heavy user of mathematics — and much of its need is not part of the traditional applied mathematics toolbox. Perhaps the most successful tool in economics and operations research is linear programming, which lives on the boundary of discrete and continuous. The applicability of linear programming in these areas, however, depends on conditions of convexity and unlimited divisibility; taking indivisibilities into account (for example, logical decisions, or individual agents) leads to integer programming and other models of combinatorial nature.

Finally, the world of computers is essentially discrete, and it is not surprising that so many discrete mathematicians are working in this science. Electronic communication and computation provides a vast array of well-formulated, difficult, and important mathematical problems on algorithms, data bases, formal languages, cryptography and computer security, VLSI layout, and much more.

In all these areas, the real understanding involves, I believe, a synthesis of the discrete and continuous, and it is an intellectually most challenging goal to develop these mathematical tools.

There are different levels of interaction between discrete and continuous mathematics, and I treat them (I believe) in the order of increasing significance.

1. We often use the finite to approximate the infinite. To discretize a complicated continuous structure has always been a basic method—from the definition of the Riemann integral through triangulating a manifold in

(say) homology theory to numerically solving a partial differential equation on a grid.

It is a slightly more subtle thought that the infinite is often (or perhaps always?) an approximation of the large finite. Continuous structures are often cleaner, more symmetric, and richer than their discrete counterparts (for example, a planar grid has a much smaller degree of symmetry than the whole euclidean plane). It is a natural and powerful method to study discrete structures by "embedding" them in the continuous world.

2. Sometimes, the key step in the proof of a purely "discrete" result is the application of a purely "continuous" theorem, or vice versa. We all have our favorite examples; I describe two in section 2.

3. In some areas of discrete mathematics, key progress has been achieved through the use of more and more sophisticated methods from analysis. This is illustrated in section 3 by describing two powerful methods in discrete optimization. (I could not find any area of "continuous" mathematics where progress would be achieved at a similar scale through the introduction of discrete methods. Perhaps algebraic topology comes closest.)

4. Connections between discrete and continuous may be the subject of mathematical study on their own right. *Numerical analysis* may be thought of this way, but *discrepancy theory* is the best example. In this article, we have to restrict ourselves to discussing two classical results in this blooming field (section 4); we refer to the book of Beck and Chen [BeC], the expository article of Beck and Sós [BeS], and the recent book of Matoušek [M].

5. The most significant level of interaction is when one and the same phenomenon appears in both the continuous and discrete setting. In such cases, intuition and insight gained from considering one of these may be extremely useful in the other. A well-known example is the connection between sequences and analytic functions, provided by the power series expansion. In this case, there is a "dictionary" between combinatorial aspects (recurrences, asymptotics) of the sequence and analytic properties (differential equations, singularities) of its generating function.

In section 5, I discuss in detail the discrete and continuous notion of "Laplacian", connected with a variety of dispersion-type processes. This notion connects topics from the heat equation to Brownian motion to random walks on graphs to linkless embeddings of graphs.

An exciting but not well understood further parallelism is connected with the fact that the iteration of very simple steps results in very complex

structures. In continuous mathematics, this idea comes up in the theory of dynamical systems and numerical analysis. In discrete mathematics, it is in a sense the basis of many algorithms, random number generation, etc. A synthesis of ideas from both sides could bring spectacular developments here.

One might mention many other areas with substantial discrete and continuous components: groups, probability, geometry... Indeed, my starting observation about this division becomes questionable if we think of some of the recent developments in these areas. I believe that further development will make it totally meaningless.

Acknowledgement. I am indebted to several of my colleagues, in particular to Alan Connes and Mike Freedman, for their valuable comments on this paper.

2 Discrete in Continuous and Continuous in Discrete

2.1 Marriage and measures. A striking application of graph theory to measure theory is the construction of the Haar measure on compact topological groups. This application was mentioned by Rota and Harper [RoH], who elaborated upon the idea in the example of the construction of a translation invariant integral for almost periodic functions on an arbitrary group. The proof is also related to Weil's proof of the existence of Haar measure. (There are other, perhaps more significant, applications of matchings to measures that could be mentioned here, for example the theorem of Ornstein [O] on the isomorphism of Bernoulli shifts; cf. also [Du], [LoM], [LoP], [St].)

Here we shall describe the construction of a translation invariant integral for continuous functions on a compact topological group, equivalent to the existence of the Haar measure [LoP]. An *invariant integration* for continuous functions on a compact topological group is a linear functional L defined on the continuous real-valued functions on G, with the following properties:

(a) $L(\alpha f + \beta g) = \alpha L(f) + \beta L(g)$ (linearity)
(b) If $f \geq 0$ then $L(f) \geq 0$ (monotonicity)
(c) If l denotes the identity function, then $L(l) = 1$ (normalization)
(d) If s and t are in G and f and g are two continuous functions such that $g(x) = f(sxt)$ for every x, then $L(g) = L(f)$ (double translation invariance).

Theorem 1. *For every compact topological group there exists an invariant integration for continuous functions.*

Proof. Let f be the function we want to integrate. The idea is to approximate the integral by the average

$$f(A) = \tfrac{1}{|A|} \sum_{a \in A} f(a)\,,$$

where A is a "uniformly dense" finite subset of group G. The question is, how to find an appropriate set A?

The key definition is the following. Let U be an open non-empty subset of G (think of it as "small"). A subset $A \subset G$ is called a *U-net* if $A \cap sUt \neq \emptyset$ for every $s, t \in G$. It follows from the compactness of G that there exists a finite U-net.

Of course, a U-net may be very unevenly distributed over the group, and accordingly, the average over it may not approximate the integral. What we show is that the simple trick of restricting ourselves to U-nets with minimum cardinality (minimum U-nets, for short), we get a sufficiently uniformly dense finite set.

To measure this uniformity we define

$$\delta(U) = \sup \left\{ |f(x)\ f(y)| \;\middle|\; x, y \in sUt \text{ for some } s, t \in G \right\}.$$

The compactness of G implies that if f is continuous then it is also uniformly continuous in the sense that for every $\epsilon > 0$ there exists a non-empty open set U such that $\delta(U) < \epsilon$.

