

Codes (Spherical) and Designs (Experimental)

R. H. Hardin and N. J. A. Sloane

ABSTRACT. An overview of the authors' work (partly in collaboration with W. D. Smith) on the problems of packing equal nonoverlapping spherical caps on a sphere in d dimensions, covering the sphere with equal caps in the most economical way, finding sets of points on a sphere with minimal energy, or whose convex hull has the greatest volume, or which can be used for numerical integration on the sphere (spherical t -designs), etc., as well as the problems of choosing good sets of points in the ball or cube to be used as experimental designs.

1. Preamble

A lot has been written about the connections between classical codes and combinatorial designs [1], [44], [58], [65]. That is *not* what this lecture is about! The codes we will discuss are sets of points on the unit sphere

$$\Omega_d = S^{d-1} = \{(x_1, \dots, x_d) \in \mathbb{R}^d : x_1^2 + \dots + x_d^2 = 1\},$$

and the designs are those used by experimenters: statisticians' designs, not mathematicians'.

There is a curious story behind our work on experimental designs. An article about the second author, dealing especially with the perhaps surprising connections between the problems of finding good packings of equal spheres in \mathbb{R}^d and of finding good signaling sets for digital communications, appeared in the October 1990 issue of the magazine *Discover* [5]. This article resulted in a large number of letters from readers. A few quotations will indicate why most of them could be tossed aside with a laugh:

"Amazing. Dr. K_____ breaks through $\sqrt{2}$. Solves one of Science's Unbreakable Enigmas ..."

"On page 697 of this report, you will find the single most important statement in the history of physics ..."

"It is convenient to assume that your 'Handbook of Integer Sequences' is the single most important mathematical 'discovery' in the last 3000

1991 *Mathematics Subject Classification*. Primary 62K05; Secondary 05B40, 11P05.

© 1995 American Mathematical Society

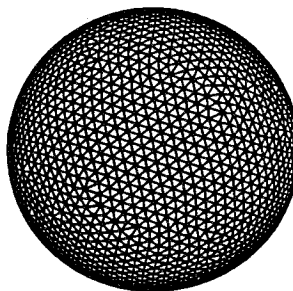
years [He is referring to [71].] ... Get ready for your annual paid vacation to Stockholm."

One letter, however, looked extremely interesting. It was written by a statistician, David Doehlert, of the Experiment Strategies Foundation in Seattle. He had come across the article by chance in an airplane, and immediately dashed off a note saying: Since you know how to pack spheres, I wish you would solve this problem, that has interested me for many years, and is of some importance in statistics: can you place 14 (or 15, or 16) points "nicely" on a sphere in four dimensions? What about 20 (or 21, or 22) points on a sphere in five dimensions, and so on?

2. Placing points on a sphere and related problems

Doehlert's letter had come to the right place. For many years the authors and Warren D. Smith have been building up tables of nice arrangements of points on spheres. The three of us call ourselves the 'Codemart' team, and our logo is shown in Figure 1. In fact we have even designed our own T-shirt, a photograph of which can be seen on page 308 of [30].

Another code from



CODEMART

R. H. Hardin, N. J. A. Sloane and W. D. Smith

**AT&T Bell Laboratories
Room 2C-376, Murray Hill, New Jersey 07974 USA**

FIGURE 1. The 'Codemart' logo.

What does a "nice" arrangement mean? We have considered several different criteria.

(P1) The packing problems. Place N points P_1, \dots, P_N on the sphere Ω_d so as to maximize the minimal distance between them. The distance is chordal distance, but an equivalent and nicer formulation is to choose P_1, \dots, P_N so as to

$$(1) \quad \text{minimize } \max_{i \neq j} P_i \cdot P_j,$$

regarding the points as unit vectors from the center of the sphere. A third formulation, also equivalent, is to maximize the minimal angle between the points, measuring angles from the center.

(P2) The covering problem. Choose P_1, \dots, P_N on Ω_d so as to minimize the maximal distance from a point of Ω_d to the nearest P_i :

$$\text{minimize } \max_{X \in \Omega_d} \max_{i=1, \dots, N} X \cdot P_i$$

One may think of the P_i as locations of fuel supplies on the moon, to be placed so that no matter where one lands on the moon, the distance to the nearest fuel supply is not too great.

The difference between the packing and covering problems is sometimes described in terms of convenience stores ("7-11" stores in some parts of the U.S.A.). From the point of view of the managers of the stores, they should be placed as far apart as possible, i.e. they should be a packing. But from the customers' point of view, the stores should be placed so that no one has too far to walk to the nearest store, i.e. they should form a covering.

(P3) The quantizing problem. Choose P_1, \dots, P_N on Ω_d so as to minimize the mean squared error when the points are used as a quantizer or analog-to-digital converter. A point X is chosen at random on Ω_d , and replaced by the closest $P_i(X)$. The P_i should be chosen so that the average error

$$\int_{X \in \Omega_d} |X - P_i(X)|^2 d\mu(X)$$

is minimized, where μ is uniform measure on the sphere.

(P4) The minimal error probability problem. Now the P_i represent signals in a communication system. When P_i is transmitted, the received signal is a neighboring point X on the sphere, where $|X - P_i|$ has a Gaussian distribution with mean 0 and variance σ^2 . The decoder replaces X by the closest P_j , and therefore makes an error if $j \neq i$. The problem is to choose $P_1, \dots, P_N \in \Omega_d$ so that the average probability of error is minimized.

(P5) The maximal volume problem. Choose $P_1, \dots, P_N \in \Omega_d$ so that the volume of their convex hull is maximized. The polytope formed by the convex hull of such a set of points has sometimes been used as an answer to the question: which N -vertex polytope inscribed in the unit sphere gives the best approximation to the sphere?

(P6) Best set of points for numerical integration. The idea is that we will approximate the integral of a function on the sphere by its average over our set of points:

$$(2) \quad \int_{\Omega_d} f(x) d\mu(x) \approx \frac{1}{N} \sum_{i=1}^N f(P_i).$$

(It is also possible to allow coefficients or *weights* on the right-hand side of (2), but we shall not consider this here.)

We say that $P_1, \dots, P_N \in \Omega_d$ form a *spherical t -design* if equality holds in (2) for all polynomials f of degree $\leq t$. For given values of d and t , the problem is

to find what values of N are possible, and in particular to determine the minimal value of N .

There are several related problems: for example, replace the sphere on the left-hand side of (2) by the ball $B_d = \{(x_1, \dots, x_d) \in \mathbb{R}^d: x_1^2 + \dots + x_d^2 \leq 1\}$, and choose $P_1, \dots, P_N \in B_d$ on the right-hand side.

Alternatively, there is the **interpolation problem**: we wish to interpolate f over Ω_d from its values at P_1, \dots, P_N . How should the P_i be chosen?

(P7) Minimal energy arrangements. Choose $P_1, \dots, P_N \in \Omega_d$ so as to minimize the potential energy

$$\sum_{i \neq j} \frac{1}{|P_i - P_j|^m}$$

for a given value of m (usually taken to be 2).

After Doehlert's letter arrived, we added another criterion:

(P8) Best statistical design. A simple version of this problem is: choose $P_1, \dots, P_N \in \Omega_d$ so that, when supplemented by an appropriate number of measurements at the center of the sphere, they form an optimal design for fitting a quadratic response surface.

Of course, as we shall discuss in Section 4, this is just one special case of a very large class of problems.

Two further problems should be mentioned, since we have successfully applied the same methods to them.

(P9) Construction of Isometric Embeddings. A set of points P_1, \dots, P_N in \mathbb{R}^d is said to form an *isometric embedding* of degree $2s$ if the following identity holds:

$$(3) \quad (x \cdot x)^{2s} = \sum_{i=1}^N (P_i \cdot x)^s,$$

where $x = (x_1, \dots, x_d)$ is a vector of indeterminates. This problem is also connected with numerical integration on the sphere (see [56], [66], [70]). We will discuss our results on this problem in Section 6.

(P10) Minimal energy arrangements of balls. Choose $P_1, \dots, P_N \in \mathbb{R}^d$ so that $|P_i - P_j| \geq 2$ for $i \neq j$ (so the balls have radius 1 and do not overlap) and the second moment

$$\sum_{i=1}^N |P_i - \bar{P}|^2$$

is minimized, where $\bar{P} = (1/N) \sum P_i$ is the centroid of the points. For our results on this problem, see [72].

The packing problem, (P1), is the most widely studied. It is sometimes known as the **Tammes problem** after the Dutch botanist P.M.L. Tammes who was led to this question by studying the distribution of pores on pollen grains [74]. The following rather long list of references has been chosen from a much greater list (see the bibliography in [42]) to illustrate the richness of this subject: [11], [13], [15], [17], [19], [28], [29], [31]–[35], [36], [50], [53], [54], [55], [57], [59], [67], [68], [69], [73], [75]–[78], [81].

