

Combinatorics of optimal designs

R. A. Bailey and Peter J. Cameron

Abstract

To a combinatorialist, a design is usually a 2-design or balanced incomplete-block design. However, 2-designs do not necessarily exist in all cases where a statistician might wish to use one to design an experiment. As a result, statisticians need to consider structures much more general than the combinatorialist's designs, and to decide which one is "best" in a given situation. This leads to the theory of optimal design. There are several concepts of optimality, and no general consensus about which one to use in any particular situation.

For block designs with fixed block size k , all these optimality criteria are determined by a graph, the *concurrence graph* of the design, and more specifically, by the eigenvalues of the Laplacian matrix of the graph. It turns out that the optimality criteria most used by statisticians correspond to properties of this graph which are interesting in other contexts: D-optimality involves maximizing the number of spanning trees; A-optimality involves minimizing the sum of resistances between all pairs of terminals (when the graph is regarded as an electrical circuit, with each edge being a one-ohm resistor); and E-optimality involves maximizing the smallest eigenvalue of the Laplacian (the corresponding graphs are likely to have good expansion and random walk properties). If you are familiar with these properties, you may expect that related "nice" properties such as regularity and large girth (or even symmetry) may tend to hold; some of our examples may come as a surprise!

The aim of this paper is to point out that the optimal design point of view unifies various topics in graph theory and design theory, and suggests some interesting open problems to which combinatorialists of all kinds might turn their expertise. We describe in some detail both the statistical background and the mathematics of various topics such as Laplace eigenvalues of graphs.

1 Preliminaries

This first section of the paper sets the scene. We look briefly at the different ways in which a combinatorialist and a statistician view a simple block design such as the Fano plane, and also introduce a running example which would not be recognised as a design under the standard combinatorial definition but is in fact the best design for a statistician in certain circumstances.

Section 2 looks at the way that information about treatment differences is recovered from the results of an experiment conducted using a block design. We look briefly at what makes a good design, and show that all criteria for this can be expressed in terms of the concurrence graph of the design. Having thus focussed our attention on graphs, we give a brief survey of the Laplacian matrix of a graph and its eigenvalues, and mention the interesting question of which graphs can be concurrence graphs of block designs with given block size.

Section 3 covers the three most important kinds of optimality, described by the letters A (average), D (determinant) and E (extreme), and how they look when expressed in terms of the concurrence matrix.

In section 4 we look at sparse graphs. For trees and unicyclic graphs, we identify the A-, D- and E-optimal graphs, observing that for unicyclic graphs the different

criteria give very different answers. We also give extensions to more general classes of graphs. These results have applications in the design of microarray experiments in genetics.

In section 5 we look at other kinds of design and optimality criterion, including the case of block designs which do not have constant block size. In the final section we look briefly at computational issues and describe how the GAP share package DESIGN can be used for some of these computations.

We note that lack of communication between mathematicians and statisticians has led to some duplication of effort in this area. For example, Kelmans and Chelnokov [62] showed in 1974 that the graph on n vertices with $\lambda n(n-1)/2$ edges with the largest number of spanning trees is the λ -fold complete graph. Cheng [32] pointed out that this follows from an early result on optimal designs by Kiefer in 1958 [65]. Wild [105] has attempted to bring the two communities closer together; this paper represents another attempt.

1.1 Introduction

Combinatorialists and statisticians understand different things by the term “block design”. To a combinatorialist, the archetypal block design is the Fano plane shown in Figure 1.

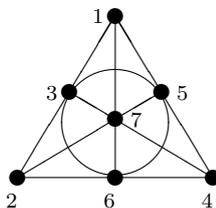


Figure 1: The Fano plane

The design has seven points and seven blocks, each block being a set of three points; any two distinct points are contained in a unique block.

A statistician might introduce this design with an example like the following. Seven different makes of fertilizer are to be tested in an experiment. Twenty-one plots of land are available for the experiment, three plots on each of seven farms in different parts of the country. We can apply one fertilizer to each plot, then grow a crop on all the plots and measure the yield. How should we allocate fertilizers to the plots? Clearly the worst thing we could do would be to put fertilizer 1 on the three plots on Farm A, fertilizer 2 on the plots on Farm B, etc. If we did this, then any difference in yield between these plots could be the result of the fertilizer, but could equally be the result of differences in fertility, soil structure or climate between the farms. It is not hard to believe that the best way to assign the fertilizers to the farms is using a scheme like the one in Table 1.

This is the Fano plane in disguise: if we identify the seven types of fertilizer with the points of the plane, and the seven farms with the lines, then a fertilizer is used on a farm precisely when the corresponding point and line are incident.

There is another step in the design of the experiment, namely *randomization*: we could number the fertilizers and farms in random order, and assign the three

	Farm A	Farm B	Farm C	Farm D	Farm E	Farm F	Farm G
Plot 1	1	1	1	2	2	3	3
Plot 2	2	4	6	4	5	4	5
Plot 3	3	5	7	6	7	7	6

Table 1: The Fano plane

fertilizers randomly to the three plots on each farm. We will not discuss further the techniques and reasons for randomization here.

The main thing to notice is that the most important things in the experimental design, the individual plots or experimental units on which the measurements are made, are not immediately visible in the mathematician's picture of the Fano plane. A little thought shows that plots correspond to the *flags* (the incident point-line pairs) in the plane.

The combinatorialist and statistician agree that in this case the best thing to do is to use the Fano plane. But what if there are only six fertilizers and six farms? There is no 2-design with six points and six blocks, having three points in each block; but the statistician still has the job of designing the experiment so as to extract as much information as possible about differences between the fertilizers. Even worse, what should we do when there are different numbers of plots available on each farm?

1.2 Combinatorial design

A combinatorialist usually defines a *block design* in one of two ways:

- a set of *points*, with a collection of subsets called *blocks*, satisfying some further conditions;
- two sets (a set of *points* and a set of *blocks*) with a binary relation of *incidence* between them.

The two views are almost equivalent. Given the first definition, we define the incidence relation between points and blocks to be the membership relation: point p and block B are incident if $p \in B$. Conversely, given two sets and an incidence relation, we can identify any block with the set of points incident with it; the difference is that we obtain a multiset of subsets of the point set, that is, we allow *repeated blocks*. However, we will freely mix the language associated with the two viewpoints.

Typical of the additional conditions imposed are those of being a t - (v, k, λ) design, where $1 \leq t \leq k \leq v$ and $\lambda > 0$: this requires that there are v points altogether, that any block contains k points, and that any set of t points is contained in exactly λ blocks. Thus the Fano plane is a 2- $(7, 3, 1)$ design.

The difference between the two viewpoints is important for the existence question. If repeated blocks are allowed, it is easy to show that t - (v, k, λ) designs in which not every k -set is a block exist for every t, k, v satisfying $t < k < v - t$ (for some λ). However, the existence of t -designs without repeated blocks is a much more difficult question: none with $t > 5$ were known until the 1980s, and their existence for all t was shown by Teirlinck [96] in 1987. A different proof was found by Ajoodani-Namini [1] in 1996. Ajoodani-Namini's proof is very much simpler, but

Teirlinck's gives designs where the value of λ (though large) is a fixed function of t , and does not grow with v . We refer to the paper by Khosrovshahi and Tayfeh-Rezaie in this volume [64] for more information on t -designs.

In the case of 2-designs, there is a powerful existence theory due to Wilson [106]. There are two easy necessary conditions on the parameters for the existence of such a design:

$$\begin{aligned} k-1 & \text{ divides } (v-1)\lambda, \\ k & \text{ divides } v(v-1)\lambda. \end{aligned} \tag{1.1}$$

For the number of blocks containing a point is equal to $(v-1)\lambda/(k-1)$, while the total number of blocks is $v(v-1)\lambda/k(k-1)$. These conditions, together with *Fisher's inequality* $b \geq v$, are known to be sufficient if $k \leq 4$ [51]. Wilson showed that, for any given k and λ , these conditions are sufficient for the existence of a 2 - (v, k, λ) design if v is large enough (in terms of k and λ).

We are far from having a comparable theory for t -designs with $t > 2$, however.

It is quite common in the combinatorial literature for the unadorned word “design” to denote a 2-design. This will not be our viewpoint; we now explain why statisticians require a much more general concept.

1.3 Statistical design

In a fairly simple kind of comparative experiment, a number v of different treatments will be applied to experimental units and the responses measured and analysed. The design is an allocation of treatments to units; because of randomization, it is better to think of the design as a partition of the set of units into v parts, so that one treatment can be allocated to each part. We use v for the number of treatments to avoid conflict with the t used in the theory of t -designs; think of v for “varieties” – in an agricultural experiment the treatments applied to the fields might be different varieties of wheat.

If the experimental units are all alike, then it seems reasonable to partition them into parts as nearly equal as possible; any such partition is as good as any other, and there is nothing interesting from a combinatorial point of view. But things are not always so simple. The next level of complication is that the units come already partitioned into “blocks”, so that the units within a block are alike, but differ systematically from those in a different block (such as the twenty-one fields on seven different farms in our example). In this case, the relationship between the two partitions gives the design a non-trivial combinatorial structure.

Accordingly, we define a *block design* to be a set of “units” or “plots” carrying two partitions: a treatment partition with v parts, and a block partition with b parts.

This looks very different from a combinatorialist's block design, but we will see that it can be recast into a very similar (but rather more general) form. To do this, we follow the combinatorialist and take the set of v treatments as the basic object. Each block now consists of a multiset of treatments, namely, the treatments applied to the plots in that block. The block design is called *binary* if no treatment occurs more than once in each block. In a binary design, a block is now identified with a set of treatments; so a binary block design can be thought of as a multiset of subsets of

the set of treatments (it is a multiset to allow the possibilities of repeated blocks). In general, however, we have a multiset of multisets of treatments.

There may be good scientific reasons for allowing a treatment to occur more than once in a block; it may have a self-interaction which is important but would not be detected if this were not done. In the simple models we consider in this paper, such interactions will not be relevant. But this does not mean that the best designs are always binary.

In the discussion of Tocher's paper [99], David Cox said

I suspect that ... balanced ternary designs are of no practical value.

(Here a design is *ternary* if the maximum number of occurrences of a fixed treatment in a block is two.) Contrary to this suspicion, it may well happen that optimal designs are non-binary; we will see examples.

However, since we will assume that we are not interested in effects involving repetition of a single treatment and simply want to compare the treatments, then a block in which only one treatment occurs (an arbitrary number of times) is simply wasted effort, since it can give no information about treatment differences. So we will make one restriction on the designs we consider: *no block can contain just a single treatment*.

Another way to represent a block design is by its *Levi graph*, defined as follows. The graph is bipartite, with vertices in one class of the bipartition being the treatments and those in the other being the blocks; for each experimental unit, the graph has an edge connecting the treatment on that unit (i.e. the part of the treatment partition containing it) with its block (the part of the block partition containing it). This graph may have multiple edges; indeed, it is simple if and only if the design is binary. Now we can obtain the "multiset of multisets" representation by taking the treatments as points, and the neighbourhoods of the blocks as the multisets. (If n_{ij} edges join point i to block j , then point i occurs n_{ij} times in the multiset corresponding to j .)

These numbers give us another simple way of representing the design. We always consider designs for v treatments in b blocks. Let N be the $v \times b$ *incidence matrix*: the entry n_{ij} is the number of times that treatment i occurs in block j . Thus the column sums of N are the block sizes, which we shall usually assume are all equal to k , while the sum r_i of the i -th row is the number of times which treatment i occurs overall, which is the *replication* of i . Thus the design is binary if $n_{ij} \in \{0, 1\}$ for all i and j .

Put $\Lambda = NN^T$, which is called the *concurrence matrix*. If the design is binary then λ_{ij} is just the number of blocks in which treatments i and j occur; in particular, $\lambda_{ii} = r_i$.

For more information on design of comparative experiments, we refer to [8].

1.4 An example

Example 1.1 Here is an example of a block design which we will use as a running example in this paper. This design was first given by Tocher [99] in the paper cited above. There are five treatments, labelled 1, 2, 3, 4, 5, and 21 units, divided into seven blocks of three, corresponding to the columns of Table 2. Its Levi graph is in Figure 2.

1	1	1	1	2	2	2
1	3	3	4	3	3	4
2	4	5	5	4	5	5

Table 2: A block design

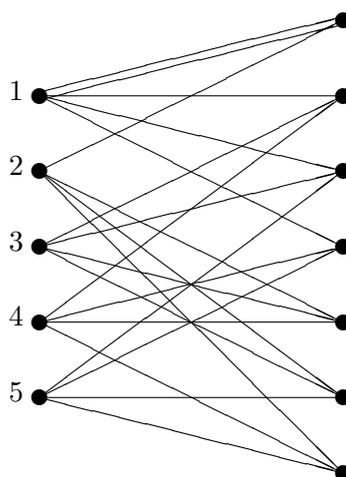


Figure 2: The Levi graph of the design in Table 2

Every block has size three, though of course the first block is a multiset. The point 1 occurs five times, including twice in the first block; the remaining points occur four times each. Every pair of points has concurrence precisely 2: the two concurrences of 1 and 2 are accounted for by the first block. So all off-diagonal elements in the concurrence matrix are equal to 2. On the other hand, the diagonal entries are $(7, 4, 4, 4, 4)$. (The two occurrences of treatment 1 in the first block contribute 4 to the first entry of NN^T .)

1.5 Graphs

As hinted above, our graphs can have multiple edges but not loops. (In Section 5.4, we briefly have a more general concept, edge-weighted graphs.) A *simple graph* is one with no multiple edges; a *regular graph* is a graph where all vertices have the same valency. Graphs will almost always be connected. The *adjacency matrix* of a graph has rows and columns indexed by vertices, with (i, j) entry equal to the number of edges connecting i and j .

A *distance-regular graph* is a simple connected regular graph of diameter d for which there exist parameters c_m, a_m, b_m (for $0 \leq m \leq d$, but c_0, a_0 and b_d are undefined) such that, if i and j are vertices at distance m , then the numbers of vertices adjacent to j and at distance $m - 1, m, m + 1$ respectively from i are c_m, a_m, b_m respectively.

There is a rich theory of distance-regular graphs, for which we refer to [16, 23, 48]. In particular, note that the eigenvalues of the adjacency matrix of a distance-regular

graph are the same as those of a tridiagonal matrix of order $d + 1$ having c_m, a_m, b_m in row $m + 1$; their multiplicities can also be calculated from these parameters. A *strongly regular graph* is a distance-regular graph with diameter 2. (Some authors allow “strongly regular graphs” to be disconnected; such graphs are disjoint unions of complete graphs of the same size.)

1.6 Partial balance

An *association scheme* of rank $s + 1$ on the set of v treatments is a set of symmetric $v \times v$ matrices A_0, \dots, A_s with entries 0 and 1 such that $A_0 = I$, $A_0 + \dots + A_s = J$, and the product $A_i A_j$ is a linear combination of A_0, \dots, A_s for all i and j . Here J denotes the $v \times v$ all-1 matrix. A pair of distinct treatments are said to be i -th associates if the corresponding entry of A_i is 1. Thus if $s = 1$ then the scheme is trivial with $A_1 = J - I$.

An association scheme of rank 2 is either a complementary pair of strongly regular graphs, or is *group-divisible* (one associate class being a uniform partition with the diagonal removed)—the second alternative is unnecessary if the more general definition of “strongly regular graph” is adopted. More generally, the distance classes in a distance-regular graph of diameter s define an association scheme of rank $s + 1$.

The set of all real linear combinations of A_0, \dots, A_s is the *Bose–Mesner algebra* of the association scheme. This algebra is commutative, and so has a basis of minimal idempotents as well as the basis formed by the adjacency matrices A_i . Many techniques for calculation in association schemes depend on transforming between these two bases: see [5].

A binary block design is defined to be *partially balanced* with respect to a given association scheme if its concurrence matrix is in the Bose–Mesner algebra of that scheme. If the scheme has rank $s + 1$, the design is said to have s associate classes. Partially balanced block designs with two associate classes were introduced by Bose and Nair in [21] when statisticians found that they needed block designs for some parameter sets for which no 2-design exists. Allowing two different concurrences seemed to be the next level of complexity; and restricting to partially balanced designs gave a simple method of calculating matrix inverses (see Section 2.1) in the days before computers were widely available.

2 Optimality

We begin this section with the statistical background: the information matrix of a block design, its generalized inverse, and the role of the latter in estimating treatment differences from experimental data. We want these estimates to be as accurate as possible: this leads to various notions of optimality. For designs with constant block size, optimality criteria depend on the concurrence graph of the design, specifically on the eigenvalues of its Laplacian matrix. Kiefer’s Theorem (Theorem 2.3) asserts that 2-designs, if they exist, are optimal in any reasonable sense. We conclude with a discussion of variance-balanced designs (with constant block size), a generalization of 2-designs.

2.1 The information matrix and its generalized inverse

Put

$$C = R - \frac{1}{k}\Lambda,$$

where $R = \text{diag}(r_1, \dots, r_v)$. This is called the *information matrix* of the design, for reasons that will become clear in Section 2.2. It is a $v \times v$ symmetric matrix with zero row and column sums.

The trace of C is equal to $(\sum_{i \neq j} \lambda_{ij})/k$. We will need to know when this is maximized (for given v, b, k). If the point i has m_i occurrences in a block, then the contribution of that block to the trace is

$$\sum_{i \neq j} m_i m_j = k^2 - \sum_i m_i^2;$$

since $\sum_i m_i = k$, the contribution is maximized when the m_i are as nearly equal as possible. We conclude:

Proposition 2.1 *A design with v points and b blocks of size k maximizes the trace of C (among all such designs) if and only if each point occurs $\lfloor k/v \rfloor$ or $\lceil k/v \rceil$ times in each block. In particular, if $k < v$, the trace of C is maximized by the binary designs; the maximum value is $b(k-1)$.*

In general, designs satisfying the conditions of this result are called *generalized binary*; such a design is obtained by adding a fixed number of occurrences of each point to each block of a binary design.

Let u be the all-1 vector in \mathbb{R}^v . (We adopt the statisticians' convention that vectors are column vectors.) Then u is an eigenvector of C with eigenvalue 0, which we call the *trivial eigenvalue*. In general, the multiplicity of the zero eigenvalue is equal to the number of connected components in the Levi graph (this follows from Proposition 2.9). This is why we say "non-trivial" rather than "non-zero" to indicate the eigenvalues on vectors orthogonal to u . Connectedness ensures that this eigenvalue has multiplicity 1. In this case, the design too is said to be *connected*. From now on, we shall assume that we are considering only connected designs.