Let A and B be minimum U-nets. The following inequality is the heart of the proof:

$$\big|f(A) - f(B)\big| \leq \delta(U)\,. \tag{1}$$

The combinatorial core of the construction lies in the proof of this inequality. Define a bipartite graph H with bipartition (A, B) by connecting $x \in A$ to $y \in B$ if and only if there exists $s, t \in G$ such that $x, y \in sUt$. We use the Marriage theorem to show that this bipartite graph has a perfect matching. We have to verify two things:

(I) $|A| = |B|$. This is trivial.

(II) Every set $X \subseteq A$ has at least $|X|$ neighbors in B. Indeed, let Y be the set of neighbors of X. We show that $T = Y \cup (A \setminus X)$ is a U-net. Let $s, t \in G$, we show that T intersects sUt. Since A is a U-net, there exists an element $x \in A \cap sUt$. If $x \notin X$ then $x \in T$ and we are done. Otherwise, we use that there exists an element $y \in B \cap sUt$. Then xy is an line of H and so $y \in Y$, and we are done again.

Thus T is a U-net. Since A is a U-net with minimum cardinality, we must have $|T| \geq |A|$, which is equivalent to $|Y| \geq |X|$.

Thus H has a perfect matching $\{a_1b_1, \dots, a_nb_n\}$. Then

$$\begin{aligned}\big|f(A) - f(B)\big| &= \frac{1}{n}\Big|\sum_{i=1}^{n}\big(f(a_i) - f(b_i)\big)\Big| \\ &\leq \frac{1}{n}\sum_{i=1}^{n}\big|f(a_i) - f(b_i)\big| \leq \frac{1}{n}\big(n\delta(U)\big) = \delta(U)\,.\end{aligned}$$

The rest of the proof is rather routine. First, we need to show that averaging over minimum U-nets for different U's gives approximately the same result. More exactly, let A be a minimum U-net and B a minimum V-net, then

$$\big|f(A) - f(B)\big| \leq \delta(U) + \delta(V)\,. \tag{2}$$

Indeed, for every $b \in B$, Ab is also a U-net with minimum cardinality, so by (1),

$$\big|f(A) - f(Ab)\big| \leq \delta(U)\,.$$

Hence

$$\big|f(A) - f(AB)\big| = \Big|f(A) - \frac{1}{|B|}\sum_{b\in B} f(Ab)\Big| \leq \frac{1}{|B|}\sum_{b\in B}\big|f(A) - f(Ab)\big| \leq \delta(U)\,.$$

Similarly,

$$\big|f(B) - f(AB)\big| \leq \delta(V)\,,$$

whence (2) follows.

Now choose a sequence U_n of open sets such that $\delta(U_n) \to 0$, and let A_n be a minimum U_n-net. Then (2) implies that the sequence $f(A_n)$ tends to a limit $L(f)$, which is independent of the choice of U_n and A_n, and so it is well defined. Conditions (a)–(d) are trivial to verify. □

2.2 Disjoint subsets and topology. Now we turn to examples demonstrating the converse direction: applications of continuous methods to a purely combinatorial problems. Our first example is a result where algebraic topology and geometry are the essential tools in the proof of a combinatorial theorem.

Theorem 2. *Let us partition the k-element subsets of an n-element set into $n - 2k + 1$ classes. Then one of the classes contains two disjoint k-subsets.*

This result was conjectured by Kneser [K], and proved in [Lo1]. The proof was simplified by Bárány [B], and we describe his version here.

Proof. We first invoke a geometric construction due to Gale *There exists a set of $2k + d$ vectors on the d-sphere S^d such that every open hemisphere contains at least k of them.* (For $d = 1$, take the vertices of a regular $(2k + 1)$-gon.)

Choosing $d = n - 2k$, we thus get a set S of n points. Suppose that all the k-subsets are partitioned into $d+1 = n-2k+1$ classes $\mathcal{P}_0, \mathcal{P}_1, \ldots, \mathcal{P}_d$. Let A_i be the set of those unit vectors $h \in S^d$ for which the open hemisphere centered at h contains a k-subset of S which belongs to $\mathcal{P}_i$. Clearly the sets $A_0, A_1, \ldots, A_d$ are open, and by the definition of S, they cover the whole sphere.

Now we turn from geometry to topology: by the Borsuk-Ulam theorem, one of the sets A_i contains two antipodal points h and $-h$. Thus the hemispheres about h and $-h$ both contain a k-subset from $\mathcal{P}_i$. Since these hemispheres are disjoint, these two k-subsets are disjoint. □

There are numerous other examples where methods from algebraic topology have been used to prove purely combinatorial statements (see [Bj] for a survey).

3 Optimization: Discrete, Linear, Semidefinite

Our next example illustrates how a whole area of important applied mathematical problems, namely discrete optimization, relies on ideas bringing more and more sophisticated tools from more traditional continuous optimization.

In a typical optimization problem, we are given a set S of *feasible solutions*, and a function $f : S \to \mathbf{R}$, called the *objective function.* The goal is to find the element of S which maximizes (or minimizes) the objective function.

In a traditional optimization problem, S is a decent continuum in a euclidean space, and f is a smooth function. In this case the optimum can be found by considering the equation $\nabla f = 0$, or by an iterative method using some version of steepest descent.

Both of these depend on the continuous structure in a neighborhood of the optimizing point, and therefore fail for a discrete (or combinatorial) optimization problem, when S is finite. This case is trivial from a classical point of view: "in principle" one could evaluate f for all elements in S, and choose the best. But in most cases of interest, S is very large in comparison with the number of data needed to specify the instance of the problem. For

example, S may be the set of spanning trees, or the set of perfect matchings, of a graph; or the set of all possible schedules of all trains in a country; or the set of states of a spin glass. In such cases, we have to use the implicit structure of S, rather than brute force, to find the optimizing element.

