Irish Mathematical Society
Cumann Matamaitice na hÉireann

Bulletin

Number 81 Summer 2018

ISSN 0791-5578

Irish Mathematical Society Bulletin

The aim of the *Bulletin* is to inform Society members, and the mathematical community at large, about the activities of the Society and about items of general mathematical interest. It appears twice each year. The *Bulletin* is supplied free of charge to members; it is sent abroad by surface mail. Libraries may subscribe to the *Bulletin* for 30 euro per annum.

The *Bulletin* seeks articles written in an expository style and likely to be of interest to the members of the Society and the wider mathematical community. We encourage informative surveys, biographical and historical articles, short research articles, classroom notes, book reviews and letters. All areas of mathematics will be considered, pure and applied, old and new. See the inside back cover for submission instructions.

Correspondence concerning the *Bulletin* should be sent, in the first instance, by e-mail to the Editor at

`ims.bulletin@gmail.com`

and only if not possible in electronic form to the address

The Editor
Irish Mathematical Society Bulletin
Department of Mathematics and Statistics
Maynooth University
Co. Kildare W23 HW31

Further information about the Irish Mathematical Society and its *Bulletin* can be obtained from the IMS webpage

`http://www.irishmathsoc.org/`

EDITORIAL

Due to circumstances beyond our control the long-standing arrangement for the printing of the Bulletin came to an end this year, and as a result the printed edition of Bulletin 80 did not appear until June. We apologise for the delay, and the consequent delay in publishing this Bulletin 81. In an attempting to make a virtue of necessity, we made a few design changes, and these appear to have been well-received by the membership.

We put the current IMS logo on the front cover, but retained Janine Mathieu's original logo on the inside title page.

As an incidental side-effect of the printing arrangements, individual copies Bulletin 80 may be purchased online at www.lulu.com for €15 plus postage. Search the Lulu site for *Irish Mathematical Society Bulletin*.

At the time of writing, it is not decided how the paper edition of this issue will be printed. If we use commercial printers again, the cost may be mitigated somewhat by revenue from an advertisement. (It has long been policy that suitable advertisements may appear in the Bulletin.)

Members may wish to take note of the revised online format of the *Basic Library List* of the MAA, which may be seen at: `maa.org/bll`. There is an account by Darren Glass in the February-March 2018 issue of MAA Focus (page 6).

This issue includes the concluding part of the survey on curvature and topology by Mark Walsh. Part I appeared in Bulletin 80.

Links for Postgraduate Study

The following are the links provided by Irish Schools for prospective research students in Mathematics:

DCU: mailto://maths@dcu.ie

DIT: mailto://chris.hills@dit.ie

NUIG: mailto://james.cruickshank@nuigalway.ie

MU: mailto://mathsstatspg@mu.ie

QUB: http://www.qub.ac.uk/puremaths/Funded_PG_2016.html

TCD: http://www.maths.tcd.ie/postgraduate/

UCC: http://www.ucc.ie/en/matsci/postgraduate/

UCD: mailto://nuria.garcia@ucd.ie

UL: mailto://sarah.mitchell@ul.ie

UU: http://www.compeng.ulster.ac.uk/rgs/

The remaining schools with Ph.D. programmes in Mathematics are invited to send their preferred link to the editor, a url that works. All links are live, and hence may be accessed by a click, in the electronic edition of this Bulletin[1].

AOF. Department of Mathematics and Statistics, Maynooth University, Co. Kildare
E-mail address: ims.bulletin@gmail.com

[1] http://www.irishmathsoc.org/bulletin/

CONTENTS

NOTICES FROM THE SOCIETY

Officers and Committee Members 2018

President	Prof S. Buckley	Maynooth University
Vice-President	Dr Pauline Mellon	UCD Dublin
Secretary	Dr D. Malone	Maynooth University
Treasurer	Prof G. Pfeiffer	NUI Galway

Dr P. Barry, Dr L. Creedon, Prof J. Gleeson, Dr R. Levene, Dr D. Mackey, Dr M. Mathieu, Dr A. Mustata, Dr J. O'Shea .

Local Representatives

Belfast	QUB	Dr M. Mathieu
Carlow	IT	Dr D. Ó Sé
Cork	IT	Dr D. Flannery
	UCC	Dr S. Wills
Dublin	DIAS	Prof T. Dorlas
	DIT	Dr D. Mackey
	DCU	Dr B. Nolan
	SPD	Dr S. Breen
	TCD	Prof R. Timoney
	UCD	Dr R. Higgs
Dundalk	IT	Mr Seamus Bellew
Galway	UCG	Dr J. Cruickshank
Limerick	MIC	Dr B. Kreussler
	UL	Mr G. Lessells
Maynooth	NUI	Prof S. Buckley
Tallaght	IT	Dr C. Stack
Tralee	IT	Dr B. Guilfoyle
Waterford	IT	Dr P. Kirwan

Applying for I.M.S. Membership

(1) The Irish Mathematical Society has reciprocity agreements with the American Mathematical Society, the Deutsche Mathematiker Vereinigung, the Irish Mathematics Teachers Association, the Moscow Mathematical Society, the New Zealand

Mathematical Society and the Real Sociedad Matemática Española.

(2) The current subscription fees are given below:

Institutional member	€160
Ordinary member	€25
Student member	€12.50
DMV, I.M.T.A., NZMS or RSME reciprocity member	€12.50
AMS reciprocity member	$15

The subscription fees listed above should be paid in euro by means of a cheque drawn on a bank in the Irish Republic, a Eurocheque, or an international money-order.

(3) The subscription fee for ordinary membership can also be paid in a currency other than euro using a cheque drawn on a foreign bank according to the following schedule:

If paid in United States currency then the subscription fee is US$ 30.00.
If paid in sterling then the subscription is £20.00.
If paid in any other currency then the subscription fee is the amount in that currency equivalent to US$ 30.00.

The amounts given in the table above have been set for the current year to allow for bank charges and possible changes in exchange rates.

(4) Any member with a bank account in the Irish Republic may pay his or her subscription by a bank standing order using the form supplied by the Society.

(5) Any ordinary member who has reached the age of 65 years and has been a fully paid up member for the previous five years may pay at the student membership rate of subscription.

(6) Subscriptions normally fall due on 1 February each year.

(7) Cheques should be made payable to the Irish Mathematical Society. If a Eurocheque is used then the card number should be written on the back of the cheque.

(8) Any application for membership must be presented to the Committee of the I.M.S. before it can be accepted. This Committee meets twice each year.

(9) Please send the completed application form with one year's subscription to:

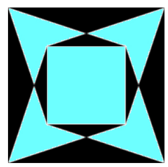
The Treasurer, IMS
School of Mathematics, Statistics and Applied Mathematics
National University of Ireland
Galway
Ireland

Deceased Members

It is with regret that we report the deaths of members:

David Simms died on 24 June 2018. He was a founder member of the Society.

E-mail address: subscriptions.ims@gmail.com



Irish Math. Soc. Bulletin
Number 81, Summer 2018, 5–22
ISSN 0791-5578

Sparsification of Matrices and Compressed Sensing

FINTAN HEGARTY, PADRAIG Ó CATHÁIN AND YUNBIN ZHAO

Abstract. Compressed sensing is a signal processing technique whereby the limits imposed by the Shannon–Nyquist theorem can be exceeded provided certain conditions are imposed on the signal. Such conditions occur in many real-world scenarios, and compressed sensing has emerging applications in medical imaging, big data, and statistics. Finding practical matrix constructions and computationally efficient recovery algorithms for compressed sensing is an area of intense research interest. Many probabilistic matrix constructions have been proposed, and it is now well known that matrices with entries drawn from a suitable probability distribution are essentially optimal for compressed sensing.

Potential applications have motivated the search for constructions of sparse compressed sensing matrices (i.e., matrices containing few non-zero entries). Various constructions have been proposed, and simulations suggest that their performance is comparable to that of dense matrices. In this paper, extensive simulations are presented which suggest that sparsification leads to a marked improvement in compressed sensing performance for a large class of matrix constructions and for many different recovery algorithms.

1. Introduction

Compressed sensing is a new paradigm in signal processing, developed in a series of ground-breaking publications by Donoho, Candès, Romberg, Tao and their collaborators over the past ten years or so [14, 8, 9]. Many real-world signals have the special property of being *sparse* — they can be stored much more concisely than a random signal. Instead of sampling the whole signal and then applying data compression algorithms, sampling and compression of sparse signals can be achieved simultaneously. This process requires dramatically

2010 *Mathematics Subject Classification.* 94A12, 90C05.

Key words and phrases. compressed sensing, optimisation, sparsification.
Received on 0-0-0000; revised 0-0-0000.

fewer measurements than the number dictated by the Shannon–Nyquist Theorem, but requires complex measurements which are *incoherent* with respect to the signal. The compressed sensing paradigm has generated an explosion of interest over the past few years within both the mathematical and electrical engineering research communities.

A particularly significant application has been to Magnetic Resonance Imaging (MRI), for which compressed sensing can speed up scans by a factor of five [23], either allowing increased resolution from a given number of samples or allowing real-time imaging at clinically useful resolutions. A major breakthrough achieved with compressed sensing has been real-time imaging of the heart [35, 24]. The US National Institute for Biomedical Imaging and Bioengineering published a news report in September 2014 describing compressed sensing as offering a "vast improvement" in paediatric MRI imaging [19]. Emerging applications of compressed sensing in data mining and computer vision were described by Candès in a plenary lecture at the 2014 International Congress of Mathematicians [7].

The central problems in compressed sensing can be framed in terms of linear algebra. In this model, a signal is a vector v in some high-dimensional vector space, $\mathbb{R}^N$. The sampling process can be described as multiplication by a specially chosen $n \times N$ matrix Φ, called the *sensing matrix*. Typically we will have $n \ll N$, so that the problem of recovering v from Φv is massively under-determined.

A vector is *k-sparse* if it has at most k non-zero entries. The set of k-sparse vectors in $\mathbb{R}^N$ plays the role of the set of compressible signals in a communication system. The problem now is to find necessary and sufficient conditions so that the inverse problem of finding v given Φ and Φv is efficiently solvable.

If u and v are distinct k-sparse vectors for which $\Phi u = \Phi v$, then one of them is not recoverable. Clearly, therefore, we require that the images of all k-sparse vectors under Φ are distinct, which is equivalent to requiring that the null-space of Φ does not contain any $2k$-sparse vectors. There is no known polynomial time algorithm to certify this property. We refer to the problem of finding the sparsest solution $\hat{x}$ to the linear system $\Phi\hat{x} = \Phi x$ as the *sparse recovery problem*. Natarajan has shown that certain instances of this problem are NP-hard [27].

Compressed sensing (CS) can be regarded as the study of methods for solving the sparse recovery problem and its generalizations (e.g., sparse approximations of non-sparse signals, solutions in the presence of noise) in a computationally efficient way. Most results in CS can be characterized either as certifications that the sparse recovery problem is solvable for a restricted class of matrices, or as the development of efficient computational methods for sparse recovery for some given class of matrices.

One of the most important early developments in CS was a series of results of Candès, Romberg, Tao and their collaborators. They established fundamental constraints for sparse recovery: one cannot hope to recover k-sparse signals of length N in less than $O(k \log N)$ measurements under any circumstances[1]. (For $k = 1$, standard results from complexity theory show that $O(\log N)$ measurements are required.) The main tools used to prove this result are the *restricted isometry parameters* (RIP), which measure how the sensing matrix Φ distorts the ℓ_2-norm of sparse vectors. Specifically, Φ has the RIP(k, ϵ) property if, for every k-sparse vector v, the following inequalities hold:

$$(1 - \epsilon)|v|_2^2 \leq |\Phi v|_2^2 \leq (1 + \epsilon)|v|_2^2.$$

Tools from Random Matrix Theory allow precise estimations of the RIP parameters of certain random matrices. In particular, it can be shown that the *random Gaussian ensemble*, which has entries drawn from a standard normal distribution, is asymptotically optimal for compressed sensing, i.e., the number of measurements required is $O(k \log N)$. A slightly weaker result is known for the *random Fourier ensemble*, a random selection of rows from the discrete Fourier transform matrix [8, 9, 33].

As well as providing examples of asymptotically optimal compressed sensing matrices, Candès et al. provided an efficient recovery algorithm: they showed that, under modest additional assumptions on the RIP parameters of a matrix, ℓ_1-minimization can be used

[1]Throughout this paper we use some standard notation for asymptotics: for functions $f, g : \mathbb{N} \to \mathbb{N}$, we write that $f = O(g)$ if there exists a constant C, not depending on n, such that $f(n) < Cg(n)$ for all sufficiently large n. Intuitively, the function g eventually dominates f, up to a constant factor. We say that $f = \Theta(g)$ if there exist two constants c, C such that $cg(n) < f(n) < Cg(n)$ for all sufficiently large n. Hence, f and g grow at the same rate.

for signal recovery. Thus efficient signal recovery is possible in large systems, making applications to real-world problems feasible.

Generation and storage of random matrices are potential obstacles to implementations of CS. It is also difficult to design efficient signal recovery algorithms capable of exploiting the structure of a random matrix. For implementation in real-world systems, it is desirable that CS constructions produce matrices that are *sparse* (possess relatively few non-zero entries), *structured*, and *deterministically constructed*. Systems with these properties can be stored implicitly, and efficient recovery algorithms can be designed to take advantage of their known structure. If Φ is $n \times N$ with d non-zero entries per column, then computing Φv takes $\Theta(dN)$ operations, which is a significant saving when $d \ll n$. In some applications, signals are frequently subject to rank-one updates (i.e., v is replaced by $v + \alpha e_i$), in which case the image vector can be updated in time $O(d)$, see [38].

Motivated by real-world applications, a number of papers have explored CS constructions where the Gaussian ensemble is replaced by a sparse random matrix (e.g., coming from an expander graph or an LDPC code) [3, 2, 18], or by a matrix obtained from a deterministic construction [11, 15, 16]. But to date, constructions meeting all three criteria have either been asymptotic in nature (i.e., the results only produce matrices that are too large for practical implementations), or are known only to exist for a very restricted range of parameters. This investigation was inspired by work of the second author on constructions of sparse CS matrices from pairwise balanced designs and complex Hadamard matrices [6, 5]. Some related work on constructing CS matrices from finite geometry is contained in [20, 37].

In this paper we take a new approach. Rather than constructing a sparse matrix and examining its CS properties, we begin with a matrix which is known to possess good CS properties (with high probability) and explore the effect of sparsification on this matrix. That is, we set many of the entries in the original matrix to zero, and compare the performance of the sparse matrix with the original. Results of Guo, Baron and Shamai suggest that sparse matrices should behave similarly to dense matrices in our regime [17]. Surprisingly, we actually observe an **improvement** in signal recovery as the sparsity increases.

First we survey some previous work on sparse compressed sensing matrices. Then in Section 3, we give a formal definition of sparsification, and describe algorithms used to generate random matrices and random vectors, as well as the recovery algorithms. In Section 4 we describe the results of extensive simulations. These provide substantial computational evidence which suggests that sparsification is a robust phenomenon, providing benefits in both recovery time and proportion of successful recoveries for a wide range of random and structured matrices occurring in the CS literature. In particular, Table 2 shows the benefits of sparsification for a range of matrix constructions, while Figure 2 illustrates how sparsification improves performance for a range of CS recovery algorithms. Finally, in Section 5 we conclude with some observations and open questions motivated by our numerical experiments.

2. Tradeoffs between sparsity and compressed sensing

A number of authors have investigated ways of replacing random ensembles with more computationally tractable sensing matrices. As previously mentioned, foundational results of Candès et al. establish asymptotically sharp results: to recover signals of length N with k non-zero entries, $n = \Theta(k \log N)$ measurements are necessary. Work of Chandar established that when $n = \Theta(k \log N)$, then the columns of Φ must contain at least $\Theta(\min\{k, N/n\})$ non-zero entries [10]. In [29], Nelson and Nguyen establish an essentially optimal result when $n = \Theta(k \log N)$ and $k < N/\log^3 N$. They show that each column of Φ necessarily contains $\Theta(k \log N)$ non-zero entries; i.e., the proportion of non-zero entries in Φ cannot tend to zero as N tends to ∞.

Observe that some restriction on k as a function of N is necessary: in the limiting case $k = N$, the identity matrix clearly suffices as a sparse sensing matrix. Furthermore, combinatorial constructions of sparse matrices are known which have near optimal recovery guarantees with a mutual incoherence property[2] [6]. In such matrices $n = O(k^2)$, and for certain infinite families of matrices (e.g., those

[2]Informally, the k-RIP property requires that k-sets of columns of Φ approximate an orthonormal basis. The incoherence of a matrix is the maximal inner product of a pair of columns, which is essentially the 2-RIP of Φ. In contrast to k-RIP, efficient constructions of matrices with near-optimal incoherence are known [6], but they have sub-optimal compressed sensing performance. Using 2-RIP alone, asymptotically one requires at least k^2 measurements to recover k-sparse

coming from projective planes) the number of non-zero entries in each column is $\Theta(k)$. Results bounding errors in the ℓ_1 norm (so-called RIP-1 guarantees) have been obtained using expander graphs. In particular, Bah and Tanner have shown that essentially optimal RIP-1 recovery can be achieved when $\lim_{n\to\infty} N/n = \alpha$ for some fixed α, with a constant number of non-zero entries per column [1]. (See also the discussion of dense versus sparse matrices in Section 3 of this paper.) These bounds are strictly weaker than RIP-2 bounds, though fast specialised algorithms have been developed for signal recovery with such matrices [32].

Since the k-RIP property is difficult to establish in practice, some authors have relaxed this in various directions. Berinde, Gilbert, Indyk, Karloff and Strauss [3] considered random binary matrices with constant column sum and related these to the incidence matrices of expander graphs. We reinterpret these matrices as sparsifications of the all-ones matrix below. Sarvotham, Baron and Baraniuk [2] and Dimakis, Smarandache and Vontobel [12] have considered the use of LDPC matrices. In particular, they have provided a strong correspondence between error-correcting performance of LDPC codes (when considered over $\mathbb{F}_2$) and CS performance of the same binary matrices (when considered over $\mathbb{R}$). While both groups obtained essentially optimal CS performance guarantees, their constructions are limited by the lack of known explicit constructions for expander graphs and LDPC codes respectively. Moghadam and Radha have previously considered a two step construction of sparse random matrices, involving construction of a random $(0, 1)$-matrix followed by replacing each entry 1 with a sample from some probability distribution, [25, 26].

If one is content with recovery of each sparse vector with high probability, then much sparser matrices become useful. A strong result in this direction is due to Gilbert, Li, Porat and Strauss, who show that there exist matrices with $n = O(k \log N)$ rows and $\Theta(\log^2 k \log N)$ non-zero entries per column which recover sparse vectors with probability 0.75 [16] (see also [34]). Their matrices also come with efficient encoding, updating and recovery algorithms. While essentially optimal results are known for sparsity bounds on

signals. No practical deterministic construction is known which uses asymptotically fewer measurements; this obstruction is known as the *square-root bottleneck*, and finding more efficient deterministic constructions is a major open problem in compressed sensing.

CS matrices with an optimal number of rows, much less is known when either some redundant rows are allowed in the construction, or when RIP is replaced with a slightly weaker condition.

Several authors have compared the performance of sparse and dense CS matrices [17, 36, 13, 21]. Guo, Baron and Shamai have essentially shown that in certain limiting cases of the sparse recovery problem, dense and sparse sensing matrices behave in a surprisingly similar manner. In particular, they consider a variant of the recovery problem: given Φ and Φx, what can one say about any single component of x? They show that, as the size of the system becomes large (in a suitably controlled way), the problem of estimating x_i becomes independent of estimating x_j. In fact, the problem is equivalent to recovering a single measurement of x_i contaminated by additive Gaussian noise. They also apply their philosophy to sparse matrices, where as the size of the matrix becomes large, estimation of **all** signal components becomes independent, and each can be recovered independently. We refer the reader to the original paper for technical details [17]. As a result, under their assumptions, there should be no essential difference between CS performance in the sparse and dense cases.

For any compressed sensing matrix Φ, denote by $\delta(\Phi)$ the proportion of non-zero entries in Φ. Suppose that the number of columns of Φ is a linear function of the number of rows, i.e., that $N = \alpha n$ for some fixed $\alpha \in \mathbb{R}$. Suppose further that $\lim_{n\to\infty} \delta(\Phi) n^{1-\epsilon} = 0$ for any $\epsilon > 0$, but that $\lim_{n\to\infty} \delta(\Phi) n$ diverges. (Consider $\delta(\Phi) = \log n/n$ for example.) Under these hypotheses, Guo, Baron and Shamai claim that sparse and dense matrices exhibit identical CS performance. In particular, there exist matrices with $k \log k$ rows and $\Theta(\log k)$ non-zero entries per row which recover vectors of sparsity $O(k)$.

This appears to be in conflict with Nelson and Nguyen's result, which requires at least $\Theta(n \log\log n/ \log n)$ non-zero entries in such a matrix. The difference is that Nelson and Nguyen's result holds only when $n = O(k \log^3 N)$, whereas Guo, Baron and Shamai consider the case $n = \Theta(N)$.

Wang, Wainwright and Ramchandran's analysis of the number of measurements required for signal recovery depends on the quantity $(1 - \delta(\Phi))k$, which can be considered a measure of how much information about x is captured in each co-ordinate of Φx. They show that if $(1-\delta(\Phi))k \to \infty$ as $N \to \infty$ (which corresponds to relatively dense matrices, where, in the limit, each component of the signal is sampled infinitely often), then sparsification has no effect on recovery, while if $(1-\delta(\Phi))k$ remains bounded (so each component of the signal is sampled only finitely many times in expectation), then what the authors term "dramatically more measurements" are required. We refer the reader to the original paper for more details [36].

Our simulations are close in spirit to those considered by Guo, Baron and Shamai. Our computations are rather surprising as they suggest a modest **improvement** in signal recovery as we apply a sparsifying process to certain families of CS matrices. This improvement seems to persist across different recovery algorithms and different matrix constructions, and does not appear to have been noted in any of the work discussed in this section. (Though Lu, Li, Kpalma and Ronsin have observed some improvement in CS performance for sparse binary matrices [22].) We also observed a substantial improvement in the running times of the recovery algorithms, which may be of interest in practical applications.

3. Sparsification

We begin with a formal definition of sparsification.

Definition 3.1. The matrix Φ' is a *sparsification* of Φ if $\Phi'_{i,j} = \Phi_{i,j}$ for every non-zero entry of Φ'. The *density* of Φ, denoted $\delta(\Phi)$, is the proportion of non-zero entries that it contains, and the *relative density* of Φ' is the ratio $\delta(\Phi')/\delta(\Phi)$. We write $\mathrm{Sp}(\Phi, s)$ for the set of all sparsifications of Φ of relative density s.

In general, we have that $\mathrm{Sp}(\Phi, 1) = \Phi$, and that $\mathrm{Sp}(\Phi, 0)$ is the zero matrix. We also have a transitive property: if $\Phi' \in \mathrm{Sp}(\Phi, s_1)$ and $\Phi'' \in \mathrm{Sp}(\Phi', s_2)$ then $\Phi'' \in \mathrm{Sp}(\Phi, s_1 s_2)$. Two independent sparsifications will not in general be comparable: there is a partial ordering on the set of sparsifications of a matrix, but not a total order.

We illustrate our notation. Consider a Bernoulli random variable which takes value 1 with probability p and value 0 with probability $1-p$ and let Φ be an $n \times N$ matrix with entries drawn from this distribution; in short, a *Bernoulli ensemble* with expected value p. Then the expected density of Φ is p. Writing J for the all-ones matrix, a randomly chosen $\Phi' \in \mathrm{Sp}(J, p)$ will also have density p, and can be considered a good approximation of a Bernoulli matrix. If Φ''

is an independent random sparsification (i.e., all non-zero entries of the matrix have an equal probability of being set to zero) of Φ' with relative density p', then Φ'' approximates the Bernoulli ensemble with expected value pp'. So we have both $\Phi' \in \mathrm{Sp}(\Phi', p')$ and $\Phi' \in \mathrm{Sp}(J, pp')$. Later, we will consider successive sparsifications where we begin with a dense matrix whose entries are drawn from, e.g., a normal distribution.

Bernoulli ensembles have previously been considered in the compressed sensing literature, see [31] for example, though note that the matrices here take values in $\{0, 1\}$, not $\{\pm 1\}$. Such $\{\pm 1\}$-matrices are an affine transformation of ours: $M' = 2M - J$; as a result, the compressed sensing performance of either matrix is essentially the same.

In this paper, we will mostly be interested in *pseudo-random* sparsifications of an $n \times N$ compressed sensing matrix Φ. Specifically, for $s = t/n$, we obtain a matrix $\Phi' \in \mathrm{Sp}(\Phi, s)$ by generating a pseudo-random $\{0, 1\}$-matrix S with sn randomly located ones per column, and returning the entry-wise product $\Phi' = \Phi * S$. We will generally re-normalize Φ' so that every column has unit ℓ_2-norm.

Given a matrix Φ, we test its CS performance by running simulations. Since many different methodologies occur in the literature, we specify ours here.

Our k-sparse vectors always contain exactly k non-zero entries, in positions chosen uniformly at random from the $\binom{N}{k}$ possible supports of this size. The entries, unless otherwise specified, are drawn from a uniform distribution on the open interval $(0, 1)$. The vector is then scaled to have unit ℓ_2-norm. Simulations where the non-zero entries were drawn from the absolute value of a Gaussian distribution produced similar results. Note that many authors use $(0, 1)$- or $(0, \pm 1)$-vectors for their simulations. Appropriate combinations of matrices and algorithms often exhibit dramatic improvements of performance on this restricted set of signals.

We recover signals using ℓ_1-minimization. In this paper we will use the `matlab` LP-solver and the implementations of *Orthogonal Matching Pursuit* (OMP) and *Compressive Sampling Matching Pursuit* (CoSaMP) algorithms which were developed by Needell and Tropp [28]. Specifically, given a matrix Φ and signal vector x, we compute $y = \Phi x$, and solve the ℓ_1-minimization problem $\Phi \hat{x} = y$ for $\hat{x}$. The objective function is the ℓ_1-norm of $\hat{x}$ and it is assumed

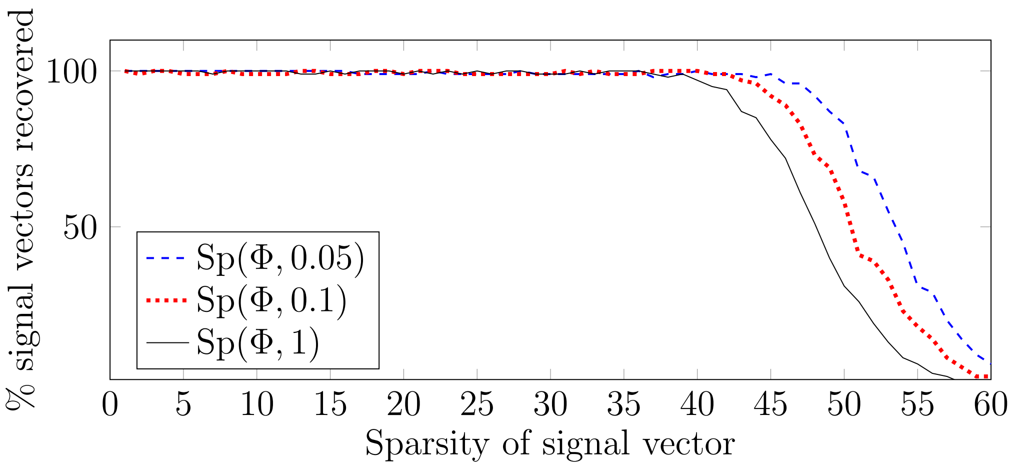


FIGURE 1. Effect of sparsification on signal recovery.

that all variables are non-negative. We consider the recovery successful if $|x - \hat{x}|_1 \leq c$ for some constant c. We take $c = 10^{-6}$ in all the simulations presented in this paper.

We conclude this section with an example illustrating the potential benefits of sparsification. In Figure 1, we explore the effect of sparsification on a 200×2000 matrix Φ with entries uniformly distributed on $(0, 1)$. The results for this case were compared with matrices drawn from $\mathrm{Sp}(\Phi, 0.1)$ and $\mathrm{Sp}(\Phi, 0.05)$. For each signal sparsity between 1 and 60, we generated 500 random vectors as described above and recorded the number of successful recoveries using the `matlab` LP-solver. To avoid bias we generated a new random matrix for each trial.

We observe that for signal vector sparsities between 45 and 55, matrices in $\mathrm{Sp}(\Phi, 0.05)$ achieve substantially better recovery than those from $\mathrm{Sp}(\Phi, 1)$. The code used to generate this simulation as well as others in this paper is available in full, along with data from multiple simulations at a webpage dedicated to this project: `http://fintanhegarty.com/compressed_sensing.html` .

4. Results

Our simulations produce large volumes of data. To highlight the interesting features of these data-sets, we propose the following measure for acceptable signal recovery in practice.

Definition 4.1. For a matrix Φ and for $0 \leq t \leq 1$, we define the *t-recovery threshold*, denoted R_t, to be the largest value of k for which Φ recovers k-sparse signal vectors with probability exceeding t.

We construct an estimate $\hat{R}_t$ for R_t by running simulations. As the number of simulations that we run increases, $\hat{R}_t$ converges to R_t. In practice this convergence is rapid. The definition of R_t generalizes naturally to a space of matrices (say $n \times N$ Gaussian ensembles): it is simply the expected value of R_t for a matrix chosen uniformly at random from the space. To estimate R_t with reasonably high confidence, we proceed as follows: beginning with signals of sparsity $k = 1$, we attempt 50 recoveries. We increment the value k by 1 and repeat until we reach the first sparsity k_0 where less than $50t$ signals are recovered. Beginning at $k_0 - 3$, we attempt 200 recoveries at each signal sparsity. When we reach a signal sparsity k_1 where less than $200t$ signals are recovered, we attempt 1000 signal recoveries at each signal sparsity starting at $k_1 - 3$. When we reach a signal sparsity k_2 where less than $1000t$ signals are recovered, we set $\hat{R}_t = k_2 - 1$.