Our own work on all these problems, and on the packing problem in particular, will be described in our book-in-preparation [42] (see also the next section).

Problems (P2), (P3) and (P4) are related in that they can all be expressed in terms of the Voronoi regions of the points. Problem (P2) asks that the greatest circumradius of any Voronoi cell should be minimized, problem (P3) that the average second moment of the Voronoi cells be minimized, and so on. Not so much has been written about these problems (see [14], [79]). For problem (P5) see [6], [42] and the following section. We defer discussion of problems (P6) and (P8) to Sections 5 and 4 respectively.

The minimal energy problem, (P7), also has an extensive literature: see [4], [24], [26], [52], [59], [64], [80], and especially [42] and the following section.

Spherical codes. Strictly speaking, by analogy with classical coding theory, only the solutions to problems (P1) and (P4) should be called *spherical codes*. However, we often use the term more generally to refer to any nice arrangement of points on Ω_d .

Comparison of problems. At first one expects that the answers to all these problems will be the same: a good packing of 24 points should also be a good covering, and so on. Nothing could be further from the truth. Except in very special cases, the answers to most of these problems are different. The case of 12 points in 3 dimensions is one of the exceptions: we believe the set of twelve vertices of a regular icosahedron solves all these problems. But in general it seems that each of problems (P1), (P2), (P5), (P6) and (P7) requires a different arrangement of points for its solution. (We have not investigated problems (P3) and (P4), and problem (P8) in the form we have stated it is very similar to the case $t = 4$ of (P6), the construction of spherical 4-designs; so we exclude problems (P3), (P4) and (P8) from this discussion.)

We illustrate by discussing the case of $N = 24$ points in 3 dimensions, since this is one of the few cases where the optimal solution to any of these problems has been established theoretically. Robinson [67] showed in 1961 that the best packing of 24 points is achieved by the vertices of a regular snub cube, one of the Archimedean solids (cf. [20]). According to our tables, however (see the next section), the 24-point solutions to the covering problem (P2), the maximal volume problem (P5), and the minimal energy configuration (P7) are all different from each other and from the regular snub cube. Furthermore the vertices of the regular snub cube form only a spherical 3-design (see Problem (P6)). We have recently discovered [41] that by modifying the shape of the snub cube slightly the vertices can be made to form a spherical 7-design: we call this the "improved snub cube".

The following table compares the different solutions, giving both the minimal angle between the points and the order of the symmetry group.

Problem	Min. angle	Group
Packing (P1)	43.691°	24
Min. energy (P7)	42.065°	24
t -design (P6)	41.376°	24
Max. volume (P5)	40.512°	4
Covering (P2)	36.673°	4

The first three solutions are all snub cubes, consisting of the 24 points that can be obtained from a single point (A, B, C) by applying any even permutation followed by changing any even number of signs, or any odd permutation followed by changing any odd number of signs. For the packing problem A, B, C are respectively .8503, .4623, .2514, giving the regular snub cube; for the minimal energy problem they are .8616, .4416, .2503; and for our new spherical 7-design, the "improved snub cube", they are .8662, .4225, .2666 (see [41] for details). The last two configurations in the table have a completely different structure, being both less symmetrical and having only triangular faces. Figure 2 shows the 24-point packing (left-hand picture), covering (center picture) and maximal volume arrangement (right-hand picture). The pictures show the convex hulls of the points. The regular snub cube in the left-hand picture has a diagonal drawn on each of its square faces. The "improved snub cube" and the 24-point minimal energy configurations are not shown, but in appearance are almost indistinguishable from the regular snub cube.

Although we do not know for certain that these are the optimal solutions to the minimal energy, maximal volume, and covering problems with 24 points, we strongly believe that they are. (It should not be difficult to show that there cannot be a single 24-point arrangement that beats the three designs mentioned for these three problems.)

We find this very convincing evidence that in general these problems have different solutions.

Incidentally it would be useful to have some theorems connecting the different problems. For example, if θ_p (resp. θ_c) is the maximal angular separation in the best N -point packing (resp. covering), then of course $\theta_c \leq \theta_p$. How small can θ_c be with respect to θ_p ? Similar questions can be asked about the other problems.

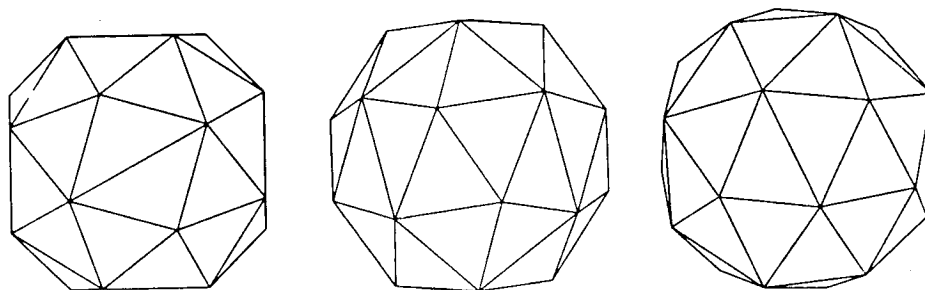


FIGURE 2. Best 24-point packing and putatively best 24-point covering and maximal-volume arrangements.

3. Our tables

Our initial response to Doehrlert's letter was to extract an appropriate collection of packings from our tables and send them to him. (We will discuss his answer in the next section.)

Our tables have been constructed using a variety of techniques, and include putative solutions to problems (P1)–(P10) (except (P2) and (P3), which we have not yet considered) for various dimensions (d) and numbers of points (N).

Our main tool, at least for 'small' numbers of points (say with $Nd \leq 500$), is an optimizer that is a modification of the Hooke and Jeeves [43] *pattern search*, and which is described in greater detail in [38] (see also [3]). We have made effective use of suitably modified versions of this optimizer for attacking all of these problems.

Our tables extend far beyond any others that have been published. We have several reasons for believing that our solutions are either optimal or very close to optimal: (a) in the case of the packing problem, where there has been a considerable amount of prior work, our solutions (with seven exceptions¹) are at least as good as the old records; (b) in the case of experimental designs for fitting linear models, where again a considerable amount was already known about optimal solutions, our optimizer was usually able to find these solutions in the range $Nd \leq 500$ (see [38] for details); and (c) as we shall discuss in Section 5, similar remarks apply to the case of quadratic response surface designs in the sphere.

We are in the process of publishing all these tables, principally in the book [42] that we are preparing. But in view of the considerable recent interest in these problems, we are making these tables available before the book is completed.

The tables are in netlib, and can be accessed via email or the netlib archive the directories

att/math/sloane/packings
att/math/sloane/coverings
att/math/sloane/maxvolumes
att/math/sloane/sphdesigns
att/math/sloane/electrons
att/math/sloane/gosset
att/math/sloane/iedir
att/math/sloane/cluster

OCT 2005. THE
TABLES ARE NOW
ON NJAS'S HOME
PAGE!

contain solutions to problems P1, P2, P5-P10 respectively, and

att/math/sloane/icosahedral

contains solutions to problems P1, P2 and P5 with icosahedral symmetry. Incidentally there are similar directories for Hadamard matrices, orthogonal arrays, minimal Lennard-Jones potential packings, and constant weight codes, as well as a directory att/math/sloane/doc that contains many of our papers.

We first describe how to access the archive by email. One can find out what is available by sending a message to netlib@research.att.com containing lines like

send index for att/math/sloane/packings
send index for att/math/sloane/packings/dim4
send index for att/math/sloane/doc

The return mail will have a list of available material. For example, here is one of the items from att/math/sloane/packings/dim4:

file att/math/sloane/packings/dim4/pack.4.24
by R.H. Hardin, N.J.A. Sloane & W.D. Smith
for File pack.4.24 contains the coords of the putatively optimal
packing of 24 points on a sphere in 4 dimensions
For more information see att/math/sloane/packings/readme
Copyright 1994 by R.H. Hardin, N.J.A. Sloane and W.D. Smith

¹For $d = 3$ and $N = 21, 27, 33, 66, 74, 81$ and 86 , Kottwitz's packings [50], [51] were better than ours. With his permission, we have added his packings, duly credited, to our tables.

To order this item, send the message

send att/math/sloane/packings/dim4/pack.4.24

or equivalently

send pack.4.24 from att/math/sloane/packings/dim4

to netlib@research.att.com. The file will be sent in one or more pieces. The mail header should be stripped from each piece to produce a shell file. The shell files may be run in any order; the original document will be created as the last shell file is run. (Of course this particular example would fit in a single file.)