Since it is symmetric, C can be written in spectral form as

$$C = \sum \theta_i P_i,$$

where θ_i are the distinct eigenvalues and P_i are the eigenprojectors. It is convenient to number the eigenvalues so that $\theta_0 = 0$; then, for a connected design, P_0 is the projector onto the space $\langle u \rangle$ spanned by u , so $P_0 = v^{-1}J$, where J is the $v \times v$ all-1 matrix.

The matrix C is singular, but we need some kind of inverse. The matrix B is said to be a *generalized inverse* of C if it satisfies

$$BCB = B \quad \text{and} \quad CBC = C. \quad (2.1)$$

Since C is invertible on the space $\langle u \rangle^\perp$ and zero on the space $\langle u \rangle$, in some ways the most natural generalized inverse is the one defined by

$$BC = CB = I - P_0 \quad \text{and} \quad BP_0 = P_0B = 0.$$

This one is called the *Moore–Penrose generalized inverse*, written C^- . It is given by

$$C^- = \sum_{i \neq 0} \frac{1}{\theta_i} P_i. \quad (2.2)$$

It is also a symmetric matrix with zero row and column sums. If B is any other generalized inverse, then $C^- = (I - P_0)B(I - P_0)$.

Statisticians who evaluate block designs have developed several methods for calculating the Moore–Penrose generalized inverse of the information matrix C . Here are some of their methods.

- (a) Express C in spectral form and use Equation (2.2). This method is good when it is easy to find the eigenspaces. Then it can work for whole families of matrices of different sizes. For example, in an association scheme the eigenprojectors are just (sums of some of) the minimal idempotents of the scheme.
- (b) Use a computer (algebra) package to compute

$$C^- = (C + P_0)^{-1} - P_0,$$

where $P_0 = v^{-1}J$. The trick is that $C + P_0$ is invertible and is the identity on $\langle u \rangle$, so its inverse is also the identity on $\langle u \rangle$.

- (c) (This is the method favoured by electrical engineers: see later.) Delete the last row and column of C , to leave an invertible matrix \tilde{C} . Put \tilde{C}^{-1} in the top left hand corner, and zeros in the last row and column, then pre- and post-multiply by $I - P_0$.
- (d) More generally, to calculate the Moore–Penrose generalized inverse of a symmetric partitioned matrix, statisticians often use the formula in Lemma A.3 of [89], which uses a result of [88]. The formula in Exercise 2.7 of [84, page 33] is similar.
- (e) Start with a symmetric matrix B with zero row sums. If there is any simultaneous permutation of the rows and columns of C that leaves it unaltered, reduce the number of unknowns in B so that it is also invariant under this permutation. Then use brute (hand) force to solve Equations (2.1).
- (f) Find powers of C until they, together with $I - P_0$, are linearly dependent. This gives an equation of the form

$$\text{polynomial in } C \text{ with no constant term} = I - P_0.$$

The degree of the left-hand side is equal to the number of distinct non-trivial eigenvalues of C . Factorize the left-hand side as CB . Pre- and post-multiply B by $I - P_0$ to get C^- . Since B has constant row and column sums, in practice this means subtracting c/v from every entry in B , where c is the row sum. If there are not many distinct eigenvalues, this method can be quicker than (a), because the eigenprojectors do not need to be calculated.

A square matrix is *completely symmetric* if its diagonal elements are constant and its off-diagonal elements are constant; that is, if it is a linear combination of I and J . If an information matrix is completely symmetric, it has the form $C = c(I - P_0)$, and so its Moore–Penrose inverse $C^- = c^{-1}(I - P_0)$ is also completely symmetric.

2.2 Estimation and variance

In the experiment, the scientist measures a response on each experimental unit. Suppose that experimental unit ω is in block j and is allocated treatment i . We assume that there are numbers τ_1, \dots, τ_v and β_1, \dots, β_b such that the expected response on unit ω is $\tau_i + \beta_j$. Notice that this is the real world here, not algebra or arithmetic: if we measure the same thing on the same treatment in the same block on two different occasions we will not usually get the same answer. This is partly because of errors of observation, and partly because of inherent variability in the underlying experimental material. For a good discussion of this, see [81].

Thus, writing Y_ω for the response on unit ω , we have to assume that Y_ω is a random variable whose expectation is $\tau_i + \beta_j$. That is,

$$E(Y_\omega) = \tau_i + \beta_j \quad \text{if } \omega \text{ is in block } j \text{ and has treatment } i. \quad (2.3)$$

We are not interested in the β parameters, but do want to estimate the τ parameters.

What is an *estimator*? It is a function of the data, but not of the unknown values which we are trying to estimate. Because the responses are random variables, each estimator is also a random variable. The most important property of an estimator is that it should be *unbiased*, which means that the expected value of the estimator is equal to the unknown parameter which we are trying to estimate.

In Equations (2.3), we can obviously add a constant to all the τ parameters and subtract it from all the β parameters without changing the model, so we cannot estimate the τ_i . In fact, we can adjust constants in this way separately in each connected component of the Levi graph. However, this is the only source of ambiguity, as we shall show. More precisely, we shall show that if the vector x is in the image of C (equivalently, if x is orthogonal to the null space of C), then there is a linear combination of the responses whose expected value is equal to $x^\top \tau$. When the design is connected, the null space of C is just $\langle u \rangle$, so x is orthogonal to the null space if and only if $x^\top u = 0$; that is, the coefficients in the vector x sum to zero. Such a vector is called a *contrast vector*, and the corresponding linear combination $x^\top \tau$ is called a *contrast* in the treatment parameters τ_1, \dots, τ_v .

Statisticians usually rewrite Equations (2.3) in the following vector form:

$$E(Y) = X\tau + Z\beta. \quad (2.4)$$

Here Y , τ and β are vectors in \mathbb{R}^{bk} , \mathbb{R}^v and \mathbb{R}^b respectively. The matrix X has bk rows and v columns: the entry in row ω and column i is equal to 1 if experimental unit ω gets treatment i ; otherwise it is zero. The $bk \times b$ matrix Z is defined similarly, using blocks instead of treatments.

The matrix $k^{-1}ZZ^\top$ is the orthogonal projector onto the subspace \mathcal{B} of \mathbb{R}^{bk} consisting of vectors which are constant on each block. (If the blocks have different sizes k_1, \dots, k_b , then the matrix is $ZK^{-1}Z^\top$, where $K = \text{diag}(k_1, \dots, k_b)$: we shall return to this form in Section 5.4.) Put $Q = I - k^{-1}ZZ^\top$. To remove the β parameters from Equation (2.4), we premultiply by Q . The expectation operator commutes with linear transformations, so we obtain

$$E(QY) = QX\tau. \quad (2.5)$$

That is, the expected value of the transformed data QY lies in the column space \mathcal{T} of QX . It is fairly natural to use the *least-squares principle* to say that our estimate of $QX\tau$ should be the vector in \mathcal{T} which is closest to the transformed data vector QY in the sense that the sum of the squares of the differences in coordinates is minimized. By Pythagoras' Theorem, we can obtain this by projecting the vector QY orthogonally onto \mathcal{T} .

Note that $X^\top X = R$ and $X^\top Z = N$. Hence $X^\top QX = C$, where C is the information matrix defined in the previous section. Consider the symmetric matrix $S = QXC^{-1}X^\top Q$. Any vector in \mathcal{T} has the form $QX\tau$, and

$$SQX\tau = QXC^{-1}X^\top QQX\tau = QXC^{-1}X^\top QX\tau = QXC^{-1}C\tau = QX(I - P_0)\tau = QX\tau$$

because the columns of XP_0 are constant vectors, hence in \mathcal{B} . On the other hand, if w is a vector in \mathbb{R}^{bk} which is orthogonal to \mathcal{T} , then $X^\top Qw = 0$ and so $Sw = 0$. It follows that S is the orthogonal projector onto \mathcal{T} .

Thus the least-squares principle gives the vector of *fitted values* as SQY . That is, our estimator of $QX\tau$ is SQY , so our estimator of $X^\top QX\tau$ is $X^\top SQY$. We have already shown that $SQX = QX$, and $QS = S$, so $X^\top SQY = X^\top QSQY = X^\top QY$. Moreover, $X^\top QX = C$, so our estimator of $C\tau$ is $X^\top QY$. If the vector x in \mathbb{R}^v is orthogonal to the null space of C , there is a vector z in \mathbb{R}^v such that $Cz = x$. Then $x^\top \tau = z^\top C\tau$, whose estimator is $z^\top X^\top QY$.

We have used the least-squares principle to obtain an estimator for $x^\top \tau$ for any vector x orthogonal to the null space of the information matrix C . In particular, if the design is connected then this gives an estimator for every contrast. The estimator is a linear combination of the data.

What are the properties of this estimator? If $Cz = x$ then

$$E(z^\top X^\top QY) = z^\top X^\top Q E(Y) = z^\top X^\top QX\tau = z^\top C\tau = x^\top \tau,$$

so the estimator is unbiased.

The expectation, or mean, of a random variable is its first moment; the second moment (about the mean) is its variance. Now we consider the variance of the estimator. The smaller the variance, the more likely it is that our estimate will be close to the true value. For example, if the estimator of the difference $\tau_1 - \tau_2$ has a normal distribution with variance V_{12} , then the probability that the modulus of the difference between the true value and the estimate is less than $\sqrt{V_{12}}$, $2\sqrt{V_{12}}$, $3\sqrt{V_{12}}$ is 0.683, 0.954, 0.997 respectively.

We assume that the variance of each response is an unknown constant σ^2 , and that different responses are independent. This information can be summarized in the variance-covariance matrix for the vector Y , as $\text{Cov}(Y) = \sigma^2 I$. Now, the variance of the estimator $z^\top X^\top QY$ of $x^\top \tau$ is given by $z^\top X^\top Q(\text{Cov}(Y))QXz$. Under our simple assumptions about $\text{Cov}(Y)$, this reduces to $(z^\top Cz)\sigma^2$, which is $(x^\top C^{-1}x)\sigma^2$.

Any other linear estimator of $x^\top \tau$ must have the form $(z^\top X^\top + w^\top)QY$, for some w in \mathbb{R}^{bk} . For this estimator to be unbiased, we need $0 = E(w^\top QY) = w^\top Q E(Y) = w^\top QX\tau$ for all τ , and so $w^\top QX = 0$. Now, the variance of this estimator is

$$\begin{aligned} (z^\top X^\top + w^\top)Q(\text{Cov}(Y))Q(w + Xz) &= (z^\top X^\top + w^\top)Q(w + Xz)\sigma^2 \\ &= (z^\top Cz + w^\top Qw + 2w^\top QXz)\sigma^2 \\ &= (z^\top Cz + w^\top Qw)\sigma^2. \end{aligned}$$

As Q is positive semi-definite, $w^\top Q w$ cannot be negative, so this variance cannot be smaller than the variance of the least-squares estimator.

In summary, we have proved the following classical theorem.

Theorem 2.2 *Suppose that $\text{Cov}(Y) = I\sigma^2$ and that Equation (2.4) holds. If the design is connected and x is any contrast vector in \mathbb{R}^v , then the least-squares estimator of $x^\top \tau$ is $x^\top C^- X^\top Q Y$, where $Q = I - k^{-1} Z Z^\top$. This estimator is unbiased, and has minimum variance among all unbiased linear estimators of $x^\top \tau$. The variance is equal to $(x^\top C^- x)\sigma^2$.*

In particular, to estimate the difference $\tau_i - \tau_j$ we use the contrast x with $x_i = 1$, $x_j = -1$ and $x_l = 0$ for $l \notin \{i, j\}$. To calculate the variance V_{ij} of this estimator, which is called a *pairwise variance*, we look at the 2×2 submatrix of C^- in the i th and j -th row and column:

$$V_{ij} = (C_{ii}^- + C_{jj}^- - C_{ij}^- - C_{ji}^-)\sigma^2. \quad (2.6)$$

If C is completely symmetric, then these pairwise variances are constant. Such a design is called *variance-balanced*. Note that 2-designs are variance-balanced.

2.3 Optimality criteria

A statistician designing a comparative experiment knows the number v of treatments to be compared and the number of experimental units available. In our case, the experimental units are partitioned into b blocks of size k , where b and k are known. Let $\mathcal{D}(v, b, k)$ denote the class of all (not necessarily binary) block designs with v treatments, and b blocks of size k . Which design in this class is best?

One obviously desirable feature is to minimize the average value of the pairwise variances V_{ij} . Since C^- has zero row and column sums, Equation (2.6) shows that, for each fixed i ,

$$\sum_{j \neq i} V_{ij} = [(v-1)C_{ii}^- + (\text{Tr}(C^-) - C_{ii}^-) + 2C_{ii}^-]\sigma^2 = (\text{Tr}(C^-) + vC_{ii}^-)\sigma^2,$$

where $\text{Tr}(C)$ is the trace of C . Hence the average value \bar{V} of the pairwise variances is equal to $2\sigma^2 \text{Tr}(C^-)/(v-1)$.

Let $\theta_1, \dots, \theta_{v-1}$ be the non-trivial eigenvalues of C , now listed with multiplicities and in non-decreasing order. Then

$$\text{Tr}(C^-) = \frac{1}{\theta_1} + \dots + \frac{1}{\theta_{v-1}}.$$

Therefore

$$\bar{V} = \frac{2\sigma^2}{v-1} \left(\frac{1}{\theta_1} + \dots + \frac{1}{\theta_{v-1}} \right) = 2\sigma^2 \times \frac{1}{\text{harmonic mean of } \theta_1, \dots, \theta_{v-1}}.$$

Thus a design is said to be *A-optimal* (in some class) if it minimizes the average pairwise variance in that class; equivalently, if it maximizes the harmonic mean of $\theta_1, \dots, \theta_{v-1}$ in that class. Here ‘A’ stands for ‘average’.

This is not the only kind of optimality. When statisticians present an estimate for the difference $\tau_i - \tau_j$, they usually also present a confidence interval for the difference. If they always present a 95% confidence interval, the true value should lie within the confidence interval in 19 cases out of 20. The smaller the confidence interval the better. Under the assumption that the responses are multivariate normal, the length of this confidence interval is proportional to $\sqrt{V_{ij}}$, and therefore minimizing the variance is the same as minimizing the length of the confidence interval. However, the volume of the confidence ellipsoid for the estimate of (τ_1, \dots, τ_v) (in the $(v-1)$ -dimensional space orthogonal to u) is proportional to the square root of the reciprocal of $\prod_{i=1}^{v-1} \theta_i$ (see [4, page 49]). A design is *D-optimal* if it minimizes this volume; equivalently, if it maximizes the geometric mean of $\theta_1, \dots, \theta_{v-1}$. Here ‘D’ stands for ‘determinant’.

In some experiments, the treatments are combinations of different quantities of inputs that can be measured on continuous scales. A multivariate polynomial model may be fitted rather than the model in Equation (2.4). Since the scales for the different variables may not be comparable, it is important to use an optimality criterion that is invariant to reparametrization of the model. The D-criterion is the only one of the popular criteria that is so invariant. However, for block designs with v qualitative treatments, the A-criterion seems more natural.

Some people are concerned about worst cases. A design is said to be *E-optimal* if it maximizes the minimum θ_1 of the non-trivial eigenvalues of C ; equivalently, it minimizes the maximum variance of the estimator of $x^\top \tau$ over all contrast vectors x with $x^\top x = 1$. Here ‘E’ stands for ‘extreme’. Somewhat similarly, a design is said to be *MV-optimal* if it minimizes the maximum of the V_{ij} . This last criterion is not a function of the eigenvalues.

The eigenvalue optimality criteria all fall under the umbrella of Φ_p -optimality. For p in $(0, \infty)$, a design is Φ_p -optimal if it minimizes

$$\left(\frac{\sum_{i=1}^{v-1} \theta_i^{-p}}{v-1} \right)^{\frac{1}{p}}.$$

A-optimality corresponds to $p = 1$; the limit as $p \rightarrow 0$ gives D-optimality; the limit as $p \rightarrow \infty$ gives E-optimality.

All these criteria depend on the design only through its information matrix, which is a $v \times v$ real symmetric matrix with row and column sums zero. Let \mathcal{X}_v denote the set of all matrices satisfying this condition. An *optimality criterion* is a function $\Phi : \mathcal{X}_v \rightarrow \mathbb{R}$. Now a design in some class \mathcal{D} of designs is said to be Φ -optimal if it minimizes the value of $\Phi(C)$ over all information matrices C of designs in \mathcal{D} .

The optimality criteria Φ that we consider all satisfy the following conditions:

- (a) Φ is convex;
- (b) for every C in \mathcal{X}_v , the function $\alpha \mapsto \Phi(\alpha C)$ is monotonic non-increasing for non-negative α ;
- (c) Φ is invariant under any simultaneous permutation of rows and columns by the same permutation—that is, re-labelling the treatments does not affect Φ .

(Any function of the eigenvalues will satisfy condition (c).)

A design is said to be *universally optimal* in some class if it minimizes every Φ satisfying the above conditions. Thus universal optimality is stronger than all of the individual criteria, but there may not be any universally optimal design in a given class of designs. Recall that a design in $\mathcal{D}(v, b, k)$ maximizes the trace of C if and only if it is generalized binary. Kiefer [66] proved the following theorem.

Theorem 2.3 *Let $\Phi : \mathcal{X}_v \rightarrow \mathbb{R}$ be a function satisfying conditions (a)–(c) above. If there is a design in the class for which C is completely symmetric and has maximum trace, then it is universally optimal. In particular, if $\mathcal{D}(v, b, k)$ contains 2-designs (BIBDs), then the minimum value of $\Phi(C)$ over information matrices of designs in $\mathcal{D}(v, b, k)$ is attained by the 2-designs.*

The A-, D- and E-criteria are all examples of functions satisfying the conditions of Kiefer's theorem. This justifies the assertion that the Fano plane is the best design to choose in our introductory example. In terms of the parameters v, b, k , the necessary conditions (1.1) for the existence of a 2-design or BIBD read

$$\begin{aligned} v & \text{ divides } bk, \\ v - 1 & \text{ divides } bk(k - 1). \end{aligned}$$

We sketch the proof of Kiefer's Theorem for functions depending on the eigenvalues, in the case $k < v$.