We typically find that $k_1 = k_2$, which gives us confidence that $\hat{R}_t = R_t$. Unless otherwise specified, we use the assumptions outlined in Section 3.

4.1. **Recovery algorithms with sparsification.** As suggested already in Figure 1, taking $\Phi' \in \mathrm{Sp}(\Phi, s)$ for some value of $s \sim 0.05$ seems to offer considerable improvements when using linear programming for signal recovery. Similar results hold for OMP and CoSaMP, though note that in each case we supply these algorithms with the sparsity of the signal vector. (While there is an option to withhold this data, the recovery performance of CoSaMP seems to suffer substantially without it — and we wish to be able to perform comparisons with linear programming.) In Figure 2, we graph $R_{0.98}$ of $\mathrm{Sp}(\Phi, s)$ as a function of s, where Φ is a 200×2000 matrix with entries drawn from the absolute values of samples from a standard normal distribution.

For each algorithm, $R_{0.98}$ appears to obtain a maximum for matrices of density between 0.15 and 0.05. It is perhaps interesting to note that the percentage improvement obtained by CoSaMP is far greater than that for either of the other algorithms.

Table 1 shows the average time taken for one hundred vector recovery attempts using 200×2000 measurement matrices with entries drawn from the absolute values of samples from a normal distribution, over a range of vector sparsities. We note an improvement in running time of an order of magnitude for linear programming when using sparsified matrices, and an improvement when using CoSaMP.

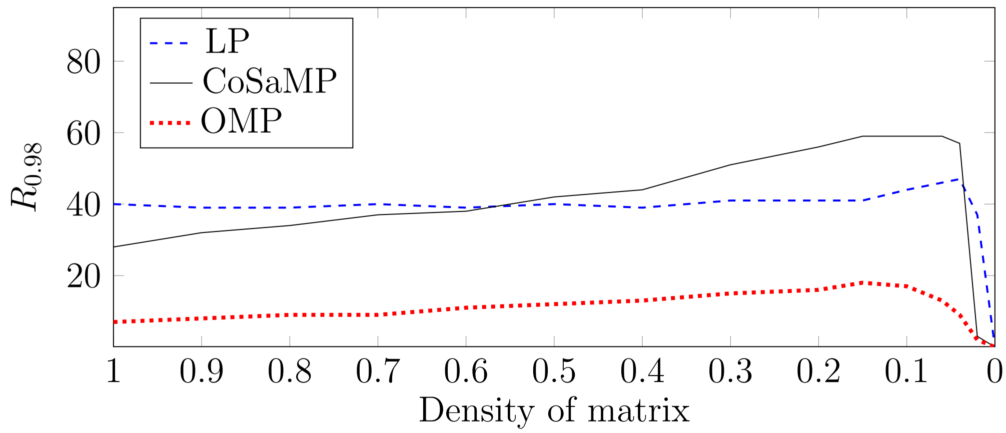


FIGURE 2. Signal recovery as a function of matrix density for LP, OMP and CoSaMP.

4.2. **Matrix constructions under sparsification.** In this section we explore the effect of sparsification on a number of different constructions proposed for CS matrices. We have already encountered the Gaussian, Uniform and Bernoulli ensembles. We will also consider some *structured random matrices*, which still have entries drawn from a probability distribution, but the matrix entries are no longer independent. The *partial circulant ensemble* [30] consists of rows sampled randomly from a circulant matrix, the first row of which contains entries drawn uniformly at random from some suitable probability distribution. Table 2 compares $R_{0.98}$ for $\mathrm{Sp}(\Phi, 1)$ and $\mathrm{Sp}(\Phi, 0.05)$ for 200×2000 matrices from each of the classes listed. Note that in the case of the Bernoulli ensemble, we actually compare $\mathrm{Sp}(J_{200,2000}, 0.5)$ with $\mathrm{Sp}(J_{200,2000}, 0.05)$, where $J_{200,2000}$ is an all-ones matrix. The entries of the partial circulant matrix were drawn from a normal distribution.

k	Time for 100 recovery attempts				% vectors successfully recovered			
	CoSaMP		LP		CoSaMP		LP	
	$\delta = 0.1$	$\delta = 1$	$\delta = 0.1$	$\delta = 1$	$\delta = 0.1$	$\delta = 1$	$\delta = 0.1$	$\delta = 1$
1	0.94	0.44	18.1	106.61	100	100	100	100
10	0.78	1.25	39.78	157.53	100	100	100	100
20	0.56	1.98	27.50	177.17	100	100	100	100
30	1.56	4.48	27.39	171.84	100	99	100	100
40	3.68	33.98	27.02	207.22	100	55	99	94
50	11.09	66.77	33.59	375.17	99	7	78	38
60	48.38	81.54	43.82	364.25	75	0	1	1
70	89.87	93.62	41.03	329.22	0	0	0	0

TABLE 1. Effect of sparsification on recovery time.

We denote by $\hat{k}$ the signal sparsity k for which the greatest difference in recovery between Φ and $\Phi' \in \mathrm{Sp}(\Phi, 0.05)$ occurs.

Construction	$R_{0.98}$		Maximal performance difference		
	$\delta = 1$	$\delta = 0.05$	$\hat{k}$	$\delta = 1$	$\delta = 0.05$
Normal	39	46	51	25	81
Uniform	39	45	51	24	73
Bernoulli	39	42	49	38	67
Partial Circulant	39	46	52	22	76

TABLE 2. Benefit of sparsification for different matrix constructions.

4.3. **Varying the matrix parameters.** Finally, we investigate the effect of sparsification on matrices of varying parameters. In particular, we explore the effect of sparsification on a family of matrices with entries drawn from the absolute value of the Gaussian distribution. First we explore the effect of sparsification as the ratio of columns to rows in the sensing matrix increases. For Figure 3, we use signal vectors whose entries were drawn from the absolute value of the normal distribution. We observe a modest improvement in performance which appears to persist.

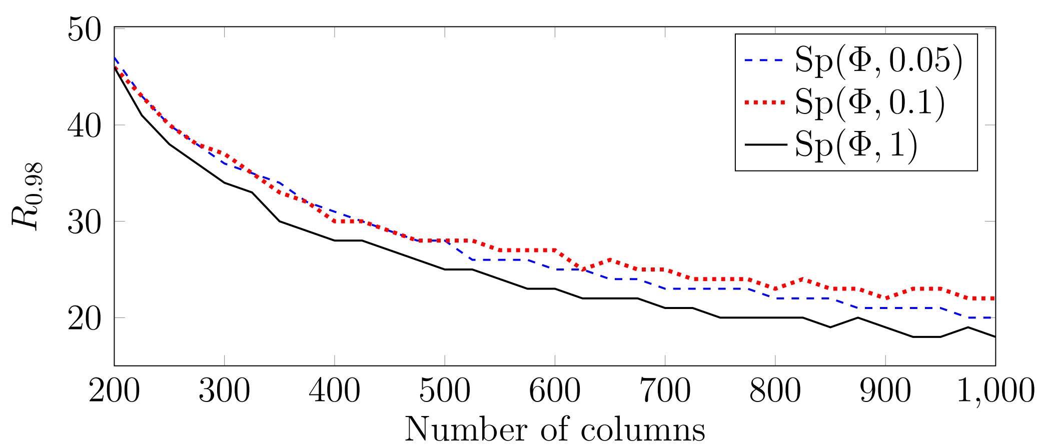


FIGURE 3. Recovery capability of matrices with 100 rows and varying number of columns under sparsification.

Now for Figure 4, we fix the ratio of columns to rows of Φ to be 10, and vary the number of rows. We know from the results of Candès et al. that $R_{0.98} = \Theta(n/\log n)$ in all cases. Nevertheless, the clear difference in slopes for recovery at different sparsities offers compelling evidence that the benefits of sparsification persist for large matrices.

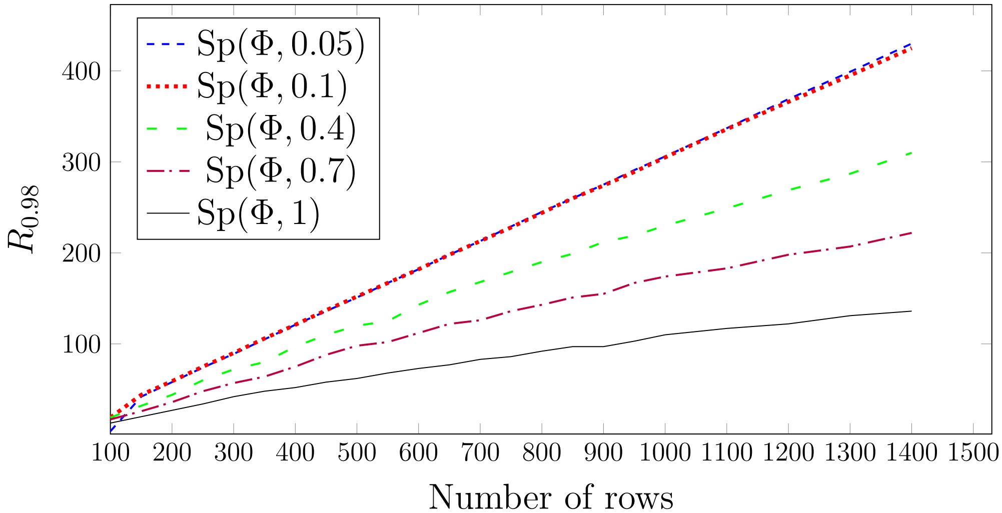


FIGURE 4. Recovery with CoSaMP for matrices with fixed row to column ratio under sparsification.

5. CONCLUSION

Some of the most important open problems in compressed sensing relate to the development of efficient matrix constructions and effective algorithms for sparse recovery. Deterministic constructions are essentially limited by the Welch bound: using known methods it is not possible to guarantee recovery of vectors of sparsity exceeding $\Theta(\sqrt{n})$, where n is the number of rows in the recovery matrix (see, e.g., [6]). Probabilistic constructions are much better: Candès and Tao's theory of restricted isometry parameters allows the provable recovery of vectors of sparsity k in dimension N with $\Theta(k \log(N))$ measurements. Such guarantees hold with overwhelming probability for Gaussian ensembles and many other classes of random matrices. But the random nature of these matrices can make the design of efficient recovery algorithms difficult. In this paper we have demonstrated that sparsification offers potential improvements for computational compressed sensing. In particular, Figures 1 and 4 show that sparsification results in the recovery of vectors of higher sparsity. Table 1 shows a substantial improvement in runtimes for linear programming arising from sparsification. These appear to be robust phenomena, which persist under a variety of recovery algorithms and matrix constructions. At the problem sizes that we explored, matrices with densities between 0.05 and 0.1 seemed to provide optimal performance.

We conclude with a small number of observations and conjectures which we believe to be suitable for further investigation. Since a Bernoulli ensemble in our terminology can be regarded as a sparsification of the all-ones matrix, it is clear that sparsification can improve CS performance. The necessary decay in CS performance as the density approaches zero shows that the effect of sparsification cannot be monotone. Extensive simulations suggest that when recovery is achieved with a general purpose linear programming solver, matrices with approximately 10% non-zero entries have substantially better CS properties than dense matrices. A catastrophic decay of compressed sensing performance occurs in many of the examples we investigated between densities of 0.05 and 0.01. We pose two questions which we think suitable for further research.

Question 5.1. *As the number of rows of Φ increases, the optimal matrix density appears to decrease. This effect does not appear to depend strongly on the matrix construction chosen. Does there exist a function $\Gamma(n, N, k)$ which describes the optimal level of sparsification for an $n \times N$ matrix recovering k-sparse vectors? Our simulations suggest that when $N < n^{\alpha}$, the optimal density of a CS matrix will be approximately $\alpha n^{-1/2}$ when $k = o(n^{1-\epsilon})$.*

Question 5.2. *We have considered pseudo-random sparsifications in this paper. In general, this should not be necessary. Are there deterministic constructions for $(0, 1)$-matrices with the property that their entry-wise product with a CS matrix improves CS performance? A natural class of candidates would be the incidence matrices of t-(v, k, λ) designs (see* [4] *for example). Some related work is contained in* [5, 6].

Acknowledgements

The authors acknowledge the comments of anonymous referees.

The first and third authors have been supported by the Engineering and Physical Sciences Research Council grant EP/K00946X/1. The second author acknowledges the support of the Australian Research Council via grant DP120103067, and Monash University where much of this work was completed. This research was partially supported by the Academy of Finland (grants #276031, #282938, and #283262).

References

[1] B. Bah and J. Tanner. Vanishingly sparse matrices and expander graphs, with application to compressed sensing. *IEEE Trans. Inform. Theory*, 59(11):7491–7508, Nov 2013.

[2] D. Baron, S. Sarvotham, and R. Baraniuk. Bayesian compressive sensing via belief propagation. *IEEE Trans. Signal Process.*, 58(1):269–280, Jan 2010.

[3] R. Berinde, A. Gilbert, P. Indyk, H. Karloff, and M. Strauss. Combining geometry and combinatorics: A unified approach to sparse signal recovery. In *Proc. 46th Ann. Allerton Conference on Communication, Control, and Computing*, pages 798–805, Sept 2008.

[4] T. Beth, D. Jungnickel, and H. Lenz. *Design theory. Vol. I*, volume 69 of *Encyclopedia Math. Appl.*. Cambridge University Press, second edition, 1999.

[5] D. Bryant, C. Colbourn, D. Horsley, and P. Ó Catháin. Compressed sensing and designs: theory and simulations. *IEEE Trans. Inform. Theory*, 63(8):4850-4859, 2017.

[6] D. Bryant and P. Ó Catháin. An asymptotic existence result on compressed sensing matrices. *Linear Algebra Appl.*, 475:134–150, 2015.

[7] E. Candès. Mathematics of sparsity (and a few other things). *International Congress of Mathematicians, Plenary Lecture 3*, Seoul, Aug 2014.

[8] E. Candès, J. Romberg, and T. Tao. Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information. *IEEE Trans. Inform. Theory*, 52(2):489–509, Feb 2006.

[9] E. J. Candès, J. Romberg, and T. Tao. Stable signal recovery from incomplete and inaccurate measurements. *Comm. Pure Appl. Math.*, 59(8):1207–1223, 2006.

[10] V. B. Chandar. Sparse graph codes for compression, sensing and secrecy. *PhD thesis*, MIT, 2010.

[11] R. A. DeVore. Deterministic constructions of compressed sensing matrices. *J. Complexity*, 23(4-6):918–925, 2007.

[12] A. Dimakis, R. Smarandache, and P. Vontobel. LDPC codes for compressed sensing. *IEEE Trans. Inform. Theory*, 58(5):3093–3114, May 2012.

[13] T. Do, L. Gan, N. Nguyen, and T. Tran. Fast and efficient compressive sensing using structurally random matrices. *IEEE Trans. Signal Process.*, 60(1):139–154, Jan 2012.

[14] D. Donoho. Compressed sensing. *IEEE Trans. Inform. Theory*, 52(4):1289–1306, April 2006.

[15] M. Fickus, D. G. Mixon, and J. C. Tremain. Steiner equiangular tight frames. *Linear Algebra Appl.*, 436(5):1014–1027, 2012.

[16] A. C. Gilbert, Y. Li, E. Porat, and M. J. Strauss. Approximate sparse recovery: Optimizing time and measurements. *SIAM J. Comput.*, 41(2):436–453, 2012.

[17] D. Guo, D. Baron, and S. Shamai. A single-letter characterization of optimal noisy compressed sensing. In *Proc. 47th Ann. Allerton Conference on Communication, Control, and Computing*, pages 52–59, 2009.

[18] S. Jafarpour, R. Willett, M. Raginsky, and R. Calderbank. Performance bounds for expander-based compressed sensing in the presence of Poisson

noise. In *Proc. 43rd Asilomar Conference on Signals, Systems and Computers*, pages 513–517, Nov 2009.

[19] M. Kern. Advanced technologies vastly improve MRI for children. *Science Highlights, National Institute for Biomedical Imaging and Bioengineering*, September 2014.

[20] S. Li and G. Ge. Deterministic construction of sparse sensing matrices via finite geometry. *IEEE Trans. Signal Process.*, 62(11):2850–2859, June 2014.

[21] W. Lu, K. Kpalma, and J. Ronsin. Semi-deterministic ternary matrix for compressed sensing. In *Proc. 22nd European Signal Processing Conference (EUSIPCO)*, pages 2230–2234, 2014.

[22] W. Lu, W. Li, K. Kpalma, and J. Ronsin. Near-optimal binary compressed sensing matrix. arXiv preprint abs/1304.4071, 2013.

[23] M. Lustig, D. Donoho, and J. Pauly. Sparse MRI: The application of compressed sensing for rapid MR imaging. *Magnetic Resonance in Medicine*, 58(6):1182–95, Dec 2007.

[24] M. Lustig, D. Donoho, J. Santos, and J. Pauly. Compressed sensing MRI. *IEEE Signal Processing Magazine*, 25(2):72–82, March 2008.

[25] A. Moghadam and H. Radha. Complex sparse projections for compressed sensing. In *Proc. 44th Ann. Conference on Information Sciences and Systems*, pages 1–6, March 2010.

[26] A. A. Moghadam and H. Radha. Sparse expander-like real-valued projection (SERP) matrices for compressed sensing. In *Proc. 2013 IEEE Global Conference on Signal and Information Processing*, pages 618–618, Dec 2013.

[27] B. K. Natarajan. Sparse approximate solutions to linear systems. *SIAM J. Comput.*, 24(2):227–234, 1995.

[28] D. Needell and J. A. Tropp. CoSaMP: iterative signal recovery from incomplete and inaccurate samples. *Appl. Comput. Harmon. Anal.*, 26(3):301–321, 2009.

[29] J. Nelson and H. L. Nguyen. Sparsity lower bounds for dimensionality reducing maps. In *Proc. 45th Ann. ACM Symposium on Theory of Computing*, pages 101–110, 2013.

[30] H. Rauhut, J. Romberg, and J. A. Tropp. Restricted isometries for partial random circulant matrices. *Appl. Comput. Harmon. Anal.*, 32(2):242–254, 2012.

[31] H. Rauhut, K. Schnass, and P. Vandergheynst. Compressed sensing and redundant dictionaries. *IEEE Trans. Inform. Theory*, 54(5):2210–2219, 2008.

[32] V. Ravanmehr, L. Danjean, B. Vasic, and D. Declercq. Interval-passing algorithm for non-negative measurement matrices: Performance and reconstruction analysis. *IEEE J. Emerging and Selected Topics in Circuits and Systems*, 2(3):424–432, Sept 2012.

[33] T. Tao. *Topics in random matrix theory*, volume 132 of *Grad. Stud. Math.*, American Mathematical Society, Providence, RI, 2012.

[34] A. Tehrani, A. Dimakis, and G. Caire. Optimal deterministic compressed sensing matrices. In *Proc.2013 IEEE Int. Conference on Acoustics, Speech and Signal Processing*, pages 5895–5899, May 2013.

[35] M. Uecker, S. Zhang, and J. Frahm. Nonlinear inverse reconstruction for real-time MRI of the human heart using undersampled radial FLASH. *Magnetic Resonance in Medicine*, 63(6):1456–62, 2010.

[36] W. Wang, M. Wainwright, and K. Ramchandran. Information-theoretic limits on sparse signal recovery: Dense versus sparse measurement matrices. *IEEE Trans. Inform. Theory*, 56(6):2967–2979, June 2010.
[37] S.-T. Xia, X.-J. Liu, Y. Jiang, and H.-T. Zheng. Deterministic constructions of binary measurement matrices from finite geometry. *IEEE Trans. Signal Process.*, 63(4):1017–1029, Feb 2015.
[38] Z. A. Zhu, R. Gelashvili, and I. Razenshteyn. Restricted isometry property for the general p-norms. *IEEE Trans. Inform. Theory*, 62:10, 5839-5854, Oct 2016.

Fintan Hegarty holds a PhD from the National University of Ireland, Galway, and spent 2013-15 as a postdoctoral researcher at the University of Birmingham. He currently works as a production editor for the Pacific Journal of Mathematics.
Padraig Ó Catháin received a PhD from the National University of Ireland, Galway in 2012. He was awarded the Kirkman medal of the Institute of Combinatorics and its Applications in 2015. He is currently an Assistant Professor at Worcester Polytechnic Institute. His research interests include combinatorial designs, Hadamard matrices and their applications.
Yunbin Zhao is a Senior Lecturer of Mathematical Optimization in the University of Birmingham. His main research interests include operations research, computational optimization, and numerical analysis.

(FH) School of Mathematics, Statistics and Applied Mathematics, National University of Ireland, Galway
fintan.hegarty@gmail.com

(PÓC) Department of Mathematical Sciences, Worcester Polytechnic Institute
p.ocathain@wpi.edu

(YZ) School of Mathematics, University of Birmingham
y.zhao.2@bham.ac.uk

Irish Math. Soc. Bulletin
Number 81, Summer 2018, 23–30
ISSN 0791-5578

Pedro Nunes and the Retrogression of the Sun

PETER LYNCH

What has been will be again, what has been done will be done again; there is nothing new under the sun. Ecclesiastes 1:9

Introduction

In northern latitudes we are used to the Sun rising in the East, following a smooth and even course through the southern sky and setting in the West. The idea that the compass bearing of the Sun might reverse seems fanciful. But that was precisely what Portuguese mathematician Pedro Nunes showed in 1537. Nunes made an amazing prediction: in certain circumstances, the shadow cast by the gnomon of a sun dial moves backwards.

Nunes' prediction was counter-intuitive. We are all familiar with the steady progress of the Sun across the sky and we expect the azimuthal angle or compass bearing to increase steadily. If the shadow on the sun dial moves backwards, the Sun must reverse direction or retrogress. Nunes' discovery came long before Newton or Galileo or Kepler, and Copernicus had not yet published his heliocentric theory. The retrogression had never been seen by anyone and it was a remarkable example of the power of mathematics to predict physical behaviour. Nunes himself had not seen the effect, nor had any of the tropical navigators or explorers whom he asked.

Nunes was aware of the link between solar regression and the biblical episode of the sun dial of Ahaz (Isaiah, 38:7–9). However, what he predicted was a natural phenomenon, requiring no miracle. It was several centuries before anyone claimed to have observed the reversal (Leitão, 2017). In a book published in Lisbon (Nunes, 1537), Nunes showed how, under certain circumstances, the azimuth of the Sun changes direction twice during the day, moving first forwards, then backwards and finally forwards again. To witness this, the observer

2010 *Mathematics Subject Classification.* 46A99, 10A10.
Key words and phrases. Truth, Heron.
Received on 0-0-0000; revised 0-0-0000.

must be located at a latitude lower than that of the Sun, that is, in the tropics with the Sun closer to the pole. Nunes was completely confident about his prediction:

> "This is something surprising but it cannot be denied because it is demonstrated with mathematical certainty and evidence." (Quoted from Leitão, 2017).

Leitão, who has made a detailed study of Nunes' works, reviewed the method used by him. While Nunes' arguments are mathematically sound, they are difficult to follow, so we will demonstrate the retrogression in a more transparent way below. But first, let us look at Pedro Nunes himself.

Pedro Nunes (1502–1578)

Pedro Nunes (also known as Petrus Nonius), a Portuguese cosmographer and one of the greatest mathematicians of his time, is best known for his contributions to navigation and to cartography. Nunes studied at the University of Salamanca in Spain, a university already 300 years old at that time. He returned to Lisbon and was later appointed Professor of Mathematics at the University of Coimbra. In 1533 he qualified as a doctor of medicine and in 1547 he was appointed Chief Royal Cosmographer.

Nunes had great skill in solving problems in spherical trigonometry. This enabled him to introduce improvements to the Ptolemaic system of astronomy, which was still current at that time (Copernicus did not publish his theory until just before his death in 1543). Nunes also worked on problems in mechanics.

Much of Nunes' research was in the area of navigation, a subject of great importance in Portugal during that period: sea trade was the main source of Portuguese wealth. Nunes understood how a ship sailing on a fixed compass bearing would not follow a great circle route but a spiraling course called a loxodrome or rhumb line that winds in decreasing loops towards the pole. Nunes taught navigation skills to some of the great Portuguese explorers. He has a place of prominence on the Monument to the Portuguese Discoveries in Lisbon, which shows several famous navigators (Figs. 1 and 2).

Analysis of Solar Retrogression

Nunes demonstrated the retrogression using spherical trigonometry. This was long before Newton's laws or differential calculus

Figure 1. Monument to the Portuguese Discoveries, Lisbon.

Figure 2. Pedro Nunes (1502–1578) (detail of Monument to Portuguese Discoveries, Lisbon).

were available. In this section we derive the condition for the phenomenon, using a simple transformation and elementary differential calculus. An expression is found for the azimuth of the Sun as a function of the time. For reversal to occur, the derivative of this function must vanish. The condition follows immediately from this. The result has been known for centuries (e.g., Morrison, 1898) but the derivation below is simpler than most previous accounts.

Frames of Reference. We begin with a cartesian frame (x, y, z) fixed relative to Earth and rotating with it. The origin is at the centre of the Earth and the x-axis passes through the point where the prime meridian intersects the equator. There is an associated polar coordinate frame (r, θ, λ) with colatitude θ and longitude λ. The latitude is $\phi = \frac{\pi}{2} - \theta$.

We assume that the Sun is at a fixed latitude ϕ_S. If its longitude at Noon is λ_O, then its longitude at time t is $\lambda_S = \lambda_O - \Omega(t - t_O)$ where Ω is the angular velocity of Earth. Given the distance A from Earth to Sun, the cartesian coordinates of the Sun are

$$(x_S, y_S, z_S) = (A \cos\lambda_S \cos\phi_S, A \sin\lambda_S \cos\phi_S, A \sin\phi_S)\,. \qquad (1)$$

The coordinates of the observation point P_O are (x_O, y_O, z_O) and from these the polar coordinates are easily found: $(a,\, \theta_O, \lambda_O)$ where a is the Earth's radius. There is no loss of generality in assuming that the observatory is on the prime meridian. Then its latitude and longitude are $(\phi_O, \lambda_O) = (\frac{\pi}{2} - \theta_O, 0)$.

We define local cartesian coordinates (X, Y, Z) at the observation point by rotating the (x, y, z) frame about the y-axis through an angle equal to the colatitude θ_O. The Z-axis then points vertically upward through P_O. Moving the origin to P_O, the (X, Y) plane becomes tangent to the Earth at this point. The cartesian coordinates of the Sun in the new system are given by the affine transformation

$$\begin{pmatrix} X_S \\ Y_S \\ Z_S \end{pmatrix} = \begin{bmatrix} \cos\theta_O & 0 & -\sin\theta_O \\ 0 & 1 & 0 \\ \sin\theta_O & 0 & \cos\theta_O \end{bmatrix} \begin{pmatrix} x_S \\ y_S \\ z_S \end{pmatrix} - \begin{pmatrix} 0 \\ 0 \\ a \end{pmatrix} .$$

In fact, since the distance to the Sun is enormous relative to the radius of the Earth, we can omit the last term $(0, 0, a)^{\mathrm{T}}$ without significant error. Then, in terms of the latitude of the observation point, the solar coordinates are

$$\begin{pmatrix} X_S \\ Y_S \\ Z_S \end{pmatrix} = \begin{bmatrix} \sin\phi_O & 0 & -\cos\phi_O \\ 0 & 1 & 0 \\ \cos\phi_O & 0 & \sin\phi_O \end{bmatrix} \begin{pmatrix} x_S \\ y_S \\ z_S \end{pmatrix} . \qquad (2)$$

The latitude and longitude of the Sun in the rotated system are

$$\Phi_S = \arcsin[Z_S/A]\,, \qquad \Lambda_S = \arctan[Y_S/X_S] \qquad (3)$$

and the azimuth and elevation (or altitude) of the Sun are

$$\alpha = \pi - \Lambda_S\,, \qquad e = \Phi_S\,. \qquad (4)$$

Variation of the Azimuthal Angle. To demonstrate the circumstances in which retrogression of the Sun occurs, we take the time derivative of the azimuthal angle of the Sun. This is given by (4). We use (3) to express the Sun's latitude and longitude in the rotated cartesian frame and then the transformation (2) for the original cartesian coordinates. Finally, (1) gives an expression for the azimuth in terms of the variables $\{\lambda_S, \phi_S, \phi_O\}$. The two latitudes are fixed in time while the longitude λ_S varies as $\lambda_S = -\Omega(t - t_O)$, where t_O is the time at Noon.

If the Sun is to retrogress, the time derivative of the azimuth must vanish. We find that

$$\tan \Lambda_S = \frac{Y_S}{X_S} = \frac{\sin \lambda_S \cos \phi_S}{\cos \lambda_S \cos \phi_S \sin \phi_O - \sin \phi_S \cos \phi_O}$$

The vanishing of the derivative leads, after some manipulation, to the equation

$$\cos \lambda_S = \frac{\tan \phi_O}{\tan \phi_S} \tag{5}$$

Clearly, there will be an azimuth at which the derivative vanishes only if the right hand side is less than unity, that is, if

$$\phi_O < \phi_S \,.$$

This means that retrogression will be seen only if the observation point is between the Equator and the latitude of the Sun. In particular, it must be in the tropics. Eq. (5) corresponds to Eq. (5) in Morrison (1898).

Numerical Results. We consider the daily path of the Sun at the time of the Summer solstice ($\phi_S = 23.5°$N) for observations from an extra-tropical point ($\phi_O = 40°$N) and a point within the tropics ($\phi_O = 20°$N). The elevation and azimuth are easily computed from the formulae above. We plot the zenith angle (the complement of the elevation) and azimuth for the extra-tropical observation in Fig. 3. The observation point is at the centre, and the course of the Sun is shown by the curve. It is clear that the azimuth increases monotonically from sunrise to sunset, as we would expect.

In Fig. 4 we plot the zenith angle and azimuth for the observation point within the tropics ($\phi_O = 20°$N). At Noon, the Sun is to the North of the central point and the azimuthal angle is *decreasing* rapidly. This is the retrogression phenomenon.