In some cases, totally combinatorial methods enable us to solve such problems. For example, let S be the set of all spanning trees of a graph G, and assume that each edge of G has a non-negative "length" associated with it. In this case a spanning tree with minimum length can be found by the *greedy algorithm* (due to Borøuvka and Kruskal): we repeatedly choose the shortest edge that does not form a cycle with the edges chosen previously, until a tree is obtained.

In many (in a sense most) cases, such a direct combinatorial algorithm is not available. A general approach is to embed the set S into a continuum S' of solutions, and also extend the objective function f to a function f' defined on S'. The problem of minimizing f' over S' is called a *relaxation* of the original problem. If we do this right (say, S' is a convex set and f' is a convex function), then the minimum of f' over S' can be found by classical tools (differentiation, steepest descent, etc.). If we are really lucky (or clever), the minimizing element of S' will belong to S, and then we have solved the original discrete problem. (If not, we may still use the solution of the relaxation to obtain a bound on the solution of the original, or to obtain an approximate solution. See later.)

3.1 Polyhedral combinatorics. The first successful realization of this scheme was worked out in the 60's and 70's, where techniques of linear programming were applied to combinatorics. Let us assume that our combinatorial optimization problem can be formulated so that S is a set of $0-1$ vectors and the objective function is linear. The set S may be specified by a variety of logical and other constraints; in most cases, it is quite easy to translate these into linear inequalities:

$$S = \left\{x \in \{0,1\}^n : a_1^{\mathsf{T}} x \le b_1, \dots, a_m^{\mathsf{T}} x \le b_m\right\}. \tag{3}$$

The objective function is

$$f(x) = c^{\mathsf{T}} x = \sum_{i=1}^{n} c_i x_i\,, \tag{4}$$

where the c_i are given real numbers. Such a formulation is typically easy to find.

Thus we have translated our combinatorial optimization problem into a linear program with integrality conditions. It is quite easy to solve this,

if we disregard the integrality conditions; the real game is to find ways to write up these linear programs in such a way that disregarding integrality conditions is justified.

A nice example is matching in bipartite graphs. Suppose that G is a bipartite graph, with node set $\{u_1, \dots, u_n, v_1, \dots, v_n\}$, where every edge connects a node u_i to a node v_j. We want to find a perfect matching, i.e., a set of edges covering each node exactly once.

Suppose that G has a perfect matching M and let

$$x_{ij} = \begin{cases} 1\,, & \text{if } u_i v_j \in M\,, \\ 0\,, & \text{otherwise}\,. \end{cases}$$

Then the defining property of perfect matchings can be expressed as follows:

$$\sum_{i=1}^{n} x_{ij} = 1 \quad \text{for all } j\,, \qquad \sum_{j=1}^{n} x_{ij} = 1 \quad \text{for all } i\,. \tag{5}$$

Conversely, we find a solution of this system of linear equations with every $x_{ij} = 0$ or 1, then we have a perfect matching.

Unfortunately, the solvability of a system of linear equations (even of a single equation) is NP-hard. What we can do is to replace the condition $x_{ij} = 0$ or 1 by the weaker condition

$$x_{ij} \geq 0\,. \tag{6}$$

To solve a linear system like (5) in non-negative real numbers is still not trivial, but doable efficiently (in polynomial time) using linear programming.

We are not done, of course, since if we find that (5)–(6) has a solution, this solution may not be in integers and hence it may not "mean" a perfect matching. There are various ways to conclude, extracting a perfect matching from a fractional solution of (5)–(6). The most elegant is the following. The set of all solutions forms a convex polytope. Now *every vertex of this convex polytope is integral*. (The proof of this fact is amusing, using Cramer's Rule and basic determinant calculations. See, e.g. [LoP].)

So we run a linear programming algorithm to see if (5)-(6) has a solution in real numbers. If not, then the graph has no perfect matching. If yes, most linear programming algorithms automatically give a basic solution, i.e., a vertex. This is an integral

solution of (5)–(6), and hence corresponds to a perfect matching.

Many variations of this idea have been developed both in theory and practice. There are ways to automatically generate new constraints and add them to (3) if these fail; there are more subtle, more efficient methods to generate new constraints in for special problems; there are ways to handle the system (3) even if it gets too large to be written up explicitly.

3.2 Semidefinite optimization. This new technique of producing relaxations of discrete optimization problems makes the problems continuous in a more sophisticated way. We illustrate it by the Maximum Cut problem. Let $G = (V, E)$ be a graph. We want to find a partition $(S, \bar{S})$ of V for which the number M of edges connecting S to $\bar{S}$ is maximum.

This problem is NP-hard, so we cannot hope to find an efficient (polynomial time) algorithm that solves it. Instead, we have to settle for less: we try to find a cut that is close to being optimal.

It is easy to find a cut that picks up half of the edges. Just process the nodes one by one, placing them in the set S or $\bar{S}$, whichever gives the larger number of new edges going between the two sets. Since no cut can have more than all edges, this simple algorithm obtains an approximation of the maximum cut with at most 50% relative error.

It turned out to be quite difficult to improve upon this simple fact, until Goemans and Williamson [GW] combined semidefinite optimization with a randomized rounding technique to obtain an approximation algorithm with a relative error of about 12%. On the other hand, it was proved by Håstad [H] (tightening earlier results of Arora et al. [ArLMSS]) that no polynomial time algorithm can produce an approximation with a relative error less than (about) 6%, unless $P = NP$.

We sketch the Goemans–Williamson algorithm. Let us describe any partition $V_1 \cup V_2$ of V by a vector $x \in \{-1, 1\}^V$ by letting $x_i = -1$ iff $i \in V_1$. Then the size of the cut corresponding to x is $(1/4)\sum_{i,j}(x_i - x_j)^2$. Hence the MAX CUT problem can be formulated as maximizing the quadratic function

$$\tfrac{1}{4}\sum_{i,j}(x_i - x_j)^2 \tag{7}$$

over all $x \in \{-1, 1\}^V$. The condition on x can also be expressed by quadratic constraints:

$$x_i^2 = 1 \qquad (i \in V)\,. \tag{8}$$

Reducing a discrete optimization problem to the problem of maximizing a quadratic function subject to a system of quadratic equations may sound great, and one might try to use Lagrange multiplicators and other classical techniques. But these won't help; in fact the new problem is very difficult (NP-hard), and it is not clear that we gain anything.