Let $\theta_1, \dots, \theta_{v-1}$ be the non-trivial eigenvalues of the the information matrix of a design in $\mathcal{D}(v, b, k)$. Then $\sum \theta_i = \text{Tr}(C) \leq b(k - 1)$, by Proposition 2.1. Under the given hypotheses, $\Phi(C)$ is minimized when the sum attains the upper bound $b(k - 1)$ (which means that the design is binary) and all eigenvalues are equal (which means that C is a linear combination of I and J , so that the design is balanced).

We say that the sequence (x_0, \dots, x_{v-1}) is *upper weakly majorized* by the sequence (y_0, \dots, y_{v-1}) if

$$\sum_{i=0}^{n-1} x_i \geq \sum_{i=0}^{n-1} y_i$$

for $n = 1, \dots, v$. If the sums are equal for $n = v$, we say that (x_0, \dots, x_{v-1}) is *majorized* by (y_0, \dots, y_{v-1}) . See Marshall and Olkin [70]. Now we say that a block design is *Schur-optimal* in a class if the eigenvalues of its information matrix (in non-decreasing order) are upper weakly majorized by those of any competing design. Clearly Schur-optimality is stronger than E-optimality. This seems to make most sense when the sum of the eigenvalues is fixed. Since upper weak majorization is a partial order, there may be no such optimal design within a given class. Giovagnoli and Wynn [47] showed:

Proposition 2.4 *If a design is Schur-optimal within any class then it is also Φ_p -optimal for all p , in particular A-, D- and E-optimal.*

Of course, this is trivial for E-optimality.

An important property of Schur-optimality is given by the following result:

Proposition 2.5 *Suppose that M_1 and M_2 are real, symmetric matrices, and that $M_1 - M_2$ is positive semi-definite. Then the eigenvalues of M_1 are upper weakly majorized by those of M_2 .*

Proof Let M be any $v \times v$ real symmetric matrix. The smallest eigenvalue μ_0 of M is equal to the minimum of $x^\top Mx$ over all unit vectors x . More generally, the n th smallest eigenvalue μ_{n-1} of M is the minimum of $x^\top Mx$ over all unit vectors x orthogonal to the eigenvectors associated with the first $n - 1$ eigenvalues. It follows immediately that

$$\sum_{i=0}^{n-1} \mu_i = \min_{x_0, \dots, x_{n-1}} \sum_{i=0}^{n-1} x_i^\top Mx_i$$

for $n = 1, \dots, v$, where the minimum is over all n -tuples of pairwise orthogonal unit vectors.

Now adding a positive semi-definite matrix to M certainly cannot decrease the expression on the right-hand side. \square

It follows that, if C_1 and C_2 are the information matrices of two designs, and $C_1 - C_2$ is positive semi-definite, then C_1 beats C_2 on the Schur criterion and hence on all the Φ_p -optimality criteria.

Example 2.6 Let \mathfrak{D}_1 be the design in Example 1.1. Since any two distinct points have concurrence 2 in this design, the information matrix is $C_1 = (10I - 2J)/3$, with eigenvalues 0, 10/3, 10/3, 10/3, 10/3. We will compare \mathfrak{D}_1 with the binary design \mathfrak{D}_2 obtained by replacing the block [1, 1, 2] by [1, 2, 3]. (We use the notation $[a, \dots]$ to denote a multiset.) A simple computation shows that the information matrix

$$C_2 = \frac{1}{3} \begin{bmatrix} 8 & -1 & -3 & -2 & -2 \\ -1 & 8 & -3 & -2 & -2 \\ -3 & -3 & 10 & -2 & -2 \\ -2 & -2 & -2 & 8 & -2 \\ -2 & -2 & -2 & -2 & 8 \end{bmatrix}$$

of \mathfrak{D}_2 has eigenvalues 0, 3, 10/3, 10/3, 13/3. So \mathfrak{D}_1 beats \mathfrak{D}_2 on the E-criterion. (We will see later that \mathfrak{D}_1 is E-optimal.) However, a further short calculation shows that \mathfrak{D}_2 wins on the A- and D-criteria. Indeed, \mathfrak{D}_1 beats \mathfrak{D}_2 on the Φ_p -criterion if and only if

$$0.9^{-p} + 1.3^{-p} > 2,$$

which holds if and only if $p > 5.32652$ (approximately).

For book-length treatments of optimality, see [4, 82, 91].

2.4 The concurrence graph

We have seen that many optimality criteria of block designs are functions of the non-trivial eigenvalues of its information matrix C . We now give this a graph-theoretic interpretation.

The *concurrence graph* of a block design is defined as follows. The vertices are the points (or treatments); distinct vertices i and j are joined by λ_{ij} edges, where

λ_{ij} is the concurrence of i and j , that is, the number of blocks in which both i and j appear (counted appropriately: that is, λ_{ij} is the (i, j) entry of NN^\top , where N is the $v \times b$ incidence matrix of the design—see Section 1.3). Note that the concurrence graph contains no loops, even if the design is not binary.

The *Laplacian matrix* $L(G)$ of a loopless graph G on v vertices is defined as follows: for $i \neq j$, its (i, j) entry is the negative of the number of edges joining i and j ; the (i, i) entry is the valency of i (the number of edges incident with i). In other words, $L(G) = D - A(G)$, where D is a diagonal matrix whose diagonal entries are the valencies, and $A(G)$ is the adjacency matrix of G . Note that $L(G)$ is a symmetric matrix with row and column sums zero.

Proposition 2.7 *If a block design has constant block size k (but is not necessarily binary) then its information matrix C is obtained by dividing the Laplacian matrix of its concurrence graph by k .*

Proof We know that $C = R - NN^\top/k$, where R is diagonal. Now NN^\top and $A(G)$ agree at all off-diagonal elements; hence C and $L(G)/k$ agree in all off-diagonal elements. But both C and $L(G)/k$ have row and column sums zero; so $C = L(G)/k$, as required. \square

This result, pointed out by Cheng [32] and cited in [91, page 30], has the important consequence that the optimality properties of a design with constant block size k are completely determined by k and its concurrence graph. Given a graph G , there may be many designs which have G as their concurrence graph; so we have potentially made a big simplification to the problem.

Which graphs are concurrence graphs of designs with block size k ? The next result answers this question.

Here is a non-standard definition. A *weighted clique* with weights w_1, \dots, w_m (where w_1, \dots, w_m are positive integers) has m vertices, numbered $1, \dots, m$, and has $w_i w_j$ edges joining vertex i to vertex j for $i \neq j$. Its *weight* is the sum of the numbers w_1, \dots, w_m . If all w_i are 1, this is just a complete graph on m vertices, and has weight m .

Proposition 2.8 *Let G be a graph. Then G is the concurrence graph of a block design with block size k if and only if the edge set of G can be partitioned into weighted cliques of weight k . In particular, G is the concurrence graph of a binary block design with block size k if and only if the edge set of G can be partitioned into complete subgraphs of size k .*

Proof From a weighted clique of weight k we can reconstruct a block, in which treatment i occurs w_i times; the concurrences between distinct treatments in this block give precisely the edges of the weighted clique. The converse is clear. \square

Remark A given graph may have many such partitions, or none. But even the partition does not uniquely determine the block design. In our running example, two edges joining the vertices 1 and 2 form a weighted clique with weights $(1, 2)$ or $(2, 1)$, and so can arise from either the block $[1, 1, 2]$ or the block $[1, 2, 2]$. Since all

optimality properties are determined by the concurrence graph, switching $[1, 1, 2]$ into $[1, 2, 2]$ preserves all such properties.

However, in the special case $k = 2$, then any graph satisfies the conditions: the blocks are the edges of the graph. (Our assumption that there is no block containing only a single treatment guarantees that every block is an edge of the graph rather than a “loop”.)

2.5 Laplace eigenvalues

Proposition 2.7 draws our attention to the study of the Laplacian eigenvalues of a graph, and their relationship to graph properties. In this section we give a brief survey of the properties of the Laplace eigenvalues of graphs. Although we speak only of graphs, the results have analogues for edge-weighted graphs (with non-negative weights). For further information and proofs we refer to Mohar’s article [72].

The Laplacian matrix is so-called because of its relationship with the Laplacian differential operator $-\nabla^2$ on a manifold. If the manifold is triangulated, then the resulting graph (with its edges weighted by a function of their lengths) is a “discrete approximation” to the manifold, and its Laplacian matrix is an approximation to the Laplacian operator of the manifold. Note that eigenvalues and eigenvectors of $-\nabla^2$ are the solutions of the partial differential equation $\nabla^2 f + \mu f = 0$; the eigenfunctions f are “vibrational modes” of the manifold, and the eigenvalues are related the vibrational frequencies. This led to Mark Kac’s famous question, “Can one hear the shape of a drum?” [61], which asks if it is possible for different manifolds or graphs to have the same Laplace eigenvalues. Not surprisingly, the answer, in both the continuous and the discrete cases, is that it is possible.

Let G be a graph with v vertices and e edges; loops are forbidden, but multiple edges are allowed except where we say otherwise. Let $d(i)$ be the valency of vertex i , and let $\delta(G)$ and $\Delta(G)$ be the smallest and largest values of $d(i)$. Recall that the Laplacian matrix $L(G)$ is the $v \times v$ matrix with (i, i) entry $d(i)$ and (i, j) entry minus the number of edges from i to j .

Note that the Laplacian matrix is the sum over edges of matrices of the following form. Each edge between vertices i and j contributes the matrix with submatrix

$$\begin{array}{cc} & \begin{array}{cc} i & j \end{array} \\ \begin{array}{c} i \\ j \end{array} & \left[\begin{array}{cc} 1 & -1 \\ -1 & 1 \end{array} \right] \end{array},$$

all other entries being zero. Thus the matrix for each edge is positive semi-definite. This gives a rather easy proof of the following facts:

Proposition 2.9 (a) *The Laplacian matrix $L(G)$ is positive semi-definite (hence so is any information matrix).*

(b) *For any vector x in \mathbb{R}^v , we have*

$$x^\top L(G)x = \sum_{\text{edges } ij} (x_i - x_j)^2.$$

(c) *The nullity of $L(G)$ is equal to the number of connected components of G . In particular, if G is connected, the all-1 vector u spans the null space.*

(d) If the graph G_2 is obtained from the graph G_1 by inserting an extra edge, then $L(G_2) - L(G_1)$ is positive semi-definite: so that we can never make a design worse on the Schur criterion (and hence any Φ_p criterion) by adding an edge to the graph (Propositions 2.4 and 2.5).

Here is another popular method of proving Proposition 2.9. Choose an arbitrary orientation of the edges of G , and (temporarily) let N be the signed vertex-edge incidence matrix of G , the $v \times e$ matrix whose (i, k) entry is

$$n_{ik} = \begin{cases} +1 & \text{if vertex } i \text{ is the head of the } k\text{th directed edge } f_k, \\ -1 & \text{if vertex } i \text{ is the tail of } f_k, \\ 0 & \text{otherwise.} \end{cases}$$

Then we have

$$NN^\top = L(G),$$

for each edge containing vertex i contributes $+1$ to the (i, i) entry of NN^\top , while each edge joining vertices i and j contributes -1 to the (i, j) entry. This shows part (a) of Proposition 2.9. Part (b), giving the particularly simple form of the quadratic form associated with $L(G)$, is easily proved using the fact that

$$x^\top L(G)x = (N^\top x)^\top (N^\top x).$$

From now on we assume that G is connected. Let μ_0, \dots, μ_{v-1} be the eigenvalues of $L(G)$, arranged in non-decreasing order, so that $\mu_0 = 0$ and $\mu_1 > 0$. (Note that this convention is different from Mohar's [72], who uses μ_1, \dots, μ_v .) We often write μ_{\max} instead of μ_{v-1} for the greatest eigenvalue. Thus we have

$$0 = \mu_0 < \mu_1 \leq \dots \leq \mu_{\max}.$$

The sum of the Laplace eigenvalues is the trace of $L(G)$, the sum of the valencies of the vertices, which is twice the number of edges; so we have

$$\mu_1 \leq 2e/(v-1) \leq \mu_{\max}.$$

These inequalities can be refined; for example, μ_1 is bounded above by $(d(i)+d(j))/2$, for any two non-adjacent vertices i and j ; and

$$\mu_1 \leq v\delta(G)/(v-1) \leq v\Delta(G)/(v-1) \leq \mu_{\max} \leq 2\Delta(G);$$

and $\mu_{\max} = 2\Delta(G)$ if and only if G is bipartite and regular. We will prove the first inequality below; for the rest, see [72].

The Laplace eigenvalues of a simple graph and its complement have a simple relation:

Proposition 2.10 *Let G be a simple graph on v vertices and \bar{G} its complement. Then $\mu_0(G) = \mu_0(\bar{G}) = 0$ and*

$$\mu_i(\bar{G}) = v - \mu_{v-i}(G)$$

for $i = 1, \dots, v-1$.

This is because $L(G) + L(\overline{G}) = vI - J$, where J is the all-1 matrix, and the all-1 vector u is the zero eigenvector for both $L(G)$ and $L(\overline{G})$, so that any other common eigenvector is perpendicular to u .

If G is regular of valency d , then $L(G) = dI - A(G)$, where $A(G)$ is the usual adjacency matrix of G , so the Laplace eigenvalues have the form $d - \alpha$, where α runs over the eigenvalues of $A(G)$.

One graph parameter closely connected with the Laplace eigenvalues is the isoperimetric number. A “good” graph (for use as a network, say) will be one without bottlenecks: any set of vertices will have many edges joining it to its complement. So, for a set S of vertices, we let $\partial(S)$ (the *boundary* of S) be the set of edges which have one vertex in S and the other in its complement, and then define the *isoperimetric number* $i(G)$ by

$$i(G) = \min \left\{ \frac{|\partial S|}{|S|} : S \subseteq V(G), 0 < |S| \leq v/2 \right\}.$$

(We need only take the size of S up to $v/2$ since a set and its complement have the same boundary.) An infinite class of simple graphs is a *family of expanders* if $i(G)$ is bounded away from zero for graphs in this class.

The following result is due to Alon and Milman [3] and Dodziuk [41].

Theorem 2.11 *Let μ_1 be the second Laplace eigenvalue of G , and $\Delta = \Delta(G)$. Then*

$$\mu_1/2 \leq i(G) \leq \sqrt{(2\Delta - \mu_1)\mu_1}.$$

We will prove the left-hand inequality, demonstrating that a graph with large μ_1 has good expansion properties. In fact we show a slightly stronger result, which will be used a number of times.

Proposition 2.12 *Let G have an edge-cutset of size c whose removal separates the graph into components of sizes m and n . Then $\mu_1 \leq c((1/m) + (1/n))$.*

Proof The second eigenvalue μ_1 is the minimum of $x^\top Lx/x^\top x$ over all vectors x perpendicular to the all-1 vector u (which is the eigenvector for the smallest eigenvalue $\mu_0 = 0$).

Now let x be the vector taking the value n on vertices on the side of the cut with size m and the value $-m$ on vertices on the other side. Then x has coordinate sum zero, so is perpendicular to the all-1 vector u . By Proposition 2.9(b), we have $x^\top Lx = c(m+n)^2$, whereas $x^\top x = mn^2 + nm^2 = mn(m+n)$. So $\mu_1 \leq c(m+n)/mn = c((1/m) + (1/n))$. \square

Now the left-hand inequality in the theorem follows: we can choose the cutset such that $c/m = i(G)$ and $m \leq n$; then $\mu_1 \leq 2c/m = 2i(G)$.

The right-hand inequality is one of a family of results called *inequalities of Cheeger type*, since they are analogues of Cheeger’s inequality in the continuous case. The analogue of Cheeger’s inequality would be the assertion $i(G) \leq \sqrt{2\Delta\mu_1}$; see [38, 72] for the proof of the stronger result quoted here and references to similar results.

Corollary 2.13 *An infinite class of simple graphs is a family of expanders if and only if μ_1 is bounded away from zero for the graphs in the class.*

The importance of expanders in computer science has resulted in a great deal of attention being paid to the graph parameter μ_1 . It is also very closely connected with the rate of convergence of a random walk on the graph.

2.6 Variance-balanced designs

Recall that a block design is variance-balanced (or VB, for short) if its information matrix is completely symmetric. Morgan and Srivastav [74] call such a design a *completely symmetric design*, or CSD; because of conflict with the term “symmetric design”, we have avoided this. We denote by $\text{VB}(v, k, \lambda)$ a variance-balanced design with v points, block size k and all pairwise concurrences λ . (Remember that they are not assumed to be binary!)

Note that VB designs are not always optimal:

Example 2.14 Two designs with $v = b = 7$, $k = 6$ can be constructed as follows:

- (a) the design whose blocks are all 6-subsets of the set of points;
- (b) the design obtained from the Fano plane by doubling each occurrence of a point in a block (so that the first block is the multiset $[1, 1, 2, 2, 3, 3]$).

Both these designs are VB; the first has $\lambda = 5$, the second has $\lambda = 4$. The first, being a 2-design, is better, according to Kiefer’s theorem.

Question 2.15 *Given k and λ , for which values of v do $\text{VB}(v, k, \lambda)$ designs exist, and what are the possible numbers of blocks of such designs?*

An asymptotic answer to this problem (that is, conditions which are necessary and sufficient for large enough v) would generalize Wilson’s existence theorem for 2-designs [106] described earlier. Two surveys of this class of designs have been given by Billington [18, 19].

Morgan and Srivastav define two new parameters of a VB design, as follows:

$$r = \left\lfloor \frac{bk}{v} \right\rfloor, \quad p = bk - vr,$$

so that $bk = vr + p$ and $0 \leq p \leq v - 1$. Thus, in a 2-design we have $p = 0$. Note that the use of r does not here imply that the design has constant replication!

Morgan and Srivastav further say that a VB design has *maximum trace* if its parameters satisfy the equation

$$r(k - 1) = (v - 1)\lambda. \tag{2.7}$$

The reason for the term is as follows. Since $bk < v(r + 1)$, some treatment occurs at most r times on the bk plots. Each occurrence contributes at most $k - 1$ edges to the concurrence graph, so the valency of this vertex is at most $r(k - 1)$. But the concurrence graph of a VB design is regular, with valency $(v - 1)\lambda$; so we have

$(v-1)\lambda \leq r(k-1)$, and the trace of the information matrix, which is $v(v-1)\lambda$, is at most $vr(k-1)$; equality for the trace implies that $(v-1)\lambda = r(k-1)$.