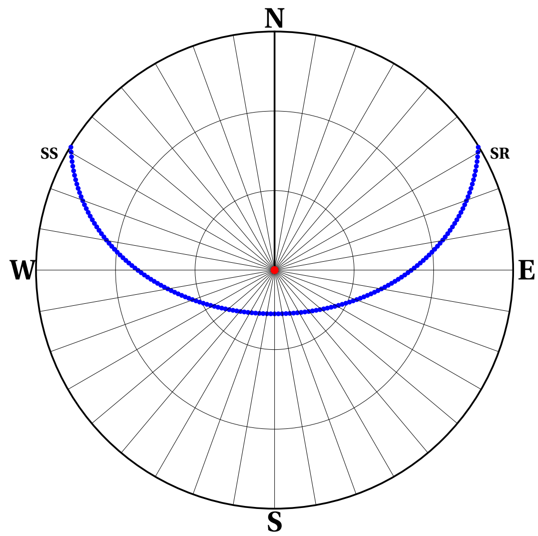


FIGURE 3. Path of the Sun at the Summer solstice for an observation point at 40°N. The angular coordinate is the azimuth or compass bearing, α. The radial coordinate is the zenith angle (the complement of the elevation). SR: Sunrise; SS: Sunset.

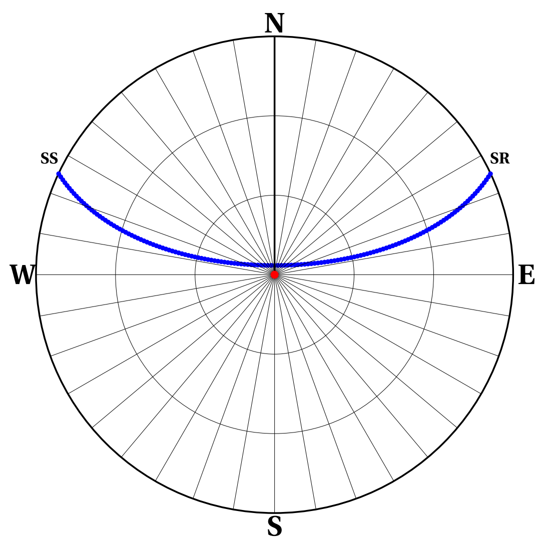


FIGURE 4. Path of the Sun at the Summer solstice for an observation point at 20°N. The angular coordinate is the azimuth or compass bearing, α. The radial coordinate is the zenith angle (the complement of the elevation). SR: Sunrise; SS: Sunset.

The azimuth and elevation of the Sun are plotted in Fig. 5 for the two observation points. It is clear that when $\phi_O = 40°$ (top right panel) the azimuth increases monotonically, while when $\phi_O = 20°$ (bottom right panel) the azimuth increases from sunrise until about 10:00, then decreases until 14:00 and finally increases until sunset.

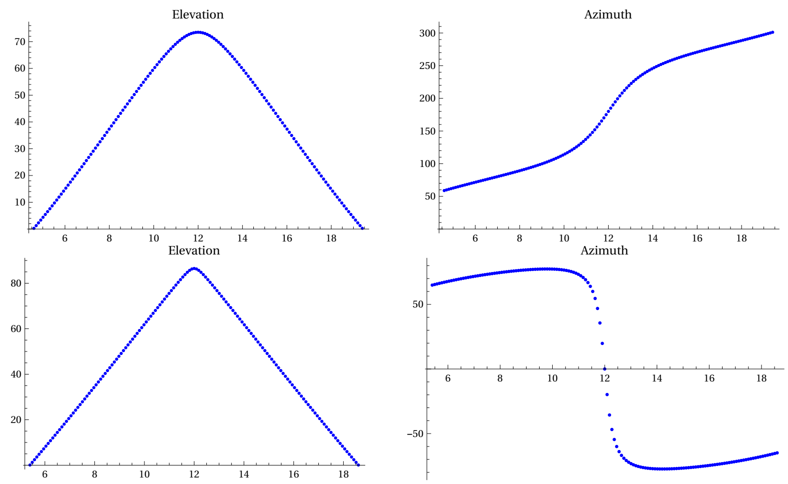


FIGURE 5. Solar elevation and azimuth, as functions of time from sunrise to sunset, for observation points at 40°N (top row) and 20°N (bottom row). Left: Solar elevation. Right: Solar azimuth.

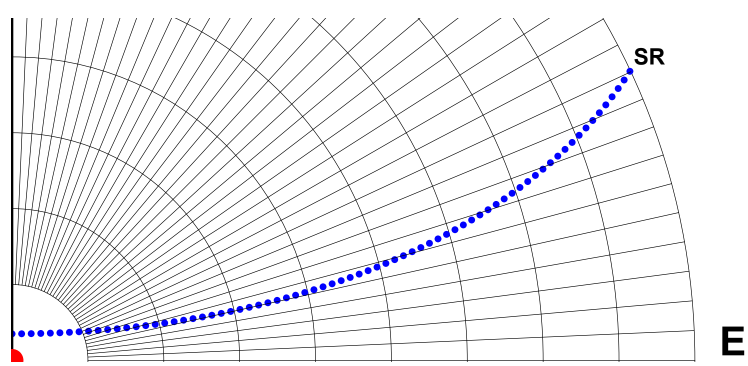


FIGURE 6. Path of the Sun at the Summer solstice for a observation point at 20°N. Angular coordinate is the azimuth or compass bearing. Radial coordinate is the zenith angle. SR: Sunrise. Only the morning segment of the Sun's track is shown.

To give more fine detail, we plot the Sun's course during the morning, as seen from the tropical observatory, in Fig. 6. The radial lines are spaced five degrees apart, and we see that the azimuth at sunrise is close to 65°. It increases to around 77° by mid-morning and then decreases to zero at Noon. For the specific values $\phi_O = 20°$ and $\phi_S = 23.5°$, the condition (5) gives the turning longitude as $\lambda_S = 33.17°$ corresponding to an azimuth of 77.4° and an elevation of 59.1°. This is in excellent agreement with the numerical solution shown in Fig. 6.

References

[1] Leitão, Henrique, 2017: A brief note on the power of mathematics: Pedro Nunes and the retrogradation of shadows. Proceedings of Recreational Mathematics Colloquium V – G4G (Europe), Ed. Jorge Nuno Silva, pp. 45–52.

[2] Morrison, J, 1898: The sun dial of Ahaz. *Popular Astronomy*, **6** (10), 537–549.

[3] Nunes, Pedro, 1537: Treatise in defence of the nautical chart. For a modern edition see *Obras de Pedro Nunes* (Pedro Nunes's Complete Works), Lisbon Acad. Sci., Vol. I, 156–157.

Peter Lynch is Emeritus Professor in the School of Mathematics and Statistics, UCD. His interests include dynamic meteorology, numerical weather prediction and recreational mathematics. He writes an occasional mathematics column *That's Maths* in the *Irish Times*; see his blog at `http://thatsmaths.com`.

School of Mathematics and Statistics, University College Dublin
E-mail address: `Peter.Lynch@udc.ie`

Irish Math. Soc. Bulletin
Number 81, Summer 2018, 31–55
ISSN 0791-5578

Leaving Cert Points: Do the Math?

DAVID MALONE, GAVIN MCCULLAGH

Abstract. In this paper we look at the Leaving Cert results from 2002–2016. We consider them in the light that many students will be choosing subjects in a way that will maximise their points. We observe that in a market for subjects, similar to the job market, a "pragmatic" student who needs points will choose subjects that offer them the highest points. As English and maths are examined across, more-or-less, the full Leaving Cert student cohort, they are uniquely comparable. Irish, which is studied by many students, also provides a useful reference point. We compare the results of these subjects over the period. In line with other authors, we identify a number of quirks of the grading system. We conclude that the statistics indicate that the mean/mode/median pragmatic student would choose to invest their effort in English over Irish and Irish over mathematics. We also highlight that normalisation of grades can cause feedback, resulting in easy subjects becoming easier and harder subjects becoming harder. We conclude that if the uptake of higher-level mathematics is to be increased, then the subject should be made more attractive in the subject marketplace. We discuss some possible ways that this could be achieved, including the awarding of bonus points.

1. Introduction

The Leaving Certificate is the final exam of the Irish second-level education system. Places in third level education are generally allocated to students based on a *points system*, where students are awarded points based on the level at which they take a subject and the grade they receive. In this system, apart from some rare exceptions, points are awarded equally for all subjects and a student's final points are the sum of the points from their best six subjects. These points are important to students. As typically observed [10],

2010 *Mathematics Subject Classification.* ?????

Key words and phrases. Leaving Certificate, points system, exam statistics, results comparison, subject choice market.

Received on 31-5-2017.

students often respond to the question "How did you do in the Leaving Cert?" by saying how many points they got, and so make efforts to maximise their points.

Students facing into the Leaving Cert have a number of choices to make. They have options regarding the subjects that they take, and options regarding the level at which they take these subjects. It seems likely that most students will aim to maximise their points, subject to the options available (e.g., subject choices and levels available at the school), the resources available (e.g., time, effort, available educational support) and information regarding how their efforts will be rewarded. This has been observed in interviews with students [20]. It is in the best interests of parents, students and teachers to be well informed about the 'points race', so while there will be rumour and hearsay, it seems likely that their information will be based on fact. As the majority of students only take the Leaving Cert examinations once, an important input to their decision will be the performance of previous cohorts of students.

Thus, we find ourselves in a situation where there may be a competitive market in Leaving Cert subjects. Students have a certain amount of effort to spend in order to get as many points as possible. All things being equal, we expect that subjects that give a greater reward for fixed effort will be chosen over more challenging subjects. Of course, directly comparing most subjects is difficult, for example, the availability of French and Ancient Greek teachers in schools will typically be quite different. Previous work [13, 9, 20] has shown that factors including choices in the junior cycle, school size and educational advantage can have an impact on course availability and students' achievement.

However, three subjects are available to all students: English, Irish and mathematics. Indeed, most students have no choice but to take all three of these subjects. Figure 1 shows how many students were taking these subjects for the years 2002–2016, based on the statistics from the `www.examinations.ie` website. We can see that the numbers taking English and maths are similar, while Irish lags by about 10% (presumably due to some students being exempt from taking Irish).

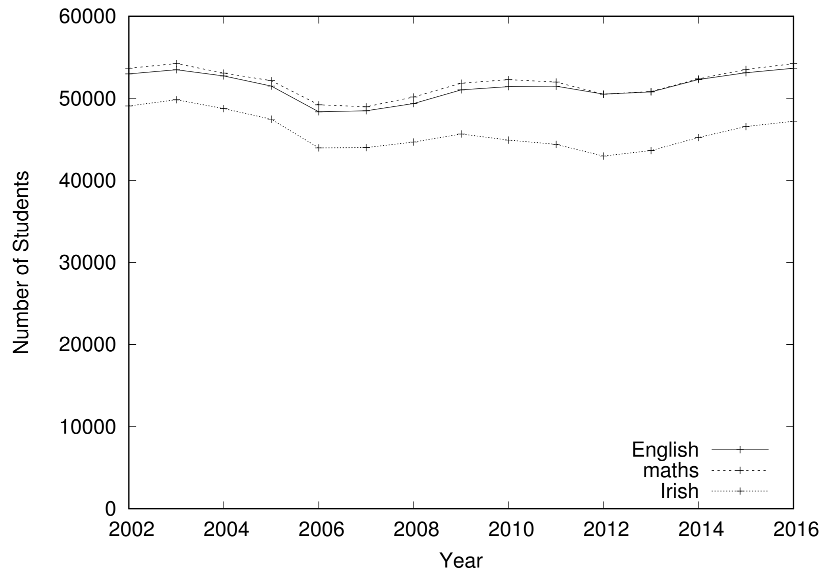


FIGURE 1. The number of students taking English, Irish and maths from 2002–2016.

These students represent the vast majority of the Leaving Cert cohort[1]. The students taking English and maths will essentially be the same people. This group will, more-or-less, overlap with all the Irish students. It seems that, by comparing performance in these subjects, we should be able to draw some conclusions about the relative rewards offered by these subjects.

The `www.examinations.ie` website publishes an annual report containing a summary of the Leaving Cert results. For each subject, it gives the number of students taking the subject at each level and the percentage breakdown of how many students achieve each grade at each level for the three years prior to the report. Historical reports are available, and results presented here are based on the results from 2002–2016. The points awarded for each grade and level are shown in Table 1. Note that from 2012 onwards, 25 "bonus points" were offered to students passing higher-level mathematics. From 2017 onwards there were some changes to the Leaving Cert grading system and points system. While the principles remain similar, the details have changed, and so we do not consider data from 2017 here.

[1]Note, a significant fraction of students, between 10–20%, do not reach the Leaving Cert [1].

Grade	A1	A2	B1	B2	B3	C1	C2	C3	D1	D2	D3	E	F	NG
Higher	100	90	85	80	75	70	65	60	55	50	45	0	0	0
Ordinary	60	50	45	40	35	30	25	20	15	10	5	0	0	0
Foundation	20	15	10	5	0	0	0	0	0	0	0	0	0	0

TABLE 1. Points offered for each grade at each level. From 2012 onwards a bonus of 25 points was given for higher-level mathematics at D grades and above. Not all institutions offer points for subjects at foundation level.

We are not the first to analyse and compare results in the Leaving Certificate. The issue of variable grading practices between subjects and levels has been previously observed [12]. In addition to quirks of the marking system, which we also identify here, this work also observes that the "difficulty" of subjects is not simply related to which grades are given out, but how easy these grades are to achieve. Earlier trends in the uptake and points awarded for the twenty most commonly taken Leaving Certificate subjects have also been studied [14]. The same work also identifies substantial variation between subjects, and particularly the significant impact of the proportion of students who choose to take subjects at ordinary and higher level.

2. A look at the data

Let us consider a number of different ways to compare the performance of students in English, Irish and mathematics. Using the points system, we will first consider the average performance, which indicates there is a significant difference in the points awarded. We will show that this difference arises because of the level at which the students have taken the different subjects, and suggest that one explanation for this is pragmatic students voting with their feet.

However, average performance may not tell the full story [12]. We will next look at the the cumulative performance, showing the number (or fraction) of students achieving at least some number of points. This shows that the quantiles of English, Irish and maths are strictly ordered[2], with English awarding higher points to more candidates.

Finally, we consider how many students achieve a particular number of points. This lets us study the mode number of points awarded,

[2] This can be termed stochastic dominance.

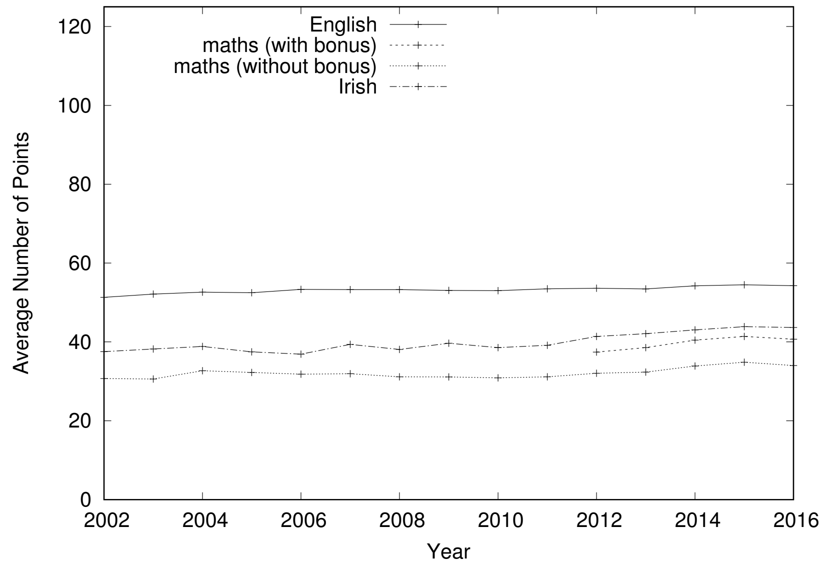


FIGURE 2. The average number of points given to students taking English, Irish and maths from 2002–2016.

and makes clearer the impact of fail grades, which may again impact the choices of pragmatic students. It also reveals a number of quirks of the Leaving Certificate marking system.

2.1. **The Average Student.** Figure 2 shows the average number of points given to students of English, Irish and maths (i.e., the total number of points given divided by the total number of students taking the subject at any level), from 2002–2016. We see that English gives about 53 points per student, Irish gives about 40 points per student and maths about 32 points per student (or 38 in 2012–2016 when the bonus is factored in). Clearly there is something different about these three subjects.

While, technically, students are not required to take English, Irish and maths, for a substantial majority of students these subjects are effectively compulsory. Consequently, the only choice available to most students regarding English, Irish and maths is the level at which they take these subjects. On the whole, students will be making choices about the level they take these subjects based on the effort to reward tradeoff. If we look at the average number of points at each level in these subjects (see Figure 3), we see that the higher-level subjects all give about 70 points per student and the ordinary-level subjects all give about 25 points per student. Foundation level

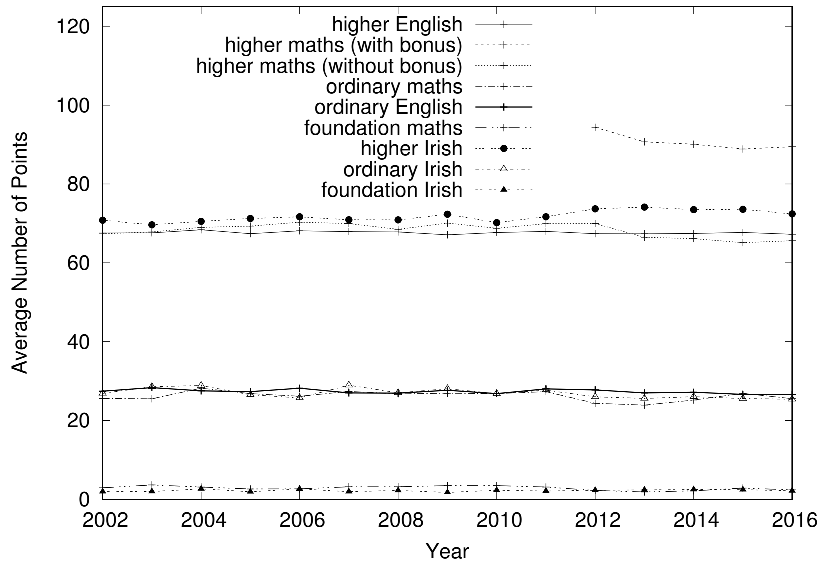


FIGURE 3. The average number of points given to students taking English, Irish and maths at a particular level from 2002–2016.

Irish and maths give about 2 points per student (English is not offered at foundation level[3]). The results at each level are quite consistent across subjects.

So, where does the difference in average reward come from? It arises from students voting with their feet. Figure 4 shows a consistent picture from year-to-year, where more than half of Leaving Cert candidates take English at higher level. From 2002–2013 around one quarter take higher-level Irish and one fifth take higher-level maths, but there has been an increase in the fraction taking higher Irish and maths since 2012, corresponding to changes in the courses and points. The smaller numbers of students at higher level for Irish and maths then result in lower overall average points. The importance of the level at which students take a subject to the average performance has also been highlighted before [14].

Why are students shying away from Irish and maths at higher level? While both Irish and maths might be considered to have an

[3]We believe it is important to include foundation level in this study, because to omit these students would make the maths and Irish cohort an artificially smaller group of stronger students. We also note that some institutions do not offer points for foundation level Irish/Maths. As we will see, considering foundation level to offer zero points would not make a significant difference to our results.

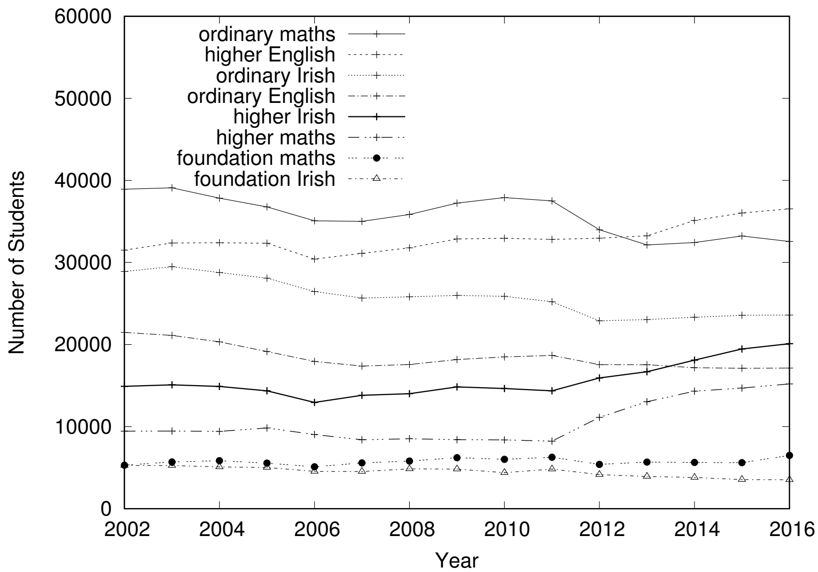


FIGURE 4. The number of students taking English, Irish and maths at a particular level from 2002–2016.

image problem, the competition for points means that students are unlikely to abandon these subjects because of image alone. Certainly, there are some differences in the availability of higher-level subjects in various schools[4], but it seems unlikely that image or availability can explain such a large difference in numbers.

One possible explanation is that students have determined that the effort required to get a high maths grade will be better rewarded by if it is spent on other subjects. This could be English (or Irish), but might also be other optional subjects. There is anecdotal evidence of students dropping from higher maths to ordinary maths and taking up an optional subject at higher level to boost their points. This behaviour has been reported in a survey of UCD Economics and Finance Students [16] and is probably more widespread [6, 15].

2.2. **On the Curve.** So far, we have looked at average points, but other statistics tell a similar story. As students aim for *at least* a particular number of points, rather than an exact target, the cumulative statistics are particularly useful. Figure 5 shows the cumulative statistics for 2003, i.e. the number of students achieving at least some

[4]Information on how many schools provide each subject is available in Section 4 of the Department of Education statistics [4] and also reported annually at `http://www.education.ie/en/Publications/Statistics/Statistical-Reports/`.

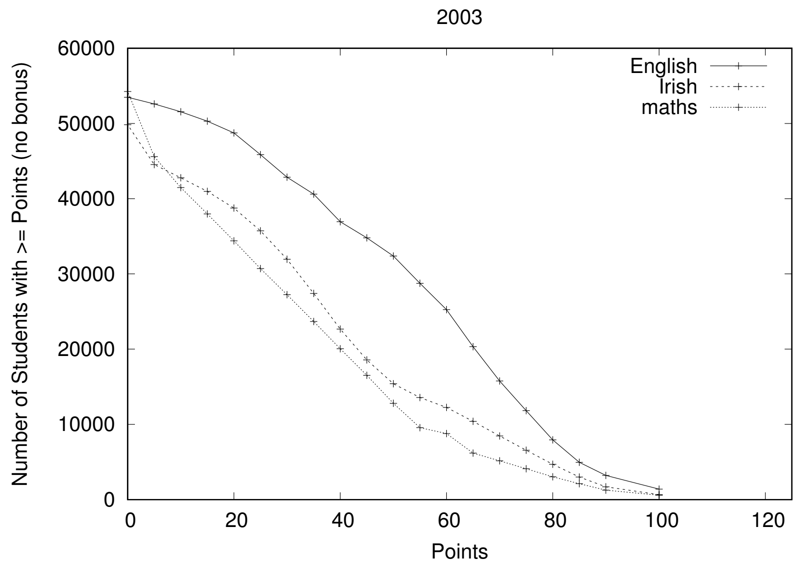


FIGURE 5. Number of points versus how many students have at least this many points by subject for 2003.

number of points at any level. For example, if we draw a vertical line at 80 points, we see that there are 7900 students on more than 80 points in English, but just 4700 in Irish and 3000 in maths.

Alternatively, if we draw a horizontal line at 10000 students, we see that the top 10000 students in English are on 75 points or more, in Irish the top 10000 students are on 65 points or more and in maths they are on 50 points or more. The mathematics curve is always under the other curves until 5 points, where mathematics rises above English/Irish, mainly because it has a larger overall number of students than other subjects.

These trends are broadly repeated from year-to-year, as can be seen from Figure 6, which shows the equivalent graph for each year between 2002–2016. From 2012 onwards, we show two curves for mathematics, corresponding to the points value with and without the bonus.

For all years, the number of students attaining at least some number of points in Irish is consistently below the number in English. Without the bonus, maths is generally lower again. The only exception is at the very lowest grades where the maths curve crosses above the others. This is simply due to the number of students sitting maths; there are more students overall studying maths than Irish or English and they all must achieve zero or more points. If the bonus points are included for mathematics, we find that the maths curve crosses the Irish curve at about 70 points and the English curve at about 80 points.

If we plot these curves in terms of the fraction of students taking the subject, rather than the total number (see Figure 7 and Figure 6 respectively), then the curves without the bonus do not cross. These graphs tell us that the median points, or indeed, any percentile for points for English is greater than that for Irish, which is greater than that for maths. For 2012-2016, where the bonus for maths applies, we see that the top 20–25% of students would get more points for maths than Irish/English.

2.3. **Making the Grade.** While harder to interpret in terms of a student's aims, we can also study the number of students with a particular number of points, as shown in Figure 9. Of course, a particular number of points may correspond to different grades at different subject levels, and we will look at these subcategories shortly.

One motivation for considering the number of students with a particular number of points is to find the mode, i.e. the most-commonly-awarded number of points. We identify the number of points where the maximum is achieved to find the mode of the points distribution. For Irish and mathematics, this is 0 points, while for English this is 60 or 65 points. Again, Figure 10 shows the curves for 2002–2016, and the trend is similar from year to year. Note that we do not show the bonus points for mathematics in these figures for two reasons. First, the students being spread over different numbers of categories make the graphs hard to compare in a meaningful way. Second, it will make no difference to the mode, as students who get 0 points do not get a bonus.

While there are rewards associated with achieving particular grades at higher level, there are also risks associated with taking an exam at higher level. Failing English, Irish or maths will preclude a student from many third level courses, careers and apprenticeships. By plotting the number of students achieving particular grades, we can get some idea of what the risks are.

To plot a curve for higher-level subjects, we simply mimic the CAO points system, taking the top percentage within that grade, i.e. A1=100, A2=90, … D3=45. Although no CAO points are awarded below the D3 grade, in order to see the tail of the curve,

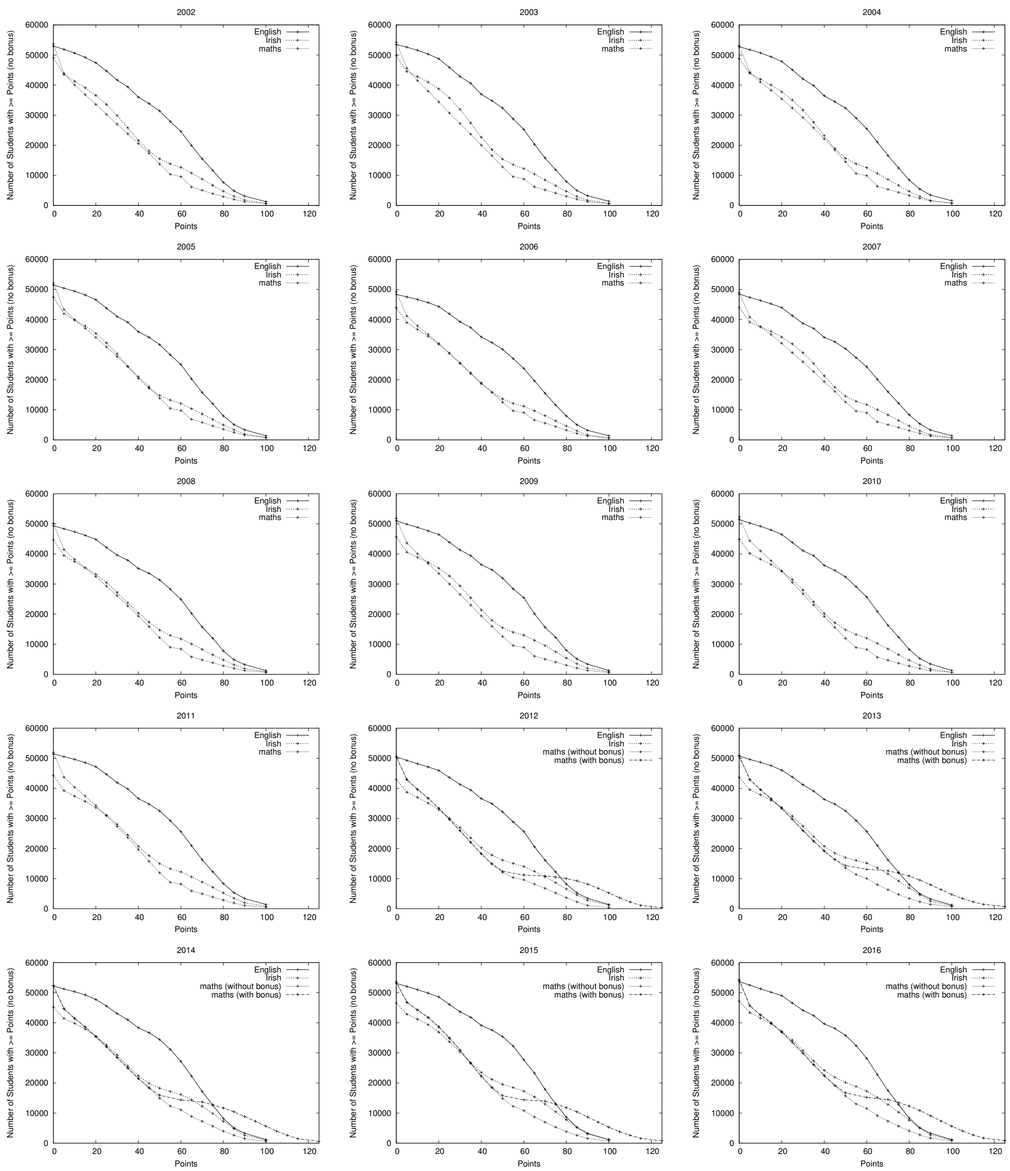


FIGURE 6. Number of points versus how many students have at least this many points by subject for 2002–2016.

we continue on assigning the lower grades their upper percentage, E=40, F=25, NG=10.