The next trick is to linearize, by introducing new variables $y_{ij} = x_i x_j$ $(1 \le i, j \le n)$. The objective function and the constraints become linear in these new variables y_{ij}:

$$\text{maximize} \qquad \tfrac{1}{4} \sum_{i,j} (y_{ii} + y_{jj} - 2y_{ij}) \tag{9}$$

$$\text{subject to} \qquad y_{11} = \cdots = y_{nn} = 1\,. \tag{10}$$

Of course, there is a catch: the variables y_{ij} are not independent! Introducing the symmetric matrix $Y = (y_{ij})_{i,j=0}^{n}$, we can note that

$$Y \text{ is positive semidefinite}\,, \tag{11}$$

and

$$Y \text{ has rank } 1\,. \tag{12}$$

The problem of solving (9)–(10) with the additional constraints (11) and (12) is equivalent to the original problem, and thus NP-hard in general. But if we drop (12), then we get a tractable relaxation! Indeed, what we get is a **semidefinite program**: *maximize a linear function of the entries of a positive semidefinite matrix, subject to linear constraints on the matrix entries.*

Since positive semidefiniteness of a matrix Y can be translated into (infinitely many) linear inequalities involving the entries of Y:

$$v^{\mathsf{T}} Y v \ge 0 \qquad \text{for all } v \in \mathbf{R}^{n+1}\,,$$

semidefinite programs can be viewed as linear programs with infinitely many constraints. However, they behave much nicer than one would expect. Among others, there is a duality theory for them (see, e.g. Wolkowitz [W]). It is also important that semidefinite programs are polynomial time solvable (up to an arbitrarily small error; note that the optimum solution may not be rational). In fact, the ellipsoid method [GrLS] and, more importantly from a practical point of view, interior point methods [Al] extend to semidefinite programs.

Coming back to the Maximum Cut problem, let Y be an optimal solution of (9)-(10)-(11), and let M^* be the optimum value of the objective function. Since every cut defines a solution, we have $M^* \ge M$.

The second trick is to observe that since Y is positive semidefinite, we can write it as a Gram matrix, i.e., there exist vectors u_i if $\mathbf{R}^n$ such that $u_i^T u_j = Y_{ij}$ for all i and j. In particular, we get that $|u_i|^2 = 1$, and

$$\sum_{i,j}(u_i - u_j)^2 = 4M^* \,. \tag{13}$$

Now choose a uniformly distributed random hyperplane H through the origin. This divides the u_i into two classes, and thereby defines a cut in G. The expected weight of this cut is

$$\sum_{i,j} C_{ij}\mathsf{P} \quad (H \text{ separates } u_i \text{ and } u_j)\,.$$

But it is trivial that the probability that H separates u_i and u_j is $(1/\pi)$ times the angle between u_i and u_j. In turn, the angle between u_i and u_j is at least $0.21964(u_i - u_j)^2$, which is easily seen by elementary calculus. Hence the expected size of the cut obtained is at least

$$\sum_{i,j} 0.21964(u_i - u_j)^2 = 0.87856M^* \geq 0.87856M\,.$$

Thus we get a cut which is at least about .88 percent of the optimal.

Except for the last part, this technique is entirely general and has been used in a number of proofs and algorithms in combinatorial optimization [LoS1].

4 Discrepancy Theory

In section 3, we started with a discrete problem, and tried to find a good continuous approximation (relaxation) of it. Let us discuss a reverse problem a bit (only to the extent of a couple of classical examples). Suppose that we have a set with a measure; how well can it be approximated by a discrete measure?

In our first example, we consider the Lebesgue measure λ on the unit square $[0,1]^2$. Suppose that we want to approximate this by the uniform measure on a finite subset T. Of course, we have to specify how the error of approximation is measured: here we define it as the maximum error on axis-parallel rectangles:

$$\Delta(T) = \sup_R \big||T \cap R| - \lambda(R)|T|\big|\,,$$

where R ranges over all sets of the form $R = [a,b] \times [c,d]$ (we scaled up by $|T|$ for convenience). We are interested in finding the best set T, and in

determining the "discrepancy"

$$\Delta_n = \inf_{|T|=n} \Delta(T)\,.$$

The question can be raised for any family of "test sets" instead of rectangles: circles, ellipses, triangles, etc. The answer depends in an intricate way on the geometric structure of the test family; see [BeC] for an exposition of these beautiful results.

The question can also be raised in other dimensions. The 1-dimensional case is easy, since the obvious choice of $T = \{1/(n+1), \ldots, n/(n+1)\}$ is optimal. But for dimensions larger than 2, the order of magnitude of Δ_n is not known.

Returning to dimension 2, the obvious first choice is to try a $\sqrt{n} \times \sqrt{n}$ grid for T. This leaves out a rectangle of size about $1/\sqrt{n}$, and so it has $\Delta(T) \approx \sqrt{n}$. There are many constructions that do better; the most elementary is the following. Let $n = 2^k$. Take all points (x, y) where both x and y are multiples of 2^{-k}, and expanding them to k bits in binary, we get these bits in reverse order: $x = 0.b_1 b_2 \ldots b_k$ and $y = 0.b_n b_{n-1} \ldots b_1$. It is a nice exercise to verify that this set has discrepancy $\Delta(T) \approx k = \log_2 n$.

It is hard to prove that one cannot do better; even to prove that $\Delta_n \to \infty$ was difficult [A]. The lower bound matching the above construction was finally proved by Schmidt [S]:

$$\Delta_n = \Theta(\log n)\,.$$

This fundamental result has many applications.

Our second example can be introduced through a statistical motivation. Suppose that we are given a $0-1$ sequence $x = x_1 a_2 \ldots x_n$ of length n. We want to test whether it comes from independent coin flips.

One approach (related to von Mises's proposal for the definition of a random sequence) could be to count 0's and 1's; their numbers should be about the same. This should also be true if we count bits in the even positions, or odd positions, or more generally, in any fixed arithmetic progression of indices.