For example, you might receive a large file in four chunks, which could arrive in any order. Strip off the mail headers, and call them say temp1, ..., temp4, in any order. Then do \$ sh temp1, ..., \$ sh temp4. At the end, the requested file will have been created.

The default limit for the size of each chunk is just under one megabyte. If this is too large for your mail program to handle, include a line such as

mailsize 40k

in your request. This will limit the size of the chunks to 40 KB.

The request

send getting.stuff from att/math

will produce further instructions.

Second, ftp access. Connect by ftp to netlib.att.com, login as anonymous, and use your email address as password. Now type

binary

cd netlib/att/math/sloane/packings

for example. You can then use "ls" to see what files are available, "get" to fetch them, "cd" to move to subdirectories, etc., and "quit" to quit.

(The files you receive will be compressed, and end with .Z, e.g. pack.4.24.Z. Do \$ uncompress pack.4.24.Z to uncompress them.)

Third, from a Mosaic document viewer, you can get directly to the material. Open an URL address such as:

ftp://netlib.att.com/netlib/att/math/sloane/doc/index.html.Z

ftp://netlib.att.com/netlib/att/math/sloane/packings/index.html.Z

ftp://netlib.att.com/netlib/att/math/sloane/packings/dim4/index.html.Z

This will give you a screen of short descriptions of all the items available. Clicking on one gives you the document, which can then be saved in a file.

At the present time the archive contains packings in dimensions 3, 4, 5 with $N \leq 130$ points, as well as an extensive table of larger 3-dimensional packings with icosahedral symmetry. We are in the process of adding a much larger table of packings in up to 24 dimensions that we have constructed by other techniques.

We would greatly appreciate hearing² of improvements to any of these tables, and will be glad to include them in the archive, giving credit to the discoverer.

Figures 3-10 show some examples of our spherical codes. Figure 3 shows our best packings, coverings and maximal volume arrangements for 72 and 100 points.

²Send them to N.J.A. Sloane, preferably by electronic mail to njas@research.att.com, or by fax to (908) 582 3340, or by regular mail to N.J.A. Sloane, Room 2C-376, AT&T Bell Labs, Murray Hill, NJ 07974, USA.

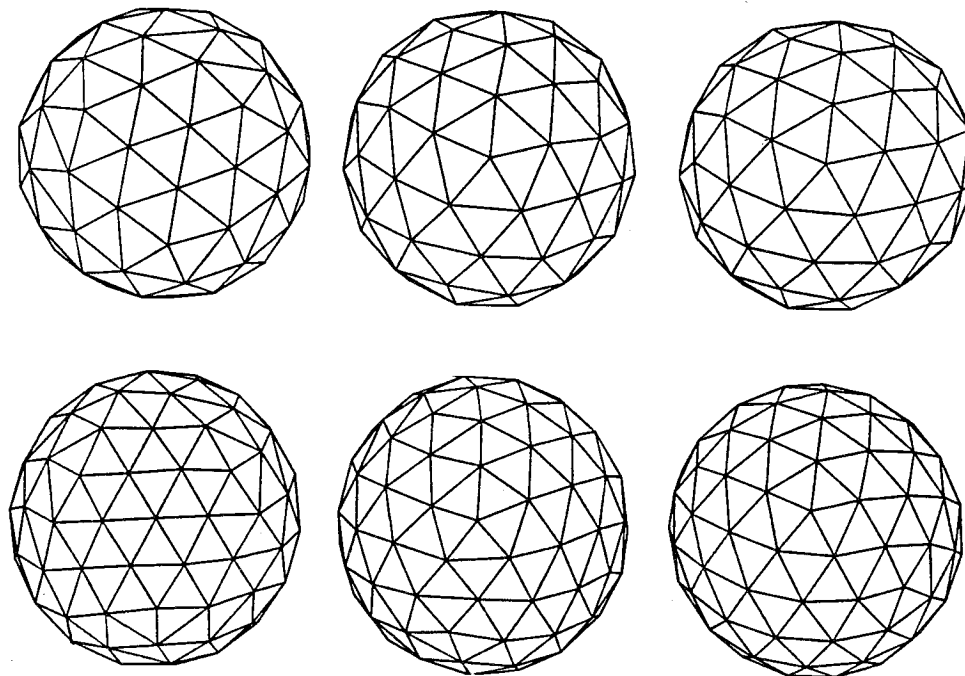


FIGURE 3. From left to right, top to bottom, the putatively best packings, coverings and maximal volume arrangements for 72 and 100 points respectively.

In Figure 3 (as in Figure 2) the pictures show the convex hulls of the arrangements. Figures 4–9 show packings with icosahedral symmetry and various numbers of points in the range 300 to 2000. In Figs. 4 and 5 we see the contact graph of the points (i.e. only points at the minimal separation are joined). The pictures have been projected stereographically with a horizon of 50° North. The 1952-point packing shown in Fig. 5 is on the front of the Codemart T-shirt. On the back is an 8192-point covering, shown in Fig. 6. Unlike the earlier pictures, which all show the best arrangements found, Fig. 6 was chosen for its aesthetic appearance, and is not the best 8192-point covering found. It has covering radius 1.408641° , which is very slightly inferior to the arrangement shown in Fig. 7, which has covering radius 1.404871° . The best 8192-*packing* found (Fig. 8) is much less pleasing to the eye.

We have found this to be true in general: coverings look nicer (and have larger symmetry groups) than packings.

After seeing Figs. 6–8 the reader may ask, why not simply cover the sphere completely with a grid of triangles, with six lines meeting at each point? Or why not simply place the points “uniformly” on the sphere, so that every point looks like every one (as in the icosahedron, for example)? The answers are that the first suggestion is simply impossible (it would violate Euler’s theorem, see [20], §3.6), and the second suggestion is only viable for small numbers of points. Jordan’s theorem (see for example [23], §30) gives an upper limit on how many points can be placed on Ω_d (for $d \geq 3$) in such a way that their symmetry group acts transitively.

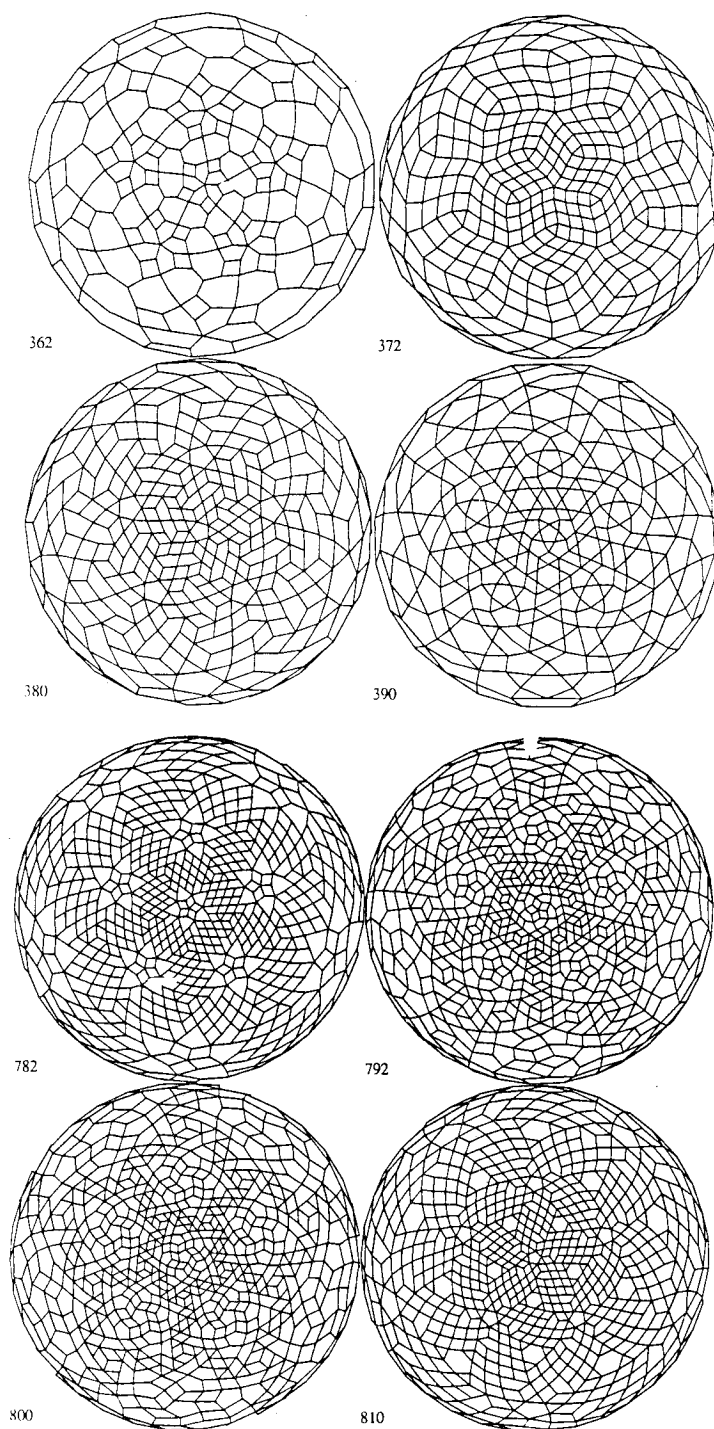


FIGURE 4. Best packings found of 362, 372, 380, 390, 782, 792, 800, 810 points assuming icosahedral symmetry.