The CAO system awards 60 points for an ordinary-level A1, effectively equating that grade with a C3 at higher level. This offset of 40 points is applied to every other grade, resulting in A2=50, … D3=5. In order to compare the curves of the different levels we do the same here. Again, we want to see the tail, so although points are not awarded, we continue on and assign E=0, F=-15, NG=-30.

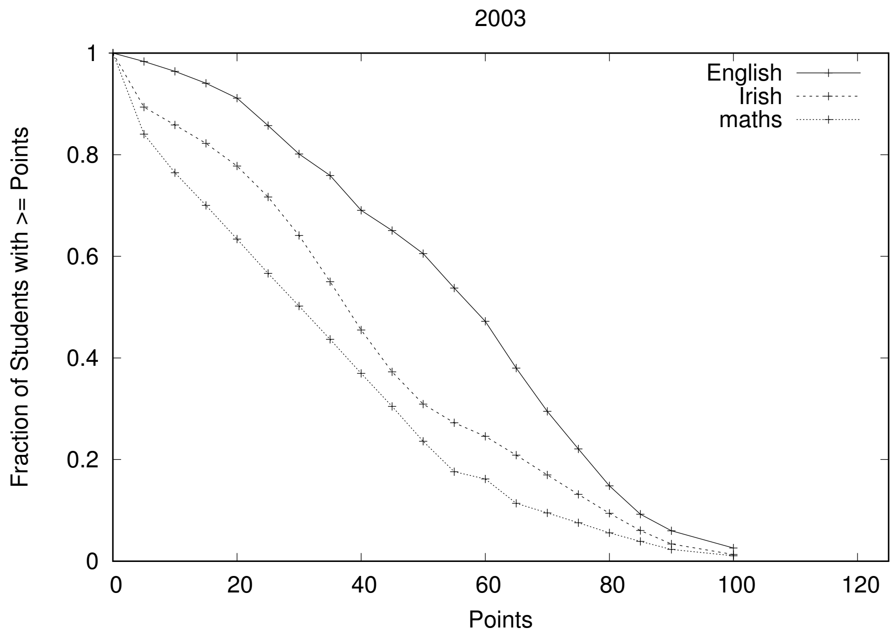


FIGURE 7. Number of points versus the fraction of students that have at least this many points by subject for 2003.

Foundation is treated similarly, with an offset of 75, as per the CAO points scheme. This formula agrees with the number of points given for all non-zero-points grades except for an A1 at foundation level[5]. Put a little more formally, we give a value for each grade at each level:

$$\text{value}(\text{grade}, \text{level}) = \text{cutoff}(\text{grade}) - \text{correction}(\text{level})$$

where the cutoff for a grade is the upper percentage that you can get for the grade and the correction is 0 for higher, 40 for ordinary and 75 for foundation.

Figure 11 shows the number of students achieving a particular value for English and maths (Irish has been omitted for clarity). Moving from the right-to-left along any curve, the first point is the number of A1s, the next A2s, then B1, B2, B3, C1, C2, C3, D1, D2, D3, E, F and NG. The E, F and NG grades are fail grades.

Looking at the right-hand side of this graph (40–100 points), we can see that the curve for higher English is far above the curve for

[5]An A1 at foundation level is actually worth only 20 points, but we give it a value of 25, so the foundation level curves are spaced in the same way as the other levels.

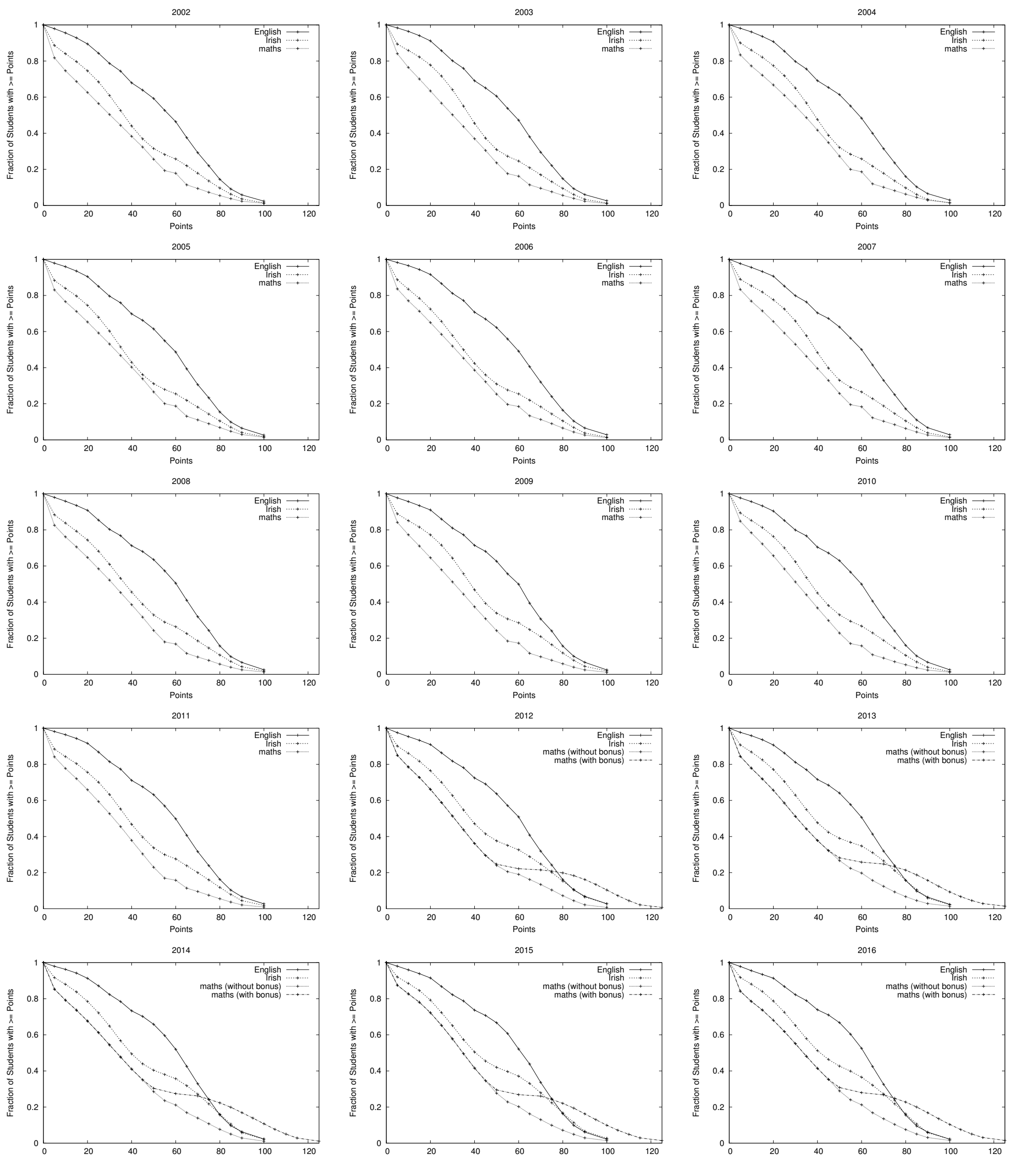


FIGURE 8. Number of points versus the fraction of students that have at least this many points by subject for 2002–2016.

higher maths, indicating the large difference in the numbers achieving particular higher-level grades in these two subjects. We have seen this in the previous figures. However, the part of the curves for higher-level English and maths between 10–40 almost over-lie one another in most years. Indeed, in some years the higher maths curve even rises above that for higher English. This means that

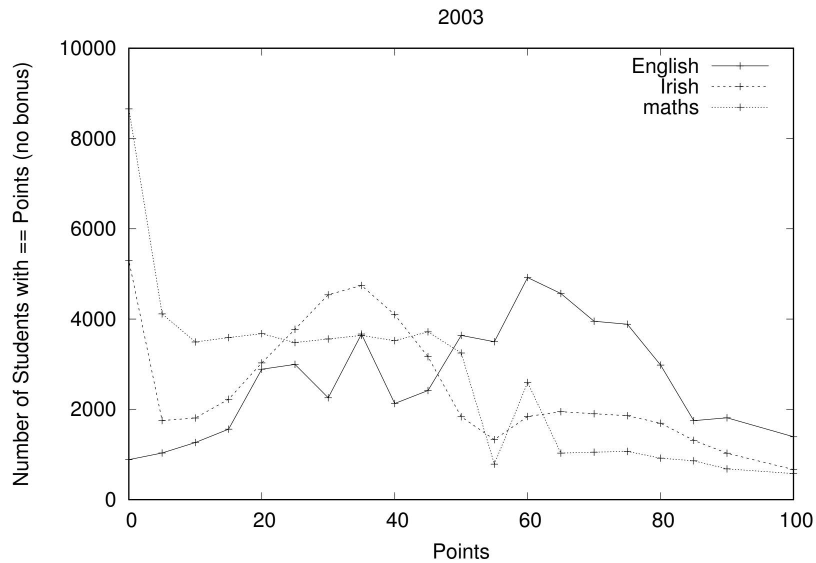


FIGURE 9. The number of students achieving a particular number of points in English, Irish and maths in 2003 (excluding bonus).

number of people failing higher maths is comparable to the number failing higher English, even though about three times as many people take higher English. One way to view this is that the risk of failing higher maths is about three times that of failing higher English. As expected, this is not peculiar to 2003, and we can see broadly similar trends repeated in Figure 12 for years 2002–2016.

The curves for ordinary English and maths are more in line with what one might expect given the greater number taking ordinary maths: the number of students achieving a particular grade in ordinary maths is usually well above the number achieving that grade in ordinary English.

There are some interesting peaks in the curves, which persist from year-to-year. Consider the English curves: there seems to be a dip where the point for A2 seems to be high relative to the point for B1; similarly, the point for B3 seems to be high relative to the point for C1. This can be observed in both the higher and ordinary curves. A likely explanation for this is that either the marking process or the appeals process is 'marking up' students across the boundary between A and B grades, and C and B grades. If this explanation is correct, it would provide evidence that the desirability of particular

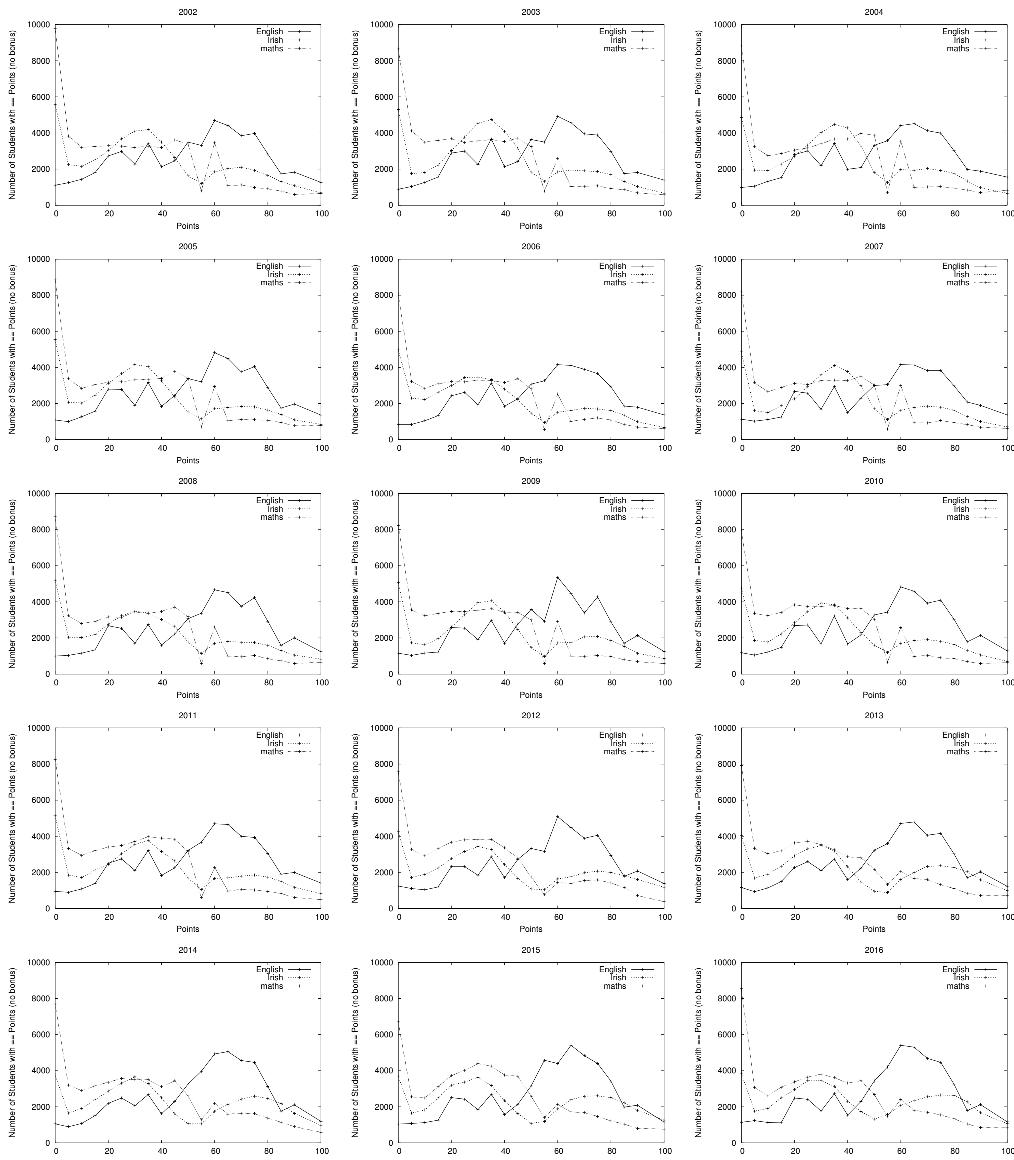


FIGURE 10. The number of students achieving a particular number of points in English, Irish and maths in 2002–2016 (excluding bonus points).

grades is having an impact on the rewards offered. Similar effects have been observed previously [14].

While the higher maths curve does not obviously show such features, possibly because of a more rigid marking scheme in mathematics, the ordinary-level maths curve does show one interesting persistent feature: the numbers achieving a D2 grade always seem to dip relative to the numbers achieving a D3 or E grade. While

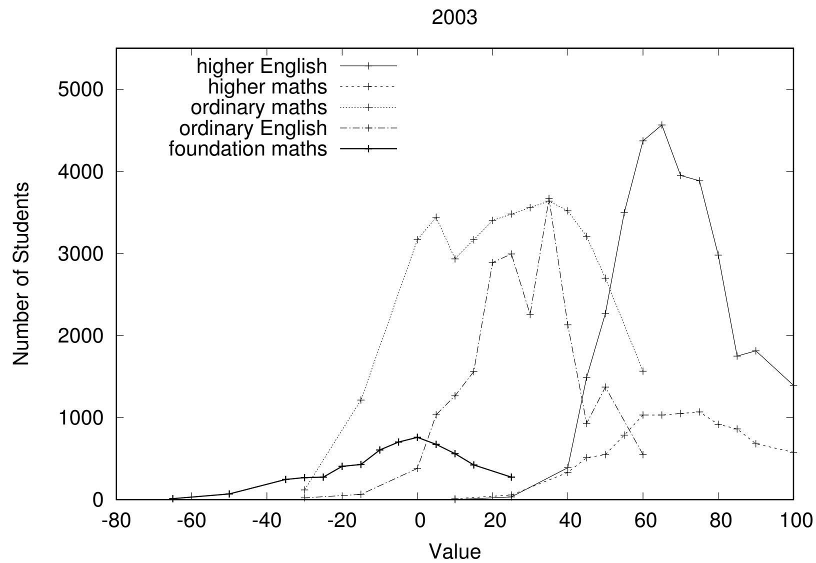


FIGURE 11. The number of students achieving different grades in each subject in English and maths in 2003.

the reason for this is not clear, it may be related to the "attempt mark" system often used for grading maths. A minimum attempt mark is assigned to all students who make a certain basic level of progress with each question, which awards about one third of the marks for the question[6]. The surge in these grades may represent students who are getting the majority of their marks from attempt marks.

3. Discussion

It is fairly clear that the three (practically) compulsory subjects are different in terms of the rewards achieved by students. Consider the following situations, in light of what we have seen.

The Price of Success: A student estimates she needs a B3 average (75 points) across 6 subjects to get her chosen college course. She believes that she will get a B3 in each of 5 optional subjects. Her Irish isn't good enough, but she can choose maths or English as her sixth points subject and take the other two at ordinary level. If she chooses English, she needs to be in the top 24% of people. If she chooses maths she needs to be in the top 8% of people in the country to get

[6]The exact operation of this system has varied slightly over the period.

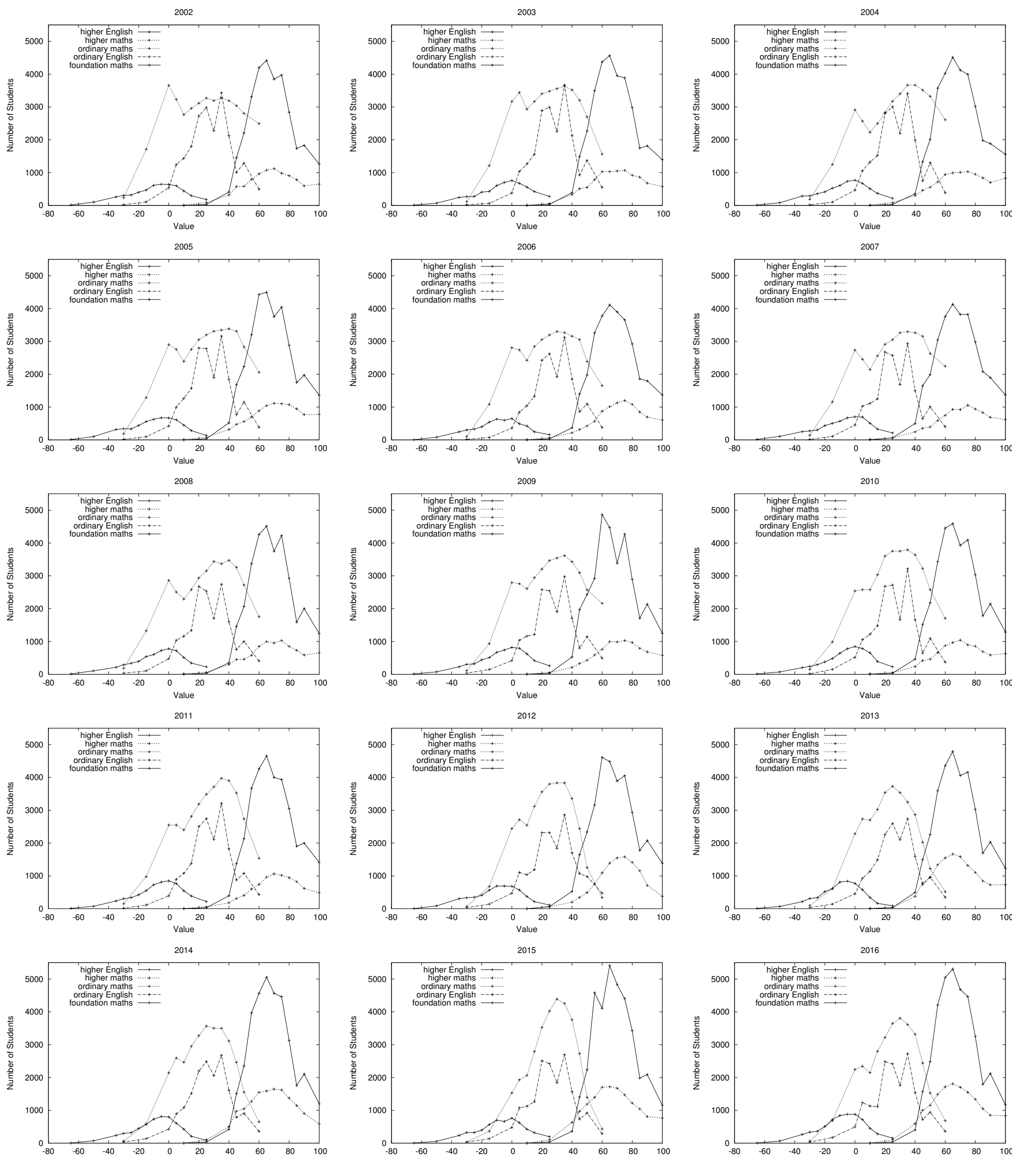


FIGURE 12. The number of students achieving different grades in each subject in English and maths from 2002–2016.

a B3, or top 22% to get 75 points with the higher level bonus. Which subject does she choose to take at higher level?

Fear of Failure: A student is borderline between higher maths and English. He *must* not fail either. He asks both teachers' advice. The maths teacher knows from experience that about 3.8% of people fail higher maths and that about 5–7% get As in ordinary maths, which are equivalent in points to a D1 or

D2 in higher level. The English teacher, from experience, has observed about a 1.4% failure rate at higher level and 7–9% of people getting an A at ordinary level.

Of course, where students can make these decisions, they may not have direct knowledge of such statistics. However the significant differences we have shown may provide anecdotal evidence that inform their decisions.

If there really is a marketplace for subjects, then this would indicate that students are not choosing higher maths because their effort can be better applied elsewhere. Increasing participation in higher maths is desirable for a range of reasons [5]. If we want to improve the take-up of maths at higher level, what can we do? Some possibilities present themselves.

(1) Reduce the effort required to study higher maths by making it easier,
(2) Reduce the effort required to study higher maths by improving educational support and accessibility,
(3) Increase the rewards associated with higher maths by increasing points offered,
(4) Decrease the risks associated with failing higher maths,
(5) Make other subjects less desirable.

We will not comment in detail on the last option, as it is unlikely to be popular with anyone.

Another proposed option [7] is to adjust the market, so that mathematics *must* be used as a points subject. This would mean that ordinary level maths becomes less attractive for students who are targeting high points.

We note that one could argue that the results we see are simply because our students are better at English and weaker at mathematics. However, this assumes that there is some absolute way to measure the difficulty of a Leaving Cert subject, other than having students study the subject and take the exam. We see no other obvious objective way of measuring the difficulty of, or level of reward associated with, a particular subject.

3.1. **Making Things Easy.** Making the mathematics course easier is unpalatable to many of the consumers of Leaving Certificate students. Universities, high-tech companies and professional bodies want students who have a certain level of mathematics, and some already find that not all students have a sufficiently high level of mathematics. There have been concerns raised about the new *Project Maths* course examined from 2012 onwards [21] and there is ongoing study on its impact on competence at third level [22].

There is also an obvious race to the bottom associated with making courses easier in order to make them more popular: if every subject repeatedly does this, all courses will gradually become easier and easier. This suggests that all Leaving Cert subjects should have some group of stakeholders who have a direct interest in the balance between the numbers taking the subject and the difficulty of the exam.

3.2. **Making Maths Better.** Improving the educational support available to students of mathematics has been on the agenda for many years. Studies have shown that there are a surprisingly large number of teachers of mathematics at second level who are teaching *out of field* [3, 19]. In particular, some commentators have suggested that the level of mathematics training required by a teacher before they can teach mathematics should be higher. Indeed, a number of courses at third level have been set up to facilitate the up-skilling of interested teachers (e.g. in UL and Maynooth). Perhaps measures that provide incentives for highly-numerate graduates to go into teaching could also be used, as they are already in high demand in the jobs marketplace. Schemes like this are in use in the UK. Certainly, it would be hard to argue against better teachers in any area, but any improvements will continue to be a long-term project.

The Project Maths course has largely been rolled out, and has been the primary examination in mathematics since 2012. It aims for students to achieve a better understanding of the material and how it is applied. The hope is that by making the applications of what is learned more obvious, that it will become more attractive to students. A more complete description of Project Maths is available [18].

It seems that changes to the emphasis and style of examination can have an impact on uptake. In 2012 the examination of Irish was changed to place more value on the Irish Oral examination. Figure 4 shows that this appears to have increased the uptake in higher Irish. While higher mathematics also shows an increase, corresponding to the introduction of Project Maths, the impact is confounded by the introduction of bonus points.

3.3. **Adjusting Points.** Increasing the points for maths has, of course, been often proposed, and bonus points were available for mathematics before[7]. When the mathematics syllabus was changed in the 1990s, bonus points were dropped because it was assumed that the new course would be more in line with other courses in terms of its difficulty. We can see that this does not seem to have been true. Indeed, looking at Leaving Cert exams from the 1990s up to 2011, there are many similarities. The latter part of each question ('part c') was still challenging and required students to combine understanding of various parts of the course. The style of questions in Project Maths is quite different, however there are still questions combining various parts of the syllabus.

Would bonus points actually encourage students to take higher maths? The evidence is not completely clear [11]. Before 2012, the University of Limerick (UL) offered bonus points for several years, but this did not seem to have significantly increased uptake among students who subsequently attend UL. However, it is not clear whether UL alone was big enough to have an impact on students' subject choices and whether those students affected ultimately went to UL. Since 2012, a bonus of 25 points for pass grades of higher maths has been in place, which we may be able to shed some light on. Indeed, it has been suggested that the points table might have to be extended to 120 or 150 points in order to account for differences between subjects [14], so it is worth considering the impact of any bonus and what aim it might have.

With the data available, we can calculate the effect of offering bonus points for maths, under the assumption demand for higher maths would remain unchanged. Say we aim to equalise the average points given by maths and English, we know that we need to raise the pre-2012 average by about 20 points. Since only about one in five students were taking higher maths, that suggests we need to give about 100 points, on average, to these higher-level students. In fact, if we give 100 extra points to all students passing higher maths, the average points for maths moves to about 50 points, which is only slightly below English. An alternative scheme, to multiply the points for higher maths by 2.5, also moves the average to around 50. This

[7]It was shown some time ago that Leaving Certificate maths was a good indicator of first year university performance [17], which provided a reason for weighting maths more highly for points up to the early 1990s.

also has some advantages over simply adding a fixed bonus, such as providing a slightly smaller step in points between a D3 and an E grade.

We expect that offering a substantial bonus would increase the numbers of students taking maths at higher level. However, this should also increase the average number of points given for maths, so equalising points using a pre-2012 year's statistics overestimates what is really required. Offering 60 bonus points at higher, or doubling points, would move the average for maths to about 40 points if the demand didn't change. This might be more acceptable and allow for change in demand. However, without understanding the feedback between the extra points and students' choices, it is impossible to say exactly what would be required to level the rewards offered by maths and English.

Our calculation suggests that the 25 bonus points offered from 2012 would fall significantly short of what would be required to equalise the average points, which we see confirmed in Figure 2. However, we do see an increase in the numbers taking higher maths in the first years after 25 bonus points were offered (see Figure 4), indicating that increased reward can indeed increase the numbers choosing to take a subject. It is worth observing that 2012 was also the first year in which Project Maths was examined on a wide scale, so the impact of this can not easily be separated from the offer of bonus points.

It is also worth noting that there might be other collateral effects of introducing bonus points for higher mathematics, from desirable effects (e.g. increasing the availability of higher maths) to the undesirable (e.g. over-influence of the third-level system on second-level studies, mathematics being used as a "booster" subject for points). Previous studies have discussed some of these issues [2, 17, 12]. Of course, there is also the option of decoupling the Leaving Cert and points, and so avoiding any impact that points could have on subject preferences. Previously, independent matriculation exams were available through various universities and admissions were based on these. However, it is unlikely there is an appetite to reintroduce separate matriculation examinations for our school leavers, and it seems likely that any anomalies created by the current points system would be inherited by the matriculation system.

3.4. **Bell Curving?** Note that there is some evidence that points are currently effectively being adjusted. Compare the consistency in points for higher, ordinary and foundation (Figure 3), even over significant course changes, to the lack of consistency seen between English, Irish and maths (Figure 2). This suggests that the grades are already being adjusted to ensure consistency at a particular level across subjects.

If this sort of adjustment is being applied, it raises a potentially serious issue when combined with a subject marketplace. If there is an easier alternative to higher level maths, some weaker students will move to ordinary level. The adjustment will then be applied, thus offering the same points to the remaining stronger students, which will effectively make higher-level maths even harder. This will cause more students to abandon higher maths, and so on. Meanwhile, the students moving from higher to ordinary are relatively strong for ordinary level, and so increase the standard there, pushing students who previously did ordinary into the lower grades there. If this positive feedback exists in practice, any difference between subjects will tend to be amplified by a combination of grade adjustment and student choices. This may even offer a explanation of the failure to achieve a 20–25% takeup rate in higher mathematics, which was the expected takeup for the syllabus introduced in 1992 [6].

We also expect that a better-rewarded subject will see the opposite happening. If a subject is seen to be an easier choice at higher level, then more students will take it at higher level. The average mark will remain the same, despite larger numbers taking it, so the marking will be adjusted to be easier, making the subject more attractive.

Normalising marking between subjects is a challenge. In optional subjects where the cohort taking the subjects is highly self-selecting, it seems that nothing short of introducing some sort of "jury service" would allow you to compare how the average student would cope with taking the course. However, given that basically everyone has to take English, Irish and maths, it should be possible to achieve better normalisation than is currently observed. One way to do this would be to adjust the points awarded for each subject so that the average across all students taking the subject is 50 points.

3.5. **Fewer Fails.** Some have suggested that an E grade at higher-level maths should be considered as a pass in certain circumstances, in order to reduce the risk associated with higher maths [8]. This has been adopted in the new Leaving Cert grading system introduced in 2017, which awards some points for a higher H7 grade (30–40% at higher level), which was previously regarded as a fail grade. The Irish University Association's guide to the grading system says this is to encourage uptake at higher level and also to reduce the risks associated with failing.

An alternative possibility that has been suggested is to allow students a second-chance exam if they fail mathematics at higher level, or to have a base-level maths exam that all students take followed by an advanced exam for those who wish to earn a higher grade.

4. Conclusion

We believe that looking at Leaving Cert subject/level choice as a market may offer interesting insights into the choices made by students. To understand subject uptake, it seems that regarding points as the reward offered by subjects, and considering the relative demands made by each subject should be a useful strategy. When we compare English, Irish and mathematics, we see that there are significant differences in the rewards offered to students, after the demands have been accounted for by student choice of subject level. In addition to quirks at grade boundaries, we also see evidence of grade normalisation, which could have a dangerous effect when combined with this subject market.