The maximum trace condition does imply some nice properties.

The above argument shows that, in a VB design of maximum trace, any point lies in at least r blocks (counted with multiplicity), with equality if and only if the point occurs at most once in each block. Since $bk = vr + p$, it follows that the number of “bad” points (which occur more than once in some block) is at most p . So if $p = 0$, the design is binary, and is a BIBD or 2-design.

In Example 2.14, we have $r = 6$, $p = 0$, and so the first design has maximum trace but the second does not.

Morgan and Srivastav [74] show that a VB design with maximum trace is E-optimal. We refer to their paper for the proof, but give here a complementary result in which maximum trace is not assumed but we require that the effect of non-binary blocks is not too large. A binary block (one in which no treatment is repeated) gives rise to $k(k-1)/2$ edges of the concurrence graph; define the *defect* of a non-binary block to be the difference between $k(k-1)/2$ and the number of edges it contributes.

Proposition 2.16 *If $k < v$, then a VB design on v points is E-optimal if the sum of defects of non-binary blocks is less than $v/2$.*

Proof Let x be the sum of defects. Then the number of edges (which we know to be $\lambda v(v-1)/2$) is $bk(k-1)/2 - x$, so that

$$b = \frac{\lambda v(v-1) + 2x}{k(k-1)}.$$

It is easily verified that the non-trivial Laplacian eigenvalues of the λ -fold complete graph are all equal to λv . So, if our design is not E-optimal, then a E-better design (with the same values of (v, b, k)) has $\mu_1 > \lambda v$. Let δ be the minimal degree of the concurrence graph of such a design. By Proposition 2.12,

$$\lambda v < \mu_1 \leq \delta(1 + 1/(v-1)) = \delta v/(v-1),$$

so that $\delta > \lambda(v-1)$, or $\delta \geq \lambda(v-1) + 1$. Hence the concurrence graph has at least $v(\lambda(v-1) + 1)/2$ edges. Since each block of this design contributes at most $k(k-1)/2$ edges, we have

$$\frac{\lambda v(v-1) + 2x}{k(k-1)} = b \geq \frac{v(\lambda(v-1) + 1)}{k(k-1)},$$

whence $x \geq v/2$. So, if $x < v/2$, then no E-better design can exist. \square

For $k = 3$, the defect of a non-binary block is 1, so a design with fewer than $v/2$ non-binary blocks is E-optimal. It follows that the design of Example 1.1 is E-optimal, as claimed earlier.

We now turn to constructions.

If we have two VB designs on the same set of v points with the same block size k , having parameters λ_1 and λ_2 , then the multiset union of the block multisets is again VB, with parameter $\lambda_1 + \lambda_2$. The new design is not necessarily of maximum

trace; but it is so if one of the VB designs we start with is a 2-design and the other is of maximum trace, or if the sum of their p parameters is less than v .

For example, suppose that $k = 3$. A VB design of maximum trace satisfies $2r = (v - 1)\lambda$, so that λ is even or v is odd. Moreover, $\lambda = 1$ is impossible (except for $2-(v, 3, 1)$ designs), since a non-binary block gives concurrence at least 2. Morgan and Srivastav [74] prove that these necessary conditions are sufficient:

Theorem 2.17 *A $VB(v, 3, \lambda)$ design of maximum trace exists whenever $\lambda(v - 1)$ is even and $\lambda > 1$.*

Proof Since a $2-(v, 3, 6)$ design exists for all v [51], it is enough to settle the existence question for λ in a complete set of non-zero residues mod 6. Now 2-designs exist in the following cases [51]:

- for $\lambda = 1$ or 5, if $v \equiv 1$ or 3 mod 6;
- for $\lambda = 2$ or 4, if $v \equiv 0$ or 1 mod 3;
- for $\lambda = 3$, if v is odd.

We give a construction of VB designs for $\lambda = 2$ and $v \equiv 2$ mod 3 below; they have $p = 1$, so the union of two copies settles $\lambda = 4$. For $\lambda = 5$ or $\lambda = 7$, with v odd, there is a 2-design unless $v \equiv 5$ mod 6; in that case we can take a 2-design with $\lambda = 3$ and a VB design with $\lambda = 2$ or $\lambda = 4$. \square

We now look further at the case $k = 3$, $\lambda = 2$. A block consisting of three points is a triangle in the concurrence graph, while a block consisting of two points (one with multiplicity 2) is a double edge. So we want to partition the edge set of doubled K_v into triangles and double edges. Clearly if a $2-(v, 3, 2)$ design exists, then this gives a VB design with the smallest number of blocks; this design exists if and only if $v \not\equiv 2$ mod 3. There is a “boring” design obtained by partitioning the edge set into double edges, which has the largest number of blocks.

The smallest case for which no 2-design exists is $v = 5$. There are four possibilities, listed in order of increasing numbers of blocks (and perhaps decreasing interest):

- Six triangles and one double edge: this is realised by the design in our running example (Example 1.1).
- Four triangles and four double edges: take a $2-(4, 3, 2)$ design (consisting of all the 3-subsets of a 4-set) and join its four points to the fifth point by four double edges.
- Two triangles and seven double edges: take a triangle twice and double the seven uncovered edges.
- Ten double edges: this is the “boring” design.

The values of (r, p) in the four cases are $(4, 1)$, $(4, 4)$, $(5, 2)$ and $(6, 0)$. So the first two have maximum trace but the others do not.

Here is a construction for $\text{VB}(v, 3, 2)$ designs having just one non-binary block. In this case, as we have seen, we must have $v \equiv 2 \pmod{3}$. Our construction uses Steiner triple systems; since there are many of these ([107]), the construction is “prolific”, producing $v^{(1+o(1))v^2/3}$ non-isomorphic designs.

Suppose first that $v \equiv 2 \pmod{6}$. In this case, there exist Steiner triple systems of orders $v \pm 1$. Take two such systems, on the point sets $\{1, \dots, v+1\}$ and $\{1, \dots, v-1\}$ respectively; let the sets of blocks be \mathcal{B}_1 and \mathcal{B}_2 . Without loss of generality, suppose that the third point of the block B of \mathcal{B}_1 containing v and $v+1$ is $v-1$.

Now we take the point set of the new design to be $\{1, \dots, v\}$. For the blocks, we first remove the block B from \mathcal{B}_1 ; then we replace each occurrence of $v+1$ in any other block with v ; the resulting blocks together with $[v-1, v-1, v]$ make up the design.

We have to check that $\{v-1, v\}$ lies only in $[v-1, v-1, v]$, while every other pair $\{i, j\}$ lies in two blocks. For the first, note that the only other candidate, namely B , has been removed. For the second, there are two cases:

- $j = v, i \neq v-1$: in \mathcal{B}_1 , there is one block containing i and v , and one containing i and $v+1$ (in which $v+1$ is replaced by v). No block of \mathcal{B}_2 can occur.
- $v \notin \{i, j\}$: one block of \mathcal{B}_1 and one of \mathcal{B}_2 contain $\{i, j\}$, and these two points are unchanged in these blocks.

There is a similar construction when $v \equiv 5 \pmod{6}$. In this case, both $v-2$ and $v+2$ are orders of Steiner triple systems; and, if $v \geq 17$, we may assume that the larger system contains a subsystem of order 7. Choose the numbering so that this system uses the points a, \dots, g , with blocks $abc, ade, afg, bdg, bef, cdf, ceg$, and abc is a block of the smaller system, while d, e, f, g are not in this system. Now:

- Delete these eight blocks from the two systems.
- Any block remaining contains at most one of d, e, f, g . Replace the point e by d and the point g by f everywhere.
- Add new blocks $bdf, cdf, abd, acd, abf, acf$ and bbc .

We leave it to the diligent reader to check that this construction works: only bbc contains b and c , while every other pair lies in two blocks. Note that it gives $(v+2)(v+1)/6 + (v-2)(v-3)/6 - 1 = (v^2 - v + 1)/3$ blocks, the correct number.

For $v = 11$, we can ask the computer to produce a design; there are plenty to choose from (see Section 6.1).

Question 2.18 *For which v, λ does a $\text{VB}(v, 3, \lambda)$ design without repeated blocks exist?*

This question is of no statistical significance; but, as explained earlier, it may be taken as a challenge by combinatorialists. More generally one could ask for the possible numbers of distinct blocks in such a design.

3 Optimal designs

In this section we first state some general results on optimality, and then look more closely at the three most important optimality criteria: D, A and E. We will see that these are related to important graph-theoretic notions for the concurrence graph: number of spanning trees, resistance as an electrical network, and expander properties.

3.1 General results

There are a few general results, showing that certain kinds of designs (if they exist) are optimal for a wide variety of block designs. The first such result (Theorem 2.3) was due to Kiefer [66], and shows that 2-designs are optimal (if they exist).

If a 2-design with the given parameters does not exist, then this theorem gives no information. An interesting question is: What happens if the necessary conditions (1.1) are satisfied but there is no design? A special case of this occurs for the values $v = 15$, $b = 21$, $k = 5$. (If a 2-design with these parameters existed, then by the Hall–Connor theorem [50] it would be the block residual of a 2-(22, 7, 2) design; but such a design does not exist, by the Bruck–Ryser–Chowla theorem [24, 36].) This parameter set has been examined by Reck and Morgan [85, 86], and the A-, D- and E-optimal designs determined.

Some extensions of Kiefer’s Theorem are known. In order to state them, we require some more definitions.

A binary equireplicate design is said to be *group-divisible* if the treatments can be divided into “groups”, all of the same size, such that two treatments in the same group are contained in λ_1 blocks while two treatments in different groups are contained in λ_2 blocks. Group-divisible designs are partially balanced with respect to the group-divisible association scheme defined earlier.

A *regular-graph design* is a binary equireplicate design in which any pair of vertices is contained in either λ or $\lambda + 1$ blocks, for some λ [58]. The name arises because its concurrence graph consists of the λ -fold complete graph together with a simple regular graph.

Now Chêng [30, 31] proved:

Theorem 3.1 *Let f be any non-negative real function which is strictly decreasing and strictly convex, such that its derivative is strictly concave and $\lim_{x \rightarrow 0^+} f(x) = \infty$. For $X \in \mathcal{X}_v$, set $\Phi(X) = \sum f(\mu_i)$, where the μ_i run over the eigenvalues of X having eigenvector orthogonal to the all-1 vector.*

- (a) *If $\mathcal{D}(v, b, k)$ contains a group-divisible design with two groups and $\lambda_2 = \lambda_1 + 1$, then the minimum of $\Phi(C)$ over information matrices C of designs in $\mathcal{D}(v, b, k)$ is attained by such a design.*
- (b) *If $\mathcal{D}(v, b, k)$ contains a group-divisible design with $\lambda_2 = \lambda_1 + 1$, then the minimum of $\Phi(C)$ over information matrices C of regular-graph designs in $\mathcal{D}(v, b, k)$ is attained by such a design.*

Again the result applies for A-optimality (with $f(x) = 1/x$) and D-optimality (with $f(x) = -\log x$). For E-optimality, we can take $f(x) = x^{-p}$ for $p > 0$ and let $p \rightarrow \infty$.

Many optimality results limit the class of competing designs to be binary and equireplicate. For example, Cheng and Bailey [34] proved the following.

Theorem 3.2 *Let f and Φ be as in Theorem 3.1. If there is a regular-graph design in $\mathcal{D}(v, b, k)$ for which (a) the graph in question is strongly regular and (b) the concurrence matrix Λ is singular, then this design minimizes $\Phi(C)$ over all binary equireplicate designs in $\mathcal{D}(v, n, k)$.*

For example, the generalized quadrangle with $v = b = 15$ and $k = 3$ is A-, D- and E-optimal among binary equireplicate designs of this size. The points (treatments) are all unordered pairs from a set of size six, and the blocks are all the partitions of that set into three pairs. A pair is in a given block if it is a part of the corresponding partition.

This result has been extended in [13].

For further general results on optimality we refer to [82, 91].

3.1.1 Equireplicate designs and their duals If design \mathfrak{D} in $\mathcal{D}(v, b, k)$ is equireplicate with replication r and incidence matrix N , then its information matrix C is $rI - k^{-1}NN^T$, whose eigenvalues lie in $[0, r]$. Statisticians usually compare such a design with the hypothetical one that would be possible in a single large block of vr plots if enough similar plots were available for the experiment. The information matrix for this design is $r(I - v^{-1}J)$, all of whose non-trivial eigenvalues are equal to r . Normalizing by this, the non-trivial eigenvalues of C are divided by r to give what are called the *canonical efficiency factors*, which lie in $(0, 1]$ for connected designs. So long as we are considering only equireplicate designs, the eigenvalues of C can be replaced by the canonical efficiency factors in all the standard optimality criteria.

Interchanging the roles of treatments and blocks in \mathfrak{D} gives an equireplicate design \mathfrak{D}^* in $\mathcal{D}(b, v, r)$, which is called the *dual design*. Its incidence matrix is N^T and its information matrix C^* is $kI - r^{-1}N^T N$, with eigenvalues in $[0, k]$. The nonzero eigenvalues of $N^T N$ and NN^T are the same, including multiplicities, so the canonical efficiency factors of a design and its dual are the same, apart from $|b - v|$ occurrences of the maximum value 1. This observation gives an easy proof of the following result [91, p. 28]:

Theorem 3.3 *On all the standard optimality criteria, a given equireplicate design is optimal among equireplicate designs if and only if its dual design is optimal among equireplicate designs.*

3.1.2 Operations on designs The following elementary lemma, whose proof we omit, allows the non-trivial eigenvalues of various designs obtained from a given one to be calculated.

Proposition 3.4 *Let \mathfrak{D}_1 and \mathfrak{D}_2 be designs on the same set of v points. Suppose that, for $n = 1, 2$, the design \mathfrak{D}_n has constant block size k_n . Suppose further that the*

concurrences $\lambda_{ij}^{(n)}$ satisfy

$$\lambda_{ij}^{(2)} = \alpha + \beta \lambda_{ij}^{(1)}$$

for $i \neq j$. Then the non-trivial eigenvalues $\theta_m^{(n)}$ of the information matrices are related by

$$\theta_m^{(2)} = \begin{cases} \frac{\alpha v}{k_2} + \frac{\beta k_1}{k_2} \theta_m^{(1)} & \text{if } \beta \geq 0, \\ \frac{\alpha v}{k_2} + \frac{\beta k_1}{k_2} \theta_{v-1-m}^{(1)} & \text{if } \beta < 0. \end{cases}$$

In particular, suppose designs \mathfrak{D}'_1 and \mathfrak{D}'_2 also have v points and block-sizes k_1 and k_2 and their concurrences stand in the same relation, with the same values of α and β . If $\beta < 0$, suppose also that the information matrices of \mathfrak{D}_1 and \mathfrak{D}'_1 have the same trace. Then \mathfrak{D}_1 beats \mathfrak{D}'_1 on the Schur criterion if and only if \mathfrak{D}_2 beats \mathfrak{D}'_2 on the Schur criterion. If \mathfrak{D}_1 is E -better than \mathfrak{D}'_1 , and $\beta > 0$, then \mathfrak{D}_2 is E -better than \mathfrak{D}'_2 .

In general, affine transformations of the form given in this Proposition do not preserve optimality criteria. We look at three special cases.

The complement of the block set Let \mathfrak{D}_1 be a simple binary design with constant block size k , and let \mathfrak{D}_2 be the design whose blocks are all k -sets which are *not* blocks of \mathfrak{D}_1 . (If $k = 2$, this is the complement of the graph \mathfrak{D}_1 .) Then the conditions of Proposition 3.4 hold, with $k_1 = k_2 = k$ and $\alpha = \binom{v-2}{k-2}$, $\beta = -1$.

The complementary design, whose blocks are the complements of the blocks of the original Suppose that \mathfrak{D}_1 is binary and equireplicate, with block size k and replication r , and let the blocks of \mathfrak{D}_2 be the complements of those of \mathfrak{D}_1 . A simple inclusion-exclusion argument shows that the conditions of Proposition 3.4 hold, with $k_1 = k$, $k_2 = v - k$, $\alpha = b - 2r$, $\beta = 1$. See also [59].

The design formed by adjoining all points to a block Let \mathfrak{D}_1 be as in the preceding paragraph. Form a new design \mathfrak{D}_2 with block size $v + k$ by adding one copy of each point to each block. Now the contributions to $\lambda_{ij}^{(2)}$ from a block B are

$$\begin{cases} 4 & \text{if } i, j \in B, \\ 2 & \text{if } i \in B, j \notin B \text{ or vice versa,} \\ 1 & \text{if } i, j \notin B. \end{cases}$$

Hence the conditions of Proposition 3.4 hold, with $k_1 = k$, $k_2 = v + k$, $\alpha = b + 2r$, and $\beta = 1$.

John and Williams [59] conjectured that, if \mathfrak{D}_1 is A -optimal, then so is \mathfrak{D}_2 . Example 2.2 of Jacroux and Whittinghill [55] shows that this is not true in general.

Example 3.5 The Petersen graph is A -optimal among equireplicate designs in $\mathcal{D}(10, 15, 2)$ but John and Mitchell [58] found by computer search that its complement (the *triangular graph* $T(5)$, the line graph of K_5) is not A -optimal. The

Laplace eigenvalues of $T(5)$ are 5 (four times) and 8 (five times), see Example 3.11; John and Mitchell found that the A-optimal equireplicate design in $\mathcal{D}(10, 30, 2)$ consists of a the bipartite graph $K_{4,6}$ into which edges have been inserted to make a 6-cycle on the larger class of the bipartition. This graph has Laplace eigenvalues 5 (twice), 6 (three times), 7 (twice), 8 and 10; so it beats $T(5)$ on the D-criterion as well as the A-criterion, and they are equal on the E-criterion. The two graphs are not Schur comparable.

3.2 D-optimality: spanning trees

D-optimality is closely connected with maximizing the number of spanning trees of the concurrence graph [31, 45]. This is a consequence of Kirchhoff's Matrix-Tree Theorem.

The Matrix-Tree Theorem asserts:

Theorem 3.6 *Let G be a graph with v vertices. Then the following are equal:*

- (a) *the number of spanning trees of G ;*
- (b) *any $(v - 1) \times (v - 1)$ cofactor of the Laplacian matrix $L(G)$;*
- (c) *the product of the Laplacian eigenvalues μ_1, \dots, μ_{v-1} , divided by v .*

The fact that if a matrix has row and column sums zero then all its cofactors are equal, and the expression in terms of the non-trivial eigenvalues, are much more general; we give the simple proof at the end of this section.