So we consider the quantity

$$\Delta(x) = \max_A \left| \sum_{i \in A} x_i - \tfrac{1}{2}|A| \right|,$$

where A ranges through all arithmetic progressions in $\{1, \ldots, n\}$. Elementary probability theory tells us that if x is generated by independent coin flips, then $\Delta(x) \approx \sqrt{n}/2$ with high probability. So if $\Delta(x)$ is larger than

this (which is the case of most non-random sequences one thinks of), then we can conclude that it is not generated by independent coin flips.

But can x fail this test in the other direction, being "too smooth"? In other words, what is

$$\Delta_n = \min_x \Delta(x)\,,$$

where x ranges over all $0-1$ sequences of length n? We can think of the question as a problem of measure approximation: we are given the measure on $\{1,\dots,n\}$ in which each atom has measure $1/2$. This is a discrete measure but not integral valued; we want to approximate it by an integral valued measure. The test family defining the error of approximation consists of arithmetic progressions.

It follows from considering random sequences that $\Delta_n = O(\sqrt{n})$ The first result in the opposite direction was proved by Roth [Rot], who showed that $\Delta_n = \Omega(n^{1/4})$. It was expected that the bound would be improved to $\sqrt{n}$ eventually, showing that random sequences are the extreme, at least in the order of magnitude (in many similar instances of combinatorial extremal problems, for example in Ramsey Theory, random choice is the best). But surprisingly, Sárközy showed that $\Delta_n = O(n^{1/3}$, which was improved by Beck [Be] to the almost sharp $\Delta_n = O(n^{1/4}\log n)$, and even the logarithmic factor was recently removed by Matoušek and Spencer [MS]. Thus there are sequences which simulate random sequences too well!

5 The Laplacian

The Laplacian, as we learn it in school, is the differential operator

$$\sum_i \frac{\partial^2}{\partial x_i^2}\,.$$

What could be more tied to the continuity of euclidean spaces, or at least of differential manifolds, than a differential operator? In this section we show that the Laplacian makes sense in graph theory, and in fact it is a basic tool. Moreover, the study of the discrete and continuous versions interact in a variety of ways, so that the use of one or the other is almost a matter of convenience in some cases.

5.1 Random walks. One of the fundamental general algorithmic problems is sampling: *generate a random element from a given distribution over a set V*. In non-trivial cases, the set V is either infinite or finite but very large, and often only implicitly described. In most cases, we want to generate an element from the uniform distribution, and this special case will

be enough for the purpose of our discussions.

I'll use as examples two sampling tasks:

(a) Given a convex body K in $\mathbf{R}^n$, generate a uniformly distributed random point of K.
(b) Given a graph G, generate a uniformly distributed random perfect matching in G.

Problem (a) comes up in many applications (volume computation, Monte-Carlo integration, optimization, etc.). Problem (b) is related to the Ising model in statistical mechanics.

In simple cases one may find elegant special methods for sampling. For example, if we need a uniformly distributed random point in the unit ball in $\mathbf{R}^n$, then we can generate n independent coordinates from the standard normal distribution, and "normalize" the obtained vector appropriately.

But in general, sampling from a large finite set, or an infinite set, is a difficult problem, and the only general approach known is the use of *Markov chains*. We define (using the structure of S) an ergodic Markov chain whose state space is S and whose stationary distribution is uniform. In the case of a finite set S, it may be easier to think of this as a connected regular non-bipartite graph G with node set V. Starting at an arbitrary node (state), we take a random walk on the graph: from a node i, we step to a next node selected uniformly from the set of neighbors of i. After a sufficiently large number T of steps, we stop: the distribution of the last node is approximately uniform.

In the case of a convex body, a rather natural Markov chain to consider is the following: starting at a convenient point in K, we move at each step to a uniformly selected random point in a ball of radius δ about the current point (if the new point is outside K, we stay where we were, and consider the step "wasted"). The step-size δ will be chosen appropriately, but typically it is about $1/\sqrt{n}$.

If we want to sample from perfect matchings in a graph G, the random walk to consider is not so obvious. Jerrum and Sinclair consider a random walk on perfect and near-perfect matchings (matchings that leave just two nodes unmatched). If we generate such a matching, it will be perfect with a small but non-negligible probability ($1/n^{\text{const}}$ if the graph is dense). So we just iterate until we get a perfect matching.

One step of the random walk is generated by picking a random edge e of G. If the current matching is perfect and contains e, then we delete e; if the current matching is near-perfect and e connects the two unmatched nodes, we add it; if e connects an unmatched node to a matched node, we add e to the matching and delete the edge it intersects.
Another way of looking at this is do a random walk on a "big" graph H whose nodes are the perfect and near-perfect matchings of the "small" graph G. Two nodes of H are adjacent if the corresponding matchings arise from each other by changing (deleting, adding, or replacing) a single edge. Loops are added to make H regular. Note that we do not want to construct H explicitly; its size may be exponentially large compared with G. The point is that the random walk on H can be implemented using only this implicit definition.

It is generally not hard to achieve that the stationary distribution of the chain is uniform, and that the chain is ergodic. The crucial question is: what is the *mixing time*, i.e., how long do we have to walk? This question leads to estimating the mixing time of Markov chains (the number of steps before the chain becomes essentially stationary). From the point of view of practical applications, it is natural to consider finite Markov chains – a computation in a computer is necessarily finite. But in the analysis, it depends on the particular application whether one prefers to use a finite, or a general measurable, state space. All the essential (and very interesting) connections that have been discovered hold in both models. In fact, the general mathematical issue is *dispersion*: we might be interested in dispersion of heat in a material, or dispersion of probability during a random walk, or many other related questions.

A Markov chain can be described by its transition matrix $M = (p_{ij})$, where p_{ij} is the probability of stepping to j, given that we are at i. In the special case of random walks on a regular undirected graph we have $M = (1/d)A$, where A is the usual adjacency matrix of the graph G. Note that the all-1 vector $\mathbf{1}$ is an eigenvector of M belonging to the eigenvalue 1.