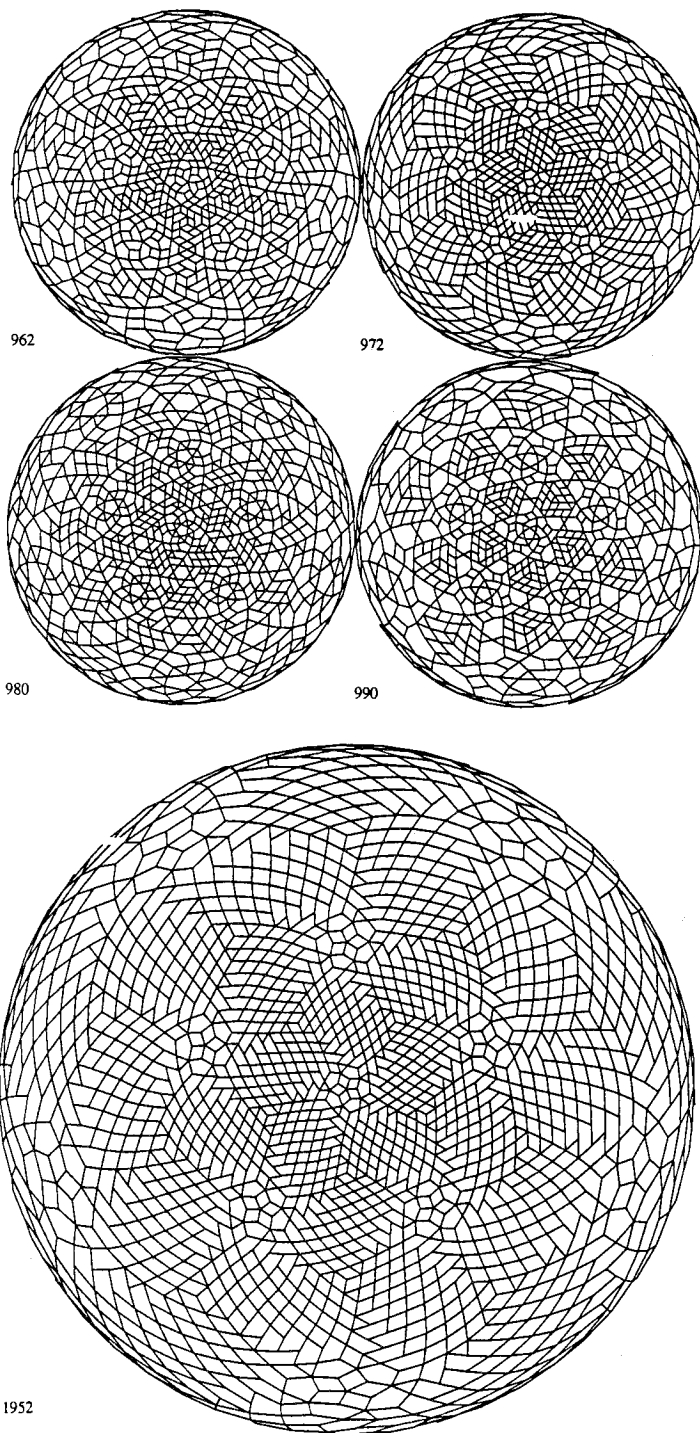


FIGURE 5. Best packings found of 962, 972, 980, 990 and 1952 points assuming icosahedral symmetry.

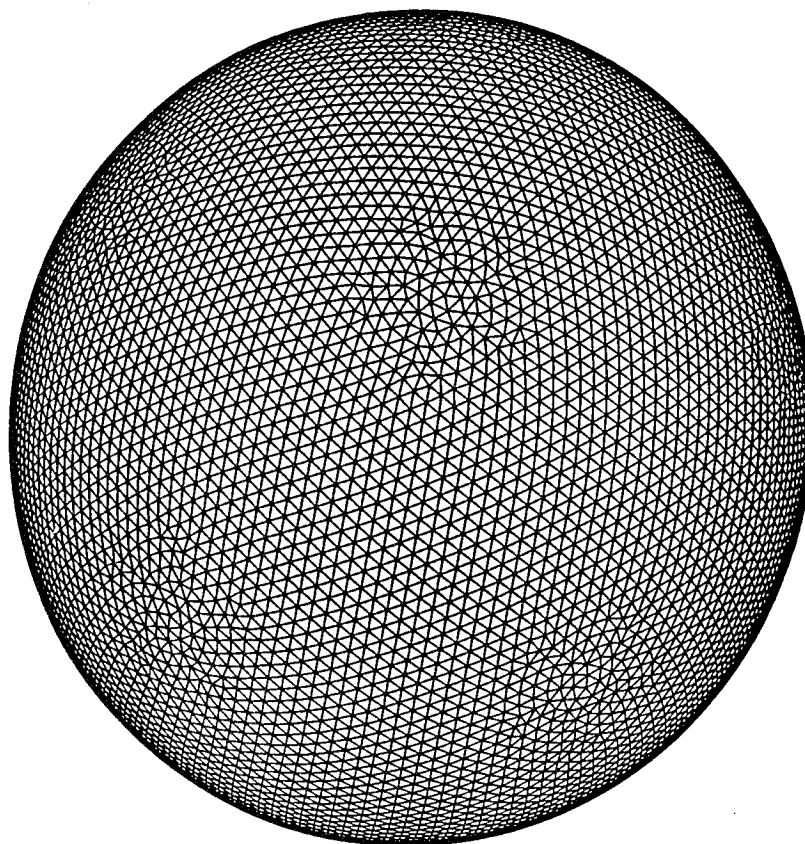


FIGURE 6. Second-best 8192-point covering with icosahedral symmetry.

For nondegenerate arrangements the limits in dimensions 3, 4, 5, 8 are respectively
120, 14400, 720, 696729600 .

Except for degenerate arrangements — such as placing them all on the equator — there is no way to place 121 or more points on Ω_3 so that they are all equivalent. Completely symmetric spherical codes with large numbers of points are impossible in any dimension above 2. On the one hand this means that the solutions aren't always pretty; on the other hand it makes the problems interesting.

4. Experimental designs and “Gosset”

We now return to our discussion of Doehlert's letter. We sent him a number of our best packings (of 14 to 20 points in 4 dimensions, 20–25 points in 5 dimensions, and so on), which he was very happy to receive and started using right away in his consulting work. However, he then explained that the problem he was really interested in was not exactly the packing problem, but rather the question of finding optimal experimental designs.

The following is a brief introduction to the subject of Design of Experiments. Further information can be found in any one of a large number of books — see for example [2], [9], [21], [46], [48], [63].

Your client is in charge of a refinery that produces oil (Fig. 9). The oil

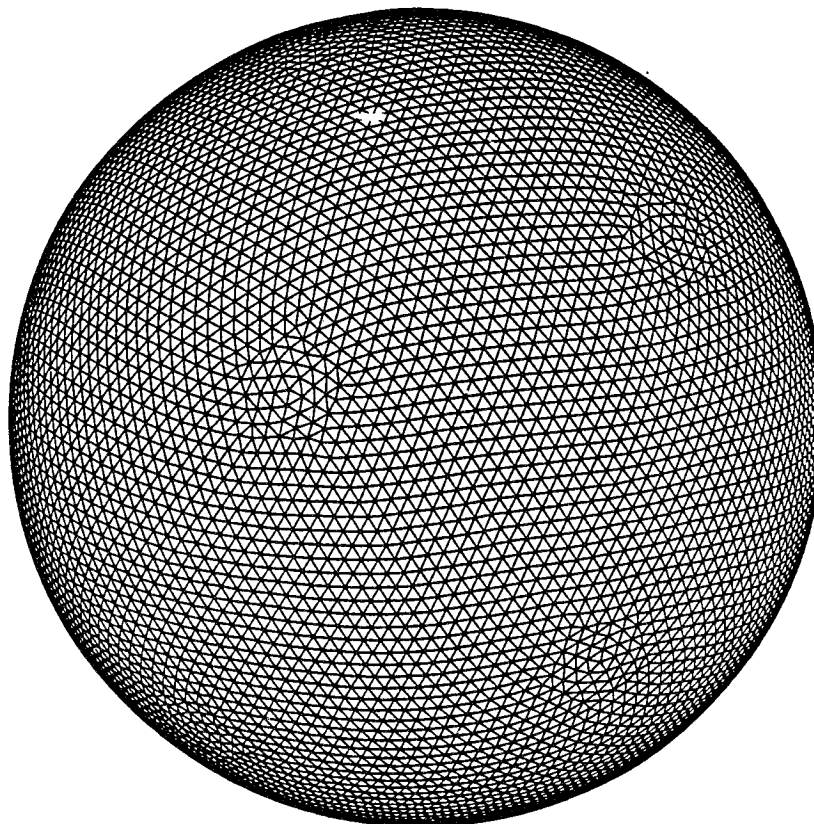


FIGURE 7. Best 8192-point covering with icosahedral symmetry.