We have also tried to identify possible courses of action to improve the uptake of higher mathematics. It seems unlikely that a realistic application of any one of these would be sufficient to put mathematics and English on a par in this points market. It is also likely that any reform that improves mathematics in the Leaving Cert through syllabus reform, improved teaching, bonus points, etc., must take account of the subject market in which maths exists to be fully effective.

Thanks. We would like to thank Peter Clifford, Doug Leith, Sharon Murphy, Bill Lynch, Niall Murphy and Delma Byrne for comments and discussion on this work. Thanks to Maria Meehan for providing a summary of the results of her survey.

Appendix A. Notes on method

The statistics presented here are derived from the following files from the http://www.examinations.ie/ website:

```
2002_LC_GradesAwardedbySubject.pdf
2004_LC_Breakdown_Candidates_by_grade_Higher_Ordinary_and_Foundation_Level.pdf
2005_LC_Breakdown_Candidates_by_grade_Higher_Ordinary_and_Foundation_Level.pdf
2006_Breakdown_Candidates_by_grade_Higher_Ordinary_and_Foundation.pdf
2007_LC_2007_breakdownResults_10_or_More.pdf
2008_LCGrade_over_3_years.pdf
2009_lc_nat_stats_2009_211009_excluding_10.pdf
Provisional_Results_2010_excluding_subjects_with_less_than_10_candidates.pdf
Provisional_Results_LC_2011_National_Stats.doc
Provisional_Results_2012_less10.doc
2011_2013_LC_Statistics_no_less_than-10619.doc
Leaving_Certificate_2015_Provisional_Results_Less_Than_10_Candidates_Word.docx
BI-ST-9486403.csv
EN-ST-20970104.csv
```

Each file presents results from multiple years. A number of discrepancies in the reported results appear, presumably due to appeals of results being resolved. Where these arise, the results from the most recent file have been used. The breakdown for each year is given as a percentage, to two decimal places. However the least significant digit is always zero, so it seems likely they have been rounded to one decimal place. The sums of the percentages are between 99.7% and 100.3%, so some small discrepancies arise between the listed total number of students and cumulative figures over grades.

References

[1] Delma Byrne and Emer Smyth. No way back? The dynamics of early school leaving. Technical report, The Liffey Press in association with The Economic & Social Research Institute, 2010. http://eprints.nuim.ie/4333/1/DB_No_Way_Back.pdf.

[2] Commission on the Points System. Final report and recommendations. Technical report, Department of Education and Science, 1999. http://odtl.dcu.ie/mirror/irlgov/educ/pointcommreport99.html.

[3] Judith Cosgrove, Gerry Shiel, Elizabeth Oldham, and Nick Sofroniou. A survey of mathematics teachers in Ireland. *The Irish Journal of Education*, 35:20–44, 2004. http://www.jstor.org/stable/30077493?__redirected.

[4] Department of Education and Skills. Statistics report. Technical report, Department of Education & Skills, 2006. http://www.education.ie/servlet/blobservlet/des_stats_report_05_06.pdf.

[5] Engineers Ireland. Report of taskforce on education of mathematics and science at second level. Technical report, Engineers Ireland, 2010. http://www.engineersireland.ie/public/20100211-Mathematics_and_Science_at_Second_Level.pdf.

[6] Expert Group on Future Skills Needs. Statement on raising national mathematical achievment. Technical report, Tánaiste & Minister for Enterprise, Trade and Employment and the Minister for Education and Science, 2008. http://www.forfas.ie/media/egfsn081215_raising_mathematical_achievement.pdf.

[7] Seán Flynn. Maths should be compulsory CAO points subject. *Irish Times*, June 2010. http://www.irishtimes.com/newspaper/frontpage/2010/0608/1224272055721.html.

[8] Seán Flynn. Minister urges college places for those failing higher maths. *Irish Times*, May 2010. http://www.irishtimes.com/newspaper/frontpage/2010/0510/1224270051148.html.

[9] Damian F. Hannan, Richard Breen, Barbara Murray, Niamh Hardiman, Dorothy Watson, and Kathleen O'Higgins. Schooling and sex roles: Sex differences in subject provision and student choice in irish post-primary schools. Technical report, 2004.

[10] Áine Hyland. Entry to higher education in Ireland in the 21st century. Technical report, HCCA/HEA Seminar, 2011. http://www.hea.ie/sites/default/files/aine_hyland2c_entry_to_higher_education_in_ireland_in_the_21st_century_1.pdf.

[11] Irish Universities Association. IUA response to proposal to award bonus points for higher level leaving certificate mathematics. Technical report, Irish Universities Association, 2008. http://www.iua.ie/publications/documents/publications/2008/Bonus_points_mathematics_IUA_response_0608.pdf.

[12] Thomas Kellaghan and David Millar. Grading in the leaving certificate examination: A discussion paper. Technical report, Educational Research Centre, 2003.

[13] Maureen Lyons, Kathleen Lynch, Sean Close, Emer Sheerin, and Philip Boland. Inside classrooms: The teaching and learning of mathematics in social context. Technical report, Institute of Public Administration, 2003.

[14] Eoghan Mac Aogáin. The points system and grading of the leaving certificate examination. *The Irish Journal of Education*, 36:3–24, 2005. http://www.erc.ie/documents/vol36chp1.pdf.

[15] Olivia McEvoy. Report on the consultation with 5th and 6th year students on the reform of the transition from 2nd level to 3rd level education. Technical report, CNAG AR AN DORAS, 2013. http://ncca.ie/en/file/post_primary/T_1.pdf.

[16] Maria Meehan. Survey of UCD first economics and finance students. Personal Communication, 2009.

[17] M.A. Moran and M.J. Crowley. The leaving certificate and first year university performance. *Journal of the Statistical and Social Inquiry Society of Ireland*, 24, 1978. http://webir.tcd.ie/bitstream/2262/3992/1/jssisiVolXXIVPartI_231266.pdf.

[18] National Council for Curriculum and Assessment. Project maths — learning and teaching for the 21st century, 2009. http://projectmaths.ie/.

[19] Máire Ní Riordáin and Ailish Hannigan. Out-of-field teaching in post-primary mathematics education: An analysis of the irish context. Technical report, NCE-MSTL Univeristy of Limerick, November

2009. http://www.nce-mstl.ie/files/Out-of-field%20teaching%20in%20post-primary%20Maths%20Education.pdf.

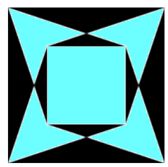
[20] Emer Smyth, Joanne Banks, and Emma Calvert. From leaving certificate to leaving school: A longitudinal study of sixth year students. Technical report, The Liffey Press in association with The Economic & Social Research Institute, National Council for Curriculum & Assessment and Department of Education & Skills, 2011. https://www.esri.ie/UserFiles/publications/BKMNEXT195.pdf.

[21] Cora Stack, Margaret Stack, and Ted Hurley. Project Maths is simply not challenging enough. *Irish Times*, October 2012. http://www.irishtimes.com/opinion/project-maths-is-simply-not-challenging-enough-1.527810.

[22] Páraic Treacy and Fiona Faulkner. Trends in basic mathematical competencies of beginning undergraduates in ireland, 2003–2013. *International Journal of Mathematical Education in Science and Technology*, 46(8):1182–1196, 2015.

(David Malone) Hamilton Institute and Department of Mathematics & Statistics, Maynooth University.

(Gavin McCullagh) Griffith College Dublin.
E-mail address: David.Malone@mu.ie



Irish Math. Soc. Bulletin
Number 81, Summer 2018, 57–95
ISSN 0791-5578

Aspects of Positive Scalar Curvature and Topology II

MARK G.WALSH

Abstract. This is the second and concluding part of a survey article. Whether or not a smooth manifold admits a Riemannian metric whose scalar curvature function is strictly positive is a problem which has been extensively studied by geometers and topologists alike. More recently, attention has shifted to another intriguing problem. Given a smooth manifold which admits metrics of positive scalar curvature, what can we say about the topology of the *space* of such metrics? We provide a brief survey, aimed at the non-expert, which is intended to provide a gentle introduction to some of the work done on these deep questions.

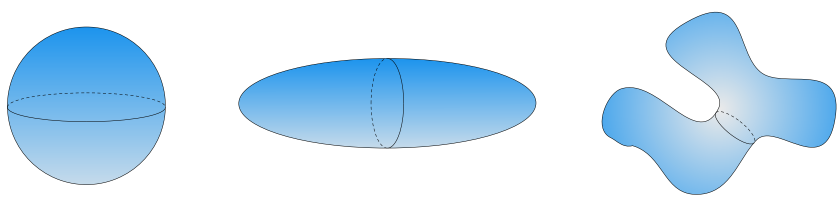


Figure 1. A selection of geometric structures on the sphere

4. The Space of Metrics of Positive Scalar Curvature

We now consider the second of our introductory questions. What can we say about the topology of the space of psc-metrics on a given smooth compact manifold? Before discussing this any further it is worth pausing to consider what we mean by a space of metrics in the first place. Recall Fig. 1, where we depict 3 distinct metrics on the 2-sphere, S^2. Each of the three images represents a point

2010 *Mathematics Subject Classification.* 46A99, 10A10.
Key words and phrases. Riemannian metrics of positive scalar curvature, spin manifolds, surgery, cobordism, Morse functions, loop spaces.
Received on 15-12-2017.
The author acknowledges support from Simons Foundation Collaboration Grant No. 280310.

in the space of metrics. One can now imagine traveling on a path in this space to consist of an animation over time which moves one such picture into another, continuously stretching and warping the sphere. Although, in our minds we usually think of these shapes extrinsically as embedded into $\mathbb{R}^3$, all of these metrics can be thought of as intrinsic geometric structures on S^2. Picturing them this way is a tougher mental exercise but worth doing, especially if one wants to consider the space of metrics on a manifold which does not embed in $\mathbb{R}^3$.

Suppose we add a constraint to our metrics on S^2. We now consider only metrics in this space with positive (Gaussian/scalar) curvature. This subspace is certainly non-empty: the round metric has positive curvature! Consider animations of the round metric which, at every frame, satisfy positive curvature. The second image in Fig.1 seems attainable, but the third metric not so as it undoubtably has some negative curvature. One question we might ask concerns the connectivity of this space. Do there exist positive curvature metrics on S^2 which cannot be connected by a path (through positive curvature metrics) to the round metric? In other words, are there distinct islands of positive curvature metrics? More generally, what can we say about higher notions of connectivity such as the fundamental group (do these islands contain lakes?) or more general homotopy groups? What about the analogous spaces for other manifolds?

It turns out that the answer to all of these questions in the case of S^2 is no. We know from work of Rosenberg and Stolz, making use of the Uniformisation Theorem, that the space of positive curvature metrics on the 2-dimensional sphere is actually a contractible space; see [45]. In a sense this is not too surprising, given the strict limitations placed by positive curvature. When imagining positive curvature geometries on S^2, it is difficult to stray too far beyond objects like the first two pictures on Fig. 1. However, if one sufficiently increases the dimension of the sphere, a great deal of non-trivial topology emerges in these spaces. For example, we know from the work of Carr in [13], that the space of positive scalar curvature metrics on the 7-sphere has infinitely many distinct path components. More generally, there are large numbers of manifolds whose space of positive scalar curvature metrics has non-trivial topological information at multiple levels. We will now attempt to give a taste of this story.

We begin with some preliminary considerations. As before, M denotes a smooth closed (compact with empty boundary) manifold of dimension $n \geqslant 2$. We let $\mathcal{R}(M)$ denote the space of all Riemannian metrics on M, under its usual C^{∞}-topology; see chapter 1 of [47] for details about this topology. This in fact gives $\mathcal{R}(M)$ the structure of an infinite dimensional Fréchet manifold; see chapter 1 of [47]. This enormous, infinite dimensional space is convex and so is, in a sense, not so interesting topologically. However, by specifying some geometric constraint, C, we can restrict to the subspace $\mathcal{R}^{C}(M) \subset \mathcal{R}(M)$ of metrics which satisfy this geometric constraint. Of course, depending on the constraint, this subspace may be empty. However, when non-empty, such subspaces may be very interesting indeed from a topological point of view. There are many geometric constraints which are of interest. For example, those interested in positive curvature may wish to study the spaces $\mathcal{R}^{\mathrm{Sec}>0}(M)$, $\mathcal{R}^{\mathrm{Ric}>0}(M)$ or $\mathcal{R}^{s>0}(M)$, the open subspaces of $\mathcal{R}(M)$ consisting of metrics with positive sectional, Ricci or scalar curvature respectively. Alternatively, one might be interested in geometric conditions such as non-negative, constant or negative curvature. Although we will say a few words later about some alternate geometric constraints,[5] our focus here is on positive scalar curvature and on understanding the topology of the space, $\mathcal{R}^{s>0}(M)$. This problem has aroused considerable attention in recent years.

At this point, we should bring up another space which is closely associated with $\mathcal{R}(M)$. This is the moduli-space of Riemannian metrics, denoted $\mathcal{M}(M)$. Before defining it we point out that two Riemannian metrics, g and g' on M, are *isometric* if there is a diffeomorphism $\phi : M \to M$ so that $g' = \phi^* g$. Here, $\phi^* g$ is the "pull-back" of the metric g under the diffeomorphism ϕ and is defined by the formula

$$\phi^* g(u, v)_x = g(d\phi_x(u), d\phi_x(v))_{\phi(x)},$$

where $x \in M$, $u, v \in T_x M$ and $d\phi_x : T_x M \to T_{\phi(x)} M$ is the derivative of ϕ. This determines an action of the group $\mathrm{Diff}(M)$, the group of self-diffeomorphisms $M \to M$, on $\mathcal{R}(M)$. The moduli space $\mathcal{M}(M)$ is then obtained as the quotient of this action on $\mathcal{R}(M)$ and, thus, is obtained from $\mathcal{R}(M)$ by identifying isometric metrics. For some, this is a more meaningful interpretation of the space of "geometries"

[5] It is also interesting to drop the compactness requirement on M, although we will not say much about non-compact manifolds here.

on M, although this is a subject of debate. Restricting the above action to a subspace of $\mathcal{R}(M)$ which satisfies a given curvature constraint leads to the moduli space of Riemannian metrics which satisfy this constraint. In particular, we will consider the moduli space of positive scalar curvature metrics: $\mathcal{M}^{s>0}(M)$. To summarise, we have the following commutative diagram, where the horizontal maps denote projections while the vertical maps are inclusions.

$$\begin{tikzcd}
\mathcal{R}(M) \arrow[r] & \mathcal{M}(M) \\
\mathcal{R}^{s>0}(M) \arrow[r] \arrow[u, hook] & \mathcal{M}^{s>0}(M) \arrow[u, hook]
\end{tikzcd}$$

The earliest results displaying topological non-triviality in the space of Riemannian metrics of positive scalar curvature are due to the landmark work of Hitchin in [29]. Given a closed smooth spin manifold M, Hitchin showed, via the α-invariant, that the action of the diffeomorphism group $\mathrm{Diff}(M)$ could be used to show that, in certain dimensions, $\mathcal{R}^{s>0}(M)$ is not path connected i.e. $\pi_0(\mathcal{R}^{s>0}(M)) \neq 0$. In fact, Hitchin also showed that $\mathcal{R}^{s>0}(M)$ may not be simply connected ($\pi_1(\mathcal{R}^{s>0}(M)) \neq 0$). We will not discuss the details of Hitchin's work here, as it will take us too far afield; for a concise discussion see [47]. One immediate consequence of Hitchin's work is the topological difference between $\mathcal{R}^{s>0}(M)$ and $\mathcal{M}^{s>0}(M)$. As the non-triviality Hitchin exhibits arises from the action of $\mathrm{Diff}(M)$, all of it disappears once we descend to the moduli space. It is important to realise that, as later results show, it is certainly possible for topological non-triviality in $\mathcal{R}^{s>0}(M)$ to survive in $\mathcal{M}^{s>0}(M)$. This is something we will discuss shortly. However it is worth pausing for a moment to consider the implication of two psc-metrics which lie in distinct path components of $\mathcal{R}^{s>0}(M)$, being projected to the same point in $\mathcal{M}^{s>0}(M)$. It is therefore perfectly possible for two Riemannian metrics of positive scalar curvature to be isometric and yet not connected by a path through psc-metrics! Intuitively, one could think of psc-metrics which are "mirror images" of each other (and thus isometric) but for which there exists no continuous animation from one to the other which, at every frame, satisfies the curvature constraint.

In recent years, we have learned that there is a great deal of topological non-triviality in the spaces $\mathcal{R}^{s>0}(M)$ and $\mathcal{M}^{s>0}(M)$. These results usually require that M be a spin manifold, although not always. Hitchin's results in particular have been significantly strengthened to deal with higher connectivity questions; see [17], [26] and [9]. Roughly, these results make use of a particular variation of the Dirac index, introduced by Hitchin in [29], and show that for certain closed spin manifolds, M, and certain psc-metrics, g, there are non-trivial homomorphisms

$$A_k(M,g) : \pi_k(\mathcal{R}^{s>0}(M), g) \longrightarrow KO_{k+n+1}.$$

The latter paper [9] by Botvinnik, Ebert and Randal-Williams contains results which are particularly powerful showing that, when the manifold dimension is at least six, this map is always non-trivial when the codomain is non-trivial. Indeed, the non-triviality detected in this paper captures not simply the non-triviality displayed in Hitchin's work but effectively all of the topological non-triviality known (for spin manifolds) up to this point. Their methods are new, highly homotopy theoretic and make use of work done by Randal-Williams and Galatius on moduli spaces of manifolds; see [23]. At the time of writing it appears that Perlmutter (whose work we will briefly mention a little later) has, using techniques developed in [44], now extended this theorem to hold for dimension 5 also. It is important to point out however that, unlike many of the results it subsumes, this is purely an existence result. In many papers, including those by Crowley and Schick [17] and by Hanke, Schick and Steimle [26], specific non-trivial elements are constructed. Indeed, the latter paper constructs an especially interesting class of examples, something we will say a few words about later on. Before we can continue this discussion however, there are some important concepts we need to introduce.

4.1. **Path Connectivity, Isotopy and Concordance.** As already suggested, the logical first question when studying the topology of the space $\mathcal{R}^{s>0}(M)$ (or $\mathcal{M}^{s>0}(M)$) concerns path connectivity. Is this space path connected? Recall that earlier we mentioned that the space of all psc-metrics on the 2-dimensional sphere S^2, $\mathcal{R}^{s>0}(S^2)$, is a contractible and therefore path-connected space; see [45]. This theorem also implies that $\mathbb{R}P^2$, the only other closed 2-dimensional manifold to admit a psc-metric, has a contractible space of psc-metrics also. This is not so surprising given the constraints positive scalar curvature place at this dimension. Indeed, in dimension 3 the situation may be similar. We know from recent work of Coda-Marques (see [16]) that the space $\mathcal{R}^{s>0}(S^3)$ is path-connected, while a number of experts have suggested that this space may well be contractible also. However, as we increase the dimension n, the scalar curvature becomes more and more flexible and so the possibility for more exotic kinds of geometric structure increases.

In dimension 4, for example, there are examples of 4-dimensional manifolds whose space of psc-metrics (as well as the corresponding moduli space) has many (even infinitely many) path components; see for example work by Ruberman in [46] and recent work by Auckly, Kim, Melvin and Ruberman in [2]. It should be said that the methods used here, such as Seiberg-Witten theory, are specific to dimension 4 and do not apply more generally. Moreover, the manifolds used here are non-trivial in their own right. In particular, they are not spheres. In fact, it is still an open question as to whether the space of psc-metrics on the 4-dimensional sphere, $\mathcal{R}^{s>0}(S^4)$, is path connected or has any non-trivial topology at all. Given that dimension 4 is very often a special case, with features and pathologies all of its own, we will focus on more general techniques for detecting disjoint path components of psc-metrics for manifolds in dimensions $\geqslant 5$.

In order to discuss the problem properly, there are a couple of helpful terms we must define: isotopy and concordance. Given the frequent use of these terms in various mathematical contexts, in particular in studying spaces of diffeomorphisms (a context which overlaps with ours), we will add the prefix "psc." Two psc-metrics g_0 and g_1 in $\mathcal{R}^{s>0}(M)$ are said to be *psc-isotopic* if they lie in the same path component of $\mathcal{R}^{s>0}(M)$, i.e. there is a continuous path $t \mapsto g_t \in \mathcal{R}^{s>0}(M)$, where $t \in [0,1]$, connecting g_0 and g_1. Such a path is called a psc-isotopy. The metrics g_0 and g_1 are said to be *psc-concordant* if there exists a psc-metric $\bar{g}$ on the cylinder $M \times [0,1]$, which near each end $M \times \{i\}$, where $i \in \{0,1\}$, respectively takes the form $g_i + dt^2$. Such a metric on the cylinder is known as a psc-concordance; see Fig. 8.

It is not difficult to verify that both notions determine equivalence relations on the set $\mathcal{R}^{s>0}(M)$, the equivalence classes of psc-isotopy being simply the path components of $\mathcal{R}^{s>0}(M)$. Moreover, it follows from a relatively straightforward calculation that psc-isotopic metrics are necessarily psc-concordant; both [21] and [25] contain

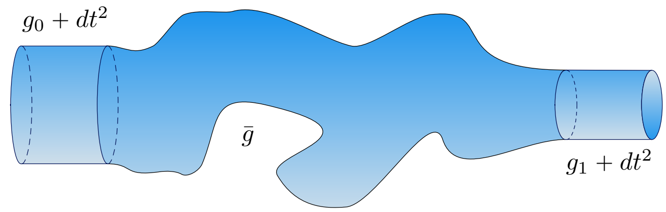


FIGURE 8. A psc-concordance $\bar{g}$ between g_0 and g_1.

versions of this calculation. The idea here is to consider the "warped product" metric $g_t + dt^2$ on the cylinder $M \times [0, 1]$ arising from a psc-isotopy $\{g_t\}_{t\in[0,1]}$ between g_0 and g_1. Although this metric may not have positive scalar curvature, since there may be unwelcome negative curvature arising in the t-direction, it can be rescaled to sufficiently slow down transition in the t-direction so that the positivity of the slices can compensate and so that the resulting metric satisfies a product structure near the ends. Thus, given a psc-isotopy, we can always construct a psc-concordance. This observation suggests at least a strategy for exhibiting distinct path components of the space $\mathcal{R}^{s>0}(M)$. Namely, show that there are psc-metrics in this space which are not psc-concordant. Given the index obstruction discussed earlier, this turns out to be a very reasonable strategy.

Before we discuss how to exhibit distinct psc-concordance classes (and thus distinct path components of $\mathcal{R}^{s>0}(M)$), let us consider the converse to the observation we have just made. Are psc-concordant metrics necessarily psc-isotopic? This is an intriguing question. The short answer is no. We know from work of Ruberman, that the Seiberg-Witten invariant detects psc-concordant metrics which are not psc-isotopic in the case of certain 4-dimensional manifolds; see [46]. But, as we noted earlier, dimension 4 has some very specific features. It might still be the case that in higher dimensions, the two notions coincide. In particular, what about the "reasonable" case of simply connected manifolds of dimension at least five (still a huge class of manifolds)?

For a long time this problem remained completely open, with little hope of progress. In [48], working in this reasonable realm of simply connected manifolds with dimension at least five, this author gave a partial affirmative answer to this question. The result held for a specific kind of psc-concordance arising via the Gromov-Lawson surgery construction, a so-called Gromov-Lawson concordance (something we will look at shortly). But it was still completely unclear as to how one would approach the general problem. We now have, from substantial work by Botvinnik in [7] and [8], the following theorem.

Theorem 4.1. (Botvinnik [8]) *Let M be a simply connected manifold of dimension at least 5 which admits psc-metrics. Then two psc-metrics on M are psc-concordant if and only if they are psc-isotopic.*

Botvinnik's work also deals with the non-simply connected case although the story here is a little more complicated. One reason is that, in this case, a certain space of diffeomorphisms on $M \times [0, 1]$ may not be path-connected. This allows for the construction of more "exotic" types of psc-concordance. The formulation of the theorem in this case must take into account the fundamental group of the manifold M and something called its Whitehead torsion, a notion we will not discuss here. The proof of Botvinnik's theorem is formidable, incorporating deep theorems in Differential Topology and Geometric Analysis. This is not at all the place for an in-depth discussion of the proof. However it is worth making one remark on why such a proof might be so difficult. Consider the possible complexity of an arbitrary psc-concordance. The sort of psc-concordance discussed earlier, obtained by stretching a warped product metric, is about the tamest kind of psc-concordance imaginable. It does, after all, have a slicewise positive scalar curvature structure. But imagine how such a tame psc-concordance could be made "wild," by the mere act of taking a connected sum, via the Surgery Theorem, with an appropriate psc-metric on the sphere. This would not change the topology of the cylinder, yet, if the psc-metric on the sphere was suitably ugly, could produce a monstrous psc-concordance; see Fig. 9 below.

Indeed, it was observed by Gromov, that the problem of deciding whether or not two psc-concordant metrics are psc-isotopic is, in fact, *algorithmically unsolvable.* Gromov's argument makes use of the well-known fact that the problem of recognising the trivial group from an arbitrary set of generators and relations is algorithmically unsolvable. The idea is to build an arbitrarily complicated "unrecognisable" psc-concordance which represents such an arbitrary set of generators and relations, by means of a geometric construction using appropriate cells and attaching relations to build a representative cellular complex. For details, see Theorem 1.1 of [8]. As Botvinnik

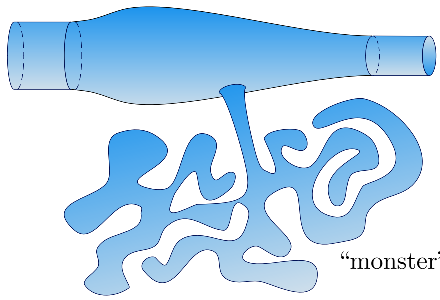


FIGURE 9. A tame psc-concordance made wild by connected sum with a "monster" psc-metric on the sphere.

demonstrates in his theorem, this does not mean that the problem is impossible to solve, just that its solution requires non-algorithmic tools, like surgery.

4.2. **Positive Scalar Curvature Cobordism.** For much of the remainder of this article, we will focus on the problem of demonstrating topological non-triviality in the space of metrics of positive scalar curvature. An important strategy in this endeavour, is to develop tools for constructing "interesting" examples of psc-metrics and, in particular, interesting families of psc-metrics. By interesting, we really mean topologically non-trivial. In the case of individual psc-metrics, this means psc-metrics which lie in distinct path components of the space of psc-metrics. More generally, we want to exhibit families of psc-metrics which represent non-trivial elements of the higher homotopy or homology groups of this space.

One approach is to make use of the action of the diffeomorphism group. As we mentioned earlier, this was part of Hitchin's technique in [29] where he exhibited such non-triviality at the level of path connectivity and the fundamental group for certain spin manifolds. We now consider another method. The principle tool we have for constructing examples of psc-metrics is the surgery technique of Gromov and Lawson. Recall that, given a manifold M and a psc-metric g on M, this technique allows us to construct explicitly a psc-metric g' on a manifold M' which is obtained from M by surgery in codimension at least three (an admissible surgery). In this section, we will consider a useful strengthening of this construction.

By way of motivation, consider a finite sequence of n-dimensional smooth compact manifolds $M_0, M_1, M_2, \cdots, M_k$, where each M_i is obtained from M_{i-1} by an admissible surgery. Thus, any psc-metric g_0 on M_0, gives rise by way of the Gromov-Lawson construction to a collection of psc-metrics $g_0, g_1, \cdots, g_k$ on the respective manifolds. Suppose furthermore that $M_0 = M_k$ (in practice we would need to identify these manifolds via some diffeomorphism $M_0 \cong M_k$, but for the sake of exposition we ignore this). This is a sort of "cyclic" condition on the sequence. It is important to realise that there are many interesting ways in which surgeries can "cancel" and the original smooth topology of M_0 be restored. Indeed it is even possible, when $n = 2p$, that a single surgery on a p-dimensional sphere in M_0 result in a manifold $M_1 = M_0$. In any case, assuming the sequence is such that M_0 is restored at the end, we may now compare the psc-metrics g_0 and g_k. Although $M_0 = M_k$, the psc-metric g_k, having possibly undergone multiple surgeries, may look very different from g_0. So how different are these psc-metrics? Could they now lie in different psc-isotopy classes? As we will shortly see, the answer to this question is yes.

To better understand the effects of surgery on psc-metrics, we need to reintroduce cobordism. Recall that a pair of smooth closed n-dimensional manifolds M_0 and M_1 are cobordant if there is a smooth compact $(n+1)$-dimensional manifold W with $\partial W = M_0 \sqcup M_1$. We now consider the following question.

Question 4. Given a psc-metric, g_0 on M_0, does this metric extend to a psc-metric, $\bar{g}$, on W which takes a product structure near the boundary of W? Thus, if it exists, the resulting metric $\bar{g}$ would satisfy $\bar{g} = g_0 + dt^2$ near M_0 and $\bar{g} = g_1 + dt^2$ near M_1, for some psc-metric g_1 on M_1.

The answer to this question depends on certain topological considerations. Before discussing this further we need to discuss some facts about cobordism and surgery.

Recall that surgery preserves the cobordism type of a manifold. Moreover, cobordant manifolds are always related by surgery. More precisely, the fact that M_0 and M_1 are cobordant means that M_1 can be obtained by successively applying finitely many surgeries to M_0 and vice versa. One way of achieving this is with a Morse function $f : W \to [0, 1]$. This is a certain smooth function satisfying $f^{-1}(i) = M_i$, where $i \in \{0, 1\}$, with only finitely many critical points all in the interior of W and satisfying the condition that, at each critical point w, $\det D^2 f(w) \neq 0$. Here, $D^2 f(w)$ is the Hessian of f at w. This latter condition means that critical points of f are of the simplest possible form and, by a lemma of Morse (see [42]), there is a choice of local coordinates $x = (x_1, \cdots, x_{n+1})$ near w where the function f takes the form

$$f(x) = c - \sum_{i=1}^{p+1} x_i^2 + \sum_{i=p+2}^{n+1} x_i^2,$$

where $f(w) = c$ and $p \in \{-1, 0, 1, \cdots, n\}$. The number $p + 1$ is referred to as the *Morse index* of the critical point w. Specifically, this is the dimension of the negative eigenspace of $D^2 f(w)$ and so is independent of any coordinate choice.