If σ is the starting distribution (which can be viewed as a vector in $\mathbf{R}^V$), then the distribution after t steps is $M^t\sigma$. From this it is easy to see that the speed of convergence to the stationary distribution (which is the eigenvector $(1/n)\mathbf{1}$), depends on the difference between the largest eigenvalue $\mathbf{1}$ and the second largest eigenvalue λ (the *spectral gap*). We could also define the

spectral gap as the smallest positive eigenvalue of the matrix $L = M - I$.

We call this matrix L the *Laplacian* of the graph. For any vector $f \in \mathbf{R}^V$, the value of Lf at node i is the average of f over the neighbors of i, minus f_i. This property shows that the Laplacian is indeed a discrete analog of the classical Laplace operator.

Some properties of the Laplace operator depend on the fine structure of differentiable manifolds, but many basic properties can be generalized easily to any graph. One can define harmonic functions, the heat kernel, prove identities like the analogue Green's formula:

$$\sum_i \big(f_i(Lg)_i - g_i(Lf)_i\big) = 0$$

and so on. Many properties of the "continuous" Laplace operator can be derived using such easy combinatorial formulas on a grid and then taking limits. See Chung [Chu] for an exposition of some of these connections.

But more significantly, the discrete Laplacian is an important tool in the study of various graph theoretic properties. As a first illustration of this fact, let us return to random walks.

Often information about the spectral gap is difficult to obtain (this is the case in both of our introductory examples). One possible remedy is to relate the dispersion speed to *isoperimetric inequalities* in the state space. To be more exact, define the *conductance* of the chain as

$$\Phi = \max_{\emptyset \subset S \subset V} \sum_{i \in S, j \in V \setminus S} \frac{\pi_i p_{ij}}{\pi(S)\pi(V \setminus S)}\,.$$

(The numerator can be viewed as the probability that choosing a random node from the stationary distribution, and then making one step, the first node is in S and the second is in $V \setminus S$. The denominator is the same probability for two independent random nodes from the stationary distribution.) The following inequality was proved by Jerrum and Sinclair [JS]:

Theorem 3.

$$\tfrac{\Phi^2}{8} \le 1 - \lambda \le \Phi\,.$$

This means that the mixing time (which we have to use informally here, since the exact definition depends on how we start, how we measure convergence, etc.) is between $1/\Phi$ and $1/\Phi^2$.

In the case of random walks in a convex body, the conductance is very closely related to the following isoperimetric theorem [LoSi], [DyF]:

Theorem 4. *Let K be a convex body in $\mathbf{R}^n$, of diameter D. Let the surface F divide K into two parts K_1 and K_2. Then*

$$\mathrm{vol}_{n-1}(F) \geq \frac{2}{D}\frac{\mathrm{vol}(K_1)\mathrm{vol}(K_2)}{\mathrm{vol}(K)}\,.$$

(A long thin cylinder shows that the bound is sharp.)
From Theorem 4, it follows that (after appropriate preprocessing and other technical difficulties which are now swept under the carpet) one can generate a sample point in an n-dimensional convex body in $O^*(n^3)$ steps.
In the case of matchings, Jerrum and Sinclair prove that for dense graphs, the conductance of the chain described is bounded from below by $1/n^{\mathrm{const}}$, and hence the mixing time is bounded by n^{const}.

How to establish isoperimetric inequalities? The generic method is to construct, explicitly or implicitly, *flows* or *multicommodity flows*. Suppose that for each subset $\emptyset \subset S \subset V$ we can construct a flow through the graph so that each node $i \in S$ is a source that produces an amount of $\pi_i\pi(V \setminus S)$ of flow, and each $j \in V \setminus S$ is a sink that consumes $\pi_j\pi(S)$ amount of flow. Suppose further that each edge ij carries at most $K\pi_i p_{ij}$ flow. Then a simple computation shows that the conductance satisfies

$$\Phi \geq \tfrac{1}{K}\,.$$

Instead of constructing a flow for each subset S of nodes, one might prefer to construct a *multicommodity flow*, i.e., a flow of value $\pi_i\pi_j$ for each i and j. If the total flow through each edge ij is at most $K\pi_i p_{ij}$, then the conductance is at least $1/K$.

How good is the bound on the conductance obtained by the multicommodity flow method? An important result of Leighton and Rao [LR] implies that for the best choice of the flows, it gets within a factor of $O(\log n)$. One of the reasons I bring this up is that this result, in turn, can be derived from the theorem of Bourgain [Bo] about the embedability of finite metric spaces in ℓ_1-spaces with small distortion. Cf. also [LiLR], [DeL].

To construct the "best" flows or multicommodity flows is often the main mathematical difficulty. In the case of the proof of Theorem 4, it can be done by a method used earlier by Payne and Weinberger [PW] in the theory of partial differential equations. We only give a rough sketch.

Suppose that we want to construct a "flow" from K_1 to K_2. By the "Ham–Sandwich" theorem, there exists a hyperplane that cuts both K_1 and K_2 into two parts with equal volume. We can separately construct the flows on both side of this hyperplane. Repeating this procedure, we can cut up K into "needles" so that each needle is split by the partition $(S, V \setminus S)$ in the same proportion, and hence it suffices to construct a flow between K_1 and K_2 inside each needle. This is easily done using the Brun–Minkowski theorem.

For the random walk on matchings, the construction of the multicommodity flow (called *canonical paths* by Jerrum and Sinclair) also goes back to one of the oldest methods in matching theory. Given (say) a perfect matching M and a near-perfect matching M', we form the union. This is a subgraph of the "small" graph G that decomposes into a path (with edges alternating between M and M'), some cycles (again alternating between M and M') and the common edges of M and M'. Now it is easy to transform M into M' by walking along each cycle and the path, and replacing the edges one by one. What this amounts to is a path in the "big" graph H, connecting nodes M and M'. If we use this path to carry the flow, then it can be shown that no edge of the graph H is overloaded, provided the graph G is dense.