Hence a block design with block size k on v vertices is D-optimal in some class if and only if its concurrence graph maximizes the number of spanning trees over this class.

For example, suppose that we are considering graphs with v vertices and $\lambda v(v - 1)/2$ edges. (This applies if we want a binary block design with v treatments and blocks of size k , where $bk(k - 1) = \lambda v(v - 1)$, for example.) By Kiefer's theorem, if a 2-design exists, then it is optimal. This is certainly the case if $k = 2$. So, for example, the λ -fold multiple of the complete graph has the largest number of spanning trees of all multigraphs with v vertices and $\lambda v(v - 1)/2$ edges, as we noted in the Introduction.

The number of spanning trees of a graph is a topic with a large literature. It was given extra impetus by a conjecture of Merino and Welsh [71], asserting that the number of spanning trees is bounded above by the maximum of the number of acyclic orientations and the number of totally cyclic orientations. A recent investigation [28] suggests that, for sparse graphs, the number of acyclic orientations is larger than the number of totally cyclic orientations, with the reverse being true for dense graphs.

In network theory, a simple graph is called *t-optimal* if it maximizes the number of spanning trees among simple graphs with given number of vertices and edges. Petingi and Rodriguez [80] extend the results in [31], but do not mention [35]. Of course, for D-optimality there is no restriction to simple graphs.

We mention here also that the number of spanning trees of a graph G is an evaluation of the Tutte polynomial of G ; in fact, it is $T(G; 1, 1)$. See [76, 94] for more information about the Tutte polynomial of a graph or matroid. The other

optimality criteria are not associated with evaluations of the Tutte polynomial. For example, all trees on n vertices have Tutte polynomial x^n ; we will see in Section 4 that they differ on the A- and E-criteria.

In the other direction, graphs with the same Laplace eigenvalues may have different Tutte polynomials. For example, the graph $L_2(4)$ (the line graph of the complete bipartite graph $K_{4,4}$) and the Shrikhande graph are strongly regular graphs with the same eigenvalues, but the former has 576 proper 4-colourings (these are just the Latin squares of order 4) while the latter has 240.

3.2.1 Matrices with row and column sums zero It may be folklore that if a square matrix has all row and column sums zero, then all its cofactors are equal. (This comes up in the Matrix-Tree Theorem.) Here is a simple proof of a generalization to arbitrary square singular matrices, which works even in non-zero characteristic with some restriction. For any (column) vector x with components x_1, \dots, x_v , set $\|x\| = \sum x_i^2$.

Proposition 3.7 *Let A be a square matrix over a field F . Suppose that there exist vectors x and z satisfying*

- $\|x\|$ and $\|z\|$ are non-zero (in F);
- $x^\top A = 0$ and $Az = 0$.

Then there is a constant d such that the (i, j) cofactor of A is equal to $x_i dz_j$.

Proof We prove the result for the $(1, 1)$ cofactor for convenience; the argument is the same for any cofactor.

Consider the matrix $A' = A + x^\top z$, with (i, j) entry $a_{ij} + x_i z_j$. Perform the following row and column operations on it:

- Let r_i be the i th row of A' . Replace r_1 by $\sum_i x_i r_i$, leaving the other entries unaltered. As a result, the $(1, j)$ entry becomes

$$\sum_i x_i (a_{ij} + x_i z_j) = \|x\| z_j,$$

while elements in rows other than the first are unaltered. The determinant is multiplied by x_1 .

- Let c_j be the j th column of this matrix. Replace c_1 by $\sum_j c_j z_j$, leaving the other entries unaltered. As a result, the $(1, 1)$ entry becomes

$$\sum_j \|x\| z_j^2 = \|x\| \cdot \|z\|;$$

for $i > 1$, the $(i, 1)$ entry becomes

$$\sum_j (a_{ij} + x_i z_j) z_j = x_i \|z\|,$$

and the other elements are unaltered. The determinant is multiplied by z_1 .

- Subtract $x_i/\|x\|$ times the first row from the i th row, for all $i > 1$. For $i > 1$, the $(i, 1)$ entry becomes 0, while for $i, j > 1$, the (i, j) entry becomes

$$a_{ij} + x_i z_j - (x_i/\|x\|)\|x\|z_j = a_{ij}.$$

The determinant is unaltered by this move.

Since we have a column (the first) in which all but one entry is zero, this entry being $\|x\| \cdot \|z\|$, and the entries in rows and columns different from the first are identical with those of A , we see that

$$x_1 z_1 \det(A') = \|x\| \cdot \|z\| \cdot A_{11},$$

where A_{11} is the $(1, 1)$ cofactor. So the Proposition is proved, with

$$d = \det(A')/(\|x\| \cdot \|z\|).$$

□

Now suppose that A is a real symmetric matrix with row and column sums zero, and eigenvalues $\mu_0 = 0, \mu_1, \dots, \mu_{v-1}$. Taking x and z to be the all-1 vector u , we see that the determinant of $A + u^\top u$ is $v \prod_{i=1}^{v-1} \mu_i$, which gives the result mentioned earlier.

Since $u^\top u = J$, this result is related to method (b) in Section 2.1 for finding a generalized inverse of A .

3.3 A-optimality: electrical networks, random walks, thickets

Given a graph G , we can consider it as an electrical network with a 1-ohm resistance in each edge. If we connect a battery across two of the vertices (or otherwise apply a potential difference to them) then current will flow in the network. Three rules are obeyed.

- (a) *Ohm's Law*: In every edge,

$$V = IR. \tag{3.1}$$

Here V denotes the voltage drop across the edge, I denotes the current flowing in the edge, and R is the resistance. Because we consider that $R = 1$ for every edge, we have $V = I$ in every edge. (When we consider designs with non-constant block size in Section 5.4 then the resistances have to be different.)

- (b) *Kirchhoff's Voltage Law*: The total voltage drop from one vertex to any other vertex is the same no matter which path we take from one to the other.
- (c) *Kirchhoff's Current Law*: At every vertex which is not one of the two connected to the battery, the total current coming in is equal to the total current going out.

Given two vertices i and j , we imagine applying voltage 0 at i and voltage V at j . We use Kirchhoff's Laws [67] to find the total current I from i to j , then use Equation (3.1) to define the *effective resistance* R_{ij} between i and j as V/I .

There are two further laws in electrical networks: the “series” and “parallel” laws, which are easy consequences of the three laws above. If two networks are connected in series at a single vertex then the effective resistance in the combined network is the sum of the two resistances. On the other hand, if they are connected in parallel, joining one “input” vertex to one “output” vertex, then the effective resistance in the combined network is the reciprocal of the sum of the reciprocals of the two resistances.

Tjur [100, 101] and Bailey [7] recognized that these two laws are precisely those that apply to the variance when different estimators are combined. Tjur imagined the Levi graph as an electrical network, Bailey the concurrence graph. In fact, standard textbooks on electrical engineering, such as [14, 39], give the result that if L is the Laplacian matrix of the graph of the (connected) electrical network, then the effective resistance R_{ij} between i and j is given by

$$R_{ij} = L_{ii}^{-} + L_{jj}^{-} - L_{ij}^{-} - L_{ji}^{-}. \quad (3.2)$$

Comparing Equations (2.6) and (3.2), and using Proposition 2.7, we see that pairwise variance = effective resistance $\times k\sigma^2$.

This observation is useful as well as interesting. For sparse graphs, or poorly connected graphs, effective resistance can be calculated by hand much faster than the information matrix can be inverted (including data-entry time).

Example 3.8 The poorly connected graph in Figure 3 is taken from [9]. Consider vertices C and D . If current is flowing from C to D then there will be no current in the part of the graph to right of E , and there will be no current in the edge AB , because we can interchange those vertices without changing the graph. This leaves current in the three disjoint paths CAD , CBD and CED . They each have length 2, so the laws permit a current of 1 amp in each of them. Now the total current flowing from C to D is 3, and the voltage drop from C to D is 2, so the effective resistance between C and D is $2/3$.

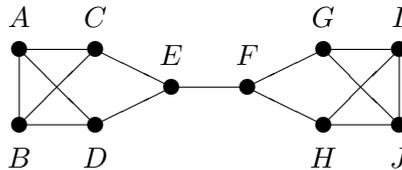


Figure 3: A poorly connected graph

For sparse graphs with a general number of vertices along a long path, such as the one in Figure 4 (taken from [7]), numerical inversion of the information matrix is impossible, but it is perfectly possible to calculate effective resistances in terms of path-lengths.

Other information known to graph theorists and to electrical engineers can now be used to simplify the calculations. For example, [20] gives a formula for the

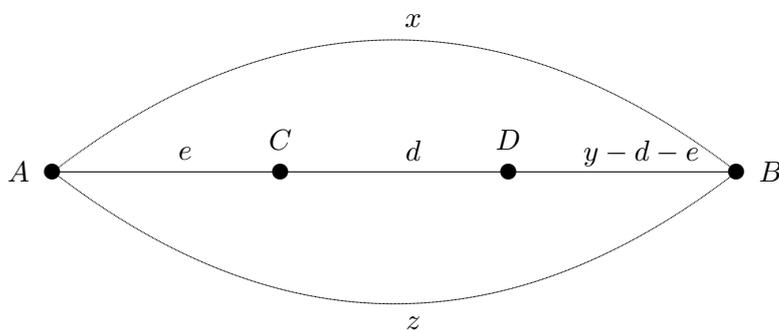


Figure 4: A graph with three parallel paths of lengths x , y and z

effective resistance between i and j in terms of thickets. A *spanning thicket* for a graph is a subset of its edges which forms a forest with two connected components, with every original vertex being in exactly one component. The formula states that

$$R_{ij} = \frac{\text{number of spanning thickets with } i, j \text{ in different parts}}{\text{number of spanning trees}}. \quad (3.3)$$

This is quite easy to calculate for sparse graphs.

Example 3.9 Consider the graph in Figure 4. The only way to make a spanning tree is to remove one edge from two of the parallel paths. To make a spanning thicket, either remove one edge from each of the three parallel paths, or remove two edges from one and one edge from another. For example, to make a spanning thicket separating C and D , remove one edge from the direct path between C and D , and either (i) remove one edge from each of the other two long paths between A and B , or (ii) remove one edge between A and C or one between B and D , and remove one edge from one of the other two long paths. The number of such thickets is $d[xz + (y - d)(x + z)]$. The number of spanning trees is $xy + xz + yz$, so the effective resistance between C and D is $[d(xy + xz + yz) - d^2(x + z)]/(xy + xz + yz)$.

Summing Equations (3.3) over all pairs of distinct vertices gives

$$\sum_{i < j} R_{ij} = \frac{\sum_{\text{thickets } F} |F_1| |F_2|}{\text{number of spanning trees}},$$

where the sum is over thickets F consisting a tree F_1 containing i and a tree F_2 containing j : see [92, 102]. This gives a relatively quick method of calculating the A-criterion for sparse graphs.

3.3.1 Foster's formulae and their generalization This section is based on the work of Emil Vaughan [103]. See also [56, 15]. In this section, we consider only simple graphs; the results can be extended to multigraphs but statements are more elaborate.

Let G be a simple connected graph with v vertices. Recall that we are interested in the quantity $\sum_{i,j} R_{ij}$, summed over all pairs of vertices. In 1948, Foster [42] proved the remarkable fact that summing over edges always gives $v - 1$. In this theorem, the notation $i \sim j$ means that we are summing over all unordered pairs $\{i, j\}$ of vertices which are adjacent in the graph.

Theorem 3.10 *For any graph G on v vertices,*

$$\sum_{i \sim j} R_{ij} = v - 1.$$

Thirteen years later, he proved an analogous formula for the sum of resistances over pairs of vertices at distance at most 2. More precisely, he showed that

$$\sum_{i \sim h \sim j} \frac{R_{ij}}{d(h)} = v - 2.$$

Note that, under the assumption that the graph is regular and that the numbers of common neighbours of two vertices at distance 1 (respectively 2) are constant, this formula gives us the sum of resistances over all pairs of vertices at distance 2. In particular, if the graph is strongly regular, we can compute the sum of resistances over all pairs of vertices.

Example 3.11 *The Petersen graph.* This graph has ten vertices, valency 3, and diameter 2; there are no triangles, and two non-adjacent vertices have a unique common neighbour. By Foster's first formula, the sum of the resistances between all pairs of adjacent vertices is 9. Also, since there are no triangles, the sum in Foster's second formula is over all pairs of non-adjacent vertices; for each such pair $\{i, j\}$ there is a unique h , and $d(h) = 3$. So the sum of resistances between all non-adjacent pairs is $3 \cdot 8 = 24$, and the total is 33. (Note that, by symmetry, the resistance between two adjacent vertices is $3/5$ while that between two non-adjacent vertices is $4/5$.)

This can also be found by considering eigenvalues. The eigenvalues of the adjacency matrix are 3, 1, -2 with multiplicities 1, 5, 4 respectively; so the Laplace eigenvalues are 0, 2, 5 with the same multiplicities. Now v times the sum of reciprocals of the non-trivial eigenvalues is $10 \cdot (5/2 + 4/5) = 33$.

New proofs were found by Tetali [97, 98] using the random walk approach, and Palacios [77] extended them to distance 3. Vaughan found a generalization for arbitrary distance, requiring suitable assumptions on the graph; we now discuss this.

Consider the uniform random walk on the graph G : in other words, at each time stage, we move from our present vertex to a neighbouring vertex, all neighbours being equiprobable. Let P_{ij} be the probability of moving from i to j in a single move; thus $P_{ij} = 1/d(i)$ if $i \sim j$, and $P_{ij} = 0$ otherwise. Let $P_{ij}^{(m)}$ be the probability of moving from i to j in exactly m steps. Note that $P_{ii}^{(0)} = 1$ and $P_{ij}^{(0)} = 0$ for $i \neq j$;

and $P_{ij}^{(1)} = P_{ij}$. Moreover,

$$P_{ij}^{(m+1)} = \sum_h P_{ih}^{(m)} P_{hj}.$$

If G is a regular graph with valency d , then $P_{ii}^{(m)} = W_{ii}^{(m)}/d^m$, where $W_{ii}^{(m)}$ is the number of walks of length m starting and ending at i .

We say that the graph G is *walk-regular up to length m* if the numbers $W_{ii}, W_{ii}^{(2)}, \dots, W_{ii}^{(m)}$ are independent of the chosen vertex i . A graph is *walk-regular* if it is walk-regular up to length m for all m . See [48, p. 190]. Note that

- all graphs are walk-regular up to length 1;
- a regular graph is walk-regular up to length 2;
- if a graph is regular, then it is walk-regular up to length m if and only if the probability $P_{ii}^{(p)}$ of returning to vertex i after p steps is independent of i for $p = 1, \dots, m$;
- a distance-regular graph is walk-regular.

Now Vaughan's Theorem states the following:

Theorem 3.12 *Suppose that G is walk-regular up to length m , for some positive integer m . Then the sum of the resistances between all pairs of vertices at distance at most $m + 1$ is*

$$d^m \left(v \left(1 + \sum_{p=1}^m P_{ii}^{(p)} \right) - m - 1 \right).$$

Note that, if $m = 0$, then d is undefined; for $m > 0$, d is the (constant) valency of the graph. In these cases we recover Foster's two formulae.

Using this theorem, we see that if the graph G is distance-regular, then the numbers $W_{ii}^{(m)}$ of closed walks, and hence the sum of resistances between all pairs of vertices of G , can be computed in terms of the parameters c_p, a_p, b_p of the graph. Calculation of the eigenvalues is not required to evaluate this.

3.4 E-optimality: spectral gap and root systems

A design is E-optimal if the smallest non-trivial eigenvalue of its information matrix is as large as possible; equivalently, if the smallest non-trivial eigenvalue μ_1 of the Laplacian of its concurrence graph is as large as possible.

We have noted the importance of μ_1 in the theory of expanders, random walks, etc. We see that maximizing μ_1 , for graphs with v vertices and e edges, is equivalent to finding the E-optimal design with block size 2 having v points and e blocks. For larger block size, we have to take graphs which are concurrence graphs, as described in Section 2.4.

There is a lot of interest in graphs of fixed valency d having large second Laplace eigenvalue. Alon and Boppana (see [2]) showed (see also [49, 90]):

Theorem 3.13 *In any infinite family of connected regular graphs of valency d , the smallest non-trivial Laplace eigenvalue μ_1 satisfies*

$$\limsup \mu_1 \leq d - 2\sqrt{d-1}.$$

A graph G of valency d is said to be a *Ramanujan graph* if $\mu_1(G) \geq d - 2\sqrt{d-1}$. Explicit Ramanujan graphs were constructed by Margulis [69] and by Lubotzky, Phillips and Sarnak [68]; an accessible account with all the background material explained appears in [38]. These graphs have good expansion properties, and also often have small diameter and large girth.

However, more often we will not be dealing with graphs with fixed valency; and, of course, if the block size of the designs is greater than 2, then the concurrence graphs will not have the large girth typical of Ramanujan graphs. Indeed, the graphs may have repeated edges!

3.4.1 Root systems As we have seen, E-optimality involves maximizing the least non-trivial Laplace eigenvalue of the concurrence graph G . If G is simple, then this is equivalent to minimizing the greatest Laplace eigenvalue of its complement \overline{G} . Moreover, if \overline{G} is regular, then this is equivalent to maximizing the smallest eigenvalue of its adjacency matrix.

There is another situation in which similar considerations apply, suggested to the authors by Ching-Shui Cheng. Suppose that a design has constant block size k , constant replication r , and has the property that all pairwise concurrences lie in the set $\{\lambda - 1, \lambda, \lambda + 1\}$ for some number λ . Then the concurrence matrix has the form $aI + bJ - M$, where M is a symmetric matrix with entries 0 and ± 1 having zero diagonal and constant row sums. Once again, maximizing the smallest Laplace eigenvalue is equivalent to maximizing the smallest eigenvalue of M .

Now, there is a structure theorem for such matrices with smallest eigenvalue -2 or greater, which we now describe. This approach follows [27] but is a bit different: that paper considered only $\{0, 1\}$ matrices and did not impose the constant row-sum condition. The approach is based on the classical theory of root systems.