Let us assume for simplicity that w is the only critical point in the level set $f^{-1}(c)$ (this can always be obtained by a minor perturbation of the function). Then, for some $\epsilon > 0$, $f^{-1}[c - \epsilon, c + \epsilon]$ contains only w as a critical point. Moreover the level sets $f^{-1}(c \pm \epsilon)$ are smooth n-dimensional manifolds (since they contain no critical points) and, most importantly, $f^{-1}(c+\epsilon)$ is obtained from $f^{-1}(c-\epsilon)$ via a surgery on an embedded p-dimensional sphere. The cobordism $f^{-1}[c-\epsilon, c+\epsilon]$ is called an *elementary cobordism*, as it involves only one surgery. Equivalently, it is also referred to as the *trace of the surgery* on $f^{-1}(c-\epsilon)$. All of this means that any cobordism W can be decomposed into a collection of elementary cobordisms. Moreover, any surgery has a corresponding cobordism, its trace, associated to it. Thus, the sort of sequence of surgeries we introduced to motivate this section, are more efficiently described using cobordisms.

Let us suppose now that W is given such a decomposition into elementary cobordisms and that each is the trace of a surgery in codimension at least three (the key hypothesis in the Surgery Theorem). This is equivalent to saying that W admits a Morse function $f : W \to [0, 1]$ in which each critical point has Morse index $\leqslant n-2$; such a Morse function is regarded as *admissible*. Given an admissible Morse function, f, on W then, it follows from a theorem of Gajer in [21] and later work by this author in [48], that the Surgery Theorem can be strengthened to extend a psc-metric g_0 on M_0 to a psc-metric on W satisfying the product structure described above; see Fig. 10.

This provides at least sufficient conditions for an affirmative answer to question 4. Although there are a number of choices made in this construction, the resulting psc-metric which is denoted $\bar{g} = \bar{g}(g_0, f)$ and known as a *Gromov-Lawson cobordism*, depends for the most part only on the initial psc-metric g_0 and the Morse function f. The extent to which the geometry of the psc-metric $\bar{g}$ (and in particular g_1) is affected by different choices of g_0 and f (given a fixed W) is an interesting problem in its own right and something we will return to shortly.

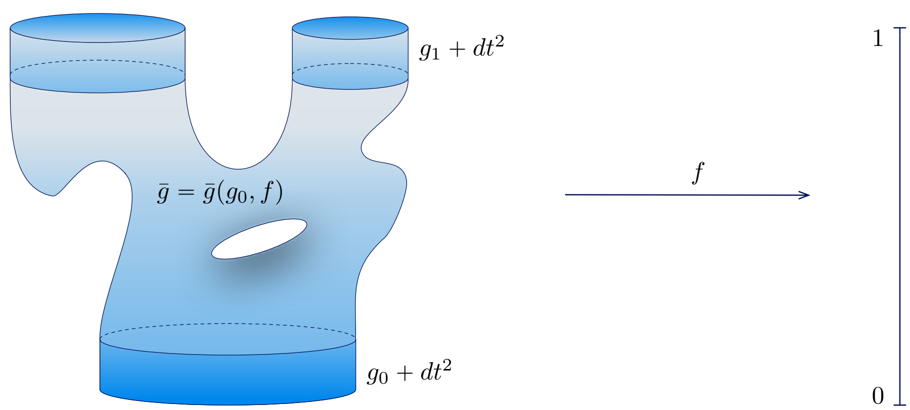


FIGURE 10. The Gromov-Lawson cobordism $\bar{g}$ arising from an admissible Morse function $f : W \to [0, 1]$ and a psc-metric g_0 on $M_0 = f^{-1}(0)$

Let us return to the problem which motivated this section. When the Gromov-Lawson construction is applied to a psc-metric over a finite sequence of admissible surgeries which result in a manifold which is the same as the starting manifold, how different are the starting and finishing psc-metrics? Equivalently, and more succinctly, if W is a cobordism with $M_0 = M_1$, $f : W \to [0, 1]$ is an admissible Morse function and g_0 is a psc-metric on M_0, how "different" can g_0 and $g_1 = \bar{g}(g_0, f)|_{M_1}$ be? In particular, could the psc-metrics g_0 and g_1 be in different path components of $\mathcal{R}^+(M_0)$? The answer to this question is yes, as we shall now demonstrate.

4.3. **Non-isotopic psc-metrics.** As we have mentioned there are many compact spin manifolds whose spin cobordism class does not lie in the kernel of the α-homomorphism and thus do **not** admit metrics of positive scalar curvature. One important example of such a manifold is a Bott manifold, named for its role in generating Bott periodicity. We consider such a manifold, denoted B, an 8-dimensional simply connected spin manifold which satisfies $\alpha([B]) = 1 \in KO_8 \cong \mathbb{Z}$. Importantly, $\alpha([B]) \neq 0$ and so B admits no psc-metrics. The topology of B is well understood; see [33] for a geometric construction. In particular, suppose we remove a pair of disjoint 8-dimensional disks, D_0 and D_1, from B. Then the resulting cobordism of 7-dimensional spheres, $W = B \setminus (D_0 \sqcup D_1)$, admits an admissible Morse function.

Recalling the Gromov-Lawson cobordism construction, we let $f : W \to [0,1]$ denote such an admissible Morse function, S_0 and S_1 denote the 7-dimensional boundary spheres of the cobordism W and $g_0 = ds_7^2$ denote the standard round metric on the S_0-boundary sphere. We now apply the Gromov-Lawson cobordism construction. The resulting psc-metric on W, $\bar{g} = \bar{g}(g_0, f)$ restricts as a psc-metric g_1 on S_1. Suppose now that g_0 and g_1 are psc-concordant and we denote by $\bar{h}$ such a psc-concordance. We now obtain a contradiction: The round psc-metric on S_0 trivially extends as a psc-metric on the disk D_0 with appropriate product structure near the boundary (a so called "torpedo" metric). By attaching the concordance $\bar{h}$ to the other end of W, at S_1, we can similarly extend g_1 to a psc-metric on D_1. But, as Fig. 11 suggests, this results in the construction of a psc-metric on B, something we know to be impossible. Thus, g_0 and g_1 are not psc-concordant and hence not psc-isotopic. Moreover, by "stacking" multiple copies of W and repeating this process, one can obtain infinitely many psc-metrics in $\mathcal{R}^{s>0}(S^7)$ which must all lie in distinct path components.

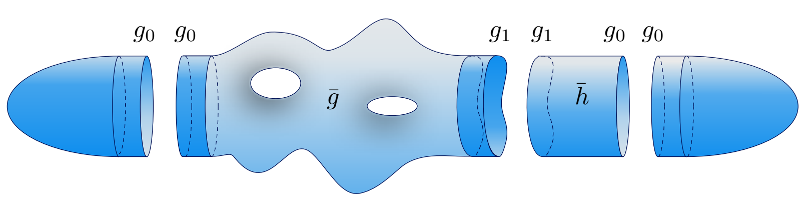


FIGURE 11. The impossible geometric decomposition of B: (left to right) the disk D_0 with torpedo metric, the cobordism W with metric $\bar{g}$, the cylinder $S_1 \times [0,1]$ equipped the proposed concordance $\bar{h}$ between g_1 and g_0 and the disk D_1 with torpedo metric

Although this is not quite the same as Carr's original proof in [13], it is very close and works for the same reasons. The argument generalises as Theorem 4.2 below. Similar arguments, making use of the α-invariant, have been used to show that $\mathcal{R}^{s>0}(M)$ has many, often infinitely many, path components for various compact spin manifolds M; see for example the work of Botvinnik and Gilkey in [12].

Theorem 4.2. (Carr [13]) *The space $\mathcal{R}^{s>0}(S^{4k-1})$ has infinitely many path components when $k \geqslant 2$.*

Interestingly, the fact that $\mathcal{R}^{s>0}(S^{4k-1})$ has *infinitely* many path components (when $k \geqslant 2$) has another important consequence, one which answers an earlier question about topological non-triviality surviving in the moduli space. It is known from work of Milnor, Kervaire [35] and Cerf [14] that the space of self-diffeomorphisms of the sphere, $\mathrm{Diff}(S^n)$, has only finitely many path components, for all n. Thus, at most finitely many of the path components demonstrated by Carr are lost when we descend to the moduli space $\mathcal{M}^{s>0}(S^n)$, meaning that this space also has infinitely many path components. Indeed this fact can be shown to hold for certain other spin manifolds by using an invariant constructed by Kreck and Stolz, called the s-invariant; see chapters 5 and 6 of [47] for a lively discussion of the s-invariant and its applications. We will not define the s-invariant save to say that it assigns a rational number $s(M, g) \in \mathbb{Q}$ to a pair consisting of a smooth closed manifold M with dimension $n = 4k - 1$ and a psc-metric g. The manifold itself must satisfy certain other topological conditions concerning the vanishing of particular cohomology classes. Under the right circumstances, $|s(M, g)|$ is actually an invariant of the path component of the moduli space of psc-metrics containing g. Thus, it can often detect when $\mathcal{M}^{s>0}(M)$ is not path connected.

4.4. **Some observations about the Gromov-Lawson Cobordism Construction.** We return once more to the general construction, given an admissible Morse function $f : W \to [0, 1]$ on a smooth compact cobordism W with $\partial W = M_0 \sqcup M_1$ and a psc-metric g_0 on M_0, of a Gromov-Lawson cobordism. Recall that this is a certain psc-metric $\bar{g} = \bar{g}(g_0, f)$ on W which extends g_0 and takes the form of a product metric near the boundary ∂W. In the next section we will discuss a "family" version of this construction where g_0 and

f are respectively replaced by certain families of psc-metrics and smooth functions. Here, we preempt this discussion by recalling the question of the dependency of this construction on the choices of psc-metric g_0 or Morse function f. Regarding g_0, it is demonstrated in [48] that the Gromov-Lawson cobordism construction goes through without a hitch for a compact continuously parameterised family of psc-metrics $t \mapsto g_0(t) \in \mathcal{R}^{s>0}(M_0)$, where $t \in K$ and K is some compact space. In particular, this means that if g_0 and g_0' are psc-isotopic metrics on M_0, the resulting psc-metrics $\bar{g} = \bar{g}(g_0, f)$ and $\bar{g}' = \bar{g}(g_0', f)$, as well as $g_1 = \bar{g}|_{M_1}$ and $g_1' = \bar{g}'|_{M_1}$, are psc-isotopic in their respective spaces of psc-metrics.

Perhaps a more interesting question concerns the choice of admissible Morse function. Before considering this, it is important to realise that a given manifold W will admit many many different Morse functions. For simplicity, let us assume that W is the cylinder $M_0 \times [0, 1]$. The projection functions schematically depicted in Fig. 12, which are composed with appropriate embeddings of this cylinder, determine two very different Morse functions. Thus, the fact that the cylinder is the (topologically) simplest cobordism does not prevent the existence of Morse functions with a great many critical points. Such Morse functions in turn lead to highly non-trivial decompositions of the cylinder into many non-cylindrical pieces. For such a non-trivial (and admissible) Morse function, f, the psc-metrics $\bar{g}(g_0, f)$ and g_1, may be very complicated, especially when compared with the analogous psc-metrics for the standard projection with no critical points.

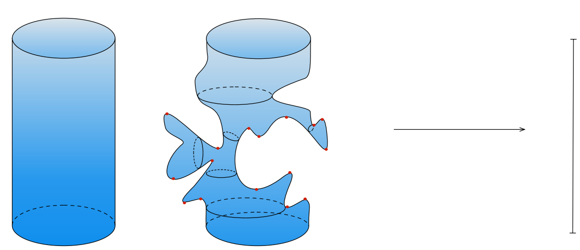


FIGURE 12. Two Morse functions on the cylinder $M \times [0, 1]$, one with no critical points, one with many "canceling" critical points

Suppose we denote by $\mathcal{M}\mathrm{or}(W)$, the space (under the usual C^∞-Whitney topology) of all Morse functions $W \to [0, 1]$. This space is always non-empty (in fact it is an open dense subspace of the space of all smooth functions $W \to [0,1]$). We will assume that W is such that the subspace $\mathcal{M}\mathrm{or}^{\mathrm{adm}}(W)$, of admissible Morse functions, is non-empty also. It is an important fact that two Morse functions lie in the same path component of $\mathcal{M}\mathrm{or}(W)$ only if they have the same number of critical points of each Morse index. Consequently, the spaces $\mathcal{M}\mathrm{or}(W)$ and $\mathcal{M}\mathrm{or}^{\mathrm{adm}}(W)$ are not path-connected. In fact each has infinitely many path components. The difference between $\mathcal{M}\mathrm{or}(W)$ and $\mathcal{M}\mathrm{or}^{\mathrm{adm}}(W)$ is simply that path components of the former, which contain functions with critical points whose Morse indices are not conducive to Gromov-Lawson surgery, are removed to obtain the latter.

Due to work of Hatcher and Igusa (see [27], [32] and [31]) it is, under reasonable hypotheses on W (assume W is simply connected and has dimension at least six), possible to "connect up" these path components. By this we mean extending the spaces $\mathcal{M}\mathrm{or}(W)$ (and $\mathcal{M}\mathrm{or}^{\mathrm{adm}}(W)$) to obtain path-connected function spaces $\mathcal{GM}\mathrm{or}(W)$ (and $\mathcal{GM}\mathrm{or}^{\mathrm{adm}}(W)$). These path-connected spaces are known respectively as the spaces of *generalised* and *admissible generalised Morse functions* on W and fit into the diagram of inclusions below.

$$\begin{tikzcd}
\mathcal{GM}\mathrm{or}^{\mathrm{adm}}(W) \arrow[r, hook] & \mathcal{GM}\mathrm{or}(W) \\
\mathcal{M}\mathrm{or}^{\mathrm{adm}}(W) \arrow[r, hook] \arrow[u, hook] & \mathcal{M}\mathrm{or}(W) \arrow[u, hook]
\end{tikzcd}$$

A *generalised Morse function* is a smooth function $W \to [0,1]$ which as well as Morse critical points (the ones with non-degenerate Hessian) is allowed to have a certain kind degenerate critical point known as a *birth-death singularity.* Recall we pointed out that near a Morse critical point, the function f took on a "quadratic form." Roughly speaking, a birth-death singularity takes on a cubical form. So, while the function $x \mapsto x^2$ has a Morse critical point at 0, the function $x \mapsto x^3$ has a birth-death critical point at 0. Birth-death singularities are places where certain pairs of regular Morse singularities can cancel along a path through smooth functions called an *unfolding.* A very simple example concerns the family of real-valued functions, $f_t : \mathbb{R} \to \mathbb{R}$ given by the formula

$$f_t(x) = x^3 + tx.$$

When $t < 0$, f_t is Morse with two critical points. When $t > 0$, f_t is Morse with no critical points. The function $f_0(x) = x^3$ is a generalised Morse function, with a lone birth-death singularity at $x = 0$. As t moves from negative to positive the critical points move closer together, collapsing at 0 only to disappear. Thus, from left to right, a death and from right to left, a birth. We see a higher dimensional variation of this in Fig. 13 below, where the projection function on the left hand image moves through a birth-death cancellation in the middle image to obtain the projection function (with no critical points) on the right.

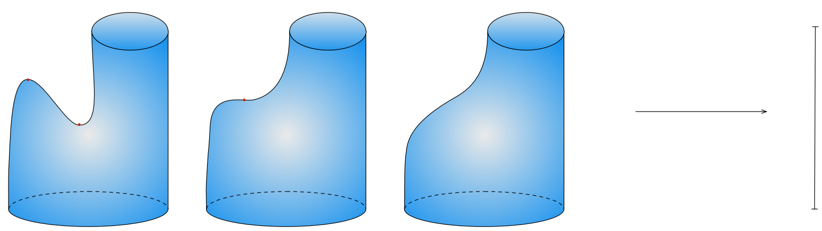


FIGURE 13. The unfolding of a birth-death singularity

We now return to the question of the dependany on the choice of admissible Morse function, f, of a Gromov-Lawson cobordism $\bar{g} = \bar{g}(g_0, f)$ on W. In [49], we make use of results by Hatcher and Igusa, to describe a parameterised version of the Gromov-Lawson cobordism construction which extends the original construction over a birth-death unfolding. In effect, we show that if f_t, with $t \in [0, 1]$, is a path in the space $\mathcal{GM}\mathrm{or}^{\mathrm{adm}}(W)$, connecting two admissible Morse functions on W, there is a corresponding isotopy $\bar{g}_t(g_0, f_t)$ through psc-metrics on W extending the original construction onto generalised Morse functions. In particular, we obtain that $\bar{g}_0 = \bar{g}(g_0, f_0)$ and $\bar{g}_1 = \bar{g}(g_0, f_1)$ are psc-isotopic. This suggests that the psc-isotopy type of the Gromov-Lawson cobordism might be independent of the choice of admissible Morse function. In order to show this of course, one needs to be able to connect with such a path, any arbitrary pair of admissible Morse functions. That is, we need that the space $\mathcal{GM}\mathrm{or}^{\mathrm{adm}}(W)$ be path-connected. Fortunately, there is a powerful theorem of Hatcher, known as the 2-Index Theorem, which sheds considerable light on this issue; see Corollary 1.4, Chapter VI of [32]. The "2-index" in the title refers to the designation of a subspace of the space of generalised Morse functions with upper and lower bounds on the indices of critical points. The theorem itself specifies levels of connectedness for such subspaces, determining that $\mathcal{GM}\mathrm{or}^{\mathrm{adm}}(W)$ is indeed path-connected provided W (along with M_0 and M_1) is simply connected and has dimension at least 6. Thus, under these conditions at least, the Gromov-Lawson cobordism construction is (up to psc-isotopy) independent of the choice of admissible Morse function. Whether or not one can find non-isotopic psc-metrics, by utilising this construction under conditions where $\mathcal{GM}\mathrm{or}^{\mathrm{adm}}(W)$ is not path connected, is an interesting open problem.

One final comment concerns the case when $W = M \times [0, 1]$. Here, a Gromov-Lawson cobordism $\bar{g} = \bar{g}(g_0, f)$ is a psc-concordance of psc-metrics g_0 and $g_1 = \bar{g}|_{M\times\{1\}}$ on M. Such a psc-concordance is known as a *Gromov-Lawson concordance*, a specific case of the more general notion. As a consequence of the construction just described, we see that for simply connected manifolds of dimension $\geqslant 5$, Gromov-Lawson concordant psc-metrics are necessarily psc-isotopic. Although the existence part of this result was later subsumed by Botvinnik's solution of the general psc-concordance problem in [7] and [8], it is worth noting that the method used in [49] involves the construction of an explicit psc-isotopy.

4.5. **Family versions of the Gromov-Lawson construction.** The ability to exhibit multiple path components in the space of psc-metrics, by application of the Gromov-Lawson construction, suggests a role for a parameterised or "family" version of this construction in possibly recognising non-trivial *higher* homotopy classes of psc-metrics. Thus, we would apply the construction to families of psc-metrics (and in the cobordism case, families of admissible Morse functions), with the aim of constructing families of psc-metrics which represent non-trivial elements in the higher homotopy groups of the space of psc-metrics.

An important first step in this regard was taken by Chernysh in [15], who makes use of the fact that the original Gromov-Lawson construction works on a compact family of psc-metrics to prove the following fact: *if M and M' are mutually obtainable from each other by surgeries in codimension $\geqslant 3$, then the spaces $\mathcal{R}^{s>0}(M)$*

and $\mathcal{R}^{s>0}(M')$ *are homotopy equivalent.*[6] The implication of this result is analogous to that of the original Surgery Theorem. It hugely increases our pool of examples. In particular, once we obtain information about the topology of the space of psc-metrics for one manifold M, we now have it for a huge class of manifolds which are related to M by appropriate surgery. It is worth mentioning that, in [52], this author proves an analogue of this result for manifolds with boundary.

The idea behind the proof of Chernysh's theorem is to consider subspaces of $\mathcal{R}^{s>0}(M)$ and $\mathcal{R}^{s>0}(M')$, consisting of psc-metrics which are already "standard" near the respective surgery spheres. These subspaces, denoted respectively $\mathcal{R}^{s>0}_{\mathrm{std}}(M)$ and $\mathcal{R}^{s>0}_{\mathrm{std}}(M')$ are easily seen to be homeomorphic via the act of attaching (or removing) the standard handle on individual psc-metrics. It then suffices to show that the inclusion $\mathcal{R}^{s>0}_{\mathrm{std}}(M) \subset \mathcal{R}^{s>0}(M)$ is a homotopy equivalence. From work of Palais in [43], we know that these spaces are dominated by CW-complexes and so by a famous theorem of Whitehead (Theorem 4.5 of [28]), it is enough to show that the relative homotopy groups $\pi_k(\mathcal{R}^{s>0}(M), \mathcal{R}^{s>0}_{\mathrm{std}}(M)) = 0$ for all k. Essentially this means showing that any continuous map $\gamma : D^k \to \mathcal{R}^{s>0}(M)$, which satisfies the condition that $\gamma|_{\partial D^k}$ maps into $\mathcal{R}^{s>0}_{\mathrm{std}}(M)$, can be continuously adjusted (via homotopy) to a map γ_{std}, whose image is contained entirely inside $\mathcal{R}^{s>0}_{\mathrm{std}}(M)$. The important catch is that, at each stage in the homotopy, the restriction to ∂D^k must always be mapped into $\mathcal{R}^{s>0}_{\mathrm{std}}(M)$. Application of the Gromov-Lawson construction to the family of psc-metrics parameterised by γ can be shown to continuously "move" this family into the standard subspace $\mathcal{R}^{s>0}_{\mathrm{std}}(M)$. Unfortunately, along the way, psc-metrics which are already standard such as those parameterised by $\gamma|_{\partial D^k}$ may be temporarily moved out of $\mathcal{R}^{s>0}_{\mathrm{std}}(M)$. As the damage to these metrics is not too severe (the Gromov-Lawson construction displaying a great deal of symmetry) this problem is solved by replacing $\mathcal{R}^{s>0}_{\mathrm{std}}(M)$ with a larger space of "almost standard" psc-metrics which captures all adjustments made to a standard psc-metric by the Gromov-Lawson

[6] The original proof of Chernysh's result was, mysteriously, never published. In [50], this author provided a shorter version of the proof, using Chernysh's method, but making use of the heavy lifting done in [48] regarding a parameterised Gromov-Lawson construction. Although this paper is rather terse, an extremely detailed version of the proof of this theorem is provided in [52].

construction. As this space is not too much larger or more complicated than $\mathcal{R}^{s>0}_{\text{std}}(M)$, it is then reasonable to show that the spaces of standard and almost standard psc-metrics are homotopy equivalent.

In [49], this author describes another family construction for positive scalar curvature metrics. The idea is to consider a smooth fibre bundle, the fibres of which are diffeomorphic to a cobordism W (of the type described above), over a smooth compact "base" manifold B. The total space of this bundle, denoted E, is equipped with a smooth function $F : E \to B \times [0,1]$, which restricts on each fibre $E_b \cong W$ over $b \in B$, as an admissible Morse function $F_b = F|_{E_b} : E_b \to \{b\} \times [0,1]$. The function F is known as a *fibrewise admissible Morse function* and is depicted schematically in Fig. 14 below. This figure also includes critical points which, in the picture, form 1-dimensional closed curves in the total space E as do their images, the critical values, in $B \times [0,1]$. Of course in practice the set of critical points (when non-empty) will have dimension the same as the base manifold, B. As such schematic pictures are limited (especially in dimension), we have depicted this bundle as if it were a trivial product bundle $E \cong B \times W$. In practice however, the bundle need not be trivial.

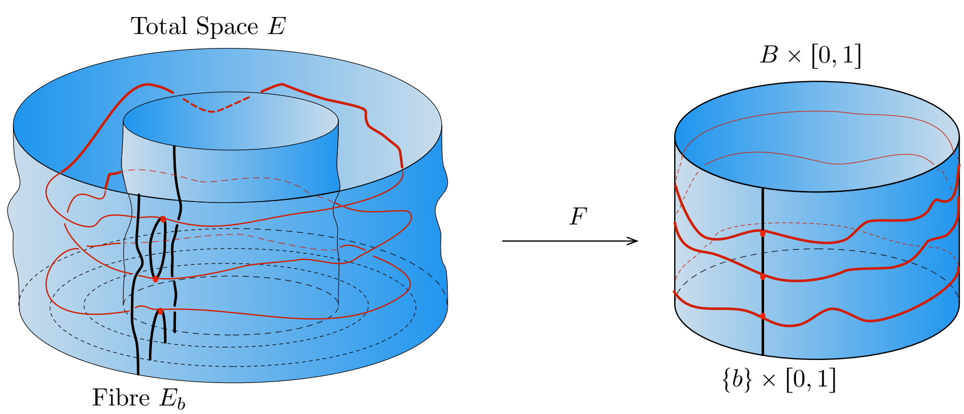


FIGURE 14. The fibrewise admissible Morse function F

Suppose we have a smooth family of psc-metrics $B \to \mathcal{R}^{s>0}(M_0)$, $b \mapsto g_0(b)$. It is then possible to construct a metric $\bar{G}$ on the total space E of the bundle which, for each $b \in B$, restricts on the fibre E_b as the Gromov-Lawson cobordism metric associated to the psc-metric $g_0(b)$ and the admissible Morse function F_b. Thus, $\bar{G}$ can be used to represent a continuous family of psc-metrics on W and by appropriate restriction, on M_1. One important point here is that, as the bundle E may be non-trivial, the fibres are only diffeomorphic to W, and not canonically so. Thus, in a sense this method is more conducive to obtaining families of psc-metrics on the moduli-spaces of psc-metrics on W and M_1. That said, a little later we will consider applying this construction to a bundle whose total space, near its boundary, takes the form of a product $M_i \times [0, \epsilon) \times B$ with $i \in \{0, 1\}$, even though the bundle itself is not trivial. This means that the metric $\bar{G}$ obtained from this construction determines, by restriction, a family of psc-metrics which unambiguously lies in $\mathcal{R}^{s>0}(M_1)$. For now, let us consider an application to the moduli space of psc-metrics.

In [10], Botvinnik, Hanke, Schick and this author exhibit non-triviality in the higher homotopy groups of the moduli space of psc-metrics, $\mathcal{M}^{s>0}(M)$, for certain manifolds M, using this construction. Initially, we work with a variation of the moduli space which is worth discussing. Recall that $\mathcal{M}(M)$, is obtained from the space of Riemannian metrics on M, as a quotient of the pull back action of $\mathrm{Diff}(M)$. We now consider a certain subgroup of $\mathrm{Diff}(M)$, denoted $\mathrm{Diff}_{x_0}(M)$ where $x_0 \in M$ is a fixed base point. This is the subgroup of diffeomorphisms which fix x_0 and whose derivative at x_0 is the identity map on $T_{x_0}M$. Thus, elements of $\mathrm{Diff}_{x_0}(M)$ all leave the point x_0 and directions emanating from this point unaltered. This point can be thought of as a sort of "observer point" on the manifold. After restricting the pull back action to this subgroup of observer respecting diffeomorphisms, we obtain $\mathcal{M}_{x_0}(M) = \mathcal{R}(M)/\mathrm{Diff}_{x_0}(M)$, the *observer moduli space of Riemannian metrics on M*. By replacing $\mathcal{R}(M)$, with $\mathcal{R}^{s>0}(M)$ (or $\mathcal{R}^{\mathrm{Ric}>0}$, $\mathcal{R}^{\mathrm{Sec}>0}(M)$), we obtain the *observer moduli space of Riemannian metrics of positive scalar (Ricci, sectional) curvature*, denoted $\mathcal{M}^{s>0}_{x_0}(M)$ ($\mathcal{M}^{\mathrm{Ric}>0}_{x_0}(M), \mathcal{M}^{\mathrm{Sec}>0}_{x_0}(M)$). Regarding this space, the main result of [10], is stated below.

Theorem 4.3. [10] *For any $k \in \mathbb{N}$, there is an integer $N(k)$ such that for all odd $n > N(k)$, and all manifolds M admitting a psc-metric, g, the group $\pi_i(\mathcal{M}^{s>0}_{x_0}(M^n), [g])$ is non-trivial when $i \leqslant 4k$ and $i \equiv 0 \mod 4$.*

In discussing the proof of this result it is important to realise that, unlike $\mathrm{Diff}(M)$, the subgroup $\mathrm{Diff}_{x_0}(M)$ acts freely on $\mathcal{R}(M)$. This gives us some useful topological information about the resulting moduli space $\mathcal{M}_{x_0}(M)$. Consider first the case when M is the sphere, S^n. We have the following calculation, due to Farrell and Hsiang in [19], which helps explain some of the hypotheses of the theorem above. For any $k \in \mathbb{N}$, there is an integer $N(k)$ such that for all odd $n > N(k)$,

$$\pi_i(\mathcal{M}_{x_0}(S^n)) \otimes \mathbb{Q} \cong \begin{cases} \mathbb{Q} & \text{if } i \equiv 0 \mod 4, \\ 0 & \text{otherwise.} \end{cases}$$

Thus, for appropriate i and n, we now have lots of non-trivial elements in the groups $\pi_i(\mathcal{M}_{x_0}(S^n))$ to work with. So how does this calculation help us? We have the inclusion

$$\mathcal{M}_{x_0}^{s>0}(S^n) \hookrightarrow \mathcal{M}_{x_0}(S^n),$$

which induces homomorphisms of rational homotopy groups

$$\pi_i(\mathcal{M}_{x_0}^{s>0}(S^n)) \otimes \mathbb{Q} \longrightarrow \pi_i(\mathcal{M}_{x_0}(S^n)) \otimes \mathbb{Q}.$$

If, for some $i = 4m$, we can show that some of the non-trivial elements in Farrell and Hsiang's calculation are in the image of such a homomorphism, then we have exhibited non-triviality in the observer moduli space by way of non-trivial elements in $\pi_{4m}(\mathcal{M}_{x_0}^{s>0}(S^n)$. Note that the dimension n (which is required to be odd) may be large.