5.2 The Cage theorem and conformal mappings. It was proved by Steinitz that every 3-connected planar graph can be represented as the skeleton of a (3-dimensional) polytope. In fact, there is a lot of freedom in choosing the geometry of this representing polytope. Among various extensions, the most interesting for us is the classical construction going back to Koebe [Ko] (first proved by Andre'ev [An]; cf. also [T]):

Theorem 5 (The Cage theorem)**.** *Let H be a 3-connected planar graph. Then H can be represented as the skeleton of a 3-dimensional polytope, all whose edges touch the unit sphere.*

We may add that the representing polytope is unique up to a projective transformation of the space that preserves the unit sphere. By considering the "horizon" from each vertex of the polytope, we obtain a representation of the nodes of the graph by openly disjoint circular disks in the plane so that adjacent nodes correspond to touching circles.

The Cage theorem may be considered as a discrete form of the Riemann

mapping theorem, in the sense that it implies the Riemann mapping theorem. Indeed, suppose that we want to construct a conformal mapping of a simply connected domain K in the complex plane onto the unit disk D. For simplicity, assume that K is bounded. Consider a triangular grid in the plane (an infinite 6-regular graph), and let G be the graph obtained by identifying all gridpoints outside K into a single node s. Consider the Koebe representation by touching circles on the sphere; we may assume that the circle representing the node s is the exterior of D. So all the other circles are inside D, and we obtain a mapping of the set of gridpoints in K into D. By letting the grid become arbitrarily fine, it was shown by Rodin and Sullivan [RS] that in the limit we get a conformal mapping of K onto D.

So the Cage theorem is indeed a beautiful bridge between discrete mathematics (graph theory) and continuous mathematics (complex analysis). But where is the Laplacian? We'll see one connection in the next section. Another one occurs in the work of Spielmann and Teng [SpT], who use the Cage theorem to show that the eigenvalue gap of the Laplacian of a planar graph is at most $1/n$. This implies, among others, that the mixing time of the random walk on a planar graph is at least linear in the number of nodes.

5.3 Colin de Verdière's invariant. In 1990, Colin de Verdière [Co] introduced a parameter $\mu(G)$ for any undirected graph G. Research concerning this parameter involves an interesting mixture of ideas from graph theory, linear algebra, and analysis.

The exact definition goes beyond the limitations of this article; roughly speaking, $\mu(G)$ is the multiplicity of the smallest positive eigenvalue of the Laplacian of G, where the edges of G are weighted so as to maximize this multiplicity.

The parameter was motivated by the study of the maximum multiplicity of the second eigenvalue of certain Laplacian-type differential operators, defined on Riemann surfaces. He approximated the surface by a sufficiently densely embedded graph G, and showed that the multiplicity of the second eigenvalue of the operator can be bounded by this value $\mu(G)$ depending only on the graph.

Colin de Verdière's invariant created much interest among graph theorists, because of its surprisingly nice graph-theoretic properties. Among others, it is minor-monotone, so that the Robertson–Seymour graph minor theory applies to it. Moreover, planarity of graphs can be characterized by

this invariant: $\mu(G) \leq 3$ *if and only if* G *is planar*.

Colin de Verdière's original proof of the "if" part of this fact was most unusual in graph theory: basically, reversing the above procedure, he showed how to reconstruct a sphere and a positive elliptic partial differential operator P on it so that $\mu(G)$ is bounded by the dimension of the null space of P, and then invoked a theorem of Cheng [C] asserting that this dimension is at most 3.

Later van der Holst [Ho] found a combinatorial proof of this fact. While this may seem as a step backward (after all, it eliminated the necessity of the only application of partial differential equations in graph theory I know of), it did open up the possibility of characterizing the next case. Verifying a conjecture of Robertson, Seymour, and Thomas, it was shown by Lovász and Schrijver [LoS2] that $\mu(G) \leq 4$ *if and only if* G *is linklessly embedable in* $\mathbf{R}^3$.

Can one go back to the original motivation from $\mu(G)$ and find a "continuous" version of this result? In what sense does a linklessly embedded graph approximate the space? These questions appear very difficult.

It turns out that graphs with large values of μ are also quite interesting. For example, for a graph G on n nodes with $\mu(G) \geq n-4$, the complement of G is planar, up to introducing "twin" points; and the converse of this assertion also holds under reasonably general conditions. The proof of the latter fact uses the Koebe–Andre'ev representation of graphs.

So the graph invariant $\mu(G)$ is related at one end of the scale to elliptic partial differential equations (Cheng's theorem); on the other, to Riemann's theorem on conformal mappings.

References

[A] T. van Aardenne-Ehrenfest, On the impossibility of a just distribution, Nederl. Akad. Wetensch. Proc. 52 (1949), 734–739; Indagationes Math. 11 (1949) 264–269.

[Al] F. Alizadeh, Interior point methods in semidefinite programming with applications to combinatorial optimization, SIAM J. on Optimization 5 (1995), 13–51.

[AloS] N. Alon, J. Spencer, The Probabilistic Method. With an appendix by Paul Erdős, Wiley, New York, 1992.

[An] E. Andre'ev, On convex polyhedra in Lobachevsky spaces, Mat. Sbornik, Nov. Ser. 81 (1970), 445–478.

[ArLMSS] S. Arora, C. Lund, R. Motwani, M. Sudan, M. Szegedy, Proof verification and hardness of approximation problems, Proc. 33rd FOCS

(1992), 14–23.

[B] I. Bárány, A short proof of Kneser's conjecture, J. Combin. Theory A 25 (1978), 325–326.

[Be] J. Beck, Roth's estimate of the discrepancy of integer sequences is nearly sharp, Combinatorica 1 (1981), 319–325.

[BeC] J. Beck, W. Chen, Irregularities of Distribution, Cambridge Univ. Press (1987).

[BeS] J. Beck, V.T. Sós, Discrepancy Theory, Chapter 26, in "Handbook of Combinatorics" (R.L. Graham, M. Grötschel, L. Lovász, eds.), North-Holland, Amsterdam (1995).

[Bj] A. Bjorner, Topological methods, in "Handbook of Combinatorics" (R.L. Graham, L. Lovász, M. Grötschel, eds.), Elsevier, Amsterdam, 1995, 1819–1872.