A *root system* is a finite non-empty set S of vectors in Euclidean space \mathbb{R}^d satisfying the following conditions:

- if $x \in S$ and $\alpha \in \mathbb{R}$, then $\alpha x \in S$ if and only if $\alpha \in \{\pm 1\}$.
- for all $x, y \in S$, we have $2(x \cdot y)/(x \cdot x) \in \mathbb{Z}$;
- for all $x \in S$, the reflection ρ_x of \mathbb{R}^d in the hyperplane perpendicular to x maps S into itself.

We can assume (and always shall do so) that S spans \mathbb{R}^d . The second condition is called the *crystallographic restriction*. The significance of the quantity referred to is that the reflection ρ_x is given by

$$\rho_x : y \mapsto y - \frac{2(x \cdot y)}{x \cdot x} x.$$

A root system is *indecomposable* if it is not contained in the union of two perpendicular subspaces of \mathbb{R}^d . There is no loss in assuming this, since an arbitrary root system is a direct sum of indecomposable ones.

The indecomposable root systems were classified by Cartan and Killing in the early twentieth century in connection with the classification of simple Lie algebras over the complex numbers. See [22, 53] for an account. An alternative proof in the spherical case appears in [27] and [48, Chapter 12].

A root system S is called *spherical* if all vectors in S have the same length. If S is spherical, we may re-scale so that $x \cdot x = 2$ for all $x \in S$; then the crystallographic condition asserts that $x \cdot y$ is an integer, and Cauchy's inequality says that $|x \cdot y| < 2$ for $y \neq \pm x$, so we have $x \cdot y \in \{0, \pm 1\}$ for $y \neq \pm x$. This says that any two such vectors lie at an angle of 60° , 90° or 120° . Moreover, if two vectors x, y make an angle of 120° , then the vector $x + y$ in their plane making an angle of 60° with each is also in S , by the reflection condition.

According to the classification, there are two infinite families $\{A_d : d \geq 1\}$ and $\{D_d : d \geq 4\}$, and three exceptional examples E_6 , E_7 and E_8 , of indecomposable spherical root systems. All of these have very explicit representations. For example, the vectors of D_d can be represented as

$$\{\pm e_i \pm e_j : 1 \leq i < j \leq d\}$$

where $\{e_1, \dots, e_d\}$ is an orthonormal basis for \mathbb{R}^d .

Now the relevance of this to eigenvalues comes from the following two observations.

Extending to a maximal set Let T be any set of vectors in \mathbb{R}^d at angles 60° , 90° and 120° . Assume that all the vectors in T have length $\sqrt{2}$. Let T' be the set of lines through the origin (1-dimensional subspaces) spanned by these vectors. These lines make angles 60° and 90° with each other. Extend T' to a maximal set S' of lines at these angles. Then S' is *star-closed*, i.e. if two lines in S' are at an angle of 60° , then the third line at 60° to both is also in S' . (This requires a short calculation, the only calculation in the whole argument.) Now let S consist of the vectors of length $\sqrt{2}$ on the lines of S' . Then S is a spherical root system (the reflection condition follows from the star-closure of S'), is indecomposable if T is, and spans \mathbb{R}^d if T does; and S contains T . So any set T satisfying our conditions is embeddable in a root system.

Representing matrices Now let A be a $v \times v$ symmetric matrix with entries $0, \pm 1$ with zero diagonal, and suppose that the smallest eigenvalue of A is -2 or greater. Then $2I + A$ is positive semi-definite symmetric, and so is the matrix of inner products of a set T of vectors spanning \mathbb{R}^d , where d is the rank of $2I + A$ (so that $v - d$ is the multiplicity of the eigenvalue -2 of A , which may be zero.) Since the diagonal entries are 2, all the vectors have length $\sqrt{2}$; the assumption on entries of A shows that the vectors have inner product 0 or ± 1 with one another, so make angles 60° , 90° and 120° . So T can be embedded into a spherical root system.

This has important consequences for the structure of such matrices.

Graphs Suppose that the entry -1 does not occur in A ; then A is the adjacency matrix of a simple graph G with smallest eigenvalue -2 or greater. It is clear that only finitely many such graphs can be "embedded" in the exceptional root systems;

moreover, a graph embeddable in A_d or D_d is a *generalized line graph* (see definition below). In particular, if G is regular, then it is one of the following:

- A *line graph*. (If H is a simple graph, then the line graph of H has one vertex for each edge of H ; two vertices of the line graph are joined if the corresponding edges of H meet.) We denote the line graph of H by $L(H)$.
- A *cocktail party graph*: this is a complete multipartite graph with parts of size 2; that is, the vertices are paired up, two vertices are joined if and only if they are not in the same pair.
- One of finitely many exceptions (embeddable in the exceptional root systems). These exceptions have all been determined; there are several hundred of them, on at most 28 vertices.

If $L(H)$ is regular, then either H itself is regular, or H is bipartite and *semiregular* (the valency of a vertex depends only on the class of the bipartition containing it) and satisfies some additional strong restrictions (but examples for this case do exist).

Here is the definition of a generalized line graph. Let H be a graph with v vertices, and let a_1, \dots, a_v be non-negative integers. Then the *generalized line graph* $L(H; a_1, \dots, a_v)$ consists of the disjoint union of $L(H)$ and cocktail parties C_i on $2a_i$ vertices for $i = 1, \dots, v$, with an edge from each vertex of C_i to each vertex of $L(H)$ indexed by an edge containing vertex i of H .

Regular signed graphs In general we can regard the matrix A as the adjacency matrix of a *signed graph* (a graph with a sign attached to each edge). Not much can be said in general, but in the case where A has constant row and column sums, see [25] for some results.

3.4.2 Discrepancy We briefly review some results of Morgan [73]. These allow the classification of E-optimal designs for a reasonable range of parameters. This extends the analysis of VB designs given in Section 2.6.

Given v, b, k , we define parameters r, λ, p, q by

$$r = \left\lfloor \frac{bk}{v} \right\rfloor, \quad p = bk - vr,$$

$$\lambda = \left\lfloor \frac{r(k-1)}{v-1} \right\rfloor, \quad q = r(k-1) - \lambda(v-1).$$

Thus, the necessary conditions (1.1) for a 2-design are satisfied if and only if $p = q = 0$; in this case, a 2-design is best if it exists. The necessary condition for a 1-design (an equireplicate design) is $p = 0$.

Given a *binary, equireplicate* design in $\mathcal{D}(v, b, k)$, with information matrix C , define its *discrepancy matrix* to be the matrix Δ defined by

$$C = \frac{v\lambda + q}{k}I - \frac{\lambda}{k}J - \frac{1}{k}\Delta.$$

Then Δ has diagonal entries zero and, for $i \neq j$, has off-diagonal entry $\lambda_{ij} - \lambda$, where λ_{ij} is the concurrence of i and j .

We see that the design is a 2-design if and only if its discrepancy matrix is zero. (This is the reason for the name: the matrix measures the “imbalance” of the design.) Also, it is a regular-graph design if and only if all entries of Δ are 0 or 1. We see that E-optimality is equivalent to maximizing the largest eigenvalue of Δ .

Using this approach, the E-optimal designs with $v \leq 15$ have been found. Not all are regular-graph designs (that is, the discrepancy matrix of the optimal design may have entries which are negative or greater than 1). The results are all available at designtheory.org.

One striking observation to emerge from the computation is that, while the E-optimal designs are not necessarily A-optimal, they are very close to optimality in terms of the A-criterion.

Morgan’s extensive catalogue should form the basis for further work on E-optimality.

4 Graphs with few edges

In this section, after looking briefly at trees, we determine the optimal unicyclic graphs (connected graphs with equally many vertices and edges), and then turn to sparse graphs (with average valency at most $5/2$). We will see that different optimality criteria select very different graphs.

4.1 Trees

As a warm-up, let us consider trees. Any tree has a unique spanning tree, so the D-criterion makes no distinction between one tree and another. Now in a tree, the resistance between two terminals is just the length of the path joining them; so A-optimality requires that the sum of distances between pairs of vertices is minimized. It is clear that the star $K_{1,v-1}$ does this, since the distance between any two non-adjacent vertices is 2 in the star, and no other tree can do as well or better.

Now we turn to E-optimality. A short calculation shows that the non-trivial Laplace eigenvalues of the star are 1 (multiplicity $v - 2$) and v (multiplicity 1). The fact that any other tree has $\mu_1 < 1$ follows from Proposition 2.12. For a tree with more than four vertices which is not a star possesses an edge whose removal leaves at least two points in one component and at least three in the other; so the Proposition gives

$$\mu_1 \leq (1/2 + 1/3) = 5/6.$$

The only other tree is the path of length 3, for which a short calculation shows that $\mu_1 = 2 - \sqrt{2}$.

So the best choice of a design in this case (v treatments, $v - 1$ blocks of size 2) is to compare all the other treatments with a fixed “reference” treatment.

4.2 Unicyclic graphs

What happens for unicyclic graphs?

Such a graph has a single cycle, of length s say, with trees attached to some vertices (possibly none) in the cycle. The only way to make a spanning tree is to remove one of the edges in the cycle, so the number of spanning trees is maximized when $s = v$; that is, when the graph is just a cycle. This proves:

Theorem 4.1 *Let $v \geq 3$. Among unicyclic graphs with v vertices, the D -optimal graph is the cycle.*

For $k = 2$, such designs are called *loop designs* in the microarray literature.

Now we turn to E -optimality, where the results are rather different. We have seen that the minimum eigenvalue for the star is 1, and Proposition 2.9(d) shows that this is not reduced if an edge is added.

The smallest eigenvalue of the Laplacian matrix of the cycle of size v is equal to $2(1 - \cos(2\pi/v))$, which is greater than 1 when $v \leq 5$, equal to 1 when $v = 6$ and less than 1 when $v \geq 7$.

Each effective resistance is an evaluation of $x^\top L^- x$ for a contrast vector x with $x^\top x = 2$. Consequently, $\mu_1 \leq 2/R_{ij}$ for all pairs of distinct vertices i and j . If vertices i and j are in trees attached to vertices m and n in the cycle (where m and n may be the same), then the effective resistance R_{ij} is the sum of the distance from i to m , the distance from j to n , and R_{mn} , which is strictly bigger than 2 (and so $\mu_1 < 1$) unless $m = n$ and $\{i, m\}$ and $\{j, n\}$ are edges. Thus a unicyclic graph cannot be E -optimal unless all the trees are leaves and they are attached to the same vertex of the cycle.

If there are at least two leaves attached to some vertex, then there is a Laplacian eigenvalue equal to 1: its eigenvector is the contrast between those two leaves. For a single leaf attached to a cycle of size 2 or 3, the least eigenvalue is $3 - \sqrt{3}$, 1 respectively.

For $s \geq 3$, the effective resistance between two vertices in the cycle at distance 2 is $2(s-2)/s$, which is bigger than 1 when $s \geq 5$. If there is a leaf attached to a vertex in this cycle, then there is an effective resistance bigger than 2, so the graph cannot be E -optimal.

For $v \geq 5$, consider the graph consisting of a 4-cycle with $v - 4$ leaves attached to one vertex i . The Laplacian has an eigenvalue 1 with multiplicity $v - 5$: the eigenvectors are the pairwise contrasts among the leaves. The contrast between the other two neighbours of i is an eigenvector with eigenvalue 2. The remaining three non-trivial eigenvalues must take one constant value on all the leaves, and a (possibly different) constant value on the other two neighbours of i . Then a short calculation shows that the remaining three non-trivial eigenvalues are the zeros of the polynomial

$$f(\mu) = \mu^3 - (v + 3)\mu^2 + (4v - 2)\mu - 2v.$$

Then $f(0) < 0$ and $f(1) = v - 4 > 0$, so one of the eigenvalues lies in $(0, 1)$.

Putting all these observations together shows:

Theorem 4.2 *Let $v \geq 3$. Among unicyclic graphs with v vertices, the E -optimal graph is*

- *the cycle, if $v \leq 5$;*
- *a triangle with $v - 3$ leaves attached to a vertex, or a star with one edge doubled, if $v \geq 7$;*
- *either of the above if $v = 6$.*

Thus the E-optimal designs are very different from the D-optimal designs when $v \geq 7$.

What about A-optimality? Again, suppose that trees are attached to two or more vertices in the cycle. If the trees are all moved to the same vertex then the sum of the pairwise effective resistances between tree vertices and cycle vertices is unchanged, as are the within-tree resistances, but the resistances between vertices in different trees is reduced. Further, suppose that a tree with more than one vertex is attached to vertex i of the cycle, in such a way that j is the only vertex of the tree joined to i . If the edge $\{i, j\}$ is contracted and vertex j is reattached to i as a leaf, then the resistances between tree vertices and i are interchanged with the resistances between tree vertices and j , resistances between tree vertices and the rest of the cycle decrease, and all other resistances are unchanged. This shows that, just as for E-optimality, the only candidates for A-optimal graphs are those consisting of a cycle of length s with $v - s$ leaves attached to one vertex, for some s .

A short calculation shows that the sum of the pairwise effective resistances in such a graph is

$$[-s^3 + 2vs^2 + s(13 - 12v) + 12v^2 - 14v]/12.$$

For $v \leq 7$, this polynomial in s is monotonic decreasing on $[0, v]$. For $v = 8$ it attains its minimum on $[0, 8]$ at 8; for higher values of v it attains its minimum on $[0, v]$ between 3 and 4, with the integer achieving the least value being 4 if $9 \leq v \leq 12$ and 3 if $v \geq 12$. This proves the following.

Theorem 4.3 *Let $v \geq 3$. Among unicyclic graphs with v vertices, the A-optimal graph is*

- *the cycle, if $v \leq 8$;*
- *a 4-cycle with $v - 4$ leaves attached to one vertex, if $9 \leq v \leq 11$;*
- *a triangle with $v - 3$ leaves attached to a vertex, if $v \geq 13$;*
- *either of the last two if $v = 12$.*

This result was obtained for $v \leq 12$ by computational search in [60, 63, 108], and by the above analysis in [7, 100].

Together, the last three results show that, for $k = 2$ and $v = b$, not only are the D-optimal designs very different from the A- and E-optimal designs for $v \geq 9$ but also the ranking on the D-criterion is very different from those on the other two criteria. This was generalized in [7] to show that, for any fixed value of $b - v$, if $k = 2$ then there is a threshold v_0 such that if $v \geq v_0$ then the D-optimal design has no leaves but the A-optimal design does have leaves. The next two results have a similar flavour.

4.3 Sparse graphs

Theorem 4.4 *Let \mathcal{G} be the set of connected graphs with v vertices and e edges, where $e > v$. If the graph G in \mathcal{G} has a vertex l which is either a leaf or, more generally, has all its edges joining it to the same other vertex, then G is not D-optimal.*

Proof First suppose that l is a leaf. Since $e > v$, G is not a tree. Therefore it contains at least one edge ij which is not an isthmus. So some spanning trees of G omit the edge ij . Note that l is cannot be equal to i or j . All spanning trees for G contain the unique edge through l . Now form the graph G' by removing the edge through l and the edge ij , and inserting new edges il and jl . Each spanning tree for G which contains ij becomes a spanning tree for G' by replacing ij and the edge through l by il and jl . Each spanning tree for G which omits ij gives two spanning trees in G' : each omits the old edge through l and includes one of il , jl . Therefore G' has more spanning trees than G , and so G is not D-optimal.

Secondly suppose that l is joined to vertex m by s edges, and to no other vertices of G , where $s \geq 2$. Let i be any other vertex of G . Form the graph G'' by removing one edge between l and m and inserting an edge between l and i . Each tree of G contains one of the edges between l and m : these all give a tree in G'' , using the replaced edge if necessary. However, the moved edge creates extra cycles, so there are some trees in G'' that contain two edges through l and do not correspond to any tree in G . Therefore G'' has more spanning trees than G , and so G is not D-optimal. \square

Theorem 4.5 *Let \mathcal{G} be the set of connected graphs with v vertices and e edges. If $20 \leq v \leq e < 5v/4$ then the E-optimal graphs in \mathcal{G} have leaves.*

Proof We know that the star $K_{1,v-1}$ has least non-trivial eigenvalue 1 and that adding edges to the star cannot decrease this. Our technique is repeated use of Proposition 2.12 to ban portions of a graph that would force a non-trivial eigenvalue smaller than 1.

Let G be an E-optimal graph in \mathcal{G} and suppose that G has no leaves. For $d \geq 2$, let n_d be the number of vertices of valency d . Call the vertices of G with valency 2 *dots*, the remaining vertices *nodes*. Call a path from one node to another a *side*: this may go through dots, and the two end nodes may be the same.

Suppose that there are three or more dots in one side. The edges at the end of the side make a cutset of size two, with parts of size s and $v - s$, for some $s \geq 3$. Hence $\mu_1 \leq 2(1/s + 1/(v - s))$, which is less than 1 if $s \geq 3$ and $v \geq 20$. So this cannot happen.

Consider a node i of valency 3. If there is a side making a loop at i , it must contain at least one dot, because G is loopless. The remaining edge at i now makes a cutset of size 1 with parts of size s and $v - s$ for $2 \leq s \leq 3$: by Proposition 2.12, this cannot happen. Therefore, the three sides at i are distinct. The edges at the far ends of those sides make a cutset of size three. Since $3(1/4 + 1/16) < 1$, the size of the part containing i is at most three, so there can be no more than two dots in total in these three sides.

Now consider a node j of valency 4. Similar arguments show that it may have one side forming a loop with precisely one dot, in which case there are no dots on the remaining two sides, and that otherwise the total number of dots in the four sides at j is at most four, because $4(1/6 + 1/14) < 1$.

Counting the dots according to nodes at the end of their sides counts them twice, so the preceding arguments show that

$$2n_2 \leq 2n_3 + 4n_4 + \sum_{d \geq 5} 2dn_d.$$

Moreover,

$$v = n_2 + n_3 + n_4 + \sum_{d \geq 5} n_d$$

and

$$2e = 2n_2 + 3n_3 + 4n_4 + \sum_{d \geq 5} dn_d.$$

Hence

$$5v - 4e = n_2 - n_3 - 3n_4 + \sum_{d \geq 5} (5 - 2d)n_d \leq -n_4 + \sum_{d \geq 5} (5 - d)n_d \leq 0.$$

This contradiction shows that G must have leaves. □

5 Further directions

In this section, after looking first at statisticians' view of the theory already presented, allowing for additional practical considerations we have not touched on, we give a number of open problems, and indicate some alternative optimality criteria and what to do if the block size is not constant. Remarkably, the case of non-constant block size leads us to consider the Laplacians of weighted graphs, for which the mathematical theory has already been developed.