output y depends on the settings of four variables, which we will pretend are $x_1 =$ temperature, $x_2 =$ pressure, $x_3 =$ cooking time, $x_4 =$ percentage of nitrogen. The client's problem is to adjust the settings of these variables so as to maximize the output. We suppose that the client has a rough idea of the best operating region, so that after rescaling the variables we may assume that the sought-for best operating point (the "sweet spot") is in a ball of radius 1 around the origin.

Let us assume now that the oil output y is well-represented by the constant, linear and quadratic terms of its Taylor series expansion in terms of the x_i . Thus our model for the refinery is

$$(4) \quad y = \beta_0 + \beta_1 x_1 + \cdots + \beta_4 x_4 + \beta_{11} x_1^2 + \cdots + \beta_{44} x_4^2 \\ + \beta_{12} x_1 x_2 + \cdots + \beta_{34} x_3 x_4 + \epsilon,$$

where the β_i are unknown coefficients, and ϵ is an error term. There are 15 coefficients β_i in this model (and more generally $\binom{n+2}{2}$ for a quadratic model involving n variables). So we will certainly have to make at least 15 measurements (or "runs") to determine the coefficients. In fact a minimal design would consist of one measurement at the center of the ball and 14 measurements on the boundary of the ball — this explains Doehlert's initial question about placing 14 points on a sphere in four dimensions!

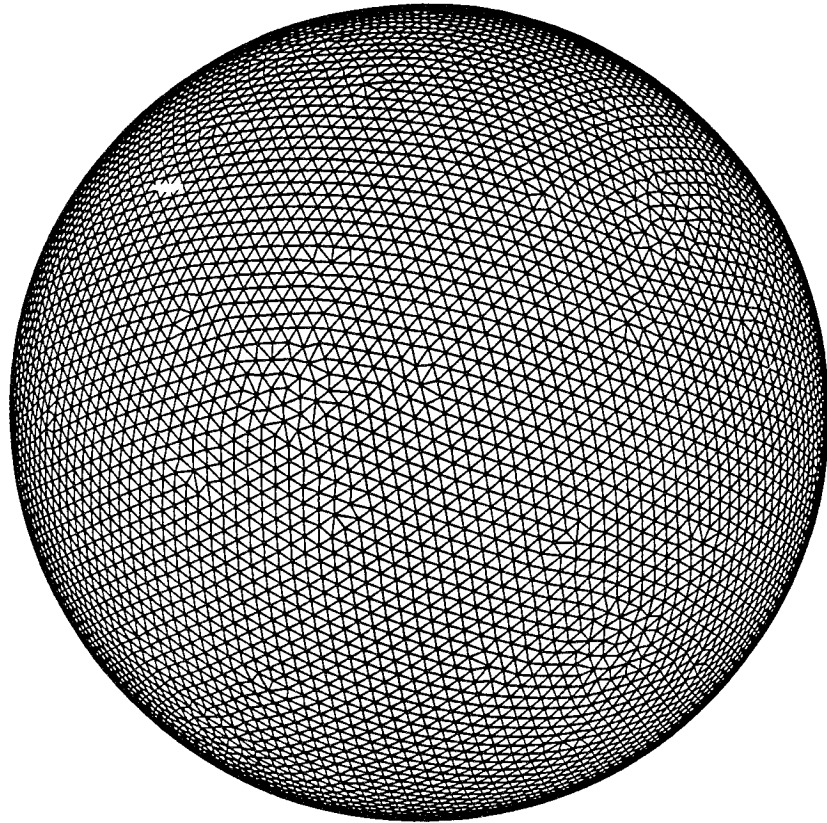


FIGURE 8. Best 8192-point packing with icosahedral symmetry.

But minimal designs are risky, for a variety of reasons, and a good design here would use some greater number p (say) of measurements, where $p \geq 15$, perhaps $p = 18$ or 20 . On the other hand p cannot be made too large, because the experiments are expensive and time consuming.

The problem in designing the experiment is to determine the "best" location for these p measurements. "Best" can be interpreted in many ways, and there is no universal agreement on which sense is best. Our own preference is very strongly in favor of what are called I -optimal designs.

In order to explain this, it is necessary to discuss what happens after the measurements have been made. We continue to use the quadratic model (4), although a similar analysis applies to other models.

Equation (4) can be written as

$$y = f(x)\beta^{tr} + \epsilon,$$

where $f(x) = (1, x_1, x_2, \dots, x_3x_4)$ is a vector listing the monomials that appear in the model, $\beta = (\beta_0, \beta_1, \dots, \beta_{34})$ is a vector of unknown coefficients, and tr denotes transpose. We will assume that the error ϵ can be represented as a random variable with mean 0 and variance σ^2 , and that the errors in different measurements are independent.

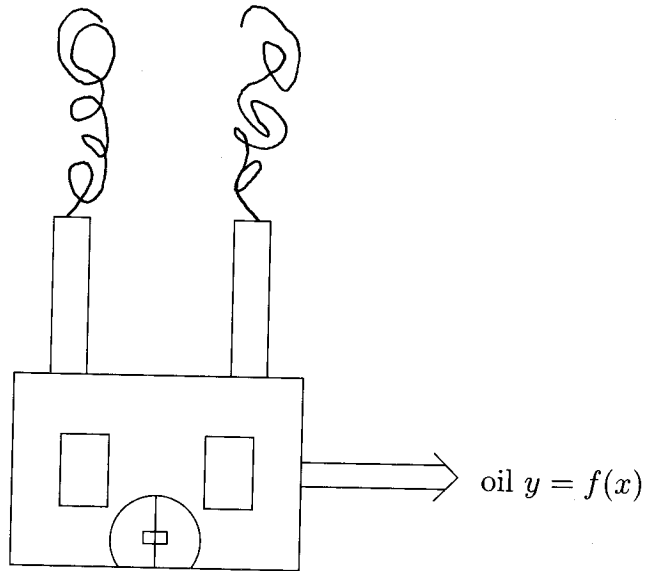


FIGURE 9

Suppose we make p measurements of y at the points

$$x^{(i)} = (x_1^{(i)}, x_2^{(i)}, x_3^{(i)}, x_4^{(i)}), \quad i = 1, \dots, p,$$

obtaining the values

$$Y = (y(x^{(1)}), \dots, y(x^{(p)})).$$

Then we have

$$Y^{tr} = X\beta^{tr} + Z^{tr},$$

where Z is a vector of errors, and the $p \times 15$ design matrix X is given by

$$X = \begin{pmatrix} f(x^{(1)}) \\ \vdots \\ f(x^{(p)}) \end{pmatrix}.$$

The best (least squares) estimate of β is then

$$\hat{\beta} = (X^{tr}X)^{-1}X^{tr}Y,$$

which has mean $E\hat{\beta} = \beta$, and covariance

$$\begin{aligned} \text{Cov}(\hat{\beta}) &= E(\hat{\beta} - \beta)^{tr}(\hat{\beta} - \beta) \\ &= \sigma^2(X^{tr}X)^{-1}. \end{aligned}$$

Our estimate for y is

$$\hat{y} = f(x)\hat{\beta}^{tr},$$

which has variance

$$\begin{aligned} \text{Var}(\hat{y}) &= E(\hat{y} - \bar{y})^2 \\ (5) \quad &= \sigma^2 f(x)(X^{tr}X)^{-1}f(x)^{tr}. \end{aligned}$$

This is the prediction variance.

An I -optimal design (see [7], [8]) is one which minimizes the *average* (or *integrated*) prediction variance over the region where we are fitting the model. Actually it is important to distinguish two difference regions: there is the *measurement* (or *operating*) region O , where we can make observations, and the *modeling region* R , where we want to fit the model. (In our program `gosset`, described below, these two regions can be specified independently.)