[Bo] J. Bourgain, On Lipschitz embedding of finite metric spaces in Hilbert space, Isr. J. Math. 52 (1985), 46–52.

[C] S.Y. Cheng, Eigenfunctions and nodal sets, Commentarii Mathematici Helvetici 51 (1976), 43–55.

[Ch] A.J. Chorin, Vorticity and Turbulence, Springer, New York, 1994.

[Co] Y. Colin de Verdière, Sur un nouvel invariant des graphes et un critère de planarité, Journal of Combinatorial Theory Series B 50 (1990), 11–21.

[Chu] F.R.K. Chung, Spectral Graph Theory, CBMS Reg. Conf. Series 92, Amer. Math. Soc., 1997.

[DP] C. Delorme, S. Poljak, Combinatorial properties and the complexity of max-cut approximations, Europ. J. Combin. 14 (1993), 313–333.

[DeL] M. Deza, M. Laurent, Geometry of Cuts and Metrics, Springer Verlag, 1997.

[Du] R.M. Dudley, Distances of probability measures and random variables, Ann. Math. Stat. 39 (1968), 1563–1572.

[DyF] M. Dyer, A. Frieze, Computing the volume of convex bodies: a case where randomness provably helps, in"Probabilistic Combinatorics and Its Applications" (Béla Bollobás, ed.), Proceedings of Symposia in Applied Mathematics, Vol. 44 (1992), 123–170.

[DyFK] M. Dyer, A. Frieze, R. Kannan, A random polynomial time algorithm for approximating the volume of convex bodies, Journal of the ACM 38 (1991), 1–17.

[GW] M.X. Goemans, D.P. Williamson, .878-Approximation algorithms for MAX CUT and MAX 2SAT, Proc. 26th ACM Symp. on Theory of Computing (1994), 422-431.

[GrLS] M. Grötschel, L. Lovász, A. Schrijver, Geometric Algorithms and Combinatorial Optimization, Springer, 1988.

[H] J. Håstad, Some optimal in-approximability results, Proc. 29th ACM Symp. on Theory of Comp. (1997), 1–10.

[Ho] H. van der Holst, A short proof of the planarity characterization of Colin de Verdière, Journal of Combinatorial Theory, Series B 65 (1995), 269–272.

[JS] M. Jerrum, A. Sinclair, Approximating the permanent, SIAM J. Computing 18 (1989), 1149-1178.

[K] M. Kneser, Aufgabe No. 360, Jber. Deutsch. Math. Ver. 58 (1955).

[Ko] P. Koebe, Kontaktprobleme der konformen Abbildung, Berichte uber die Verhandlungen d. Sächs. Akad. d. Wiss., Math.–Phys. Klasse 88 (1936), 141–164.

[LR] F.T. Leighton, S. Rao, An approximate max-flow min-cut theorem for uniform multicommodity flow problems with applications to approximation algorithms, Proc. 29th Annual Symp. on Found. of Computer Science, IEEE Computer Soc. (1988), 422-431.

[LiLR] N. Linial, E. London, Y. Rabinovich, The geometry of graphs and some of its algebraic applications, Combinatorica 15 (1995), 215–245.

[Lo1] L. Lovász, Kneser's conjecture, chromatic number, and homotopy, J. Comb. Theory A 25 (1978), 319-324.

[Lo2] L. Lovász, On the Shannon capacity of graphs, IEEE Trans. Inform. Theory 25 (1979), 1–7.

[LoM] L. Lovász, P. Major, A note on a paper of Dudley, Studia Sci. Math. Hung. 8 (1973), 151–152.

[LoP] L. Lovász, M.D. Plummer, Matching Theory, Akadémiai Kiadó - North Holland, Budapest, 1986.

[LoS1] L. Lovász, A. Schrijver, Cones of matrices and set-functions, and 0-1 optimization, SIAM J. on Optimization 1 (1990), 166–190.

[LoS2] L. Lovász, A. Schrijver, A Borsuk theorem for antipodal links and a spectral characterization of linklessly embeddable graphs, Proceedings of the AMS 126 (1998), 1275–1285.

[LoSi] L. Lovász, M. Simonovits, Random walks in a convex body and an improved volume algorithm, Random Structures and Alg. 4 (1993), 359–412.

[M] J. Matoušek, Geometric Discrepancy, Springer Verlag, 1999.

[MS] J. Matoušek, J. Spencer, Discrepancy in arithmetic progressions, J. Amer. Math. Soc. 9 (1996), 195–204.

[O] D. Ornstein, Bernoulli shifts with the same entropy are isomorphic, Advances in Math. 4 (1970), 337–352.

[PW] L.E. Payne, H.F. Weinberger, An optimal Poincaré inequality for convex domains, Arch. Rat. mech. Anal. 5 (1960), 286–292.

[RS] B. Rodin, D. Sullivan, The convergence of circle packings to the Riemann mapping, J. Differential Geom. 26 (1987), 349–360.

[RoH] G.-C. Rota, L.H. Harper, Matching theory, an introduction, in "Advances in Probability Theory and Related Topics" (P. Ney, ed.) Vol I,

Marcel Dekker, New York (1971) 169–215.

[Rot] K.F. Roth, Remark concerning integer sequences, Acta Arith. 9 (1964), 257–260.

[S] W. Schmidt, Irregularities of distribution, Quart. J. Math. Oxford 19 (1968), 181–191.

[SpT] D.A. Spielman, S.-H. Teng, Spectral partitioning works: planar graphs and finite element meshes, Proc. 37th Ann. Symp. Found. of Comp. Sci., IEEE (1996), 96–105.

[St] V. Strassen, The existence of probability measures with given marginals, Ann. Math. Statist 36 (1965), 423–439.

[T] W. Thurston, The Geometry and Topology of Three-manifolds, Princeton Lecture Notes, Chapter 13, Princeton, 1985.

[W] H. Wolkowitz, Some applications of optimization in matrix theory, Linear Algebra and its Applications 40 (1981), 101–118.

László Lovász, Microsoft Research, One Microsoft Way, Redmond, WA 98052, USA lovasz@microsoft.com