5.1 Statistical and practical considerations

When there are no blocks, the information matrix is just the diagonal matrix of replications R . Then the optimal design on all the criteria is as equally replicate as possible. People with even a little statistical training are used to this, and so tend to favour equireplicate designs. One argument in favour of equal replication is that, if the model in Equation (2.4) is fitted and it is found that the block parameters differ little from each other, then the blocks can be ignored and the data re-analysed, at which stage an equireplicate design is best. However, this argument is controversial, and not accepted by all statisticians.

Many blocked experiments have so-called *complete blocks*: $k = v$ and every treatment occurs once in each block. In some areas of science, both practical experimenters and regulatory authorities expect experiments to be like this.

When blocks are smaller, it may happen that blocks are naturally grouped into larger "superblocks". For the design in Table 1, it may be that the farms can be grouped by soil-type; in another experiment, different superblocks may be dealt with at different times. If each superblock contains v experimental units then it is often advocated that a *resolved* block design should be used; that is, one in which each treatment occurs once in each superblock. Since any differences between superblocks are already accounted for in Equation (2.4), the arguments in favour of resolved designs are that (i) if a whole superblock is lost then the remaining design is equireplicate, which begs the question; (ii) if the first analysis suggests that blocks do not differ within each superblock, then the data can be reanalysed with parameters for superblocks but not blocks, which is as controversial as the previous re-analysis strategy. Nevertheless, resolved designs are requested sufficiently often in practice that some work has been done on optimal designs within this class: see [12].

Because of this background, statisticians tend to be horrified at the A- and E-optimal designs found in Section 4, where most treatments have replication 1. On the other hand, some scientists find them completely natural, having absorbed the dictum that all treatments should be compared with a reference, or control, treatment. Even then, there is concern that if the reference treatment is very different from the others then Equation (2.4) may not hold.

Robustness is a relevant issue here. If it is possible that one or two observations may be lost from the experiment, and if it is important to compare all treatments, then a design is robust if all comparisons can be made no matter which two observations are lost, so it must have every treatment replicated at least three times. On the other hand, in an experiment to compare 300 mutants of yeast to the “wild-type”, losing two observations from a star-like design does not compromise the information about the remaining 298 mutants, while losing two observations from a 301-cycle makes many comparisons impossible and gives others very large variances.

Sometimes experimenters may be forced to use given, unequal replication because some treatments are more easily available than others, or because safer ones must be tested before riskier ones. Usually it is still true that all comparisons are interesting, but the search for an optimal design must be restricted to one whose replications are feasible.

In many situations, an optimal design for the given restrictions is not known. However, there may be little practical difference between designs which are close to optimal, so it can be useful to have a realistic bound on the optimality criterion.

One block design may be compared with another. The *efficiency factor* for contrast x in design \mathfrak{D}_1 relative to \mathfrak{D}_2 is the ratio $x^\top C_2^- x / x^\top C_1^- x$, where C_i is the information matrix of \mathfrak{D}_i . To decide which design is best, this ratio needs to be compared with the ratio of the variances of the individual responses for the two designs, which may be different if the block sizes differ.

Often the comparator is taken to be the unblocked design with equal replication \bar{r} , defined to be bk/v even if this is not an integer. However, if the replications are forced to be unequal, then the comparator design may be taken to be the unblocked design with the same replications, whose information matrix is just R . Then the efficiency factor for x is $x^\top R^{-1} x / x^\top C^{-1} x$. This leads to the notion of a design being *efficiency balanced* if the non-trivial eigenvalues of $R^{-1/2} C R^{-1/2}$ are all equal, where $R^{1/2}$ is the diagonal matrix whose entries are the positive square roots of the replications. This approach takes no account of the possibility of finding a design that can get the most information *in spite of* being constrained in some way. On the other hand, if the low replications are not forced but correspond to less interesting treatments, then the appropriate approach would seem to be to incorporate weights (on contrasts, or on treatments) into the optimality criterion. This is already done in an extreme way (weights of zero and one) in two situations: (i) one treatment is a control and only comparisons with it are of interest; (ii) the treatments are all combinations of levels of several factors, and higher-order interactions are ignored.

5.2 Conjectures and refutations

It seems so obvious that balanced incomplete-block designs must be optimal on every reasonable criterion that everyone believed Kiefer’s theorem long before it was

proved. Other conjectures about optimality seem equally obvious, but all too often turn out to be wrong.

Given Kiefer's theorem, surely it must be true that, when there is no 2-design in $\mathcal{D}(v, b, k)$, there must be an optimal design which is binary or generalized binary, which is equireplicate or nearly so, and whose concurrences are as equal as possible? Surely there will be optimal designs with a high degree of symmetry, and surely a design which is optimal on one criterion will not be far from optimal on another? Our running example, and the results in Section 4, show that these plausible conjectures are all wrong.

What about the concurrence graph? Is it true that pairwise variance V_{ij} (equivalently, effective resistance R_{ij}) increases as the concurrence λ_{ij} decreases or as the distance in the graph increases? Example 3.8 shows that this is not true. Vertices C and D are at distance 2, with $\lambda_{CD} = 0$ and $R_{CD} = 2/3$, while vertices E and F are at distance 1, with $\lambda_{EF} = 1$ and $R_{EF} = 1$. Nonetheless, the vertices at the furthest distance do have the largest resistance.

Question 5.1 *In a loopless graph, is it true that the maximal value of the effective resistance R_{ij} is achieved for some pair of vertices $\{i, j\}$ whose distance apart in the graph is maximal?*

We have already seen that distance-regular graphs can be interesting. In [9] it is proved that, for a regular-graph design for which the graph in question is distance-regular, pairwise variance is a monotonic increasing function of distance in this graph. Biggs [17] proved the equivalent result for effective resistances in an electrical network.

Some of the false conjectures came from experience with partially balanced incomplete-block designs with two associate classes. These designs have two distinct non-trivial eigenvalues. It is shown in [5] that if a design has two distinct non-trivial eigenvalues then pairwise variance is a decreasing linear function of concurrence.

An association scheme is defined to be *amorphic* if it remains an association scheme after any fusion of its associate classes. If a design is partially balanced with respect to an amorphic association scheme then pairwise variance is again a decreasing function of concurrence [9].

The graph in Figure 3 can be regarded as a design in $\mathcal{D}(10, 15, 2)$. It is a fruitful source of counter-examples, but that may be because it is such a bad design. The Petersen graph in Example 3.5 also gives a design in $\mathcal{D}(10, 15, 2)$: it is much better than the previous design on the A-, D- and E-criteria. In fact, the Petersen design is A-optimal among equireplicate designs. Perhaps the nonintuitive behaviour of variance occurs only in designs that are far from optimal. Tjur [100] gave some heuristic reasons for believing this.

Question 5.2 *Suppose that a design is optimal in $\mathcal{D}(v, b, k)$, or in the equireplicate subclass of that. Is it true that, for that design, if $\lambda_{ij} > \lambda_{mn}$ then $V_{ij} < V_{mn}$? If the design is a regular-graph design, is it true that, if the distance between i and j in the graph in question is less than the distance between m and n then $V_{ij} < V_{mn}$?*

If a design is equireplicate then its information matrix is $r(I - (rk)^{-1}\Lambda)$. Paterson [78] expanded $(I - (rk)^{-1}\Lambda)^{-1}$ by the binomial theorem to give an expression for

$\text{Tr}(C^-)$ as a convergent sum over $m \geq 2$ of the quantities \mathcal{W}_m/r^m , where $\mathcal{W}_m = \sum_i W_{ii}^{(m)}$ is the number of closed walks of length m in the concurrence graph. He proposed ordering designs lexicographically by their sequences $(\mathcal{W}_2, \mathcal{W}_3, \mathcal{W}_4, \dots)$. He argued heuristically that an A-optimal design should have the earliest sequence. However, he showed that, among non-optimal designs, the ranking on the A-criterion could be the reverse of that on the sequences.

Question 5.3 *Is it true, that among equireplicate designs, a design with the earliest sequence is A-optimal?*

Cheng [33], building on work in [35], gave an asymptotic result, showing that this question has a positive answer when b is sufficiently large. More precisely, he proved the following statement. We call a design *nearly balanced* if its replication numbers r_i differ by at most one and, for fixed i , its concurrences λ_{ij} differ by at most one. Now, given v and k , there is a number $b_0 = b_0(v, k)$ such that, if $b \geq b_0$, then any nearly balanced block design in $\mathcal{D}(v, b, k)$ is better than any design which is not nearly balanced on a wide range of criteria, including all Φ_p criteria.

The earlier result in [35] asserts this for $k = 2$ and D-optimality: that is, given v , there is a number $e_0 = e_0(v)$ such that, if $e \geq e_0$, then a graph with v vertices and e edges which is nearly balanced has more spanning trees than one which is not. This extends asymptotically the result of Kelmans and Chelnokov [62] cited in the Introduction.

This approach was used in [79] to obtain lower bounds for the A-criterion in terms of \mathcal{W}_3 . The results in [6] support the idea that the concurrence graph of a good equireplicate design should have few triangles, but it is possible that this is true only within the range of replications that statisticians typically consider. See also [78, 104].

To achieve the necessary conditions for Kiefer's theorem, the number of blocks must be a quadratic function of the number of treatments. Theorem 4.5 suggests that if b is merely a linear function of v then the optimality criteria may conflict if v is large enough.

Question 5.4 *Let c and p be positive constants. Is it true that there is a threshold v_p such that the Φ_p -optimal designs in $\mathcal{D}(v, cv, k)$ are very different from the D-optimal designs when $v > v_p$? If so,*

- (a) *are the rankings on the D-criterion and the Φ_p -criterion very different when $v > v_p$?*
- (b) *is it true that the D-optimal designs remain (nearly) equireplicate for all v but the Φ_p -optimal designs have one treatment with much larger replication than the rest when $v > v_p$?*
- (c) *is v_p a decreasing function of p ?*

5.3 Other kinds of optimality

A large number of other optimality criteria have been proposed for designs where the treatments are combinations of amounts of quantitative variables: see [4]. These are not related to combinatorial design, so we do consider them further.

However, there is a refinement of E-optimality that seems worth pursuing. We saw in Section 4 that about $v/2$ edges need to be added to the star $K_{1,v-1}$ before its least non-trivial eigenvalue changes: as these edges are added, it is the *multiplicity* of the eigenvalue that changes. Among designs with the same least Laplacian eigenvalue, it would be sensible to say that one design is better than another if the multiplicity of that eigenvalue is lower.

Such a refinement is analogous to using *minimum aberration* [44] in factorial designs. These designs are the duals of linear codes. Resolution in the factorial design corresponds to minimum weight in the code; minimum aberration to the smallest number of words that have that weight.

Somewhat analogously, one could modify the Schur criterion so that only the first few partial sums $\sum_{i=0}^{n-1} \theta_i$ of eigenvalues are compared. For example, in the Petersen graph, the partial sums of the Laplace eigenvalues are

$$2, 4, 6, 8, 10, 15, 20, 25, 30,$$

whereas for the graph of Figure 3 they are approximately

$$0.22, 2.22, 5.22, 8.22, 11.51, 15.51, 19.51, 24.51, 30;$$

while neither is Schur-better than the other, the Petersen graph is ahead for the first three steps.

5.4 Non-constant block size

So far we have only considered block designs with constant block size. The theory can in principle be extended to remove this assumption. However, there are situations where it is not reasonable to assume that individual responses in different blocks have the same variance if the blocks have different sizes.

Now we have to replace the matrix Q in Section 2.2 by $I - ZK^{-1}Z^\top$, where K is the diagonal matrix of block sizes. Then the information matrix becomes

$$C = R - NK^{-1}N^\top,$$

where R is the diagonal matrix of replication numbers, and N is the point-block incidence matrix. We define the *weighted concurrence graph* to be the complete graph on the vertex set $\{1, \dots, v\}$, where the edge $\{i, j\}$ is weighted by the sum of the reciprocals of the sizes of blocks containing both i and j ; the *weighted adjacency matrix* of this graph has (i, j) entry equal to this weight for $i \neq j$, and diagonal entries zero. Now an argument identical to that of Proposition 2.7 shows:

Proposition 5.5 *Let A be the weighted adjacency matrix of a block design. Then the information matrix is $D - A$, where D is a diagonal matrix whose entries are chosen so that $D - A$ has row and column sums zero.*

In other words, the information matrix is the Laplacian matrix of the weighted concurrence graph.

The theory of Laplacian eigenvalues extends in a fairly straightforward way to weighted graphs [72]. So in principle the earlier results on optimal designs could

be extended to optimality in the class $\mathcal{D}'(v; k_1, \dots, k_b)$ of designs with v points and blocks of sizes k_1, \dots, k_b . Not much is known in generality about this.

There is a simple connection between pairwise balanced designs and variance-balanced designs (without assuming the designs are binary or have constant block size). Any *pairwise balanced design* (one in which all pairwise concurrences are equal) can be converted to a variance-balanced design by repeating each block a number of times proportional to its size; the inverse operation converts a variance-balanced design into a pairwise balanced design. (This result is due to Hedayat and Stufken [52].) So the existence questions for the two types of design with prescribed number of points and set of block sizes are equivalent.

6 Computing with designs

Some commercial statistical computing packages will find optimal designs. For example, CycDesign [37] gives A-optimal designs while JMP [57] gives D-optimal designs. The publicly available package R [83] includes some procedures to find A- or D-optimal designs. The designs are usually found by iteration to improve the criterion rather than by exhaustive search.

However, any rational symmetric function of the eigenvalues of a rational matrix (such as the A- and D-parameters) is invariant under the Galois group of its characteristic polynomial, and hence is a rational number. A mathematician will naturally prefer to compute such parameters exactly as rationals. For this, it may be best to turn to a computer algebra package such as GAP [46].

6.1 The DESIGN package

Apart from the fact that it is free, a great advantage of GAP is that it has been extended by many refereed shareware packages, including the DESIGN package [93], for computing with block designs. This package can in principle find all designs satisfying given specifications, up to isomorphism (or perhaps up to a restricted notion of isomorphism where a “structure group” is specified), or it can be set just to return a single design if one exists.

Although it deals directly only with binary block designs, many other types of design can be fitted into this framework. For example, a Latin square of order n can be regarded as a group divisible design with three “groups” of size n , where each block contains one point from each group, and any two points in different groups belong to a single block.

The DESIGN package is well adapted for finding all binary designs with constant block size k having a given concurrence graph. As input to the graph we can give the numbers of points and blocks, the block size, and the number of blocks containing any given pair of points. In particular, if we have computed the $n \times n$ adjacency matrix M of the concurrence graph, then the command

```
BlockDesigns(rec(v:=n, t:=2, blockSizes:=[k],
  tSubsetStructure:=rec(t:=2, lambdas:=1, partition:=p)));
```

will find all such designs, up to isomorphism; here p is the partition of the 2-element subsets $\{i, j\}$ of $\{1, \dots, n\}$ according to the entry M_{ij} of the matrix M , and 1 the

list of corresponding entries. If we also put `isoLevel:=0` in the record passed to the function, then it will return just one such design (if one exists).

Example 6.1 *Variance balanced designs with block size 3.* We have seen that, if $v \equiv 2 \pmod{3}$, the best we can do is to take a block $[1,1,2]$, and to choose the remaining $(v+1)(v-2)/3$ blocks to have all points distinct and to cover the pairs other than $\{1,2\}$ twice. The following DESIGN code produces a list of all such designs (up to isomorphism).

```
designs:=function(pts) # argument must be congruent to 2 mod 3
  local pn;
  pn:=[[1,2]];
  Add(pn, Difference(Combinations([1..pts],2),[[1,2]]));
  return BlockDesigns(rec(v:=pts, blockSizes:=[3],
    tSubsetStructure:=rec(t:=2, lambdas:=[0,2], partition:=pn)));
end;
```

The argument to the function `BlockDesigns` is a record, of which the component `v` is the number of points, `blockSizes` a list of the allowed sizes of blocks, and `tSubsetStructure` describes how many blocks should contain a given t -subset; the first component `t` is the value of t , the second `lambdas` a list of the required concurrence numbers, and the third `partition` a partition of the set of t -subsets with the same number of parts as the number of components of `lambdas`; the designs found will have the i th component of `lambdas` as the number of blocks containing a t -set in the i th part of the partition. So the above code specifies that the blocks should have size 3, that no block should contain $\{1,2\}$, and that any other set of size 2 should be contained in two blocks. We then add the block $[1,1,2]$ to get the required design.

For $v = 5$, there is a unique design, the one we have used as a running example. For $v = 8$, there are 10 designs up to isomorphism. Of these, seven have repeated blocks; the sizes of the automorphism groups of the ten designs are 12, 6 (three times), 4 (three times), 2 (twice) and 1. (The program returns a generating set for each automorphism group.) The computation takes a couple of seconds on a desktop PC.

There is a reason why knowledge of the automorphism group matters. To the output of the program, we can add either $[1,1,2]$ or $[1,2,2]$ to form a VB design. The resulting designs will be isomorphic if and only if the binary part has an automorphism interchanging 1 and 2. Since just two of the designs on 8 points have such an automorphism, we find altogether 18 isomorphism classes of $\text{VB}(8,3,2)$ designs with just one non-binary block. Of these 18 designs, 13 have repeated blocks.

For $v = 11$, the computation is much more substantial, since many designs exist. The DESIGN package finds that, if we prescribe that the design should have an automorphism of order 2 fixing the special points 1 and 2 and three further points, there are 3998 choices for the binary part of the design, leading to 7814 designs altogether. We have not computed the total number for $v = 11$ because of memory limitations.

Some other examples of the use of the package include:

- Construction and analysis of *regular-graph designs* [73]. This class includes many optimal designs in cases where 2-designs do not exist.
- Construction and analysis of 2-designs with repeated blocks [40].
- An implementation of the Jacobson–Matthews Markov chain method [54] for choosing a uniform random Latin square, and an adaptation of it to other kinds of design such as Steiner triple systems [26].
- Analysis of small *gerechte designs* [10]. These are a particular type of Latin square adapted for dealing with spatial inhomogeneities. Solutions to Sudoku puzzles are the best-known examples.