More precisely, an I -optimal design is one which minimizes the "integrated variance"

$$\begin{aligned} IV &= \int_R \frac{1}{\sigma^2} \text{Var } \hat{y}(x) d\mu(x) \\ &= \int f(x)(X^{tr}X)^{-1}f(x)^{tr}d\mu(x) \\ &= \text{trace} \{ M(X^{tr}X)^{-1} \}, \end{aligned}$$

where

$$M = \int_R f(x)^{tr}f(x)d\mu(x)$$

is the moment matrix of the modeling region. The I -efficiency of a design is the ratio IV_∞/IV , where IV_∞ denotes the limiting value of IV as $p \rightarrow \infty$.

In short, for an I -optimal design, we wish to choose p points in O so as to minimize

$$IV = \text{trace } M(X^{tr}X)^{-1}.$$

Other criteria that have often been used in the past include D -optimality, which minimizes $\det(X^{tr}X)^{-1}$, and A -optimality, which minimizes $\text{trace}(X^{tr}X)^{-1}$. Neither of these involve M , and we have found them to be unsatisfactory (except when the model is a polynomial of degree 1) for a number of reasons (see [38]). For later use we define the D -value of the design to be

$$D = (\det X^{tr}X)^{-1/p},$$

and the D -efficiency to be D_∞/D , where D_∞ is the limiting value of D as $p \rightarrow \infty$.

Doehlert's initial letter was followed by many subsequent ones, asking about optimal placements of points in the cube, in the simplex, then in more complicated regions, and for models which were not the full quadratic model shown in (4) but also quadratics with some terms omitted, cubics, etc. Over the course of the next three years we built up an ever-more complicated program that will attempt to find optimal designs for a very wide range of problems.

The program is called `gosset`, is written in C, and runs on any UnixTM platform. It is still being developed even today. However, we would like to hear from readers who might be interested in testing the present version of the program for us.

The program is named after the amateur mathematician *Thorold Gosset* (1869–1962), who was one of the first to study polytopes in six, seven and eight dimensions [16, p. 164], and his contemporary, the statistician *William Seally Gosset* (1876–1937), who was one of the first to use statistical methods in the planning and interpretation of agricultural experiments [62]. Although from our geometric viewpoint their work is related, we do not know if the paths of Thorold (Cambridge, London, lawyer), and William Seally (Oxford, Dublin, brewer) ever crossed.

Some of the program's features are the following. Only a few can be described here — for further information see [38], [39].

Problem specification. Problems are presented to the program using a very simple language, rather like "BASIC" — examples can be seen below. Because the program contains a built-in parser, the format is extremely flexible and user-friendly.

Variables. Variables may be discrete or continuous. Continuous variables may range over a sphere or a cube (or both). Discrete variables may be quantitative or qualitative (or both). Example:

10 range Temp Pressure 100 200

means that Temp and Pressure are continuous variables ranging between 100 and 200.

Constraints. The variables may be required to satisfy linear equalities and/or inequalities, so mixtures and constraints present no difficulty. Example:

20 constraint Temp + Pressure < 150

or

30 constraint $A + B + C + D = 100$

Models. The user can specify a model in a quite general way. In principle, almost any model at all can be used, provided it can be described by a C program! In most applications so far the models used have been low degree polynomials, but we expect this to change as this feature becomes more widely known. Example:

40 model (1 + Temp + Pressure) ↑ 2

defines a full quadratic model of the form

$$\begin{aligned} &\beta_0 + \beta_1 \text{Temp} + \beta_2 \text{Pressure} \\ &+ \beta_{11} \text{Temp}^2 + \beta_{12} \text{Temp Pressure} + \beta_{22} \text{Pressure}^2 \\ &+ \text{error term,} \end{aligned}$$

while

40 model (1 + x + sin(y)) ↑ 2 - sin(y) ↑ 2

specifies that the terms in the model are 1, x , x^2 , sin(y) and x sin(y).

If no model is available: packings. If no model is known, the user can simply ask for a "packing", that is, request the program to place N points in the operating region so that they are well-separated. As far as we know, this is the first time a program has been available that will search for packings in fairly arbitrary regions. Here is a simple 2-dimensional example, showing 100 points packed into a wedge-shaped region. The region was specified by

10 range a b 0 1

20 constraint $a < 4 * b$

The result is shown in Fig. 10.

Sequential designs. The design can be required to include a specified set of points. Thus a sequence of optimal designs can be constructed, each one building on the previous experiments.

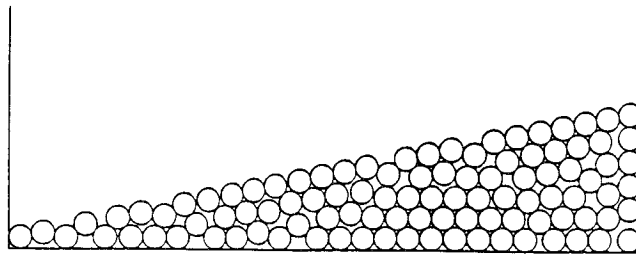


FIGURE 10

Measuring region and modeling region can be different: There are many situations where this feature is useful:

- some variables may take only discrete values (e.g. 5, 10, 25 or 100 grams), but it is desired to fit a model over the whole range
- measurements can be made over a large region, but we want an accurate model fitted over a smaller region
- extrapolation or prediction: we can make measurements only in a small region, but want to fit a model over a larger region.

Number of runs. The user can specify how many runs (or observations) the design will include. The user also has control over how hard the program works in attempting to find a good design. Example:

$$\text{design runs} = 20 \quad n = 30 \quad \text{processors} = 6$$

would instruct the program to look for a 20-run design, taking the best of 30 tries, and running 6 processors simultaneously. (Of course the latter feature is useful only on a multiprocessor machine.)

Optimality criteria. The program can search for *A*-, *D*-, *E*- or *I*-optimal designs. As already mentioned, our experience indicates that *I*-optimality is the most useful of these criteria, since all the others have serious drawbacks.

We have also introduced some further optimality criteria, useful in situations where it is possible that one of the measurements will fail, or otherwise be lost for some reason (we are told by experimenters that in practice this is not at all unusual).

For example, a *J*-optimal design with N runs has the property that the worst of the $N+1$ designs (the original design and those obtained by dropping one run in all possible ways) is *I*-optimized. In other words, *J*-optimal is "*I*-optimal given that an experiment may be lost". There are similar criteria for *A*-, *D*- and *E*-optimality.

Designs for situations where the errors are correlated. If the errors in the measurements, instead of being independent, have a known (or estimated) correlation, the program can take this into account. For example, Fig. 11 shows some designs intended for a situation in which successive pairs of measurements are correlated (the application was to an experiment making measurements on eyes, first the left eye, then the right eye). The design is specified by

$$10 \text{ sphere } X \ Y \ - \ 1 \ 1$$

$$20 \text{ model } (1 + X + Y) \uparrow 2$$

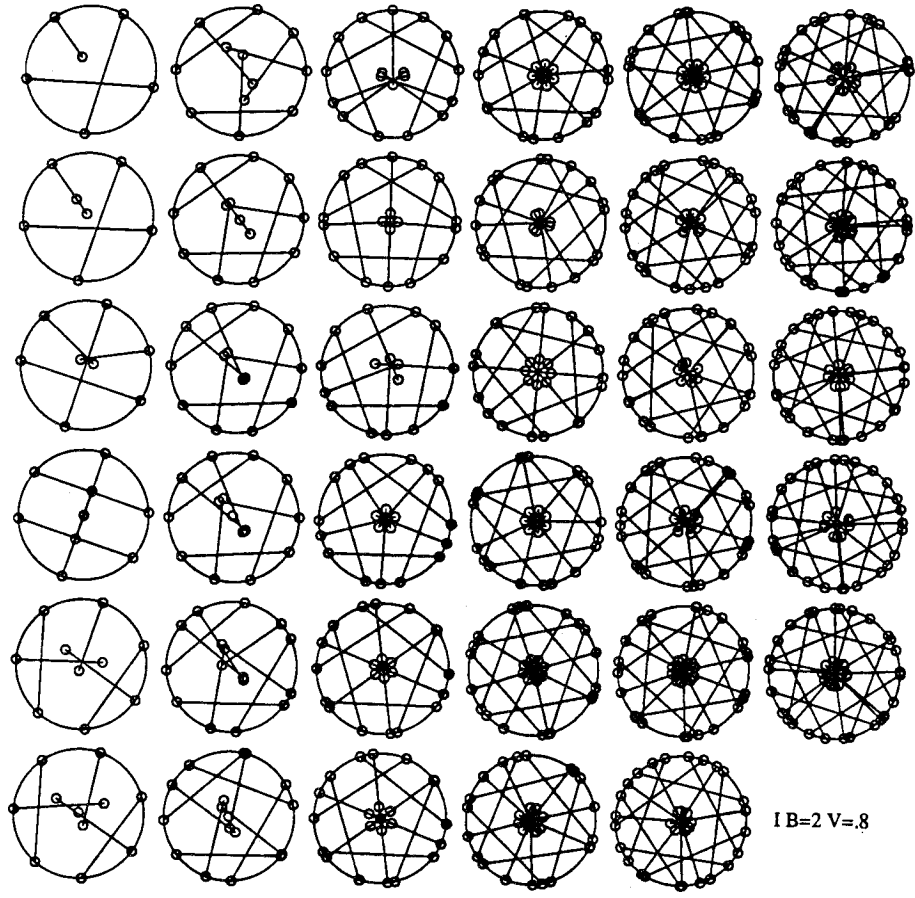


FIGURE 11

that is, a full quadratic model in a circle, in which the covariance matrix for the samples is

$$\begin{bmatrix} 1 & .8 & 0 & 0 & \dots \\ .8 & 1 & 0 & 0 & \dots \\ 0 & 0 & 1 & .8 & \dots \\ 0 & 0 & .8 & 1 & \dots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix}$$