In [10], we show that all of the Farrell-Hsiang elements are in the image of such a homomorphism. The idea is as follows. From work of Hatcher, we know that every element of $\pi_i(\mathcal{M}_{x_0}(S^n)) \otimes \mathbb{Q}$ determines a specific S^n bundle over S^i. These "Hatcher bundles" come naturally equipped with fibrewise admissible Morse functions on their total spaces; a comprehensive description is given by Götte in [24]. The fact that the fibres are spheres (and not manifolds with boundary) is not a problem here. One defines the fibrewise Morse function first on a bundle of "southern" hemispherical disks with global minima on the south poles. On the remainder of the disk, the function has a pair of "canceling" critical points; see Fig. 15 for a depiction of the gradient flow of such a function on the disk.

We then form a fibrewise "doubling" of this disk bundle and its fibrewise Morse function to obtain a sphere bundle, the Hatcher bundle. Away from the polar maxima and mimina, each fibre function has four remaining critical points with appropriate cancellation properties. Regarding the metric construction we equip, in a fibrewise sense, a neighbourhood of each south pole with a "torpedo" metric, before using the Gromov-Lawson cobordism construction to extend past the critical points to a family of psc-metrics on the

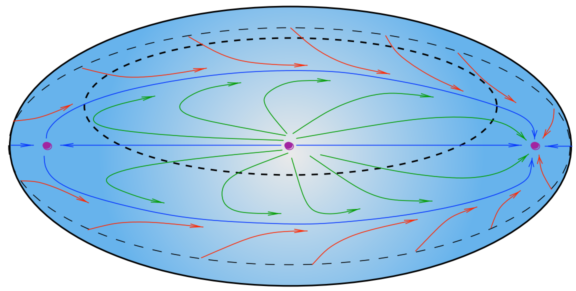


FIGURE 15. The gradient flow of the restriction of the Hatcher fibrewise Morse function to one hemisphere of one spherical fibre

hemispherical disk bundle. This is then "doubled" to obtain the desired fibrewise psc-metric on the total space of the Hatcher bundle, representing a "lift" of a non-trivial element of $\pi_i(\mathcal{M}_{x_0}(S^n))$ to a necessarily non-trivial element of $\pi_i(\mathcal{M}_{x_0}^{s>0}(S^n))$. This proves Theorem 4.3 in the case when M is the sphere. To complete the proof, the authors show that taking a fibrewise connected sum of such a representative bundle with an n-dimensional manifold M (equipped with a psc-metric), gives rise to non-trivial elements in $\pi_i(\mathcal{M}_{x_0}^{s>0}(M))$.

One important observation about the above result is that the technique used to prove it does not make use of the index obstruction α. Indeed, the method applies to both spin and non-spin manifolds, provided they admit psc-metrics and satisfy the appropriate dimension requirements. As far as we are aware, this is the first result displaying such non-triviality in any space of psc-metrics which does this. Returning to the regular moduli-space, we have the following. Using some deep arguments from algebraic topology (which we will not discuss here), the authors show that for certain suitable manifolds M, this non-triviality carries over to the traditional moduli space $\mathcal{M}^{s>0}(M)$. Interestingly, these suitable odd-dimensional manifolds are all *non-spin*.

We close this section with a brief discussion of a useful strengthening of this family Gromov-Lawson cobordism construction. Recall our earlier discussion on *generalised* Morse functions. We now reconsider the earlier smooth fibre bundle, with fibres diffeomorphic to a cobordism W, over a smooth compact "base" manifold B. The total space of this bundle, denoted E, is now equipped with a smooth function $F : E \to B \times [0,1]$, which restricts on each fibre $E_b \cong W$ over $b \in B$, as an admissible *generalised* Morse function $F_b = F|_{E_b} : E_b \to \{b\} \times [0,1]$. The function F is known as a *fibrewise admissible generalised Morse function* and is depicted schematically in Fig. 16 below. This allows for a variation in the numbers of critical points on fibres as certain Morse critical points may now cancel. In this figure, which suppresses the critical points depicted in Fig. 14, the fibres E_a, E_b and E_c depict stages of the unfolding of a birth-death critical point, with two critical points at E_a canceling at E_b. We now consider an application.

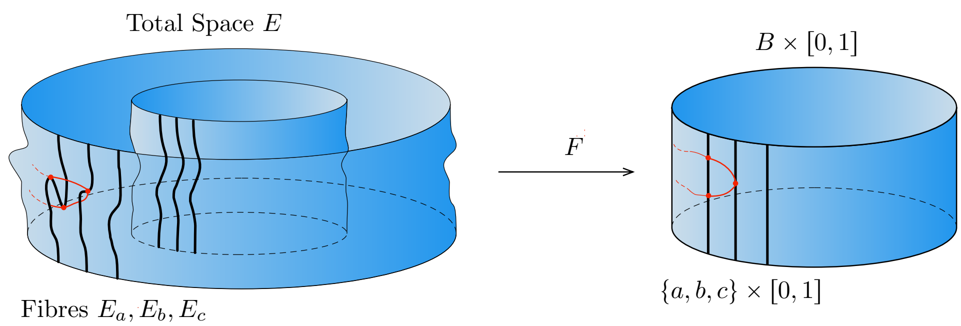


FIGURE 16. The fibrewise admissible generalised Morse function F

In [26], Hanke, Schick and Steimle construct a rather fascinating collection of objects. Recall, for any $4k$-dimensional spin manifold, we can associate a number $\hat{A}(M) \in \mathbb{Z}$, the so-called $\hat{A}$-genus. It is well known that this topological invariant, which depends only on the cobordism type of M, is multiplicative. That is, given a product of manifolds $M \times N$, we know that $\hat{A}(M \times N) = \hat{A}(M)\hat{A}(N)$. What was not clear was whether this multiplicitivity held in manifold bundles. A fibre bundle (unless it is trivial) is of course a "twisted product" and, like the Möbius band, only locally behaves like a regular product. In [26], the authors demonstrate that there exist certain manifold bundles, the fibres and base manifolds of which are compact spin manifolds but where the $\hat{A}$-genus of the total space is **not** the product of the $\hat{A}$-genera of the fibre and base. In particular, they show that there are bundles over the sphere, the total space of which is a spin manifold with non-zero $\hat{A}$-genus, but with fibres having $\hat{A} = 0$. Thus, we have a spin manifold with non-zero $\hat{A}$-genus which decomposes as a "twisted product" of spin manifolds each with zero $\hat{A}$-genus!

This allows for an intriguing analogue of the Bott manifold construction we performed earlier. The authors show that one can construct bundles of the following form. Each bundle has base manifold a sphere, S^k, and fibre a cobordism of n-dimensional spheres which we denote S_0 and S_1 to distinguish the bottom and top ends of the cobordism. There are certain dimension requirements on n and k which we will ignore. The bundle also has the property that, near its boundary, the total space E has a product structure. Thus, the total space has a well-defined bottom and top which take the form $S_i \times [0, \epsilon) \times S^k$ with $i \in \{0, 1\}$. There is thus an obvious way of capping off the ends of the total space with appropriate disk products, $D_i \times S^k$, where D_0 and D_1 are $(n+1)$-dimensional hemispheres, to form a closed manifold $\bar{E}$, from the total space. Crucially, the closed spin manifold $\bar{E}$ is such that $\hat{A}(\bar{E}) \neq 0$.

Now, the fibres of this bundle are such that the total space admits a fibrewise admissible *generalised* Morse function. Thus, given a family of psc-metrics on S_0, parameterised by the base manifold S^k, we can apply the family Gromov-Lawson cobordism construction (strengthened for generalised Morse functions) above to obtain a psc-metric $\bar{G}$ on the total space E, which restricts to a family of psc-metrics on S_1 also parameterised by S^k. For simplicity, let us assume that the family of psc-metrics on S_0 is trivial, i.e. is constantly the round metric. Thus, at the lower end of the total space E we have a standard cylindrical product of round metrics which easily extends as a positive scalar curvature metric on the "southern cap" of $\bar{E}$. As with the Bott manifold example earlier, we know that no such extension is possible at the northern cap of $\bar{E}$. This is because $\hat{A}(\bar{E}) \neq 0$ and so $\bar{E}$ admits no psc-metrics. Hence, the family of psc-metrics obtained on the sphere S_1 is homotopically distinct from the trivial one on S_0 and thus constitutes a non-trivial homotopy class in $\pi_k(\mathcal{R}^{s>0}(S^n))$. The authors go on to obtain a number of very interesting results concerning non-triviality in both the space of psc-metrics and its moduli space for certain manifolds. In particular, they prove the following theorem.

Theorem 4.4. (Hanke, Schick, Steimle [26]) *Given $k \in \mathbb{N} \cup \{0\}$, there is a natural number $N(k)$ such that for all $n \geqslant N(k)$ and* *each spin manifold* M^{4n-k+1} *admitting a psc-metric* g_0*, the homotopy group* $\pi_k(\mathcal{R}^{s>0}(M); g_0)$ *contains elements of infinite order when* $k \geqslant 1$ *and infinitely many distinct elements when* $k = 0$*.*

To distinguish some of the non-trivial elements constructed in this paper from those constructed by Hitchin in [29] or by Crowley and Schick in [17] (all of which become trivial in the moduli space), the authors introduce the notion of "geometrically significant" elements. Essentially, elements of $\pi_k(\mathcal{R}^{s>0}(M); g_0)$ are *not* geometrically significant if they can be obtained from a single fixed psc-metric on M via pull-back over an S^k-parameterised family of oriented diffeomorphisms $M \to M$. Otherwise such an element is *geometrically significant.* Obviously, non-geometrically significant elements become trivial in the homotopy groups of the moduli space $\mathcal{M}^{s>0}(M)$. The authors go on to show that many of the elements they construct are in fact geometrically significant. In particular, in the case when the manifold M satisfies the additional hypothesis of being the fibre of an oriented fibre bundle over the sphere S^{k+1} whose total space has vanishing $\hat{A}$-genus, then the groups $\pi_k(\mathcal{M}^{s>0}(M); [g_0])$ also contain elements of infinite order.

In closing this section, we should mention a very significant recent result of Nathan Perlmutter concerning spaces of Morse functions. In [44], he constructs something called a *cobordism category* for Morse functions. We will not attempt to define the term cobordism category here except to say that it (and in particular its associated classifying space) allows us to view the set all manifolds of a particular dimension as a single "space of manifolds" and from this distill aspects of the structure which are stable under certain operations such as surgery. This idea, which is still relatively new, was developed by Galatius, Madsen, Tillmann and Weiss in [22]. Here the authors use it to provide a new proof of a famous problem called the Mumford Conjecture, following the original proof by Madsen and Weiss in [39].

Various versions of cobordism category exist which deal not simply with manifolds, but with manifolds equipped with extra structure, such as Riemannian metrics, complex or symplectic structures. In Perlmutter's case, he considers a category which, very roughly, has as objects n-dimensional manifolds (embedded in a dimensionally large Euclidan space) and as morphisms, $(n+1)$-dimensional manifolds with boundary which form cobordisms between objects.

Importantly these cobordisms come with a Morse function which arises as the projection, onto a fixed axis, of an embedding of the cobordism in Euclidean space (such as that depicted in Fig. 12). Perlmutter's results extend work of Madsen and Weiss in [39] on understanding the homotopy type of the classifying space of this category, to deal with subcategories where the Morse functions have bounds placed on the indices of their critical points. In particular, Perlmutter sheds a great deal of light on the case of *admissible* Morse functions. This is something which, as Perlmutter explains in his paper, has great significance for positive scalar curvature, particularly in the construction and analysis of a positive scalar curvature cobordism category.

4.6. **H-Space and Loop Space Structure.** We close this section with a very brief discussion regarding another aspect of the topology of the space of positive scalar curvature metrics. Up to now, our search for topological information has essentially meant a search for non-trivial elements of the homotopy groups of the space. We should mention that, although we have not discussed the homology or cohomology of spaces of psc-metrics, there is certainly non-trivial topological information there also, much of it following from that found in the homotopy groups. Indeed, in the paper by Hanke, Schick and Steimle above, [26], the authors explicitly show that the infinite order elements they construct in $\pi_k(\mathcal{R}^{s>0}(M); g_0)$ have infinite order images in the corresponding homology groups, under the Hurewicz homomorphism: $\pi_k(\mathcal{R}^{s>0}(M); g_0) \to H_k(\mathcal{R}^{s>0}(M))$. We now examine the space of psc-metrics for a layer of structure which has substantial homotopy theoretic implications. This concerns the question of whether or not $\mathcal{R}^{s>0}(M)$ admits a multiplicative *H-space* structure or, more significantly, whether this space has the structure of a *loop space.*

We begin with multiplication. A topological space, Z, is an H-*space* if Z is equipped with a continuous multiplication map $\mu : Z \times Z \to Z$ and an identity element $e \in Z$ so that the maps from Z to Z given by $x \mapsto \mu(x, e)$ and $x \mapsto \mu(e, x)$ are both homotopy equivalent to the identity map $x \mapsto x$.[7] An H-space Z is said to be homotopy commutative if the maps μ and $\mu \circ \omega$, where $\omega : Z \times Z \to Z \times Z$ is the "flip" map defined by $\omega(x, y) = (y, x)$, are homotopy

[7] There are stronger versions of this definition (see section 3.C. of [28]), however all of these versions coincide with regard to the spaces we will consider.

equivalent. Finally, Z is a homotopy associative H-space if the maps from $Z \times Z \times Z$ to Z given by $(x, y, z) \mapsto \mu(\mu(x, y), z)$ and $(x, y, z) \mapsto \mu(x, \mu(y, z))$ are homotopy equivalent. The condition that a topological space is an H-space has various implications for its homotopy type and specifically in homology; see section 3.C. of [28]. For example, H-space multiplication gives the cohomology ring the structure of an algebra, a so-called *Hopf algebra.* Another consequence is that the fundamental group of an H-space is always abelian. Hence, it is worth investigating if such structure can be found in our spaces of psc-metrics.

Recall Corollary 3.3of the Surgery Theorem. Manifolds of dimension $\geqslant 3$ admitting psc-metrics may be combined by the process of connected sum to obtain new manifolds admitting psc-metrics. Thus, we have a "geometric connected sum" construction as opposed to a purely topological one. Consider now a pair of psc-metrics on the sphere, S^n, with $n \geqslant 3$. Given that a connected sum of spheres is still a sphere, we now have a way of combining these psc-metrics, via this geometric connected sum, to obtain a new psc-metric on the sphere. This suggests a possible multiplicative structure on the space of psc-metrics on S^n. There are a number technical issues to overcome. In particular, the various choices involved in the connected sum construction mean that, as it stands, this operation is far from well-defined. However, it is shown by this author in [51] that a careful refinement of this geometric connected sum construction leads to the following result.

Theorem 4.5. [51] *When $n \geqslant 3$, the space $\mathcal{R}^{s>0}(S^n)$ is homotopy equivalent to a homotopy commutative, homotopy associative H-space.*

The following corollary follows from a standard fact about H-spaces (mentioned above) when $n \geqslant 3$ and from the fact that $\mathcal{R}^{s>0}(S^n)$ is contractible when $n = 2$.

Corollary 4.6. [51] *When $n \geqslant 2$, the space $\mathcal{R}^{s>0}(S^n)$ has abelian fundamental group.*

The idea behind the proof of Theorem 4.5 is to replace $\mathcal{R}^{s>0}(S^n)$ with a particular subspace, $\mathcal{R}^{s>0}_{\text{std}}(S^n)$, of metrics which take the form of a standard torpedo metric near a fixed base point (the north pole say). As discussed earlier in the proof of Chernysh's theorem, the inclusion $\mathcal{R}^{s>0}_{\text{std}}(S^n) \subset \mathcal{R}^{s>0}(S^n)$ is a homotopy equivalence and so it suffices to work with the subspace. At least now there seems to be a well-defined way of taking connected sums: just remove the caps near the north poles of two such standard psc-metrics and glue. Of course the problem here is that the resulting psc-metric no longer has a standard torpedo at its north pole. We get around this problem by using an intermediary "tripod" metric, as shown in Fig. 17. This metric, as well as having a north pole torpedo, also has two identical torpedo attachments in its southern hemisphere. These southern torpedoes are used to connect distinct psc-metrics in the space, while the northern torpedo ensures that the multiplication is closed. The reader should view Fig. 17 as an equation with the highlighted tripod metric playing the role of a multiplication sign. Verifying that this multiplication is homotopy commutative is not so difficult, the homotopy in question being a simple rotation. Ensuring homotopy associativity is a little more complicated and involves a very careful sequence of geometric manouvres; details can be found in [51].

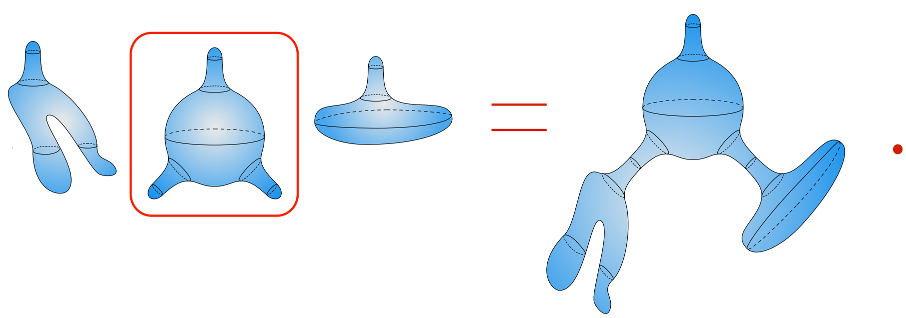


Figure 17. Mulitplying two metrics in $\mathcal{R}^{s>0}_{\mathrm{std}}(S^n)$

There is another level of structure, one which goes deeper than H-space structure, which we now consider. Given a topological space Y with a fixed base point $y_0 \in Y$, we define the *loop space of* Y denoted ΩY, as the space of all continuous maps $\gamma : [0,1] \to Y$ so that $\gamma(0) = \gamma(1) = y_0$. Repeated application of this construction yields the k-th iterated loop space $\Omega^k Y = \Omega(\Omega \cdots (\Omega Y))$ where at each stage the new base point is simply the constant loop at the old base point. It is a fact that every loop space is also an H-space with the multiplication determined by concatenation of loops. However the condition of being a loop space is stronger and an H-space may not even be homotopy equivalent to a loop space. As before, such structure has important topological consequences and knowing for example that a topological space Z has the homotopy type of a loop space (that is Z is homotopy equivalent to ΩY for some topological space Y) or better yet an iterated loop space is very helpful in understanding homotopy type.

Bearing in mind the case of $\mathcal{R}^{s>0}(S^n)$, we consider the problem of how to tell if a given H-space has the structure of a loop space. In general, this is a complicated problem concerning certain "coherence" conditions on the homotopy associativity of the multiplication. Roughly speaking, given an H-space Z, we compare the different maps obtainable from a k-fold product, Z^k, to Z by prioritizing (with appropriate brackets) the multiplication of components. For example, in the case when $k = 4$, two such maps are given by:

$$(a, b, c, d) \mapsto (ab)(cd) \quad \text{and} \quad (\mathrm{a}, \mathrm{b}, \mathrm{c}, \mathrm{d}) \mapsto ((\mathrm{ab})\mathrm{c})\mathrm{d},$$

with H-space multiplication now denoted by juxtaposition. These maps may not agree on the nose, but we would like that they are at least homotopic. In the case when $k = 3$, this is precisely the homotopy associativity condition defined above. Returning to the case when $k = 4$, there are 5 maps to consider. These can be denoted in specific fashion by vertices of a pentagonal polyhedron P; see Fig. 18. Suppose we can specify a map

$$P \times Z^4 \longrightarrow Z,$$

which restricts as these various "rebracketing" maps on the vertices and, on the edges, specifies homotopies between specific vertex maps. We then say that Z is an A_4-space. More generally, for each k there is a corresponding polyhedron which describes these higher associativity relations, leading to the notion of an A_k space. This approach was developed by James Stasheff and so these shapes are known as *Stasheff polyhedra* (or sometimes as "associahedra"). A space which is A_k for all k is known as an A_∞-space. It is a theorem of Stasheff that a space is an A_∞-space, if and only if it is a loop space; see Theorem 4.18 in [40].

When it comes to deciding whether or not a space is an iterated loop space, these coherence conditions are more efficiently described using the notion of an *operad*. An operad is not something we will define here except to say that it is a collection of topological spaces with combinatorial data. Operads (potentially) act on topological spaces in a way which captures at a deeper level, the sort of associativity information described above. The idea of an operad arose out

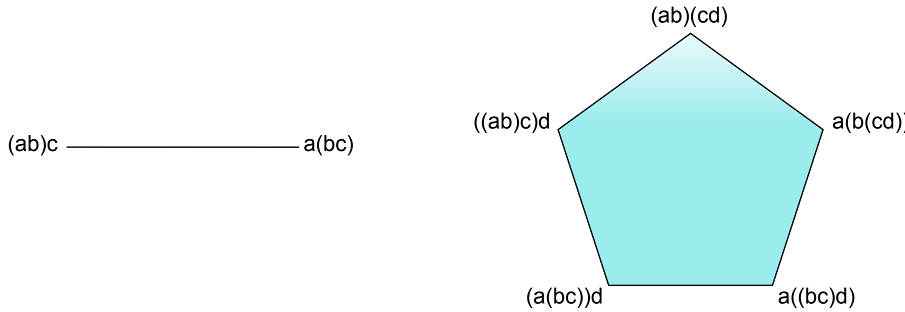


Figure 18. Stasheff polyheda for threefold and fourfold multiplication

of work by Boardman, Vogt and May on the problem of recognising whether a given topological space is an iterated loop space and has since seen wide application in various areas including graph theory and theoretical physics; see [40] for a comprehensive guide. Using their so-called *Recognition Principle* (Theorem 13.1 from [41]), this author was able to prove the following strengthening of Theorem 4.5 above.

Theorem 4.7. [51] *When $n \geqslant 3$, $\mathcal{R}^{s>0}(S^n)$ is weakly homotopy equivalent to an n-fold loop space.*

The proof involves demonstrating an action on $\mathcal{R}^{s>0}(S^n)$ of a certain operad known as the *operad of n-dimensional little disks* and makes use of something called the "bar-construction" on operads and in particular Theorem 4.37 of [6]. Roughly, the bar construction allows us to replace the original operad with something which is a little more "flexible" regarding the various geometric manoeuvres on metrics needed to satisfy the associativity conditions required by the operad action. One important point is that, although this works fine for path connected spaces, $\mathcal{R}^{s>0}(S^n)$ is usually not path connected. Thus, initially our result held only for the path component of $\mathcal{R}^{s>0}(S^n)$ containing the round metric. There is however, a way around this. The theorem holds for all of $\mathcal{R}^{s>0}(S^n)$ provided the operad action induces a group structure on the set of path components, $\pi_0(\mathcal{R}^{s>0}(S^n))$; see Theorem 13.1 of [41]. In the end, the proof that such a group structure existed (specifically that elements had inverses) turned out to be a substantially difficult problem. It is true however, but requires as heavyweight a result as the Concordance Theorem of Botvinnik, Theorem 4.1 above.

5. Some Related Work

As we have previously stated, positive scalar curvature is just one of many curvature constraints we might consider. It is important to realise that a good deal of work has been done on understanding the topology of spaces (and moduli spaces) of metrics which satisfy other geometric constraints. We will finish this article by taking a brief look at some of the results concerning these other spaces, before closing with some words on a recent and highly significant result by Sebastian Hoelzel concerning surgery. Much of what we discuss below is drawn from Tuschmann and Wraith's extremely useful book on moduli spaces of Riemannian metrics, [47].

5.1. **Spaces of metrics satisfying other curvature constraints.**

Early in this article we discussed the general notion of a geometric (in particular curvature) constraint C on Riemannian metrics on a smooth manifold M, leading to a subspace, $\mathcal{R}^C(M) \subset \mathcal{R}(M)$, of Riemannian metrics which satisfy C. So far we have only considered the case when C is the condition of having positive scalar curvature. So what about other curvature constraints?

Without even leaving the scalar curvature, one might consider for example the space, $\mathcal{R}^{s<0}(M)$, of all Riemannian metrics of negative scalar curvature on a smooth manifold M. Temporarily, we will drop the compactness assumption on M and allow that M may be compact or not. We know from a previously mentioned result of Lohkamp [38], that for any such M with dimension $n \geqslant 3$, the space $\mathcal{R}^{s<0}(M)$ is non-empty. From the Gauss-Bonnet theorem we know that there are 2-dimensional closed manifolds for which this does not apply: the sphere, projective plane, torus and Klein bottle. Lohkamp goes on to prove a great deal more. In particular, he proves that these non-empty spaces are actually all contractible! Even more amazingly, these results hold just as well if we replace negative scalar with negative *Ricci* curvature and the same is also true of the corresponding negative scalar and negative Ricci curvature moduli spaces. The proofs behind these facts are highly nontrivial but at their geometric heart is the fact that negative Ricci and scalar curvature display a great deal more flexiblity regarding local metric adjustment than do their positive counterparts.

Continuing with the theme of negative curvature, and given the comprehensive results by Lohkamp in the scalar and Ricci case, it remains to consider negative sectional curvature. This case is a little more interesting. In the case when M is a 2-dimensional surface, the story is intimately connected with Teichmüller theory. Without opening a discussion of this subject here, it can be shown that the space, $\mathcal{R}^{\mathrm{Sec}<0}(M)$, of negative sectional curvature metrics on a closed oriented surface, M, with genus at least 2, is homotopy equivalent to an object called the *Teichmüller space* of M, denoted $\mathcal{T}(M)$. This space, which is a complex manifold obtained as a certain quotient of the space of complex differentiable structures on M, is known to be contractible; see section 9 of [47]. In higher dimensions, the story is a little different and requires an analogue of $\mathcal{T}(M)$ known as the *Teichmüller space of Riemannian metrics* on M. There are a number of recent results on this subject, due to Farrell and Ontaneda (see for example [20]), which show a multiplicity of path components and non-triviality in the higher homotopy groups of the space $\mathcal{R}^{\mathrm{Sec}<0}(M)$ (and its moduli space), for certain closed manifolds admitting hyperbolic metrics. Typically, the dimension of these manifolds is at least 10.

Another interesting case concerns that of the positive Ricci curvature. The earliest result concerning topological non-triviality in spaces of such metrics concerns the moduli space of positive Ricci curvature metrics. It is due to Kreck and Stolz and utilises their s-invariant. Recall that, under reasonable circumstances, the s-invariant $s(M,g) \in \mathbb{Q}$ is an invariant of the path component of $[g] \in \mathcal{M}^{s>0}(M)$. Thus, it suffices to find a manifold M which has metrics of positive Ricci curvature (such metrics necessarily have positive scalar curvature) with distinct s-values to demonstrate that $\mathcal{M}^{\mathrm{Ric}>0}(M)$ is not path-connected. In [36], Kreck and Stolz show that there exist closed 7-dimensional manifolds for which the moduli space of positive Ricci curvature metrics has in fact infinitely many path components. The manifolds themselves are part of a class of examples, constructed by Wang and Ziller in [53], of bundles over the manifold $\mathbb{C}P^2 \times \mathbb{C}P^1$ with fibre S^1. Each such manifold admits infinitely many Einstein metrics (metrics of constant Ricci curvature) with positive Einstein constant (thus *positive* constant Ricci curvature). Importantly, the authors compute that for each of these manifolds, every element of this infinite collection of positive Ricci curvature metrics has a different s-invariant.

A further application of the s-invariant was used by Wraith in [58] to extend Carr's theorem on the space of psc-metrics on spheres of dimension $4k-1$, with $k \geqslant 2$, to the positive Ricci case. In particular, Wraith's result applies not just to standard spheres but to certain exotic spheres also. In dimensions $4k-1$ with $k \geqslant 2$, Wraith considers certain collections, denoted bP_{4k}, of smooth spheres which are topologically the same as but not necessarily diffeomorphic to the standard sphere. These spheres are the boundaries of certain $4k$-dimensional manifolds which are *parallelisable* (have trivial tangent bundle), constructed via a process called *plumbing* (a topological construction with certain similarities to surgery). Building on some of his earlier constructive results concerning the existence of positive Ricci curvature metrics in [55], [56] and [57], Wraith proves the following.

Theorem 5.1. (Wraith [58]) *For any sphere $\Sigma^{4k-1} \in bP_{4k}$, with $k \geqslant 2$, each of the spaces $\mathcal{R}^{\mathrm{Ric}>0}(\Sigma^{4k-1})$ and $\mathcal{M}^{\mathrm{Ric}>0}(\Sigma^{4k-1})$ has infinitely many path components.*

Regarding positive Ricci curvature, we should finally mention one very recent result concerning the observer moduli space of positive Ricci curvature metrics. Botvinnik, Wraith and this author have recently proved a positive Ricci version of Theorem 4.3, in the case when the underlying manifold M is the sphere S^n; see [11]. Thus, in appropriate dimensions, the space $\mathcal{M}_{x_0}^{\mathrm{Ric}>0}(S^n)$ has many non-trivial higher homotopy groups. As yet, it is unclear how this result might be extended to other manifolds or to the regular moduli space. The proof works for essentially the same topological reasons as the original, however the geometric construction is very different and relies on certain gluing results of Perelman.

There are a number of results concerning positive and also non-negative sectional curvature for closed manifolds. Regarding positive sectional curvature, Kreck and Stolz in [36] have demonstrated that for a certain class of closed 7-dimensional manifolds known as Aloff-Wallach spaces (see [1] for a description), the corresponding moduli spaces of positive sectional curvature metrics are not-path connected. The strategy here, which again makes use of the s-invariant, is similar to that used in the positive Ricci case for the Wang-Ziller examples described above. Indeed, regarding this positive Ricci result, it was later shown by Kapovitch, Petrunin and Tuschmann in [34] that there are closed 7-dimensional manifolds (of the class constructed by Wang and Ziller) for which the moduli space of *non-negative* sectional curvature metrics has infinitely many path components. A more recent example can be found in [18]. We should also point out that there are a number of interesting results concerning topological non-triviality in the moduli space of non-negative sectional curvature metrics for certain open manifolds; see in particular work by Belegradek, Kwasik and Schultz [5], Belegradek and Hu [4] and very recently Belegradek, Farrell and Kapovitch [3].