6.2 The `designtheory.org` project

The DESIGN package was produced as part of a wider project, available from `designtheory.org`, and described in the paper [11]. The project was restricted to binary block designs in the first instance, though the above example shows how to work around this restriction in some cases.

Among the aims of the project are:

- An XML specification for the *external representation* of block designs, in which the design and a wide variety of its combinatorial and statistical properties are described in a uniform computer-readable form for exchange between different programs and databases;
- The DESIGN package for GAP, for constructing and analysing block designs, and reading and writing them in the external representation format;
- A growing database of designs, with interfaces to make it easily usable by both combinatorialists and statisticians;
- The *Encyclopaedia of Design Theory*, an on-line resource including descriptions of several classes of design, essays on various topics in design theory and its mathematical and statistical underpinning, and an extensive glossary, as well as links to other resources;
- A collection of preprints on design theory.

Contributions to some or all of these components of the project are welcomed.

Acknowledgements

The authors are especially grateful to Ching-Shui Cheng and J. P. Morgan, whose extensive and detailed comments have greatly improved the accuracy and completeness of this paper. We also thank Emil Vaughan, to whose researches much of the information on the graph-theoretic interpretation of A-optimality are due; Francesca Merola for a number of suggestions; and Angela Dean, Heiko Großmann and other participants at the Isaac Newton Institute programme on Design of Experiments in July–August 2008.

References

- [1] S. Ajoodani-Namini, Extending large sets of t -designs, *J. Combinatorial Theory Series A* **76** (1996), 139–144.
- [2] N. Alon, Eigenvalues and expanders, *Combinatorica* **6** (1986), 83–96.
- [3] N. Alon & V. Milman, λ_1 , isoperimetric inequalities for graphs, and superconcentrators, *J. Combinatorial Theory Series B* **38** (1985), 73–88.
- [4] A. C. Atkinson & A. N. Donev, *Optimum Experimental Designs*, Oxford University Press, Oxford, 1992.
- [5] R. A. Bailey, *Association Schemes*, Cambridge Studies in Advanced Mathematics **84**, Cambridge University Press, Cambridge, 2004.
- [6] R. A. Bailey, Six families of efficient resolvable designs in three replicates. *Metrika*, **62** (2005), 161–173.
- [7] R. A. Bailey, Designs for two-colour microarray experiments, *Applied Statistics* **56** (2007), 365–394.
- [8] R. A. Bailey, *Design of Comparative Experiments*, Cambridge Series in Statistical and Probabilistic Mathematics **25**, Cambridge University Press, Cambridge, 2008.
- [9] R. A. Bailey, Variance and concurrence in block designs, and distance in the corresponding graphs, *Michigan Math. J.*, to appear.
- [10] R. A. Bailey, P. J. Cameron & R. Connelly, Sudoku, gerechte designs, resolutions, affine space, spreads, reguli, and Hamming codes, *American Math. Monthly* **115** (2008), 383–404.
- [11] R. A. Bailey, P. J. Cameron, P. Dobcsányi, J. P. Morgan & L. H. Soicher, Designs on the Web, *Discrete Math.* **306** (2006), 3014–3027.
- [12] R. A. Bailey, H. Monod & J. P. Morgan, Construction and optimality of affine-resolvable designs, *Biometrika* **82** (1995), 187–200.
- [13] B. Bagchi & S. Bagchi, Optimality of partial geometric designs, *Annals of Statistics* **29** (2001), 577–594.
- [14] N. Balabanian & T. A. Bickart, *Electrical Network Theory*, Wiley, 1969.
- [15] E. Bendito, A. Carmona, A. M. Encinas & J. M. Gesto, A formula for the Kirchhoff index, *Internat. J. Quantum Chemistry* **108** (2008), 1200–1206.
- [16] N. Biggs, *Algebraic Graph Theory*, 2nd ed., Cambridge University Press, Cambridge, 1993.
- [17] N. L. Biggs, Potential theory on distance-regular graphs. *Combinatorics, Probability and Computing*, **2** (1993), 107–119.

- [18] E. J. Billington, Balanced n -ary designs: a combinatorial survey and some new results, *Ars Combinatoria* **17** (1984), 37–72.
- [19] E. J. Billington, Designs with repeated elements in blocks: a survey and some recent results, *Congressus Numerantium* **68** (1989), 123–146.
- [20] B. Bollobás, *Modern Graph Theory*, Graduate Texts in Mathematics **184**, Springer, New York, 1998.
- [21] R. C. Bose & K. R. Nair, Partially balanced incomplete block designs, *Sankhyā* **4** (1939), 337–372.
- [22] N. Bourbaki, *Lie Groups and Lie Algebras* (transl. A. Pressley), Springer, 1998, 2002, 2005.
- [23] A. E. Brouwer, A. M. Cohen & A. Neumaier, *Distance-Regular Graphs*, Ergebnisse der Mathematik und ihrer Grenzgebiete (3), **18**, Springer-Verlag, Berlin, 1989.
- [24] R. H. Bruck & H. J. Ryser, The nonexistence of certain finite projective planes, *Canadian J. Math.* **1** (1949), 88–93.
- [25] P. J. Cameron, Optimal designs and root systems, *Michigan Math. J.*, to appear.
- [26] P. J. Cameron, A generalisation of t -designs, *Discrete Math.*, in press.
- [27] P. J. Cameron, J.-M. Goethals, J. J. Seidel & E. E. Shult, Line graphs, root systems and elliptic geometry, *J. Algebra* **43** (1976), 305–327.
- [28] P. J. Cameron & C. Thomassen, Spanning trees and orientations of graphs, in preparation.
- [29] A. K. Chandra, R. Raghavan, W. L. Ruzzo, R. Smolensky & P. Tiwari, The electrical resistance of a graph captures its commute and cover times, *Computational Complexity* **6** (1996), 312–340.
- [30] C.-S. Chêng, Optimality of certain asymmetrical experimental designs, *Annals of Statistics* **6** (1978), 1239–1261.
- [31] C.-S. Chêng, Maximizing the total number of spanning trees in a graph: two related problems in graph theory and optimum design theory. *J. Combinatorial Theory Series B* **31** (1981), 240–248.
- [32] C.-S. Cheng, Graph and optimum design theories—some connections and examples, *Bull. Internat. Statistical Institute* **49** (1) Proceedings of the 43rd Session (Buenos Aires) (1981), 580–590.
- [33] C.-S. Cheng, On the optimality of (M.S)-optimal designs in large systems, *Sankhyā* **54** (1992), 117–125.
- [34] C.-S. Cheng & R. A. Bailey, Optimality of some two-associate-class partially balanced incomplete-block designs, *Annals of Statistics* **19** (1991), 1667–1671.

- [35] C.-S. Cheng, J. C. Masaro & C. S. Wong, Do nearly balanced multigraphs have more spanning trees? *J. Graph Theory* **8** (1985), 342–345.
- [36] S. Chowla & H. J. Ryser, Combinatorial problems, *Canadian J. Math.* **2** (1950), 93–99.
- [37] CycDesignN 3.0, <http://cycdesign.co.nz>
- [38] G. Davidoff, P. Sarnak & A. Valette, *Elementary Number Theory, Group Theory, and Ramanujan Graphs*, London Mathematical Society Student Texts **55**, Cambridge University Press, Cambridge, 2003.
- [39] N. Deo, *Graph Theory with Applications to Engineering and Computer Science*, Prentice Hall, New Delhi, 1980.
- [40] P. Dobcsányi, D. A. Preece & L. H. Soicher, On balanced incomplete-block designs with repeated blocks, *European J. Combinatorics* **28** (2007), 1955–1970.
- [41] J. Dodziuk, Difference equations, isoperimetric inequality and transience of certain random walks, *Trans. American Math. Soc.* **284** (1984), 787–794.
- [42] R. M. Foster, The average impedance of an electrical network, in *Reissner Anniversary Volume, Contributions to Applied Mechanics*, J. W. Edwards, Ann Arbor, Michigan, 1948, pp. 333–340.
- [43] R. M. Foster, An extension of a network theorem, *IRE Trans. Circuit Theory* **8** (1961), 75–76.
- [44] A. Fries & W. G. Hunter, Minimum aberration 2^{k-p} designs, *Technometrics* **22** (1980), 601–608.
- [45] N. Gaffke, *Optimale Versuchsplanung für linear Zwei-Faktor Modelle*, PhD thesis, Rheinisch-Westfälische Technische Hochschule, Aachen, 1978.
- [46] The GAP Group, GAP – Groups, Algorithms, and Programming, Version 4.4.10, 2007, <http://www.gap-system.org>
- [47] A. Giovagnoli & H. P. Wynn, Optimum continuous block designs, *Proc. Royal Soc. London Series A* **377** (1981), 405–416.
- [48] C. Godsil & G. Royle, *Algebraic Graph Theory*, Springer, New York, 2001.
- [49] R. I. Grigorchuk & A. Zuk, On the asymptotic spectrum of random walks on infinite families of graphs, in *Random Walks and Discrete Potential Theory* (ed. M. Picardello *et al.*), Symposia Mathematica **39**, Cambridge University Press, Cambridge, 1999, pp. 188–204.
- [50] M. Hall, Jr. & W. S. Connor, An imbedding theorem for balanced incomplete block designs, *Canadian J. Math.* **6** (1954), 35–41.
- [51] H. Hanani, Balanced incomplete block designs and related designs, *Discrete Math.* **11** (1975), 255–369.

- [52] A. Hedayat & J. Stufken, A relation between pairwise balanced and variance balanced block designs, *J. American Statistical Association* **84** (1989), 753–755.
- [53] J. E. Humphreys, *Reflection Groups and Coxeter Groups*, Cambridge Studies in Advanced Mathematics **29**, Cambridge University Press, Cambridge, 1990.
- [54] M. T. Jacobson & P. Matthews, Generating uniformly distributed random Latin squares, *J. Combinatorial Designs* **4** (1996), 405–437.
- [55] M. Jacroux & D. C. Whittinghill III, On the E- and MV-optimality of block designs having $k \geq v$, *Annals of the Institute of Statistical Math.* **40** (1988), 407–418.
- [56] M. A. Jafarizadeh, R. Sufani & S. Jafarizadeh, Calculating two-point resistances in distance-regular resistor networks, *J. Physics Series A (Mathematical and Theoretical)* **40** (2007), 4949–4972.
- [57] JMP ®, <http://www.jmp.com>
- [58] J. A. John & T. J. Mitchell, Optimal incomplete block designs, *J. Royal Statistical Soc. Series B* **39** (1977), 39–43.
- [59] J. A. John & E. R. Williams, Conjectures for optimal block designs, *J. Royal Statistical Soc. Series B* **44** (1982), 221–225.
- [60] B. Jones & J. A. Eccleston, Exchange and interchange procedures to search for optimal designs, *J. Royal Statistical Soc. Series B* **42** (1980), 238–243.
- [61] M. Kac, Can one hear the shape of a drum? *American Math. Monthly* **73** (1966), 1–23.
- [62] A. K. Kelmans & V. M. Chelnokov, A certain polynomial of a graph and graphs with an extremal number of trees, *J. Combinatorial Theory Series B* **16** (1974), 197–214.
- [63] M. K. Kerr & G. A. Churchill, Experimental design for gene expression microarrays, *Biostatistics* **2** (2001), 183–201.
- [64] G. B. Khosrovshahi & B. Tayfeh-Rezaie, Trades and t -designs, this volume, pp. ??–??.
- [65] J. Kiefer, On the nonrandomized optimality and randomized nonoptimality of symmetrical designs, *Annals of Mathematical Statistics* **29** (1958), 675–699.
- [66] J. Kiefer, Construction and optimality of generalized Youden designs, in *A Survey of Statistical Design and Linear Models* (ed. J. N. Srivastava), North-Holland, Amsterdam (1975), pp. 333–353.
- [67] G. Kirchhoff, Über die Auflösung der Gleichung, auf welche man bei der Untersuchung der linearen Verteilung galvanischer Ströme geführt wird. *Annals of Physical Chemistry*, **72** (1847), 497–508.

- [68] A. Lubotzky, R. Phillips & P. Sarnak, Ramanujan graphs, *Combinatorica* **8** (1988), 261–277.
- [69] G. A. Margulis, Explicit constructions of graphs without short cycles and low density codes, *Combinatorica* **2** (1982), 71–78.
- [70] A. W. Marshall & I. Olkin, *Inequalities: Theory of Majorization and its Applications*, Mathematics in Science and Engineering, **143**, Academic Press, Inc., New York–London, 1979.
- [71] C. Merino & D. J. A. Welsh, Forests, colorings and acyclic orientations of the square lattice, *Annals of Combinatorics* **3** (1999), 417–429.
- [72] B. Mohar, Some applications of Laplace eigenvalues of graphs, in *Graph Symmetry: Algebraic Methods and Applications* (ed. G. Hahn and G. Sabidussi), NATO ASI Series C, Kluwer, 1997, pp. 227–276.
- [73] J. P. Morgan, Optimal incomplete block designs, *J. American Statistical Association* **102** (2007), 655–663.
- [74] J. P. Morgan & S. K. Srivastav, The completely symmetric designs with block-size three, *J. Statistical Planning and Inference* **106** (2002), 21–30.
- [75] J. P. Morgan & N. Uddin, Optimal, non-binary, variance balanced designs, *Statistica Sinica* **5** (1995), 535–546.
- [76] J. G. Oxley, *Matroid Theory*, Oxford University Press, New York, 1992.
- [77] J. L. Palacios, Foster’s formulas via probability and the Kirchhoff index, *Methodology and Computing in Applied Probability* **6** (2004), 381–387.
- [78] L. Paterson, Circuits and efficiency in incomplete block designs, *Biometrika* **70** (1983), 215–225.
- [79] L. J. Paterson & P. Wild, Triangles and efficiency factors, *Biometrika* **73** (1986), 289–299.
- [80] L. Petingi & J. Rodriguez, A new technique for the characterization of graphs with a maximum number of spanning trees, *Discrete Math.* **244** (2002), 351–373.
- [81] C. Pritchard, Mistakes concerning a chance encounter between Francis Galton and John Venn, *British Society for the History of Mathematics Bulletin* **23** (2008), 103–108.
- [82] F. Pukelsheim, *Optimal Design of Experiments*, Wiley, New York, 1993.
- [83] The R project for statistical computing, <http://www.r-project.org>
- [84] C. R. Rao, *Linear Statistical Inference and its Applications*, 2nd ed., Wiley, New York, 1973.

- [85] B. Reck & J. P. Morgan, Optimal design in irregular BIBD settings. *J. Statistical Planning and Inference* **129** (2005), 59–84.
- [86] B. Reck & J. P. Morgan, E-optimal design in irregular BIBD settings, *J. Statistical Planning and Inference* **137** (2007), 1658–1668.
- [87] O. Reingold, S. Vadhan & A. Wigderson, Entropy waves, the zig-zag graph product and constant-degree expanders, *Annals of Math. (2)* **155** (2002), 157–187.
- [88] C. A. Rohde, Generalized inverses of partitioned matrices, *J. Soc. Industrial and Applied Math.* **13** (1965), 1033–1035.
- [89] R. Schwabe, *Optimum Designs for Multi-Factor Models*, Lecture Notes in Statistics **113**, Springer, New York, 1996.
- [90] J.-P. Serre, Répartition asymptotique des valeurs propres de l’opérateur de Hecke T_p , *J. American Math. Soc.* **10** (1997), 75–102.
- [91] K. R. Shah & B. K. Sinha, *Theory of Optimal Designs*, Lecture Notes in Statistics **54**, Springer-Verlag, New York, 1989.
- [92] L. W. Shapiro, An electrical lemma, *Math. Magazine* **60** (1987) 36–38.
- [93] L. H. Soicher, The DESIGN package for GAP, Version 1.3, 2006, <http://designtheory.org/software/gap-design/>
- [94] A. D. Sokal, The multivariate Tutte polynomial (alias Potts model) for graphs and matroids, *Surveys in Combinatorics 2005* (ed. B. S. Webb), London Mathematical Society Lecture Note Series **327**, Cambridge University Press, Cambridge, 2005, pp. 173–226.
- [95] S. K. Srivastav & J. P. Morgan, Optimality of designs with generalized group divisible structure, *J. Statistical Planning and Inference* **71** (1998), 313–330.
- [96] L. Teirlinck, Nontrivial t -designs without repeated blocks exist for all t , *Discrete Math.* **65** (1987), 301–311.
- [97] P. Tetali, Random walks and the effective resistance of networks, *J. Theoretical Probability* **4** (1991), 101–109.
- [98] P. Tetali, An extension of Foster’s network theorem, *Combinatorics, Probability and Computing* **3** (1994), 421–427.
- [99] K. D. Tocher, The design and analysis of block experiments, *J. Royal Statistical Soc. Series B* **14** (1952), 45–100.
- [100] T. Tjur, Block designs and electrical networks, *Annals of Statistics* **19** (1991) 1010–1027.
- [101] T. Tjur, An algorithm for optimization of block designs. *J. Statistical Planning and Inference* **36** (1993) 277–282.

- [102] E. Vaughan, personal communication.
- [103] E. Vaughan, Foster's formulæ for distance-regular graphs, preprint, February 7, 2007.
- [104] P. Wild, On circuits and optimality conjectures for block designs, *J. Royal Statistical Soc. Series B* **49** (1987), 90–94.
- [105] P. Wild, Statistics, *Graph Connections*, Oxford Lecture Series in Mathematics and Applications **5**, Oxford University Press, New York, 1997, pp. 208–226.
- [106] R. M. Wilson, An existence theory for pairwise balanced designs: I, Composition theorems and morphisms, *J. Combinatorial Theory Series A* **13** (1972), 220–245; II, The structure of PBD-closed sets and the existence conjectures, *ibid.* **13** (1972), 246–273; III, A proof of the existence conjectures, *ibid.* **18** (1975), 71–79.
- [107] R. M. Wilson, Non-isomorphic Steiner triple systems, *Math. Zeitschrift* **135** (1974), 303–313.
- [108] E. Wit, A. Nobile & R. Khanin, Near-optimal designs for dual channel microarray studies, *Applied Statistics* **54** (2005), 817–830.

School of Mathematical Sciences
Queen Mary, University of London
London E1 4NS, UK
r.a.bailey@qmul.ac.uk, p.j.cameron@qmul.ac.uk