The putatively *I*-optimal designs with 6, 7, 8, ... runs are as follows. (Pairs of successive design points are joined by lines.)

Blocked designs. A feature that has had many applications is the ability to look for a design in which the measurements are made in blocks. See the manual [39] for examples.

Applications. So far there have been two main uses for *gosset*.
 (i) We have attempted to construct optimal (especially *I*-optimal) designs for a number of "classical" situations, for example linear, quadratic or cubic response-surface designs with *n* continuous variables in a cube or ball with *p* experiments, over quite a large range of values of *n* and *p*, typically $1 \leq n \leq 12$ and *p* ranging

from the minimal value to 6 (or more) greater than the minimal value. We have also computed designs for similar models and regions in which the variables are discrete. An extensive library of these designs is now built into `gosset`. Our work on these "classical" problems can be regarded as an attempt to provide optimal "exact" designs with small numbers of runs to complement the "asymptotic" designs of Kiefer et al. (see [48]).

We have also used this collection of designs as data for theoretical investigations. Two results are worth mentioning here.

- (a) There is a simple lower bound on the integrated variance of an I -optimal design for quadratic models with p measurements in an n -dimensional ball (see the next section). A number of interesting designs meet the bound. There is a similar bound for D -optimal designs.
 - (b) It is known that for large numbers of experiments D - and G -optimal designs are equivalent [49]. Our results show that I -optimal designs are strictly different. For example, I -optimal designs make more measurements at the center of the region and fewer at the boundary. For quadratic models in an n -dimensional ball, n large, an I -optimal design makes about $4/n^2$ of the measurements at the center of the sphere, compared with about $2/n^2$ for D - and G -optimal designs. I -optimality also appears to be a more strict condition than D -optimality. In situations where the criteria produce similar designs (such as certain linear designs), we commonly find that although I -optimal designs are D -optimal, the converse is not necessarily true.
- (ii) We have constructed designs for a large number of industrial applications. These include:
- optimizing the production of wafers for integrated circuits (see for example [60])
 - placing laser beams for treatment of tumors in the brain
 - growing protein crystals
 - maximizing the thermal conductivity of heat-spreading diamond film
 - designing potted meat
 - designing a cellular ceramic substrate used in catalytic converters
 - designing coated photographic paper

5. The construction of spherical 4-designs

One of the most interesting things to have come out of our work with `gosset` is a series of discoveries that has led to the construction of many new spherical 4-designs (see Problem (P6) of Section 2). These results are described in more detail in [37] and [38].

These discoveries came about when we were comparing I - and D -optimal designs in the ball for the full quadratic model of (4), using the extensive library of designs that we had computed using `gosset`. Let $\mathcal{I}(n, p)$ (resp. $\mathcal{D}(n, p)$) denote any I -optimal (resp. D -optimal) design for this model, where n is the number of variables and p is the number of experiments, with $p \geq \binom{n+2}{2}$.

First, it appears that there is a unique design $\mathcal{I}(3, 14)$ (apart from orthogonal transformations), which consists of three copies of a point distant .003622 from the center of the ball and eleven points on the surface of the ball. The precise design may be found in the `gosset` library. However, restricting $\mathcal{I}(3, 14)$ to points

at the center and on the surface of the ball incurs a loss of only about .00006% in I -efficiency. A similar phenomenon occurs for other values of n and p .

In contrast, for D -optimal designs, it seems that the only points that occur are at the center and on the surface. We formalize these observations in the following conjecture, still unproved.

Conjecture. *For all n and p , $D(n, p)$ contains only points at the center and on the surface of the ball. This is not true for $I(n, p)$, but restricting $I(n, p)$ to designs with this property incurs a loss in I -efficiency of less than .01%.*

In the rest of this section we therefore restrict attention to designs supported only at the center and the surface of the ball. For any such design, not necessarily optimal, let B and C respectively denote the number of points on the surface (or boundary) and at the center, so that $p = B + C$.

Our second discovery was that the IV -value of the best designs was often given by the formula

$$(6) \quad \frac{1}{(n+2)(n+4)} \left\{ \frac{n^2(n^2+5n+10)}{2\beta} + \frac{8}{\gamma} \right\},$$

where $\beta = B/p$, $\gamma = C/p$. This was an empirical formula, found simply by examining our collection of designs (we observed that obviously rational IV -values were arising from many geometrically distinct designs, suggesting that there was a formula to be found). This formula was the key to everything that followed.

Our observations were that this formula is always a lower bound to the IV -value of a design, and that it could be attained once the number of experiments was sufficiently large. We were then able to prove these results, by considering the moments of the design points. Let ξ denote the discrete measure, normalized to have total measure 1, defined by the B surface points of the design. The B surface points form a spherical 4-design if and only if the moments of ξ up through order 4 agree with the moments of uniform measure on the sphere. The fact that (6) is a lower bound on IV is a consequence of the following result

THEOREM 5.1. *For fixed values of n , p , B and C , both the I - and D -efficiencies are maximized if the B surface points can be arranged to form a spherical 4-design.*

The proof for I -efficiency is given in [37], while for D -efficiency this follows from the work of Box and Hunter [10] and Kiefer [47] (see also [27], [61]).

If the surface points do form a spherical 4-design, the IV - and D -values of the design can be calculated analytically, and turn out to be given by (6) for the IV -value, thus confirming our empirical observations, and by

$$(7) \quad \frac{n}{\beta}(n+2)^{(n-1)/(n+2)} \left\{ \frac{\beta}{n\gamma 2^{n-1}} \right\}^{2/((n+1)(n+2))}$$

for the D -value. It follows from Theorem 5.1 that these are indeed lower bounds.

One can now select the values of β and γ to minimize (6) or (7) for any fixed value of n . For large n , (6) implies that in an I -optimal design the fraction of points at the center is

$$(8) \quad \gamma = \frac{4n\sqrt{n^2+5n+10}-16}{(n-1)(n+2)(n^2+4n+8)}.$$

(7) implies that in a D -optimal design this fraction is

$$(9) \quad \gamma = \frac{2}{(n+2)(n+4)},$$

which is a theorem of Kiefer [47].

By substituting these expressions in (6) and (7) we obtain the limiting values

$$(10) \quad IV_{\infty} = \frac{n+2}{2(n+4)} \left\{ \frac{(n-1)(n^2+4n+8)}{n\sqrt{n^2+5n+10}-4} \right\}^2,$$

$$(11) \quad D_{\infty} = \frac{(n+1)(n+2)^{2n/(n+1)}}{n+3} \left(\frac{n+3}{2^n} \right)^{2/((n+1)(n+2))},$$

which enable us to determine the I - and D -efficiencies of any given design.

It follows from a result of Neumaier and Seidel [61] that a design has D -value given by (7) if and only if the B surface points form a spherical 4-design. There is a similar result for I -values.

THEOREM 5.2. *A quadratic design in the ball has IV -value given by (6) if and only if the B surface points form a spherical 4-design.*

Thus whenever gosset is able to find a design that meets the bound (6), the surface points form a spherical 4-design. And the nonexistence of designs meeting the bound imply the nonexistence of the 4-design. Since our program is very good at finding these designs, we take this as compelling evidence that the corresponding 4-designs do not exist.

In summary, we believe that spherical 4-designs containing B points in n dimensions exist for the following values of B and n :

n	B
1	2, 4, 6, 8, ...
2	≥ 5
3	12, 14, ≥ 16
4	≥ 20
5	≥ 29
6	27, 36, ≥ 39
7	≥ 53
8	≥ 69

See [37] and [38] for further information about these spherical 4-designs, as well as [41] for our more recent work on the extension of this investigation to spherical t -designs for larger values of t .