5.2. **Extending the Surgery Theorem to other curvatures.** Let us step back once more to consider the general problem of understanding the topology of a space $\mathcal{R}^C(M)$ of Riemannian metrics on a smooth manifold M which satisfy a given curvature constraint C. One of the key tools in understanding this problem in the case where C is positive scalar cuvature, is surgery and, in particular, the Surgery Theorem of Gromov-Lawson and Schoen-Yau. Typically, stronger curvature notions do not behave as well under surgery. The Surgery Theorem as currently stated is simply false if we replace positive scalar curvature with positive Ricci or sectional curvatures. However, this does not mean that there are not certain circumstances under which a curvature condition might be preserved by appropriate surgeries. Indeed, given the power of surgery in constructing examples, understanding how the Surgery Theorem might extend for other curvature notions is an obvious priority.

One example of such an extension is due to Labbi in [37] and concerns a notion called p-curvature. Given a smooth Riemannian manifold M of dimension n and an integer p satisfying $0 \leqslant p \leqslant n-2$, there is a type of curvature s_p defined as follows. For any $x \in M$ and any p-dimensional plane $V \subset T_xM$, we define the p curvature $s_p(x, V)$ to be the scalar curvature at x of the *locally* specified $(n-p)$-dimensional submanifold determined by $V^\perp$. Thus, when $p = n-2$, $V^\perp$ is a 2-dimensional subspace of T_xM and so this is precisely the definition of the sectional curvature. When $p = 0$, $V = \{0\} \in T_xM$ implying that $V^\perp = T_xM$ and $s_0(x, \{0\}) = s(x)$, the scalar curvature at x. The p-curvatures therefore, are a collection of increasingly stronger curvature notions from scalar (when $p = 0$) all the way up to sectional (when $p = n-2$). The Ricci curvature does not appear exactly as one of the p-curvatures but can be described in an equation involving the scalar curvature, s_0, and s_1.[8] In [37], by a careful analysis of the the construction of Gromov and Lawson in [25], Labbi proves the following extension of the Surgery Theorem.

Theorem 5.2. (Labbi [37]) *Let M and M' be smooth manifolds of dimension n with M' obtained from M by a surgery in codimension $\geqslant 3+p$, where $0 \leqslant p \leqslant n-2$. Then if M admits a Riemannian metric of positive p-curvature, so does M'.*

Later, an analogous extension of the Surgery Theorem with appropriate codimension hypotheses was worked out by Wolfson in [54] in regard to a curvature constraint called *positive k-Ricci curvature.* As these results share a common origin in the work of Gromov-Lawson and Schoen-Yau, a natural idea would be to extrapolate the general principle and find a theorem which subsumes all of these individual cases. Precisely this was done in a remarkable paper by Sebastian Hoelzel; see [30]. Hoelzel defines the idea of a curvature condition C which is stable under surgeries of a certain codimension and proves a comprehensive generalisation of the original Surgery Theorem covering all the previous curvature extensions and many more. He further goes on to prove an extension of the *classification of simply connected manifolds of positive scalar curvature,* Theorem 3.4above, to this general setting. In better understanding spaces of metrics which satisfy a curvature constraint C, an especially nice next step would be a "family" or paramaterised version of Hoelzel's theorem.

References

[1] S. Aloff and N. Wallach, *An infinite family of distinct 7-manifolds admitting positively curved Riemannian structures,* Bull. Amer. Math. Soc. 81 (1975), 93-97

[2] D. Auckly, H. Jung Kim, P. Melvin and D. Ruberman, *Stable isotopy in four dimensions,* J London Math Soc (2015) 91 (2): 439-463.

[3] I. Belegradek, F. T. Farrell and V. Kapovitch, *Spaces of non-negatively curved manifolds and pseudoisotopies,* Preprint at arxiv:1501.03475.

[4] I. Belegradek and J. Hu, *Connectedness properties of the space of complete non-negatively curved planes,* Math. Ann. 362 (2015), no. 3-4, 1273-1286

[5] I. Belegradek, S. Kwasik and R. Schultz, *Moduli spaces of non-negative sectional curvature and non-unique souls,* J. Diff. Geom. 89 (2011), 49-86

[8] The s_1-curvature is also known as the *Einstein curvature* in that it is the quadratic form associated to an object called the Einstein tensor of the metric g, $\frac{s}{2}g - \mathrm{Ric}$.

[6] J.M. Boardman and R. M. Vogt *Homotopy Invariant Algebraic Structures on Topological Spaces*, Lecture Notes in Mathematics Number 347 (1973)

[7] B. Botvinnik, *Concordance and Isotopy of Metrics with Positive Scalar Curvature,* Geom. Funct. Anal. Vol. 23 (2013) 1099-1144

[8] B. Botvinnik, *Concordance and Isotopy of Metrics with Positive Scalar Curvature II,* arXiv:1604.07466

[9] B. Botvinnik, Johannes Ebert and Oscar Randal-Williams, *Infinite Loop Spaces and Positive Scalar Curvature,* Invent. math. (2017). doi:10.1007/s00222-017-0719-3

[10] B. Botvinnik, B. Hanke, T. Schick and M. Walsh, *Homotopy Groups of the Moduli Space of Metrics of Positive Scalar Curvature,* Geometry and Topology 14 (2010) 20472076

[11] B. Botvinnik, M. Walsh and D.J. Wraith *The Observer Moduli Space of Metrics of Positive Ricci Curvature* in preparation, Summary of Results contained in Oberwolfach Report No. 3/2017 on Mini Workshop: Spaces and Moduli Spaces of Riemannian Metrics

[12] B. Botvinnik and P.B. Gilkey, *The eta invariant and metrics of positive scalar curvature,* Math. Ann **302**, no. 3, 1995, 507–517.

[13] R. Carr, *Construction of manifolds of positive scalar curvature*, Trans. A.M.S. **307**, no. 1, May 1988, 63–74.

[14] J. Cerf, *La stratification naturelle des espaces de fonctions differentiables reelles et le theoreme de la pseudo-isotopie,* Publications mathematiques de l'I.H.E.S., tome 39 (1970), p. 5–173.

[15] V. Chernysh, *On the homotopy type of the space* $\mathcal{R}^+(M)$. Preprint: arXiv 0405235

[16] F. Coda Marques, *Deforming three-manifolds with positive scalar curvature,* Annals of Math. Pages 815-863, Volume 176 (2012), Issue 2

[17] D. Crowley and T. Schick, *The Gromoll filtration, KO-characteristic classes and metrics of positive scalar curvature*, Preprint: arXiv 1204.6474

[18] A. Dessai, S. Klaus, W. Tuschmann, *Nonconnected moduli spaces of nonnegative sectional curvature metrics on simply-connected manifolds*, Bull. London Math. Soc. (to appear).

[19] F. T. Farrell and W. C. Hsiang, *On the rational homotopy groups of the diffeomorphism groups of discs, spheres and aspherical manifolds,* Proc. Sympos. Pure Math. XXXII AMS, Providence R.I. (1978)

[20] F. T. Farrell and P. Ontaneda, *On the topology of the space of negatively curved metrics,* J. Diff. Geom. 86, (2010), no. 2, 273-301

[21] P. Gajer, *Riemannian metrics of positive scalar curvature on compact manifolds with boundary,* Ann. Global Anal. Geom. **5**, no. 3 (1987) 179–191.

[22] S. Galatius, I. Madsen, U. Tilmann and M. Weiss, *The homotopy type of the cobordism category* , Acta Mathematica, June 2009, Volume 202, Issue 2, pp 195239

[23] S. Galatius and O. Randal-Williams *Stable moduli spaces of high-dimensional manifolds,* Acta Math., 212 (2) :257377, 2014.

[24] S. Götte, *Morse Theory and higher torsion invariants I* Preprint, arXiv:math.DG/0111222

[25] M. Gromov and H. B. Lawson, Jr., *The classification of simply-connected manifolds of positive scalar curvature,* Ann. of Math. **111** (1980), 423–434.

[26] B. Hanke, T. Schick and W. Steimle, *The Space of Metrics of Positive Scalar Curvature,* Publ. Math. Inst. Hautes Etudes Sci. 120 (2014), 335-367

[27] A. Hatcher and J. Wagoner, *Pseudo-Isotopies of Compact Manifolds,* Societe Mathematique de France (1973).

[28] A. Hatcher, *Algebraic Topology,* Cambridge University Press (2002)

[29] N. Hitchin, *Harmonic spinors,* Adv. in Math., **14** (1974), 1–55.

[30] S. Hoelzel *Surgery stable curvature conditions,* Math. Ann. June 2016, Volume 365, Issue 1, pp 1347

[31] K. Igusa, *Higher Franz-Reidemeister Torsion,* AMS/IP, Studies in Advanced Mathematics, Vol. 31, (2002).

[32] K. Igusa, *The stability theorem for smooth pseudoisotopies,* K-Theory **2** (1988) 1–355

[33] D.D. Joyce, *Compact 8-manifoles with holonomy Spin(7),* Invent. Math. **123** (1996), no. 3, 507–552.

[34] V. Kapovitch, A. Petrunin and W. Tuschmann, *Non-negative pinching, moduli spaces and bundles with infinitely many souls,* J. Diff. Geom. 71 (2005) no. 3, 365-383

[35] M.A. Kervaire, J.W. Milnor, *Groups of homotopy spheres: I,* (1963) Annals of Mathematics. Princeton University Press. 77 (3): 504537.

[36] M. Kreck, S. Stolz, *Non-connected moduli spaces of positive sectional curvature metrics,* J. Am. Math, Soc. **6** (1993), 825-850.

[37] M. Labbi *Stability of the p-Curvature Positivity under Surgeries and Manifolds with Positive Einstein Tensor* , Annals of Global Analysis and Geometry Volume 15, Number 4 (1997), 299-312.

[38] J. Lohkamp, *The space of negative scalar curvature metrics*, Invent. Math **110** (1992), no. 2, 403–407.

[39] I. Madsen and M. Weiss, *The Stable Moduli Space of Riemann Surfaces: Mumford's Conjecture*, Ann. of Math. (2) 165 (2007), no. 3, 843-941

[40] M. Markl, S. Shnider, J. Stasheff, *Operads in Algebra, Topology and Physics*, Mathematical Surveys and Monographs A.M.S. Vol.96 (2002)

[41] P. May, *The Geometry of Iterated Loop Spaces*, Lecture Notes in Mathematics, Volume 271 (1972)

[42] J. Milnor, *Morse Theory*, Princeton Univ. Press (1963).

[43] R. S. Palais, *Homotopy theory of infinite dimensional manifolds*, Topology **5** (1966), 1–16.

[44] N. Perlmutter, *Cobordism Categories and Parametrized Morse Theory,* Prepint at arXiv:1703.01047

[45] J. Rosenberg and S. Stolz, *Metrics of positive scalar curvature and connections with surgery,* Surveys on Surgery Theory, Vol. 2, Ann. of Math. Studies 149, Princeton Univ. Press, 2001.

[46] D. Ruberman, *Positive scalar curvature, diffeomorphisms and the Seiberg-Witten invariants.* Geom. Topol. 5 (2001), 895–924

[47] W. Tuschmann, D.J. Wraith, *Moduli Spaces of Riemannian Metrics,* Oberwolfach Seminars Volume 46 (2015), Birkhäuser

[48] M. Walsh, *Metrics of positive scalar curvature and generalised Morse functions, Part 1*,
Mem. Am. Math. Soc 209 (2011), no. 983

[49] M. Walsh, *Metrics of positive scalar curvature and generalised Morse functions, Part 2*,
Trans. Am. Math. Soc. 366 (2014), 150

[50] M. Walsh, *Cobordism Invariance of the Homotopy Type of the Space of Positive Scalar Curvature Metrics*, Proc. Am. Math. 141 (2013), no. 7, 2475-2484

[51] M. Walsh, *H-spaces, loop spaces and positive scalar curvature* Geom. Top. 18, (2014), no. 4, 2189-2243

[52] M. Walsh, *The space of positive scalar curvature metrics on a manifold with boundary* Preprint at arxiv.org

[53] M. Wang, W. Ziller, *Einsten metrics with positive scalar curvature* Curvature and Topology of Riemannian Manifolds, Proceedings, Lecture Notes in Math. 1021, Springer-Verlag (1986), 319-336

[54] J. Wolfson, *Manifolds with k-positive Ricci Curvature*, Preprint: www.math.msu.edu/ wolfson/k-positive-Ricci.pdf

[55] D. J. Wraith, *Exotic Spheres with Positive Ricci Curvature,* J. Diff. Geom. 45 (1997), 638-649.

[56] D. J. Wraith, *Surgery on Ricci positive manifolds,* J. reine angew. Math. 501 (1998), 99-113.

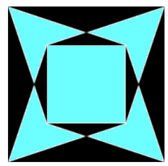
[57] D. J. Wraith, *New connected sums with positive Ricci curvature,* Ann. Glob. Anal. Geom. 32 (2007), 343-360

[58] D. J. Wraith, *On the moduli space of positive Ricci curvature metrics on homotopy spheres* Geom. Top. 15 (2011), 1983-2015

Mark Walsh is a lecturer in Mathematics at Maynooth University. Formerly, he was associate professor of Mathematics at Wichita State University, Kansas, USA. His research interests are in Differential Geometry, in particular the relationship between curvature and topology. He obtained his PhD from the University of Oregon and a masters and undergraduate degrees from Maynooth University.

(M. Walsh) Department of Mathematics and Statistics, Maynooth University, Maynooth, Co. Kildare, Ireland

E-mail address: Mark.Walsh@mu.ie



Irish Math. Soc. Bulletin
Number 81, Summer 2018, 97–100
ISSN 0791-5578

Jean-Pierre Serre: Finite Groups: an introduction, International Press, 2016. ISBN:978-1-57146-327-2, USD 79, 178 pp.

REVIEWED BY JOHN MURRAY

In the preface Serre tells us that he based Finite Groups on handwritten notes from a course he taught in the late 1970's, available at arXiv:math/0503154 (in French), recently translated, revised and expanded. As one might expect from the author, the book is beautifully written. Results are either proved in full, or the gaps highlighted and referenced. However it is certainly not an introduction to group theory and might be more accurately titled 'Finite Groups: an apprenticeship'. The text proceeds at break-neck speed and important results and methods are relegated to the exercises. As well as being of interest to graduate students, non-specialists can dip into the book to learn about particular topics. Specialists will appreciate the efficient development of the theory and the elegance of the proofs.

Serre does not define what a group is. Instead, he begins with the notion of a G-set, which reflects the fact that every group is the group of symmetries of some set. By page 6 he has defined normal and characteristic subgroup, simple group (a key topic of interest throughout the book) and proved the Jordan-Hölder theorem on the uniqueness of composition factors. His choice of results of Goursat and Ribet in Section 1.4 is a bit idiosyncratic. The exercises indicate the sophistication expected of the reader: they cover the n-transitivity of G-sets and showing that the alternating groups A_n are simple, for $n \geq 5$.

Chapter 2 gives a conventional but efficient treatment of Sylow's three theorems. In addition to Burnside's fusion theorem, Serre proves Alperin fusion theorem: that the conjugation of p-elements is p-locally controlled (i.e. in the normalizers of non-trivial p-subgroups).

1991 *Mathematics Subject Classification.* 20-01.
Key words and phrases. Group.
Received on 13-3-2018.

This anticipates the current theory of Fusion Systems. Generalizations of Sylow subgroups are described, including Hall subgroups of solvable groups (the subject of a later chapter), tori of compact Lie groups and Borel subgroups of linear algebraic groups. The exercise include results on normal p-complements, and determining the groups of order pq and the Sylow subgroups of symmetric groups.

The next chapter deals with solvable and nilpotent groups. It is disappointing that the structure theorem for finite abelian groups, although stated, is not proved. Applications of solvable groups to ruler and compass constructions, Galois theory and proving that $\mathbb{C}$ is algebraically closed are described. Hall-Burnside's theorem on p'-groups of automorphisms of p-groups is also covered. J. Thompson proved that a finite group is solvable if and only if all its two generator subgroups are solvable. The proof (omitted!) reduces to checking 'minimal' simple groups. This prefigures the use of the classification of finite simple groups to prove results about finite groups. Some of the exercises are strenuous: the reader is asked to prove Iwasawa's simplicity theorem and use it to prove the simplicity of (most) finite projective special linear groups. They also cover such varied topics as supersolvable groups, towers of quadratic field extensions and torsion in nilpotent groups.

Serre covers group cohomology and its application to the existence and uniqueness of group extensions (both abelian and non-abelian) with commendable clarity and completeness in Chapter 4. In particular he proves Zassenhaus theorem on the existence and uniqueness of complements. In one section he shows that every representation of a finite group over the field with p-elements lifts to a respresentation over the field of p-adic numbers. The exercises in this chapter include a useful exploration of the group-theoretic interpretations of low degree cohomology groups.

P. Hall proved that if G is a finite solvable group and π is a set of primes, then G has a subgroup of order $|G|_\pi$. Moreover all Hall π-subgroups of are G-conjugate and every π-subgroup of G is contained in a Hall π-subgroup. Proving this is the main task of Chapter 5.

A finite group is said to be Frobenius if it acts transitively on a set so that every non-trivial group element fixes at most one element of the set. For a Frobenius group G, the stabilizer H of a point is called a Frobenius Complement. G. Frobenius used character induction to show that the identity and the set of elements of G which lie in no conjugate of H forms a normal complement N to H, now called a Frobenius Kernel. Serre defers proving this to a later chapter, and instead focuses in Chapter 6 on proving interesting properties of Frobenius groups. He gives elementary proofs that N is nilpotent if it possesses a fixed-point-free automorphism of order 2 or 3 and discusses Thompson's theorem that N is always nilpotent. Frobenius groups are a rich source of interesting examples, and this is reflected in the exercises.

Given a group G and a finite index subgroup H, Transfer is a homomorphism from G to the abelianization of H. One important application is Hall's Focal Subgroup Theorem. This characterises the group generated by all commutators in a finite group which lie in a fixed Sylow subgroup of the group (Exercise 10 in chapter 7). Transfer also provides the 'correct' framework to prove Gauss's lemma in quadratic reciprocity. Serre describes other applications of transfer to number theory and topology and demonstrates its usefulness in classifying 'small' finite simple groups.

Each complex representation of a finite or compact group has an associated character. This is a complex valued function on the group which is constant on conjugacy classes. The theory of characters was used in the Odd Order paper of Feit-Thompson, and in conjunction with modular characters, in the classification of the simple groups with small 2-rank. Therefore it was instrumental in kick-starting the classification of all finite simple groups. Chapter 8 runs briskly through the existence of invariant Hermitian forms and Maschke's theorem, orthogonality relations, Wedderburn's theorem on the structure of the group algebra and applications of integrality properties such as Burnside's $p^a q^b$ theorem. But there is much more: induction, the existence of Frobenius Kernels, Adams operations on characters, Galois conjugation of characters and fields of values of characters. I particularly liked Serre's thorough treatment of Frobenius-Schur indicators, which includes using the theory of positive definite hermitian forms. The 34 exercises at the end of the chapter indicate Serre's deep interest in characters.

The penultimate Chapter 9 deals with the problem of bounding the order of a finite matrix group. This is not standard textbook material and the approach strays well beyond mere algebra. Minkowski demonstrated that the order of a finite group of rational matrices is bounded by an explicit function in the size n of the matrices. Later Jordan proved that a finite group of complex matrices contains an abelian normal subgroup whose index is bounded by a function of n. The chapter displays the breath of Serre's interests in algebra and geometry.

The final chapter is called 'Small Groups'. This walks the reader through an array of interesting examples and outlines many of the exceptional isomorphisms between low order almost-simple groups, including $\mathcal{S}_4 \cong \mathrm{PGL}_2(3)$, $\mathcal{A}_5 \cong \mathrm{SL}_2(4)$, $\mathcal{S}_5 \cong \mathrm{PGL}_2(5)$, $2.\mathcal{A}_5 \cong \mathrm{SL}_2(5)$, $\mathrm{SL}_3(2) \cong \mathrm{PSL}_2(7)$, $\mathcal{A}_6 \cong \mathrm{PSL}_2(9)$, $\mathcal{S}_6 \cong \mathrm{Sp}_4(2)$ and (of course) $\mathcal{A}_8 \cong \mathrm{SO}_6(2) \cong \mathrm{SL}_4(2)$. There is also a section on embeddings of $\mathcal{A}_4$, $\mathcal{S}_4$ and $\mathcal{A}_5$ in $\mathrm{PGL}_2(q)$, anticipating modular representation theory.

Serre includes an extensive bibliography of 41 books and 40 academic papers on group theory, representation theory, number theory and homological algebra. In addition, there is a list of books dealing with relevant mathematical 'Topics' and a handy index of names. This slim volume will sit handsomely on any mathematician's bookshelf.

Mathematics and Statistics, Maynooth University
E-mail address: `john.murray@mu.ie`

Irish Math. Soc. Bulletin
Number 81, Summer 2018, 101–104
ISSN 0791-5578

PROBLEMS

IAN SHORT

Problems

The first problem this issue was contributed by Peter Danchev of the Institute of Mathematics & Informatics of the Bulgarian Academy of Sciences.

Problem 81.1. Find a homogenous linear ordinary differential equation of order two that is satisfied by the function

$$y(x) = \int_0^\pi \sin(x\cos t)\,dt.$$

The second problem was suggested by Finbarr Holland of University College Cork. To state this problem, we use the standard notation

$$f(x) \sim g(x) \quad \text{as} \quad x \to \infty,$$

where f and g are positive functions, to mean that

$$\frac{f(x)}{g(x)} \to 1 \quad \text{as} \quad x \to \infty.$$

Problem 81.2. Let

$$a_n = \sum_{k=0}^{n} \binom{n}{k}^2, \quad n = 0, 1, 2, \dots.$$

Prove that

$$\sum_{n=0}^{\infty} \frac{a_n x^n}{(n!)^2} \sim \frac{e^{4\sqrt{x}}}{4\pi\sqrt{x}} \quad \text{as} \quad x \to \infty.$$

I learned the last problem from a friend recently.

Problem 81.3. Find a function $f\colon \mathbb{R} \longrightarrow \mathbb{R}$ such that the restriction of f to any open interval I is a surjective function from I to $\mathbb{R}$.

Received on 10-3-2011; revised 7-4-11.

Solutions

Here are solutions to the problems from *Bulletin* Number 79. The first problem was solved by Henry Ricardo of the Westchester Area Math Circle, New York, USA, as well as the North Kildare Mathematics Problem Club and the proposer, Peter Danchev. The solution we give is an amalgamation of these solutions.

Problem 79.1. Suppose that k and n are positive integers with $1 \leqslant k \leqslant n$. Find the largest integer m such that the binomial coefficient $\binom{2^n}{k}$ is divisible by 2^m.

Solution 79.1. We have

$$\binom{2^n}{k} = \frac{2^n}{k}\binom{2^n-1}{k-1}.$$

Observe that, for $1 \leqslant a < 2^n$, the highest power of 2 that is a factor of a is equal to the highest power of 2 that is a factor of $2^n - a$. From this we see that the integer

$$\binom{2^n-1}{k-1} = \frac{(2^n-1)(2^n-2)\cdots(2^n-(k-1))}{1\cdot 2\cdot\cdots\cdot(k-1)}$$

is odd. Let us write $k = 2^r s$, where r is a nonnegative integer and s is an odd positive integer. It follows that $\binom{2^n}{k}$ is divisible by 2^{n-r}, and it is divisible by no higher power of 2, so $m = n - r$. □

The second problem was solved by Finbarr Holland, Ángel Plaza of Universidad de Las Palmas de Gran Canaria, Spain, Henry Ricardo, the North Kildare Mathematics Problem Club, and the proposer, Prithwijit De of the Homi Bhabha Centre for Science Education, Mumbai, India. We present Henry Ricardo's solution. The other solutions were similar, with some differences in how the trigonometric integral was evaluated.

Problem 79.2. Let f be a function that is continuous on the interval $[0, \pi/2]$ and that satisfies $f(x)+f(\pi/2-x) = 1$ for each x in $[0, \pi/2]$. Evaluate the integral

$$I = \int_0^{\pi/2} \frac{f(x)}{(\sin^3 x + \cos^3 x)^2}\,dx.$$

Solution 79.2. By substituting $u = \pi/2 - x$, we see that

$$I = \int_0^{\pi/2} \frac{f(\pi/2 - u)}{\left(\cos^3 u + \sin^3 u\right)^2}\, du = \int_0^{\pi/2} \frac{1 - f(u)}{\left(\cos^3 u + \sin^3 u\right)^2}\, du.$$

Hence

$$I = \frac{1}{2} \int_0^{\pi/2} \frac{1}{\left(\cos^3 u + \sin^3 u\right)^2}\, du.$$

Next let $t = \tan u$. Then

$$I = \frac{1}{2} \int_0^{\infty} \frac{t^4 + 2t^2 + 1}{1 + 2t^3 + t^6}\, dt.$$

Expanding the integrand using partial fractions, we obtain

$$I = \frac{5}{18} \int_0^{\infty} \frac{dt}{t^2 - t + 1} + \frac{4}{18} \int_0^{\infty} \frac{dt}{(t+1)^2} + \frac{1}{6} \int_0^{\infty} \frac{t}{(t^2 - t + 1)^2}\, dt.$$

These are standard integrals, which can easily be evaluated to give

$$I = \left(\frac{10}{81}\pi\sqrt{3}\right) + \frac{2}{9} + \left(\frac{1}{9} + \frac{2}{81}\pi\sqrt{3}\right) = \frac{4}{27}\pi\sqrt{3} + \frac{1}{3}. \qquad \square$$

The third problem was solved by Yagub Aliyev of ADA University, Azerbaijan, Henry Ricardo, Ángel Plaza, and the North Kildare Mathematics Problem Club. All solutions used either induction or quoted known identities involving sums of powers. The solution we present is similar to that submitted by Ángel Plaza.

Problem 79.3. Prove that, for any positive integer n,

$$(1^5 + \cdots + n^5) + (1^7 + \cdots + n^7) = 2(1 + \cdots + n)^4.$$

Solution 79.3. In brief, let L_n and R_n denote the expressions on the left-hand side and right-hand side, respectively, for $n = 0, 1, 2, \ldots$. Then

$$L_{n+1} - L_n = (n+1)^5 + (n+1)^7 = (n+1)^5(n^2 + 2n + 2).$$

Another calculation shows that $R_{n+1} - R_n = (n+1)^5(n^2 + 2n + 2)$. Since $L_0 = R_0 = 2$, we deduce that

$$L_m = 2 + \sum_{n=0}^{m-1} (L_{n+1} - L_n) = 2 + \sum_{n=0}^{m-1} (R_{n+1} - R_n) = R_m,$$

for $m = 1, 2, \ldots$, as required. $\square$

We invite readers to submit problems and solutions. Please email submissions to imsproblems@gmail.com in any format (we prefer Latex). Submissions for the summer Bulletin should arrive before the end of April, and submissions for the winter Bulletin should arrive by October. The solution to a problem is published two issues after the issue in which the problem first appeared. Please include solutions to any problems you submit, if you have them.

School of Mathematics and Statistics, The Open University, Milton Keynes MK7 6AA, United Kingdom

Books from Logic Press

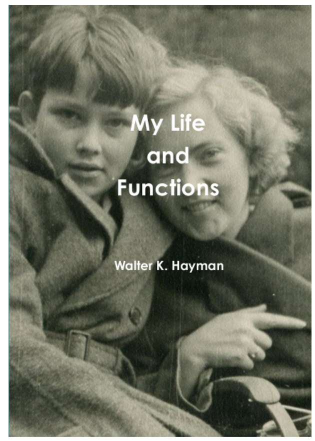
www.logicpress.ie

Walter K. Hayman
My Life and Functions.
Hardcover: 978-1-326-03224-1.
Paperback: 978-1-326-03020-9.

Paul Barry
Riordan Arrays: a Primer
Hardcover: 978-1-326-85523-9.

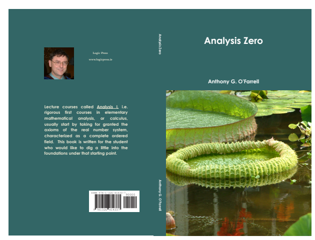
Maria Roginskaya
Advanced Basics of Geometric Measure Theory
Paperback: 978-1-326-36743-5

Stephen M. Buckley
Fractals and Real Dynamical Systems (in Irish or English)



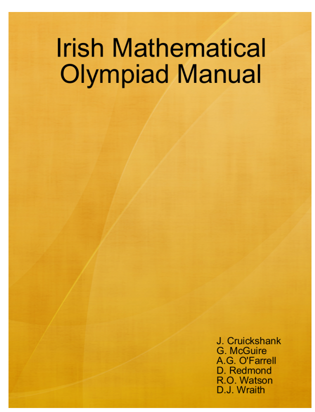
Anthony G. O'Farrell
Analysis Zero
Hardcover: 978-1-326-41433-7.
Paperback: 978-1-326-41581-5.

Analysis One
Hardcover: 978-0-244-02530-4.



Irish Mathematical Olympiad Manual
by J. Cruickshank, G. McGuire, A.G. O'Farrell, D. Redmond, R.O. Watson and D.J. Wraith.
Paperback: 978-095442698-9

Notaí an Bhráthar Mac Craith le Antóin Ó Fearghail.
Paperback: 1-8991-6223-2.



Instructions to Authors

Papers should be submitted by email to the address:

`ims.bulletin@gmail.com`

In the first instance, authors may submit a pdf version of their paper. Other formats such as MS/Word or RTF are not acceptable. The *Bulletin* is typeset using PDF files produced from LaTeX source; therefore, authors must be ready to supply LaTeX source files (and any ancillary graphics or other files needed) when their paper is accepted. Source files should be suitable for processing using `pdflatex`.

Once a paper is accepted in final form, the author(s) will be responsible for producing an output according to the *Bulletin's* standard layout. Standard template files for articles, abstracts and reviews, and the necessary class and style files may be downloaded from the IMS website, or obtained from the editor in case of difficulty.

Since normally no proofs are sent out before publication, it is the author's responsibility to check carefully for any misprints or other errors.

After publication, authors will be sent a free copy of the issue in which their article(s) appeared as well as a pdf file containing their article(s). The Society grants authors free use of *this* pdf file; hence they may post it on personal web sites or electronic archives. They may reuse the content in other publications, but should explicitly acknowledge that this is being done with the permission of the IMS.