6. Isometric embeddings

Finally, we discuss problem (P9), the construction of isometric embeddings. We will give just one example, and show that there exists a 23-point isometric embedding in \mathbb{R}^4 . Equivalently, we will show that $(x_1^2 + x_2^2 + x_3^2 + x_4^2)^3$ can be written as a sum of 23 sixth powers of linear forms. This is one less than is required in a famous 1912 identity of Kempner. The discussion is based on [40], while further examples can be found in [12].

In connection with Waring's problem of expressing integers as a sum of sixth powers, Lucas published in 1876 the incorrect identity

$$(x_1^2 + x_2^2 + x_3^2 + x_4^2)^3 = \frac{1}{10} \sum_{i < j}^{(12)} (x_i \pm x_j)^6,$$

and in the following year replaced it by

$$(x_1^2 + x_2^2 + x_3^2 + x_4^2)^3 = \frac{2}{5} \sum^{(4)} x_i^6 + \frac{1}{10} \sum_{i < j}^{(12)} (x_i \pm x_j)^6,$$

also incorrect (cf. Dickson [22], p. 718). A correct version,

$$(13) \quad (x_1^2 + x_2^2 + x_3^2 + x_4^2)^3 = \frac{8}{15} \sum^{(4)} x_i^6 + \frac{1}{15} \sum_{i < j}^{(12)} (x_i \pm x_j)^6 + \frac{1}{120} \sum^{(8)} (x_1 \pm x_2 \pm x_3 \pm x_4)^6$$

was apparently given for the first time by Kempner [45] in 1912.

More generally, for any even number q , we look for N points $P^{(i)} = [P_1^{(i)}, \dots, P_d^{(i)}] \in \mathbb{R}^d$ and positive coefficients $c_i \in \mathbb{R}$, $1 \leq i \leq N$, such that there is an identity

$$(14) \quad (x_1^2 + \dots + x_d^2)^{q/2} = \sum_{i=1}^N c_i \left(\sum_{j=1}^d P_j^{(i)} x_j \right)^q.$$

The existence of such identities, for all d and even q , with c_i and $P_j^{(i)}$ rational, was a key step in Hilbert's (1909) general solution of Waring's problem (Ellison [25]). Chapter XXV of Dickson [22] surveys the classical results concerning such identities. They have recently resurfaced in works by Lyubich and Vaserstein [56], Reznick [66] and Seidel [70], where many new results are given as well as connections with numerical quadrature, designs, and isometric embeddings of one space in another.

Let $N(d, q)$ denote the smallest value of N for which an identity (14) is possible. Equation (13) shows that $N(4, 6) \leq 24$, and Proposition 9.2 of [66] shows that $N(4, 6) \geq 21$. We will present an identity of the form

$$(15) \quad (x_1^2 + x_2^2 + x_3^2 + x_4^2)^3 = \sum_{i=1}^{23} c_i \left(\sum_{j=1}^4 P_j^{(i)} x_j \right)^6,$$

which establishes $N(4, 6) \leq 23$. The points $P^{(i)}$ and coefficients c_i are given in Table 1. (However, since some of the c_i and $P^{(i)}$ are irrational, (15) does not contribute to Waring's problem.)

Equation (15) was discovered in the following way. Let us rewrite it as $(\sum_{i=1}^4 x_i^2)^3 = \sum_{i=1}^{23} (\sum_{j=1}^4 Q_j^{(i)} x_j)^6$, where $Q_j^{(i)} = c_i^{1/6} P_j^{(i)}$. We used the computer to find sets of $Q_j^{(i)}$'s which satisfied this identity to within a small tolerance. The algorithm was simply a further modification of the one mentioned in Section 3. These sets of 23 points were then analyzed by hand. There appear to be many inequivalent solutions, and it was not easy to find a set of points with enough symmetry to identify their coordinates $Q_j^{(i)}$ as algebraic numbers. Having identified the $Q_j^{(i)}$, there is still the problem of finding a convenient way to factorize them

TABLE 1(A). The points $P^{(1)}, \dots, P^{(23)}$ in the identity (15). Double parentheses indicate that all cyclic shifts of the last three components are to be included.

[1	0	0	0]	=	[1	0	0	0]		
$[\alpha$	((0	± 1	$\pm \tau$))	=	[1.473370	((0	± 1	± 1.618034))
[0	1	1	1]	=	[0	1	1	1]		
[0	((1	$\tau - 1$	$-\tau$))	=	[0	((1	.618034	-1.618034))
[0	(($-\tau$	β	$\beta\tau + 1$))	=	[0	((-1.618034	1.868517	4.023324))
[0	(($-\tau$	γ	$\gamma\tau + 1$))	=	[0	((-1.618034	-0.535184	.134054))

as $c_i^{1/6} P_j^{(i)}$. There is no unique way to do this, and our aim was to make the final answer as simple as possible. The computer took a couple of hours to find each approximate solution, but "beautification" of the points by hand took a couple of weeks. We will not attempt to describe this fairly mysterious process, but just give our best answer, which emerged after a series of miraculous simplifications, in Table 1.

The $P^{(i)}$ and c_i are listed in Table 1, both exactly (as algebraic numbers) and approximately (as decimal numbers, rounded to six significant figures). The geometric structure of the points is as follows (there is a picture in [37]). $P^{(1)}$ is at the north pole, and $P^{(2)}$ to $P^{(13)}$ form a regular icosahedron (with x_1 coordinate $\sqrt{3\tau} 5^{-1/4}$, where $\tau = (1 + \sqrt{5})/2$). The remaining ten points lie on the equatorial hyperplane $x_1 = 0$ and consist of a singleton and three equilateral triangles. The vertices of those equilateral triangles lie on the edges (possibly produced) of the icosahedron (with x_1 coordinate omitted). The only symmetries of the configuration are the cyclic shifts of the last three components. However, if the negatives or antipodes of the points are included, the resulting 46-point configuration has symmetry group of structure $C_2 \times C_2 \times D_6$, where D_6 is a dihedral group of order 6. This group contains the cyclic shift just mentioned and the reflection in $[0, 1, 1/\tau, -\tau]$, which together generate a three-dimensional triangular antiprismatic group (denoted by $[2^+, 6]$ in the notation of [18], by $2 * 3$ in the Conway-Thurston orbifold notation, and with structure $C_2 \times D_6$). The full group is generated by this subgroup and negation of the first coordinate, where $\alpha = \sqrt{3\tau} 5^{-1/4}$, $\beta = (2 + \sqrt{13})/3$, $\gamma = (2 - \sqrt{13})/3$.

The points $P^{(i)}$ lie on six different spheres, one for each of the six rows of Table 1(a). (So in the notation of [66], equation (15) is a sixth-caliber representation of $(x_1^2 + \dots + x_4^2)^3$.)

As to the proof that (15) is an identity: the verification could be carried out by hand, but would be rather tedious. The c_i and $P^{(i)}$ are now simple enough, however, that a computer algebra system such as Maple is able to verify the identity exactly in a few seconds. Additional confirmation is provided by the fact that floating point evaluations of the coefficients on each side of (15) agree very closely.

Our computer investigations have failed to produce a 22-term identity, and we conjecture that $N(4, 6) = 23$.

TABLE 1(B). The coefficients c_1, \dots, c_{23} in the identity (15).

$$\begin{aligned} \frac{16}{25} &= .64 \text{ (once) ,} \\ \frac{1}{367^3\sqrt{5}} &= .00293258 \text{ (twelve times) ,} \\ \frac{16}{735} &= .0217687 \text{ (once) ,} \\ \frac{1}{120} &= .00833333 \text{ (three times) ,} \\ -\frac{1460 + \sqrt{5} - \sqrt{13}(577 - 76\sqrt{5})}{2^4 3^2 5 7^2 \sqrt{13}} &= .0000427301 \text{ (three times)} \\ \frac{1460 + \sqrt{5} + \sqrt{13}(577 - 76\sqrt{5})}{2^4 3^2 5 7^2 \sqrt{13}} &= .0230332 \text{ (three times)} \end{aligned}$$

We have also found sets of points which suggest that

$$N(3, 10) \leq 24, N(3, 12) \leq 32, N(3, 14) \leq 41, N(3, 16) \leq 52, N(3, 18) \leq 66,$$

$$N(4, 8) \leq 43, N(6, 6) \leq 63, N(7, 6) \leq 91, N(8, 4) \leq 45, N(9, 4) \leq 59 .$$

We are currently in the process of formally establishing these identities. Details will be found in [12].